ISSN 2522-3488 Vol. 6, Issue 2: 1-5, DOI: 10.25236/IJNDES.2022.060201

Analysis of influencing factors of Fiscal revenue in Beijing based on Ridge regression and Lasso regression model

Nie Ruichao

School of Management of Science and Engineering, Anhui University of Finance and Economic, Bengbu, 233030, China
**Corresponding Author*

Abstract: *Based on the fiscal revenue and other relevant economic index data of Beijing from 1995 to 2020, this study uses the research methods of Ridge regression and Lasso regression to explore the influencing factors of Beijing's fiscal revenue. Considering that the traditional linear regression model will produce strong multicollinearity among many variables. Therefore, ridge regression and Lasso regression model were firstly used to reduce the influence of multicollinearity between variables, and then variable selection was carried out. Finally, the two models were compared according to the analysis results, and the optimal analysis model was selected. The results show that compared with ridge regression model, lasso regression model has better goodness of fit, smaller error and better model. The added value of the second industry, power generation, resident population, urban per capita disposable income and total retail sales of social consumer goods has a positive impact on fiscal revenue, and the whole social fixed assets investment, employment in cities and towns, per capita consumption expenditure of urban households is has a certain negative impact on fiscal income level.*

Keywords: *Ridge regression, Lasso regression, multicollinearity, financial revenue*

1. Theories and methods

1.1 Ridge regression model

Ridge regression was proposed by statistician Hoerl in 1962. In 1970, Hoerl worked with Kennard to further develop the ridge regression method systematically. Ridge regression is a biased estimation method based on least square estimation method, which is an improvement of least square estimation method. Some reliable and practical regression coefficients are not obtained, some information is lost, the accuracy is reduced, and the unbiasedness of the least square method is abandoned. The main solution of ridge regression is to add a small disturbance term on the basis of X^TX, so that the determinant is no longer 0, that is, the solution formula of parameters can be changed into:

$$\theta = (X^TX + \gamma I)^{-1}X^TY$$

In fact, ridge regression is the regular term of adding the sum of squares of parameters on the basis of the sum of squares of residuals, as shown below:

$$J(\theta) = \sum_{i=1}^{m}(y_i - \sum_{j=0}^{P}\theta_j x_{ij})^2 + \gamma\sum_{j=1}^{P}\theta_j^2$$

1.2 Lasso regression model

Lasso regression algorithm was proposed by Tibshirani in 1996. Lasso regression avoids the occurrence of too large regression coefficient, so as not to cause overfitting. Different from ridge regression, Lasso is a regular term that adds the absolute value sum of parameters to the cost function, as shown below:

$$J(\theta) = \sum_{i=1}^{m}(y_i - \sum_{j=0}^{P}\theta_j x_{ij})^2 + \gamma\sum_{j=1}^{P}|\theta_j^2|$$

ISSN 2522-3488 Vol. 6, Issue 2: 1-5, DOI: 10.25236/IJNDES.2022.060201

Suppose the regression equation $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = \sum_{j=0}^{p} \beta_j x_{ij}$, Mn and Lasso regression are equivalent to adding constraints on the estimated coefficients into the square sum of residual errors estimated by ordinary linear regression. Lasso regression adds constraint penalty term of absolute value sum of parameters (also called L_1 regularization), which can be expressed as:

$$\hat{\beta}^{ridge} = \underbrace{argmin}_{\hat{\beta}} \left\{ \sum_{i=1}^{m} (y_i - \sum_{j=0}^{p} \beta_j x_{ij})^2 \right\} subject\ to \sum_{j=0}^{p} |\beta_j| \le t$$

2. Construction of index system

Table 1: Revaluation form of influencing factors of fiscal revenue

index	variable	max	min	mean	median
Fiscal revenue	Y	5817.1	115.26	2297.467692	1664.98
GDP	X_1	36102.55	1516.2	14081.10192	11119.3
Fiscal expenditure	X_2	7471.43	154.4	2772.990769	1804.395
Added value of the secondary industry	X_3	5716.37	638.3	2783.783462	2470.05
Total volume of foreign trade	X_4	428995812	29318330	210632628.6	203866407
Generating capacity	X_5	464.09	132.21	261.7430769	235.365
Investment in fixed assets throughout society	X_6	9283	841.5	4387.889231	3860.965
Urban working population	X_7	1198.9	456.35	820.1253846	822.6
Permanent resident population	X_8	2195	1240	1743.380769	1723.5
Urban per capita disposable incom	X_9	78372	6235	30541.46038	23356.8
Per capita consumption expenditure of urban residents	X_{10}	49382	5019.76	20784.44577	15895.35
Total retail sales of consumer goods	X_{11}	15063.7	826.98	6034.976923	4240.35

