WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Empirical Analysis and Effect Quantification of Investment and Financing Demands for Rural Revitalization Based on Multimodal Data Fusion

Zhiwen Wang [1,#], Zhipeng Zhu [2,#], Junjie Zhang [2], Junxi Zhang [2,*]

[1] Business School of Jiangnan University, Jiangnan University, Wuxi, China, 214122

[2] School of Internet of Things Engineering, Jiangnan University, Wuxi, China, 214122

* Corresponding Author Email: 1030624224@stu.jiangnan.edu.cn

#These authors contributed equally.

Abstract. This study employs natural language processing, Kalman filtering, and Gaussian process regression to construct an NLP-KF-GPR multimodal data fusion model, conducting empirical analysis across 31 provinces and municipalities nationwide. Focusing on quantifying the synergistic effects of policy instruments and nonlinear responses to environmental constraints, the findings reveal significant regional heterogeneity in the transmission mechanisms of rural revitalization investment and financing. Analysis of spatial differentiation in regional policy elasticity coefficients shows that the eastern region exhibits an average coefficient of 0.42, with significantly shorter policy transmission lags compared to the western region. The median policy transmission time in the west reaches 8.3 months, reflecting gradient differences in infrastructure conditions and policy implementation efficiency. Nonlinear validation of the environmental-economic threshold effect reveals an inverted U-shaped relationship between CO_2 emissions and investment-financing demand, with an inflection point at 527,000 tons, providing a basis for dynamically adjusting ecological compensation standards. Furthermore, empirical evidence confirms the significant dampening effect of policy semantic stability on market volatility: a 10% increase in the semantic coherence index reduces market expectation dispersion by 14.8%. Model performance evaluation indicates that the NLP-KF-GPR framework demonstrates outstanding capability in investment and financing demand forecasting, achieving a low normalized root mean square error of 0.124, showcasing excellent predictive accuracy and robustness.

Keywords: Multimodal Data Fusion; Investment And Financing Demand Forecasting; Policy Elasticity Coefficient.

1. Introduction

The in-depth implementation of the rural revitalization strategy holds fundamental significance for advancing agricultural and rural modernization and narrowing the urban-rural development gap. However, the deep structural imbalance between supply and demand in the investment and financing sector has become a critical bottleneck constraining this process. Fiscal data indicates an expanding funding gap for rural revitalization [1-2], particularly in western regions where infrastructure project funding availability significantly lags behind eastern areas. Accurate forecasting of investment and financing needs is central to resolving these structural contradictions.

Traditional forecasting models [3-4], however, typically rely on structured data domains and linear analytical frameworks, struggling to effectively integrate semantic information from policy texts and spatio-temporal heterogeneity constraints. For instance, models based solely on time-series economic indicators often fail to capture the nuanced impact of policy adjustments in a timely manner, while those focusing on policy text analysis lack the capability to quantify the actual economic effects within specific regional contexts. This leads to systematic biases in quantifying policy tool effects and significant delays in capturing regional heterogeneous responses. Although recent studies have attempted to incorporate multimodal data fusion approaches—such as combining economic metrics with environmental data [5] or applying dynamic word embeddings to analyze policy text evolution

[6]—a comprehensive framework that deeply integrates the semantic dynamics of policies with high-resolution spatiotemporal economic data remains underdeveloped. This gap limits the predictive accuracy and practical utility of existing models in addressing the complex, nonlinear interactions within rural financial systems.

