Voice of Climate: Focus on GGDP

Shijia Liu
Mathematics and Artificial Intelligence Department, Qilu University of Technology, Jinan,Shandong province250399

Abstract: Green development has become a global goal for sustainable development. However, traditional GDP is unable to effectively reflect the level of economic growth of an economy, let alone consider its relationship with the natural environment and ecosystems.

Regarding the first issue, this paper proposes a suitable GGDP calculation method, which includes traditional GDP, the value of natural resource depletion, the value of environmental pollution, as well as the benefits of resource and environmental improvement. We use partial least squares regression analysis (PLS) to model and accurately quantify the impact of GGDP variables on climate response indicators. The results show that the selected GGDP method can significantly correlate and reflect climate change.

Regarding the second issue, this paper uses dynamic multivariate time series models (ARIMAX) and vector error correction models (VECM) to predict the impact of China's climate mitigation. Cointegration tests were performed to determine the long-term equilibrium relationship among these indicators, and residual stationary white noise tests were conducted. The future 10-year GGDP was estimated using the quadratic curve estimation method, and future changes in climate indicators for the next 10 years were predicted using a multivariate time series model. The research findings indicate that using GGDP instead of GDP has a positive impact on global climate mitigation.

Keywords: PLS, ARIMAX, VECM, Residuals, Stationarity and White Noise Test t

1. Introduction

1.1 Background

Our task This problem requires us to mathematically model and analyze the predicted impact of GGDP on climate, and select the best GGDP calculation model. Therefore, our work includes the following:

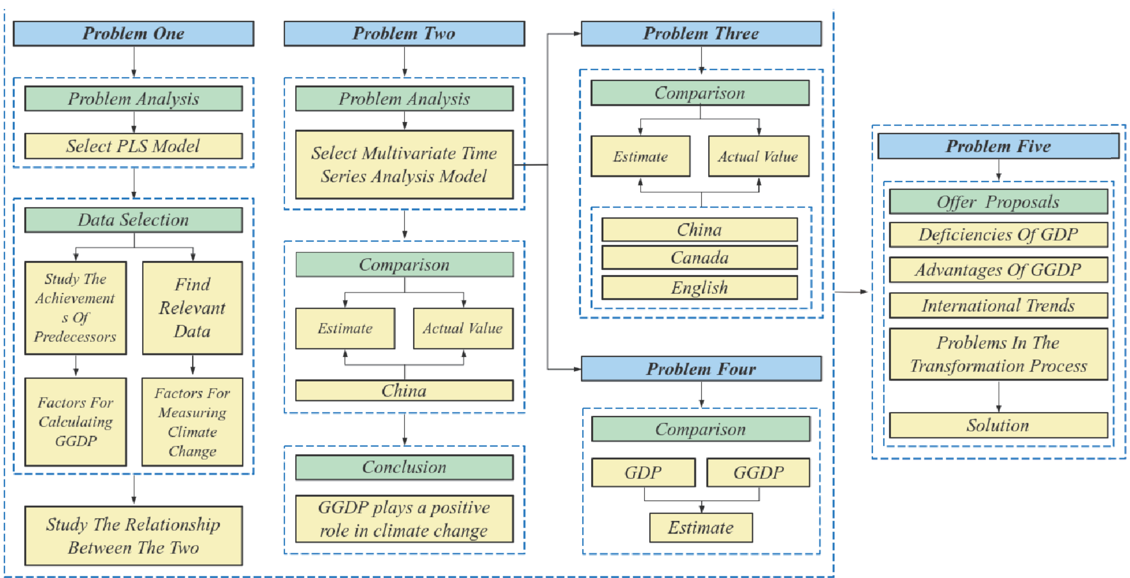


Fig.1Technology roadmap for this paper

2. The Description of the Problem

2.1 Problem statement

●Choose the most suitable method for calculating GGDP among the methods already proposed, replacing GDP as the primary measure of economic health, and discuss the measurable impact it may have on mitigating climate change.

