Modeling and Forecasting Gross Domestic Product in Different Regions of China

Lingyun Duan[1,2,3,4,5], Ziyuan Liu[1,2,3,4], Wen Yu[1,2,3,4,5,*] , Wei Chen[1,2,3,4], Dongyan Jin[1,2,3,4], Jiajia Liu[1,2,3,4], Han Zhou[1,2,3,4], Suhua Sun[1,2,3,4,5] and Ruixi Dai[1,2,3,4,5]

[1]*Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, P.R.China*
[2]*Key Laboratory of Agricultural Information Service Technology, Ministry of Agriculture and Rural Affairs, Beijing 100081, P.R.China*
[3]*Key Laboratory of Intelligent Agricultural Early Warning Technology and System, Chinese Academy of Agricultural Sciences, Beijing 100081, P.R.China*
[4]*Beijing Engineering Research Center for Agricultural Monitoring and Early Warning, Beijing 100081, P.R.China*
[5]*Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, P.R.China*
[*]*Corresponding author. Email: yuwen@caas.cn*

ABSTRACT

Gross Domestic Product (GDP) is an important indicator to measure a country's economic operation in a certain period. By understanding the speed and quality of economic development, it plays a guiding role in policy formulation and infrastructure planning. This article selects sample data of fixed asset investment in various regions from 1990 to 2017, per capita years of education in each region from 1990 to 2017, and GDP of each region from 1990 to 2019 to predict the GDP of each region from 2020 to 2050. The modeling ideas are as follows: firstly, through the cointegration test, it is judged that the regional GDP model with fixed asset investment and per capita education years as explanatory variables has a long-term equilibrium relationship. Secondly, the ECM is established, and the ARIMA model is used to fit the predictive values of the explanatory variables. Finally, the predictive value of the explanatory variables is used to calculate the GDP of each region. Result: the GDP of all regions is on the rise, and GDP maintains stable and rapid development; the average years of education per capita has become an influencing factor for economic growth, and investment in fixed assets is still an important means of stimulating the economy. Conclusion: the government should properly adjust the investment structure of fixed assets, cultivate scientific and technological innovation talents, develop knowledge-based economy and promote better economic development.

Keywords: GDP, ECM, Cointegration test, ARIMA.

1. INTRODUCTION

GDP is an important feature for government policymakers to accurately grasp the health of national economic operation, GDP is an important indicator for correctly formulating economic policies, and an important standard to judge whether the current economic policies are effective. Similarly, regional GDP shows policy makers the operation of the whole region. Only when each region maintains healthy and stable development, the whole country can develop more stably towards economic goals. Therefore, accurately and timely grasp the development of regional GDP, so as to better promote national development.

Many scholars have used different methods to study the forecast of the total GDP in the future. The models and methods of GDP prediction are mainly time series analysis method, artificial neural network method and grey theory method. Luo Sen et al.[1] established a VAR model containing multiple variables and successfully predicted China's total GDP and GDP growth rate, and the prediction effect was good. Guo Qingchun et al.[2] used BP artificial neural network to predict GDP, and the predicted value fitted well with the actual value, which is practical. Shi Ping et al. [3] used the Grey Prediction GM (1, n) model to predict the GDP development of Baotou City, confirming that the Grey Prediction GM (1, n) is effective and feasible in economic forecasting. In

addition to the above methods, domestic and foreign scholars have used other models to obtain good prediction results. Jiang Tong et al.[4] used the Cobb-Douglas economic forecasting model to predict the economic development of all provinces in the context of multiple shared social economic paths (SSPs), and predicted that under five paths, Jiangsu, Guangdong and Shandong rank in the top three in terms of total GDP. The GDP growth rate of each province has different responses under different paths. Each province should rationally arrange its economic development pattern based on its own resource situation. Li Ping et al. [5] constructed a computable general equilibrium (CN3ET-DCGE) model taking into account a variety of main influencing factors, and concluded that China's GDP growth rate will slow down after 2025, and the economic growth pattern will change from investment driven to consumption driven. Geng Peng et al.[6] used the MIDAS model and real-time GDP data to predict, which can effectively eliminate the lag information and noise information of the model, and obtain more accurate estimation results.

