

The Application of Time Series Analysis in Financial Revenue of China

Jianna Zhao

Department of Economics and Management, North China
Electric Power University, China
No.689 Huadian Road, Beishi district, Baoding, 071003,
China
zhjnzhf@163.com

Di Xin

Department of Economics and Management, North China
Electric Power University, China
No.689 Huadian Road, Beishi district, Baoding, 071003,
China
xindilove@163.com

Abstract—The impact tax revenue on fiscal revenue is getting bigger. It is of great significance to predict fiscal revenue and analyze the impact of tax on China's economic development according to historical data. The time series analysis and prediction method is to reveal the rule that the phenomenon changes with time based on the historical statistical data of the sequence, and then predict the future trend according to this rule. The purpose of this paper is to predict base on the historical data of fiscal revenue and to analyze the impact of tax revenue on fiscal revenue. It analyzes the modeling process of the time series model, and selects the financial revenue and tax revenue data of the National Bureau of statistics for the last 20 years. First of all, we only use the historical data of fiscal revenue to predict it, then combine the tax revenue data to make integration test, and establish the error correction model with the help of Eviews6.0 software. In this paper, the application of time series and co-integration test are the innovation point, finally, we draw a conclusion that tax revenue plays an important role in the government revenue of long and short time.

Keywords—time series; fiscal revenue; tax revenue; forecast

I. INTRODUCTION

In the process of classical econometrics modeling, time series data are directly used for regression analysis. In fact, the process implies the stationary assumption of time series and on the premise of these assumptions, t , F , χ^2 and other tests are reliable. However, most of the time series are nonstationary in the actual economic analysis. If the nonstationary time series is analyzed as the stationary, the "pseudo regression" problem will be brought^[1] The time series analysis solves this problem and there are two characteristics relative to the econometric model in the model: the one is that it searches the changing rules of the time series themselves rather than using other variables that have causal relationships to model and predict based on the economic theory.

II. MODELING STEPS

A. Analysis of scale and structure of environmental protection investment

If a time series is preprocessed and transformed into a stationary time series, the time series can be modeled. The

establishment of time series model usually includes four steps: one is the identification of models, two is the estimation of model parameters, three is the inspection and diagnosis of models, and the four is prediction.^[2]

One, Model identification: according to the autocorrelation coefficient (ACF) and the partial autocorrelation coefficient (PACF) of the sample, we can find the appropriate values of P, D and Q.

Two, Estimation of model parameters: we identify the appropriate values of P, D, and Q, and the next step is to estimate the parameters of the autoregressive and moving average items in the model.

Three, Diagnosis and test of the model: after selecting the ARIM model and estimating its parameters, it is necessary to see whether the selected models are good enough for data fitting, because another ARIM model may do it equally well. If the selected model is fitted well enough, the residual error estimated from the model should be a white noise sequence. If it is, we can accept the model; if not, we have to do it again.

Four, Forecast: a major reason why ARIM modeling methods is universal is that it can predict the future trend of the sequence, especially in the short term prediction.

III. FORECAST OF FISCAL REVENUE

With the rapid development of China's economy and the increase of fiscal revenue, next, we take the 1996-2015 years' Government Revenue (GR) and Tax Revenue (TR) data (see table I) as an example to predict the government revenue from the two aspects. First of all, we use the method of time series analysis and establish a mathematical model, to predict it with a reasonable prediction method. Then, combined with government revenue to make co-integration test, we set up error correction model and forecast again.^[3]

TABLE I. GR AND TR (UNIT: BILLION YUAN)

year	GR	TR	year	GR	TR
1996	7407.99	6909.82	2006	38760.2	34804.35
1997	8651.14	8234.04	2007	51321.78	45621.97
1998	9875.95	9262.8	2008	61330.35	54223.79
1999	11444.08	10682.58	2009	68518.3	59521.59
2000	13395.23	12581.51	2010	83101.51	73210.79
2001	16386.04	15301.38	2011	103874.43	89738.89
2002	18903.64	17636.45	2012	117253.52	100614.28
2003	21715.25	20017.31	2013	129209.64	110530.7
2004	26396.47	24165.68	2014	140370.03	119175.31
2005	31649.29	28778.54	2015	152269.23	124922.2

A. Test and treatment of stability

Drawing the path map of the time series of government revenue, we initially judge the stationarity of data. From Fig.1, we can see that China's fiscal revenue shows an obvious upward trend, so the original series is non-stationary.

