

Analysis and Forecast of Shanghai Financial Revenue Based on Data Mining

Wei Deng

Baoshan Campus of Shanghai University in China

m18351996052@163.com

Keywords: Shanghai fiscal revenue; Adaptive-Lasso; Grey prediction; Neural network

Abstract. After 1994, the fiscal management system that China began to implement was a tax-sharing system. Local fiscal revenue was an important fund for local governments to carry out macroeconomic regulation and control. Based on the data of Shanghai's fiscal revenue and its influencing factors from 1994 to 2016, based on the Adaptive-Lasso variable selection method, the combination of grey prediction and neural network is used to fit and predict Shanghai's fiscal revenue. Providing data support for the Shanghai Municipal Government, and also providing reference for other cities with faster economic development.

Introduction

As a coastal developed city, Shanghai has experienced rapid economic development since the reform and opening up. All economic indicators are far ahead in the country and are known as China's "financial center". The growth of Shanghai's fiscal revenue is closely related to the growth of Shanghai's economy. At present, although Shanghai's income scale and structure have reached the national top level, compared with cities in developed countries, there is still room for continuous improvement. Therefore, how to accurately predict Shanghai's fiscal revenue and provide information for government management decisions has become a matter of great concern to everyone.

The significances of this paper include: On the one hand, establish a forecast model of fiscal revenue, and use the data of Shanghai Statistical Yearbook for empirical analysis to supplement and support the study of Shanghai's fiscal revenue forecasting model. On the other hand, use advanced methods to improve the accuracy of the fiscal revenue forecasting model. At present, most of the local fiscal revenue forecasting models are based on empirical estimates, and the scientific and reproducible results are not high. Based on advanced statistical analysis methods, this paper establishes a reasonable model of Shanghai's fiscal revenue, which not only provides data support for the macro-control of the Shanghai government, but also provides reference for other cities with faster economic development.

Literature References

Cointegration theory. The cointegration theory mainly uses the mutual influence relationship between GDP and fiscal revenue to predict tax revenue and fiscal revenue. Wei Bangrong and Yang Yusheng [1] established an error correction model for the relationship between fiscal revenue and GDP in China based on the cointegration model. The results show that although China's fiscal revenue and GDP are both non-stationary sequences, there is a long-term stable cointegration relationship between them. Cao Haijuan, Hu Yijian, Feng Suzhen [2] Based on the cointegration model to study the cointegration relationship between Shanghai's fiscal revenue and industrial structure and Shanghai's GDP, the study shows that Shanghai's fiscal revenue and industrial structure and Shanghai's GDP have long-term co-integration. Relationship.

Time series prediction. Zhang Lunjun [3] built and predicted the tax revenue by establishing autoregressive model, time series model and multiple linear regression model. The three models were compared and analyzed, and the original model was compared by comparing the results. Improvement.

Regression model. Cheng Maolin and Zhang Lunjun [4] established and introduced two linear models of the relationship between fiscal revenue and the added value of the three industries. Bai Ping [5] introduced the process and steps of establishing an econometric model in detail through an example, and used multiple linear regression analysis to analyze the practical problems. By qualitative analysis, several influencing factors have been selected which have strong correlation with China's fiscal revenue, and use it as an explanatory variable to establish a linear model with fiscal revenue.

Methods

Adaptive-Lasso. Lasso is a compression estimate that allows parameter estimation to be performed simultaneously with variable selection and has demonstrated Lasso's parameter estimation consistency. Its parameter estimate is defined as:

$$\hat{\beta}(lasso) = \underset{\beta}{\operatorname{argmin}} \left\| y - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (1)$$

Among them, λ is a non-negative regular parameter and $\sum_{j=1}^p |\beta_j|$ is a penalty.

