

Analysis of Regional Financial Risk Factors under Random Forest

Yishan Liu *, Yuhan Zhang

Department of International Economics, China Foreign Affairs University, Beijing, China

* Corresponding Author Email: 15510685598@163.com

Abstract. Objectively understanding the current state of China's financial risks and their external drivers is a crucial guarantee for maintaining financial security and stability, and promoting high-quality financial development. This paper breaks through the limitations of traditional indicators in monitoring shadow banking, cross-border capital, and the marginal effects of debt risk. It employs the entropy weight method to construct a comprehensive regional risk assessment framework covering four sectors, and empirically analyzes the main external influencing factors of regional financial risks based on the random forest model. The research indicates that: (1) China's regional financial risks exhibit significant spatial differentiation, with major influencing factors forming a four-tier risk driver system. (2) The risk sources in different strategic regions stem from the reshaping of regional economic structures due to national strategic positioning. This is manifested in three regional dynamics: the Beijing-Tianjin-Hebei region demonstrates dual constraints through state-owned enterprises' dominance in major project investments; the Yangtze River Delta, as an export-oriented economic hub, exhibits heightened susceptibility to international market fluctuations; while the Yangtze River Economic Belt illustrates the critical equilibrium between large-scale infrastructure investment and local fiscal capacity in cross-regional coordination. (3) Regional risk governance needs to emphasize the asymmetry of dynamic evolution, providing a scientific basis for differentiated policy measures.

Keywords: Random forest model, Entropy weight method, Regional financial risks, regional heterogeneity.

1. Introduction

The restructuring of global financial governance has precipitated regional financial risks characterized by cross-border spillovers and multifaceted complexity. Internationally, the spillover effects of monetary policies in developed economies exacerbate systemic fragility, geopolitical tensions disrupt supply chains in commodity markets coupled with capital outflows from emerging markets, alongside coexisting sovereign debt risks and policy divergence pressures. Domestically, heightened interregional risk correlations demand the improvement of a 'comprehensive monitoring indicator framework', targeted controls for real estate debt, local fiscal stress, and small-to-medium financial institution vulnerabilities, as well as building cross-regional coordination mechanisms for critical economic zones. This study aims to identify major influencing factors in different regions, designing differentiated prevention and control schemes to reinforce China's financial security infrastructure for the new development pattern.

In recent years, research in this field has gradually expanded. Early scholars mostly used traditional econometric methods. Cao Y. and Cai Z. (2013) revealed the cross-domain bidirectional contagion effect of China's regional risks during the financial crisis based on VAR models [1]. With the popularization of ML algorithms, Giudici et al. (2020) described how correlation network models with multivariate network structure measure systemic risk in financial networks [3]. Zhang L. and Zhang S. (2022) emphasized the robustness of random forest in regional heterogeneous data analysis [16]. Xu (2023) selected multiple machine learning methods to identify factors affecting China's systemic financial risk index, verifying the high accuracy of integrated learning and deep learning [13]. Tang et al. (2024) constructed a dual-view regional financial risk warning system, embedding the ReliefF algorithm into an improved RF model for feature selection [12].

Theoretical academia mainly focuses on traditional financial risk management or financial system risk control. However, the existing ML research often focuses on single risk types and lacks a scientific classification of regional risk levels. Moreover, traditional indicators like loan-to-deposit ratio and non-performing loan rate struggle to capture the risk spillover effects of shadow banking and cross-border capital flows. Compared to existing research, this paper incorporates innovative indicators covering four sectors such as foreign exchange loan-to-deposit ratio, corporate bond increment and household debt leverage ratio into the panel analysis framework. Through the integrated application of entropy weight (EW) method and random forest (RF) model, it promotes a paradigm shift in regulatory technology from linear analysis to dynamic weight measurement. At the macro level, it quantifies the formation mechanism of the 'real estate-foreign exchange' risk resonance channel, while at the micro level, it verifies that the financial health of market entity is the leading indicator of risk transmission.

