Journal of Economics and Management Sciences; Vol. 8, No. 1; 2025
ISSN 2576-3008 E-ISSN 2576-3016
https://doi.org/10.30560/jems.v8n1p1

# China's Economic Growth from a Comprehensive Analysis Perspective: A Synergistic Application of GRA, PCA, and GRU Models

Min Juanhua[1] & Xu Donglin[1]

[1] Guangdong Baiyun University, China

Correspondence: Min Juanhua, Guangdong Baiyun University, Guangzhou, Guangdong 510550, China.

Received: December 3, 2024; Accepted: January 5, 2025; Published: January 6, 2025

**Abstract**

In the context of globalization and economic integration, accurately predicting China's quarterly GDP growth rate is crucial for macroeconomic decision-making. This paper proposes a comprehensive forecasting framework that combines Grey Relational Analysis (GRA), Principal Component Analysis (PCA), and a Gated Recurrent Unit (GRU)-based neural network model to improve prediction accuracy. The study first identifies economic indicators highly correlated with GDP growth through GRA. It then extracts key variability from multidimensional data using PCA, constructing a "China Macroeconomic Vitality Index." Simultaneously, a GRU model, combined with error correction and Bayesian optimization techniques, significantly enhances prediction accuracy. The results demonstrate that the optimized GRU model exhibits significant reductions in evaluation metrics such as SMAPE, MAE, and MSE, indicating good forecasting performance. Additionally, SHAP value analysis reveals the specific impact of each economic indicator on GDP growth, providing a basis for macroeconomic decisions.

**Keywords:** GDP Growth Prediction, Grey Relational Analysis, Principal Component Analysis, GRU Neural Network

## 1. Introduction

In the current complex and volatile global economic landscape, accurately predicting macroeconomic indicators, especially the Gross Domestic Product (GDP) growth rate, is of great significance for government policy formulation, corporate strategic planning, and investor decision-making. However, existing forecasting methods still have limitations, which affect the accuracy and practicality of prediction results. For example, traditional forecasting models rely on conventional statistical data, overlooking the potential of big data sources such as internet search data. This data provides real-time market information that can offer new perspectives for GDP prediction, but how to integrate and process these non-traditional data sources, and ensure their data quality and reliability, remains a challenge. Traditional models may face overfitting problems, especially when dealing with complex economic cycles and external shocks. Models need to have strong generalization capabilities to avoid prediction biases. In addition, model parameter selection and optimization are also critical, especially for neural network models that require a large number of parameter adjustments, which is a complex and time-consuming process.

In view of this, this paper conducts an immediate forecasting study on China's quarterly year-on-year GDP growth rate through the comprehensive use of Grey Relational Analysis, Principal Component Analysis (PCA), and Gated Recurrent Unit (GRU)-based neural network models. It aims to build a high-precision forecasting model, deeply analyze the key economic indicators affecting GDP growth, and provide a scientific basis for macroeconomic decision-making.

## 2. Literature Review

### *2.1 Traditional GDP Growth Forecasting Models*

Traditional economic forecasting models mainly rely on mathematical relationships in historical data, such as Autoregressive Integrated Moving Average (ARIMA) [1], Seasonal Autoregressive Integrated Moving Average (SARIMA) [2], etc. These models have certain advantages in dealing with the time series characteristics of data, but also have shortcomings such as high requirements for data quality, strong subjectivity in parameter selection, and insufficient ability to capture non-linear relationships[14]. In addition, traditional models generally lack the ability to process and analyze data in a big data environment, and it is difficult to adapt to the challenges brought by the rapid growth and structural changes of economic data [3].

In 2024, Yi Yanping, Huang Dejin, and others proposed a mixed-frequency dynamic multi-factor model based on the Lasso method combined with the EM algorithm, and used a large number of monthly macroeconomic variables to make an instant prediction of China's quarterly year-on-year GDP growth rate [4]. Xiang Ping [5] and Wang Shasha [6] respectively applied the combined model forecasting method to the prediction of China's GDP, thereby improving the prediction accuracy of the model. Zhang Jiangcheng [7] and Li Na and Xue Junqiang [8] used Eviews software to model China's GDP data and made reasonable short-term predictions.

### *2.2 Prediction Research Based on Neural Networks*

In recent years, significant progress has been made in prediction research based on neural networks. In 2024, He Yating, Sun Ying, and others constructed a SARIMA-LSTM combined model based on residual correction in a non-linear combination method to predict the national quarterly GDP [9]. In 2024, Wu Yang and Luo Jikang studied the use of Long Short-Term Memory Neural Networks (LSTM) to build two-step and three-step prediction models to predict the annual GDP data of Inner Mongolia Autonomous Region from 1992 to 2022 [10]. Guo Qiuyan and He Yue established a model for predicting GDP by combining the DFA method and BP neural network. The conclusion is that the GDP prediction model based on the DFA method and BP neural network is effective and has high prediction accuracy. This model can be applied to actual economic forecasting to provide valuable information for related decisions [11]. By comparing and analyzing the predictive capabilities of the BVAR model and the LSTM model with and without expansion indicators for GDP, Xiao Zhengyan and others concluded that the LSTM neural network model shows broad prospects in the field of GDP prediction. Its strong self-learning ability and good generalization ability make it have obvious advantages in short-term GDP prediction [12]. In addition, Liang Longyue used wavelet analysis technology (WA) to decompose the selected macroeconomic variables for quarterly GDP data, and constructed and proved that the LSTM&WA model has a good effect on quarterly GDP prediction [13].

In addition, mixed forecasting methods that combine neural networks with other economic models are also constantly emerging. By comprehensively utilizing the advantages of different models, the prediction accuracy is further improved. These advances not only enrich the theoretical methods of economic forecasting, but also provide powerful tools for actual policy formulation and economic analysis.

## 3. Research Design

### *3.1 Problem Description and Analysis*

In the research on immediate GDP forecasting, the first challenge we face is how to effectively process and analyze high-dimensional mixed-frequency unbalanced panel data. The characteristics of this type of data lie in its high dimensionality, inconsistent frequencies, and the presence of missing data. These issues pose higher requirements for the construction and accuracy of prediction models.

Problem 1: Data preprocessing and solving unbalanced panel data problems.
Problem 2: Correlation analysis between economic indicators and GDP.
Problem 3: Setting of indicator factor parameters.
Problem 4: Establishment and application of an immediate prediction model.
Problem 5: Calculation of indicator contribution values and impact analysis.

Throughout the research process, our goal is to build a model that can accurately predict GDP growth while being able to identify the key economic indicators affecting GDP. This requires not only precise processing and analysis of data, but also careful design and verification of the model. Through this method, we can provide a scientific and accurate basis for macroeconomic decision-making.

### *3.2 Data Source*

The original data comes from the original dataset of GDP prediction-related indicators provided by the "Big Data Innovation and Creativity Track" of the 2024 Guoyan Economic Big Data Modeling Competition. After checking with the data source "National Bureau of Statistics of China", it is confirmed that the data is reliable and complete. The original dataset contains a total of 57 indicators, of which 56 are exogenous variables.

At the same time, the original data uses the time format of 2014-01-01 as the data index, and the full set records a total of 10,575 lines of time series data from 1978-12-31 to 2024-5-14. However, due to the wide variety of data, the update time and update frequency of various data are different, resulting in high-dimensional mixed-frequency unbalanced panel problems. This puts higher demands on the forecasting work.

