

The Establishment of Tax Revenue Forecast Model and Its Empirical Test and Analysis

Rui Cao

Accounting and Finance, University of Sussex, BN1 9SN, U.K.

rc452@sussex.ac.uk

Abstract. Taxes are generated with the creation of the state, to satisfy the state to fulfill its functions material needs. The scale of taxation is an important indicator to measure the national financial strength and the function scope of the government in social and economic life. Tax forecasting is of great significance to tax planning and budget management. The current economic form requires the establishment of a scientific forecasting system based on tax revenue forecasting as soon as possible, so as to grasp the initiative of organizing tax revenue. Therefore, this paper adopts the methods of econometrics and time series to establish a tax revenue forecast model and conduct empirical test and economic analysis on it, trying to provide a reference for the national and local taxation bureaus to formulate scientific tax collection plans and promote a more harmonious development of the regional economy.

Keywords: Tax revenue; Regression equation; Model establishment.

1. Introduction

1.1 The significance of tax revenue forecasting research

Since the date of taxation, the organization of fiscal revenue for the country to perform its functions has become its duty to organize fiscal revenue, and it is the basic role of taxation. Tax forecasting is of great practical significance for the formulation of tax plans, improving the management of tax plans, giving play to the predictability of taxation, rationally arranging taxation tasks, correctly handling the relationship between taxation and economic development, and promoting the sustainable and healthy development of local taxation. With the construction and development of my country's market economy, the trend and changes of the national economy are no longer completely controlled by the plan. Both the speed of economic development and the adjustment of economic structure can only be driven by changes in market demand, resulting in tax revenue. The traditional model of the base plus projected growth rate can no longer continue on the old way of planning and management. How to establish a scientific forecast system based on tax revenue forecast as soon as possible under the current economic form, and then grasp the initiative of organizing tax revenue, has become the focus of this paper.

1.2 Problems that should be paid attention to in the study of tax revenue forecast

Tax revenue forecasting models mainly include regression forecasting methods and deterministic time series forecasting methods. From the perspective of some research results of tax forecasting, there are some problems, mainly in:

(1) Most of the tax revenue forecasting models in my country are established by traditional econometric methods. Most of them are simple linear models, including univariate models with GDP as the white variable and time as the white variable, and multi-variate linear models with multiple macroeconomic indicators as independent variables. The development of econometrics tells us that when several time series variables are non-stationary variables, the established linear model may have the problem of pseudo-regression. But so far, few literatures have done research on this aspect.

(2) Most of the research on tax revenue forecasting model in our country is a national research or a general theoretical research. Tax revenue is realized under a specific economic environment, and differences in economic development patterns and industrial structures in various regions have a great impact on tax revenue. The tax revenue generated by the same GDP varies greatly in different regions.

Therefore, it is necessary for each province to establish a tax revenue forecast model suitable for itself according to the actual macroeconomic situation of the province. But at present in our country, it is rare to establish a more comprehensive tax revenue forecasting model based on the actual economic situation of a certain province or a certain region.

(3) Most of the tax revenue forecasting models in my country mainly focus on discussing the fitting effect of the model itself, and rarely use historical data for empirical testing. The effect of the prediction model needs to be tested by practice. Statistics show that the model with good fitting effect may not have good prediction accuracy. Therefore, in order to ensure the practical significance of the prediction model, the empirical test of the model is a section of the study of the prediction model. But at present, except for a few literatures, most of the researches on the tax revenue forecasting model have not carried out this step. This research needs to be further improved.

2. Error Correction Model (ECM) for Taxation

2.1 Cointegration theory and its test

2.2.1 Cointegration Theory

The concept of cointegration was proposed by Engle & Granger in the 1980s. It is used to test the long-term stable relationship between time series variables and can effectively overcome "pseudo-regression". The so-called "pseudo-regression" refers to the wrong conclusion that there is no interdependence between variables, but the regression results show that there is such a relationship between them.

