A New Method -- Multi-factor Trend Regression and Its Application to Economy Forecast in Jiangxi*

DING Yuechao
College of Computer Engineering, Jimei University, Xiamen, Fujian 361021, China
E-mail:ding@jmu.edu.cn

Abstract

The principle of a new method called Trend Regression is introduced and applied to the economy forecast of Jiangxi Province. The method improved previous time series forecasting method in which only self-extension is done and multiple factors (variables) are not taken into consideration. Also, it got over the weakness of forecasting by general regression analysis that relies on simultaneous independent variables. A time series is the function of multiple factors. The values (independent variables) in a period may affect the value (dependent variable) to be predicated in the next period. The nearer the sample time to the predicted time, the more important the sample to the predict value. By shifting the dependent variable to establish models, sequential regression and prediction can be realized. In this way the trend of information can be mined.

Key Words: *Economy Forecast; Time Series; Trend Regression*

1. Introduction

Forecast means predicting what will happen in the future based on accurate statistical data, by combing scientific methods with the perspective of history, actual conditions and objective laws. The data from economy development is a type of time series information. Time series information, which varies with time, is the function of time, making it possible to forecast the future value of time series.

In general, for single indicator forecast, time series smooth method is applied. It includes shifting average method, index smooth method, difference index smooth method, self-adaptive filtering method, linear model method, multinomial model method, exponential curve model method, revision exponential curve model method, growth curve model method, and seasonal variation method [1] [2].

However, the generating of time series data is the synthetically result of multiple factors. By using simple linear regression, neither the linear extension nor the curved extension of a single factor can accurately predict the new value. By using multivariate regression, the value of dependent variables can be achieved only when the simultaneous independent variables are known. For general matters, the values of independent variables and dependent variable are produced simultaneously. Therefore, it is not necessary to predict the dependent variable's value while it is known.

Meanwhile, in regression analysis, the sequence of samples' values is not considered in the model. In other words, there is no method of sequential regression in math. So it's necessary to improve ordinary multivariate regression analysis to make it applicable to prediction or trend analysis.

2. The forecasting method -- regression

The time series data varies with a number of influence factors, and it is a function of influencing factors. So regression analysis [3] [4] method is usually applied. The relationship between the related variables can be linear or non-linear. Only linear regression analysis is discussed here.

When the function contains only one independent variable, it is called unary linear regression, and its model is:

$$Y_i = a + bx_i + \varepsilon_i \quad i=1, 2, ...$$

where a and b are called regression coefficients, and εi represents the total sum of the effect all kinds of stochastic factors have on the dependent variable.

* The project was supported by the Natural Science Foundation of Fujian Province of China (No.A0640008)

E-FORENSICS 2008, January 21-23, Adelaide, Australia

DOI 10.4108/wkdd.2008.2652

When there are multiple independent variables, it is called multiple regression. Suppose $x_1, x_2, ..., x_p$ are measurable or controllable variables, and if variable y has a linear relation with $x_1, x_2, ..., x_p$, then n group of data can be collected after n times of experiments:

$y_i, x_{i1}, x_{i2}, ..., x_{ip}, \quad i = 1, 2, ..., n$

Multiple regression models can be represented as:

$$y_1 = b_0 + b_1x_{11} + b_2x_{12} + ... + b_px_{1p} + \varepsilon_1$$
$$y_2 = b_0 + b_1x_{21} + b_2x_{22} + ... + b_px_{2p} + \varepsilon_2$$
$$... ...$$
$$y_n = b_0 + b_1x_{n1} + b_2x_{n2} + ... + b_px_{np} + \varepsilon_n$$

where b_0, b_1, b_2, ..., b_p are p+1 undetermined parameters, also known as regression coefficient. ε_i denotes random factor's influence on y_i during the experiment, and it is usually ignored.

