ANALYSIS AND DECISION-MAKING OF REGIONAL ECONOMIC VITALITY AND ITS INFLUENCING FACTORS

ShaSha Zhang
School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China.
Corresponding Email: sszhangq@qq.com

Abstract: This paper takes Henan Province as the research object, selects GDP as the index of regional economic vitality, and selects the index that affects GDP from the aspects of economic benefit, opening up, population, government regulation, residents' quality of life and enterprise vitality. Firstly, a multiple linear regression model is established to test the multicollinearity of the independent variables. Then, the ridge regression and LASSO regression models are established to correct them, and the multicollinearity problem between independent variables is solved. By comparing and analyzing the two models, the LASSO regression model is the optimal regression model. Secondly, the Holt exponential smoothing model is used to predict the time series of the five variables showing a linear trend in the LASSO regression equation. The simple exponential smoothing model is used to predict the four variables of random fluctuation, and the data of each variable in 2024 are predicted. The predicted value of GDP in Henan Province in 2024 is 61441.55 billion yuan. Finally, some suggestions are put forward for Henan Province from several aspects, so that the economic development of the region can form a virtuous circle, so as to enhance the competitiveness of its regional economy and promote the sustainable development of the economy.
Keywords: Multivariate linear regression model; Ridge regression model; LASSO regression model; Time series prediction; Regional economic vitality

1 INTRODUCTION

1.1 Research Background

China implemented reform and opening up in 1978. In the early stage, the development model of "first rich and then rich" showed great superiority, which made the economy of the eastern coastal areas of China develop rapidly, broke the obstacles brought by human resources and production resources in the process of economic development, and achieved great results.

However, in this development process, the regional economic development gap between the eastern, central, western and northeastern regions of China is increasing, which is not conducive to the rational allocation of various resources, nor to the long-term stable and healthy development of China's economy. The development of China's regional economy is unbalanced. This phenomenon not only hinders the development of China's economic level, but also has a negative impact on social stability and prosperity. After the implementation of the reform and opening up, China has formulated many corresponding development strategies to shorten the imbalance of regional economic development in China, such as the strategy of "western development" and "central rise". Although various development strategies have been formulated, this phenomenon in China has not been well improved, and this phenomenon has become a challenge that China must deal with in economic development in the new era[1].

1.2 Research Meaning

China is a large developing country with numerous economic regions, and the conditions of each region are very different. Therefore, the level and status of regional economic development are very different. How to make the economy of all regions of our country develop, the regional economy becomes more active, and the overall efficiency becomes higher is a difficult problem that China's current development economics needs to explore and overcome[2]. The region and industry are closely linked, they are inseparable, regional economic development will inevitably lead to the development of some industries. Only by building a unified, open, competitive and orderly Chinese market, uniting all regions, complementing each other's advantages and cooperating with each other, can the whole national economy be put on the agenda.

Industrialization, marketization and socialization are not only the historical task that China has to complete, but also the great historical mission entrusted by the era of realizing China's modernization. The task is very difficult. Economic development plays a very important role in this task, which is a very important link that we must go through to complete this dual important historical mission. To realize modernization development requires a complete process, which should be phased, planned and have complete steps. Reform and opening up is the most fundamental feature of China in the new era. China is currently undergoing a transition from a traditional planned economic system to a modern socialist market economy. Therefore, it is of great significance to develop China's regional economy and shorten the gap between regions, not only to realize China's modernization development, but also to maintain the stable development of the motherland, maintain the unity of the motherland, and establish a great image of a powerful country.

2 RESEARCH THOUGHT

2.1 The Overall Framework

Taking a region (or city or province) as an example, this paper determines the indicators that affect the regional economic vitality by consulting the statistical yearbook, establishes the relationship model of the factors affecting the economic vitality, compares the models, selects the optimal model, and puts forward some suggestions to improve the economic vitality of Henan Province according to the model results, and analyzes the influence and function of the regional economic vitality from various angles. Some policy suggestions are put forward to improve the economic vitality of the region (or city or province) discussed in the analysis, so as to improve the economic vitality of the region and strengthen the sustainable development of the economy. The specific framework is shown in Figure 1.
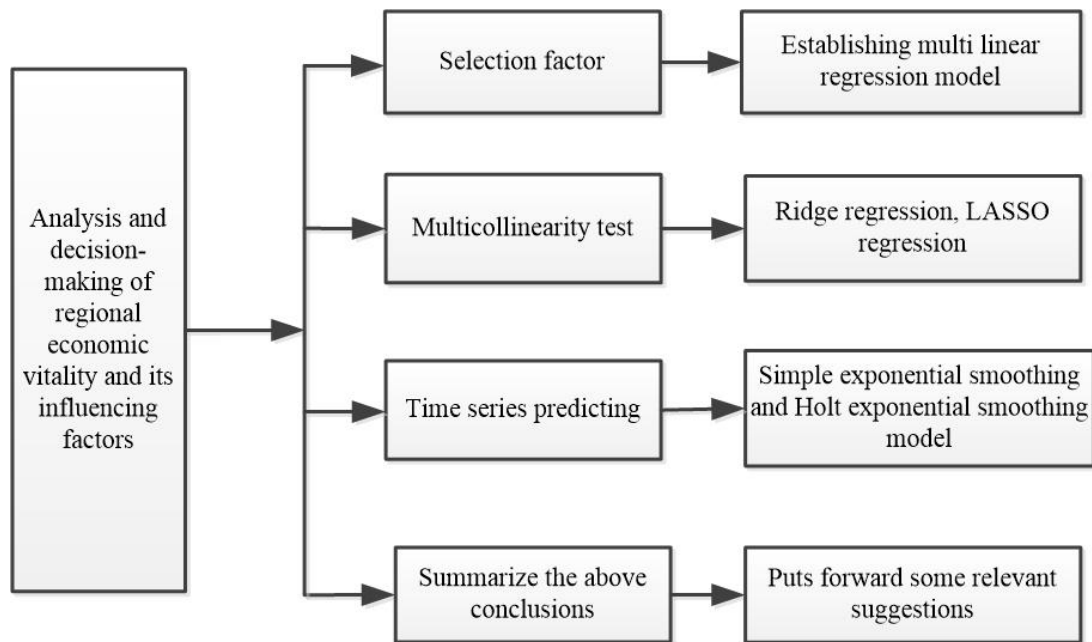