### *3.3 Data Preprocessing*

#### 3.3.1 Unbalanced Panel Data Transformation

After uniformly processing the data in the time dimension through Python, we obtain time series data with a uniform time dimension of quarters. Since there are still many missing values in the dataset after balancing the dimensions, the dataset is further analyzed for missing values.

#### 3.3.2 Missing Value Analysis and Handling

Since there are continuous missing values for endogenous variables from December 31, 1978 to December 31, 1997, it is chosen to remove the data before December 31, 1997, and use SPSS 25.0 to analyze the missing values of the remaining 130 lines of data. Remove exogenous variables with a missing value ratio greater than 40%. Then, use the EM method and regression method to estimate the remaining 48 indicators. The mean, covariance, and correlation of the EM method and the regression method can be obtained. At the same time, it can be seen from the correlation analysis MACR test results in the table below that the significance P is less than 0.05, which means that the missing data rejects the original hypothesis that the missing value is MACR (Missing Completely at Random).

Table 1. Correlation MACR Test Analysis

| **EM Estimation Statistics** | | | |
|---|---|---|---|
| a. Little's MCAR Test: | Chi-square=1784.859 | Degrees of freedom=1145 | Significance=.000 |

This means that the missing data may be related to some observed variables, and it is appropriate to use the regression method to fill in the missing values. Therefore, the regression method adjusted by residual estimation is used by SPSS 25.0 to complete the missing values to obtain the final dataset.

#### 3.3.3 Data Standardization

By scaling the data to the [0, 1] interval, the impact of feature scale differences on model performance is reduced. The data is then normalized, thereby improving the convergence speed and prediction accuracy of the model. The specific calculation formula is as follows:

$$x' = \frac{x - min}{max - min} \text{ (Formula 1)}$$

### *3.4 Data Variable Naming*

In the processed dataset, "China: GDP: Constant Price: Year-on-Year Growth in Current Quarter" is the prediction indicator for this study, namely the endogenous variable Y, and the remaining 47 indicators are possible influencing factors, namely exogenous variables, which are named sequentially as $x_i$.

## 4. Further Data Analysis

### *4.1 Correlation Analysis of GDP Growth Rate Based on Grey Relational Degree*

In order to study the degree of influence and correlation between various economic indicators on the quarterly year-on-year growth rate of GDP, this paper selects the quarterly year-on-year growth rate of GDP as the reference sequence for grey relational analysis $X_0$, and other economic indicators as the comparison sequence $X_i$. The gray relational coefficient is calculated according to the following formula to measure the similarity between two sequences.

$$\xi_{ij}(k) = \frac{min_i min_k |x_0(k) - x_i(k)| + \rho max_i max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho max_i max_k |x_0(k) - x_i(k)|} \text{ (Formula 2)}$$

Where ρ is the resolution coefficient, usually taken as 0.5. The relationship coefficient matrix obtained through the above formula is calculated and visualized, as shown in the figure below.

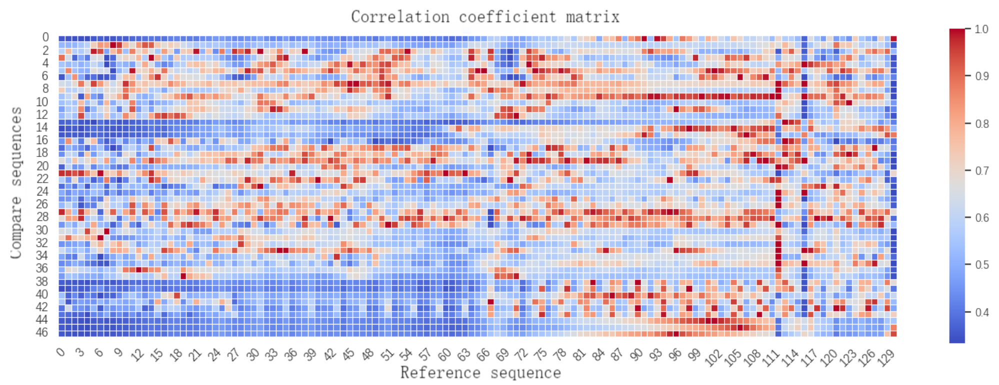


Figure 1. Heat Map of Grey Relational Coefficients

In order to better compare the correlation between each exogenous variable and endogenous variable, we will calculate the gray relational degree based on the relational coefficient. The relational degree is to concentrate the relational coefficients of each moment into one value, that is, find its average value as the quantity representation of the degree of association between the comparison sequence and the reference sequence. The calculation formula is:

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi(X_i) \quad \text{(Formula 3)}$$

The calculation results of gray relational degree show that the correlation between sequence 14 (Philadelphia Semiconductor Index) and the reference sequence is 0.48 (<0.5). The correlation between the remaining sequences and the reference sequence is greater than 0.5, indicating that except for sequence 14, the rest of the calculated sequences selected in this study have a medium or higher correlation with the reference sequence. The gray relational degree results are shown in Figure 2 after being sorted.

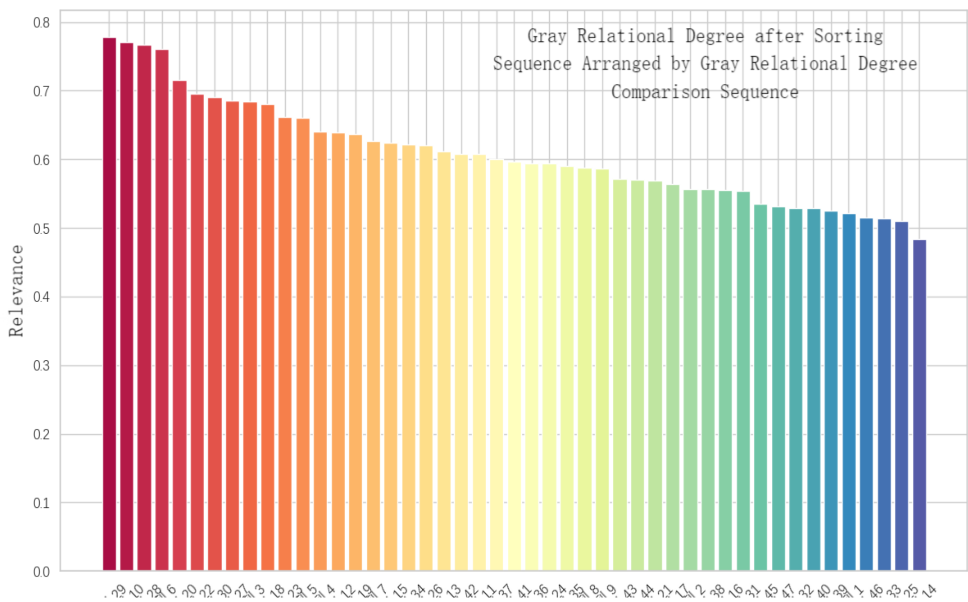


Figure 2. Gray Relational Degree after Sorting

The correlation between China's Manufacturing PMI new orders and the quarterly year-on-year growth rate of GDP exceeds 0.7, showing a strong correlation, indicating that manufacturing new orders are an important leading indicator of economic growth. An increase in new orders indicates an increase in market demand, which promotes production activities and promotes economic growth through supply chain effects. The year-on-year growth of total retail sales of consumer goods is closely related to GDP growth, emphasizing the core role of consumption in economic growth.

