

Analysis of Factors Affecting Fiscal Revenue Based on Stepwise Regression Method

Tiantian Gu

School of Economics and Management

Zaozhuang University

Zaozhuang, China

Abstract—The financial revenue is a country's cornerstone of the economy. This article analyzes the related influence factor of the financial revenue through the multiple regression model. The research discovers that the total retail sales of social consumer goods, the total agricultural output and the value volume of technology market have the remarkable influence to the financial revenue.

Keywords—the financial revenue; multiple regression; the influence factor

I. INTRODUCTION

The growth of fiscal revenue is related to the country's economic development and social progress. The state can effectively regulate resource allocation through fiscal revenue. The national public infrastructure construction and basic industrial construction all require financial support. The science, education, culture and health undertakings also need financial support. The fiscal revenue relates to the improvement of national security, economic level and people's living standards. Since the reform and opening up, the government of our country has applied a number of policy measures, mainly fiscal and monetary policies, to regulate and control the macro economy to ease the periodic economic fluctuations and promote economic growth. The fiscal revenue plays an important role in promoting the sustained, rapid, stable and healthy development of the national economy. The fiscal revenue of a country will be affected by many factors. Then, what are the important factors that affect the scale of fiscal revenue, and to what extent do the factors that affect the scale of fiscal revenue affect fiscal revenue?

II. MODEL BUILDING

The regression analysis is a statistical method for studying the correlation between random variables, the intention of which is to study the statistical relationship between an explained variable (also known as a dependent variable) and one or more explanatory variables (also known as independent variables).

The basic model of multiple linear regression is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \mu$$

Basic hypotheses: (1) μ is a random vector

$$E(\mu) = 0$$

$Cov(\mu) = E(\mu\mu^T) = \sigma_\mu^2 I_n$, which essentially includes two hypotheses: one is that there is no sequence correlation, and the other is that it has the homoscedasticity.

$$\mu \sim N(0, \sigma_\mu^2 I_n)$$

$$E(X^T \mu) = 0, \text{ or } X \text{ is deterministic matrix.}$$

$$\text{Sequence } \rho(X) = k(k < n)$$

Determinant $|X^T X|$ is away from zero.

A. Parameter Estimation

The parameter estimation is always achieved by using the ordinary least square (OLS) method, and the parameter estimation value is $\hat{\beta}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

In which the precondition requires that $(X^T X)^{-1}$ exists (satisfy the third of the basic assumptions). The results are unbiased and effective.

The parameter β_j indicates the degree of influence on the dependent variable y for each unit of change of x_j , and its positive and negative indicate the directions of action: β_j is positive, which indicates that when x_j increases, y also increases, and vice versa.

B. Model Test

The model test mainly includes the significance test of the equation (F test) and the significance test of the regression coefficient (t test). The F test is mainly aimed for the overall effect of the model fitting sample, and the significance test of the regression coefficient reflects the rationality of each independent variable.

F test

Hypothesis: $H_0: \beta_j = 0$

Alternative hypothesis: $H_1 : \beta_j \neq 0$

$$F = \frac{SSR/(k-1)}{SSE/(n-k)}$$

Test statistics:

The bigger F is the better. When the calculated statistical value is $f > f_\alpha(k-1, n-k)$, which indicates that the regression effect is good and above the level α , and the explained variance is significantly larger than the unexplained variance.

1) Significance test of the regression coefficient:

Hypothesis: $H_0 : \beta_j = 0$

Alternative hypothesis: $H_1 : \beta_j \neq 0$

$$t = \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}}$$

Test statistics: follows t(n-k) distribution

When $|t| > t_{\frac{\alpha}{2}}(n-k)$, H_1 is true, that is, β_j is significantly different from 0. Each regression coefficient passes the t test, which indicates that each of the independent variables in the model is significant.

2) DW test:

DW test is used to test the autocorrelation of the residual sequence in order to verify whether the basic hypothesis 2 is true or not. ρ is marked to be the autocorrelation coefficient of the residual sequence.

Hypothesis: $H_0 : \rho = 0$

$$DW = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \approx 2(1 - \hat{\rho})$$

DW statistics: , in which $e_t = \rho e_{t-1} + v$

While $|\hat{\rho}| < 1$, DW is more than 0 and less than 4.

