ATLANTIS PRESS
Advances in Economics, Business and Management Research, volume 82
International Conference on Management, Education Technology and Economics (ICMETE 2019)

Establishment of Multivariate Linear and Logarithmic Function Mixed Model to Predict Economic Trend Based on 2005-2015 Data

Huimin Zou, Jiayi Wu
Beijing Polytechnic
Beijing,100176, China

***Abstract*—According to the China Statistical Yearbook, the data of 2005-2015 were found. Eviews was used to establish a multiple linear regression model. The model did not pass the t test. Then a mixed model of multiple linear and logarithmic functions was established. The economic significance test, statistical test, econometric test and prediction test of the model were all passed, and the goodness of fit was close to 1. Secondly, we used this model to forecast the economic growth trend in the coming years. In the future, the decline of population growth rate (aging) would lead to the slowdown of economic growth. From the model, it can be seen that the economic growth rate was closely related to population growth. The downward trend of economic growth rate was irreversible and can only be delayed. Finally, from the perspective of the model, length of education was the second factor affecting economic development, so we suggested to the government: (1) Open the multi-child policy (or the three-child policy); (2) Implement general education to increase the length of education for all Reach to delay the downward trend of GDP.**

Keywords—Mixed model of multiple linear and logarithmic functions; eviews; Population size; Length of education Educational appropriations

I. INTRODUCTION

As the world's most populous country, education is also affecting economic growth. Considering that China's education has entered a new stage of educational reform today. Appropriate measures should be taken to carry out education reform in a timely manner, which is a good way for future economic growth. The educational level of the people depends on many aspects, including the important length of education.

According to the data in recent years, although the labor force population has been increasing, the growth rate of the labor force has been decreasing. With the passage of time, the effective labor force will also be decreasing. The term "demographic dividend" may no longer exist.

In our later calculation, there will be two results, one is the result of neglecting the demographic dividend, the other is the result of calculating the demographic dividend.

There is also the size of the population. As there are many cities in our country, there may be huge differences in economic level, population, economy and other aspects. It is not suitable for uniform comparison, nor for average calculation. It is unfair to over-strong or over-poor areas.

So, in response to this problem, we have found a relatively poor region, Gansu, to compare with the national data.

Therefore, we set up a model to discuss how the level of education will affect China's economic growth in the next 20 years. Especially the future economic growth rate and income level. Whether the average annual per capita economic growth in the next 20 years can be about 7%.

II. DATA COLLECTION

We need to collect population size. Average length of education GDP; Data such as education expenditure can be used to predict future economic growth rate and income level.

These data are the population data and other information involved in the content of China Statistical Yearbook 2016 published by the National Bureau of Statistics of the People's Republic of China.

However, in view of the differences between the cities mentioned before, we cannot collect local statistical data based on these data. We have collected the data of Gansu Statistical Yearbook and the sixth census to make up for the data we did not collect.

Although the data of China Statistical Yearbook 2016 is insufficient in timeliness, it is more detailed than the sixth census report and can meet the needs of numerical simulation. It is more comprehensive than Gansu Statistical Yearbook.

The model used in this paper is compared with the data collected from China Statistical Yearbook 2016 and Gansu Statistical Yearbook. The sixth census was compensated with some data that had not yet been collected.

III. ESTABLISHMENT OF MODEL

In order to analyze the economic growth of our country, we can use the economic index of gross national product (Y) as the research object, use the total population of the whole country (X_1) to measure the population, use the total investment in education (X_2) to measure the degree of emphasis on education, and use the length of education per capita (X_3) to measure the education level of the people. The data mentioned above are used for regression analysis.

The explanatory variable here is Y: Gross national product (GNP)

The economic factors closely related to Y-GNP are three possible explanatory variables of the model.

X_1 represents the total population of the whole country

X_2 represents the total investment in education, and

X_3 represents the length of education per capita.

C is the random disturbance item

The establishment of the model can be roughly divided into several steps: theoretical model setting, parameter estimation, model checking and model updating. If the model conforms to the actual economic theory and passes all levels of tests, then the model can be used as the final model for structural analysis and economic prediction.