Through the solution of the linear equation, the values of parameter b_0, b_1, b_2, ..., b_p can be obtained, and p-variable regression equation is shown below:

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_px_p$$

The goal of establishing regression equation is to utilize it to forecast and control. In reality, the relationship between random parameter y and $x_1, x_2, ..., x_p$ cannot be judged. Linear regression model is only an assumption before regression equation is solved. Therefore, it needs to be tested statistically.

The regression analysis does not consider the order of time. The value of dependent variable is obtained merely based on the linear expression of influencing factors. Furthermore, some values of dependent variables, such as economic indicators, can only be obtained on condition that the values of other variables during the same period are known. Therefore, it is not forecast in the strict sense. The true forecast means the values are forecast before actual events occur.

3. A new method of forecasting

Based on the view that the values of the indicators at a time is the foundation of the next value of the indicator to be forecast, the dependent variable at a later time can be deem to be a function of all the independent variables at a previous time. Thus the regression model can be improved.

A. The model shifting dependent 1 period forward

$$Y_2 = b_0 + b_1x_{11} + b_2x_{12} + ... + b_px_{1p} + \varepsilon_1$$
$$Y_3 = b_0 + b_1x_{21} + b_2x_{22} + ... + b_px_{2p} + \varepsilon_2$$
$$... ...$$
$$Y_n = b_0 + b_1x_{m1} + b_2x_{m2} + ... + b_px_{mp} + \varepsilon_m$$

where m = n – 1. The model is one line less than the common regression model and the independent variables' values of the last sample is not taken into account while processing data, which has little effect on the solution to the equation when a large number of samples are available. The last sample (in the time n) may be put into the equation to forecast the value Y_{n+1} in the time n+1.

B. The model shifting dependent 2 periods forward

$$Y_3 = b_0 + b_1x_{11} + b_2x_{12} + ... + b_px_{1p} + \varepsilon_1$$
$$Y_4 = b_0 + b_1x_{21} + b_2x_{22} + ... + b_px_{2p} + \varepsilon_2$$
$$... ...$$
$$Y_n = b_0 + b_1x_{m1} + b_2x_{m2} + ... + b_px_{mp} + \varepsilon_m$$

where m = n – 2. The model is two lines less than the common regression model and the independent variables' values of the last two samples are not taken into account while processing data, which has little effect on the solution to the equation when a large number of samples are available. The last sample but two (in the time n-1) may be put into the equation to forecast the value Y_{n+1} in the time n+1.

C. The model shifting dependent 3 periods forward

$$Y_4 = b_0 + b_1x_{11} + b_2x_{12} + ... + b_px_{1p} + \varepsilon_1$$
$$Y_5 = b_0 + b_1x_{21} + b_2x_{22} + ... + b_px_{2p} + \varepsilon_2$$
$$... ...$$
$$Y_n = b_0 + b_1x_{m1} + b_2x_{m2} + ... + b_px_{mp} + \varepsilon_m$$

where m = n – 3. The model is three lines less than the common regression model and the independent variables' values of the last three samples are not taken into account while processing data, which has little effect on the solution to the equation when a large number of samples are available. The last sample but two (in the time n-2) may be put into the equation to forecast the value Y_{n+1} in the time n+1.

The above three models emphasize the influence of i, i-1, i-2 on Y_{i+1}. Forecasting Yn+1 produces three predicted values. Therefore, the average value can serve as the final predicted value. Weighted average method can also be applied here: the weight of predicted value in the model of shifting dependent 1, 2, 3 periods forward is 0.5, 0.3 and 0.2 respectively. To predict the next values of dependent, the nearest sample has the greatest impact on the value to be predicted.

Combining the above three models, the method can reflect the developing trend of the multiple time series. Thus this new method may be called "Multi-factor Trend Regression".

In 2005, we applied the method of "shifting dependent 1 period forward" [5] to predict the per capita GDP of Hubei Province in 2005 as 12087 Yuan. The real value released by statistical bulletin is 11431 Yuan in Feb. 2006, with the error of 5.74%. Now by using "multiple-factor trend regression" method to analyze the data in that time, the predicted value of 2 periods forward is 11921 Yuan and the predicted value of 3 periods forward is 11576 Yuan. The average value by the methods shifting dependent 1, 2 and 3 periods forward is 11861 Yuan with the error of 3.76%, reducing about 2% error. The weighted average with the weights of 0.5, 0.3 and 0.2 is 11935 Yuan, which has 4.41% error compared with the real value. Therefore, this method is quite applicable.