$\hat{\rho} = 1$ and $DW=0$, which indicates completely positive autocorrelation. As $\hat{\rho}$ only can be close to 1, e_t and e_{t-1} can only be close to a completely positive autocorrelation.

$\hat{\rho} = 1$ and $DW=2$, which indicates that e_t and e_{t-1} completely have no positive autocorrelation.

$\hat{\rho} = -1$ and $DW=2$, which indicates completely negative autocorrelation. As $\hat{\rho}$ only can be close to -1, e_t and e_{t-1} can only be close to a completely negative autocorrelation.

C. Model Evaluation

Judgment of Coefficient R^2 and Corrected R^2

R^2 reflects the regression effect.

$$R^2 = \frac{\| \hat{y} \|^2}{\| y \|^2} = \frac{\| y \|^2 - \| e \|^2}{\| y \|^2} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n y_i^2}$$

R^2 reflects the overall regression effect, which is better if closer to 1.

In order to overcome the effect of R^2 by the number of independent variables, the adjusted sample determination coefficient \bar{R}^2 is introduced, and is the corrected R^2 .

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} (1 - R^2)$$

The corrected R^2 is also related to the participation number k except of having the original index characteristics. The index is used for judging goodness of fit is more effective than R^2 both based on the precision and number of variables.

This paper uses a stepwise regression method. The basic idea of the stepwise regression method is to gradually select the variable that has the largest contribution to the explained variation (that is, partial interpretation variation) from the variable X_2, \dots, X_k when considering the regression of Y to a known group of variables (X_2, \dots, X_k), and enter the regression equation. The discriminant basis for the contribution to the explained variation is the F statistic F_j containing the partial interpretation variation. The maximum x_j of the value f_j of the statistic F_j first enters the equation; the last independent variable x_j entering the equation should also satisfy: the significance probability P of the value f_j of the statistic F_j is less than or equal to the selected significance probability level α .

III. CASE ANALYSIS

This paper mainly analyzes the factors influencing the fiscal revenue; and taking fiscal revenue as the dependent variable, the choice of independent variables takes into account the availability and possible impact of the data. A multivariate regression analysis model is built respectively with the total output value of construction industry (100 million yuan) X1, the total agricultural output value (100 million yuan) X2, total retail sales of social consumer goods (100 million yuan) X3, international tourism (foreign exchange) income (100 million yuan) X4, population (10,000 people) X5, number of employees (10,000 people) X6, technology market turnover (100 million yuan) X7. The sample is the data of 31 provinces and municipalities in our country in 2008, and the sample data is from China Statistical Yearbook 2008. The specific data can be found in the attached table.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \mu$$

μ is the residual residual and $E(\mu)=0$, independent of the independent variable.

The results are analyzed in "Table I":

TABLE I. CORRECTION COEFFICIENT MATRIX

		Correlation							
		Y	X1	X2	X3	X4	X5	X6	X7
Pearson correlation	Y	1.000	.789	.434	.911	.817	.563	.824	.469
	X1	.789	1.000	.455	.777	.718	.529	.718	.244
	X2	.434	.455	1.000	.717	.699	.937	.778	-.174
	X3	.911	.777	.717	1.000	.926	.789	.952	.243
	X4	.817	.718	.699	.926	1.000	.766	.873	.013
	X5	.563	.529	.937	.789	.766	1.000	.853	-.106
	X6	.824	.718	.778	.952	.873	.853	1.000	.234
	X7	.469	.244	-.174	.243	.013	-.106	.234	1.000

"Table I" is a table of correlation coefficients for each variable. It can be seen from the table that fiscal revenue has significant correlation with various factors; total output value of construction industry, total agricultural output value, total

retail sales of social consumer goods, international tourism (foreign exchange) income, and population, the number of employees and the technology market turnover all have certain influence on the fiscal revenue.