A. Determination of theoretical model

Through the selection of variables, the following variables are finally determined to establish the regression model.

Explained variable Y: GDP explanatory variable X_: total population of the whole country

X_2 : Total investment in education

X_3 : Length of education per capita

In addition, from the economic point of view, the total population of the whole country, the total investment in education and the length of education per capita determine the level of economic growth to a great extent. In the sense of single economy, the choice of variables is correct. Intuitively speaking, the explanatory variables are related to the explained variables, and the three explanations are all

"Benign" variables of economic growth, and their growth has a positive development in China.

TABLE I 2005-2015 Explained Variables and Explanatory Variables Data

Year	GNI /RMB 100 million	Total population of the whole country /100,000,000 people	Education expenditure/RMB 100 million	Length of education per capita/year
2005	185998.9	13.0756	8418.8391	7.831
2006	219028.5	13.1448	9815.3087	8.04
2007	270844	13.2129	12148.066	8.186
2008	321500.5	13.2802	14500.737	8.27
2009	348498.5	13.345	16502.707	8.38
2010	411265.2	13.4091	19561.847	8.51
2011	484753.2	13.4735	23869.294	8.846
2012	539116.5	13.5404	28655.305	8.942
2013	590422.4	13.6072	30364.718	8.98
2014	644791.1	13.6782	32806.469	9.037
2015	686449.6	13.7462	36129.193	9.077

The above data are from China Statistical Yearbook.

Firstly, check whether the linear relationship between the explained variable and the explanatory variable is established. Observe the scatter diagram between the explained variable and the explanatory variable.

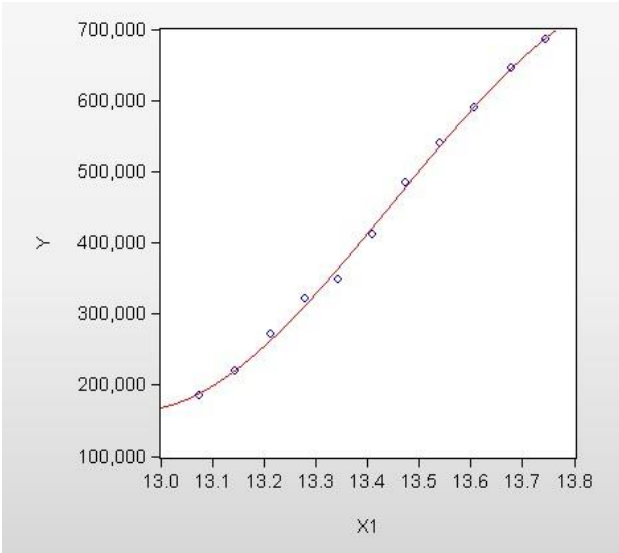

Fig. 1 Scatter diagram between the explained variable Y and the explanatory variable X_1

Judging from the trend line in the graph, the relationship between the explained variable and the explanatory variable is basically linear.

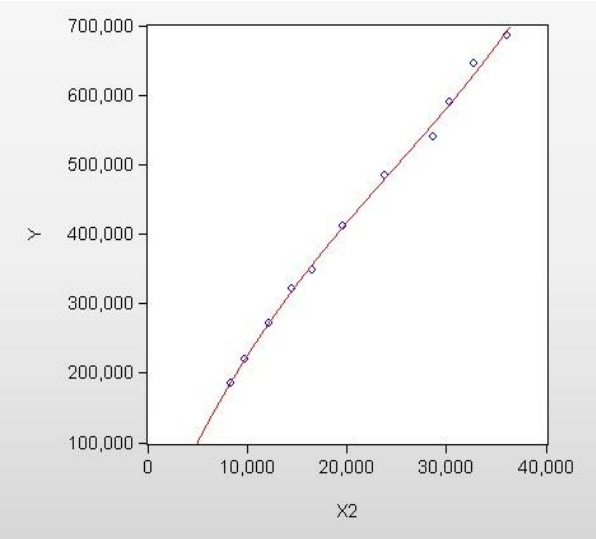

Fig. 2 Scatter diagram between the explained variable Y and the explanatory variable X_2

Judging from the trend line in the graph, can be judged from the trend line in the figure, and the relationship between the explained variable and the explanatory variable X_2 is basically linear.