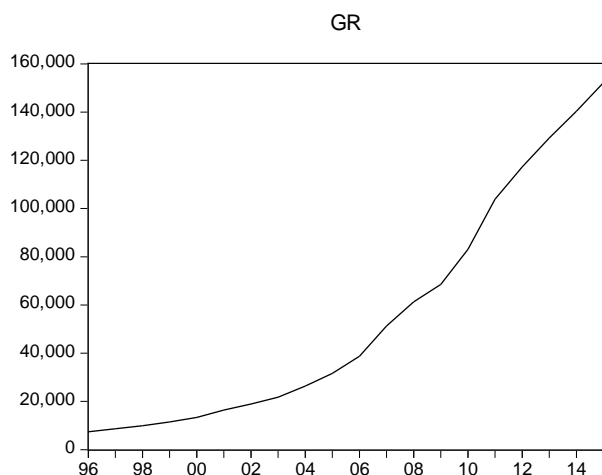


Fig.1 China's government revenue path map

Next, ADF unit root test is used. From Fig.2, we can see that under the model with intercept and trend item, the ADF test value of GR is greater than the level of 95% confidence level, so the original fiscal revenue sequence is nonstationary.

Augmented Dickey-Fuller Unit Root Test on GR

Null Hypothesis: GR has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.933514	0.9302
Test critical values:	1% level	-4.532598
	5% level	-3.673616
	10% level	-3.277364

Fig.2 ADF test results of sequence GR level values (including intercept and trend terms)

We need to further examine the models that contain intercept or intercept and trend terms neither. The results are shown to be non-stationary in Fig.3 and Fig.4

Augmented Dickey-Fuller Unit Root Test on GR

Null Hypothesis: GR has a unit root Exogenous: Constant Lag Length: 1 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.820499	0.9914
Test critical values:	1% level	-3.857386
	5% level	-3.040391
	10% level	-2.660551

Fig.3 The ADF test result of the sequence GR level value (only containing the intercept term)

Augmented Dickey-Fuller Unit Root Test on GR

Null Hypothesis: GR has a unit root Exogenous: None Lag Length: 1 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.982155	0.9063
Test critical values:	1% level	-2.699769
	5% level	-1.961409
	10% level	-1.606610

Fig. 4 ADF test results of sequence GR level values (excluding intercept and trend terms)

At this point, the test results show that the GR level of China's fiscal revenue series is nonstationary from 1996 to 2015. [4] Next, the original data need to be stabilized. In the process of stabilization, we make the first order difference of the original sequence, as shown in Fig. 5.

Augmented Dickey-Fuller Unit Root Test on D(GR)

Null Hypothesis: D(GR) has a unit root Exogenous: Constant, Linear Trend Lag Length: 1 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.220309	0.1134
Test critical values:	1% level	-4.616209
	5% level	-3.710482
	10% level	-3.297799

Fig. 5 Sequence GR first order difference ADF test results (including intercept and trend terms)

From the above diagram, we can see that under the model containing intercept and trend items, the ADF test value of GR is larger than the critical value of 90%, 95% and 99% confidence level, and the p value is larger, so the sequence is still non-stationary at this time, and then two order difference is achieved. The result is shown in Fig. 6:

Augmented Dickey-Fuller Unit Root Test on D(GR,2)		
Null Hypothesis: D(GR,2) has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 1 (Automatic based on SIC, MAXLAG=4)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.081062	0.0049
Test critical values: 1% level	-4.667883	
5% level	-3.733200	
10% level	-3.310349	

Fig.6 Sequence GR two order difference ADF test results (including intercept and trend terms)

From the above diagram, we can see that under the model containing intercept and trend item, the ADF test value of GR is less than the critical value of 99% confidence level, so we reject the original hypothesis and the sequence is stationary. Based on the above unit root test results, the GR sequence is a non-stationary time series. After two order difference, it can become a stationary sequence, namely GR-I (2) and d=2 in the corresponding ARIMA model. [5]

B. Establishing time series model of government revenue

The recognition and order of the ARMA (p, q) model can be obtained through the observation of the autocorrelation and partial autocorrelation function of the sample. [6] The post difference autocorrelation and partial autocorrelation coefficient of the two order difference of the GR sequence as shown in Fig. 7

Correlogram of D(GR,2)						
Date: 06/20/17 Time: 21:42						
Sample: 1996 2015						
Included observations: 18						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1 -0.051	-0.051	0.0545	0.815	
		2 -0.585	-0.589	7.7509	0.021	
		3 0.113	0.055	8.0556	0.045	
		4 0.268	-0.102	9.8961	0.042	
		5 -0.141	-0.052	10.449	0.063	
		6 -0.010	0.140	10.452	0.107	
		7 -0.018	-0.204	10.463	0.164	
		8 -0.116	-0.053	10.949	0.205	
		9 -0.022	-0.209	10.969	0.278	
		10 0.057	-0.139	11.113	0.349	
		11 -0.008	-0.067	11.116	0.434	
		12 -0.017	-0.094	11.133	0.518	

Fig. 7 Sequence GR two order post difference autocorrelation graph

It can be seen from the autocorrelation and partial autocorrelation analysis chart, coefficient of attenuation in two order tends to 0, so p, q value can be 2, but taking into account the principle of simplification model recognition and seeking the best, following on p, q values were compared, as shown in table II, and then determine the most optimal model.