2.2.2 Stationarity and Unity and Their Tests

Generally speaking, a time series is an infinite-dimensional random vector, and its probability distribution is a generalization of the finite-dimensional random vector probability distribution. Let $\{Y_t\}(t = 1, 2, \dots)$ be a time series and consider its finite-dimensional probability distribution:

$$F(t_1, \dots, t_n; y_1, \dots, y_n) = P\{Y_{t_1} < y_1, \dots, Y_{t_n} < y_n\}$$

The purpose of stationary time series is to predict the future based on the past, so it is necessary to assume what the past development looked like and what the future development should look like.

$$EY_t = EY_{t+m} = EY_{t-m} \equiv \mu$$

$$\gamma(t_1, t_2) = \text{Cov}(Y_{t_1}, Y_{t_2}) = \text{Cov}(Y_{t_1+m}, Y_{t_2+m})$$

$$= \gamma(t_1 + m, t_2 + m) = \gamma(t_2 - t_1, 0) = \gamma_{t_2-t_1}$$

That is, the probability distribution of the sequence does not change over time.

2.1.3 Empirical Analysis

Taking LNTAX to represent the logarithm of TAX's timing chart, it is clear that both LNTAX and LNGDP have a very clear upward trend.

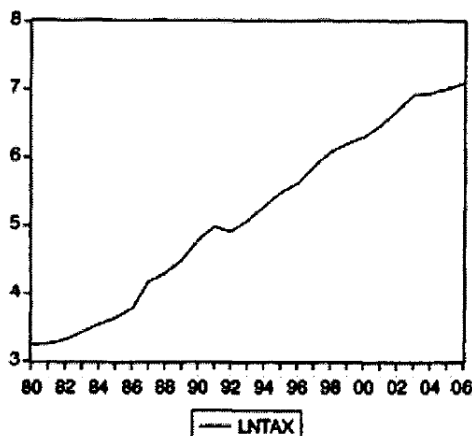


Figure 1. TAX logarithmic timing diagram

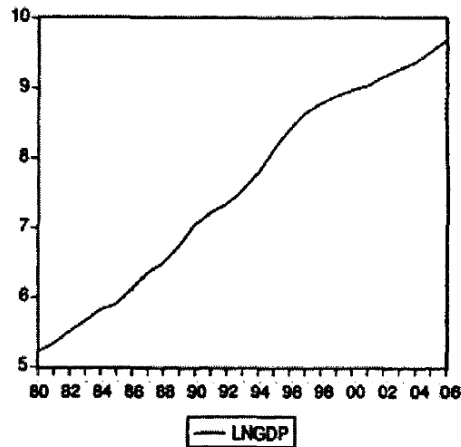


Figure 2. GDP logarithmic time series diagram

Therefore, the ADF test model with constant term and trend term is used for stationarity test. The ADF test T statistic values of LNGDP and LNTAX are 1.638982 and 2.849169, respectively, which are both larger than the critical value with a significance level of 10%. Therefore, the null hypothesis cannot be rejected, that is, the two sequences have a unit root, and they are both non-stationary sequences.

After testing the non-stationarity of the two variables, according to the cointegration theory, the single integer order of them should also be determined before performing the cointegration test. The method is to differentiate the variable series and test the stationarity of the difference series. The three models of ADF are still used for testing. The first-order difference of LNTAX is represented by DTAX, and the first-order difference of LNGDP is represented by DNGDP. The timing diagram is as follows:

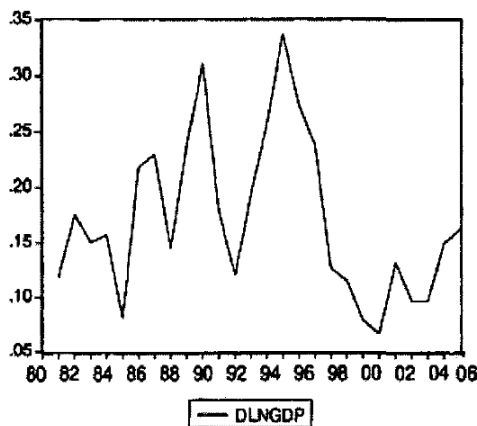


Figure 3. First-order differential timing diagram of LNGDP

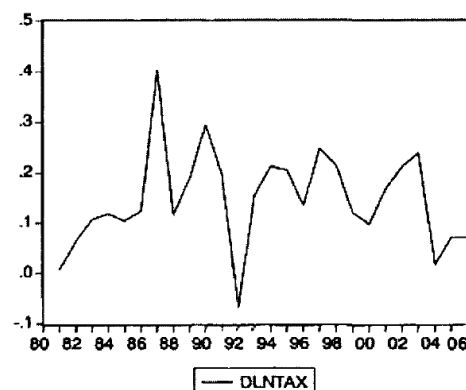


Figure 4. First-order differential timing diagram of LNTAX

2.2 Tax revenue error correction model

The above empirical analysis shows that there is a long-term stable cointegration relationship between LNTAX and LNGDP. According to Grange's theorem, if there is a co-integration relationship between non-stationary variables, an error correction model must be established. In this section, an error correction model between tax GDP is established.

2.2.1 Overview of Error Correction Model

Sa Yigan first proposed the concept of error correction model, namely ECM. When using error correction model to calculate the adjustment value of the explained variable, it is not based on the observed value of the explanatory variable, but the deviation value of the explanatory variable from the equilibrium state. Error correction models not only solve many problems that plague the econometrics community.

2.2.2 Empirical test

The data in this article are compiled from the "Fifty Years of Statistical Data Compilation of New China" and the website of the National Bureau of Statistics. In order to test the prediction effect and accuracy of the prediction model, the software is used to obtain the predicted value of tax revenue (TAX') over the years, the actual value of tax revenue (TAX), and the percentage of prediction error $\eta = \frac{TAX' - TAX}{TAX} \%$:

Table 1. Comparison between the predicted value of the ECM prediction model and the actual value

Year	Tax Income Value (Logarithmic Value)	fitted value	Error percentage (%)
1999	3.5223	3.4072	-1.50%
2001	3.6512	3.6545	-0.00%
1988	3.7748	3.8812	-2.81%
2002	4.0936	4.0997	-0.14%
2003	4.3549	4.4776	-2.81%
2004	4.5418	4.5922	-1.11%
2005	4.6292	4.6545	3.54%
2006	4.7154	4.7706	-1.17%
2007	4.7923	4.8311	-0.81%
2008	4.8506	4.9313	-8.07%
2009	4.9639	4.9698	-0.12%
2010	5.0861	5.0744	0.23%
2011	5.2432	5.2631	-0.38%
2012	5.4087	5.5204	-2.07%
2013	5.6881	5.6192	1.21%
2014	5.8111	5.8626	-0.89%
2015	5.8683	6.0017	-2.28%
2016	5.9501	6.0569	-1.80%
2017	6.2376	6.1774	0.97%
2018	6.4400	6.4390	-0.00%
2019	6.6172	6.6435	0.39%
2020	6.8533	6.8394	-0.025%

2.2.3 Conclusion of the model

The table generally reflects that the overall prediction effect of the model is good, and the prediction accuracy is high. The model says that the current period's tax is not only related to the current period's GDP, but also to the previous period's TAX and GDP. There is also only selected data after 1999 for analysis, considering that the tax system reform was implemented in 2000, and the most significant point was the introduction of consumption tax; while the implementation of the tax sharing system in 2008 had little impact on tax revenue .

3. Multiple regression prediction model

3.1 Introduction to Multiple Linear Regression Model

The relationship between phenomena in the objective world is complex, and the changes of many phenomena involve the quantitative relationship between multiple variables . The theory and method of studying the relationship between a dependent variable and multiple independent variables is multivariate. Linear regression forecasting .

The multiple linear regression model has the form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n$$

3.2 Selection of independent variables of regression equation

(1) Factor analysis method. It is based on the understanding of the dependent variable, according to the basic economic theory and experience, to find out all the factors that affect the dependent variable, and then according to the conditional limitations of the sample data, select the independent variable that has a greater impact on the research object. Using factor analysis to select independent variables should consider the factors that affect the dependent variables as comprehensively as possible, so as to avoid omitting some factors that do have important influences .