As a comprehensive indicator, China's Manufacturing PMI highlights the important position of the manufacturing industry in the economy. Its fluctuations reflect economic stability and growth rate in a timely manner. The high correlation between the year-on-year growth of import amount and GDP growth shows that domestic demand is growing, usually in sync with economic expansion. The year-on-year increase in the production of cement and natural crude oil in China indicates an increase in infrastructure construction and energy demand, reflecting the expansion of economic activities.

The high correlation between manufacturing employment and PMI reflects the expansion of production activities, promoting income and consumption growth. The growth of public fiscal revenue reflects an increase in economic activity and helps government investment in infrastructure and public services. Rising prices of non-food consumer goods reflect changes in consumer purchasing power, reflecting the health of the economy. The year-on-year growth of power generation is usually synchronized with economic activity, showing an increase in production and consumption activities.

Through gray relational degree analysis, we can identify key economic factors. Although the analysis itself cannot directly predict economic trends, it helps formulate economic policies and understand the factors affecting economic changes.

*4.2 Setting of Indicator Factor Parameters Based on Principal Component Analysis*

By using SPSS 25.0 to perform principal component analysis on 47 exogenous variables, we use the correlation matrix to observe the correlation between each variable, and perform KMO test and Bartlett test.

Table 2. KMO & Bartlett Test

| **KMO and Bartlett Test** | | |
|---|---|---|
| KMO Measure of Sampling Adequacy | | .833 |
| Bartlett Test of Sphericity | Approximate Chi-Square | 8718.850 |
| | Degrees of freedom | 1081 |
| | Significance | .000 |

The test results are shown in Table 2 above. It can be seen that the KMO value is 0.833. At the same time, the results of Bartlett's sphericity test show that the significance P value is 0.000*, which is significant at the level, rejecting the original hypothesis. The variables are related and the principal component analysis is effective, and the degree is very suitable.

At the same time, the variance of all variables can be explained by the principal component is more than 65%, and most of them are more than 80%, indicating that most variables can be well explained by the principal component.

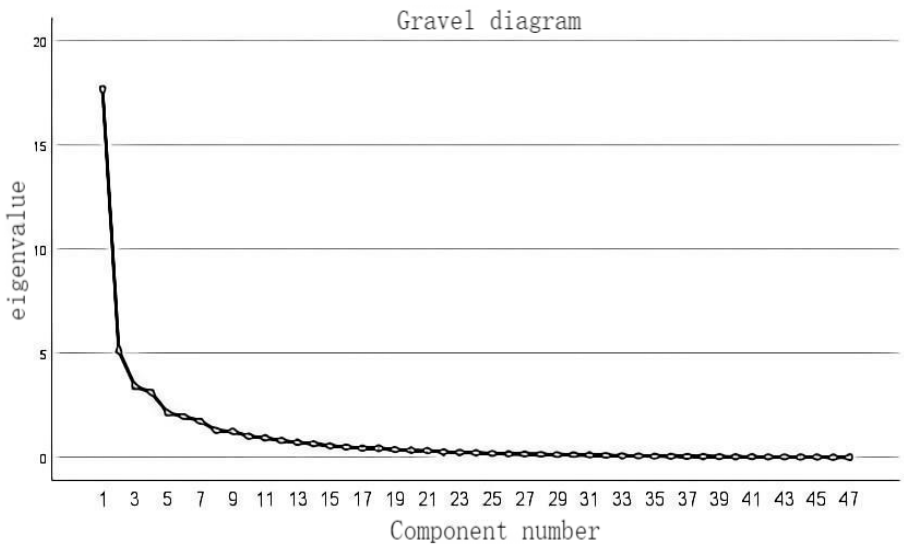


Figure 3. Principal Component Scree Plot

According to Figure 3. Principal Component Scree Plot, it can be determined to extract 10 principal components, and the cumulative variance explanation percentage of the 10 principal components reaches 82.347% (> 80%). This indicates that Principal Component Analysis (PCA) effectively captures the main variability in the data, and the extracted principal components are sufficient to represent most of the variability in the dataset.

According to Formula 4, the linear combination relationship of each principal component on the original variable can be calculated, where $\mathrm{a_{ki}}$ represents the loading of the k-th principal component on $\mathrm{x_i}$, which represents the degree of contribution of $\mathrm{x_i}$ to the principal component $\mathrm{PC_k}$. The larger the absolute value of the loading $\mathrm{a_{ki}}$, the more important the variable is in that principal component. The positive and negative values indicate whether $\mathrm{x_i}$ has a positive or negative impact on $\mathrm{PC_k}$.

$$PC_k = a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{ki}x_{\mathrm{i}} \quad \text{(Formula 4)}$$

Due to space limitations, the linear combination formula of the first principal component is given here, and the linear combination formulas of the remaining principal components are the same.

$$PC_1 = 0.118x_1 + 0.012x_2 + 0.009x_3 - 0.004x_4 - 0.026x_5 + 0.003x_6 - 0.027x_7 + 0.009x_8 + 0.023x_{10} + 0.017x_{11} + 0.020x_{12} + 0.072x_{12} + 0.037x_{13} - 0.069x_{14} - 0.005x_{15} + 0.015x_{16} - 0.068x_{17} - 0.007x_{18} - 0.013x_{19} + 0.021x_{20} + 0.028x_{21} + 0.040x_{22} + 0.015x_{23} + 0.023x_{24} - 0.001x_{25} - 0.023x_{26} - 0.007x_{27} + 0.000x_{28} + 0.009x_{29} + 0.036x_{30} + 0.024x_{31} + 0.028x_{32} + 0.012x_{33} + 0.008x_{34} - 0.016x_{35} + 0.019x_{36} + 0.001x_{37} + 0.038x_{38} + 0.176x_{39} + 0.191x_{40} + 0.041x_{41} + 0.239x_{42} + 0.240x_{43} + 0.208x_{44} - 0.018x_{45} - 0.025x_{46} + 0.001x_{47} \quad \text{(Formula 5)}$$

By observing the contribution values of each original variable to $PC_1$, it can be seen that the total contribution value of all original variables to $PC_1$ is 2.042 (the sum of the absolute values of the contribution values of all original variables). As shown in Figure 4 , the original variables with an absolute cumulative contribution rate of more than 60% are selected, which are $x_{43}$ (11.75%), $x_{42}$ (11.70%), $x_{44}$ (10.19%), $x_{40}$ (9.35%), $x_{39}$ (8.62%), $x_1$ (5.78%), $x_{12}$ (3.53%), namely, the cumulative value of the completed floor area of houses in China, the cumulative value of the newly started floor area of houses in China, the cumulative value of the sales area of commercial houses in China, the cumulative value of the total profit of industrial enterprises in China, the cumulative value of the completed investment in fixed assets in China, the monthly value of total retail sales of consumer goods in China, and China's M1 year-on-year, with a cumulative contribution rate of 60.92%. At the same time, according to the positive and negative nature of the component score, it can be known that these variables have a positive impact on $PC_1$.

