ISSN: 2688-9323

Research on Influencing Factors of Tax Revenue

-- Based on Multiple Linear Regression Model

Zhuofan Hou

Zhongnan University of Economics and Law, Wuhan, 410000, China

Abstract

As the main source of our country's fiscal revenue, taxation is closely related to our country's economic development. On the one hand, taxation benefits from a good economic situation. On the other hand, it also plays a vital role in economic development. It is of great practical significance to explore the influencing factors of tax revenue for China's economic development and tax revenue increase. This paper collects the relevant data from 1990 to 2020, establishes a multiple linear regression model for empirical analysis, and explores the influencing factors of China's tax revenue. Finally, it is concluded that the added value of the tertiary industry and the whole society's investment in fixed assets have a significant positive impact on China's tax revenue, while the impact of GDP and the total import and export, on tax revenue is not significant.

Keywords

Tax Revenue; Influencing Factors; Multiple Linear Regression; Econometrics.

1. Introduction

Tax revenue is defined as revenues that the state compulsorily collects from taxpayers by wielding its political power. Taxation has three main characteristics: compulsory, non- remunerative and fixed. In addition to the function of organizing revenues, taxation has an important regulatory role in the operation of the economy and society and the allocation of resources. Tax revenue is the main source of China's fiscal revenue and the cornerstone of economic development, and tax growth is the weather vane for the smooth operation and sound development of China's economy. Taxation and the economy interact with each other: on the one hand, taxation benefits from a good economic situation, and economic growth will promote the growth of taxation revenue; on the other hand, taxation acts in turn on the economy and is an important way to regulate the macro economy. Therefore, research on tax revenue has always received close attention from both academia and industry, especially after the COVID-19 took a turn for the better in 2020 and the nation's efforts to restore economic development, while tax revenue is closely linked to economic development and there are a wide range of economic factors affecting tax revenue.

It is of practical significance to explore the influencing factors of tax revenue, which will help promote tax growth, accelerate the restoration of economic order, boost rapid economic development and facilitate the regulation of national income distribution. Furthermore, the analysis of the influencing factors of taxation helps the government regulate the macro economy and provides a theoretical basis for improving the taxation system and further reducing taxes and burdens.

In view of this, this paper put forward reasonable hypotheses based on economic theories and initially selected GDP, total import and export, added value of tertiary industry and total social fixed asset investment as possible influencing factors. Relevant data from 1990 to 2020 were

ISSN: 2688-9323

collected, and a multiple linear regression model was constructed using R software programming to explore the influencing factors of tax revenue in China.

2. Literature Review

Most of the early studies focused on the empirical analysis of the relationship between tax revenue growth and economic growth. In 2000, Fan and Zhang analyzed how economic growth in different segments affected tax revenue, which was the first time that GDP was decomposed into minor categories and the mechanism of different segments were studied. In 2006, Sun concluded that economic growth did have a huge impact on tax revenue by analyzing data from 1994 to 2006. In 2001, Hu and Pan questioned for the first time that the tax growth rate was mainly influenced by the GDP growth rate, arguing that while the influence existed, it was not the main cause. It was also suggested that the tax growth rate was mainly influenced by tax policies, tax expectations and the structure of tax sources. In 2008, Zhang found that social investment in fixed assets, fiscal expenditure and net exports were conducive to the growth of tax revenue. In 2014, Yang's research showed that the direction of the change of tax revenue was the same as that of the tax source quality factor, and the tax structure would not be affected by these two factors. The paper combines regression analysis and factor analysis, and it is the first study to use provincial panel data. In 2015, Huang and Zhu applied a segmented regression approach to figure out how different factors affect tax revenue. The influencing factors they considered were fiscal expenditure, economic growth and retail commodity price index, and the tax reform factor was also taken into account. In 2018, Guo collected data of Yunnan province from 1988 to 2016. He took the year of 1994 as the segmentation point and applied segmented regression to analyses the impact of three types of factors (economy, government and macro level) on tax revenue in Yunnan province. In 2020, Zhu proposed the view that both the value added of tertiary industry and the investment in fixed assets of the whole society are positive factors causing the increase of tax revenue. Also, GDP, total import and export, and fiscal expenditure are not significant influencing factors. In 2020, Liu studied the prediction of tax revenue, and the prediction model was based on multiple linear regression method. First, he concluded that GDP, total retail sales of consumer goods, and fiscal expenditure were significant factors, with GDP and fiscal expenditure acting positively and total retail sales of consumer goods the other way around. Tax systems and macroeconomic systems change all the time, as do the transmission mechanisms of different factors on tax revenues, and the significance and direction of action of the same factor sometimes varies when the samples are different. Any research has limitations and only current findings can be adapted to the current national and economic situation.

