The Analysis to Tertiary-industry with ARIMAX Model

Jing Fan (Corresponding author)

Department of Science, Yanshan University

Qin Huangdao 066004, China

E-mail: fanjing850328@163.com

Rui Shan

Department of Science, Yanshan University

Qin Huangdao 066004, China

E-mail: srlhy@hotmail.com

Xiaoqin Cao & Peiliang Li

Department of Science, Yanshan University

Qin Huangdao 066004, China

Abstract

The application of multivariate time series is so large,it can be used in many systems, like ecnomic systems,biological systems, and so on.This paper introduced the method's building and the structure of ARIMAX model (auto-regressive integrated moving average model with explanatory variables) and its SAS realizing. The paper analysed the tertiary-industry in China with the realty business to be input variable and proved that there had been co-integration relationship between the two time serieses. Then, the paper modeled an appropriate ARIMAX model to tertiary-industry and fit this model with the real statistics(the tertiary-industry's production values in China from 1978 to 2007). And the result showed that ARIMAX, applied ARIMAX model to analyzing and forecasting of tertiary-industry, it is a model with high prediction precision.

Keywords: ARIMAX model, Least Squares Estimation, Tertiary-industry, Fit

1. Introduction

Recently, the trendy of tertiary-industry has been studied by many scholars with the time-series, but mostly,they used ARIMA model with one time series which was an effective method to study time series with one time series in systems.However, since the ARIMA is only with one variable, it couldn't express the relationships among the variables well in systems. The complex systems always are expressed by more than one variable in practice.According to observation or operation, the multivariate time series could be obtained, such as recording a few different variables from a studied system, like atmospheric pressure, temperatures, humidities in the meteorology; Cardiac rate, blood pressure, breathe and oxygen saturation of blood in the physiology; Rates of exchange in some currencies, sub-index in securities markets; Or in space expanded systems, more than one records could be obtained from the different space states, like overfalls, satellite data, electrocardiograms, electroencephalograms and so on. For these multivariate time serieses, they would be impacted by other variables except they have changing rules themselves. It is not possible to express the multivariate time serieses' changing rules well using ARIMA model with only one time series since the mensuration model is not enough completed. So it is necessary to model a model with multi-variables-ARIMAX model.

The analysis to multivariate time series has began early. Cox and Jenkins modeled for steady multivariate time series using ARIMA model with input variables.Technically, it is necessary to steady for input series and studied series, which is serious. And the condition limited the development of the analysis to multivariate time series. Engle and Granger put forward the concept of cointegration. Based on cointegraty theory, the regression residuals time series of input time series and output time series were needed steady only and it is not necessary to steady for themselves. Putting forward the concept of cointegration improved the development of the analysis to multivariate time series. The analysis of multi-regression and time series was integrated in the concept which improved the precision of forecasting.

In Section 1, the structure of ARIMAX model will be introduced; In Section 2, the basic thought of ARIMAX model will be raised; And the detailed steps to this model will be summaried in Section 3; The last, the simulant study to real data(tertiary-industry's production values in China from 1978 to 2007) with ARIMAX model will be studied.

2. The Structure

The model with the structure as follows were defined dynamic regreession model,simpled as ARIMAX model.

$$
\begin{cases}
y_t = \mu + \sum_{i=1}^{k} \dfrac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it} + \varepsilon_t \\
\varepsilon_t = \dfrac{\Theta(B)}{\Phi(B)} a_t
\end{cases}
\tag{1}
$$

Where, $\Phi_i(B)$ stands for the auto regression coefficients' multinomial of the I_{th} input time series; $\Theta_i(B)$ stands for the average coefficients' multinomial of the i_{th} input time series; l_i represents the lag degree of the i_{th} input variables; $\{\varepsilon_t\}$ represents the regession residule time series; $\Phi(B)$ stands for residual series's auto-regression coefficients' multinomial; stands for residule series' moving average coefficients' multinomial; and $\{a_t\}$ is white noise time-series with zero average.