Grey prediction. Let the variable $X^{(0)} = (X^{(0)}(i), i = 1, 2, \dots, n)$ be a non-negative monotonic original data sequence, first for $X^{(0)}$ The accumulation is obtained once to obtain the sequence $X^{(1)} = (X^{(1)}(k), k = 1, 2, \dots, n)$. For $X^{(1)}$, the following first-order linear differential equation can be established:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \quad (2)$$

Solve the prediction model:

$$\hat{X}^{(1)}(k+1) = \left[\hat{X}^{(1)}(0) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \quad (3)$$

BP neural network. BP neural network is a multi-layer feedforward network trained by error inverse propagation algorithm. The learning algorithm is a delta learning rule, which is widely used. The objective function of BP neural network is adopted:

$$E = \sum_{k=1}^N [Y_k - T_k]^2 \quad (4)$$

The BP algorithm uses the error after output to estimate the error of the previous layer, and then uses the error of the previous layer to estimate the error of the previous layer. This layer has been propagated backwards one by one, so that the errors of all other layers can be obtained. estimate. This results in a process of transferring the error obtained by the output layer to the original input layer in a reverse direction.

Results

Variable description. Table 1 is a description of the variables:

Table 1 Description of the variables

| Variable meaning | Abbreviation |
|--|--------------|
| Revenue | y |
| Average wage of employees in Shanghai | x1 |
| Shanghai total retail sales of consumer goods | x2 |
| Total population at the end of the year | x3 |
| Shanghai fixed assets investment | x4 |
| Shanghai Gross Domestic Product | x5 |
| Shanghai's primary industry output value | x6 |
| tax | x7 |
| Shanghai Residents' Consumption Index | x8 |
| Shanghai residents' consumption level | x9 |
| Ratio of output value of tertiary industry to secondary industry | x10 |
| Shanghai urban per capita disposable income | x11 |

Descriptive statistics. Descriptive statistics are used as the basis for analytical modeling, and Table 2 gives descriptive statistics for independent and dependent variables. As can be seen from Table 2, the average value of Shanghai's fiscal revenue (y) is 1636.77, and the standard deviation is 1418.82, which indicates that Shanghai's fiscal revenue experienced a huge leap from 1994 to 2014.

Table 2 Descriptive statistics

| | Min | Max | Mean | Sd |
|-----|----------|----------|----------|----------|
| y | 175.33 | 4585.55 | 1636.77 | 1418.82 |
| x1 | 5973 | 100623 | 40619.71 | 29787.7 |
| x2 | 770.8 | 9303.5 | 3649.26 | 2695.8 |
| x3 | 12167000 | 14386900 | 13547867 | 563924.8 |
| x4 | 1123.29 | 6016.43 | 3364.94 | 1578.31 |
| x5 | 1990.86 | 23567.7 | 10129.78 | 7001.55 |
| x6 | 47.61 | 127.8 | 91.57 | 23.78 |
| x7 | 179.95 | 4219.05 | 1492.96 | 1326.7 |
| x8 | 99.6 | 123.9 | 104.2 | 6.16 |
| x9 | 5081 | 43007 | 19577.24 | 11892.88 |
| x10 | 0.69 | 1.87 | 1.18 | 0.31 |
| x11 | 5889 | 47710 | 20812.9 | 12882.2 |

Model result. In this paper, the LARS algorithm is used to solve the Adaptive-Lasso estimation. In Eq.1, $\lambda = 1$ is used, and the R language is used to program the operation. The results are shown in Table 3 below.

Table 3 The coefficient of the Adaptive-Lasso variable selection model

| | | | | | |
|---------|--------|--------|--------|--------|-----------|
| x1 | x2 | x3 | x4 | x5 | x6 |
| -0.0094 | 0.0310 | 0.0000 | 0.0000 | 0.0763 | -2.6709 |
| x7 | x8 | x9 | x10 | x11 | Intercept |
| 0.5192 | 0.0000 | 0.0000 | 0.0000 | 0.0353 | -132.3675 |

In order to investigate the prediction accuracy of gray prediction, this paper compares the posterior difference ratio test of each variable gray prediction with the posterior difference test

discriminant reference, as shown in Table 4, using gray prediction for the 2015 and 2016 independent variables. Prediction has good accuracy.