3. Theoretical Analysis and Research Hypotheses

3.1. Gross Domestic Product (GDP)

GDP is a core indicator of a country's overall economic situation over a certain period of time, and there is a strong link between it and tax revenue. This is because the level of the economy determines the proportion of tax revenue in the social distribution, which in turn determines the tax structure. When the economic situation is good, the rapid growth of GDP will drive the growth of tax revenue, and there is a certain tax elasticity between GDP and tax revenue, with tax revenue growing faster than GDP at times.

Hypothesis 1: Gross Domestic Product (GDP) has a significant positive impact on tax revenues.

3.2. Total Imports and Exports

The total amount of imports and exports is usually the total amount of goods that enter or leave the country. Generally speaking, the larger the exports, the higher the tax revenue will be.

ISSN: 2688-9323

Import duties, which are one of the almost mandatory hurdles for goods to enter countries, increase with the total amount of imports, leading to an increase in tax revenue.

Hypothesis 2: Total imports have a significant positive impact on tax revenue.

3.3. Value Added of the Tertiary Sector

The reform of the economic system and the change of the industrial structure have gone hand in hand, and the share of the tertiary sector in GDP has increased year by year, undergoing a transition from "two one three" to "two three one" to "three two one". Nowadays, the tertiary sector is a leading industry and its growth has a significant impact on China's tax revenue.

Hypothesis 3: Value added of the tertiary sector has a significant positive impact on tax revenue.

3.4. Total Social Investment in Fixed Asset

As an important component of investment, investment in fixed assets plays an irreplaceable role in promoting economic development. Investment is one of the key factors in promoting socio-economic development. We can presume that investment in fixed assets promotes tax revenue by facilitating economic development.

Hypothesis 4: Total social investment in fixed assets has a significant positive impact on tax revenue.

4. Data Sources and Model Building

Based on economic theory and actual situation, this paper uses Chinese tax revenue (y) as the predicted variable; and GDP, total imports and exports (), value added of the tertiary industry () and total social investment in fixed asset () as the explanatory variables. Data on total tax revenue, total imports and exports, value added of the tertiary sector, GDP and total social investment in fixed asset were collected for a total of 31 years from 1990 to 2020, according to the National Bureau of Statistics and China's statistical yearbooks for each year.

Preliminary multiple linear regression modeling:

+++

where, respectively, the regression parameters corresponding to the explanatory variables are independent random error terms.

The basic conditions for building the model: (1) The model is set up correctly. (2) The random error term has conditional zero mean: E(/x) = 0. (3) The random error term has conditional homoscedasticity: Var(/x) =. (4) There is no serial autocorrelation between the random error terms and they follow a normal distribution. (5) There is no multicollinearity in the explanatory variables.

4.1. The Parameter Estimation - Stepwise Regression based on the AIC Criterion

When the regression analysis was carried out on the four explanatory variables, the output showed that the coefficients of the variables in the regression equation were not significant and did not pass the significance test, indicating that the impact of GDP on China's tax revenue is not significant and it is not reasonable to choose all the variables to construct the regression model. Therefore, the stepwise regression method based on the AIC criterion is adopted below in order to establish the an "optimal" regression model.

The results of the stepwise regression show that the value of the AIC statistic is 484.38 when the regression equation includes all variables, and 482.61, 503.75, 497.93 and 514.95 if the variables, and are removed in turn. Since AIC takes the minimum value when the above variables are delected, if and are removed, the AIC increases. In this regard, the optimal regression equation is based on, as the explanatory variable. In other words, the impact of GDP

ISSN: 2688-9323

on China's tax revenue is not significant, while the impact of total import and export, value added of the tertiary industry and total social fixed asset investment on tax revenue is more significant. Therefore, we construct the regression model with total import and export, value added of the tertiary industry and total social fixed asset investment as the explanatory variables.

The least squares method was used to estimate the parameters and the R software was run to obtain the following estimates of the parameters.

+ (1)

5. Model Testing and Correction

5.1. Economic Tests

According to the estimation results of the regression model, it can be known that, ...and... are all greater than 0, indicating that the total imports and exports, the added value of the tertiary industry and the total social investment in fixed assets all have a positive driving effect on China's tax revenue. In other words, when the scale of China's foreign economic trade expands and the total amount of imports and exports increases, China's tax revenue will increase significantly; when the tertiary industry rises rapidly and accelerates its development, China's tax revenue will increase significantly; when the whole society actively participates in fixed asset investment and the total amount of fixed asset investment increases, China's tax revenue will increase significantly. All the above are in line with economic theory. Therefore, the regression model is reasonable and the model passes the economics test.