●Develop a robust model to estimate the expected impact of using GGDP as the primary indicator for measuring a country's economic health on mitigating climate change.

●Compare the potential benefits of using GGDP versus GDP as the primary indicator for measuring a country's economic health in mitigating the impact of climate change .

2.2 Analysis of Specific Issues

●Question 1 requires selecting the most suitable method for calculating GGDP and discussing its potential measurable impact on mitigating climate change. Since calculating GGDP involves multiple variables, and the indicators for measuring climate change also involve multiple variables, we use the Partial Least Squares Regression Analysis Method (PLS) of multivariate regression to model. This method allows for the precise quantification of the influence of independent variables on dependent variables.

●Question 2 requires the creation of a stable model to estimate the expected impact of the main indicator of measuring a country's economic health, GGDP, on mitigating climate change. Since climate change indicators and green GGDP indicators both vary over time, China's GGDP from 2002 to 2014 was selected as the input variable, and carbon dioxide emissions and average annual temperature were selected as the output variables for multivariate time series analysis.

3. Basic assumption

●It is assumed that there are no Black Swan events such as economic crises or systemic risks that would significantly impact a country's economic level.

●It is assumed that the calculation of GGDP does not consider the social impact, such as different energy policies and environmental policies between different countries.

●It is assumed that a country's climate change is mainly influenced by its annual average temperature, annual precipitation, and annual carbon dioxide emissions.

4. Model Preparation

4.1 Symbols

Table 1 Notations used in this paper

Symbols	Definition	Units
G	GDP	Billion RMB
G_1	GGDP	Billion RMB
G_2	Improved GGDP	Billion RMB
x_1	coal expenditure	Billion RMB
x_2	Water environment protection expenditure.	Billion RMB
x_3	Atmospheric pollution expenses	Billion RMB
x_4	Wastewater treatment cost	Billion RMB
x_5	Total economic benefits of afforestation.	Billion RMB
y_1	annual average temperature	°C
y_2	Annual precipitation	ml
y_3	Annual carbon dioxide emissions	10000 tons
a_{ij}	Observational data of the independent variable	
$\tilde{a}_{ij}$	standardized indicator value of independent variable	
$\{x_k\}$	independent variable sequence	
$\hat{\beta}_k$	Least squares estimates	
$\{\varepsilon_t\}$	Regression residual series	
$\theta(B)$	moving average coefficient polynomial	
$\phi(B)$	Autoregressive coefficient polynomia	
Cov_k	Delayintercorrelation function	
$C\rho_k$	Delayed Cross-Correlation Coefficients	

5. Models

5.1 Analysis and Solving of Question 1

Different definitions of green GDP will result in changes in accounting ideas and calculation formulas, which will also lead to significantly different accounting results. Currently, a problem in research is the significant difference in the proportion of green GDP to GDP. The unstable and inconsistent results of green GDP accounting have caused academic discussions on the motivation of green accounting.The following are some of the calculation methods for green GDP proposed by scholars, as shown in Table 2:

Table 2Calculation formulas of various green GDP

Researcher	Calculation formula
MiankunYang(2001)	Green GDP = GDP - external diseconomies + external economies
TaoPeng(2010)	Green GDP = Economic System Output - Natural Resource Depletion -Environmental quality degradation + Integrated waste utilization
Lianying Ge(2013)	Green GDP = Traditional GDP - Resource Depletion Costs -Environmental degradation loss costs + Environmental improvement gains
GuoxiaMa(2015)	Green National Accounting = GDP - Cost of Environmental Degradation + Gross Ecosystem Product

As shown in Table 3, we construct GGDP indicators including coal expenditures, water environmental protection expenditures, atmospheric pollution costs, wastewater treatment costs, and afforestation total economic benefits in China from 2002 to 2014. We also use annual average temperature (in Celsius), annual precipitation (in milliliters), and annual carbon dioxide emissions (in 10,000 tons) as variables that affect climate change and perform partial least squares regression modeling.