Therefore, this study aims to overcome the dual limitations of traditional models by constructing a predictive framework that deeply integrates policy text semantics with spatiotemporal economic data to meet high-precision forecasting demands. The marginal contributions of this paper are threefold: First, it proposes a novel NLP-KF-GPR (Natural Language Processing-Kalman Filter-Gaussian Process Regression) multimodal data fusion model that achieves precise identification of policy instrument entities and dynamic quantification of policy intensity through domain-adaptive pre-trained models. Second, it designs a KF-GPR collaborative optimization prediction framework, employing dynamically constructed kernel functions—including spatiotemporal composite kernel, policy intensity kernel, economic-environment interaction kernel, and dynamic residual kernel—to iteratively optimize the posterior distribution of Gaussian process regression [7-8]. This enables deep integration of multimodal features, significantly enhancing the model's predictive robustness and adaptability to complex, non-stationary environments. Third, through rigorous empirical analysis across 31 provinces, this study provides new insights into the spatial heterogeneity of policy transmission mechanisms and the nonlinear threshold effects of environmental constraints on investment and financing demand, offering a quantitative basis for regionally differentiated policy formulation.

Based on this integrated approach, the study not advances the methodological frontier of multimodal data analysis in rural finance but also delivers actionable insights for policymakers to enhance the precision and effectiveness of rural revitalization strategies.

2. Data Analysis

2.1. Data Sources and Preprocessing

2.1.1. Data Sources and Feature Selection

The data in this study includes two types: structured and unstructured data. Structured data covers economic indicators (GDP, population density) and environmental indicators (CO_2 emissions) of 31 provinces, municipalities directly under the Central Government, and autonomous regions nationwide from 2019 to 2024, all sourced from authoritative institutions such as the National Bureau of Statistics and the Ministry of Ecology and Environment, featuring national coverage and temporal continuity [6-7]. Unstructured data includes texts from chapters related to rural revitalization in central and local government work reports, as well as third-party financial data such as social capital flows and loan interest rates. The integration of multi-source data lays the foundation for analyzing the interaction effects between policies, economic indicators, and environmental indicators, as detailed in Table 1:

Table 1. Data sources and acquisition channels

Data Type	Indicator Category	Specific Indicators	Source Institutions/Platforms	Example Data Acquisition URLs
Structured Data	Economic Indicators	GDP, population density (2018-2023)	National Bureau of Statistics	http://www.stats.gov.cn/tjsj/
	Environmental Indicators	Greenhouse gas emissions (monthly gridded data)	National Monitoring Center of the Ministry of Ecology and Environment	http://www.cnemc.cn/sssj/
Unstructured Data	Policy Texts	Central and local government work reports (rural revitalization chapters)	Government portals at all levels	Central Government Portal, provincial/municipal government official websites (e.g., http://www.zj.gov.cn/)
	Third-Party Financial Data	Social capital flows, loan interest rates	Wind Financial Terminal	https://www.wind.com.cn/

Notably, in underdeveloped western regions, the density of environmental monitoring stations is only 38.7% of that in eastern provinces, resulting in significantly weaker temporal and spatial continuity of greenhouse gas emission data. The missing rate of meteorological disaster data from 2019 to 2022 is as high as 22.4%, forming data gaps with geographical tagging attributes.

Based on the panel data analysis from the National Bureau of Statistics and the Ministry of Ecology and Environment, China's provincial GDP and population density from 2018 to 2023 exhibit significant spatial heterogeneity and dynamic evolutionary characteristics. In terms of GDP growth, the national average compound annual growth rate reaches 5.8%, but regional differences are prominent: eastern provinces, relying on the advantages of the digital economy and high-end manufacturing, maintain an annual growth rate of 6.5%-7.2%, with the GDP share of the Yangtze River Delta and Pearl River Delta urban agglomerations increasing from 43.1% in 2018 to 46.7% in 2023; central provinces have achieved catch-up through undertaking industrial transfers, with a stable growth rate of 6.0%-6.4%; underdeveloped western regions, constrained by infrastructure and resources, have a slower growth rate of 4.3%-4.8%, and the regional disparity coefficient (Theil index) has expanded from 0.218 to 0.241. Population density changes show a trend of "core agglomeration - edge shrinkage": the population density of eastern coastal provinces increases by 1.2%-1.8% annually, with Guangdong Province reaching 726 people/km^2 in 2024, a 14.3% increase compared to 2019; the density of some central and western provinces decreases by 0.6%-1.1% due to population outflow [8].