In order to study the factors affecting the fiscal revenue of Beijing, the general public budget revenue is taken as the explained variable. In this paper, on the basis of reference to related literature, improvements on the index system of innovation, put forward the influencing factors, respectively from the level of economic development, the degree of industrial structure, trade, energy consumption and residents's expenditure to income, specific indicators of GDP respectively (X_1), fiscal expenditure (X_2), the added value of the second industry (X_3), the total import and export (X_4) , power generation (X_5), total investment in fixed assets (X_6), number of employed persons in urban units (X_7), number of permanent resident population (X_8), per capita disposable income of urban residents (X_9), per capita consumption expenditure of urban residents (X_{10}) and total retail sales of social goods (X_{11}).

3. Analysis of influencing factors of Fiscal revenue in Beijing

3.1 Multicollinearity test

Table 2: Multicollinearity test

variable	VIF	variable	VIF
X_1	7070.31	X_7	43.10
X_2	1483.36	X_8	182.13
X_3	1717.26	X_9	8443.29
X_4	110.75	X_{10}	2145.65
X_5	82.83	X_{11}	1150.94
X_6	962.5	Mean VIF	2199.29

Linear regression was carried out for all variables, and the regression results showed that the goodness of fit of the equation was good, but most of the variables did not pass the significance test. For this reason, variance inflation factor (VIF) is used to test the strength of collinearity of variables. If the value is greater than 10, it indicates the existence of serious multicollinearity. It can be seen from the following table that all variables have significant multicollinearity. Therefore, it is difficult to

ISSN 2522-3488 Vol. 6, Issue 2: 1-5, DOI: 10.25236/IJNDES.2022.060201

eliminate the influence of multicollinearity by using the traditional least square estimation method. Therefore, in this study, ridge regression and Lasso regression are used to analyze the influencing factors of Beijing's fiscal revenue.

3.2 Ridge regression analysis

Linear Model Ridge () is included in the SciKit-learn package in Python, so it is possible to implement Ridge regression directly using general commands.

The model equation after ridge regression is as follows:

$$Y = -1221.1329 + 0.4710X_3 + 1.6338X_5 - 0.1815X_6 - 0.9081X_7 + 0.9447X_8 + 0.0074X_9 - 0.0320X_{10} + 0.3491X_{11}$$

From the above model, it can be seen intuitively that the added value of the secondary industry (X3), power generation (X5), permanent resident population (X8), per capita disposable income of urban residents (X9) and total retail sales of social consumer goods (X11) change in the same direction as the financial revenue of Beijing. The total social investment in fixed assets (X6), urban employment (X7), and per capita consumption expenditure of urban households (X10) have a negative impact on Beijing's fiscal revenue.

The actual and predicted values of corresponding variables in the test set were compared and the ridge regression estimation graph was drawn as follows. As can be seen from the following figure, the predicted value of Y in the test set of ridge regression estimation is basically the same as the real value, but there is a large deviation between the recent and long-term time periods.