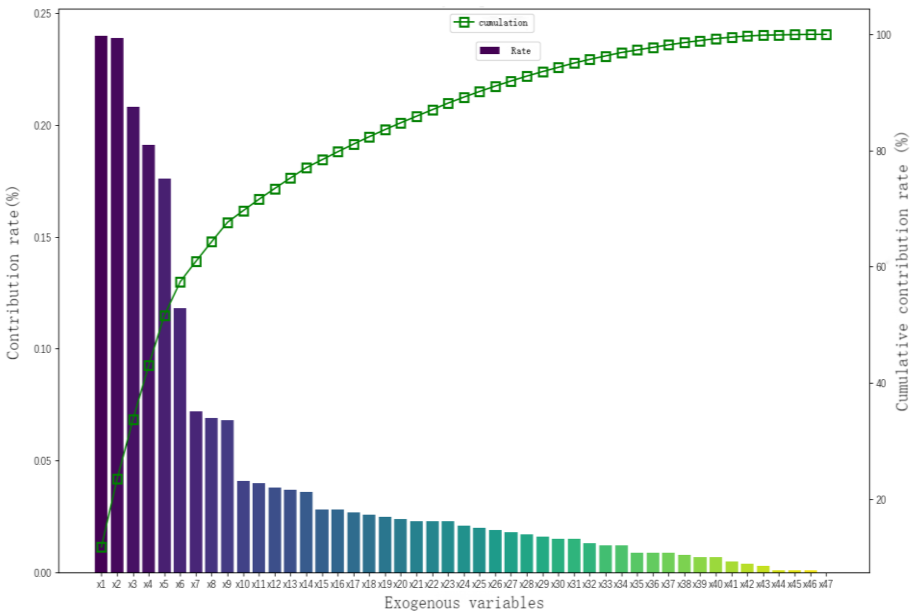


Figure 4. Pareto Chart of Variable Contribution Rate of $PC_1$

Combining the economic meaning and practical significance of the original variables, it is considered that the first principal component can be interpreted as the China Macroeconomic Vitality Index. Overall, PC1 may represent a comprehensive indicator of China's real estate market, industrial production, consumer market, and money supply. The positive coefficients of X43 (cumulative value of completed floor area of houses) and X42 (cumulative value of newly started floor area of houses) indicate that construction activities in the real estate market are positively correlated with the increase of PC1, which may reflect the activity of the real estate market. X44 (cumulative value of the sales area of commercial houses) also has a positive coefficient, which may be related to the sales situation of the real estate market, further emphasizing the importance of the real estate market in PC1. The positive coefficients of X40 (cumulative value of completed investment in fixed assets) and X39 (cumulative value of total profit of industrial enterprises) may be related to the health of industry and manufacturing, as fixed asset investment and corporate profits are key indicators for measuring these industries. The positive coefficient of X1 (the monthly value of total retail sales of consumer goods) may reflect the state of the consumer market, which is another important driver of economic growth. As an indicator of the money supply, the positive coefficient of X12 (M1 year-on-year) may be related to the increase in economic liquidity, which may have an impact on overall economic activity. An increase in M1 may mean that more liquidity is circulating in the economy, which may stimulate investment and consumption, thus having a positive impact on PC1.

Based on the economic indicators contained in PC1 and their component score coefficients, we can summarize PC1 as the "China Macroeconomic Vitality Index". This index comprehensively reflects the economic activities in key areas such as the real estate market, industrial production, consumer market, and money supply. It is an important indicator for measuring China's overall economic vitality and growth potential.

Similarly, the main influencing variables and principal component meanings of the remaining principal components are shown in the table below.

Table 3. Principal Component Explanation Table

| Principal Component | Principal Component Meaning | Main Influencing Variables | Cumulative Contribution Rate |
|---|---|---|---|
| $PC_1$ | China Macroeconomic Vitality Index | $x_{43}$ 、 $x_{42}$ 、 $x_{44}$ 、 $x_{40}$ 、 $x_{39}$ 、 $x_1$ 、 $x_{12}$ | 60.92% |
| $PC_2$ | China Manufacturing Vitality and Macroeconomic Health Index | $x_{28}$ 、 $x_{19}$ 、 $x_{29}$ 、 $x_{18}$ 、 $x_{30}$ 、 $x_{35}$ 、 $x_{21}$ 、 $x_{15}$ 、 $x_{25}$ 、 $x_{12}$ 、 $x_{33}$ 、 $x_6$ | 62.57% |
| $PC_3$ | China Economic Liquidity and Market Comprehensive Index | $x_{38}$ 、 $x_{37}$ 、 $x_{13}$ 、 $x_{15}$ 、 $x_{34}$ 、 $x_{26}$ 、 $x_2$ 、 $x_{32}$ 、 $x_{17}$ 、 $x_{43}$ 、 $x_{20}$ | 60.69% |
| $PC_4$ | China Industrial and Financial Market Comprehensive Index | $x_{17}$ 、 $x_{14}$ 、 $x_{46}$ 、 $x_{11}$ 、 $x_{45}$ 、 $x_{47}$ 、 $x_{42}$ 、 $x_{43}$ 、 $x_{25}$ 、 $x_{41}$ 、 $x_{16}$ 、 $x_{44}$ 、 $x_{21}$ 、 $x_3$ | 61.51% |
| $PC_5$ | China Economic Activity and Market Trend Index | $x_{31}$ 、 $x_{32}$ 、 $x_9$ 、 $x_{38}$ 、 $x_{37}$ 、 $x_8$ 、 $x_{17}$ 、 $x_{20}$ 、 $x_{23}$ 、 $x_{11}$ 、 $x_2$ 、 $x_{35}$ | 60.47% |
| $PC_6$ | China Economic Domestic and External Demand Balance and Price Stability Index | $x_7$ 、 $x_4$ 、 $x_3$ 、 $x_5$ 、 $x_{21}$ 、 $x_{23}$ 、 $x_2$ 、 $x_{25}$ 、 $x_{35}$ 、 $x_{10}$ | 61.34% |
| $PC_7$ | China Economic Activity and Fiscal Dynamic Index | $x_{35}$ 、 $x_{36}$ 、 $x_{10}$ 、 $x_{23}$ 、 $x_{11}$ 、 $x_{29}$ 、 $x_{12}$ 、 $x_{27}$ 、 $x_{26}$ 、 $x_{20}$ 、 $x_{25}$ | 60.54% |
| $PC_8$ $PC_8$ | China Macroeconomic Policy and Market Vitality Index | $x_{26}$ 、 $x_{11}$ 、 $x_{27}$ 、 $x_{25}$ 、 $x_{15}$ 、 $x_{34}$ 、 $x_5$ 、 $x_{12}$ 、 $x_{19}$ 、 $x_{37}$ 、 $x_{23}$ 、 $x_{38}$ | 62.38% |
| $PC_9$ $PC_9$ | China Macroeconomic Comprehensive Index | $x_{24}$ 、 $x_{19}$ 、 $x_{33}$ 、 $x_6$ 、 $x_{34}$ 、 $x_{12}$ 、 $x_9$ 、 $x_7$ 、 $x_{35}$ 、 $x_{20}$ 、 $x_{26}$ 、 $x_{11}$ 、 $x_{31}$ 、 $x_{15}$ | 62.36% |
| $PC_{10}$ $PC_{10}$ | China Economic Energy and Market Dynamic Index | $x_{22}$ 、 $x_{34}$ 、 $x_{23}$ 、 $x_{35}$ 、 $x_{11}$ 、 $x_3$ 、 $x_5$ 、 $x_2$ 、 $x_{15}$ 、 $x_{33}$ 、 $x_{27}$ | 60.44% |