In the past, the literature mostly forecasted the national economic aggregate or the economic aggregate of a certain province, but less predicted the GDP growth of all provinces. Therefore, this article selects GDP of each region, investment in social fixed assets and per capita education degree of each region as the influencing factors to predict the total economic volume of each province and explore the development trend of GDP, and make suggestions.

This paper first summarizes the various methods and models used by previous scholars to predict GDP, and then introduces the models used in this paper to predict GDP, such as cointegration test, error correction model and autoregressive moving average model, etc., then obtains the prediction results, analyzes the prediction results, and finally puts forward some suggestions.

2. MODELING

2.1. Data Selection and Processing

Select the GDP of each region from 1990 to 2019, per capita education years of each region from 1990 to 2017, and the fixed asset investment data of each region from 1990 to 2017 are from the Data Center of the National Bureau of Statistics. Regional GDP, regional investment in fixed assets, and per capita years of education in the region are respectively expressed as gdp, inv, and edu. In order to eliminate the heteroscedasticity between variables, the regional GDP and regional investment in fixed assets are processed in logarithm and named lngdp and lninv respectively. This article first uses the unit root test to

test the stability of the variables; then conducts the cointegration test to determine the long-term relationship between the variables and establishes the error correction model, and finally uses the autoregressive moving average model to obtain the predictive data of the independent variables and combines with the error correction model. The model predicts the forecast value of the regional GDP from 2020 to 2050.

2.2. Unit Root Test

Before testing whether there is a long-term equilibrium relationship between time series variables, ADF test needs to be used to test the stationarity of time series variables. Cointegration test can be carried out only if the variables are single integration of the same order. The ADF test values of lnGDP, lninv and edu in each region are not listed. The results show that the ADF test values of the first-order difference time series of lngdp, lninv, and edu are all less than 5% critical value, -2.992, so the null hypothesis that there is a unit root is rejected at the 5% significance level, that is, the first-order difference time series of lnGDP, lninv and edu are stationary.

2.3. Cointegration Test

According to the ADF test results, lngdp, lninv, and edu are all first-order single integration time series, which meet the prerequisites of the cointegration test, and then perform the cointegration test to determine whether there is a long-term equilibrium relationship between the variables. Using the EG two-step test method:

The first step is to perform cointegration regression on lngdp-lninv-edu, and use the least square method to estimate the regression coefficients and the regression equation is shown in Table 1. The cointegration regression equation of the three variables is:

$$Y_t = a_0 + a_1 X_t + a_2 Z_t \qquad (1)$$

In the above formula, Y is lngdp, X is lninv, and Z is edu. In Table 1, Adj-R^2 is the adjusted goodness of fit. The closer Adj-R^2 is to 1, the better the fit; α_1 and α_2 are the long-term elasticity of Y with respect to X and Z.

The second step is to test the stationarity of the residual sequence (e_t) of each regression equation, and use ADF to test the stationarity of e_t. If the residual sequence is stationary, there is a long-term equilibrium relationship between the variables. The critical value of ADF test for cointegration test of residual series (e_t) is calculated by the method of cointegration test in *Econometrics* edited by Zinai Li.