TABLE II. GENERAL PARAMETER TABLE

Model summary ^d

Model	R	R square	Adjusted R square	Standard estimated error	Change of statistics					Durbin-Watson
					Modification of R square	F modification	df1	df2	Sig. F modification	
1	.911a	.829	.823	338.2901812	.829	140.639	1	29	.000	
2	.963b	.928	.923	223.4020320	.099	38.497	1	28	.000	
3	.970c	.940	.933	207.6310891	.012	5.415	1	27	.028	1.844

^a. Predictive variable: (constant), X3

^b. Predictive variable: (constant), X3, X2

^c. Predictive variable: (constant), X3, X2, X7

^d. Dependent variable: Y

"Table II" is the general parameter table of the three models in the stepwise regression process. In the stepwise regression analysis, we choose to enter $F (\alpha =0.50)$, and it is $F (\alpha =0.10)$ if removal, and three stepwise regression analysis modes are obtained. The independent variables orderly entering the regression analysis equation are: total retail sales of social consumer goods, total agricultural output value, and technology market turnover. It can be seen that with the gradual entry of the independent variables, the model's determination coefficient R Square gradually increases, and the R Square of the model three is 0.940, reflecting that the overall regression effect is better, that is, 94% of the variation of the dependent variable can be explained by the change in the independent variable. The increasing R Square also indicates that the model three has better regression effects than the first two models. For multivariate linear regression models, the model fit is generally judged by its adjusted determination coefficient. In this case, the value is 0.933, indicating that the fit is good. The last column is a test related to the DW sequence; the DW value of the model three is 1.844, which is close to 2,

indicating that there is no autocorrelation between the samples in this model.

TABLE III. REGRESSION COEFFICIENT AND SIGNIFICANCE AND COLLINEARITY TEST TABLE

		Coefficient ^a								
Model		Non-standardized coefficient		Standard coefficient	t	Sig.	95.0% confidence interval of B		Collinearity statistics	
		B	Standard error	Trial version			Lower limit	Upper limit	Tolerance	VIF
1	(constant)	54.276	95.249		.570	.573	-140.530	249.083		
	X3	.240	.020	.911	11.859	.000	.199	.282	1.000	1.000
2	(constant)	218.667	68.253		3.204	.003	78.856	358.477		
	X3	.326	.019	1.234	16.963	.000	.287	.365	.485	2.060
	X2	-.253	.041	-.452	-6.205	.000	-.337	-.170	.485	2.060
3	(constant)	169.597	66.848		2.537	.017	32.435	306.758		
	X3	.300	.021	1.135	14.165	.000	.256	.343	.346	2.889
	X2	-.200	.044	-.357	-4.524	.000	-.291	-.109	.357	2.803
	X7	.556	.239	.132	2.327	.028	.066	1.046	.691	1.446

^a. Dependent variable : Y

"Table III" is the regression coefficient and the significance and collinearity test table of the two models in the stepwise regression process. It can be seen from the table that the t-test significance probability of the constant items and the variables including total retail sales of social consumer goods, the total agricultural output value, and the technology market turnover are both less than 0.05, which means that there is a significant difference from 0, so it should appear in the equation. In the process of stepwise regression, X3 first enters the equation, indicating that the

total retail sales of social consumer goods have the most significant influence on fiscal revenue. The parameter β_j is 0.3, indicating that if every 100 million yuan in total retail sales of social consumer goods is increased, the fiscal revenue is increased by 0.3 billion yuan. The last two columns are collinearity tests between independent variables, with VIFs less than 10 and no collinearity problems. The model equation can be written as:

$$Y=169.597+0.3X3+0.2X2+0.556X7$$

TABLE IV. RESIDUAL STATISTICS ^A

	Minimum	Maximum	Mean	Standard error	N
Predicted value	172.770309	3447.421875	924.186761	779.9468267	31
Residual	-392.5804138	671.5601196	.0000000	196.9761463	31
Standard predicted value	-.963	3.235	.000	1.000	31
Standard residual	-1.891	3.234	.000	.949	31

^a. Dependent variable: Y

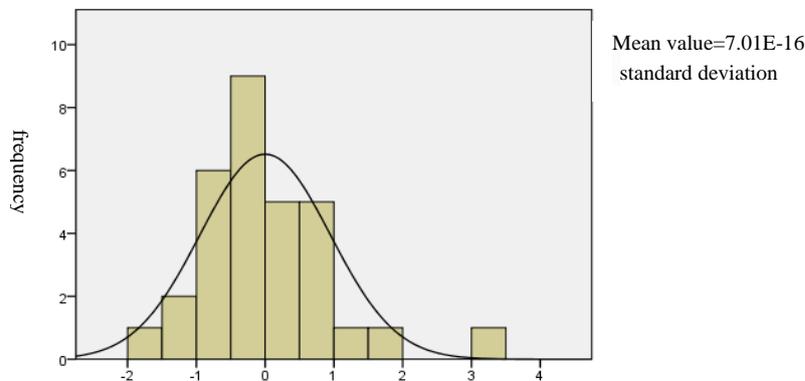


Fig. 1. Regression standardized residual of dependent variable: Y.