TABLE II. ARIMA MODEL COMPARISON

model	adjusted R ²	AIC	SC	Residual sequence correlation
ARIMA (1,2,1)	0.081	19.24	19.53	relevant
ARIMA (1,2,2)	0.302	19.00	19.20	inccorrelate
ARIMA (2,2,1)	0.194	19.23	19.42	inccorrelate
ARIMA (2,2,2)	0.275	19.16	19.40	significant correlation

The adjusted R² represents the overall goodness of fit of the model, which is between 0-1, and the bigger of it, the better fitting effect it is. Both AIC and SC represent the information criteria, and the smaller the better for the model. Residual sequence correlation is one of the keys to model selection. If there is serious residual sequence correlation, that is, residual sequence correlation test significantly rejecting the null hypothesis, even if other indicators are superior, we should reject this model. As you can see from table 3-2, the model ARIMA (1,2,2) is optimal. So, the optimal model for estimating GR is ARIMA (1,2,2).

The following parameters are estimated for the model ARIMA (1,2,2), and the results are shown in Fig. 8:

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: DGR									
Method: Least Squares									
Date: 06/21/17 Time: 16:31									
Sample (adjusted): 1999 2015									
Included observations: 17 after adjustments									
Convergence achieved after 78 iterations									
MA Backcast: 1997 1998									
Variable		Coefficient	Std. Error	t-Statistic	Prob.				
C		924.8670	241.8784	3.823685	0.0021				
AR(1)		-0.258182	0.277259	-0.931194	0.3687				
MA(1)		-0.009166	0.090092	-0.101737	0.9205				
MA(2)		-0.861954	0.078989	-10.91233	0.0000				
R-squared		0.432726	Mean dependent var		627.9053				
Adjusted R-squared		0.301817	S.D. dependent var		3511.468				
S.E. of regression		2934.090	Akaike info criterion		19.00851				
Sum squared resid		1.12E+08	Schwarz criterion		19.20456				
Log likelihood		-157.5723	Hannan-Quinn criter.		19.02799				
F-statistic		3.305541	Durbin-Watson stat		2.010350				
Prob(F-statistic)		0.054279							
Inverted AR Roots		-.26							
Inverted MA Roots		.93	-.92						

Fig. 8 Model (1,2,2) parameter estimation

The results of the model estimate are as follows:

$$\Delta^2 GR_t = 924.867 - 0.258\Delta^2 GR_{t-1} - 0.009\mu_{t-2} - 0.862\mu_{t-2} + \mu_t$$

According to the model estimate, we can know that if the government revenue in 2015 is 152269.23billion and it is 140370.03 billion in 2014, we will use the static prediction method of the above models to predict the fiscal revenue in 2016,

$$\Delta^2 \hat{GR} = 924.876 - 0.258 \times 738.81 = 734.254$$

The forecast value of China's government revenue in 2016 is as follows:

$$GR = 734.254 + (152269.23 - 140370.03) + 152269.2 \\ = 164902.154 \text{ billion}$$

C. The influence of TR on GR

The above is to establish a model to predict future income by analyzing the historical data of government revenue. Following the tax revenue, we analyze the two relations of fiscal revenue and tax revenue through the co-integration test and error correction.

On the one hand, for the co-integration test, variables often appear in logarithmic form in practical analysis, because the difference of logarithm of variables is approximately equal to the rate of change of this variable, and the rate of change of economic variables is often a stable sequence, so it is suitable for inclusion in classical regression equation. First, we take the logarithm of the government revenue (GR) and the tax revenue (TR) and make a graph:

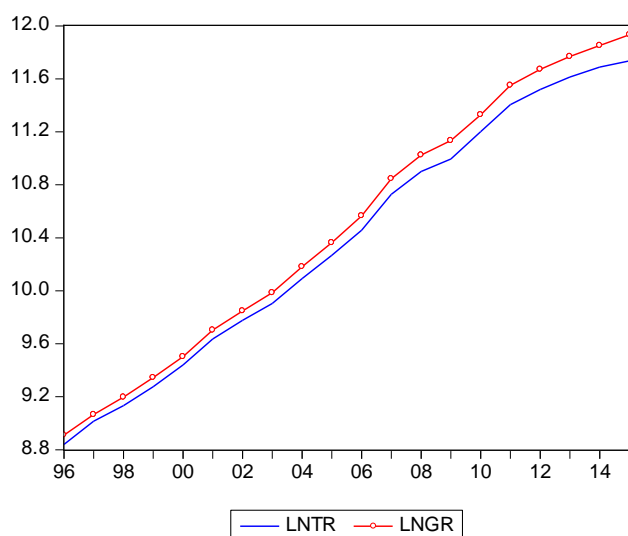


Fig. 9 Graphs of sequence lnTR and lnGR

As can be seen from Fig.9, lnTR and lnGR have roughly the same growth and change trends, indicating that there may be co-integration relationships between the two.

Thirdly, it is found that the stationarity of lnTR and lnGR is nonstationary. After first order difference, the stationarity is still poor. After two order differences, it can become a stationary sequence, so the two sequences are all same order and satisfy the premise of co-integration.

Then we use the E-G two step method to make co-integration test. The first step, the logarithmic sequence lnTR of the tax revenue is interpreted as the variable, and the logarithm of the government revenue as the explanatory variable, making least square regression between two. The second step is to test the stability of the residual sequence. According to the calculating resulting, Mackinnon's p value is 0.0005. At 1% significant level, we can reject the original

hypothesis of "unit root of residual sequence" and think that the residual sequence is a stationary random process. So then, it shows that there is a long-term stable equilibrium relationship between lnTR and lnGR, which can be expressed by co-integration regression model

$$\ln TR = 0.3101 + 0.9602 \ln GR$$

$$t = (10.9274) \quad (356.4175) \quad R^2 = 0.9998$$

That is, the long-term resilience of government revenue to tax revenue is 0.9602.

On the other hand, setting up an error correction model. It is proved that there is a co-integration relationship between lnTR and lnGR, so an error correction model (ECM) can be established. [7] The estimation equation of error correction model is obtained by Eviews6.0 software:

$$\Delta^2 \ln TR_t = 0.0013 + 0.9935 \Delta^2 \ln GR_t - 1.345 e_{t-1} + \varepsilon_t$$

Because the co-integration equation is $e_{t-1} = \ln \hat{TR}_{t-1} - 0.3101 - 0.9602 \ln GR_{t-1}$, so the upper form can also be written as:

$$\Delta^2 \ln TR_t = 0.0013 + 0.9935 \Delta^2 \ln GR_t - \\ 1.345 (\ln \hat{TR}_{t-1} - 0.3101 - 0.9602 \ln GR_{t-1}) + \varepsilon_t$$

According to the error correction model, the long term elasticity of lnTR for lnGR is 0.9602 and the short-term elasticity is 0.9935.

IV. CONCLUSION

Government revenue is an important index to measure the financial capacity of a country's government. The scope and quantity of public goods and services provided by the government in social and economic activities, to a large extent, is determined by the abundant state of financial revenue. Government revenue is the premise of fiscal expenditure, and it is also a financial guarantee for the realization of state functions. It can optimize resource allocation and fully mobilize all aspects of enthusiasm through correctly handling various material interests in the process of organization. Tax revenue is an important source of financial revenue. This paper uses the time series model to analyze the historical data of fiscal revenue, and has carried on the short-term forecast. After that, we analyzed the impact of tax revenue on fiscal revenue through co-integration test and error correction model. It was found that the logarithmic series had little change in long-term and short-term impact, indicating that tax revenue has always played an important role in fiscal revenue.

REFERENCES

- [1] Lishouli. Application of time series model to GDP prediction of city level city [D], 2013.
- [2] Wang Yan. Analysis of the time series of [M]. Chinese people's humanities press, 2005:1-239.

- [3] Chenyao. China's GDP prediction model based on time series analysis [D], 2015.
- [4] Yuping. The application of time series analysis to the prediction of air quality index _AQI_ [D], 2015.
- [5] Wangdefa. Econometrics[M]. Shanghai University of Finance and Economics press, 2016.
- [6] Shimeijuan. The application of ARIMA model in the prediction of fixed assets investment in Shanghai [J], 2005.
- [7] Liyanyi, Liurong, Dingweidai and so on. Statistical analysis and application of Eviews[M]. Electronic Industry Press. 2012.