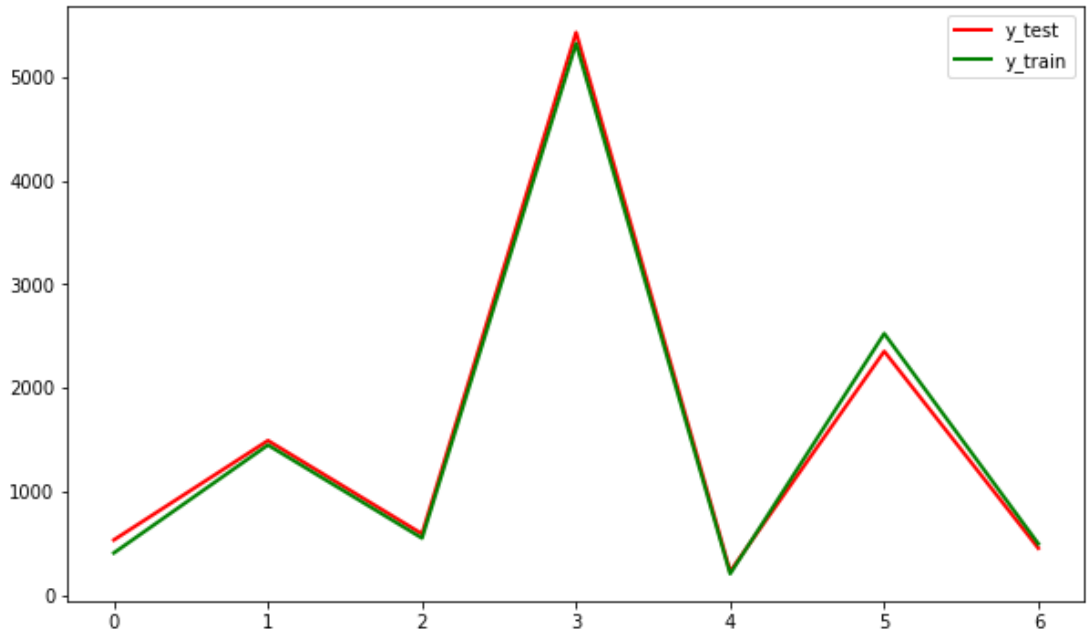

Figure 1: Rridge regression prediction

3.3 Lasso regression analysis

Lasso regression uses the coordinate descent method to solve the parameter estimates. Coordinate descent method selects parameters of one dimension each time for one-dimensional optimization, and then continuously selects and updates multiple dimensions until the function converges.

The Lasso regression model equation is as follows:

$$Y = -1191.7628 + 0.4344X_3 + 0.6370X_5 - 0.1857X_6 - 0.7362X_7 + 0.9447X_8 + 0.0057X_9 - 0.0217X_{10} + 0.3581X_{11}$$

From the above model, it can be seen intuitively that the added value of the secondary industry (X3), power generation (X5), permanent resident population (X8), per capita disposable income of urban residents (X9) and total retail sales of social consumer goods (X11) change in the same direction with

ISSN 2522-3488 Vol. 6, Issue 2: 1-5, DOI: 10.25236/IJNDES.2022.060201

the financial revenue of Beijing. The total social investment in fixed assets (X6), urban employment (X7), and per capita consumption expenditure of urban households (X10) have a negative impact on Beijing's fiscal revenue. The analysis results are basically the same as those of ridge regression model. The actual and predicted values of corresponding variables in the test set were compared to draw the following LasSO regression estimation graph.

As can be seen from the figure below, the predicted value of Y in the test set of ridge regression estimation is basically the same as the real value, and the two tend to be the same in the short-term, mid-term and long-term prediction, with good fitting.

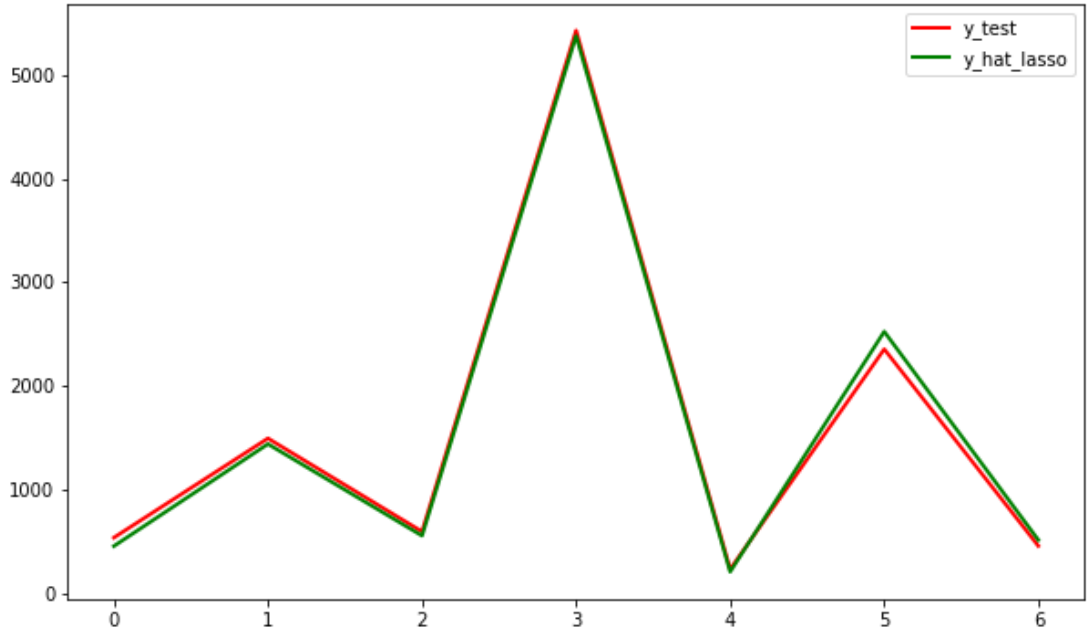

Figure 2: Lasso regression prediction

3.4 Comparative analysis of results

After using ridge regression model and Lasso regression model to select variables, the multicollinearity between variables was eliminated, and the better model was selected by comparing model test and parameter test. The following table shows the analysis results of Ridge regression model and Lasso regression model intuitively, and the fitting results of lasso regression model are better according to goodness of fit R^2. At the same time, the prediction errors obtained by Lasso regression on MAE, MSE and RMSE are all smaller than those obtained by linear model of ridge regression.

