Analysis and decision-making of regional economic vitality and its influencing factors: Taking Beijing as an example

Xinyu Yang[1], Bailing Xu[1], Fanyi You[2]

[1] Institue of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, Gansu, 730100

[2] School of Management Engineering, Capital University of Economics and Businesses, Beijing, 100071

Abstract: This article focuses on how to improve the vitality of the regional economy, starting from Beijing, combining with the actual situation in Beijing in the past 15 years, using the mathematical methods or principles of multiple regression analysis, gray-scale prediction and analytic hierarchy process, and establishing a mathematical model of influencing factors. The economic vitality was analyzed and calculated, and reasonable conclusions were made using MATLAB software. Finally, the error analysis of the model results is carried out.

1. Background Knowledge

The economic vitality of a region (province or city) is an important part of a region's comprehensive competitiveness. In recent years, in order to improve economic vitality, some regions have launched many preferential policies to stimulate economic vitality, such as reducing the investment invitation and approval process, providing financial support for start-up enterprises, reducing the threshold of settlement, etc., in order to achieve the purpose of attracting foreign investment. However, due to the different allocation of urban resources, these policies have different impacts in different regions.

Take Beijing as an example: according to the statistics of Beijing Municipal Bureau of market supervision and administration, in 2018, Beijing issued a series of policies to improve the business environment, greatly reduced the threshold of market access, significantly reduced the start-up time of enterprises to 5 days, achieved outstanding results in the reform of registration facilitation, steadily increased the total number of market subjects, led the growth of the number of enterprises, and ranked the first in the country in terms of average capital per household The main structure of the dynamic market continued to be optimized. By the end of 2018, there were 21511000 real market entities (including branches) in Beijing, an increase of 55700 over the same period of last year, a year-on-year increase of 2.7%. The total amount of real capital is [1] 38.42 trillion yuan, an increase of 3.40 trillion yuan over the same period of last year, a year-on-year increase of 9.7%. The total amount of capital ranks the third in all provinces and cities of China after Guangdong and Zhejiang. The average household capital scale of the main market entities is 17.8587 million yuan, an increase of 6.9% year on year. The average household capital scale ranks first in all provinces and cities in China. Among all kinds of entities, the growth rate of enterprises is the fastest. By the end of 2018, there were 1654500 enterprises in the city, with a year-on-year growth of 6.5%, 3.8 percentage points higher than the average growth rate of market entities. There are 489000 individual businesses, down 8.4% year on year. The rapid growth of enterprises and the decline of individual businesses have promoted the continuous optimization of the main structure of the market. The individual ratio of enterprises increased from 2.91:1 in the same period last year to 3.38:1.

In order to study how to improve the regional economic vitality, we have obtained some data and established a suitable model, and solved the problems according to the data in the appendix and the data obtained through the survey.

First we take a region as an example to establish a model of influencing factors of economic vitality. We consider the population age structure, population migration, the number of newly added enterprises and the number of continued enterprises to reflect the trend of population change and the trend of enterprise vitality change respectively. Through the method of multiple regression analysis, we regard one variable as a dependent variable and the other one or more as a dependent variable Four variables are regarded as independent variables. The linear or non-linear mathematical model is established and the sample data is used for analysis.

Second by modifying the model of the first question, according to the relevant data of Beijing's economic vitality in the past 15 years and the transformation of economic policy obtained from the survey, the grey prediction method is used to identify the different degree of the development trend between the system factors, that is, the correlation analysis is carried out, and the original data is generated and processed to find the law of the system changes.

2. Model of factors affecting economic vitality

2.1. Analysis

Regression analysis is one of the most commonly used data analysis methods. It is based on the obtained test results and previous experience to establish a statistical model, and study the correlation between variables, establish the approximate expression of the relationship between variables, namely empirical formula, and then predict and control the corresponding variables.

For the requirement of question 1, we take Beijing as an example to establish a model of influencing factors of economic vitality. We consider the population age structure, population migration, the number of newly added enterprises and the number of continued enterprises to reflect the trend of population change and the trend of enterprise vitality change respectively. Through multiple regression analysis, we regard one variable as a dependent variable in the relevant variables, and the other one or Multiple variables are regarded as independent variables. The linear or non-linear mathematical model of multiple variables is established and the sample data is used for analysis. The final result is that the population change trend and the enterprise vitality change trend are positively correlated with the regional economic vitality change, but the former has a greater impact on the regional economic vitality than the latter.