Table 3 Construction of green GDP indicators

Tier 1 Indicators	Secondary indicators	Unit
Depletion value of natural resources	Coal Expenses	Billion RMB
Environmental pollution depletion value	Water environmental protection expenditure Atmospheric pollution expenses Wastewater treatment costs	Billion RMB
Resource and environmental improvement benefits	Total economic return from afforestation	Billion RMB

Use x_4 x_3 x_5 x_2 x_1, to denote the independent variables coal expenditure, water environmental protection expenditure, air pollution cost, wastewater treatment cost, and total economic return from afforestation respectively.

Also, the observed data matrix of the independent variable is noted as $A=\left(a_{ij}\right)_{12*5}$ and the observed data matrix of the dependent variable is noted as $B=(b_{ij})_{12*3}$.

Step 1: Data normalization is performed to convert each indicator value to a standard indicator value $\tilde{a}_{ij}$ with.

$$\tilde{a}_{ij}=\frac{a_{ij}-\mu_j^{(1)}}{s_j^{(1)}}\quad i=1,2,...,20,\ j=1,2,3 \tag{1}$$

Among them,

$$\mu_j^{(1)}=\frac{1}{20}\sum_{i=1}^{20}a_{ij} \tag{2}$$

$$s_j^{(1)}=\sqrt{\frac{1}{20-1}\sum_{i=1}^{20}(a_{ij}-\mu_j^{(1)})^2}\ ,\ j=1,2,3, \tag{3}$$

That is, $\mu_j^{(1)}$ $s_j^{(1)}$, are the sample mean and sample standard deviation of the first independent variables x_j.

Correspondingly, call :

$$\tilde{x}_j=\frac{x_j-\mu_j^{(1)}}{s_j^{(1)}},\ j=1,2,3 \tag{4}$$

as a standardized indicator variable.

Step 2: Calculate the correlation coefficient matrix. As shown in Fig. 2, a heatmap of the correlation coefficient matrix is plotted. Fig. 2 shows that there is a strong correlation between the selected indicators. In addition, except for a negative correlation between annual precipitation and other indicators, the other indicators are positively correlated with each other.

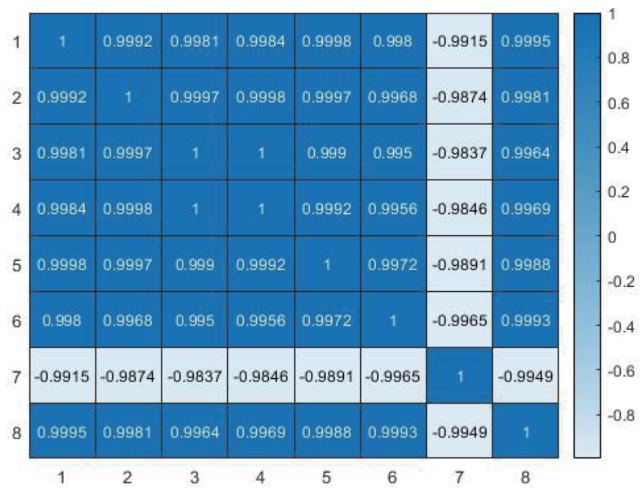


Fig.2Thermal diagram of correlation coefficient matrix

In summary, taking into account the availability and practicality of the data, the formula for calculating GGDP is defined:

$$G_1 = G - x_1 - x_2 - x_3 - x_4 + x_5 \tag{5}$$

In the formula G_1 is GGDP , x_1 is Depletion value of natural resources, x_2 + x_3 + x_4 is Depletion value of natural resources, x_5 is Resource and environmental improvement benefits.

5.2 Analysis and Solving of Question 2

Problem 2 requires the creation of a stable model to estimate the expected impact of the main indicator of measuring a country's economic health, GGDP, on mitigating climate change. Since climate change indicators and green GGDP indicators both vary over time, China's GGDP from 2002 to 2014 was selected as the input variable, and carbon dioxide emissions and average annual temperature were selected as the output variables for multivariate time series analysis. 3.2.2 Modeling for Problem 2

Step 1: Model Analysis.