Through an in-depth analysis of the ten principal components of the Chinese economy, we can gain a comprehensive perspective. These principal components together outline the multi-dimensional profile of the Chinese economy, reflecting different economic trends and dynamics. PC1, as the China Macroeconomic Vitality Index, focuses on the activity of the real estate market, the growth of industrial production, the expansion of the consumer market, and the changes in the money supply. These factors together depict the overall growth potential and health of the Chinese economy. PC2 represents a comprehensive indicator of China's manufacturing industry and economic activity, reflecting the overall health and growth potential of the economy, especially the growth of manufacturing activities and market demand. PC3 covers a comprehensive trend of many aspects such as financial market liquidity, money supply, stock market performance, real estate market, government spending, inflation, industrial product prices, and construction industry activities. It is an important indicator for measuring China's economic liquidity and market vitality. PC4 integrates indicators of industrial product prices, the semiconductor industry, financial system liquidity, real estate market activities, automobile industry sales, stock market performance, and non-food consumer product prices, reflecting the overall situation and trend of industrial and financial markets in China's economy. PC5 is related to the real estate market, fixed asset investment, financial credit, money supply, inflation, and passenger traffic, revealing how these factors jointly affect economic trends. PC6 reflects the comprehensive trend in the Chinese economy related to export-oriented industries, price stability, industrial production, and consumer market dynamics. PC7 focuses on logistics and supply chain activities, consumer market dynamics, industrial production, fiscal policy, and the situation in the construction and automobile industries. PC8 integrates indicators from multiple aspects such as fiscal policy, money supply, financial markets, real estate markets, export activities, and industrial production to form an index that reflects macroeconomic policies and market vitality. PC9 covers comprehensive trends in industrial enterprise profits, real estate market, import activities, money supply, fixed asset investment, export prices, passenger traffic, and public fiscal expenditures. Finally, PC10, as the China Economic Energy and Market Dynamic Index, highlights the impact of energy supply, the real estate market, industrial production, and consumer prices on economic stability.

### *4.3 Summary*

Through the comprehensive use of Grey Relational Analysis (GRA) and Principal Component Analysis (PCA), the key economic indicators affecting the quarterly year-on-year growth rate of China's GDP are deeply identified and evaluated. The analysis results show that most economic indicators have a correlation degree greater than 0.5 with GDP growth rate, indicating that there is a medium or above correlation between them and GDP growth rate. Among them, indicators such as the Manufacturing PMI new orders and the year-on-year growth of total retail sales of consumer goods have a correlation degree exceeding 0.7 with GDP growth rate, showing a strong correlation.

The principal components extracted by PCA construct an index that reflects China's economic vitality. The first principal component is defined as the "China Macroeconomic Vitality Index", which focuses on economic activities in areas such as real estate, industrial production, consumer market, and money supply. This index is an important tool for measuring China's overall economic vitality and growth potential. The other principal components extracted by PCA also reveal different aspects of the Chinese economy, providing a multi-dimensional perspective.

Combining the results of GRA and PCA, it is possible to identify the most critical economic indicators for predicting GDP growth. Indicators such as the Manufacturing PMI new orders and the total retail sales of consumer goods reveal their importance as leading indicators, especially emphasizing the core role of consumption in economic growth. At the same time, indicators such as the Manufacturing PMI, the import amount, and the production of cement highlight the core position of manufacturing and infrastructure construction.

## 5. Instant Prediction Analysis Model Based on Neural Network

### *5.1 Selection and Establishment of Instant Prediction Model*

According to the characteristics of the data, this paper constructs a RNN neural network model represented by GRU through Python 3.8 to fit and evaluate the results, so as to obtain an instant prediction model for predicting the "quarterly year-on-year growth rate of China's GDP".

#### 5.1.1 Determination of Input and Output Variables

For the GRU neural network, which is suitable for small samples and can handle multi-dimensional input, by comparing the impact of data variables before and after input dimensionality reduction on the model fitting and prediction effect, we conclude that using the data before dimensionality reduction for GRU neural network modeling can more accurately fit and predict the quarterly year-on-year growth rate of China's GDP. Therefore, for GRU, we use the data before dimensionality reduction.

Retaining the original time series and combining the past values of the target time series for modeling analysis can better analyze and explain the influence relationship between various economic indicators and China's quarterly year-on-year GDP growth rate, better reduce information loss and errors, and increase the model fitting effect. Therefore, the input variable is selected as the past values of the quarterly year-on-year growth rate of China's GDP $x_0$ and the 47 economic indicators of the dataset $x_1 \sim x_{47}$, and the output variable of the model, which is the predicted variable, is the quarterly year-on-year growth rate of China's GDP y.

5.1.2 Hyperparameter Presetting

For the construction of the GRU neural network, the setting of various hyperparameters is extremely important, and they all more or less affect the fitting effect of the neural network on the data. The preliminary settings and meanings of related hyperparameters are shown in the table below.

Table 4. Hyperparameter Presetting

| Hyperparameter Name | Meaning | Preliminary Presetting |
|---|---|---|
| Hidden Layer Size （Hidden Size） | One of the most important parameters, it determines the dimensionality of the hidden state | 128 |
| Number of Layers （Number of Layers） | Stack multiple layers to capture more complex sequence features. However, increasing the number of layers may also lead to overfitting of the model | 2 |
| Learning Rate （Learning Rate） | A parameter that controls the step size of model weight updates. A small learning rate may lead to slow model training, while a large learning rate may lead to training instability | 0.0001 |
| Batch Size （Batch Size） | Determines the number of samples used each time the model weights are updated. Smaller can provide more noise, which helps the model to generalize, but it may also make the training process more noisy. Larger can speed up training, but requires more memory and may cause the model to fall into a local minimum. | 40 |
| Sequence Length （Sequence Length） | Sequence length is a crucial parameter when dealing with time series data. It determines the number of time steps included in each input sequence. The choice of sequence length should be based on the characteristics of the data and the complexity of the model. | 20 |
| L1/L2 Regularization （Regularization） | To prevent overfitting, L1/L2 regularization terms can be added to the model. | L2(0.01) |
| Dropout | It represents the proportion of neurons that are randomly dropped (i.e., set to zero) during the training process. This forces the network to learn more robust features, thereby improving the model's generalization ability. | 0.2 |

Based on the table above, this study initially sets the number of nodes in the hidden layer to 128. This value is derived from experimental verification and literature review. The selected number of network layers is 2, and the initial learning rate is 0.001. These are conventional starting parameter values and can be adjusted according to the model's performance during training. At the same time, this study selects a batch size of 18. This value is determined based on the availability of computing resources and empirical judgments, and the purpose is to achieve a balance between training speed and model generalization performance. Based on the characteristics of the dataset and the experimental results, this study chooses a sequence length of 20 to ensure that the model can fully capture the dependencies in the time series data while avoiding the introduction of too many parameters. In addition, this study adopts an L2 regularization strategy, and its regularization coefficient is set to 0.01, aiming to achieve a balance between preventing overfitting and maintaining model performance through experiments and literature review. Finally, this study chooses a dropout ratio of 0.2, that is, 20% of neurons are randomly discarded in each iterative training process to improve the generalization ability of the model.