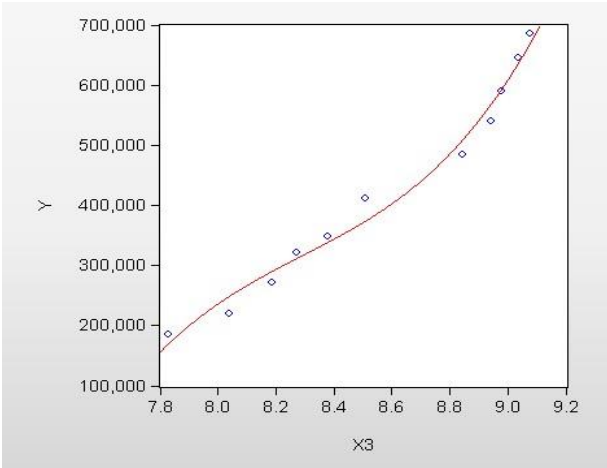

Fig. 3 Scatter diagram between the explained variable Y and the explanatory variable X_2

Judging from the trend line in the graph can be judged from the trend line in the figure, and the relationship between the explained variable Y and the explanatory variable X_2 is basically linear.

It can be judged through the relevant coefficient between variables.

Covariance Analysis: Ordinary

Date: June 2, 2018 Time: 05:16

Sample: 2005 2015

Included observations: 11

TABLE II Matrix table of relevant coefficient of the explained variable and the explanatory variable

Correlation	Y	X1	X2	X3
Y	2.73E+10 1.000000			
X1	34680.45	0.044269		
	0.997067	1.000000		
X2	1.53E+09	1939.533	86255279	
	0.997824	0.992552	1.000000	
X3	55166.49	0.070379	3100.729	0.183038
	0.780003	0.781851	0.780371	1.000000

It can be seen from the table that there is a high correlation between the explained variable Y and the explanatory variables X_1, $X_2$and X_3. Through the judgment of scatter diagram and correlation coefficient matrix table, it can be judged that there is an obvious linear correlation between explained variables and explanatory variables. At the same time, through the relevant graph analysis of the explained variables and the explanatory variables, the model is set up as follows:

$$y = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + C$$

B. Establish the initial model

1) Parameter estimation

Dependent Variable: Y

Method: Least Squares

Date: June 2, 2018 Time: 05:24

TABLE III Output result of least square method (LS) parameter estimation

Sample: 2005 2015 Included observations: 11				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4529503.	1152866.	-3.928906	0.0057
X1	355338.6	89473.54	3.971438	0.0054
X2	9.840890	2.020992	4.869335	0.0018
X3	-1944.610	8572.339	-0.226847	0.8270
R-squared	0.998665	Mean dependent var		427515.3
Adjusted R-squared	0.998093	S. D. dependent var		173382.6
S. E. of regression	7571.239	Akaike info criterion		20.97739
Sum squared resid	4.01E+08	Schwarz criterion		21.12208
Log likelihood	-111.3756	Hannan-Quinn criter.		20.88618
F-statistic	1745.724	Durbin-Watson stat		2.765301
Prob(F-statistic)	0.000000			

The initial model obtained is as follows:

$Y=4529503.13462+355338.641102X_1+9.84089010994X_2-1944.60951388X_3$

2) Model test

Four levels of tests including economic significance test, statistical test, econometric test and prediction test are carried out for the established model.

a) Economic significance test

The coefficients of explanatory variables are α1=355338.6411024277, α2=9.840890109935449 respectively.

The coefficients of the two variables are positive, which accords with the positive correlation between the explained variables and the explanatory variables, accords with the economic reality that the explanatory variable growth drives the explained variable growth, α3=1944.60951387848, accords with the negative correlation between the explained variable and the explanatory variable, and accords with the real economic significance. Therefore, the model passes the economic significance test.

b) Statistical check

(1) Test of goodness of fit R-squared=0.9986651848765925, Adjusted R-

Squared=0.998093 degree of fitting is very high, and the two items are basically closed to 1, and the degree of fitting of equation is very good.