2.2. Establishment of Model

Step 1: establish multiple regression analysis model
The model of multiple regression analysis is

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

Where $\beta_0, \beta_1, \cdots, \beta_m, \sigma^2$ is an unknown parameter independent of $x_1, x_2, \cdots, x_m$, and $\beta_0, \beta_1, \cdots, \beta_m$ is called regression coefficient.

Now, we get n independent observation data $(y_i, x_{i1}, \cdots, x_{im}), i = 1, \cdots, n, n > m$, which is obtained by formula (1)

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), i = 1, \cdots, n \end{cases}$$

record as

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\varepsilon = [\varepsilon_1 \quad \cdots \quad \varepsilon_n]^T, \beta = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_m]^T$$

Where E_n is the unit matrix of order n.
Step 2: parameter estimation

the parameter $\beta_0, \beta_1, \cdots, \beta_m$ in the formula is still estimated by the least square method, i.e. the estimated value $\hat{\beta}_j$ should be selected so that when $\beta_j = \hat{\beta}_j, j = 0,1,2,\cdots,m$, the sum of the squares of the errors

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2$$

To a minior this purpose,

$$\frac{\Box Q}{\Box \beta_j} = 0, j = 0,1,2,\cdots,n$$

Get

$$\begin{cases} \frac{\Box Q}{\Box \beta_0} = -2\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_m x_{im}) = 0 \\ \frac{\Box Q}{\Box \beta_j} = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_m x_{im})x_{ij} = 0, j = 0,1,2,\cdots,m \end{cases}$$

After finishing, it is transformed into the following normal equations

$$\begin{cases} \beta_0 n + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \cdots + \beta_m \sum_{i=1}^n x_{im} = \sum_{i=1}^n y_i \\ \beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1}x_{i2} + \cdots + \beta_m \sum_{i=1}^n x_{i1}x_{im} = \sum_{i=1}^n x_{i1}y_i \\ \beta_0 \sum_{i=1}^n x_{im} + \beta_1 \sum_{i=1}^n x_{im}x_{i1} + \beta_2 \sum_{i=1}^n x_{im}x_{i2} + \cdots + \beta_m \sum_{i=1}^n x_{im}^2 = \sum_{i=1}^n x_{im}y_i \end{cases}$$

When the X column of the matrix is full rank, $X^T Y$ is a reversible square matrix, and the solution of formula (8) is

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Replace $\hat{\beta}$ with the original model to get the estimated value of y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_m x_m$$

The fitting value of this group of data is $\hat{Y} = X\hat{\beta}$, and the fitting error $e = Y - \hat{Y}$ is called residual, which can be used as the estimation of random error ε

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

is the sum of the squares of the residuals (or the sum of the squares of the residuals), $Q(\hat{\beta})$.

Step 3: statistical analysis

(i)$\hat{\beta}$ is the linear unbiased minimum variance estimate of β. It means that $\hat{\beta}$ is a linear function of Y; $\hat{\beta}$'s expectation is equal to β; in F's linear unbiased estimation, $\hat{\beta}$'s variance is the smallest.

(ii)$\hat{\beta}$ obeys normal distribution

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

record as $(X^T X)^{-1} = (c_{ij})_{n \times n}$.

(iii)For the sum of squared residuals Q, $EQ = (n - m - 1)\sigma^2$, and

$$\frac{Q}{\sigma^2} \sim x^2(n - m - 1)$$

The unbiased estimation of σ^2 is obtained

$$s^2 = \frac{Q}{n - m - 1} = \hat{\sigma}^2$$

s^2 is the residual variance (the variance of the residual), and S is called the residual standard deviation.

(iv)decompose the total sum of squares $SST = \sum_{i=1}^n (y_i - \bar{y})^2$, including

$$SST = Q + U, \quad U = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Where q is the sum of residual squares defined by formula (15), reflecting the influence of random error on y, u is called the sum of regression squares, reflecting the influence of independent variables on y. The normal equations are used in the above decomposition.

Step 4: hypothesis test of regression model

It is necessary to test whether there is a linear relationship between dependent variable y and independent variable $x_1, \cdots, x_m$ as shown in model (1). Obviously, if all 's$|\hat{\beta}_j|(j = 1, \cdots, m)$are very small, the linear relationship between Y and $x_1, \cdots, x_m$ is not obvious, so the original assumption can be

$$H_0: \beta_j = 0 (j = 1, \cdots, m)$$

When H_0 holds, u and Q defined by decomposition (15) satisfy

$$F = \frac{\dfrac{U}{m}}{\dfrac{Q}{(n-m-1)}} \sim F(m, n-m-1)$$

There is upper α quantile $F_\alpha(m, n - m - 1)$ under significance level α, if $F < F_\alpha(m, n - m - 1)$, accept H_0; otherwise, reject. The acceptance of H_0 only indicates that the linear relationship between Y and $x_1, \cdots, x_m$ is not obvious, and there may be a nonlinear relationship, such as the square relationship. There are also some indicators to measure the degree of correlation between Y and $x_1, \cdots, x_m$, such as the ratio of regression square sum to the total square sum to define the complex determination coefficient

$$R^2 = \frac{U}{S}$$

$R = \sqrt{R^2}$ is called complex correlation coefficient. The larger R is, the closer the correlation between Y and $x_1, \cdots, x_m$ is. Generally, if R is greater than 0.8 (or 0.9), the correlation is considered to be valid.