Assuming that the sequence of independent variables is $\{x_1\}$,…, $\{x_k\}$,and the sequence of response variables is $\{y_t\}$, construct the regression model:

$$y_t = \hat{\beta}_0 + \beta_1 x_{1t} + ... + \beta_k x_{kt} + \varepsilon_t \tag{6}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$,…, $\hat{\beta}_k$ is the least squares estimate.

If the regression residual series $\{\varepsilon_t\}$ passes the smoothness test, i.e., $\varepsilon_t \sim I(0)$, it means that there is a cointegrating relationship between the response series and the input series, i.e., the independent variable series is $\{x_1\}$,…, $\{x_k\}$ and the response variable series $\{y_t\}$ for having a long-run equilibrium relationship.

This equilibrium relationship can be expressed using the regression model established in the first step of the Engle-Granger (EG) test:

$$y_t = \hat{\beta}_0 + \beta_1 x_{1t} + \beta_2 x_{2t} ... + \beta_k x_{kt} \tag{7}$$

Regression residual series.

$$\varepsilon_t = y_t - (\hat{\beta}_0 + \beta_1 x_{1t} + \beta_2 x_{2t} ... + \beta_k x_{kt}) \tag{8}$$

Including random fluctuations that cannot be explained by the input sequence, the response sequence may still contain correlation between historical information. Therefore, we can further examine the autocorrelation and partial autocorrelation information of $\{y_t\}$, and construct an ARMA model:

$$\varepsilon_t = \frac{\theta(B)}{\phi(B)} a_t \tag{9}$$

In the formula, $\theta(B)$ is the qth order moving average coefficient polynomial; $\phi(B)$ is the order autoregressive coefficient polynomial; is the white noise series $a_t \sim N(0, \sigma^2)$.

Completing the above parts of the analysis, we can finally obtain the cointegration fitting model for the response series $\{y_t\}$ with the influence of the input series:

$$y_t = \hat{\beta}_0 + \beta_1 x_{1t} + \beta_2 x_{2t} ... + \beta_k x_{kt} + \frac{\theta(B)}{\phi(B)} a_t \tag{10}$$

Step 2:Data preprocessing:

After calculating the specific values of China's GDP from 2002 to 2014 using the method chosen in problem 1, it was found that the values of GDP were larger compared to the temperature values. Therefore, the logarithm of GDP values was taken as the input sequence, and the temperature values were used as the response sequence. A time series plot was then created based on the average temperature values from 2002 to 2014 and the specific values of China's GDP from 2002 to 2014.

Fig.3shows the time series plot of China's GDP from 2002 to 2014, and Fig.3shows the time series plot of China's average temperature from 2002 to 2014.

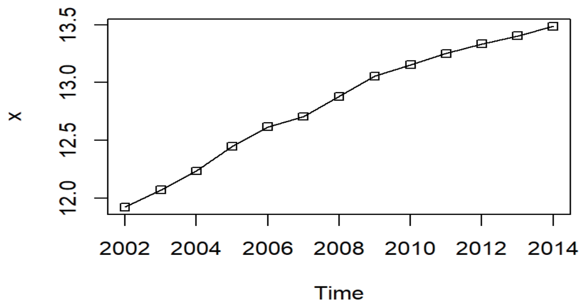


Fig.3Sequence Diagram of Chinese GDP from 2002 to 2014

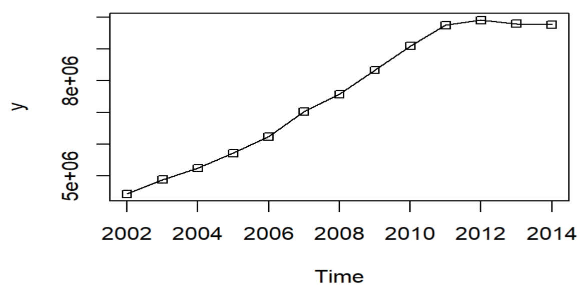


Fig.4Sequence Diagram of annual average carbon dioxide in China from 2002 to 2014

Step 3: Model Building,From Fig.3and Fig.4, it can be seen that the average temperature in China from 2002 to 2004 and the calculated GGDP from 2002 to 2014 are not stationary sequences and both have an increasing trend, indicating non-stationary sequences. Therefore, considering these two sequences together, by observing the time series plot (Fig.5), it can be found that they have a very stable linear correlation. When the GGDP of China from 2002 to 2014 increases, the average temperature from 2002 to 2014 in China also increases, and their rates of change are almost the same. To effectively measure whether there is a long-term equilibrium relationship between the sequences, we introduce the concept of cointegration.