(2) Significance test of variable

TABLE IV Significance test of model coefficient

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4529503.	1152866.	-3.928906	0.0057
X1	355338.6	89473.54	3.971438	0.0054
X2	9.840890	2.020992	4.869335	0.0018
X3	-1944.610	8572.339	-0.226847	0.8270

The test results show that one of them fails to pass the test, so we use exponential or logarithmic fitting.

The test results show that one of the constant terms c (1) has not passed the test, so it can use logarithmic fitting.

TABLE I

c) Correction model

TABLE V Significance test of exponential fitting

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	96.40078	47.68696	2.021533	0.0739
C(2)	2.643926	0.147986	17.86605	0.0000

TABLE VI Significance test of logarithmic fitting

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.203687	0.794328	-6.551055	0.0001
LOG(X3)	8.430698	0.370178	22.77470	0.0000

The test results show that the adjoin probability of t test for all explanatory variable coefficients including the constant term is less than 5%, and the coefficient of X_3 is not 0, so it passes the significance test. The final model is:

$$Y=\frac{1}{3}(355338.64X_1+9.84\ X_2)+\frac{1}{2}(-5.20+8.43\log X_3)$$

The model retains two digits for each coefficient item.

① In the test of goodness of fit R-squared=0.982944 Adjusted R-squared=0.981049, it can be seen that the degree of fitting is very high, close to 1, and the degree of fitting of equation is very good.

(1) Heteroscedasticity test is carried out: We use Brunspagan test to test heteroscedasticity, and the results are as follows:

d) Econometric test

TABLE VII Heteroscedasticity test result

F-statistic	0.321676 Prob. F(3,7)	0.8099
Obs*R-squared	1.332739 Prob. Chi-Square(3)	0.7214
Scaled explained SS	0.288008 Prob. Chi-Square(3)	0.9623

The test results show that there is no heteroscedasticity in the model, so there is no need to correct the model.

Multiple collinearity test: The stepwise regression method is used to test as follows:

TABLE VIII Estimation result of the least square product of explained variables Y and X_1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10078074	268813.7	-37.49092	0.0000
X1	783398.1	20042.88	39.08611	0.0000
R-squared	0.994143	Mean dependent var		427515.3
Adjusted R-squared	0.993493	S. D. dependent var		173382.6
S. E. of regression	13986.45	Akaike info criterion		22.09253
Sum squared resid	1.76E+09	Schwarz criterion		22.16488
Log likelihood	-119.5089	Hannan-Quinn criter.		22.04693
F-statistic	1527.724	Durbin-Watson stat		1.050580
Prob(F-statistic)	0.000000			

TABLE IX Estimation result of the least square product of explained variable Y and X_2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51669.93	9040.176	5.715589	0.0003
X2	17.76112	0.391189	45.40293	0.0000
R-squared	0.995653	Mean dependent var		427515.3
Adjusted R-squared	0.995170	S. D. dependent var		173382.6
S. E. of regression	12049.68	Akaike info criterion		21.79443
Sum squared resid	1.31E+09	Schwarz criterion		21.86677
Log likelihood	-117.8694	Hannan-Quinn criter.		21.74883
F-statistic	2061.426	Durbin-Watson stat		1.689311
Prob(F-statistic)	0.000000			

TABLE X Estimation result of the least square product of explained variable Y and X_3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2863543.	210185.4	-13.62389	0.0000
X3	384718.7	24540.43	15.67693	0.0000
R-squared	0.964673	Mean dependent var		427515.3
Adjusted R-squared	0.960748	S. D. dependent var		173382.6
S. E. of regression	34350.62	Akaike info criterion		23.88959
Sum squared resid	1.06E+10	Schwarz criterion		23.96194
Log likelihood	-129.3928	Hannan-Quinn criter.		23.84399
F-statistic	245.7661	Durbin-Watson stat		0.894614
Prob(F-statistic)	0.000000			