Step 5: hypothesis test and interval estimation of regression coefficient

When H_0 above is rejected, β_j is not all zero, but several of them are equal to zero. Therefore, the following M tests should be carried out $(j = 0, 1, \cdots, m)$:

$$H_0^{(j)}: \beta_j = 0$$

From formula, $\hat{\beta} \sim N(\beta_j, \sigma^2 c_{jj})$ and c_{jj} are the (j, j) elements in $(X^T X)^{-1}$, s^2 is used instead of σ^2, from formula (12) - (14), when $H_0^{(j)}$ is established

$$t_j = \frac{\dfrac{\hat{\beta}}{\sqrt{c_{jj}}}}{\sqrt{\dfrac{Q}{(n-m-1)}}} \sim t(n-m-1)$$

For a given α, if $|t_j| < t_{\frac{\alpha}{2}}(n - m - 1)$, accept $H_0^{(j)}$; otherwise, reject.

Formula (18) can also be used to estimate interval $(j = 0, 1, \cdots, m)$ for β_j. At the confidence level, the confidence interval of $1 - \alpha$ and β_j is

$$\left[\hat{\beta}_j - t_{\frac{\alpha}{2}}(n-m-1)s\sqrt{c_{jj}}, \hat{\beta}_j + t_{\frac{\alpha}{2}}(n-m-1)s\sqrt{c_{jj}}\right]$$

inside $s = \sqrt{\dfrac{Q}{n-m-1}}$.

Step 6: use regression model to predict

When the regression model and coefficient pass the test, the given $x_0 = (x_{01}, \cdots, x_{0m})$ can predict that y_0 and y_0 are random. Obviously, the predicted value (point estimation) is

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_m x_{0m}$$

Given α, the prediction interval (interval estimation) of y_0 can be calculated, and the result is more complex. However, when n is large and x_{01} is close to the average value $\bar{x}_i$, the prediction interval of y_0 can be simplified as

$$\left[\hat{y}_0 - z_{\frac{\alpha}{2}} s, \hat{y}_0 + z_{\frac{\alpha}{2}} s \right]$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of the standard normal distribution.

The interval estimation method of y_0 can be used to give the confidence interval of the known data residual $e_i = y_i - \hat{y}_i (i = 1, \cdots, n)$. e_i obeys the normal distribution with the mean value of zero. Therefore, if the confidence interval of a e_i does not contain zero, the data is considered abnormal and can be `eliminated`.

2.3. Solution of Model

We use the proportion of 19-54-year-old population to reflect the age structure, the net migration to reflect the relationship between the population moving in and the population moving out, and the number of existing enterprises to reflect the relationship between the number of new enterprises. Using MATLAB software, draw the scatter diagram of these factors:
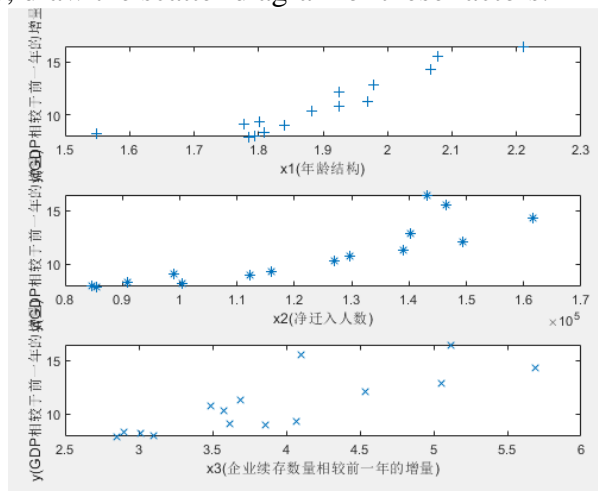
According to the scatter diagram, we establish a linear equation of three variables. According to the formula of the first step and the second step, we get four parameters are -13.7382, 9.7811, 0.3667, 3.3856, respectively, and the confidence interval of each parameter does not include 0, indicating that these parameters are normal. Therefore, the ternary linear equation is:

$$y = -13.7382 + 9.7811 x_1 + 0.3667 x_2 + 3.3856 x_3$$

Through the third and fourth steps, we can use matlab to get: the determinable coefficient of the model is 0.8966, which shows that 89.66% of the cases can be solved by this model; the observation value of F statistic is 31.8037, which is far greater than the critical value of F test; the probability p = 0.0001, which is far less than 0.05; the square sum of the model residual is 0.0342, which shows that the residual is basically normal.