3. The Thought

The basic thought of ARIMAX model is as follows: Assume both output time series $\{y_t\}$ and input time serieses $\{x_{it}\}(i = 1, 2, \cdots, k)$ are steady. Firstly,model regression model for output time series and input time serieses:

$$
y_t = \mu + \sum_{i=1}^{l} \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it} + \varepsilon_t
$$

The linear combinations of steady time-series are also steady, so the residule time-series $\{\varepsilon_t\}$ is steady because $\{y_t\}$ and $\{x_{it}\}(i = 1, 2, \cdots, k)$ are steady:

$$
\varepsilon_t = y_t - (\mu + \sum_{i=1}^{k} \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it})
$$

To pick up the interrelated information from the residule time-series $\{\varepsilon_t\}$, the model (1) will be obtained.

4. The Steps

(a) Firstly, the steady test to the output time-series $\{y_t\}$ and input time-series $\{x_{it}\}$;

(b) Secondly, the ARMA model's building for the steady input time-serieses $\{x_{it}\}$ diffed appropriately to produce white noise time-serieses $\{\varepsilon_{xit}\}$:

$$
\varepsilon_{xit} = \frac{\Theta_{xi}(B)}{\Phi_{xi}(B)} x_{(i}t) \ (i = 1, 2, \cdots, k)
$$

(c) Thirdly, the ARMA model's building for output time-series $\{y_t\}$ diffed appropriately to produce a white noise time-series $\{\varepsilon_{yit}\}$:

$$
\varepsilon_{yit} = \frac{\Theta_{xi}(B)}{\Phi_{xi}(B)} y_t \ (i = 1, 2, \cdots, k)
$$

(d) Finally, consideration to residule time-series $\{\varepsilon_t\}$, and fitting to it with appropriate model

$$
\varepsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t
$$

($\{a_t\}$ is a white noise time-series).

5. The Application

This paper takes the data of tertiary-industry's production value in China from 1978 to 2007 (from China Statistical Yearbook) for example to implement the building process of ARIMAX model and its SAS realizing.

5.1 Classification of the data

Let time-series $\{x_t\}$ and $\{y_t\}$ stand for the realty business and tertiary-industry's production value in China from 1978 to 2007 respectively. Firstly, make the sequence charts for the primary time-series $\{x_t\}$ and $\{y_t\}$ logarithmic time-series $\{\ln y_t\}$ and $\{\ln x_t\}$, one-degree diffed logarithmic time-series $\{\nabla \ln y_t\}$ and $\{\nabla \ln x_t\}$ to Fig.1.

Fig.1 shows the primary time-series and logarithmic time-series are trendy non-steady time-series, but one-degree diffed logarithmic time-series are hoping steady.

5.2 The Steady Test

The results of autocorrelation check for white noise to time-series $\{\nabla \ln y_t\}$ and $\{\triangle \ln x_t\}$ as follows:

(1) Fig. 2 shows the result of check for white noise to time-series $\{\triangle \ln x_t\}$, $\{\nabla \ln x_t\}$ is a steady non-white noise time-series,so it can be modeled.

(2) Fig. 3 shows the result of check for white noise to time-series $\{\triangle \ln y_t\}$, $\{\triangle \ln y_t\}$ is a steady non-white noise time-series,so it also can be modeled.

5.3 The Model's Building

(1) The autocorrelation check of residuals

First, model for input time-series $\{\nabla \ln x_t\}$, after ordered and tested repeatedly,determination the model as AR(1). Fig. 4 shows the parametric estimation and testation,the parameters are non-zero remarkably. The autocorrelation check of residuals shows the fitting model is appropriate. (Fig. 5)

The concrete form of fitting model as:

$$\nabla \ln x_t = 0.15536 + \frac{1}{1 - 0.47038} \varepsilon_t$$

(2) The relationship between $\{\nabla \ln x_t\}$ and $\{\nabla \ln y_t\}$

According to their cross-correlations (Fig. 6), there are only the lag 1 and -1 cross-correlation coefficient which are zero remarkably, which expresses there is 1 lag effection between input and output time-series.