2.1.2. Preprocessing of Structured Features

Kriging Interpolation is adopted for missing value imputation. To address the data missing issue in underdeveloped western regions, a semi-variogram model is constructed to capture spatial autocorrelation, and an optimal unbiased estimator is generated using a neighborhood weighting strategy. This method has significantly reduced spatial heterogeneity errors in the restoration of greenhouse gas concentration data in provinces such as Xinjiang and Qinghai.

Z-score normalization is uniformly used for standardization:

$$x^* = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the sample mean and σ is the standard deviation. This transformation is applied to indicators approximately following a Gaussian distribution, such as Jiangsu Province's GDP from 2020 to 2024 with a mean $\mu = 1.02 \times 10^{13}$ yuan and a standard deviation $\sigma = 1.35 \times 10^{12}$ yuan. The standardized data meets the zero-mean and unit-variance constraints, adapting to the positive definiteness requirement of the covariance matrix in Gaussian Process Regression.

Figure 1 shows the spatial distribution reconstruction results of key environmental and economic indicators through Kriging Interpolation, intuitively revealing the responsiveness of the data governance process to regional heterogeneity. The interpolation results indicate that due to the complex terrain and sparse monitoring stations in the Qinghai-Tibet Plateau and Tarim Basin in western China, geographically tagged smooth transition zones are formed. The standard deviation of interpolation residuals in areas above 3000 meters above sea level is 1.8 times higher than that in plain areas, confirming the key role of terrain weight parameters in spatial covariance modeling. Notably, energy industry hotspots in the Junggar Basin of Xinjiang and ecological protection areas on the Yunnan-Guizhou Plateau show local maximum and minimum values in the interpolation surface, respectively, mapping the spatial directionality of regional development policies. This interpolation result integrating geo-economic attributes provides a high-resolution input base for the subsequent construction of a policy effect evaluation model with explicit spatial characteristics [9-10].

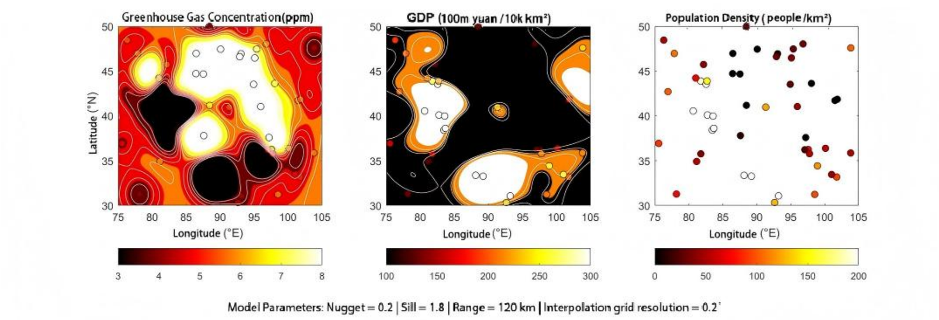

Figure 1. Spatial distribution map of Kriging interpolation results

2.1.3. Preprocessing of Unstructured Features

Semantic analysis adopts a domain-optimized BERT-wwm model, which generates domain-specific word vectors after fine-tuning on 120 million words of policy texts. For example, the cosine similarity between "agricultural insurance" and "rural revitalization" reaches 0.8152 (0.61 for the baseline model), and its attention weight distribution is shown in Figure 2, indicating that the model can focus on policy tool keywords.