4. The Application of new method to the forecast of indicators in Jiangxi Province

The model "shifting dependent 1 period forward" has been proved to be effective in forecasting the economy indicators of Hubei [5], Fujian [6] and Yunnan [7] provinces. Here, trend regression method is applied to forecasting the economy indicators of Jiangxi province.

A. Original Data

The original data [8] [9] are listed in Table 1. These indicators include: Added Value of Primary Industry (AVPI), Added Value of Secondary Industry (AVSI), Added Value of Tertiary Industry (AVTI) , Freight Ton-Kilometers (FTK), Passenger-Kilometers (PK), Total Investment in Fixed Assets (TIFA), Consumer Price Index (CPI), Retail Sale of Consumer Goods (RSCG), per Capita Annual Income of Rural Households (CAIRH), Natural Growth Rate of Population (NGRP), and per Capita Gross Domestic Product (CGDP). These variables are represented by x1, x2, …, x10 and Y respectively.

Table 1. Eleven economic indicators of Jiangxi province in 1985-2006

Year	AVPI	AVSI	AVTI	FTK	PK	TIFA	CPI	RSCG	CAIRH	NGRP	CGDP
	100M Yuan	100M Yuan	100M Yuan	100M T-km	M P-km	100M Yuan	**	100M Yuan	Yuan	‰	Yuan
1985	84.1	76.1	47.8	233.6	13669	51.2	109.0	85.7	377.3	14.9	597
1986	90.3	83.6	57.0	259.1	14679	53.3	106.6	96.5	395.6	18.6	652
1987	104.6	92.4	65.8	281.5	16124	69.3	106.6	108.8	429.3	15.7	729
1988	119.2	117.4	89.3	296.5	18275	81.6	121.8	140.0	488.2	14.1	891
1989	133.2	131.2	112.0	313.9	17961	75.5	118.5	152.7	558.6	16.8	1013
1990	176.0	133.6	119.1	299.3	17037	71.6	102.1	151.9	669.9	17.1	1134
1991	183.3	154.8	127.1	312.3	17615	91.1	102.8	169.2	702.5	14.1	1212
1992	200.8	199.4	159.3	343.6	21357	123.9	105.7	197.6	768.4	12.5	1439
1993	225.6	282.5	193.9	360.5	20964	196.3	114.6	243.6	869.8	13.4	1835
1994	314.4	338.2	295.6	392.0	23144	237.4	126.9	331.0	1218.2	12.4	2376
1995	374.6	451.1	379.4	411.7	23906	282.5	116.9	410.9	1537.4	11.7	3083
1996	440.0	588.8	488.4	443.6	24716	317.3	108.4	490.4	1869.6	10.5	3715
1997	475.2	658.3	581.8	500.2	30488	329.5	102.0	558.5	2107.3	10.9	4155
1998	450.4	740.3	661.2	569.4	33398	400.6	101.0	605.1	2048.0	9.8	4484
1999	464.4	758.2	740.4	603.9	36900	454.4	98.6	650.5	2129.5	9.5	4661
2000	485.1	700.8	817.2	645.0	41890	516.1	100.3	704.9	2135.3	9.5	4851
2001	506.0	788.1	881.6	653.6	44640	631.8	99.5	763.3	2231.6	9.4	5221
2002	536.0	951.8	962.7	717.6	48910	889.0	100.1	832.7	2306.5	8.7	5829
2003	560.0	1227.4	1043.1	768.6	48210	1303.2	100.8	923.2	2457.5	8.1	6678
2004	711.7	1595.7	1188.5	870.1	55460	1713.2	103.5	1074.5	2786.8	7.6	8189
2005	727.4	1917.5	1411.9	885.2	59240	2176.6	101.7	1236.2	3128.9	7.8	9440
2006	786.3	2319.0	1513.5	935.8	65280	2683.2	101.2	1428.0	3585.0	7.8	10679

** Take the value in previous year as 100.

According to the number of samples and variables and the estimated number of variables to be selected into the model, the F-statistic for significance test of entering and removing variable would be determined while doing stepwise regression. We can establish regression models by shifting the dependent variable 1, 2 and 3 periods forward respectively.