5.2. Statistical Tests

5.2.1. Goodness-of-fit Test

The output yields =0.9985 and the adjusted =0.9983, which is close to 1, indicating that the model explains 99.83% of the tax revenue, and that the econometric model is reasonably well constructed and that the multiple linear regression model is a good fit for the sample.

5.2.2. F-statistic Test

Original assumptions: ===0

According to the regression results, the value of the F-statistic is 6009, where the explanatory variable is 3 and the sample size is 27. Under the condition of the given significant level, the value of the critical value (3,27) is obtained by checking the table, and the value of the F-statistic is much greater than 2.96 after comparison. Also, Prob (F -statistic) = 0.0000, and the p-value of the F-test is close to 0, so the original hypothesis is rejected. Therefore, at the 95% level of significance, the linearity of the model holds up quite well.

5.2.3. T-statistic Test

Original assumption: =0 (j=2,3,4)

According to the regression results, the T-statistic values of the variables and were 6.285, 4.316 and 28.711 respectively. Also given the significance level, a check of the table shows that (27) = 2.0555, and upon comparison the T-statistic test values of each explanatory variable were found to be greater than the critical value. And the p-value of the t-test of the coefficients of each variable, are close to 0, which is less than the confidence level coefficient of t = 0.05. Therefore, the original hypothesis is rejected, indicating that the effect of the explanatory variables and on the explained variables is significant. Thus, the effects of total imports and exports, social investment in fixed asset and the value added of tertiary industry on total tax revenue are all significant.

ISSN: 2688-9323

5.3. Econometric Tests

5.3.1. Heteroskedasticity Test

The residuals, normalized residuals, predicted values were calculated and a scatter plot of the residuals was made, as shown in Figure 1. The observation of the image shows that the scatter is randomly distributed and the model has no heteroskedasticity.

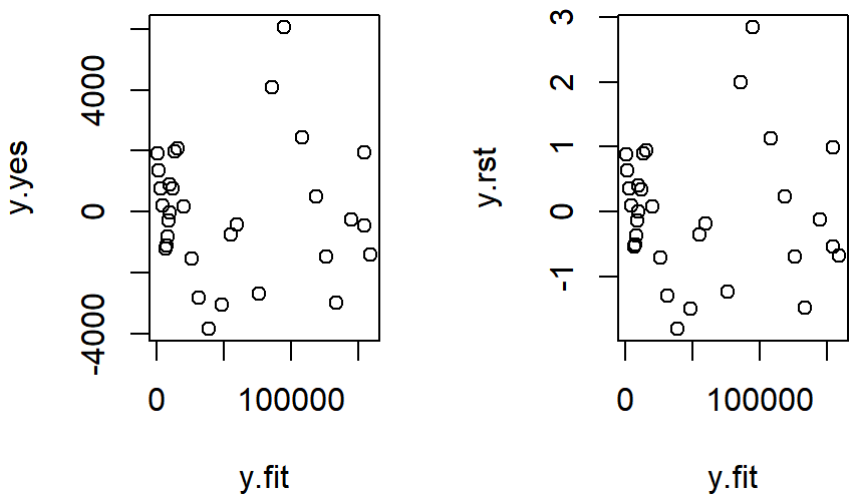

Figure 1. Residual Pots for Model (1)

Combining Figures 2 and 3, it is easy to see from Figure 2 that the scatter points are roughly concentrated on the straight line of the QQ diagram, indicating that the random error term is well normalized and satisfies the assumptions set by the model. From Figure 3, it is easy to see that the scatter points are randomly distributed around the curve, which means that the assumption that Var (/x) = holds and the model does not have heteroskedasticity.

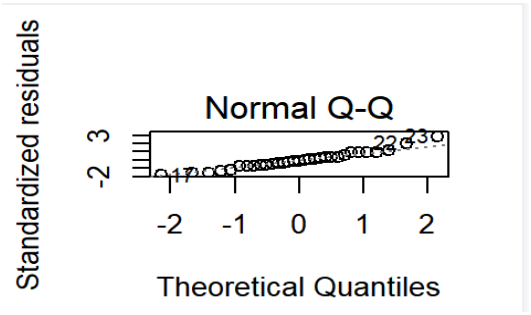

Figure 2. Theoretical quantiles

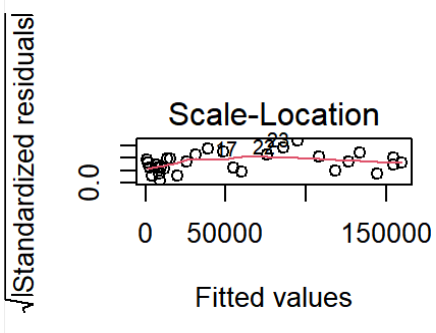

Figure 3. Fitted values

ISSN: 2688-9323

In addition to graphing to test for heteroskedasticity, hypothesis testing can also be performed. For the hypothesis: Var(/x) =, based on the output, it can be seen that p=0.21917, greater than 0.05 and can be considered to satisfy the hypothesis of equal variance.