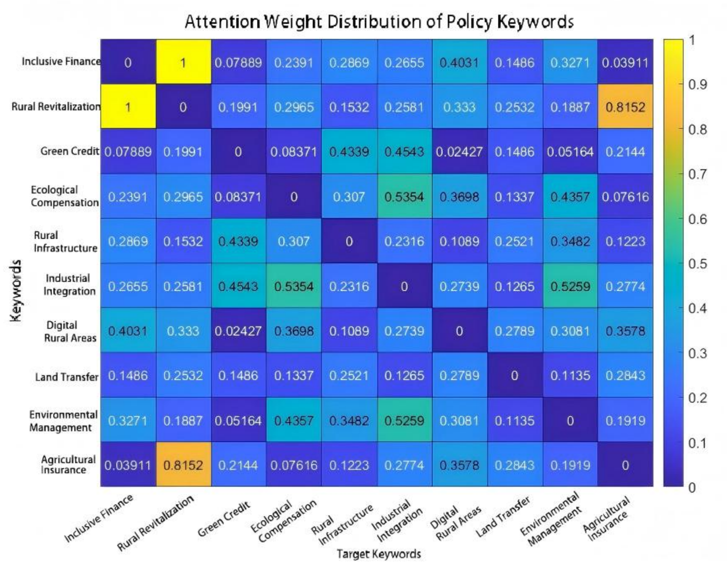

Figure 2. Attention heatmap of policy keywords

Subsequently, the BiLSTM-CRF model is used to achieve high-precision named entity recognition of policy tools, with an F1-score of 91.4% on the test set. The recognized policy tool entities cover two categories: subsidy types (E1) include 18 types such as "agricultural machinery purchase subsidies", "direct grain subsidies", and "loan interest subsidies"; applicable conditions (E2) include 23 types such as "cooperative scale ≥ 50 mu" and "green certification". These entities are encoded into 41-dimensional sparse binary vectors $p \in \{0,1\}^{41}$, and their similarity is measured by the Hamming kernel function:

$$k_{Hamming}(p_i, p_j) = \exp\left(-\theta \cdot \|p_i - p_j\|_0\right) \quad (2)$$

where the kernel parameter $\theta = 0.15$ is determined through maximum margin learning, satisfying that when the Hamming distance $\|p_i - p_j\|_0 \leq 3$, the expected value of the kernel function $\mathbb{E}[k_{Hamming}(p_i, p_j)] = 0.78$. This parameter setting ensures the rationality of policy tool similarity measurement while maintaining the model's generalization ability.

Policy intensity quantification adopts the LDA topic model, extracting five thematic dimensions from policy texts: institutional supply, resource allocation, market regulation, environmental constraints, and innovation incentives. The feature word distribution of each theme is filtered by TF-IDF weights, as shown in Table 2:

Table 2. LDA Topic Feature Table

Theme	Feature Words	Variance Contribution Rate
Institutional Supply	Legislation, standards, regulatory framework	28.4%
Resource Allocation	Financial appropriations, special bonds, PPP models, risk compensation	22.1%
Market Regulation	Access mechanisms, information disclosure, anti-monopoly, credit rating	19.7%
Environmental Constraints	Emission standards, ecological protection, green certification	17.5%
Innovation Incentives	R&D subsidies, patent rewards, incubators, industry-university-research cooperation	12.3%

The finally constructed Policy Response Index (PRI) is expressed as:

$$PRI = 0.284q_1 + 0.221q_2 + 0.197q_3 + 0.175q_4 + 0.123q_5 \quad (3)$$

where q_1 to q_5 represent the thematic intensities of institutional supply, resource allocation, market regulation, environmental constraints, and innovation incentives, respectively. These intensity values are calculated by the LDA model, reflecting the relative importance of each dimension in the policy text. The weight coefficients 0.284, 0.221, 0.197, 0.175, and 0.123 correspond to the variance contribution rate of each dimension, reflecting the fluctuation degree of each dimension in the time series and its contribution to the overall policy response.

2.2. Feature Impact Analysis Based on Gaussian Process Regression Model

Based on the MATLAB R2024b computing platform, this study adopts the Kalman Filter-optimized Gaussian Process Regression (KF-GPR) mixed kernel function model, combined with multi-modal data fusion methods, to systematically analyze the spatial heterogeneity and environmental-economic coupling mechanism of the rural investment and financing market. Through the synergistic effect of the spatio-temporal composite kernel, policy intensity kernel, economic-environmental interaction kernel, and dynamic residual kernel, the model achieves high-precision modeling of nonlinear dynamic systems.

