Analysis of Factors Affecting Fiscal Revenue in Guangxi Zhuang Autonomous Region

Xuehan Song

School of Mathematics and Statistics, Guangxi Normal University, Guilin 541006, Guangxi, China
DOI: 10.32629/memf.v5i5.2876

Abstract: Based on the fiscal revenue data and related economic indicators of Guangxi Zhuang Autonomous Region from 1994 to 2022, this paper explores the main influencing factors of fiscal revenue in Guangxi Zhuang Autonomous Region. The first step is to test the correlation and multicollinearity between the variables. According to the test results, there is a strong multicollinearity among the independent variables. Therefore, the variables need to be screened. Ridge regression, lasso regression and adaptive lasso regression are used for variable selection in this paper. The three models are then evaluated and it is concluded that the lasso regression model provides the best fit. According to the lasso model, the factors that have a more pronounced impact on fiscal revenue are: tax revenue, total retail sales of consumer goods, education expenditure and the number of college graduates. Finally, relevant suggestions are made for the selected key factors to increase Guangxi's fiscal revenue.

***Keywords*:** ridge regression; lasso regression; adaptive lasso regression; fiscal revenue

1. Introduction

In 2023, the general public budget revenue of Guangxi Zhuang Autonomous Region was 178.380 billion yuan, the general public budget expenditure was 610.258 billion yuan, and the balance of revenue and expenditure was 431.878 billion yuan. There is a serious imbalance between fiscal revenue and expenditure. Increasing fiscal revenue has become the top priority at present. Therefore, it is of great importance to study the influencing factors of fiscal revenue in Guangxi and conduct effective screening. In the analysis process, due to the involvement of multiple independent variables, it is necessary to screen the independent variables to avoid excessive model complexity. For this purpose, ridge regression, lasso regression and adaptive lasso regression are used to screen variables and reduce errors through modelling. Finally, based on the characteristics of economic development in Guangxi, policy suggestions are made to promote the increase of fiscal revenue.

2. Basic principles of the model

Ridge regression, lasso regression and adaptive lasso regression are all linear regression methods used to solve multicollinearity problems and perform feature selection. They control model complexity by introducing different regularization terms into the regression model.

2.1 Ridge regression

Ridge regression is a variant of linear regression that suppresses the complexity of the model by adding an L_2 regularization term to the loss function. The L_2 regularization term is the sum of the squares of the parameters. The loss function takes the form:

$$Q(\beta)=(y-X\beta)^T(y-X\beta)+\lambda\sum_{j=1}^{p}\beta_j^2$$

Where, λ is the regularization parameter, which controls the strength of the penalty term.

2.2 Lasso regression

Lasso regression performs feature selection and parameter estimation by adding an L_1 regularization term to the loss function. The L_1 regularization term is the sum of the absolute values of the parameters. Its loss function is:

$$Q(\beta)=(y-X\beta)^T(y-X\beta)+\lambda\sum_{j=1}^{p}|\beta_j|$$

Where, λ is the regularization parameter, which is used to control the strength of the penalty term.

2.3 Adaptive lasso regression

Adaptive lasso regression is an improved method based on lasso regression. By applying different weights to different regression coefficients, lasso regression is more robust and selective. The loss function of adaptive lasso is as follows:

$$Q(\beta)=(y-X\beta)^T(y-X\beta)+\lambda\sum_{j=1}^{p}\omega_j|\beta_j|$$

Where, λ is the regularization parameter, ω_j is the weight related to the regression coefficient, usually $\omega_j=\frac{1}{|\hat{\beta}_j|^{\gamma}}$, $\hat{\beta}_j$ is the initial estimate, and $\gamma>0$ is a control parameter.

3. Empirical analysis

3.1 Data sources and selection of indicators

To improve the accuracy and credibility of the model, as many samples as possible should be selected. However, blind selection may increase the error due to the inconsistent statistical quality of the previous and subsequent data. Therefore, this paper selects data from 1994, when the tax sharing system was introduced, to 2022 for analysis.

In order to study the factors affecting the fiscal revenue of Guangxi Zhuang Autonomous Region, this paper takes the general public budget revenue of Guangxi Zhuang Autonomous Region as the dependent variable. By combining the current economic situation of Guangxi Zhuang Autonomous Region, starting from the financial field, national economy, import and export, and in order to consider the influence of various factors on the fiscal revenue of Guangxi Zhuang Autonomous Region as much as possible, choose x_1 to represent tax revenue, x_2 GDP of the primary industry, x_3 GDP of the secondary industry, x_4 GDP of the tertiary industry, x_5 GDP per capita, x_6 total retail sales of consumer goods, x_7 per capita disposable income of urban residents, x_8 per capita disposable income of rural residents, x_9 education expenditure, x_{10} total export value, x_{11} total import value , x_{12} the number of college graduates.