Therefore, considering the analysis results of the two models comprehensively, the Lasso regression model is better.

Table 3: Model comparative analysis

model	R^2	MAE	MSE	RMSE
Ridge	0.9969	78.8245	8895.0621	94.3136
Lasso	0.9977	68.6551	6664.7105	81.6377

4. Conclusions and Suggestions

4.1 Conclusions

Through the analysis of ling regression model and Lasso regression model, it can be concluded that the added value of the secondary industry, electricity generation, permanent population, per capita disposable income of urban residents and total retail sales of social consumer goods have a positive impact on the fiscal revenue of Beijing. At the same time, electricity generation and resident population have a high degree of impact on Beijing's fiscal revenue, which directly reflects the regional energy

ISSN 2522-3488 Vol. 6, Issue 2: 1-5, DOI: 10.25236/IJNDES.2022.060201

development and the number of regional population will have a great impact on Beijing's fiscal revenue.

In addition, the three major factors of total social investment in fixed assets, urban employment and per capita consumption expenditure of urban households have a negative impact on Beijing's fiscal revenue. Social fixed assets reflect the total amount of fixed assets currently owned by the society, while urban employment and consumption expenditure of urban residents reflect people's living standards. The state should also strengthen macro-control to maintain the positive growth of Beijing's fiscal revenue.

4.2 Relevant suggestions

(1) Strengthen population attraction, especially the introduction of high-quality labor and human resources. The resident population of Beijing is an important force to promote the regional fiscal revenue. Meanwhile, it is necessary to strengthen the introduction of all kinds of human resources and high-quality labor force to meet the social demand for labor force and improve social productivity. At the same time, the use of regional talents to carry out industrial innovation in various industries, comprehensively improve production efficiency, improve the financial revenue of Beijing.

(2) Promote the reform of industrial structure and the mutual transformation among the three industries. The three major industries are the important force to promote the regional economic development, especially the secondary industry and a large number of new tertiary industries. The social value created by the two industries is rising with the development of time. At the same time, while stabilizing the proportion of the primary industry, we will strengthen the steady development of a series of manufacturing industries included in the secondary industry. Actively encourage the development of the tertiary industry, especially the green industry and some high-tech industries, which will have a huge social influence in the future.

(3) Introduce relevant policies to improve people's living standards and stabilize the social consumption environment. Policy support and guarantee of people's living standard are the basic requirements for stabilizing Beijing's financial revenue, especially to ensure the stability of social consumption environment, maintain normal price level and guarantee people's quality of life.

References

[1] Bian Yunuo, Research on the Influence of China's Fiscal Revenue based on ANN-MLP model. Economic Research Guide,2022(06):67-69.]

[2] Lu W J, A study on inter-governmental fiscal revenue distribution and its Impact in China [D]. Zhejiang university, 2021. DOI: 10.27461 /, dc nki. Gzidx. 2021.002691. E

[3] Qin Quan. Analysis and prediction of influencing factors of Fiscal Revenue in Hunan Province based on Python software. The Chinese market, 2021 (29): 40 and 41. DO1:10.13939 / j.carol carroll nki. ZGSC. 2021.29.040.

[4] Ke Li. The dynamic analysis of local finance income gap of our country [D]. Chinese finance research institute, 2021. The DOI: 10.26975 /, dc nki. GCCKS. 2021.000057.

[5] Li Luxiu, Dong Xiyuan, Empirical analysis of influencing factors of China's fiscal revenue based on EViews. China Collective Economy,2021(16):92-94.]

[6] Zhang Yingchun, resource tax reform's influence on the fiscal revenue in shandong province research [D]. Shandong university of finance and economics, 2021, DOI: 10.27274 /, dc nki. GSDJC. 2021.000410.

[7] NiJie, shandong finance income factors and forecast analysis [D]. Shandong normal university, 2021. The DOI: 10.27280 /, dc nki. Gsdsu. 2021.001185.

[8] Zhao Hui, Correlation analysis between Fiscal Revenue and GDP in Liangcheng County []. Inner Mongolia Science & Technology and Economy,2021(08):67-68.

[9] Wang Ran, Li Tingshan, Zhan Xiaoyu, Empirical Analysis of Factors Affecting China's Fiscal Revenue [C]// 2021 International Conference on Innovative Talent Training and Sustainable Development proceedings (Chinese).. 2021:77-79. DOI: 10.26914 / Arthur c. nkihy. 2021.011622.

[10] Xie Yi and Analysis on the impact of Guangdong's Foreign Trade import and Export on Fiscal Revenue [1]. Economic Forum,2021(03):46-55.