(3) The order for model

According to the cross-correlations's figure and experiments,the structure of model can be obtained as follows:

$$\nabla \ln y_t = \omega \nabla \ln x_{t-1} + \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} a_t \tag{2}$$

Fitted this ARIMAX model and used least squares estimation,the parametric estimations are non-zero remarkbly (Fig. 7).

The concrete model can be obtained by least square method as follows:

$$\nabla \ln y_t = 0.90469 \nabla \ln x_{t-1} + \frac{1 + 1.80732B + 0.94382B^2}{1 + 1.71249B + 0.80396B^2} a_t$$

Or

$$\ln y_t = 0.90469 \ln x_{t-1} + \frac{1 + 1.80732B + 0.94382B^2}{(1 - B)(1 + 1.71249B + 0.80396B^2)} a_t \tag{3}$$

5.4 The effect of model fitting

Fit with ARIMAX model,the fitting result to logarithms in Fig.8.

Obviously, the fitting effect is very good.

6. Conclusion

The application of multi-time series is so large in many systems, this paper only analyzed tertiary-industry's production value in China in ecnomic systems with ARIMAX model. The tertiary-industry's production value in China would be influenced by many elements,so the fitted effects with ARIMA model were not very well for the changing rules of the tertiary-industry's total output value in China. This paper introduced the basic thought and the concrete building steps of ARIMAX model, and realized it with SAS software. Note the influence the realty business to the tertiary-industry's production value, the realty business time-series was pulled in the tertiary-industry's production value's as an input time-series, and the real data of the tertiary-industry's production value in China from 1978 to 2007 were analyzed. It can be seen that both the two time-serieses are non-steady, but the diffed logrithm time-series are steady, so the diffed log tertiary-industry's production value can be modeled for ARIMAX model. The study analyzed them to indicate there was cointegration relationship between them and then be sure that the multivariate time series' model could be built. This paper gave the steady tests, the white noise tests and the parametrical least squre estimations, then built an appropriate ARIMAX model; finally, fitted the model with figure according to that the fitted effect were well.

References

Charles S. Bos, Philip Hans Franses, Marius Ooms. (2002). Inflation, forecast intervals and long memory regression models. *International Journal of Forecasting*, 18. 243-264.

Fan, J. Q. & Yao, J. W. (2005). *Nonlinear Time Series*. Beijing: Higher Education Press, 21-33.

Gao, H. X. (1998). *SAS System*. Beijing: China Statistic Press, 65-106.

Volkan S. Ediger, Sertac Akar. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, (35). 1701-1708.

Wang, Yan. (2005). *The analysis of applied time series*. Beijing: China Renmin University Press, 208-239.