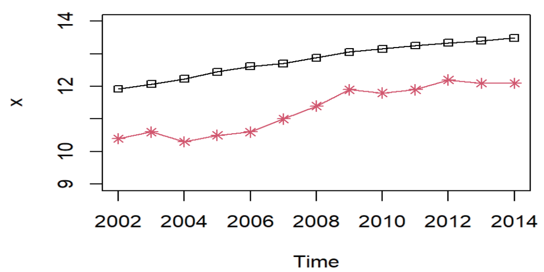


Fig.5Sequence Diagram

Considering that there is a close relationship between the input series GGDP and the response series annual average temperature, a diagonal correlation plot between them is made to examine the structure of the regression model.

The delayed K-order intercorrelation function is defined as:

$$Cov_k = Cov(y_t, x_{t-k}) = E\left[\left(y_t - E(y_t)\right)(x_{t-k} - E(x_{t-k}))\right] \tag{11}$$

Delay The order interrelationship number is defined as:

$$C\rho_k = \frac{Cov(y_t, x_{t-k})}{\sqrt{Var(y_t)}\sqrt{Var(x_{t-k})}} \tag{12}$$

where, can be positive or negative.

Like autocorrelation coefficient and partial autocorrelation coefficient, according to Bartlett's theorem, the mutual relationship number approximately obeys zero-mean normal distribution:

$$C\rho_k \sim N\left(0, \frac{1}{n-|x|}\right) \tag{13}$$

The number of interrelationships exceeding two times the standard deviation can be considered significant and non-zero, i.e., there is a significant correlation between $\{y_t\}$ and $\{x_{t-k}\}$:

The intercorrelation numbers for each order of delay are plotted from the data processed by the input and response series as intercorrelation numbers (see Fig.6)

$$C\rho_k > \frac{2}{\sqrt{n-|k|}} \tag{14}$$

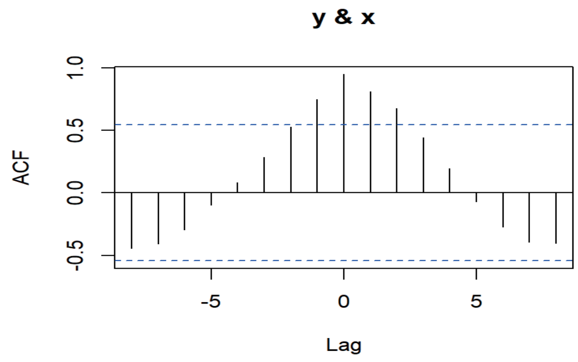


Fig.6Cross-correlation coefficient of output sequence and input sequence

Fig.6 shows the cross-correlation coefficient between the output sequence and input sequence. Since the causal relationship to be examined is clear, the input sequence is the independent variable, and the output sequence is the dependent variable, we choose the peak 0 as the lag period, which indicates that the contemporaneous lag effect between the input and output sequences is the strongest. The regression model obtained by ARIMA can be expressed as:

The is less than for the pure randomness test and passes the test.

The nature of the regression residual series is then examined, and the autocorrelation and partial autocorrelation plots of the residual series (Fig.7) show the characteristics of trailing autocorrelation coefficients and first-order truncated tails of partial autocorrelation coefficients, so the $AR(1)$ model is fitted to the residual series.