3. Association analysis based on grey forecast

3.1. Analysis of Model

From 2004 to 2018, we selected population changes and enterprise vitality changes caused by economic policy changes in Beijing to reflect economic vitality through GDP growth compared with the previous year. Therefore, the short-term and long-term effects of economic policy changes on regional economic vitality can be concluded.

We use the grey time series forecast, that is, we use the observed time series which reflect the characteristics of the forecast object to construct the grey forecast model to predict the increment of GDP compared with the previous year. In this paper, the tested data series are accumulated once to generate one accumulated data series. Then, the least square linear regression model is used to select the best fitting curve to minimize the total fitting error, so as to obtain the estimated value of development coefficient and the estimated value of ash consumption. Finally, regression analysis and the white differential equation are used to obtain the prediction value of GDP increment compared with the previous year. Through analysis, the final conclusion is drawn.

3.2. Establishment of Model

In order to ensure the feasibility of GM (1,1) modeling method, it is necessary to test and process the known data. Set the original data sequence as $x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right)$, and calculate the rank ratio of the sequence

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2,3, \cdots, n$$

If all the stage ratios fall within the range of accommodation coverage

$$X = \left(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}}\right)$$

Then the data column $x^{(0)}$ can build GM (1,1) model and make grey prediction. Otherwise, the data shall be transformed appropriately, such as translation transformation:

$$y^{(0)}(k) = x^{(0)}(k) + c, k = 1,2, \cdots, n$$

Take c to make the ratio of data columns fall within the range of tolerance coverage.

The process of accumulating data in sequence is called accumulation generation process (ago). The sequence obtained from the accumulation generation process is called the accumulation generation sequence. Order

$$x^{(0)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1,2,3, \cdots, n$$

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right)$$

The new sequence $x^{(0)}$ obtained is called the first accumulation generating sequence of the sequence.

The timetable is $k = 2,3, \cdots, n$, and the grey derivative of definition $x^{(1)}$ is

$$d(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$$

Let $z^{(1)}(k)$ be the adjacent value of the sequence $x^{(1)}$ to generate the sequence, that is

$$z^{(1)}(k) = ax^{(1)}(k) + (1-a)x^{(1)}(k-1)$$

So the grey differential equation model of GM (1,1) is defined as

$$d(k) = az^{(1)}(k) = b$$

Or

$$x^{(0)}(k) + az^{(1)}(k) = b$$

In formula, $x^{(0)}(k)$ is called grey derivative, a is called development coefficient, $z^{(1)}(k)$ is called whitening background value, and b is called grey dosage.

Substituting schedule $k = 2,3,\cdots,n$ into formula (11) has

$$\begin{cases} x^{(0)}(2) + az^{(1)}(2) = b, \\ x^{(0)}(3) + az^{(1)}(3) = b, \\ \cdots\cdots \\ x^{(0)}(n) + az^{(1)}(n) = b \end{cases}$$

So GM (1,1) model can be expressed as

$$Y = Bu$$

Now the problem comes down to finding the values of a and B. Using linear regression of one variable, i.e. least square method, their estimated values are

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y$$

Whitening type of GM (1,1):

For t h e grey differential equation of GM (1,1), I f the time $k = 2,3,\cdots,n$ of grey derivative $x^{(0)}(k)$ is regarded as the continuous variable t, then $x^{(1)}$ is regarded as the function $x^{(1)}(t)$ of time t, then $x^{(0)}(k)$ corresponds to the derivative magnitude $\frac{dx^{(1)}(t)}{dt}$, and the whitened background value $z^{(1)}(k)$ corresponds to the derivative $x^{(1)}(t)$. So the grey differential equation of GM (1,1) corresponds to the white differential equation

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$$

The regression analysis is used to obtain the estimated values of a and B, so the corresponding whitening model solution is

$$x^{(1)}(t) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} + \frac{b}{a}$$

So we get the predicted value

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}, k = 1,2,\cdots,n-1$$

Accordingly, the predicted value is obtained:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 1,2,\cdots,n-1$$

Residual test: calculate relative residual

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, k = 1,2,\cdots,n$$

If all $|\varepsilon(k)| < 0.1$'s, the higher requirements are considered to be met; otherwise, if all $|\varepsilon(k)| < 0.2$'s, the general requirements are considered to be met.