Zhang, S. Y. & Fan, Zhi. (2004). *The cointegration theory and fluctuate model*. Beijing: Tsinghua University Press, 12-30.
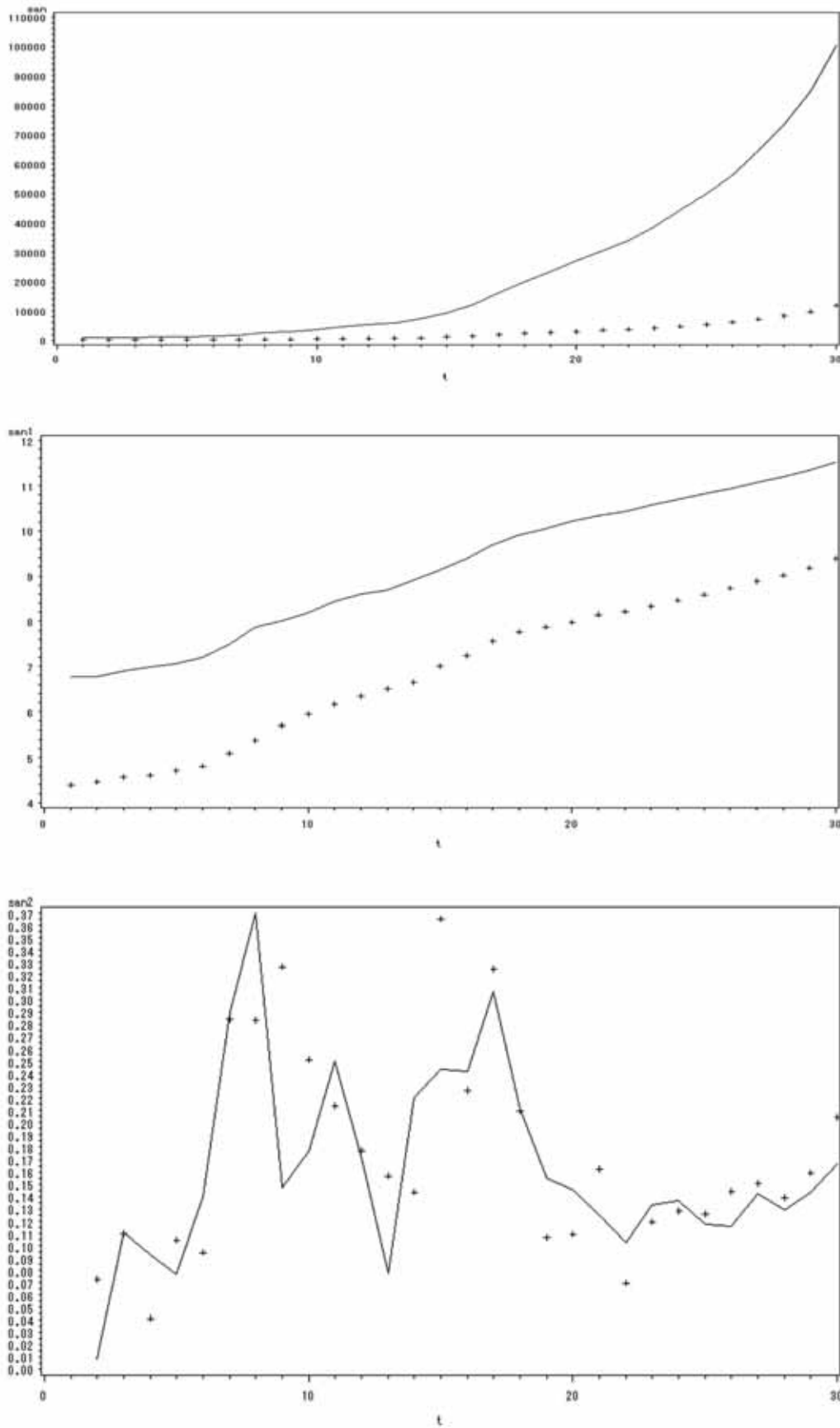
Figure 1. The sequence charts of primary time-series, logarithmic time-series and one-degree diffed logarithmic time-series

			Autocorrelation Check for White Noise						
To	Chi-		Pr >						
Lag	Square	DF	ChiSq	--------------------Autocorrelations--------------------------------					
6	13.44	6	0.0365	0.457	0.087	0.131	-0.039	-0.329	-0.214

Figure 2. The check for white noise to time-serieses {$\nabla \ln x_t$

			Autocorrelation Check for White Noise						
To	Chi-		Pr >						
Lag	Square	DF	ChiSq	--------------------Autocorrelations----------------------------					
6	17.65	6	0.0072	0.516	0.387	-0.012	-0.179	-0.279	-0.028