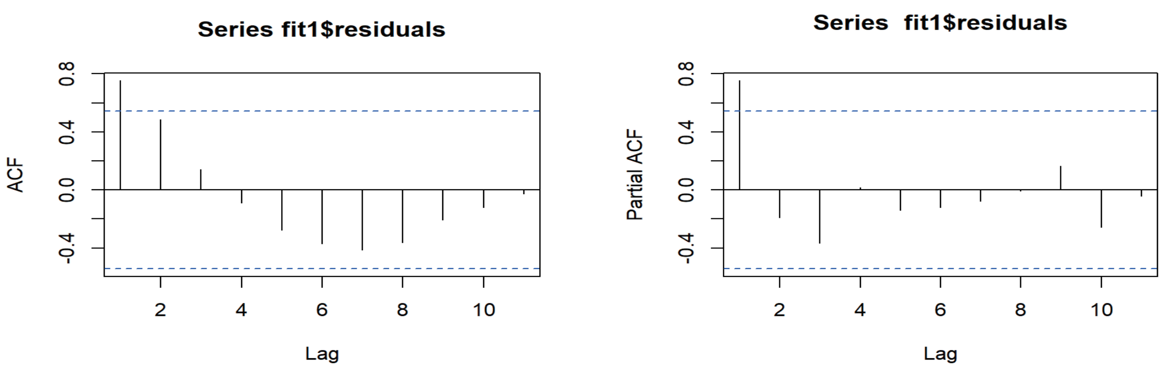


Fig.7Autocorrelation (Left) and Partial Correlation(Right) of egression residual sequence

The final fitting model is as follows;

$$y_t = 0.8835 \ln x_t + \frac{\varepsilon_t}{1 - 0.7349B} \tag{15}$$

The quadratic curve of itself was established to predict itself, and then combined with the already established mathematical model of ARIMA correlation cointegration to predict the annual average temperature of China from 2017 to 2026 data , and also combined with the annual average temperature data of China from 2002 to 2014 to make the fitted prediction graphFig.8.

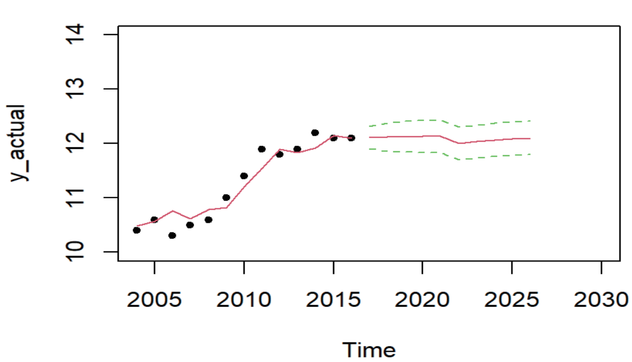


Fig.8Fitting prediction chart of Chineseannual average temperature from 2017 to 2026

Similarly, we can generate a prediction graph for China's annual carbon dioxide emissions:

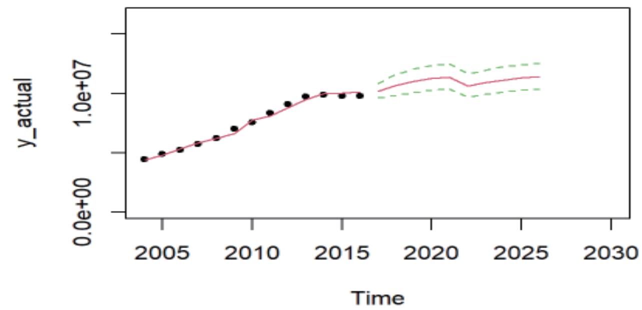
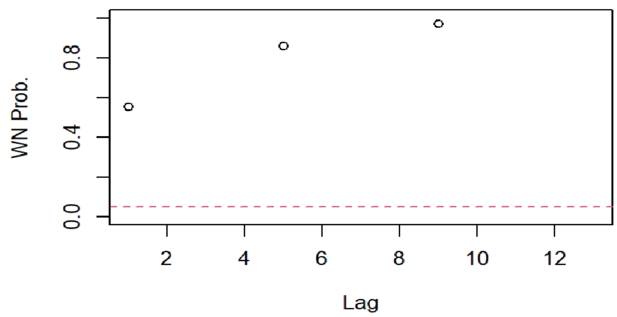


Fig.9Fitting prediction chart of ChineseAnnual carbon dioxide emissions from 2017 to 2026

As can be seen from the figures, after using GGDP instead of GDP as the main indicator to measure a country's economic health, the annual average temperature and annual CO2 emissions have both decreased. It can be considered that using GGDP as a measure of a country's economic health has a positive impact on solving climate change issues.