Table 4 Gray prediction accuracy table

| | x1 | x2 | x5 | x6 | x7 | x11 |
|---------------------------|-----------|----------|----------|--------|---------|----------|
| 2015 forecast | 127275.80 | 10761.77 | 29592.16 | 137.98 | 6383.90 | 53359.51 |
| 2016 forecast | 143410.20 | 12200.61 | 33115.20 | 143.50 | 7393.90 | 59176.05 |
| Prediction accuracy level | good | good | good | good | good | good |

The predicted values of the above-mentioned independent variables obtained by the gray prediction are brought into the trained neural network prediction model. The R language has multiple packages specifically for the neural network. This paper uses the nnet package in the R language and sets its parameters to The number of learning is 10,000, the number of neurons is five, and the error accuracy is 10^{-7} . Fig. 1 shows a comparison of the original and predicted fiscal revenues of Shanghai. It can be seen that the predicted values are very close to the real values, and the fitting accuracy is very high.

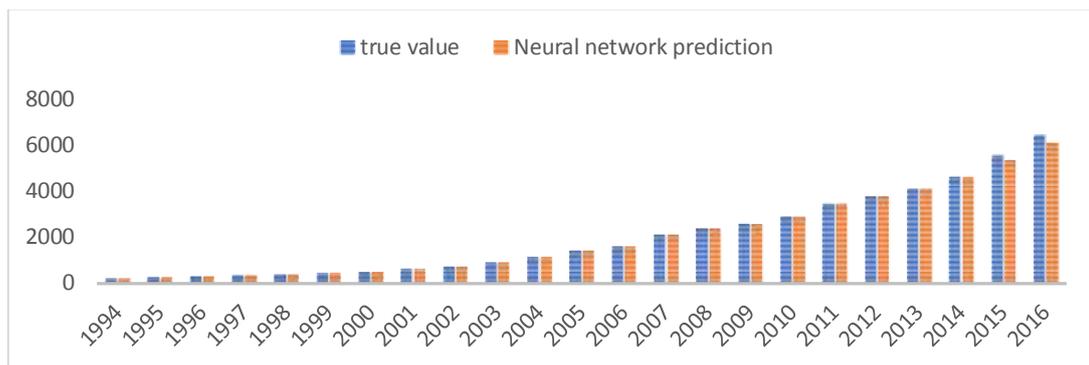


Figure 1. The true and predicted value of Shanghai's fiscal revenue

Conclusion

In the model studied in this paper, 11 variables that may affect Shanghai's fiscal revenue were included. Five variables were excluded by the Adaptive-Lasso method, namely the total population at the end of the year (x3), the amount of Shanghai fixed assets investment (x4), and Shanghai residents. Consumption index (x8), tertiary industry and secondary industry output ratio (x9), Shanghai urban per capita disposable income (x10), that is, in the model construction process, the coefficients of these five variables are all 0. The remaining variables are the average wage of employees on the job (x1), total retail sales of social consumer goods (x2), total production value of Shanghai (x5), output value of primary industry (x6), taxation (x7), and disposable income of Shanghai urban residents (x11).

Discussion of culling variables. Due to the large number of migrant workers from Shanghai, the total population of Shanghai at the end of the year (x3) has almost negligible impact on fiscal revenue; due to Shanghai's fixed asset investment (x4), tertiary industry and secondary industry output ratio (x9) and Shanghai's gross domestic product (x5) has obvious collinearity. The Adaptive-Lasso method will only screen out Shanghai's gross domestic product (x5), which has the biggest impact on fiscal revenue, and will invest in Shanghai's fixed assets (x4). The industry-to-secondary industry output ratio (x9) is excluded; since the correlation between household consumption level (x8) and fiscal revenue is too small, that is, there is no obvious linear relationship, the Adaptive-Lasso method will eliminate the variables that have no significant linear

relationship; Shanghai residents' consumption level (x9) and Shanghai urban per capita disposable income (x11) have obvious collinearity with Shanghai's GDP. This shows that using the Adaptive-Lasso method can eliminate the explanatory variables that have no significant influence on the explanatory variables. You can also delete the independent variables that have a collinear relationship, which also reflects the excellent performance of this method in modeling multiple variables.