Figure 1 Research Idea Diagram

2.2 Research Specific Ideas

With the continuous advancement of China's modernization drive, the regional economy has developed rapidly under the impetus of various driving forces. Under the stimulation of various preferential policies issued by various regions, effectively improving regional economic vitality has become an irreplaceable part of regional comprehensive competitiveness.

Taking Henan Province as an example, this paper analyzes and studies the economic vitality of Henan Province. Firstly, by consulting the statistical yearbooks of Henan Province over the years, it determines the indicators that affect the regional economic vitality of Henan Province, and explains the rationality and correctness of selecting these indicators. Through the website of the National Bureau of Statistics, the data corresponding to each index of Henan Province from 2009 to 2023 were searched and sorted out. According to the selected indicators, a multiple linear regression model is established. According to the solution results, suggestions are put forward to help improve the economic vitality of Henan Province, and its impact on regional economic vitality is analyzed from various aspects.

From the results of the multiple linear regression model, there is a great correlation between the independent variables. Then the multicollinearity test is carried out on the independent variables, and the variables with multicollinearity are obtained. Then the ridge regression model and the LASSO regression model are established, and the regression equations of each regression model are obtained. According to the regression equation, the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, is calculated, and the comparison diagram between the fitted value and the actual value is obtained. By comparing the two models, the optimal regression model is obtained. Through the time series diagram of the independent variables in the optimal regression equation, the time series prediction of each index is judged by establishing what model, and the standardized data of each index in 2024 is predicted, and the data is substituted into the optimal regression equation, and the predicted value of GDP in Henan Province in 2024 is obtained.

Finally, combined with the analysis results, starting from the actual economic development of Henan Province, comply with the basic principles of regional economic development, break through various constraints that hinder regional economic development, and put forward some constructive policy recommendations for Henan Province, so that the economic vitality of Henan Province forms a virtuous cycle, economic sustainable development, and constantly strengthen the comprehensive competitiveness of the region.

3 MODEL ASSUMPTIONS AND SYMBOLIC DESCRIPTION

3.1 Model Assumptions

It is assumed that the data in the statistical yearbooks of Henan Province over the years are true and effective; assume that the data in the National Bureau of Statistics are true and effective; it is assumed that the selected indicators can fully reflect the changes in regional economic vitality; assume that human error is not considered. All the above assumptions are true.

3.2 Symbol Description

The symbols used in this article, the relevant instructions are shown in Table 1.

Table 1 Symbol Definition

Sign	Definition
β_0	Regression constants
$\beta_j(j=1,2,\cdots k)$	Regression coefficient
y	Dependent variable
ZX_i	Represents the standardized variable of X_i
$E(X_i)$	Represents the mean value of variable X_i
$Var(X_i)$	Represents the variance of variable X_i
r_{ij}	Correlation coefficient between the variables X_j representing the variable X_i
P_{ij}	Denotes partial correlation coefficient

From the above table, we can clearly see the relevant definitions of the symbols used in this article.

4 ESTABLISHMENT AND SOLUTION OF MODEL

4.1 The Establishment and Solution of Multiple Linear Regression Model

4.1.1 The construction of economic vitality index system

Henan Province is the largest grain production base in China, and its agricultural food and animal products rank first in the national ranking. In addition, Henan Province is not only outstanding, but also has a very rich mineral tourism resources and its transportation is also very convenient. However, there are other problems in Henan Province, such as the lack of excellent talents and the attraction of outstanding talents outside the province, which leads to the weak economic strength of Henan Province. Compared with other neighboring provinces, its regional comprehensive competitiveness is relatively small. Therefore, this paper selects Henan Province as the research object, and systematically and deeply studies the regional economic vitality of Henan Province.

First of all, in the selection of the impact indicators of economic vitality in Henan Province, it must be based on the principles of scientificity and integrity, and the selected indicators can accurately and truly show the impact of various factors on the economic vitality of Henan Province. Because the regional economic vitality is a complex and pluralistic system, the indicators selected in this paper should reflect the development status and trend of regional economic vitality from multiple levels, multiple perspectives and multiple dimensions. Then, when establishing an index system that affects economic vitality, it is mainly based on theoretical analysis, but due to the limitations of data sources and data availability when constructing indicators. Therefore, the constructed indicators should be as easy to understand as possible, the evaluation methods should be combined with the old and the new, and the selected indicators should ensure the authenticity and reliability of the data sources. Finally, regional economic vitality is an abstract phenomenon that is difficult to present in detail. When selecting indicators, it is necessary to pay attention to whether there is an inevitable connection with it. Based on the research of other scholars, the index system constructed for different periods is not exactly the same, so it is necessary to select indicators according to the actual situation of the research object of this topic[3].