After initially setting these hyperparameters, we were able to build a high-performance Gated Recurrent Unit (GRU) model, which is used to predict the quarterly year-on-year growth rate of China's Gross Domestic Product (GDP). Although these parameter settings provide an initial good foundation for model optimization, the determination of the actual optimal parameter values still depends on subsequent experimental adjustments and optimization processes.

5.1.3 Model Prediction Evaluation and Correction

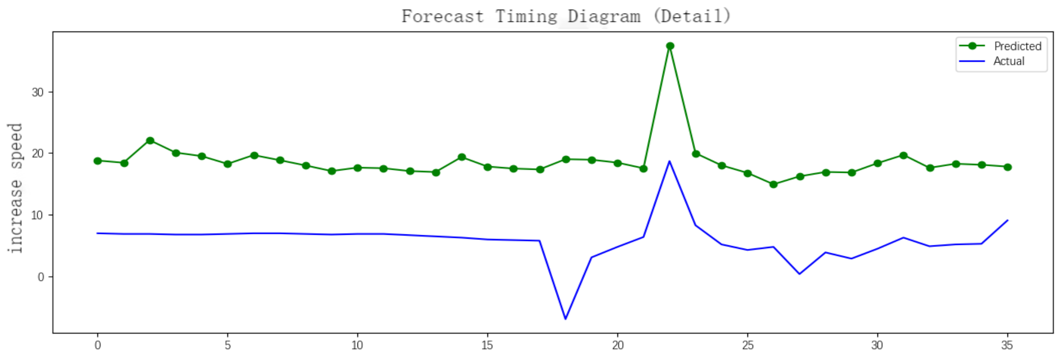


Figure 5. GRU Prediction Plot Before Correction

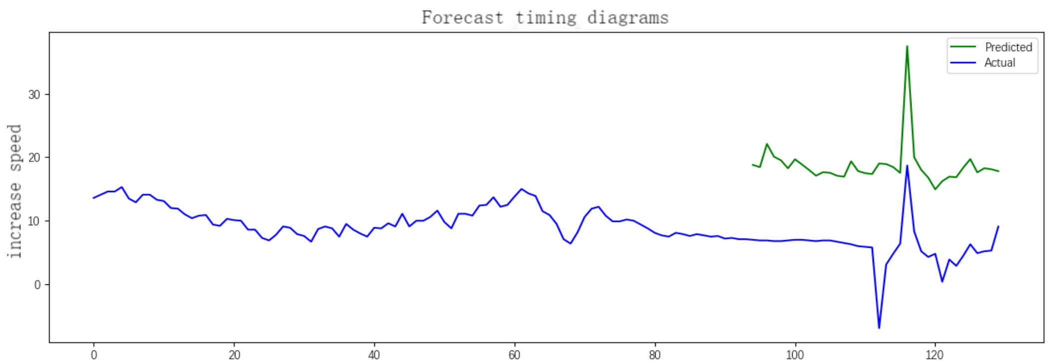


Figure 6. GRU Prediction Plot Before Correction

It can be seen from the figure that the prediction results of GRU well fit the fluctuations of China's quarterly year-on-year GDP growth rate. However, the prediction time series line does not accurately fit the predicted value, but is on average higher than the actual time series line. Therefore, we added error correction to the GRU neural network model. By calculating the MSE of the prediction model, the final error-corrected predicted value is obtained by subtracting the MSE from the predicted value of the model.

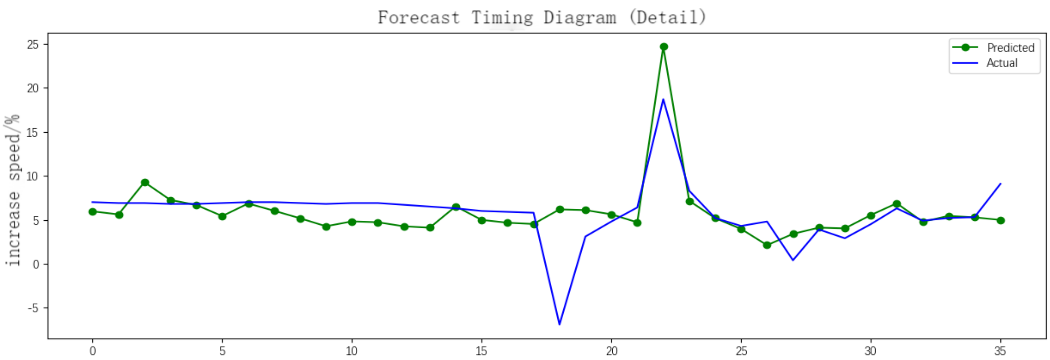


Figure 7. Prediction Plot After Correction (Partial)

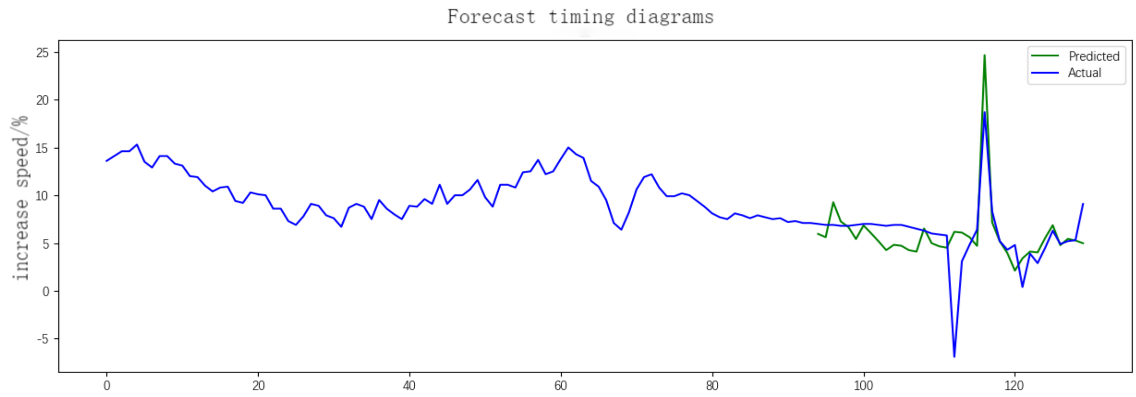


Figure 8. Prediction Plot After Correction

As can be seen from the figure, the predicted time series line after error correction basically overlaps with the actual time series line, which not only better fits the volatility of China's quarterly GDP growth rate time series, but also more accurately predicts future values. In order to more objectively evaluate the model fitting effect, we selected the following evaluation indicators: SMAPE, MAE, and RMSE.