It can be seen from the figure that the goodness of fit between Y and X_2 is the largest. R-squared=0.995653

. Make regression model of Y and X_1, X_2

TABLE XI Estimation result of the least square product of explained variable Y and X_1, X_2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4519309.	1081542.	-4.178581	0.0031
X1	353395.6	83616.09	4.226406	0.0029
X2	9.814676	1.894297	5.181170	0.0008
R-squared	0.998655	Mean dependent var		427515.3
Adjusted R-squared	0.998319	S. D. dependent var		173382.6
S. E. of regression	7108.230	Akaike info criterion		20.80290
Sum squared resid	4.04E+08	Schwarz criterion		20.91141
Log likelihood	-111.4159	Hannan-Quinn criter.		20.73449
F-statistic	2970.801	Durbin-Watson stat		2.759490
Prob(F-statistic)	0.000000			

TABLE XII Estimation result of the least square product of explained variable Y and X_2, X_3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-335264.0	311468.6	-1.076397	0.3131
X2	15.50028	1.858408	8.340625	0.0000
X3	50824.48	40895.73	1.242782	0.2491
R-squared	0.996356	Mean dependent var		427515.3
Adjusted R-squared	0.995446	S.D. dependent var		173382.6
S.E. of regression	11700.92	Akaike info criterion		21.79972
Sum squared resid	1.10E+09	Schwarz criterion		21.90824
Log likelihood	-116.8985	Hannan-Quinn criter.		21.73132
F-statistic	1093.844	Durbin-Watson stat		1.737729
Prob(F-statistic)	0.000000			

TABLE XIII Estimation result of the least square product of explained variable Y and X_1, X_2 and X_3 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4396414.	1143634.	-3.844250	0.0063
X1	333915.8	92626.08	3.604987	0.0087
X2	9.498578	2.037996	4.660745	0.0023
X3	16952.93	27520.91	0.616002	0.5574
R-squared	0.998725	Mean dependent var		427515.3
Adjusted R-squared	0.998178	S. D. dependent var		173382.6
S. E. of regression	7401.065	Akaike info criterion		20.93192
Sum squared resid	3.83E+08	Schwarz criterion		21.07661
Log likelihood	-111.1256	Hannan-Quinn criter.		20.84072
F-statistic	1827.035	Durbin-Watson stat		2.741038
Prob(F-statistic)	0.000000			

The goodness of fit (R-squared=0.998655) of least square product estimation of Y and X_1 is observed. Compared with goodness of fit (R-squared=0.964673) of least square product estimation of Y and X_1, the influence of X_1 on Y is not obvious. The goodness of fit (R-squared=0.996356) of least square product estimation of Y and $X_2$and X_3 is observed. Compared with goodness of fit (R-squared=0.964673) of least square product estimation of Y and X_3, the change is obvious, indicating that the influence of X_3 on Y is obvious. Through the above tests, the final initial model is:

$$\boxed{Y = \frac{1}{2}(355338.64X_1 + 9.84X_2) + \frac{1}{2}(-5.20 + 8.43\log X_3)}$$

C. Conclusion

So the final model we establish

$$\boxed{Y = \frac{1}{2}(355338.64X_1 + 9.84X_2) + \frac{1}{2}(-5.20 +}$$

8.43 logX model undergoes and passes the four-level tests including economic significance test, statistical test, econometric test and prediction test, and the goodness of fit is close to 1.

IV. Use Model to Predict

A. Prediction

Use Forcast to predict GDP of the future 20 years.