The purpose of this paper is to focus on the economic vitality of Henan Province, and to establish an index system based on the relevant data and the actual economic development of Henan Province. In this paper, GDP (billion yuan) is selected as the representative index of economic vitality in Henan Province, and the relevant indexes affecting regional economic vitality are selected from six aspects: economic benefit, opening up, population, government regulation, residents' quality of life and enterprise vitality. They are per capita GDP (yuan / time), the annual cumulative number of tourists (millions of people), the total amount of foreign-funded enterprises in and out (thousands of dollars), the natural population growth rate (%), the consumer price index (last year = 100), the urban unemployment rate (%), fiscal

revenue (billions of yuan), fiscal expenditure (billions of yuan), per capita disposable income (yuan), total retail sales of social consumer goods (billions of yuan), R&D funds for industrial enterprises above designated size (ten thousand yuan) and the number of industrial enterprises above designated size (units)[4]. The details are shown in Table 2.

Table 2 The Selection of Economic Vitality Indicators in Henan Province

Influencing factor	Index
economic benefit	GDP per capita (yuan / person)
opening to the outside world	Cumulative annual number of visitors (millions)
	Total import and export volume of foreign-invested enterprises (USD 1000)
population	Natural population growth rate (%)
	Consumer Price Index (last year = 100)
	Urban unemployment rate (%)
government regulation	Fiscal revenue (billion yuan)
	Fiscal expenditure (billion yuan)
residents' quality of life	Per capita disposable income (yuan)
	Total retail sales of social consumer goods (billion yuan)
enterprise vitality	R&D expenditure of industrial enterprises above designated size (ten thousand yuan)
	Number of industrial enterprises above designated size (units)

4.1.2 Based on the establishment of multiple linear regression model

Let the multivariate linear regression model of random variable y and variable $x_1, x_2, \cdots, x_k$ ($k \geq 2$) be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{1}$$

Among them, β_0 is the regression constant, $\beta_j \left(j = 1, 2, \cdots, k \right)$ is the regression coefficient, y is the dependent variable, $x_1, x_2, \cdots, x_k$ is k independent variables that have a significant impact on y, and ε is a random term that shows the comprehensive impact of various errors on the dependent variable[5].

If n sets of observation data $x_{i1}, x_{i2}, \cdots, x_{ik}, y_i \left(i = 1, 2, \cdots, n \right)$ are obtained, the above equation can be expressed as the following equation.

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\ \quad\quad\quad\quad\quad \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n \end{cases} \tag{2}$$

4.1.3 Based on the solution of multiple linear regression model

Through the website of the National Bureau of Statistics, the data corresponding to each index of Henan Province from 2009 to 2023 are searched and sorted out. The 12 indexes selected from the six aspects of economic benefits, opening up, population, government regulation, residents' quality of life and enterprise vitality are expressed in turn by A, B,...,L. Firstly, the correlation between GDP, the representative index of economic vitality, and these 12 indexes is analyzed[6]. It is concluded that the correlation coefficients between GDP and per capita GDP, fiscal revenue, fiscal expenditure, per capita disposable income, total retail sales of social consumer goods and R&D funds of industrial enterprises above designated size are all higher than 0.9. There is a strong positive correlation; the correlation coefficients between GDP and the total number of tourists received throughout the year, the total amount of foreign-invested enterprises in and out, the natural population growth rate and the number of industrial enterprises above designated size are all higher than 0.6, showing a strong positive correlation. However, the correlation between GDP and consumer price index and urban unemployment rate is not very high, and the correlation coefficients are -0.27 and -0.04 respectively.

Therefore, this paper considers GDP, the representative index of regional economic vitality in Henan Province, as the dependent variable (y)[7]. From the six aspects of economic benefits, opening up, population, government regulation, residents' quality of life and enterprise vitality, 10 indicators affecting economic vitality are selected as independent variables, which are set as per capita GDP (x_1), total number of tourists received throughout the year (x_2), total inflow and outflow of foreign-invested enterprises (x_3), natural population growth rate (x_4), fiscal revenue (x_5), fiscal

expenditure (x_6), per capita disposable income (x_7), total retail sales of social consumer goods (x_8), R&D funds of industrial enterprises above designated size (x_9), and number of industrial enterprises above designated size (x_{10}). Establish a multiple linear regression model, use R software to write code, and use the least squares estimation method to obtain the OLS regression model as follows:

$$y = 0.837x_1 - 130.1x_2 + 0.000027x_3 - 309.7x_4 - 0.586x_5 + 0.277x_6$$
$$-0.727x_7 + 0.773x_8 + 0.0006x_9 + 0.0056x_{10} + 3625 \tag{3}$$

Through the above equation, we get that the indicators that have the greatest impact on the economic vitality of Henan Province are: the annual cumulative number of tourists and the natural population growth rate, and these independent variables are negatively correlated with GDP. The total amount of foreign-invested enterprises, the R&D funds of industrial enterprises above designated size and the number of industrial enterprises above designated size have the least impact on the economic vitality of Henan Province.

We use R software to write code to test the OLS regression model. The correlation coefficient of the OLS regression model is 0.999, indicating that the fitting effect of the model is very good. And the corresponding P value is less than 0.01, indicating that the OLS regression model is highly significant and the overall fitting effect is good.

It can be seen from the t-test results that the per capita GDP and the total retail sales of social consumer goods pass the t-test, while other variables do not pass the t-test, and the impact on the dependent variables is not significant[8-12]. Combined with the above analysis, it can be seen that there is a great correlation between these independent variables that do not pass the t test, so this paper should consider that the variable does not pass the t test may be the reason for the multicollinearity of these variables.