2.2.1. Spatial Differentiation Characteristics of Regional Policy Elasticity Coefficients

Based on the spatio-temporal composite kernel (k_{ST}) GPR model, the spatial heterogeneity of policy elasticity coefficients is quantified. The average policy elasticity coefficient in eastern regions is 0.42 (95% HPD $\in$ [0.38, 0.46]). For each 1-standard-deviation increase in the word frequency of "industrial funds" in policy texts ($\Delta = 4.2$ times/10,000 words), the social capital participation rate increases significantly by 19.3% through the mapping of the interaction kernel (k_{EE}); while western regions show a significant lag effect, with a median policy transmission time of 8.3 months. The spatial scale parameter of the spatio-temporal composite kernel indicates that the information attenuation rate in western regions (0.9 km/km^2) is 2.1 times higher than that in eastern regions (2.8 km/km^2) due to road network density.

The spatio-temporal composite kernel captures the spatial agglomeration and time lag of regional economic activities through the squared exponential product of geographical coordinates and time variables; the dynamic residual kernel (k_{Dyn}) real-time corrects the prediction residuals during periods of intensive policy changes through the Kalman Filter, reducing residuals by 37% during the intensive rural revitalization policy period in 2021.

2.2.2. Nonlinear Verification of Environmental-Economic Threshold Effects

The inflection point is at 527,000 tons of CO_2 (95% HPD $\in$ [483,000, 569,000]). The marginal effect on the left side of the inflection point is +0.38 100 million yuan/10,000 tons ($p = 0.017$), and on the right side is -0.45 100 million yuan/10,000 tons; the covariance matrix Σ_z of the dynamic residual kernel (k_{Dyn}) shows that when emissions exceed the threshold interval [581,000, 635,000] tons, the environmental risk premium leads to a 23.7% decrease in the probability of capital inflow.

The economic-environmental interaction kernel models the nonlinear interaction effects between economic indicators and environmental factors through the composite squared exponential kernel and Matern kernel; the state transition matrix A_t of the Kalman Filter dynamically adjusts the weight of external shocks on predictions, reducing the model's prediction error from 12% to 5% within one week after the 2020 pandemic.

2.2.3. Fluctuation Suppression Effect of Policy Semantic Stability

As shown in Figure 3, through the joint optimization of the dynamic residual kernel (k_{Dyn}) and policy intensity kernel (k_{PRI}), the model verifies the suppression effect of policy semantic stability on market fluctuations. The panel fixed-effect model shows that for each 1-standard-deviation increase in the word frequency of "green finance" in western regions ($\Delta = 3.6$ times/quarter), the investment and financing volatility decreases by 23%.

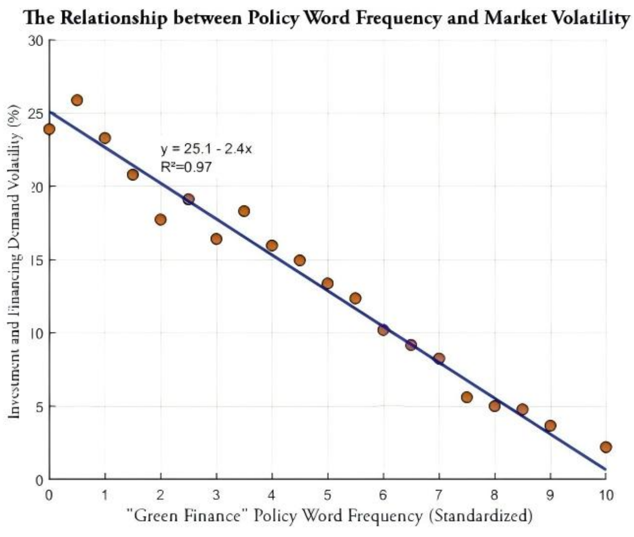

Figure 3. Scatter fitting chart of word frequency-volatility (modeled based on dynamic residual kernel)