Discussion of reserved variables. The average wage of employees on the job (x1) has a certain impact on Shanghai's fiscal revenue. Personal income tax is an important component of fiscal revenue. The average wage of employees on the job directly affects personal income tax, which indirectly affects fiscal revenue. Shanghai's total retail sales of consumer goods (x2) has a strong impact on fiscal revenue. Value-added tax is an important part of fiscal revenue. When the total retail sales of social consumer goods increases, it will directly affect the value-added tax in fiscal revenue, which indirectly affects fiscal revenue. Shanghai's GDP (x5) has a positive impact on Shanghai's fiscal revenue, which means that when Shanghai's GDP increases, Shanghai's fiscal revenue will also increase. Therefore, an important means for the government to increase fiscal revenue is to develop GDP in a stable and green way. The output value of the primary industry (x6) has a negative impact on Shanghai's fiscal revenue, that is, the greater the output value of the primary industry and the smaller the fiscal revenue, which means that Shanghai should take advantage of the "financial center" and focus on its development. Second, the tertiary industry is tilted. Taxation (x7) has a strong positive impact on Shanghai's fiscal revenue. Tax revenue constitutes an important factor in Shanghai's fiscal revenue. When the tax revenue is larger, the fiscal revenue will increase. Shanghai urban per capita disposable income (x11) has a strong impact on Shanghai's fiscal revenue. The higher the disposable income of urban residents, the stronger their spending power and the growth of fiscal revenue.

References

- [1] Wei Bangrong, Yang Yusheng. Cointegration Analysis and Error Correction Model of the Relationship between Fiscal Revenue and GDP in China [J]. *Statistics and Information Forum*, 2016.
- [2] Cao Haijuan, Hu Yijian, Feng Suzhen. An Empirical Analysis of Factors Affecting Shanghai's Fiscal Revenue Growth during Urban Transition[J]. *Journal of Shanghai University of Finance and Economics*, 2015.
- [3] Zhang Lunjun. Fitting and Analysis of Tax Forecasting Model [J] *Mathematical Statistics and Management*, 2009, (1): 5-8.
- [4] Cheng Maolin, Zhang Lunjun. The Establishment of Multivariate Nonlinear Economic Forecasting Model [J]. *Statistics & Decision Making*, 2005, (5): 20-21.
- [5] Bai Ping. Multiple Linear Regression Model Affecting China's Fiscal Revenue [J]. *Statistics & Decisions*. 2005, (5): 92-94
- [6] Yan Tiejun, Jin Shuanghua. Application of Grey System Theory in Tax Forecasting [J]. *Quantum Economics and Technology Economics Research*, 2001, (12).
- [7] Lin Guojun, Xuan Huiyu. Application of Genetic Algorithm and BP Artificial Neural Network in Tax Forecasting [J]. *Systems Engineering - Theory Methodology*, 2005, (2).
- [8] GU Junhua, SONG Lijuan, ZHAO Wenhai, SONG Jie. A Tax Revenue Forecasting Model Based on Improved BP Neural Network [J]. *Journal of Hebei University of Technology*, 2003, (1).
- [9] Si Sukui. *Mathematical Modeling Algorithm and Application* [M]. Second Edition. Beijing: National Defense Industry Press, 2016.
- [10] Gorgeous, Zhang Liangjun. Comprehensive fiscal and regression model for local fiscal revenue forecasting. *China Management Informationization* [J]. 2016(5) : 145 - 148.