4.1.4 Analysis of effect

According to the results of the multiple linear regression model, this paper gives the action plan to improve the economic vitality: from the perspective of opening up, Henan Province can improve the vitality of economic growth by reducing the number of tourists. From the perspective of population, Henan Province can appropriately reduce the number of unemployed people and increase the consumer price index of residents by controlling the natural growth rate of the population, reducing the population base, increasing the total per capita GDP and increasing the setting of posts, so as to improve the vitality of economic growth. From the perspective of government regulation and control, Henan Province can appropriately promote a virtuous cycle of economic vitality by reducing fiscal revenue and increasing fiscal expenditure; from the perspective of residents' quality of life, Henan Province can appropriately reduce per capita disposable income and increase the sales of social consumer goods to improve the vitality of economic growth.

4.1.5 Multiple collinearity diagnosis

Considering that the independent variables in the OLS regression model may have multicollinearity, this paper uses the most common eigenvalue determination method to diagnose the multicollinearity of the independent variables.

Using R software to write the code, the condition number is 32151, indicating that there is a serious multicollinearity between the independent variables. In order to determine which independent variables have multicollinearity, the eigenvalue of each variable is obtained by using R software to write code.

$$\zeta = \left(7.97, 1.07, 0.62, 0.24, 0.09, 0.006, 0.003, 0.0009, 0.0008, 0.0002\right) \tag{4}$$

It can be seen that the eigenvalues of x_6 , x_7 , x_8 , x_9 and x_{10} are close to 0, so there is multicollinearity between x_1 , x_2 , x_3 , x_4 and x_5 . When there is multicollinearity between independent variables, the accuracy of parameter estimation of the model will decline sharply, so that the estimated value cannot be explained from the perspective of economy and society, so that the application value of the model will decline sharply. In the following, we will use ridge regression and LASSO regression to modify the multiple linear regression model respectively, and at the same time solve the problem of multicollinearity between independent variables.

4.2 The Establishment and Solution of Ridge Regression Model

4.2.1 The establishment of ridge regression model

When there is multicollinearity between independent variables, ridge regression is a coefficient estimation method proposed by Hoerl and Kennard. Although ridge regression will lose some information and reduce the accuracy of fitting, the obtained regression coefficient is more realistic and reliable, and the fitting of ill-conditioned data is better than the least square method[13].

The linear regression model is established as follows:

$$Y = Z\theta + \varepsilon \tag{5}$$

In the above formula, Y is the dependent variable, Z is the independent variable, θ is the standard regression coefficient, ε is a random error, P is the number of independent variables, n is the number of samples. Where,

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{11} & \cdots & Z_{p1} \\ \vdots & \ddots & \vdots \\ Z_{1n} & \cdots & Z_{pn} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{6}$$

Due to the different dimensions or the large numerical gap, the solution results will cause great errors. It is necessary to use the following formula to centralize and unit the length of Y and Z respectively.

$$Z_{ij} = \frac{X_{ij} - \overline{X}_j}{\sqrt{\sum (X_{ij} - \overline{x}_j)^2}} \tag{7}$$

The least squares estimate of θ can be expressed as:

$$\hat{\theta} = (Z^T Z)^{-1} Z^T Y \tag{8}$$

When there is multicollinearity between independent variables, matrix $Z^T Z$ is a singular matrix, and its corresponding eigenvalues are very small. The elements on the diagonal of matrix $(Z^T Z)^{-1}$ are very large, which will lead to very unstable parameter estimation. The slight change of data will make the estimated value of parameters change greatly, and the regression coefficient cannot accurately and objectively reflect the influence of independent variables on dependent variables. Ridge regression is to correct the shortcomings of the least squares method. A diagonal matrix kI is added on the basis of matrix $Z^T Z$, so that the eigenvalues of the matrix become larger, and the singular matrix becomes a non-singular matrix, so as to improve the stability of parameter estimation and make the regression coefficient more accurately reflect the objective reality. The standardized coefficient $\Theta_{(k)}$ of ridge regression estimation is:

$$\hat{\Theta}_{(k)} = (Z^T Z + kI)^{-1} Z^T Y = (Z^T Z + kI)^{-1} Z^T \hat{\Theta} \tag{9}$$

Where, k is the ridge regression parameter, and the value range is 0~2. When $k=0$, it is the least squares estimation; when the $k \neq 0$ and k values are larger, the predicted variance is larger, and the influence of multicollinearity on the stability of the regression coefficient becomes smaller.

4.2.2 The solution of ridge regression model

Ridge regression is a biased estimation regression method used to solve the multicollinearity between independent variables. By giving up the unbiasedness of least squares estimation, the regression coefficient obtained at the cost of losing some information and reducing the accuracy of fitting is more realistic and more reliable. The regression method is suitable for the fitting of ill-conditioned data.