Table 1. Coefficients of Cointegration test and ECM model in different regions

| province | Cointegration test | | | | | ECM | | | | | |
| | | | | | | inv | | edu | | ecm_{t-1} | |
	α_0	α_1_lninv	α_2_edu	Adj R^2	ADF_e_t	β_1	P-value	γ_1	P-value	λ	P-value
Beijing	-0.245	0.770	0.285	0.990	-3.128	0.203	0.003	0.007	0.010	0.150	0.000
Tianjin	1.481	0.574	0.238	0.993	-3.636	0.357	0.008	0.084	0.021	0.110	0.005
Hebei	2.491	0.562	0.225	0.997	-3.512	0.316	0.004	0.022	0.012	0.114	0.003
Shanxi	0.561	0.408	0.547	0.981	-3.306	0.264	0.029	0.048	0.011	0.066	0.006
InnerMongolia	0.570	0.363	0.589	0.988	-3.948	0.104	0.034	0.013	0.008	0.366	0.000
Liaoning	1.212	0.187	0.703	0.968	-3.839	0.106	0.006	0.011	0.005	0.015	0.000
Jilin	1.284	0.432	0.409	0.989	-3.755	0.304	0.003	0.115	0.010	0.054	0.002
Heilongjiang	1.385	0.355	0.511	0.989	-3.847	0.219	0.067	0.168	0.004	0.243	0.000
Shanghai	-0.542	0.564	0.515	0.968	-3.576	0.305	0.000	0.008	0.013	0.059	0.000
Jiangsu	2.223	0.718	0.139	0.996	-3.834	0.346	0.000	0.108	0.009	0.050	0.000
Zhejiang	2.015	0.705	0.169	0.996	-3.851	0.468	0.000	0.023	0.002	0.227	0.000
Anhui	2.721	0.562	0.207	0.990	-3.847	0.301	0.006	0.019	0.002	0.158	0.000
Fujian	2.268	0.651	0.179	0.993	-3.103	0.486	0.000	0.017	0.007	0.127	0.000
Jiangxi	3.139	0.640	0.040	0.993	-3.627	0.169	0.016	0.037	0.005	0.117	0.000
Shandong	3.261	0.706	0.012	0.991	-3.806	0.335	0.032	0.083	0.018	0.066	0.000
Henan	2.457	0.585	0.229	0.993	-3.997	0.553	0.000	0.037	0.005	0.030	0.000
Hubei	2.008	0.556	0.293	0.998	-4.119	0.409	0.000	0.058	0.009	0.183	0.000
Hunan	3.068	0.661	0.060	0.997	-3.978	0.469	0.001	0.024	0.007	-0.007	0.000
Guangdong	1.156	0.661	0.352	0.991	-3.837	0.270	0.021	0.027	0.009	0.126	0.001
Guangxi	2.821	0.574	0.148	0.994	-3.964	0.213	0.003	0.058	0.031	0.482	0.001
Hainan	0.490	0.522	0.388	0.986	-3.957	0.339	0.000	0.015	0.015	0.074	0.000
Chongqing	1.140	0.399	0.536	0.980	-4.209	0.047	0.009	0.220	0.007	0.439	0.000
Sichuan	2.680	0.653	0.124	0.997	-3.872	0.338	0.020	0.062	0.030	0.231	0.000
Guizhou	2.882	0.715	0.032	0.997	-3.668	0.361	0.006	0.010	0.008	0.132	0.000
Yunnan	2.727	0.595	0.159	0.998	-3.990	0.269	0.002	0.072	0.016	0.570	0.004
Tibet	1.532	0.700	0.083	0.995	-3.742	0.040	0.005	0.011	0.009	-0.020	0.000
Shaanxi	2.483	0.685	0.068	0.998	-3.743	0.397	0.001	0.027	0.008	0.342	0.000
Gansu	2.300	0.537	0.218	0.986	-3.932	0.157	0.009	0.037	0.016	0.022	0.000
Qinghai	2.038	0.632	0.073	0.994	-3.483	0.130	0.012	0.015	0.003	0.286	0.000
Ningxia	1.119	0.628	0.201	0.991	-3.715	0.246	0.008	0.043	0.010	0.145	0.000
Xinjiang	1.505	0.645	0.200	0.992	-3.542	0.279	0.007	0.024	0.010	0.318	0.000

The critical value of ADF test for cointegration test with 3 variables and 30 sample sizes at 5% significance level is -3.461.The unit root test results of the residual sequence are shown in ADF_e_t in Table 1. It can be seen from Table 1 that the residual sequence of each regression equation is stationary, and there is a co-integration relationship between lngdp-lninv-edu.

2.4. Error Correction Model (ECM)

The cointegration test proves that there is a long-term equilibrium relationship between variables. In order to avoid short-term fluctuations in the long-term equilibrium relationship, reduce errors, and test whether there is a short-term equilibrium relationship between the variables, the Error Correction Model came into being. The stable residual series of regression equation with long-term equilibrium relationship is put into the model as error correction term, and the model is established together with other explanatory variables. Specific modeling steps:

The first step is to find the residual sequence as the error correction term (ecm) from the cointegration equation of lngdp-lninv-edu. The residual formula of the three variables is:

$$ecm_{t-1} = Y_{t-1} - a_0 - a_1 X_{t-1} - a_2 Z_{t-1} \qquad (2)$$

Table 2. Order and coefficient estimation of ARIMA model for investment and education in different regions

province	inv_ARIMA(p,d,q)										edu_ARIMA(p,d,q)									
	p	d	q	β_0	Zvalue	β_1	Zvalue	β_2	Zvalue	Wald	p	d	q	β_0	Zvalue	β_1	Zvalue	β_2	Zvalue	Wald
Beijing	1	1	0			0.801	8.24			67.96	1	2	0			-0.519	-4.27			18.21
Tianjin	1	1	0			0.931	16.3			265.83	1	2	0			-0.766	-6.78			45.99
Hebei	1	1	0			0.931	17.11			292.82	1	2	0			-0.490	-2.23			4.98
Shanxi	1	1	0			0.479	2.48			6.17	1	2	0			-0.634	-4.22			17.8
InnerMongolia	1	2	0			-0.731	-4.97			24.69	1	2	0			-0.417	-2.59			6.72
Liaoning	1	1	0			0.596	7.14			50.96	1	2	0			-0.554	-3.66			13.37
Jilin	2	1	0			0.766	2.93	0.376	1.89	45.06	1	2	0			-0.600	-5.4			29.21
Heilongjiang	1	1	0			0.802	5.84			34.09	1	2	0			-0.536	-2.53			6.39
Shanghai	1	1	0			0.864	9.69			93.95	1	1	0			-0.499	-3.37			11.36
Jiangsu	1	3	0			-0.440	-3.22			10.37	1	2	0			-0.534	-2.91			8.49
Zhejiang	2	3	0			-0.611	-2.45	-0.419	-2.350	7.82	1	2	0			-0.544	-2.72			7.4
Anhui	1	3	0			-0.506	-4.29			18.42	2	2	0			-0.868	-3.35	-0.548	-2.27	11.23
Fujian	1	1	0			0.922	13.16			173.24	1	2	0			-0.708	-4.51			20.37
Jiangxi	1	1	0			0.919	12.44			154.84	1	2	0			-0.668	-5.23			27.3
Shandong	1	1	0			0.926	16.31			266.05	1	2	0			-0.481	-2.8			3.23
Henan	1	1	0			0.926	9.26			85.67	1	2	0			-0.573	-2.92			8.55
Hubei	1	1	0			0.934	16.12			259.75	1	2	0			-0.610	-4.27			18.22
Hunan	1	1	0			0.940	15.26			232.95	1	2	0			-0.701	-3.8			14.45
Guangdong	1	1	0			0.839	8.78			77.03	1	2	0			-0.588	-3.55			12.58
Guangxi	1	1	0			0.875	15.4			237.12	1	2	0			-0.707	-5.03			25.32
Hainan	1	1	0			0.779	12.06			145.51	1	2	0			-0.616	-2.88			8.27
Chongqing	1	3	0			0.916	-5.82			33.88	1	2	0			-0.611	-4.46			19.85
Sichuan	1	1	0			0.877	13.87			192.43	1	2	0			-0.463	-2.38			5.65
Guizhou	1	1	0			0.941	16.52			272.86	1	2	0			-0.527	-2.97			8.83
Yunnan	1	1	0			0.891	15.99			255.64	1	2	0			-0.609	-3.63			13.17
Tibet	1	1	0	0.205	13.13	0.664	5.15			26.52	1	2	0			-0.497	-2.68			7.18
Shaanxi	1	1	0			0.890	11.19			125.32	1	1	0	0.103	4.4	-0.603	-6.38			19.04
Gansu	1	1	0			0.728	3.8			14.47	1	2	0			-0.582	-3.41			11.65
Qinghai	1	1	0			0.820	8.66			74.97	1	1	0			-0.625	-3.85			14.79
Ningxia	1	1	0			0.914	13.63			185.87	1	2	0			-0.646	-4.1			16.81
Xinjiang	1	1	0			0.922	12.57			157.91	1	2	0			-0.595	-4.91			24.16

The second step is to establish the error correction model of lngdp-lninv-edu, and the first-order error correction model formula of three variables is:

$$\Delta Y_t = \beta_1 \Delta X_t + \gamma_1 \Delta Z_t - \lambda ecm_{t-1} + \mu_t \qquad (3)$$

Substitute ecm as a new variable and the data corresponding to other independent variables into the above formula, and use ordinary least squares to estimate the corresponding parameters. The results are shown in Table 1.