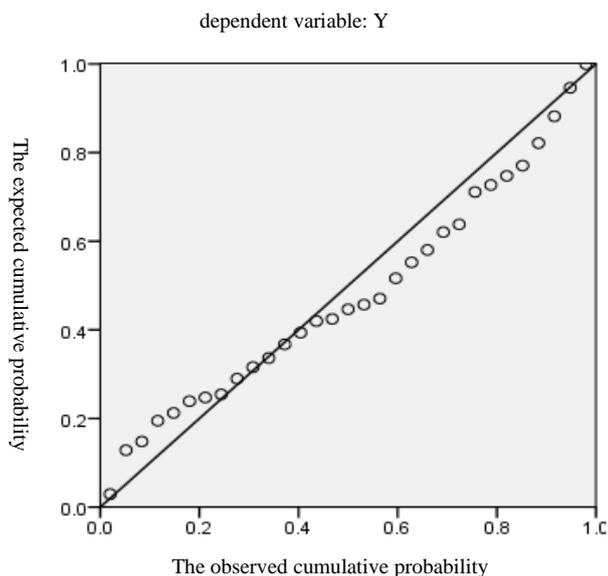


Fig. 2. Cumulative probability graph.

The "Fig. 1" and "Fig. 2" are the normalized residual histogram and the normalized residual normal probability graph of the model. As can be seen from the above charts, the histogram is normally distributed, and the cumulative probability points of the probability graph are distributed along the diagonal, which indicates that the data distribution is good and the degree of fitting is high. Therefore, we can write the equation of this model:

$$Y=169.597+0.3X3+0.2X2+0.556X7$$

It can be seen from the equation that the total retail sales of social consumer goods, the total agricultural output value, the technology market turnover and the fiscal revenue are positively related. For every 100 million yuan increase in the total retail sales of social goods, the fiscal revenue will increase by 30 million yuan. The total agricultural output value and the technology market turnover will be in the same way.

IV. CONCLUSION

It can be seen from the above analysis that factors with the obvious influence on fiscal revenue are the total retail sales of social consumer goods, the total agricultural output value and the technology market turnover. The first to enter the regression is the total retail sales of consumer goods, which can be seen that fiscal revenue is still very sensitive to the total retail sales of social goods. Therefore, our country should adopt various positive policy measures, strive to expand domestic demand, continue to stimulate consumption, and make the domestic market active, and drive the steady growth of the national economy. China is a large agricultural country, and peasants account for the vast majority of the population. Therefore, the growth of total agricultural output has also significantly promoted the growth of fiscal revenue. It can be seen from the model that for every 100 million yuan increase in the technology market turnover, the fiscal revenue will increase by 0.556 billion yuan, which indicates

that the influence of the technology market turnover on the fiscal revenue is still very great, and that increasing technical investment will significantly increase fiscal revenue. This is also a good illustration of "technology is the primary productivity." Therefore, our country should vigorously increase its technical investment, enhance its scientific and technological research and development capabilities, and at the same time perfect intellectual property protection and promote the sustainable and healthy development of the technology market.

REFERENCES

- [1] Ma Qingguo: Management Statistics [M]. Beijing: Science Press, 2003.
- [2] Deng Jie. Empirical Analysis of Factors Affecting China's Fiscal Revenue[J]. Economics and Law, 2009(9).
- [3] Zhang Zhenqiang. Research on Fiscal Revenue Model Based on Stepwise Regression Analysis[J]. Economic Research Guide, 2009(5).
- [4] Zhou Dan. A Stepwise Regression Analysis of the Factors Affecting the Development of Real Estate Industry in Regions of China [J]. Market Modernization, 2009 (1).
- [5] Zhang Xiaogang. Analysis of the Factors Affecting the Total Import and Export Trade of Gansu Province Based on Stepwise Regression[J]. Modern Agriculture.
- [6] Zhang Long, Jia Mingde. An Empirical Analysis of the Impact of Fiscal Expenditure and Fiscal Revenue on Economic Growth[J]. Forecast, 2009(1).
- [7] Yang You, Li Xiaohong. Stepwise Regression Test Analysis of Multicollinearity[J].Journal of Chongqing Threccorges University,2006(3).