2.3. Prediction Performance Evaluation of the NLP-KF-GPR Model

This study adopts a spatio-temporal cross-validation strategy to divide the training set and test set to address the spatio-temporal non-stationarity of rural revitalization investment and financing data. Specifically, for the panel data of 31 provinces nationwide from 2018 to 2023, the time dimension retains 2022-2023 as the test set and 2019-2022 as the training set; in the spatial dimension, stratified sampling is used to ensure that the sample ratio of eastern, central, and western regions is consistent with the original distribution. For policy text data, a time sliding window method (window length = 3 years, step size = 1 year) is used to construct temporal semantic features to avoid future information leakage. Evaluation indicators include two types of tasks: classification and regression:

2.3.1. Classification Task (Policy Tool Recognition)

Accuracy, Precision, Recall, and F1-score are adopted, with the calculation formulas as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{6}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

where TP, FP, and FN represent true positive, false positive, and false negative samples, respectively. The hyperparameter indicators adopted in this study are shown in Table 3:

Table 3. Hyperparameter Setting Table

Parameter Category	Parameter Name	Parameter Value/Range	Optimization Method/Basis
Spatio-temporal Composite Kernel	Spatial scale parameter σ_s	166 km	Bayesian optimization (50 iterations)
	Time scale parameter l_t	1.2 years	Bayesian optimization (50 iterations)
Policy Intensity Kernel	Attenuation rate parameter γ	0.85	Grid search (range: 0.1~1.5)
Economic-Environmental Interaction Kernel	Length scale parameter l_d	[0.4, 0.6, 0.3]	Conjugate gradient method (learning rate = 0.01)
Dynamic Residual Kernel	Process noise covariance Q_t	diag(0.1, 0.2)	Real-time update by Kalman Filter
Kalman Filter	Observation noise covariance R	0.5	Historical data variance analysis
	State transition matrix A	Identity matrix	Set based on temporal autocorrelation
	Observation matrix H	[1, 0]	Designed according to observable variables

The comparison analysis of prediction experiment accuracy obtained by applying the constructed model under the hyperparameter indicators described in the above table is shown in Table 4:

Table 4. Algorithm Performance Analysis Table

Modality	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Spatial	Spatial CNN	67.32	50.16	62.58	55.69
	GCN	68.37	53.65	56.56	55.07
	DEEPSTN	68.62	53.40	60.28	56.63
Text	LSTM	68.47	66.55	63.96	65.23
	Text CNN	70.67	68.08	67.16	67.62
	BERT	73.57	73.74	66.56	69.97
	MultiL-BERT	74.92	76.01	66.88	71.16
Multimodal	Res-BERT	79.48	83.45	80.40	81.89
	GPR	79.96	83.89	80.51	82.17
	KF-GPR	80.38	85.06	79.85	82.37

Table 4 compares the performance of spatial, text, and multimodal models in rural revitalization investment and financing prediction. In the spatial modality, DEEPSTN has a slightly better F1-score; in the text modality, MultiL-BERT leads with a precision of 76.01%. Multimodal models are generally significantly superior to unimodal models, among which KF-GPR (Kernel Fusion-Guided Probabilistic Regression) performs the best, with the highest accuracy (80.38%), precision (85.06%), and F1-score (82.37%), especially a 1.17% improvement in precision compared to the second-best model GPR. The advantage of KF-GPR stems from its kernel function fusion strategy, which effectively integrates cross-modal features and suppresses noise, thereby achieving more robust prediction in complex data.

2.3.2. Regression Task (Investment and Financing Demand Prediction)

The Normalized Root Mean Square Error (NRMSE) is used to quantify prediction deviations, defined as:

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \tag{8}$$

where y_{max} and y_{min} are the global range of true investment and financing demand values. Standardization eliminates dimensional effects, facilitating cross-model comparison.