3.3. Solution of Model

The range of accommodation coverage is $X = (0.8824969, 1.1331485)$. it can be seen that the order ratio of the sequence exceeds this range. Therefore, the data columns need to be translated. According to formula (24), set C = 3, and find the new data column as the original data column $x^{(0)}$ of the model.

If we make one accumulation of the original sequence $x^{(0)}$, we have one accumulation of the original sequence $x^{(1)}$.
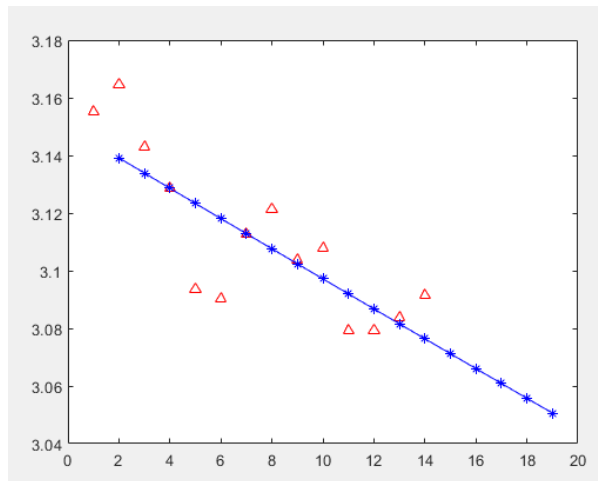

According to formula, the grey derivative of the first cumulative generating sequence is obtained. According to formula, the adjacent value of sequence is calculated to generate sequence. According to to formula, is called grey derivative, a is called development coefficient, is called whitening background value, and b is called grey dosage.

In this case, the matrix vector marks u, y, B are introduced, and the GM (1, 1) model can be expressed as , then the problem is reduced to the value of a, B. By using linear regression of one variable, i.e. the least square method, their estimated values can be obtained, a = 0.0017, B = 3.1471.

After calculating the values of a and B, the predicted value of   can be obtained according to the whitening type of GM (1,1). The predicted values are 3.1552, 3.1391, 3.1339, 3.1286 and 3.1233, respectively. Through the translation transformation, minus the value of C, the GDP of the next five years is 0.1252, 0.1191, 0.1139, 0.1186, 0.1133 respectively compared with the prediction of the previous year's increment.

According to formula (36), we can get all the relative residuals.

It can be seen from the above table that for all $|\varepsilon(k)| < 0.1$'s, we can think that our model meets the higher requirements and achieves the accurate prediction of economic vitality.
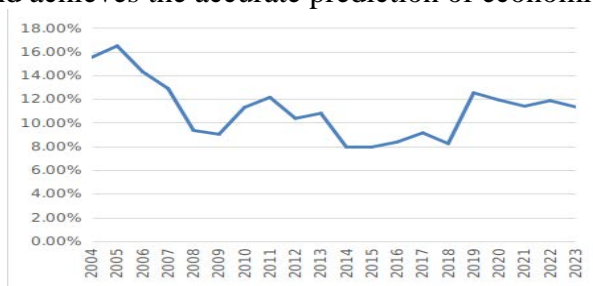
According to the above analysis and the above table, we can draw the conclusion that in the short term, the current economic policy change will make Beijing's economic vitality soar; but in the long term, the current economic policy change will lead to a downward trend of Beijing's economic vitality.

References

[1] Ma Zhengfei, application of mathematical calculation method and software engineering [M], chemical industry press, 2002.12.

[2] Zhao jingdanqi, mathematical modeling and mathematical experiment (3rd Edition) [M], higher education press, 2008.1.

[3] Lin Qining, decision analysis [M], Beijing, Beijing University of Posts and Telecommunications Press, February 2003.

[4] Sun Guomao, macro economy and capital market [M], Shandong people's publishing house, 2001.

[5] Stiglitz, Economics (Volume I and II) [M], China Renmin University Press, 1998.

[6] mankun, Macroeconomics (Fifth Edition) [M], China Renmin University Press, 2005.

[7] Gan Yingai, MATLAB 7.0 practical guide [M], Beijing, electronic industry press, 2005.

[8] Gan Yingai, operations research [M], Beijing, Tsinghua University Press, 2009.

[9] Qin Yuanyuan. Comprehensive evaluation of regional economic development vitality in China [J]. China Statistics, 1998 (12): 26-27