2. Regional financial risk concept and index system

2.1. Concept of regional financial risk

This study defines regional financial risk as potential losses arising from individual institutions or specific financial activities within a defined geographic area. The analytical framework integrates three core risk dimensions - credit risk, liquidity risk, and market risk - to establish cross-sectoral and cross-market risk interaction pathways through critical financial subsystems.

2.2. Regional financial risk measurement indicator system

Based on the theoretical framework of systemic risk, the index system evaluates four institutional sectors: financial institutions, government entities, corporate enterprises, and households [11].

The financial institution sector indicators cover four dimensions of monitoring: the robustness of the banking system, the development level of the bond market, the fluctuation pressure of the capital market and the foreign exchange liquidity risk.

(1) The regional Bank_liquidity_ratio and Non_performing_loan ratio are respectively used to depict banking system soundness from the perspective of short-term solvency and long-term asset quality.

(2) Bond_market_development and Stock_market_development indicators monitor the systemic accumulation of debt default risk and the bubble tendency of market valuations deviating from fundamentals.

(3) The introduction of the Foreign_exchange_liquidity ratio monitors currency mismatch risks triggered by abnormal cross-border capital flows.

(4) Premium_depth ratio characterizes the corresponding insurance industry risk buffer capacity, and Trust_loan_increment ratio warns of the credit overheating risks.

Enterprise sector indicators select the Asset_liability_ratio of listed companies to reflect corporate debt risk, the Industrial_loss rate to reflect corporate operating risk, the non-financial Corporate_bonds_increment rate to reflect corporate debt scale expansion and solvency, and the Real_estate_growth rate to reflect industry concentration risk.

Government sector indicators set the Gov_debt and Fiscal_gap rate to reflect the revenue-expenditure gap when the government allocates social resources. The quadratic design of the Fiscal_gap ratio effectively quantifies the marginal increasing effect of local government debt risk.

Household sector uses Household_leverage rate to reflect the accumulation of household leverage risk.

In summary, the regional financial risk index parameters constructed in this paper are shown in table 1:

Table 1. Regional financial risk composite index table

Primary Indicator	Secondary Indicator	Tertiary Indicator	Calculation Method	Pos/Neg Effect	Ref
Financial Institutions	Market Risk	Stock_market_development	Total Stock Market Value / GDP	+	[6]
		Bond_market_development	Total Bond Financing / GDP	+	[4]
	Credit Risk	Premium_depth	Total Bond Financing / GDP	-	[7]
		Non_performing_loan	Non-performing Loans / Total Loans	+	[7]
		Trust_loan_increment	Trust Loan Increment	+	[10]
	Liquidity Risk	Bank_liquidity_ratio	Liquid Assets / Liquid Liabilities	+	[10]
		Foreign_exchange_liquidity	Financial Institution FX Loans / Financial Institution FX Deposits	+	[2]
Non-financial Enterprises	Credit Risk	Asset_liability_ratio	Listed Company Asset-liability Ratio	+	[10]
		Industrial_loss	Enterprise Loss Amount / Total Assets	+	[7]
		Corporate_bonds_Increment	Corporate Bond Increment	+	[5]
	Market Risk	Real_estate_growth	Δ Real Estate Investment / Previous Period Real Estate Investment	+	[8]
Governments	Credit Risk	Gov_debt	Government Bond Issuance / GDP	+	[7]
		Fiscal_gap	$(\text{Fiscal Gap} / \text{GDP})^2$	+	[10]
Households	Credit Risk	Household_leverage	Household Debt Leverage Ratio	+	[9]

2.3. Influencing factors of regional financial risk

Firstly, macro factors mainly refer to the operating conditions of the external macroeconomic environment of the region, specifically covering three levels of national economic operation quality, price stability and investment structure:

(1) Economic fundamentals support uses actual GDP_growth to measure regional economic operation.

(2) CPI index serves as a price stability mechanism to monitor regional inflation.