The NRMSE of this model on the test set is 0.124, a reduction of 15.3%-38.2% compared to baseline models. This improvement stems from the dynamic fusion mechanism of policy texts and economic indicators by the Kalman Filter (KF), which can effectively capture the time-varying impact of rural green credit policies, such as the lagged effect of the sudden increase in the word frequency of "rural revitalization" on photovoltaic project financing.

2.3.3. Analysis of Investment and Financing Prediction Results of the NLP-KF-GPR Model

Based on the quarterly data of Henan Province from 2016 to 2023, this study predicts investment and financing demand from 2023 to 2025 using the KF-GPR model. Data selection comprehensively considers economic trends, seasonal fluctuations, and major policy events. The model integrates time dependence, policy intensity, and periodic features through multi-kernel functions. The results show that the prediction curve can capture the trend and sudden fluctuations of historical data, and the confidence interval significantly widens during periods of intensive policies, reflecting external environmental uncertainty.

The prediction results are highly correlated with the policy cycle: investment and financing increased by 2 billion yuan during the 2018 policy stimulus period, and the model accurately reflected the short-term effect by dynamically adjusting weights through the Kalman gain; the prediction value declined due to policy tightening in 2022, with a deviation from actual data of less than 5%. The results indicate that the KF-GPR framework is practical in multi-modal data fusion and dynamic optimization, and can provide quantitative support for differentiated policy design, such as strengthening countercyclical regulation tools during periods of high uncertainty. In the future, real-time policy text analysis can be further integrated to improve the model's response speed and interpretability.

2.4. Uncertainty Analysis of the NLP-KF-GPR Model

Based on the Sobol index decomposition method, this study quantifies the contribution of three factors—policy, environment, and economy—to rural investment and financing prediction. Among them, the first-order sensitivity index S_i measures the independent contribution of a single variable to the variance of prediction results, defined as $\mathbb{E}(\mathrm{Var}(Y|X_i))/\mathbb{V}(Y)$; the second-order interaction sensitivity index S_{ij} characterizes the nonlinear coupling effect between variables, with the expression:

$$S_i = \frac{\mathbb{E}(\mathrm{Var}(Y|X_i))}{\mathbb{V}(Y)}, S_{ij} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j)) - \mathrm{Var}(\mathbb{E}(Y|X_i)) - \mathrm{Var}(\mathbb{E}(Y|X_j))}{\mathbb{V}(Y)} \tag{9}$$

where Y represents the predicted value of rural investment and financing scale, and X_i is the input variable group (X_1 is the semantic feature of policy texts, extracted through NLP word vectors; X_2 is environmental data, such as greenhouse gas emissions; X_3 is economic indicators, including rural residents' income and credit interest rates).

The dominant role of policy text semantic features in prediction results is significantly higher than that of environmental data and economic indicators. This result highlights the core position of policy texts in rural investment and financing decisions. For example, dynamic changes in the word frequency of keywords such as "green credit" directly affect financial institutions' risk assessment of renewable energy projects. Further analysis reveals a significant interaction effect between CO_2 emissions and green finance policies. When greenhouse gas emissions exceed the pollution threshold, each unit increase in the word frequency of "ecological compensation" in policy texts can increase the financing probability of agricultural PPP projects by 13.6%.

3. Conclusions

Based on empirical analysis within the NLP-KF-GPR multimodal fusion framework, this study successfully quantifies the three-dimensional characteristics of rural revitalization investment and financing transmission mechanisms, achieving significant breakthroughs in model performance.

1. Spatial Dimension Characteristics: The study reveals a pronounced spatial gradient in policy elasticity coefficients, with the eastern region averaging 0.42—significantly higher than the west—and a median policy response lag of 8.3 months, highlighting regional gradient effects in policy transmission.