Table 5. GRU Evaluation Indicators Before and After Error Correction

| Comparison Before and After Error Correction | SMAPE | MAE | RMSE |
|---|---|---|---|
| GRU Before Correction | 47.4537% | 6.718234 | 10.902488 |
| GRU After Correction | 25.5428% | 1.525912 | 3.082478 |

As can be seen from the table above, the SMAPE, MAE, and RMSE of the GRU neural network instant prediction model after error correction are all lower than the corresponding evaluation indicators of the GRU neural network instant prediction model before correction. Among them, SMAPE decreased by 21.9109%, MAE decreased by 5.192322, and RMSE decreased by 7.82001. It can be seen that the instant prediction model after error correction has a better fitting effect on China's quarterly year-on-year GDP growth rate. And because the model evaluation indicators after error correction are SMAPE of 25.5428%, MAE of 1.525912, and RMSE of 3.082478, the GRU neural network instant prediction model after error correction can be considered to be relatively good.

### *5.2 GRU Neural Network Based on Bayesian Optimization*

#### 5.2.1 Neural Network Architecture after Bayesian Optimization

Table 6. Hyperparameters of the Bayes-GRU Neural Network Architecture

| **Hyperparameter** | Hidden Layer | Number of Layers | Learning Rate | Batch Size | Sequence Length | L1/L2 Regularization | Dropout |
|---|---|---|---|---|---|---|---|
| **Setting** | 2304 | 2 | 0.008 | 40 | 20 | L2(0.01) | 0.2 |

As shown in the table above, Bayesian optimization performs 200 hyperparameter tuning on the GRU neural network and selects the local optimal hyperparameters for setting. It can be seen that the hyperparameter settings of the Bayes-GRU neural network are hidden layer size (128), number of layers (2), learning rate (0.008), batch size (40), sequence length (20), L2 (0.01), Dropout (0.2).

#### 5.2.2 Bayes-GRU Model Prediction Analysis

Apply the Bayes-GRU instant prediction model with Bayesian hyperparameter selection to fit and train the quarterly year-on-year growth rate of China's GDP. Similarly, use a 7:3 ratio to divide the training set and the verification set, visualize the predicted values of the prediction model on the verification set, and compare the prediction plot of Bayes-GRU with the prediction plot of GRU before Bayesian optimization, as shown in the figure below.

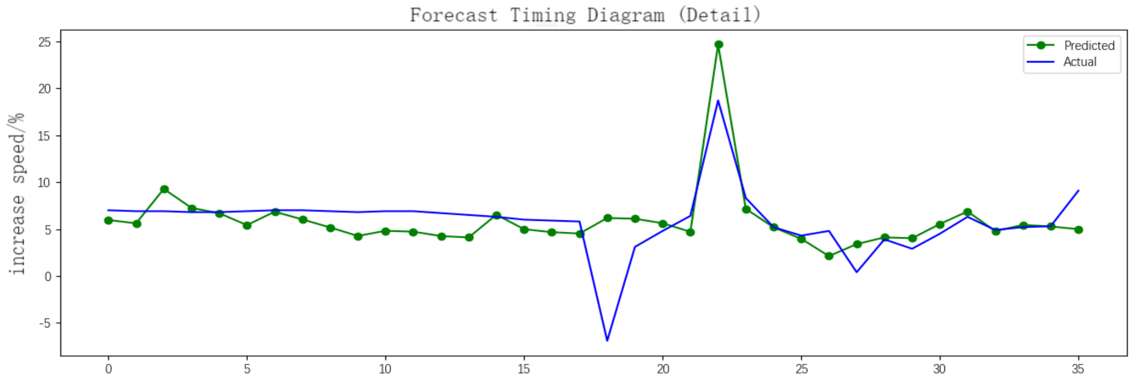


Figure 9. GRU Prediction Time Series Plot Before Optimization (Partial)

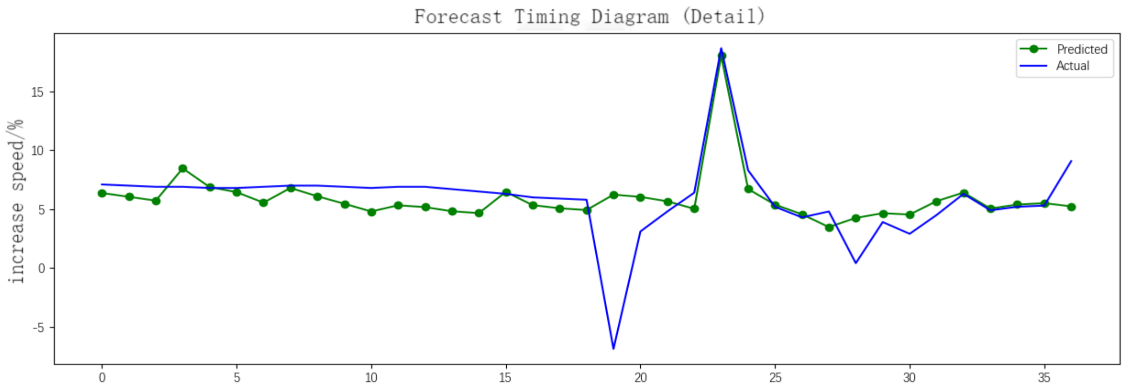


Figure 10. Bayes-GRU Prediction Time Series Plot (Partial)

From Figure 9 and Figure 10, it can be seen that the Bayes-GRU instant prediction model maintains good performance in effectively capturing the volatility of the quarterly year-on-year growth rate of China's Gross Domestic Product (GDP) compared with the previous model. At the same time, the model reduces the average error between the predicted value and the actual value to a certain extent. It can be seen that the model optimized by Bayes has shown some improvement in the prediction effect.

Table 7. GRU Evaluation Indicators Before and After Bayesian Optimization

| Comparison Before and After Bayes Optimization | SMAPE | MAE | RMSE |
|---|---|---|---|
| GRU Before Optimization | 25.5428% | 1.525912 | 3.082478 |
| GRU After Optimization | 23.3746% | 1.389350 | 2.623492 |

We compare the evaluation indicators of the models before and after Bayesian optimization. It can be seen from the table above that the prediction indicators of the optimized instant prediction model have been significantly optimized. Among them, the optimized SMAPE is 23.3756%, a decrease of 2.1682%; MAE is 1.389350, a decrease of 0.136562; RMSE is 2.623492, a decrease of 0.458986. All indicators show that the instant prediction model is good.

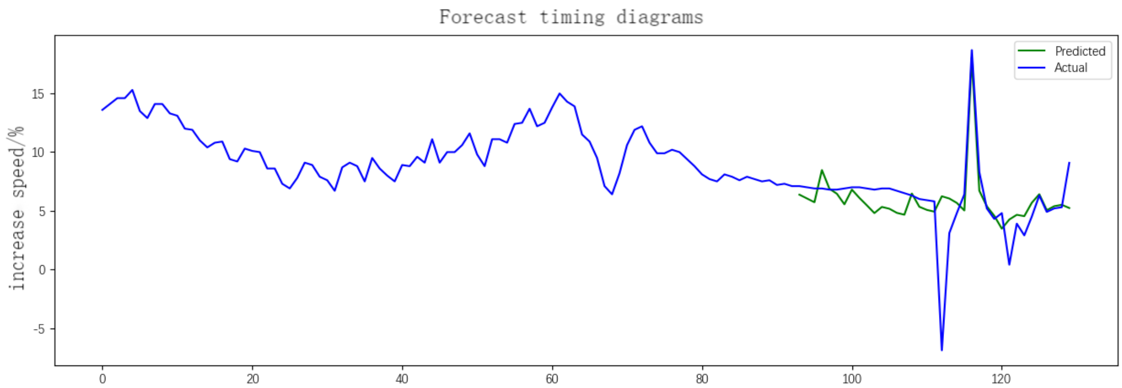


Figure 11. Bayes-GRU Prediction Time Series Plot

We combine Figure 11: Bayes-GRU Prediction Time Series Plot analysis and see that in the first half of the verification set, the time series trend of China's quarterly year-on-year GDP growth rate is relatively flat, without obvious volatility, but the prediction time series line shows a certain degree of volatility during this period, and the fluctuation interval is within $\pm$ 1.3%. In the second half of the verification set, the volatility of the time series is well captured, and the prediction is more accurate.