3.2 Descriptive analysis

First, the original data of local fiscal revenue and influencing factors in Guangxi Zhuang Autonomous Region are analyzed descriptively, and the basic statistical analysis is made in the form of tables to find out the internal laws of these data. Table 1 shows the results of the four main statistics of minimum, maximum, mean and standard deviation of the variable data.

Table 1. Descriptive statistics of variables

Variable	*Min*	*Max*	*Mean*	*SD*
y	62.3	1811.9	783.9	669.2
x_1	47.3	1191.1	509.4	437.6
x_2	333.8	4269.8	1701.9	1205.3
x_3	468.6	8938.6	3297.8	2716.7
x_4	395.9	13092.5	4299.5	4112.7
x_5	2675	52164	19224	15973.1
x_6	395.8	8539.1	3397.6	2911.9
x_7	3981	39703	17172	11817.7
x_8	1107	17433	5988	5057.7
x_9	27.4	1141.7	416.2	399.2
x_{10}	102.3	3705.4	963.9	1029.6
x_{11}	41.9	2991.5	809.9	922
x_{12}	1.3	36.2	12	9.6

The general public budget revenue of Guangxi Zhuang Autonomous Region increased from 6.23 billion yuan to 181.19 billion yuan between 1994 and 2022, indicating that the increase is very large and that Guangxi's economy is developing rapidly. Tax revenue increased from 4.73 billion yuan to 119.11 billion yuan, indicating that the income of individuals and enterprises in Guangxi has increased rapidly. Comparing the GDP of primary industry, secondary industry and tertiary industry in Guangxi, the tertiary industry has increased the most, indicating that Guangxi's tertiary industry has the best development and construction, and tourism is developing rapidly. The per capita disposable income of both urban and rural residents has increased significantly, but the per capita disposable income of rural residents is much lower than that of urban residents, indicating that Guangxi should strengthen rural construction and vigorously develop rural revitalization. The number of college graduates in Guangxi has also increased from 13,000 to 362,000, indicating that the construction of colleges and universities has achieved good results.

3.3 Correlation analysis and multicollinearity test

As the dimensions of the 12 selected variables are different, it is necessary to standardize the original data. After standardizing the data, the correlation between the variables is tested using the person correlation coefficient, and we can see that the correlation coefficients between the variables are all greater than 0.8. From this we can conclude that there may be multiple collinear problems between the explanatory variables.

The condition number is used to assess the presence of multicollinearity. When the condition number is large (generally greater than 30), it indicates the existence of multicollinearity. The condition number of the variables studied in this paper can be calculated directly by the kappa() function of the R software as follows $k = 412510 \geq 31$, That is, there is serious multicollinearity among various explanatory variables, so it is impossible to establish a simple multiple regression model to analyze the main factors affecting fiscal revenue.

A review of the literature shows that collinearity between independent variables is very common in the process of modelling economic problems. In this paper, ridge regression, lasso regression and adaptive lasso regression are used to deal with multicollinearity.

3.4 Ridge regression

Solved by glmnet package in R, the ridge regression coefficients are:

Table 2. Ridge regression coefficients

Variable	Ridge regression coefficient	Variable	Ridge regression coefficient	Variable	Ridge regression coefficient
x_1	0.3244	x_5	0.0822	x_9	0.1542
x_2	0.0698	x_6	0.1428	x_{10}	-0.0364
x_3	0.1016	x_7	0.1157	x_{11}	-0.0290
x_4	0.0214	x_8	0.0053	x_{12}	0.0309

As can be seen from Table 2, the ridge regression coefficient of the variable x_1 is the largest, at 0.3244, indicating that tax revenue has the greatest impact on fiscal revenue. For every unit increase in tax revenue, general public budget revenue increases by 0.3244 units. The second and the third largest are x_9 education expenditure and x_6 total retail sales of consumer goods, with ridge regression coefficients of 0.1542 and 0.1428 respectively, followed by x_7 per capita disposable income of urban residents and x_3 GDP of the secondary industry.

3.5 Lasso regression

Using lasso regression to screen variables, and also using the glmnet package of R language, the lasso regression coefficients are:

Table 3. Lasso regression coefficients

Variable	x_1	x_6	x_9	x_{12}
Lasso regression coefficient	0.6671	0.0174	0.2574	0.0588

Table 3 shows that the lasso regression selects four variables: x_1, x_6, x_9 and x_{12}. Among these, x_1 has the largest regression coefficient for tax revenue, at 0.6671, followed by x_9, which has a regression coefficient of 0.2574 for education expenditure. The regression coefficients for x_6 and x_{12} are both less than 0.1, indicating that tax revenue and education

expenditure have the greatest impact on fiscal revenue.