(3) Foreign_trade_level index measures the degree of financial openness, the Fix_assets_investment ratio warns of overcapacity and local debt accumulation, and the commercial Housing_sales_price_index measures real estate bubble risk.

Secondly, the meso factors mainly refer to the intervention of local governments in financial activities. The Gov_supervision index is used to measure the intensity of government supervision.

Third, the micro factors mainly include characteristics of the household and enterprise sectors. The Income_growth rate reflects the solvency of residents and measures household debt default risk. The Unemployment_rate index directly reflects the pressure of economic recession. The Corporate_liquidity_ratio reflects the short-term solvency of the enterprise.

Table 2. Influencing factors of regional financial risk

Impact Dimensions	Driving Factors	Measurement Metrics	Variable Selection	Variable Description	Effect
Macro Level	National Economy	Economic Operating Status	GDP_growth	$\Delta \text{GDP} / \text{Previous Period GDP}$	-
	Consumption Level	Inflation	CPI	CPI Index	+
	Trade and investment	Financial Openness	Foreign_trade_level	$(\text{Import} + \text{Export}) / \text{GDP}$	-
		Fixed Assets	Fix_assets_investment	Fixed Assets Investment / GDP	+
		Real Estate Market Hotness	Housing_sales_price_index	Commercial Housing Sales / Sales Area	+
Meso Level	Local Government Intervention	Financial Regulatory Intensity	Gov_supervision	Fiscal Financial Regulatory Expenditure / Total Fiscal Expenditure	-
Micro Level	Households	Living Standard	Income_growth	Per Capita Income Growth Rate	-
		Unemployment situation	Unemployment_rate	Unemployed Population / Labor Force	+
	Enterprise	Corporate Liquidity Ratio	Corporate_liquidity_ratio	Liquid Assets / Liquid Liabilities	-

3. Empirical study

3.1. Selection of study area

This study selects 31 provinces (autonomous regions, municipalities) in mainland China as basic research units, excluding Hong Kong, Macao, and Taiwan. Referring to Zhou & He, 2024, Chinese provinces are divided into ‘East-Central-West’ three major regions, with the Northeastern region included in the Eastern region [17]. Three major regional strategic clusters are also delineated, as shown in Figure 1.



Figure 1. Study area and boundaries of sub regions

Note: Based on the standard map service system of the Ministry of Natural Resources, the base map review number is GS (2024)0650. The base map boundaries are unmodified.

3.2. Data collection

This paper is based on provincial panel data from 2013 to 2023, using 14 core indicators to construct a comprehensive regional financial risk index. Simultaneously, 9 influencing factor indicators were selected from macro, meso, and micro levels. Missing values were imputed using interpolation.

Data sources include the National Bureau of Statistics, EPS DATA, and Wind Financial Terminal. Core financial regulatory indicators are taken from publicly disclosed information by the People's Bank of China and the China Securities Regulatory Commission. Capital market data is obtained through the official websites of the Shanghai and Shenzhen Stock Exchanges and industry reports from the China Association for Public Companies.

Descriptive statistics are shown in Table 3:

Table 3. Statistical table of index characteristic value

Indicator	Max	Min	Mean	SD	Mid
Stock_market_development	56.24	0.01	1.39	3.49	0.67
Bond_market_development	117.77	0.18	7.63	11.33	4.99
Premium_depth	7.56	1.40	3.71	1.14	3.52
Non_performing_loan	11.90	0.23	1.91	1.39	1.60
Trust_loan_increment	6561.00	-4428.70	111.49	1401.48	0.00
Bank_liquidity_ratio	1.25	0.37	0.82	0.16	0.81
Foreign_exchange_liquidity	17.45	0.04	1.91	2.48	1.09
Asset_liability_ratio	80.00	47.20	65.00	6.54	65.70
Industrial_loss	13.08	0.21	1.07	0.88	0.87
Corporate_bonds_Increment	22191.80	-13221.50	2028.02	3672.25	929.80
Real_estate_growth	9.98	-1.29	0.56	0.77	0.56
Gov_debt	22.96	0.05	6.96	4.07	6.04
Fiscal_gap	122.08	1.34	17.79	20.01	12.64
Household_leverage	120.51	18.32	52.78	19.03	48.69

3.3. Identification and mechanism of influencing factors

This paper uses the EW method to calculate the objective weights of 14 indicators and construct a regional financial risk index.