(2) Simple correlation analysis. Whether the factors selected by factor analysis can be entered into the regression model as independent variables, but also through quantitative analysis. A simple and feasible method is to carry out correlation analysis, that is, to calculate the simple correlation coefficient between the dependent variable and each influencing factor, and select those with a high degree of correlation with the dependent variable as the independent variable. Generally, factors with a correlation coefficient less than 0.5 are regarded as having a low degree of correlation, and are discarded and not included in the regression equation. For some factors that are uncertain whether they should be discarded, they can also be temporarily retained until the model is tested before making a decision .

(3) Whether the white variables selected by simple correlation analysis can all enter the regression equation also depends on whether there is a strong linear relationship between the independent variables. If there is a strong linear relationship between the independent variables, the variable X; Changes, the independent variables that are highly correlated with them will change accordingly, and the regression coefficient cannot be explained. A linear constraint between independent variables is called multicollinearity, and it results from two or more independent variables (combinations of independent variables) being highly correlated with each other. When the independent variables have multicollinearity, only one of the collinear variables can be retained, and the rest are omitted. Usually, the variables with the smaller correlation coefficient with the dependent variable are omitted.

(4) Stepwise regression method. Selecting variables based on simple correlation analysis often ignores the interaction between independent variables. Therefore, stepwise regression method is often used to select independent variables in practice. The specific method is to introduce the dependent variables one by one. After each independent variable is introduced, the selected variables should be tested one by one. When the originally introduced variable becomes no longer significant due to the introduction of the latter variable, it will be eliminated. . Then, an F-test was performed to ensure that the regression equation included only significant variables each time a new variable was introduced.

This process is repeated until the variables included in the regression equation are all significant factors affecting the dependent variable .

3.3 Empirical research

In addition to the variables given in the previous section, my country has experienced two relatively large tax reforms in 2008 and 2012. We will not discuss the impact of tax policy changes for the time being. As can be seen from the introduction in Section 2, in order to reasonably select explanatory variables and effectively eliminate multicollinearity, stepwise regression can be used to select independent variables that have a major impact on the research object. The process of stepwise regression is completed by SPSS software. In the SPSS software, input the data of the variables obtained by the above analysis, and use the Stepwise method for regression, and get the results as shown in Table 2.

Table 2. Regression coefficient table

Lodel	Chstandardized	Coefficients		Standardized Coefficients		
		B	Std Error	Beat	t	Sig
1	(Constand)	10.853	10.475		1.232	.168
	TIAV	.205	0.004	.944	47.385	.000
2	(Constand)	32.677	8.827		5.703	.000
	TIAV	.288	0.013	1.397	21.843	.000
	HPCA	.070	0.011	-.411	-6.430	.000
3	(Constand)	14.576	6.052		5.400	.000
	TIAV	.359	0.014	1.743	26.212	.000
	HPCA	.212	0.023	-1.428	-9.183	.000
	SIAV	.470	0.073	.506	6.429	.000
4	(Constand)	-32418	38.538		-1.516	.134
	TIAV	.367	0.013	1.780	28.459	.000
	HPCA	.218	0.021	-1.278	-10.305	.000
	SIAV	.478	0.067	.515	7.170	.000
	CPI	.849	0.356	.019	2.388	.026
5	(Constand)	-29.762	34.337		-2.515	.020
	TIAV	.460	0.012	1.843	23.491	.000
	HPCA	.195	0.02	-1.152	16.871	.000
	SIAV	.478	0.064	.423	10.251	.000
	CPI	1.176	0.324	.027	3.631	.002
	TIAB	-.065	0.015	-.110	-2.989	.007

It can be seen from the regression coefficient table 2 that the fifth model is the final model through step-by-step regression. The final tax revenue is mainly affected by the value added of the secondary industry (TIAV), the value added of the tertiary industry (SIAV), and the per capita consumption level of residents (HPCA). , Consumer Price Index (CPI), Total Imports and Exports (TIAE). From

the fifth regression model, the P values of these independent variables are all close to zero, that is, less than 0.05, that is, T The test is significant. Thus, the model is obtained:

$$\text{TAX} = -29.762 + 0.46\text{TIAV} + 0.478\text{SIAV} + 0.195\text{HPCA} + 1.176\text{CPI} - 0.065\text{TIAE}$$

4. Conclusion

From different theoretical perspectives, this paper uses the stepwise regression model and the error correction model to establish a single prediction model. Among them, the stepwise regression model is based on two levels of consideration: First, the choice of the multiple linear regression model. The selection of the best multiple linear regression model is discussed, and the stepwise regression method is used to build the model with SAS software. Furthermore, considering that taxation is jointly affected by a variety of economic variables, and there may be multicollinearity among these variables, the established prediction model is given. The error correction model is based on the co-integration test. It analyzes the long-term stable relationship between the logarithm of tax revenue and the logarithm of GDP, and then establishes the corresponding error correction model.