6. Analysis and inspection of the model



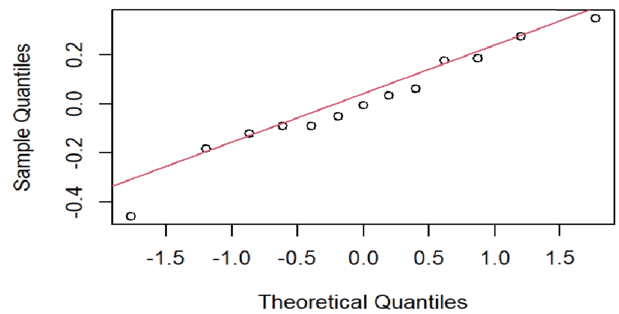


Fig.10The WN Prob graph **Fig.11The QQ graph**

The WN Prob graph is a white noise probability graph, which can be used to test whether the residual sequence conforms to a white noise distribution. The horizontal axis of this graph represents the quantiles of the standard normal distribution, while the vertical axis represents the sorted values of the residuals. The p-value of the residual white noise test is greater than 0.05, indicating that the model passes the test. The residual sequence conforms to a white noise distribution, with residual values approximately distributed along the diagonal line.

The QQ graph is a normal probability graph of the residuals, used to test whether the residual sequence conforms to a normal distribution. The horizontal axis of this graph represents the quantiles of the standard normal distribution, while the vertical axis represents the quantiles of the residuals. This graph indicates that the residual sequence conforms to a normal distribution, with points distributed along a straight line.

7. Evaluation and Promotion of Model

7.1 Strength

●Partial least squares regression analysis can perform regression of multiple dependent variables on multiple independent variables simultaneously. It is also applicable to small sample sizes and can obtain accurate regression equations.

●Improving prediction accuracy: Multivariate time series analysis with cointegration models can use information from multiple variables to improve future prediction accuracy.

●The established model can closely relate to reality, solve the proposed requirements in combination with the actual situation, and have strong universality and promotability.

7.2 Weakness:

●The regression coefficients used in problem 1 with partial least squares regression analysis are difficult to interpret.

●The multivariate time series analysis method used in problem 2 requires that the residual of the input and output sequences are both stationary white noise sequences, and the application conditions are relatively strict. It is necessary to establish a more suitable mathematical model based on the data.

7.3Promotion

●Consider more variables:In practical applications, there may be multiple factors related to the target variable. Considering more variables in the model may lead to a more accurate prediction of the target variable.

●Introduce time trend:Time factors may also have a significant impact on variable changes in practical applications. Introducing time trend items, such as seasonal factors or long-term trends, can enhance the predictive ability of the model.

References

[1] Harvey Alexander. Nature's services: societal dependence on natural ecosystems[J].Corporate Environmental Strategy,1999,6(02) :219.

[2] Costanza. R .Ecological economies: reintegrating the study of humans and nature[J].Nature Sciences Science, 1997,5(02) : 90.

[3]Su Shiliang,Xiao,Rui,Lan Zhen,et al. Characterizing landscape pattern and ecosystem service value changes for urbanization impacts at an ecoregional scale[J].Applied Geography,2012,(34): 295-305.

[4] Sun C,Wu ZF,Lv ZQ,et al. Quantifying different types of urban growth and the change dynamic in Guangzhou using multitemporal remote sensing data[J]. International Journal of Applied Earth Observations and Geoinformation,2013,(21):409-417.