Considering the different dimensions of each variable, the R software is used to write the code. Firstly, the original data is standardized to eliminate the influence of the dimension on the accuracy of the model. Then, the ridge regression analysis is carried out. The ridge parameter k is 0-2, and the step size is 0.2. The ridge trace diagram corresponding to the 10 ridge parameter values is shown in Figure 2.
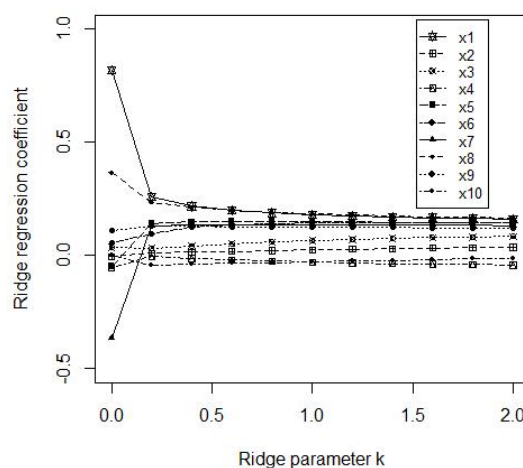

Figure 2 Ridge Trace Map

It can be seen from the above figure that according to the general principle of ridge regression k value selection, the independent variables x_2, x_3, x_4, and x_{10} with relatively stable coefficients and small absolute values are eliminated, and the new ridge trace is obtained, as shown in Figure 3.
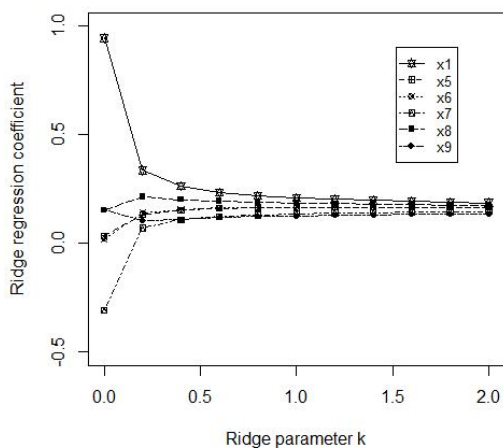
Figure 3 Ridge Trace Map

It can be seen from the above figure that the ridge regression coefficient changes more and more slowly after removing some variables. In summary, when $k > 0.8$, the value of the ridge regression coefficient basically reaches a stable state. Therefore, this paper uses R software to write code and select the ridge regression coefficient corresponding to each other when the ridge parameter $k=0.8$ is selected. The standardized ridge regression equation is:

$$\hat{y} = 0.220x_1 + 0.165x_5 + 0.166x_6 + 0.132x_7 + 0.189x_8 + 0.124x_9 \tag{10}$$

The F value of the ridge regression model is 10090, and the p value is less than 0.01 when the t test is performed, indicating that the model is very effective. The correlation coefficient is 0.999, the fitting effect is good, and the goodness is high. This model can be used for research and analysis.

From the standardized ridge regression equation, it can be seen that per capita GDP, fiscal revenue, fiscal expenditure, per capita disposable income, total retail sales of social consumer goods and R&D expenditure of industrial enterprises above designated size are positively correlated with the economic vitality of Henan Province. At the same time, the importance of the main six factors affecting the economic vitality of Henan Province is ranked from large to small: per capita GDP, total retail sales of social consumer goods, fiscal expenditure, fiscal revenue, per capita disposable income, and R&D expenditure of industrial enterprises above designated size.

Using R software to write the code, the unstandardized ridge regression equation is:

$$\hat{y} = 0.967x_1 + 0.429x_5 + 0.1x_6 - 0.609x_7 + 0.324x_8 + 0.0008x_9 + 199.9 \tag{11}$$

In order to further better observe the fitting effect of the ridge regression model, this paper calculates the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, according to the above equation, and uses R software to write the code to draw a comparison diagram, as shown in Figure 4.
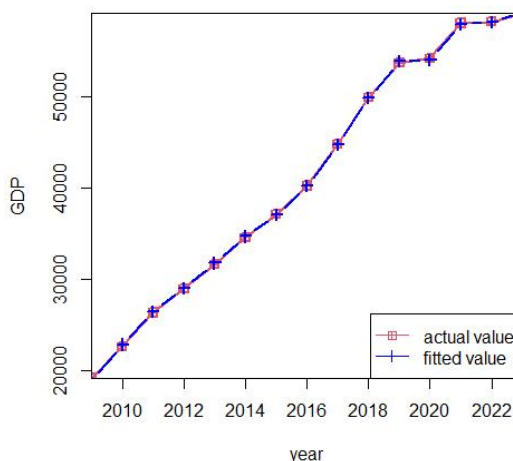

Figure 4 The Comparison Chart Between the Actual Value of GDP and the Fitting Value of Ridge Regression

From the above figure, we can see the intuitive effect of the comparison between the actual value of GDP and the ridge regression fitting value, which shows that the fitting effect of the model is very good.

4.3 The Establishment and Solution of LASSO Regression Model

4.3.1 The establishment of LASSO regression model

LASSO model is designed to give priority to the importance of independent variables. LASSO regression determines the best model by compressing the estimation. Its fundamental property is to obtain a more accurate model by establishing a penalty function. Under the constraint that the sum of the absolute values of the regression coefficients is less than a constant, the sum of the residual squares is minimized, and the variables whose regression coefficients are close to or equal to 0 are eliminated, so as to solve the problem of multicollinearity[14].

LASSO regression is based on ordinary linear regression with additional penalty terms, and its estimation is:

$$\hat{\beta}_{Lasso} = \arg\min_{\beta \in R^p} \left\| Y - X\beta \right\|^2 \tag{12}$$

$$s.t. \sum_{j=1}^{p} \left| \beta_j \right| \le t, t \ge 0 \tag{13}$$

Equivalent to:

$$\hat{\beta}_{Lasso} = \arg\min \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right\} \tag{14}$$

In the above formula, $\arg\min(\cdot)$ is the function of finding the minimum value of the parameter, and $\hat{\beta}_{Lasso}$ is the objective function of minimization; λ is the adjustment parameter, and $\lambda \ge 0$; $\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$ represents the effect of model fitting, and $\lambda \sum_{j=1}^{p} \left| \beta_j \right|$ represents the penalty of parameters[15].