References

- [1] Multiple Linear Regression Modeling and MATLAB and SPSS Solutions[J]. Chen Yongsheng. Journal of Suihua University. 2007(06).
- [2] Tax prediction model based on Elman neural network [J]. Zhang Jigang, Liang Na. Statistics and Decision. 2007(11).
- [3] Coordination between taxation and economic growth in my country in 2004 [J]. Wang Jiabin, Zhang Lunjun. Finance and Trade Economy. 2005(05)
- [4] Application of spline function in statistical regression model [J]. Lin Shanghua, Wang Guozhao. Zhejiang Statistics. 2005(02).
- [5] 2004 China Tax Revenue Forecast Research [J]. Guo Jue, Qian Xin, Cao Hua. Finance and Economics. 2004(S1).
- [6] Views on a tax forecast model [J]. Tang Yan. Shanxi Finance and Taxation. 2004(08).
- [7] Theoretical Analysis and Empirical Research on China's Tax Increase - Case Analysis of China's Tax Increase from 1992 to 2002 [J]. Qian Sheng, Lu Lingbo. Journal of Central University of Finance and Economics. 2003(09).
- [8] Selection of Linear Regression Model in SAS System [J]. Qin Zhangjing, Hu Guihua, Wan Weiguo. Journal of Jingzhou Normal University. 2003(02).
- [9] Model design and empirical analysis of local fiscal revenue forecast [J]. Cheng Jun. Reference for Economic Research. 2003(88).
- [10] Tax revenue prediction model based on improved BP neural network [J]. Gu Junhua, Song Lijuan, Song Jie, Zhao Wenhui. Journal of Hebei University of Technology. 2003(01).
- [11] Prediction and Analysis of Per Capita GDP Based on ARIMA-PLS Dynamic Combination Model [J]. Wang Qin, Lou Xinyi, Ou Pan. Modern Business. 2018(25).
- [12] Prediction of per capita GDP in Shanxi Province based on the grey-Markov model [J]. Zhang Yaqing. Journal of Taiyuan University (Natural Science Edition). 2018(01).
- [13] Analysis and prediction of Shandong's GDP based on ARIMA model [J]. Yan Yanwen. Practice and Understanding of Mathematics. 2018(04).
- [14] Application of GM (1,1) model based on background value optimization in GDP forecasting of Mudanjiang [J]. Zu Peifu, Zhao Wenyong, Xie Wei, Liu Fang, Zhang Yujiao. Practice and Understanding of Mathematics. 2018(01).
- [15] Application of PSO-BP algorithm based on principal component analysis in GDP and CPI forecast [J]. Wang Yongjie, Bai Yanping. Journal of Chongqing University of Technology (Natural Science). 2017(04)
- [16] Quantitative analysis of sewage treatment problem based on neural network [J]. Yuan Zeyi, Zhu Jiaming, Li Zhenguang, Li Dongling. Journal of Jiamusi University (Natural Science Edition). 2016(02).

- [17] Research on the performance of ACO-BP neural network based on ant colony algorithm [J]. Yin Shuyan, Du Qingdong. Heilongjiang Science and Technology Information. 2015(22).
- [18] Research and application of curve trend extrapolation extension method [J]. Wu Xiubin, Sun Zhaorui, Sun Guibin, Li Guoping. Manufacturing Automation. 2014(08).
- [19] Kashgar GDP forecast based on variable weight GM (1,1) combination model [J]. Wang Fengxiao. Practice and Understanding of Mathematics. 2013(24).
- [20] GDP prediction based on principal component analysis and Bayesian regularized BP neural network [J]. Yu Shenghua, Deng Juan. Journal of Hunan University (Social Science Edition). 2011(06).