*5.3 Analysis of Indicator Contribution Based on Prediction Model*

The SHAP contribution values of each influencing indicator to the predicted value in the prediction model were obtained by establishing a Gated Recurrent Unit (GRU) neural network model to predict China's quarterly year-on-year GDP growth rate.

The top ten most significant economic indicators in the model play an important role in the analysis. Their changes directly or indirectly affect multiple levels of the economy. These indicators reveal the complexity of economic dynamics, from money supply to consumer demand, and from production costs to industrial profitability.

M1 growth reflects monetary liquidity. Higher growth usually indicates economic expansion, while lower growth may indicate an economic slowdown. The Nanhua Industrial Products Index measures industrial activity and price fluctuations, directly affecting production costs and predicting economic cycle turning points. The CPI non-food item year-on-year accurately reflects changes in consumer prices. High CPI indicates high inflationary pressure, which may suppress consumption, while low growth shows weak demand. The PPI reflects changes in production prices. Higher PPI indicates stronger industrial profitability, while lower PPI may indicate insufficient demand or overcapacity, affecting production decisions.

The PMI employment data reflects the employment situation in the manufacturing industry. Increasing the number of employees usually means an improvement in the prosperity of the manufacturing industry. The loan balance of financial institutions reflects the financing demand. An increase in the balance usually indicates economic expansion. GDP payment items reveal the distribution of economic output and help understand changes in economic structure. The year-on-year growth of commercial housing sales area reflects fluctuations in real estate demand, which affects related industries and consumer confidence.

These indicators not only directly reflect the current economic situation, but also provide important clues for predicting future trends, helping policymakers and market decision-makers make scientific decisions and foresee economic turning points.

## 6. Conclusion

This paper analyzes China’s quarterly year-on-year GDP growth rate through Grey Relational Analysis, Principal Component Analysis (PCA), and a prediction model based on the GRU neural network. The study first identifies economic indicators that are highly correlated with GDP growth rate through Grey Relational Analysis (GRA), and then uses Principal Component Analysis (PCA) to extract key variables from multi-dimensional data and construct a "China Macroeconomic Vitality Index". Based on the GRU model, combined with error correction and Bayesian optimization techniques, the prediction accuracy is significantly improved.

The results show that the optimized GRU model has significant improvements in evaluation indicators such as SMAPE, MAE, and MSE, and the prediction effect is significantly better than traditional methods. In addition, SHAP value analysis further reveals the specific impact of each economic indicator on GDP growth, which provides strong support for macroeconomic decision-making. This study provides a new methodological framework for GDP growth forecasting, and also provides data support and theoretical basis for macroeconomic analysis and policy formulation.

## References

[1] Song, J. (2016). Application of time series in GDP prediction in Anhui Province—Based on ARIMA model. *Modern Business, (21)*, 134–136.

[2] Sun, W., Yang, J., Yang, Y., et al. (2018). Analysis and prediction of China's GDP based on SARIMA model. *China Collective Economy, 2018*(36), 78–80.

[3] He, Q. (2018). Exploring the idea of using big data to predict quarterly GDP growth rate. *China Statistics, 2018*(09), 9–12.

[4] Yi, Y., Huang, D., & Wang, X. (2024). GDP nowcasting based on macro big data. *China Economic Quarterly, 24*(03), 843–860. https://doi.org/10.13821/j.cnki.ceq.2024.03.10

[5] You, X., Sun, D., & Lin, F. (2010). Application of combined model in China's GDP prediction. *Journal of Langfang Normal College (Natural Science Edition), 10*(2), 87–89.

[6] Wang, S., Chen, A., Su, J., et al. (2009). Application of combined prediction model in China's GDP prediction. *Journal of Shandong University (Natural Science Edition), 44*(2), 56–59.

[7] Zhang, J. (2016). Short-term prediction of China's GDP based on time series. *Commerce (Natural Science Edition), (15)*, 204–206.

[8] Li, N., & Xue, J. (2013). Prediction of China's GDP growth based on the optimal ARIMA model. *Statistics and Decision, (9)*, 23–26.

[9] He, Y., Sun, Y., & Gao, Y. (2024). Quarterly GDP prediction based on SARIMA-LSTM model. *Times Economy and Trade, 21*(10), 28–32. https://doi.org/10.19463/j.cnki.sdjm.2024.10.011

[10] Wu, Y., Luo, J., Zhao, Z., et al. (2024). GDP prediction of Inner Mongolia Autonomous Region based on LSTM neural network. *Journal of Inner Mongolia University of Technology (Natural Science Edition), 43*(04), 296–301. https://doi.org/10.13785/j.cnki.nmggydxxbzrkxb.2024.04.002

[11] Guo, Q., & He, Y. (2014). GDP prediction model based on DFA method and BP neural network. *Statistics and Decision, (08)*, 82–84. https://doi.org/10.13546/j.cnki.tjyjc.2014.08.030

[12] Xiao, Z., Liu, L., Zhao, T., et al. (2020). Can deep learning neural network improve GDP prediction ability? *Research on Economics and Management, 41*(07), 3–17.

[13] Liang, L., & Chen, Y. (2023). Quarterly GDP prediction based on deep learning neural network. *Statistics and Decision, 39*(02), 24–29. https://doi.org/10.13546/j.cnki.tjyjc.2023.02.005

[14] Yu, L. (2017). Application of ARIMA model in China's GDP prediction. *Times Finance, (21)*, 180+185.

[15] Xue, Q., Mou, F., & Tu, Z. (2017). Application of combined prediction methods in Chongqing GDP prediction. *Journal of Chongqing Technology and Business University (Natural Science Edition), 34*(1), 56–63.

[16] Yuan, Y. (2021). Research on the measurement of China's quarterly GDP growth rate based on the MIDAS model (Doctoral dissertation). Jinan University.

[17] Fu, X. (2022). Research on nowcasting China's quarterly GDP based on dynamic factor model (Doctoral dissertation). Southwestern University of Finance and Economics.

[18] Ding, L. (2010). Time series model and prediction of GDP per capita in Heilongjiang Province. *Journal of Harbin Normal University (Natural Science Edition), 26*(6), 16–18.

## Author Biographies

1. Min Juanhua (1988- ), female, Master of Economics, Lecturer of the School of Applied Economics of Guangdong Baiyun University, China. Mailing Address: No. 7, Jing'an Street, Binjiang Middle Road, Haizhu District, Guangzhou City, Guangdong Province, A2402. Postal Code: 510220. Tel: 086-136-3136-7165. E-mail: min5612at163.com

2. Xu Donglin (2004- ), male, Guangdong Baiyun University, China.