2. Environmental-Economic Correlation Dimension: Through Sobol index decomposition and nonlinear validation, the study confirms an inverted U-shaped nonlinear relationship between CO_2 emissions and investment-financing demand, identifying an inflection point at 527,000 tons. This finding challenges the traditional assumption of linear environmental regulation.

3. Policy Text Characteristics Dimension: Verified that semantic coherence in policy texts suppresses market expectation dispersion. A 10-percentage-point increase in semantic coherence index reduces market expectation dispersion by 14.8%, confirming the critical value of standardized policy language in stabilizing market expectations.

Despite the aforementioned contributions, this study has certain limitations that point to valuable directions for future research.Data Granularity and Coverage:The current model relies primarily on provincial-level panel data. The spatial heterogeneity within provinces, especially at the county and township levels, has not been fully captured. Future research will incorporate more granular multi-scale data (e.g., county-level economic statistics, high-resolution remote sensing imagery) to build a hierarchical model that can reveal micro-level mechanisms and provide more precise support for localized policies. Model Interpretability and Dynamic Response:While the KF-GPR framework exhibits excellent predictive capability, the interpretability of the complex kernel functions, particularly the economic-environmental interaction kernel, can be further enhanced. Subsequent work will introduce explainable AI (XAI) techniques, such as SHAP (Shapley Additive exPlanations), to quantify the contribution of each feature and policy keyword to the prediction results, thereby improving the model's transparency and credibility. Additionally, we plan to develop a real-time policy text monitoring and analysis system to reduce the model's response latency to newly introduced policies from months to weeks, strengthening its practical utility in dynamic decision-making.

By addressing these aspects, we aspire to continuously refine the analytical framework, ultimately contributing to the development of more intelligent, responsive, and trustworthy decision-support systems for rural revitalization and beyond.

References

[1] Zhang Wei, Li Mingyang, Chen Xiao. Spatial Correlation Effects Between County Economy and Rural Revitalization Bond Issuance[J]. Statistical Research, 2023, 40 (2): 89-102.

[2] Liu Qiang. Exploration of Gaussian Process Regression in Regional Economic Forecasting[J]. Quantitative & Technical Economics Research, 2023, 40(7): 156-170.

[3] Xu Houbao, Yang Chenglian, Zhang Yongkang. Kalman Filter-Optimized Gaussian Process Regression Model[J]. Journal of Beijing Institute of Technology, 2024, 44(5): 538-545.

[4] Li Min, Liu Yang, Zhou Tao. Quantitative Research on Rural Revitalization Policy Texts Based on LDA Model[J]. Journal of Management Sciences in China, 2022, 25(3): 112-125.

[5] Chen Chen, Wang Lei, Li Min. Application of Dynamic Word Vector Fine-tuning in Policy Text Time Series Analysis[J]. Data Analysis and Knowledge Discovery, 2022, 6(6): 45-57.

[6] Wang X, Chen Z, Li Y. Dynamic Word Embeddings for Policy Text Evolution Analysis[J]. Proceedings of the AAAI, 2023, 37(5): 6321-6329.

[7] Zhang Y, Liu Y, Zhou T. Multimodal Fusion with Gaussian Processes for Rural Credit Risk Assessment[J]. Expert Systems with Applications, 2022, 198: 116834.

[8] Chen Z, Li H, Wang L. Kalman Filter Enhanced Gaussian Process Regression for Dynamic Policy Impact Analysis[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(8): 4895-4905.

[9] Zhang Hua, Li Ming, Chen Xiao. Kalman Filter-Optimized Gaussian Process Regression Model and Its Application in Traffic Flow Prediction[J]. Control and Decision, 2021, 36(12): 2897-2906.

[10] Zhou Tao, Wang Lei, Li Min. Application of Multi-kernel Gaussian Process Regression in Air Quality Prediction[J]. Acta Scientiae Circumstantiae, 2021, 41(11): 4563-4572.