Regional financial risk is affected by the combined influence of multiple external factors, while traditional regression analysis methods cannot provide effective explanations under numerous factors. Referring to Rahman Ali (2021) and Zilin Wang (2022), this paper uses a RF model to rank the importance of multi-dimensional drivers [14-15]. RF integrates Bootstrap Bagging and Random Subspace dual randomness, and has unique advantages in high-dimensional data processing, non-linear relationship capture, prediction accuracy, and model robustness by constructing the 'collective wisdom' of multiple decision trees.

The calculation of factor importance mainly involves three steps:

STEP1: Sample splitting. Randomly divide the original sample into N parts, with one part as the model's test set and the remaining parts as the training set for the training model.

STEP2: Model parameter optimization and adjustment mainly for the Max_features, Max_depth, and Min_samples_leaf.

STEP3: Use the full sample data to estimate the model under the optimal parameter combination, and finally obtain the importance of each factor. R^2 and MAPE were used to evaluate the model's goodness of fit.

This paper uses Python language to implement the construction of the RF model, and based on the sample area division, derives the importance ranking of influencing factors for China's "East-Central-West" regions and major regional strategic clusters.

4. Regional heterogeneity analysis based on RF

The provincial panel data analysis spanning 2013-2023 (Column 1 of Table 5) reveals Real_estate_growth as the predominant risk driver, accounting for 55.66% of explanatory power, a pattern consistent with China's financial vulnerability shaped by high leverage ratios and asset price dependencies. Foreign_trade_level (10.19%), Corporate_liquidity_ratio (10.03%), and Unemployment_rate (8.25%) constitute the second-tier influencing factors, collectively illustrating systemic constraints in liquidity management within China's export-oriented economy.

Regional decomposition (Columns 2-5) demonstrates marked spatial variations. Eastern provinces exhibit the Corporate_liquidity_ratio (30.59%) as the primary determinant, reflecting advanced regions' reliance on balance sheet optimization. This is followed by Foreign_trade_level (17.42%), Real_estate_growth (11.91%), and Unemployment_rate (10.63%). Northeastern region displays parity between Corporate_liquidity_ratio (28.30%) and Income_growth (27.48%), indicating transitional risks during industrial upgrading. In contrast, central provinces remain investment-driven, with Fix_assets_investment (26.36%) predominating, complemented by Real_estate_growth (16.89%) and Foreign_trade_level (16.18%), forming an 'infrastructure-property-exports' tripartite risk transmission chain. Western region exhibits extreme Real_estate_growth (65.21%), highlighting the risk of regional economic monolithicity. While other factors demonstrate marginal influence, ranging from 2.58% to 7.67%, signaling underdeveloped risk mitigation channels.

Regional major strategic clusters (Columns 6-8) share common drivers: Foreign_trade_level, Fix_assets_investment, Corporate_liquidity_ratio, and Gov_supervision. Yet these contributing factors demonstrate divergent ranking significance due to north-south geographical disparities and land-sea economic configuration variances. BTH cluster prioritizes Corporate_liquidity_ratio (34.27%) and Fix_assets_investment (32.31%), reflecting the dual characteristics of 'concentration of central enterprise headquarters - intensive major project investment' during the construction of the capital economic circle. Among secondary determinants, Foreign_trade_level (11.16%), Income_growth (7.57%), and Gov_supervision (5.59%) exhibit comparable weights, reflecting the imbalanced effects of industrial specialization restructuring, demographic mobility patterns, and consumption market interconnectivity imbalances. The Yangtze River Delta is significantly affected by the dispersed