Figure 3. The check for white noise to time-serieses {$\nabla \ln y_t$

	The ARIMA Procedure Conditional Least Squares Estimation				
		Standard		Approx	
Parameter	Estimate	Error	t Value	Pr > |t|	Lag
MU	0.15536	0.02348	6.62	<.0001	0
AR1,1	0.47038	0.17073	2.76	0.0104	1

Figure 4. The parametric estimation to time-series {$\nabla \ln x_t$} for model AR(1)

			Autocorrelation Check of Residuals						
To	Chi-		Pr >						
Lag	Square	DF	ChiSq	--------------------Autocorrelations--------------------					
6	8.52	5	0.1300	0.068	-0.208	0.182	0.070	-0.343	-0.175
12	12.94	11	0.2973	0.227	0.008	0.077	0.059	-0.017	-0.188
18	16.00	17	0.5241	-0.073	-0.121	-0.150	0.052	0.004	-0.054
24	17.01	23	0.8086	0.030	0.040	-0.023	0.026	0.052	0.035

Figure 5. Autocorrelation Check of Residuals

		Cross correlations	
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-7	-0.0001833	-.02970	| . *| . |
-6	-0.0015387	-.24928	| . *****| . |
-5	-0.0021331	-.34558	| *******| . |
-4	-0.0005420	-.08781	| . **| . |
-3	0.00042153	0.06829	| . |* . |
-2	0.0017384	0.28163	| . |******. |
-1	0.0039997	0.64799	| . |************* |
0	0.0046255	0.74936	| . |*************** |
1	0.0029071	0.47097	| . |********* |
2	0.0014618	0.23682	| . |***** . |
3	0.00017073	0.02766	| . |* . |
4	-0.0009949	-.16118	| . ***| . |
5	-0.0009319	-.15097	| . ***| . |
6	-0.0003980	-.06447	| . *| . |
7	0.00085346	0.13827	| . |*** . |

Figure 6. The Cross-correlations between $\{\nabla \ln x_t\}$ and $\{\nabla \ln y_t\}$

		The ARIMA Procedure					
		Conditional Least Squares Estimation					
		Standard		Approx			
Parameter	Estimate	Error	t Value	Pr > |t|	Lag	Variable	Shift
MA1,1	-1.80732	0.16961	-10.66	<.0001	1	san2	0
MA1,2	-0.94382	0.16770	-5.63	<.0001	2	san2	0
AR1,1	-1.71249	0.25669	-6.67	<.0001	1	san2	0
AR1,2	-0.80396	0.25999	-3.09	0.0050	2	san2	0
NUM1	0.90469	0.05871	15.41	<.0001	0	fd2	0

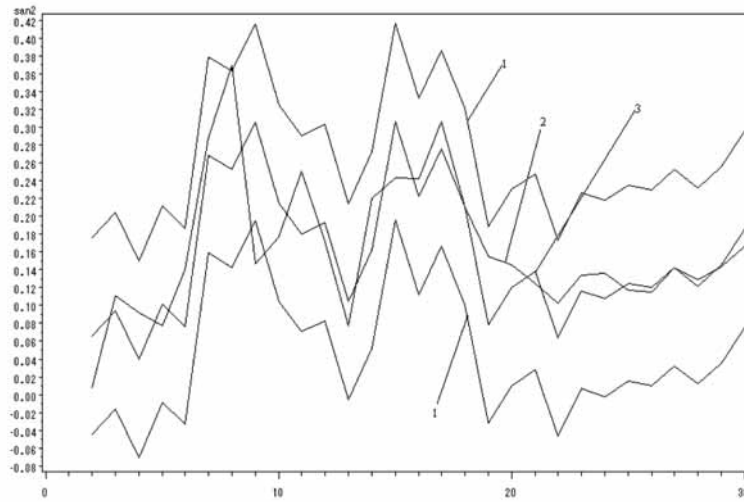Figure 7. The parametric estimations to fitted model (2)

Figure 8. The effect to fit model (3) for the logarithms

Note:1-95% confidence interval; 2-the real values; 3-predicted value