Non-constant Variance Score Test Variance

formula: ~ fitted.values

Chisquare = 1.509774, Df = 1, p = 0.21917

In summary, the model does not suffer from heteroskedasticity.

5.3.2. Multicollinearity Testing and Correction

According to the output, the variance expansion factor VIF is less than 10, and the variance expansion factor VIF of, is greater than 10, so there may be multicollinearity between, and. The linear correlation coefficient between, and, is 0.955, so it is determined that there is serious co-linearity between, and.

```
>library(DAAG)
>vif(lm.step,digits=5)
   x2        x3    x4
   14.5650 7.6993 12.1340
   >cor(x2,   x4)   [1]
   0.9552341
```

In order to solve the multicollinearity problem, we adopt the full subset regression method to filter the variables, and get the adjusted graph, as shown in Figure 4. From the graph, we can see that when the model contains only the intercept term and, the Adjusted r- squared reaches 0.87; when the model contains the intercept term and, the Adjusted r-squared reaches above 0.9; when the model excludes the variables and contains only the intercept term, the Adjusted r-squared reaches 0.95; when the model excludes the variables and contains only the intercept term, the Adjusted r-squared rises to 1. When the model excludes the variables and includes only the intercept term, the Adjusted r-squared reaches 0.95; when the model excludes the variables and includes only the intercept term, the Adjusted r-squared rises to 1; in addition, the Adjusted r-squared of the model with the variables excluded is also 1, and the Adjusted r-squared of the model with both the intercept term and, is also is 1.

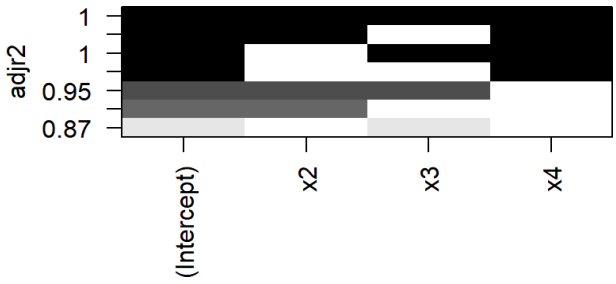

Figure 4. Filter variables using full subset regression

ISSN: 2688-9323

Since the VIF value of, is greater than 10, to solve the multicollinearity problem, we try to eliminate the variable or. When the variables were kept in the model and thevariables were removed, the Adjusted r-squared of the model was 0.95; when the variables were kept and the variables were removed, the Adjusted r-squared increased to 1. Therefore, we considered that the model with the variables kept and the variables removed was superior to the other models. The new model was subjected to regression analysis and then diagnosed for multicollinearity, and the results were as follows, with variance inflation factor VIF values less than 10, so there was no multicollinearity in the model.

lm.reg2<-lm (y~x3+x4, data=tax)

summary(lm.reg2)

library (DAAG)

vif(lm.reg2, digits=5) x3

x4

6.0396 6.0396

At this point the modified model reads

+ (2)

t= (1.483) (5.30) (30.50)

=0.9963 =0.9961 F=3787 (explanatory variable 2, sample size 28)

The coefficient of determination of this model is close to 1, which indicates that the model has a good fit; it passes the F-test, which indicates that the linear relationship of the model is significant and the model setting is reasonable; the coefficients of all variables pass the t-test, which indicates that the influence of the value added of the tertiary industry and the total social fixed asset investment on China's tax revenue is significant; and the influence of all explanatory variables on the explained variables is in line with the economic theory and the actual situation, that is, when the value added of the tertiary industry and the total investment in fixed assets of the whole society increase, China's tax revenue increases accordingly; the residual plot of the model is shown in Figure 5, with a scattered random distribution, which is considered that there is no heteroskedasticity in the model.