B. 1 period forward

By F-statistic significance test, only AVTI, TIFA and RSCG have remarkable effects on Y, i.e. CGDP, the other indicators having no obvious effects. Regression coefficient can be obtained, as shown in table 2.

The regression equation:

$$CGDP_t = -528.65 - 8.75AVTI_{t-1} + 0.795TIFA_{t-1} + 17.787\ RSCG_{t-1}$$

where t represents period, i.e. year

According to this equation, CGDP in 2007 is predicted as 13760 Yuan, and CGDP between 1986 and 2006 is listed in Table 3, in which only two years' errors exceed 10%, five years' errors between 4% and 8%, the rest below 2.3%. The average error is only 3.6%. There are four years during which the difference between the predicted value and the real value is less than 8 Yuan. Therefore, the method is effective.

C. 2 periods forward

By F-statistic significance test, only AVSI, AVTI, TIFA and RSCG have remarkable significance to CGDP, the other indicators having no obvious effects. Regression coefficient can be obtained, as shown in table 4.

The regression equation:

$$CGDP_t = -857 - 5.358AVSI_{t-1} - 13.937AVTI_{t-1} + 3.457TIFA_{t-1} + 28.883RSCG_{t-1}$$

According to this equation, CGDP in 2007 is predicted as 12420 Yuan. To reduce space of this paper, the rest are not given unnecessary details here.

D. 3 periods forward

By F-statistic significance test, only AVPI, PK, TIFA and CAIRH have remarkable effects on CGDP, the other indicators having no obvious effects. Regression coefficient can be obtained, as shown in table 5.

The regression equation:

$$CGDP_t = -1292 + 29.07AVPI_{t-1} + 0.106PK_{t-1} + 3.435TIFA_{t-1} - 5.63\ CAIRH_{t-1}$$

According to this equation, CGDP in 2007 is predicted as 15472 Yuan. The rest are not given unnecessary details here.

Combing 1, 2 and 3 periods forward methods, $CGDP_{2007}$ is :

$$CGDP_{2007} = (13760 + 12420 + 15472)/3 = 13884\ \text{Yuan}$$

Suppose the weight of three methods is 0.5, 0.3 and 0.2 respectively, $CGDP_{2007}$ is:

$$CGDP_{2007} = 13760*0.5 + 12420*0.3 + 15472*0.2 = 13700\ \text{Yuan}$$

In total, the above independent and dependent variables are 11 indicators. As the development of any event is determined by multiple factors, any indicator can serve as a dependent variable, the rest as the independent variables to make prediction. Applying the same method, we could get the predicted values of all the other indicators of the year 2007. As CPI and NGRP change little every year, they are not discussed here. The predicted results of the rest 9 main economy indicators are shown in Table 6, and their regression models are not listed here. Taking the predicted values of indicators as known sample, the indicators' values of the year 2008 can be forecasted. The results are listed in Table 6, and the accuracy remains to be verified based on the next two years' statistical bulletin.

5. Conclusion

The new method -- trend regression improved the ordinary multivariate regression method so that it can predict time series information. It also emphases the last three period samples' influence on the prediction. The predicted effect based on this method is therefore superior to the shifting 1 period forward regression. The predicted values of main economy indicators of Jiangxi province in 2007-2008 remain to be conformed by the statistical data in the future. Further research is required to really implement sequential multivariate regression and to predict multi-factor time series.