The smaller the value of λ , the smaller the punishment will be, and the more variables will be retained in the model; the greater the value of λ is, the greater the punishment will be, and the fewer variables are retained in the model. Through the selection of parameter λ , variable selection can be achieved[16].

4.3.2 The solution of LASSO regression model

Considering the different dimensions of each variable, R software is used to write code. Firstly, the original data is standardized to eliminate the influence of dimension on the accuracy of the model. Then, LASSO regression analysis is carried out to obtain Figure 5.
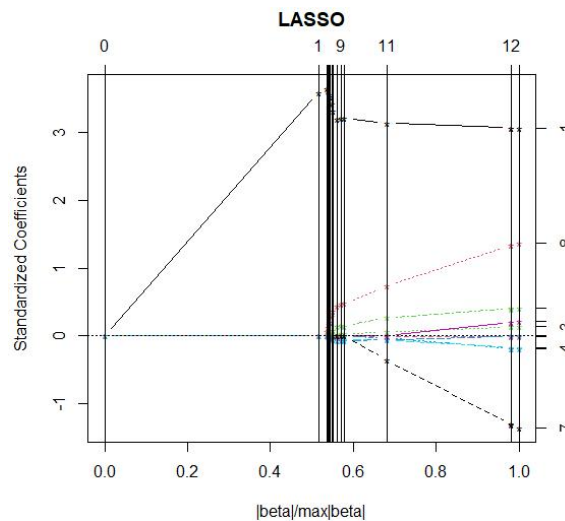

Figure 5 LASSO Regression Diagram

The coefficient selection methods include the cv method of k-fold cross-validation and the method of using Cp statistics to evaluate regression for coefficient selection. In this paper, the Mallows Cp statistic selection coefficient is selected, and the standardized LASSO regression equation is obtained by using R software to write code.

$$\hat{y} = 0.840x_1 - 0.003x_2 + 0.015x_3 - 0.009x_4 - 0.017x_5 - 0.096x_7 + 0.196x_8 + 0.071x_9 - 0.015x_{10} \tag{15}$$

From the above equation, it can be seen that the coefficient of the fiscal expenditure variable is zero, and this variable is eliminated. The importance of the main nine independent variables affecting the economic vitality of Henan Province is ranked from large to small: per capita GDP, total retail sales of social consumer goods, per capita disposable income, R&D expenditure of industrial enterprises above designated size, fiscal revenue, total inflow and outflow of foreign-invested enterprises, number of units of industrial enterprises above designated size, natural population growth rate and cumulative number of tourists received throughout the year.

In order to further observe the fitting effect of the LASSO regression model, this paper calculates the fitting value of GDP, the representative index of economic vitality in Henan Province from 2009 to 2023, according to the above equation, and uses R software to write the code to draw a comparison chart, as shown in Figure 6.
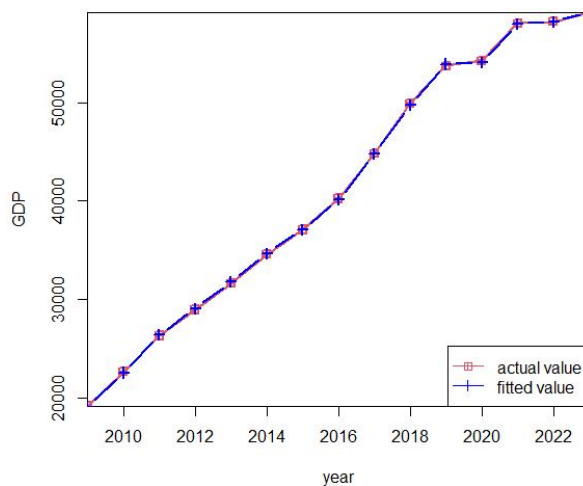
Figure 6 The Comparison Between the Actual Value of GDP and the Fitting Value of LASSO Regression

From the above diagram, we can see the intuitive effect of the comparison between the actual value of GDP and the fitting value of LASSO regression, and the fitting effect of the model is very good.

4.4 Model Analysis

In this paper, ridge regression and LASSO regression model are used to eliminate the effect of multicollinearity between variables. By comparing the results of the test and parameter test of each model, the optimal regression model is selected. Using R software to write code, the comparative analysis of ridge regression and LASSO regression model is shown in Table 3.

Table 3 Model Analysis

Model	RMSE	R^2
Ridge regression	0.011106	0.999
LASSO regression	0.008112	0.999

From the above table, it can be seen that the RMSE of the LASSO regression model is small, indicating that the deviation between the fitting value of the LASSO regression model and the actual value is smaller, and the effect is optimal; by comparing the goodness of fit, it can be concluded that the values of the ridge regression and the LASSO regression model are the same, and the fitting effect is very good. In summary, the LASSO regression model is a relatively better model.