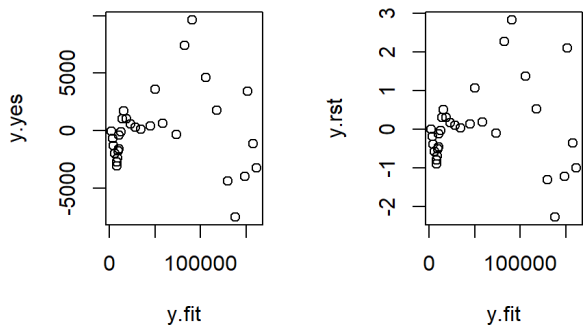

Figure 5. Residual plots for model (2)

6. Conclusion and Suggestions

This paper analyses the changes in China's tax revenue from 1990 to 2020 from the perspective of exploring the influencing factors. The influencing factors initially selected in this paper are

ISSN: 2688-9323

GDP, total imports and exports, value added of the tertiary industry, and social investment in fixed assets. According to the results of the empirical analysis, it can be seen that the total imports and exports, the value added of the tertiary industry and the social investment in fixed assets have a significant positive impact on China's tax revenue, with the value added of the tertiary industry having a greater impact on tax revenue. While GDP and total imports and exports have insignificant impact on China's tax revenue. That is, (1) when the tertiary industry rises rapidly and accelerates its pace of development, our tax revenue increases significantly with it. (2) When the whole society actively participates in fixed asset investment and the total amount of fixed asset investment increases, our tax revenue increases significantly. (3) The value added of the tertiary industry has a greater impact on tax revenue than the fixed asset investment of the whole society, and is a significant influencing factor on tax revenue. (4) The impact of GDP and total imports and exports on China's tax revenue is not significant, but these two factors do not have no impact on tax revenue at all.

The change in tax revenue reflects China's economic operation and development situation. If government wants to maintain stability or increase tax revenue, they should vigorously develop the tertiary industry, promote its accelerated growth and speed up the pace of development; encourage the nationals to actively participate in fixed asset investment and support benign investment in society; keep the domestic economy in a good state of development and stable operation in line with the trend of the times of economic globalization, expand scale of foreign economic trade and increase the total amount of China's imports and exports, instead of unilaterally raising tariffs and ignoring the changes in the quantity of imports and exports.

References

[1] Huang, L. Yan, W.T. Zou, M.N. Yuan, W.Q. Xiao, Y. Xie, B.F. 2021. Analysis of factors influencing house prices in Ganzhou City based on multiple linear regression[J]. Inner Mongolia Science Technology & Economy(20). p77-78+82.

[2] Xiao, X. Wu, X.G. 2021. Geometric Interpretation of Multicollinearity in Linear Regression[J]. Statistics & Decision (21). p 46-51. DOI:10.13546/j.cnki.tjyjc.2021.21.009.

[3] He, X.Y. 2021. An Econometric Analysis on Influencing the Growth of Tax Income[J]. Taxpaying(20). p23-24+54.

[4] Cao, B.X. Li, Z.P. 2021. The Confidence Interval of Fiscal Revenue Based on the Total Least Squares Regression[J]. Journal of Gannan Normal University (42). p40- 43. DOI: 10.13698/j.cnki.cn36-1346/ c.03.007.

[5] Cao, Q.Q. Zhu, J.M. Jin, J. 2020. Quantitative Analysis of Factors Influencing Taxation in China Based on Multiple Linear Regression[J]. Natural Science Journal of Harbin Normal University(04). p9-15.

[6] Zuo, Y.T. Wu, L. 2020. Research on the Influencing Factors of Tax Revenue in Anhui Province-- An Empirical Analysis based on Data from 2001-2018[J]. Journal of Chifeng University (05). p66-71. DOI: 10.13398/j.cnki.issn1673- 260x. 05.020.

[7] Liu, J.H. Ren, D. 2020. Multiple Regression Analysis and Prediction of Factors Affecting Tax Revenue [J]. Journal of Gansu Sciences(01). p129- 136. DOI:10.16468/j.cnki.issn1004-0366.2020.01.024.

[8] Fan, F. & Wang, C.L. 2015. Analysis of Influencing Factors of Tax Revenue and Forecast Model[J]. Journal of Logistical Engineering University(06). p92-96.

[9] Tang, X.Y. Zhang, T. 2008. The Economic Factors on China's Tax Revenue Growth & the Forecast of Trend [J]. Journal of Central University of Finance and Economics (11). p6-11.

[10] Sun, Y.D. 2008. Factors of Influencing Rapid Increase of Tax Revenue in China and Their Quantity Analysis [J]. Economic Theory and Business Management (06). p31-35.

[11] Fan, L.M. & Zhang, B. 2000. The Correlation Analysis of Economic Growth and Tax Revenue[J]. Taxation. Research Journal (02). p4-11. DOI:10.19376/j.cnki.cn11- 1011/f.2000.02.001.