Acknowledgments

The authors wish to thank the Technology Bureau of Fujian Province for the project was supported by the Natural Science Foundation of Fujian Province of China (No.A0640008).

Table 2. Coefficients of regression equation of shifting dependent 1 period forward

bi	b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
value	-528.65	0	0	-8.75	0	0	0.795	0	17.787	0	0

Table 3. Comparison of estimated values and real values

Year	Real Value (Yuan)	Estimated Value (Yuan)	Error (Yuan)	Absolute Error (%)	Year	Real Value (Yuan)	EstimatedValue (Yuan)	Error (Yuan)	Absolute Error (%)
1986	652	618	-33.6	5.1	1997	4155	4172	17.0	0.4
1987	729	732	2.8	0.4	1998	4484	4576	92.4	2.1
1988	891	886	-5.4	0.6	1999	4661	4767	105.5	2.3
1989	1013	1245	232.2	22.9	2000	4851	4924	72.6	1.5
1990	1134	1267	133.0	11.7	2001	5221	5269	47.7	0.9
1991	1212	1188	-24.1	2.0	2002	5829	5836	7.0	0.1
1992	1439	1441	2.4	0.2	2003	6678	6565	-113.4	1.7
1993	1835	1690	-144.6	7.9	2004	8189	7801	-388.5	4.7
1994	2376	2264	-111.6	4.7	2005	9440	9545	105.1	1.1
1995	3083	2961	-122.0	4.0	2006	10679	10835	155.6	1.5
1996	3715	3685	-30.0	0.8	2007		13760		

Table 4. Coefficients of regression equation of shifting dependent 2 periods forward

bi	b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
value	-857	0	-5.358	-13.937	0	0	3.457	0	28.883	0	0

Table 5. Coefficients of regression equation of shifting dependent 3 periods forward

bi	b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
value	-1292	29.07			0	0.106	3.435	0		-5.63	0

Table 6. The forecast of economical indicators of Jiangxi province in 2007 and 2008

Year	AVPI	AVSI	AVTI	FTK	PK	TIFA	RSCG	CAIRH	CGDP
	100M Yuan	100M Yuan	100M Yuan	100M T-km	M P-km	100M Yuan	100M Yuan	Yuan	Yuan
2007	952	2871	1765	1057	70333	3198	1707	3854	13700
2008	1064	3408	2074	1164	78536	4067	2030	4457	16224

References

[1] Peng Nianyi, Hu Zongyi, “Practical Economy forecast and Dicision”, Central South Polytechnical University Press, Changsha, 1997.3

[2] Qian Zhongwei, Li Shengda, “Quantitative Method of Economic Forecast”, Chongqing Press, Chongqing, 1994

[3] James M.Lattin,J.Douglas Carroll,Paul E.Green, “Analyzing Multivariate Data, ” Brooks/Cole, 2003

[4] Bruce L Bowerman, Richard T O’Connell, “Forecasting and Time Series ”, Brooks/Cole, 1979

[5] Shen Jun,Ding Yuechao,Zhang Jiemin, “Some Statistic Characteristics of Economy Development in Hubei Province and the Prediction in 2005”, Modern Economy, 2005, vol. 168, pp. 70-72.

[6] Shen Jun, Ding Yuechao, “Some Statistic Characteristics of Economy Development in Fujian and its Developing Trend”, Journal of Jimei University (Philosophy and Social Science), 2006, vol. 33, pp.38-43

[7] Shen Jun, Ding Yuechao, “A New Method -- forward Shifting Regression Analysis and Its Application in Economy Forecast”, Statistics and Decision, 2006, vol. 223, pp.22-23

[8] Bureau of Statistics of Jiangxi, “Jngxi Province Statistics Yearbook 2006”, China Statistics Press, Beijing, 2007.7

[9] Bureau of Statistics of JiangXi, “Statistical Bulletin 2006 of Jiangxi Province National Economy and Society Development”, http://www.stats.gov.cn/tjgb/ndtjgb/dfndtjgb/t20070330402395521.htm, 2007-03-30