Figure 1 GDP development trend of China's provinces from 1990 to 2050

Note: In the figure 1, the yellow part is the trend of GDP development from 1990 to 2019 and the red part

is the forecast trend of GDP development from 2020 to 2050.

2.5. Autoregressive Moving Average Model (ARMA)

The ARIMA model is a classic model for studying time series data. It uses the potential connections between variables to predict variables. It has a good fitting effect and is widely used in research. The model first performs stationarity test on the time series. When the test is a stationary time series data, it uses the ACF and PACF functions, combined with the AIC and BIC criteria to determine the lag order of the model, and establishes the model prediction formula. The formula of ARMA (p, q) model is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \qquad (4)$$

From the Equation (3) of the Error Correction Model, the data of $\triangle X_t$, $\triangle Z_t$ and ecm_{t-1} are needed to calculate the short-term fluctuation $\triangle Y_t$. Therefore, to predict the short-term fluctuations of lngdp, we need to obtain the forecast data of lninv, edu and ecm. The ARIMA model is used to obtain the predicted values of lninv and edu in each region from 2018 to 2050. The prediction coefficients of ARIMA model of lninv and edu are shown in Table 2.

2.6. Forecast Results

Substitute lninv and edu into Equation (4) and ecm_{t-1} into Equation (2) to calculate the forecast data. Substitute the forecast data corresponding to lninv, edu, ecm_{t-1} into Equation (3) to find the annual growth value of lngdp, and finally, calculate the annual GDP. The forecast results are shown in Figure 1.

3. CONCLUSIONS AND SUGGESTIONS

Judging from the forecast results, the GDP of each province from 2020 to 2050 will be on the rise. China's economy will continue to develop steadily and rapidly, and the development speed of the eastern coastal regions is faster. Beijing and Shanghai still maintain their current growth rates, the GDP of Guangdong, Zhejiang and Jiangsu developed rapidly. The annual growth rate of GDP in most provinces of China is more than 5%. It is estimated that China's per capita GDP in 2030 will be about 110000 yuan, reaching 15948 US dollars at the average exchange rate in 2020; in 2050, China's per capita GDP is about 300000 yuan, which will reach 43494 US dollars according to the average exchange rate in 2020.

Policy suggestions also could be in the following as: firstly, it can be seen that the development of regional GDP in the future is closely related to the development of regional GDP in the current period.

Therefore, the government should develop the economy and maintain better economic development. Secondly, with the development of economy, fixed asset investment is still an effective means to stimulate economic growth. Increasing fixed asset investment can not only produce current GDP increment, but also bring long-term GDP increment. Therefore, fixed asset investment has dual effects. In the future, we should appropriately increase investment in fixed assets and focus on advantageous industries and emerging industries, so as to increase employment, stimulate current production and consumption, promote consumption growth and improve people's living standards. Thirdly, talents are the most potential resources, and talents are the key factor for the development of national undertakings. In the future, the government should increase investment in talent training, actively train high-quality talents, strengthen the popularization of basic education and encourage the training of skilled workers; at the same time, the government should encourage scientific and technological innovation, cultivate innovative talents and develop knowledge-based economy.

ACKNOWLEDGMENTS

REFERENCES

[1] Luo Sen, Situ Xueying. Research on China's quarterly GDP forecast based on VAR model [J]. China Price, 2020(02): 25-28.

[2] Guo Qingchun, He Zhenfang. Economic forecasting model based on artificial neural network [J]. Computing Technology and Automation, 2014, 33(01): 132-136.

[3] Shi Ping, Tang Jun. The application of gray GM (1, n) model in economic forecasting——taking the economic development of Baotou as an example [J]. Journal of Inner Mongolia University (Natural Science Edition), 2015, 46(01): 43-47.

[4] Jiang Tong, Zhao Jing, Cao Lige, Wang Yanjun, Subuda, Jing Cheng, Wang Run, Gao Chao. Forecast of economic changes in China and its provinces under the shared social economic path [J]. Progress in Climate Change Research, 2018, 14(01):50-58.

[5] Li Ping, Lou Feng, Wang Hongwei. Analysis and forecast of China's economic aggregate and its structure from 2016 to 2035[J]. China Engineering Science, 2017, 19(01): 13-20.

[6] Geng Peng, Qi Hongqian. Prediction and evaluation of real time quarterly GDP data in China [J]. Statistical Research, 2012, 29(01): 8-14.