influence of multiple factors, with Foreign_trade_level (20.56%), Unemployment_rate (19.53%), and CPI (16.10%) forming a risk trifecta, exposing external demand sensitivity. Contrastingly, the Yangtze Economic Belt is dominated by Fix_assets_investment (21.02%), amplified by Gov_supervision (19.42%), Corporate_liquidity_ratio (19.09%), and Real_estate_growth (13.22%). This stems from the Yangtze River Delta acting as an important economic corridor spanning the ‘East-Central-West’ regions, revealing debt transmission risks along the intensive construction of coastal port clusters and crossregional major projects.

Comparing the differences among the three major strategic clusters, the state-owned enterprise dominance in BTH region diverges from the diversity of market entities in the Yangtze River Delta, while the Yangtze Economic Belt emphasizes active government regulatory intervention and coordinated development across the entire basin, due to differences in national strategic positioning despite shared geographical lineage with the Delta. These spatial heterogeneities underscore the necessity for regionally differentiated risk governance framework.

Table 4. Importance of Influencing Factors for Regional Financial Risk

Indicator	Full Sample	Regional Division				Major regional strategic clusters		
		Eastern Region		Central Region	Western Region	BTH Region	Yangtze River delta	Yangtze River Economic belt
		Other	Northeastern Region					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP_growth	0.043	0.064	0.086	0.104	0.052	0.031	0.127	0.055
CPI	0.031	0.045	0.041	0.071	0.030	0.006	0.161	0.026
Foreign_trade_level	0.102	0.174	0.044	0.162	0.077	0.112	0.206	0.066
Fix_assets_investment	0.034	0.023	0.056	0.264	0.026	0.323	0.057	0.210
Housing_sales_price_index	0.557	0.119	0.144	0.169	0.652	0.021	0.072	0.132
Gov_supervision	0.024	0.077	0.065	0.093	0.029	0.056	0.088	0.194
Income_growth	0.027	0.086	0.275	0.047	0.036	0.076	0.022	0.044
Unemployment_rate	0.083	0.106	0.005	0.048	0.035	0.034	0.195	0.081
Corporate_liquidity_ratio	0.100	0.306	0.283	0.042	0.063	0.343	0.071	0.191
R ²	0.255	0.003	0.069	0.825	0.554	0.376	0.311	0.127
MAPE	0.078	0.112	0.023	0.016	0.067	0.081	0.243	0.112
N	341	143	33	66	132	33	44	121

5. Conclusion

This study utilizes panel data from 31 Chinese provinces (2013–2023) to construct a regional financial risk index system through EW, with empirical analysis conducted via RF models to identify key external drivers. Key findings reveal:

First, China’s regional financial risks exhibit multi-layered characteristics. Nationally, the main influencing factors constitute a four-tier driver system: the real estate market as the foundation layer, foreign trade and corporate liquidity as the transmission layer, unemployment rate and other social indicators as the buffer layer, and government regulation as the adjustment layer. Among them, the real estate market, as the core source of systemic risk, tightly couples asset prices with credit expansion in China's economic development. Regionally, eastern areas face complex risk interactions between income growth and corporate debt servicing during industrial transitions, central regions show cyclical dependencies on infrastructure-real estate linkages, while western regions reveal structural fragility from overreliance on single industries.

Second, regional major strategic clusters demonstrate risk profiles reshaped by national strategic positioning. The BTH region demonstrates dual constraints between major project investment and corporate operational efficiency, reflecting tensions between administrative resource allocation and market-oriented processes. The Yangtze River Delta, as an export-oriented economy, exhibits heightened sensitivity to international market demand fluctuations and price transmission mechanisms. The Yangtze River Economic Belt highlights debt transmission risks from cross-regional infrastructure

coordination, exposing the dynamic balancing challenge between large-scale infrastructure investments and local fiscal capacities.