4.5 Time Series Predicting

From the comparative analysis results of the model, it can be seen that the LASSO regression model is a relatively better model, and in the LASSO regression equation, the fiscal expenditure variable is eliminated. Therefore, we first use R software to write the code, and get the time series diagram of other independent variables in the equation, as shown in Figure 7, Figure 8, Figure 9.
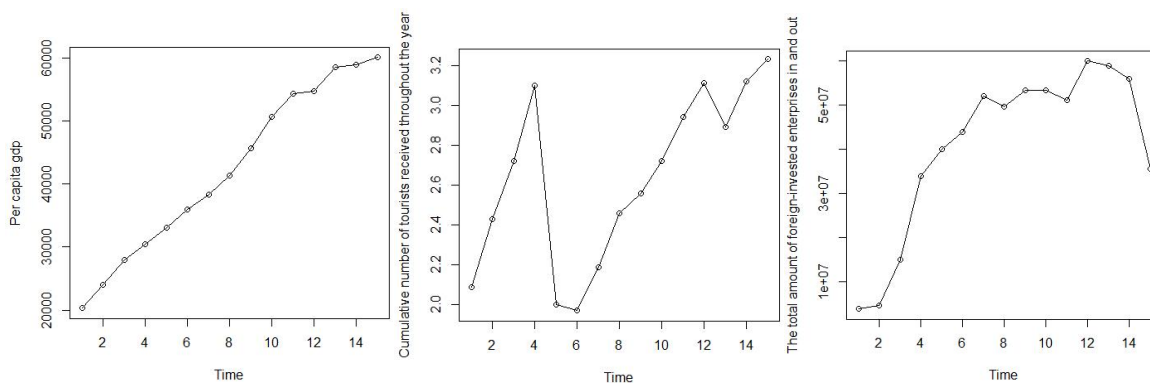

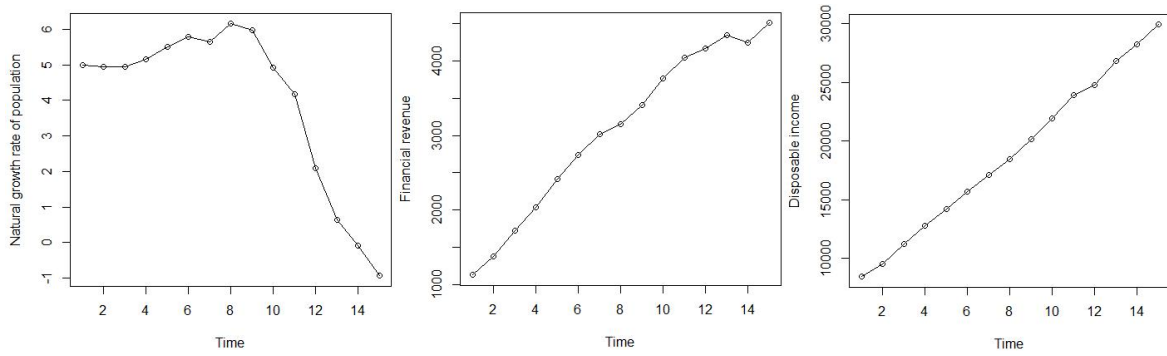**Figure 7** x_1, x_2, x_3 Time Series Diagram

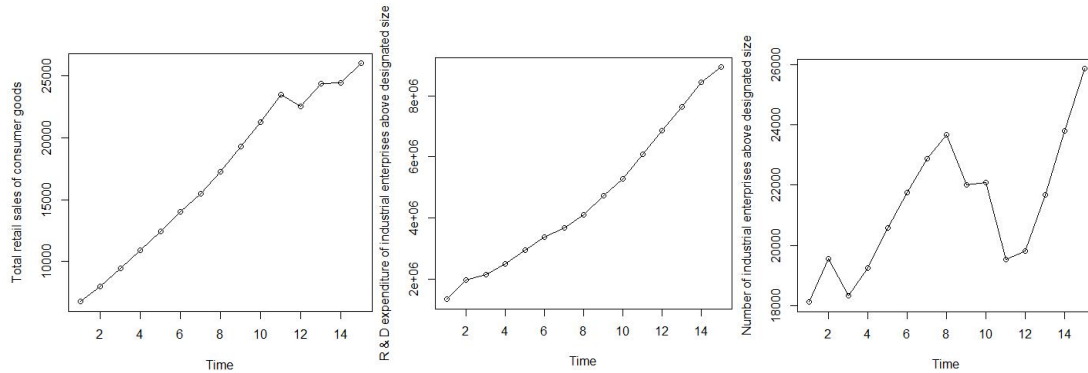**Figure 8** x_4, x_5, x_7 Time Series Diagram

Figure 9 x_8, x_9, x_{10} Time Series Diagram

From the time series diagram of the above variables, it can be seen that the five variables of per capita GDP, fiscal revenue, per capita disposable income, total retail sales of social consumer goods, and R&D expenditure of industrial enterprises above designated size show a linear trend, and the remaining four variables show random fluctuations without significant changes.

For the five variables showing a linear trend, this paper uses the Holt exponential smoothing model for time series prediction. For these four variables that show random fluctuations without significant changes, a simple exponential smoothing model is used for time series prediction. The standardized data of each index in 2024 are shown in Table 4.

Table 4 The Predictive Value of Independent Variables in 2024

Year	x_1	x_2	x_3	x_4	x_5	x_7	x_8	x_9	x_{10}
2024	1.472	1.355	-0.276	-2.095	1.497	1.822	1.560	1.985	2.076

From the results of the LASSO model, the standardized LASSO regression equation is:

$$\hat{y} = 0.840x_1 - 0.003x_2 + 0.015x_3 - 0.009x_4 - 0.017x_5 - 0.096x_7 + 0.196x_8 + 0.071x_9 - 0.015x_{10} \tag{16}$$

The standardized data of each index predicted in the above table in 2024 are substituted into the above regression model, and then the predicted value of GDP in Henan Province in 2024 is calculated to be 61441.55 billion yuan.