3.6 Adaptive lasso regression

According to the glmnet package, the two variables selected are x_1 tax revenue and x_9 education expenditure, with coefficients of 0.7964 and 0.1982 respectively, indicating that the screening intensity of the adaptive lasso is too large.

In order to compare the fitting effects of ridge regression, lasso regression and adaptive lasso regression, the RMSE and R2 of the above three methods are compared, and the results are shown in Table 4:

Table 4. Analysis table of fitting degree of three methods

Method	RMSE	R^2
Ridge regression	0.09322599	0.9913741
Lasso	0.04083455	0.9983156
Adaptive Lasso	0.05540671	0.9968813

By comparing the RMSE and R2 of the three methods, we can see that the lasso regression is the best fit.

4. Conclusions and recommendations

4.1 Conclusion

Through lasso regression, the key factors affecting Guangxi's fiscal revenue include tax revenue, total retail sales of consumer goods, education expenditure and the number of college graduates. Tax revenue has the greatest impact on fiscal revenue, education expenditure promotes fiscal revenue, and total retail sales of consumer goods reflects the change in consumption level.

4.2 Suggestions

(1) Increase tax revenue: Under the condition of tax reduction, reasonably increase tax revenue by stimulating the economy and reducing non-essential expenditures.

(2) Increase investment in education: focus on the effective investment of education funds, support talent cultivation, and promote economic development and fiscal revenue growth.

(3) Promote consumption growth: encourage emerging consumption forms, promote rural consumption, and expand domestic demand to increase fiscal revenue.

References

[1] Jiang Xin, Zhao Yang. An empirical study on the influencing factors of China's fiscal revenue [J]. China Township Enterprise Accounting, 2019, 08: 105-106.

[2] Ding Xianwen, Yuan Hong. Fiscal revenue estimation of Jiangsu Province based on ridge regression [J]. Journal of Jiangsu Institute of Technology, 2020, 26 (6): 5-10.

[3] He Xueping, Li Xingxu. Analysis of factors affecting fiscal revenue in Yunnan Province [J]. China Market, 2017, 19:47-51.

[4] Dong Xiaogang, Diao Yajing, Li Huiling, et al. Analysis of fiscal revenue factors under Ling regression, LASSO regression and Adaptive-LASSO regression [J]. Journal of Jilin Normal University, 2018, 39 (2): 45-53.

[5] Wu Xizhi. Statistical methods of complex data [M]. Beijing: Renmin University of China Press, 2012.

[6] Feng Haiqi, Zhao Yumei, Wang Linlin. Analysis of influencing factors of retail sales of social consumer goods in Shandong Province-based on Lasso regression model [J]. Journal of Ludong University (Natural Science Edition), 2021, 37 (4): 310-314.

[7] He Xiaoqun, Liu Wenqing. Applied Regression Analysis [M]. 4th edition, Beijing: Renmin University of China Press, 2015.

[8] Hu Yuwen. Analysis of influencing factors of R&D investment intensity in universities in Jiangxi Province based on LASSO regression [J]. Science and Technology and Industry, 2020, 20 (5): 84-88.

[9] Zhu Hailong, Li Pingping. Analysis of influencing factors of fiscal revenue in Anhui Province based on Ridge regression and LASSO regression [J]. Journal of Jiangxi University of Science and Technology, 2022, 43 (1): 59-65.

[10] Ji Chao. Evaluation of influencing factors of my country's fiscal revenue and its forecasting methods [J]. Local Finance Research, 2016 (2): 41-46.

[11] Wang Qi, Guo Shuang. Forecast analysis of fiscal revenue in Gansu Province [J]. Chinese Market, 2018 (28): 39-40.

[12] Zhao Xufang, Lu Wei. Empirical study on taxation influencing factors based on stepwise regression method [J]. China Township Enterprise Accounting, 2020 (08): 25-27.
[13] Yu Li. Statistical analysis of factors affecting China's fiscal revenue [J]. Journal of Qinghai University (Natural Science Edition), 2015, 33 (03): 90-93+100.
[14] Shuanghua. Fiscal revenue forecast of Hebei Province based on grey GM (1, N) model [J]. Journal of Shijiazhuang University, 2018, 20 (01): 18-23.
[15] Sun Yuan, Lu Ning. Local fiscal general budget revenue forecast model and empirical analysis [J]. Research on Quantitative Economics and Technology Economics, 2007 (01): 38-45.