Third, regional risk governance requires dynamic asymmetry recognition. Market self-correction dominates risk mitigation in eastern regions, whereas central and western regions remain reliant on policy interventions. Transitional economies face unique risk pathways: Northeastern region industrial upgrading pressures inversely correlate with household income growth, while western regions accumulate contradictions between industrial monolithicity and risk diversification capacity. These findings necessitate regionally adaptive policy frameworks incorporating real-time risk assessment mechanisms.

References

- [1] CAO Y, CAI Z. The Empirical Analysis on the Contagion Effect of the Regional Financial Risks Based on the VAR Model -- Illustrated by the Data Before and After the International Financial Crisis [J]. *On Economic Problems*, 2013, (10): 59 - 64.
- [2] D'Amico, G., Basilio, B.D. and Petroni, F. Drawdown-based risk indicators for high-frequency financial volumes [J]. *Financial Innovation*, 2024, 10 (1).
- [3] GIUDICI P, SARLIN P, SPELTA A. The interconnected nature of financial systems: direct and common exposures [J]. *Journal of Banking & Finance*, 2020, 112: 105149.
- [4] Kim, K. Liquidity basis between credit default swaps and corporate bonds markets [J]. *International Review of Economics & Finance*, 2017, 48: 98 – 115.
- [5] Lai, J., and Lyu, J. L. Measurement of Financial Systemic Risk Based on Financial Stress Index [J]. *Statistics and Decision*, 2010, (19): 128 - 131.
- [6] Li, F. and Perez-Saiz, H. Measuring systemic risk across financial market infrastructures [J]. *Journal of Financial Stability*, 2018, 34: 1 – 11.
- [7] Li, L. H., and Han, J. W. Does Financial Innovation Exacerbate Regional Financial Risk? --A Study Based on Social Network and Dynamic Panel Threshold Effects [J]. *Securities Market Herald*, 2023, (3): 68 - 79.
- [8] Li, N. Measurement of Systemic Financial Risk Based on Real Estate Market [J]. *Frontiers in Business, Economics and Management*, 2024, 17 (3): 185 – 188.
- [9] Li, Q. and YUSOFF, M.E. The Impact of Household Financial Asset Allocation on Regional Bank Risk Spillovers [J]. *Proceedings of the International Conference on Information Economy, Data Modeling and Cloud Computing, ICIDC 2022*, 17 - 19.
- [10] Shen L, Liu Y, Li W. China's Regional Financial Risk Spatial Correlation Network and Regional Contagion Effect: 2009-2016 [J]. *Management Review*, 2019, 31 (8): 35 - 48.
- [11] Shen, L., Zhang, Y., Li, W., and Liu, Y. Spatial-Temporal Evolution and Driving Mechanisms of Regional Financial Risks in China: A Perspective of Four Economic Sectors [J]. *Southern Economic Journal*, 2019, (9).
- [12] Tang E, Bai C, Qin T. Digital finance and regional systemic financial risk Forecasting [J]. *Applied Mathematics and Nonlinear Sciences*, 2024, 9 (1).
- [13] Xu Y. Empirical Study on the Measurement and Influencing Factors of China's Systemic Financial Risk Based on Machine Learning [J]. *Financial Development Review*, 2023, (10): 44 - 61.
- [14] WANG, Z. A Study on Early Warning of Financial Indicators of Listed Companies Based on Random Forest [J]. *Scientific Programming*, 2022.
- [15] YANGYUDONGNANXIN, G., ALI, R. Financial Credit Risk Control Strategy Based on Weighted Random Forest Algorithm [J]. *Scientific Programming*, 2021, (2).
- [16] ZHANG L, ZHANG S. Progress on Machine Learning for Regional Financial Risk Prevention [J]. *Journal of Frontiers of Computer Science and Technology*, 2022, 16 (9): 1969 - 1989.
- [17] ZHOU W, He Z. Study on spatial distribution, regional differences and dynamic evolution of rural financial risk in China [J]. *PLOS ONE*, 2024, 19 (5): e0301977.