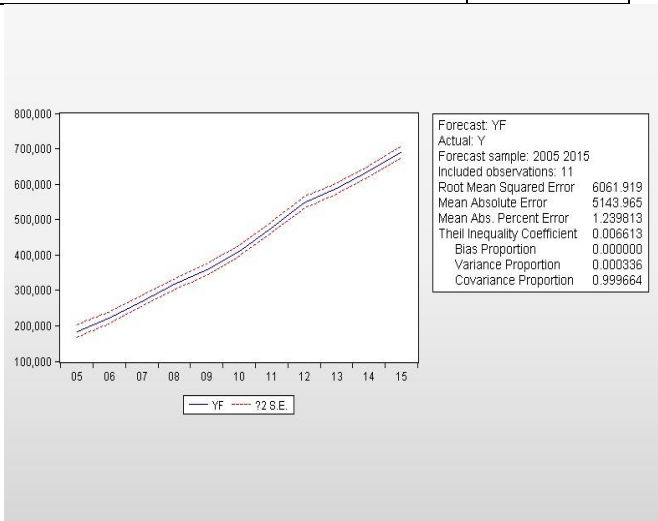

Fig. 4 Changes of prediction curve of Y and X

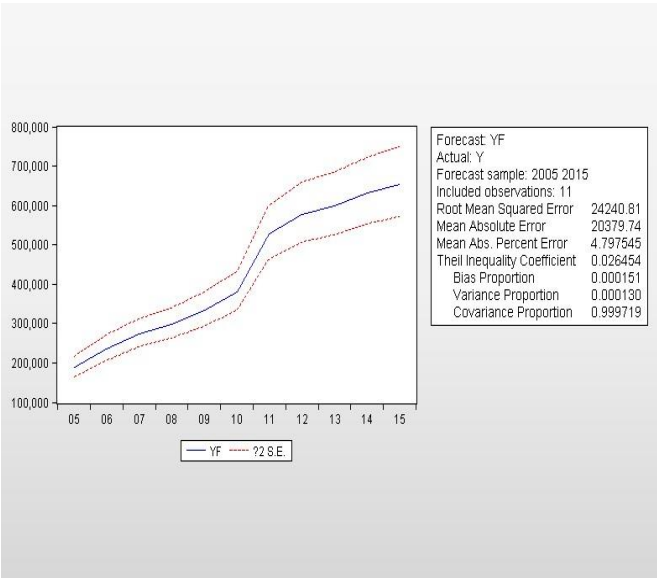

Fig. 5 Changes of prediction curve of Y and X

In the figure, the data show that the MAPE values are less than 10%, so the error is small and the accuracy is high. It can be predicted directly without correction.

B. Data results

The final prediction results are calculated in combination with Excel as follows:

TABLE XIV Prediction of changes of the gross domestic product (GDP) in the next 20 years

Year	GDP / RMB 100 million	GDP Growth Rate
2016	740362.2	0.075762
2017	792503.3	0.070426
2018	844644.5	0.065793
2019	896785.6	0.061731
2020	948926.7	0.058142
2021	1001067.9	0.054948
2022	1053209	0.052085
2023	1105350.2	0.049507
2024	1157491.3	0.047172
2025	1209632.5	0.045047
2026	1261773.6	0.043105
2027	1313914.8	0.041324
2028	1366055.9	0.039684
2029	1418197	0.038169
2030	1470338.2	0.036766
2031	1522479.3	0.035462
2032	1574620.5	0.034248
2033	1626761.6	0.033113
2034	1678902.8	0.032052
2035	1731403.9	0.031057

From the model, we can see that the downward trend of economic growth rate is irreversible, and it is impossible to achieve 7% per capita economic growth rate in the next 20 years.

V. POLICY SUGGESTIONS

The policies adopted in education make all students equal and learn the same knowledge together, to raise the level of education by one level. The improvement of education level has also led to the improvement of the quality of total national income. Special education also needs to be sustained by the economy, which implies that the welfare of education owned by us has increased. The policy of economic growth has strengthened the market power, and the foreign trade and other fields have also been promoted, which has greatly enhanced the economic growth.

REFERENCES

[1] China Statistical Yearbook published by National Bureau of Statistics;

[2] Gansu Statistical Yearbook published by Gansu Province Bureau of Statistics

[3] Research on the Meaning and Nature of Econometrics: Liu Liyan

[4] The Nineteenth National Congress Report from

[5] Priority to the Development of Education http://cpc.people.com.cn/19th/n1/2017/1021/c414305 - 29600097.html

[6] National Population Size Forecast: Xiaoyu ZENG, Juan ZHENG, Wenjun CAO, Hanwen XU