5 SUGGESTIONS ON PROMOTING THE ECONOMIC DEVELOPMENT OF HENAN PROVINCE

5.1 Increase Investment in Domestic Enterprises to Form a Sustainable Industrial Chain

The government of Henan Province should increase the investment attraction and investment attraction of enterprises, speed up the implementation of excellent projects as soon as possible, especially introduce and train some leading enterprises and leading industries, so as to improve the scientific and technological technology of Henan Province, make the industrial chain longer, improve the comprehensive benefits of Henan Province, and improve the vitality of regional economy in an all-round way. It can also start from the regional financial field, through continuous improvement and broadening the financing channels of private enterprises and state-owned enterprises, in order to create a good financial environment and achieve a sustainable financial chain.

5.2 Improve the Efficiency of Government Work

The government of Henan Province can integrate the resources within the scope of Henan Province through policy intervention, which can provide more excellent resource services for enterprises, so that enterprises have a broader space for development. The government can also relatively reduce the more complicated procedures and reduce

unnecessary operations for the establishment process of enterprises. The government should improve work efficiency, because its work efficiency has a direct impact on the cost of enterprises. Low efficiency work will bring bad results to the competitiveness of enterprises, and enterprises play a very important role in regional economic competitiveness. The competition among regions is the competition among enterprises in different regions.

5.3 Pay Attention to System and Mechanism Innovation, Establish Market Competition Mechanism

The development trend of the economy is influenced by the system of our country. The quality of the system plays a decisive role in the economic growth of various places, especially in the degree of modern informationization, the development degree of the factor market and the quality of the market environment in our country. This shows that a good market environment can attract many foreign excellent resources to the region, and the more it can promote the rapid growth of the local economy.

In addition, by improving the environment of the market, elements can be attracted to take root locally, so that these excellent resources can be retained locally, forming the local core strength and improving the local economic development level. Competition plays a vital role in economic growth. Competition with other enterprises can make enterprises improve their fighting spirit, constantly improve themselves, improve their competitiveness, and make themselves better. Two-way competition can enable enterprises to continuously innovate. In general, the economy can be improved as a whole through mutual competition.

5.4 Strengthen Regional Cooperation and Achieve Coordinated Development of Regional Economy

Regional cooperation plays a vital role in the coordinated development of economy. Good regional cooperation not only helps to improve the status and role of all regions in their respective division of labor, but also helps market players to obtain the required factor supply and product demand from a wider range. Henan Province is close to Anhui Province and Shandong Province in the east, and is closely linked to Shaanxi Province in the west. Anhui Province is rich in labor resources, tourism resources and tourism resources. Shandong Province not only has very large reserves of coal and oil, but also has very convenient transportation. Shaanxi Province has a wide variety of mineral resources. Strengthening cooperation with neighboring provinces can not only effectively improve the advantages of each region, but also effectively play the special capabilities of each region, forming a highly competitive regional economy with strong competitiveness according to local conditions, division of labor and cooperation, complementary advantages and common development.

COMPETING INTERESTS

The author have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Hu Jinyana, Cai Yongfanga, Cheng Jiaminga. Evaluation Model of the Regional Economic Vitality. Journal of Physics: Conference Series, 2020(3).
[2] Zhao Dan, Zhao Zhi, Chen Zening. Construction of Regional Economic Vitality Model. Journal of Economic Science Research, 2020(2): 19-23.
[3] Xin Siqian, Xin Sihan, Hu Buen, Wang Penghui, Fang Kang. Analysis and Decision of Regional Economic Vitality and its Influencing Factors. Frontiers in Economics and Management, 2020(1): 76-85.
[4] Pang Jinfeng, He Jintao, Luo Lan, Li Yao, Mao Yiren, Zhang Hanrui. Analysis of China's Regional Economic Vitality and Its Influential Factors Based on System Cluster Model and Computer Multiple Regression Model. Journal of Physics: Conference Series, 2020.
[5] Zhong Zhenfang. Factors Influencing Regional Economic Vitality Based on Regression Analysis. E3S Web of Conferences, 2020.
[6] Liu Yi, You Xiaoyu, Zhang Chunshuo. Regional Economic Vitality Based on Weighted Grey Relational Analysis. Journal of Economic Science Research, 2020(2): 12-18.
[7] Gan Mingli. Empirical Analysis on Influencing Factors of Financial Revenue in Sichuan Province. Journal of Economics and Public Finance, 2022(8): 68-74.
[8] Zhang Enquan. Empirical analysis of the influencing factors of fiscal revenue in Xinjiang. Economic Forum, 2015(5): 26-30.
[9] Guo Lin, Guo Li, Zhang Na, Tang Biao. Empirical Analysis on the Influencing Factors of Fiscal Revenue in China. Modern Business Trade Industry, 2016(27): 135-136.
[10] Li Lvxiu, Dong Xiyuan. Empirical analysis based on Eviews on China's fiscal revenue factors. China Collective Economy, 2021(16): 92-94.
[11] Xie Liying. Empirical Analysis on the Influencing Factors of Fiscal Revenue in Anhui Province. Modern Business Trade Industry, 2019(9): 115-116.
[12] You Jiawei. Empirical analysis of the influencing factors of China's fiscal revenue. China Collective Economy, 2019(7): 100-103.
[13] McDonald Gary C. Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 2009(1): 93-100.

[14] Tibshirani Rob. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B: Methodological, 1996(58): 267-288.

[15] Zou H. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association, 2006(101): 1418-1429.

[16] Nie Ruichao. Analysis of influencing factors of Fiscal revenue in Beijing based on Ridge regression and Lasso regression model. International Journal of New Developments in Engineering and Society, 2022(6): 1-5.