

The Research of ARMA Model in CPI Time Series

Fengwang ZHANG, Wengang CHE, Beibei XU
The School of Information Engineer and Automation
Kunming University of Science and Technology
Kunming, China
E-mail: zfw88888@163.com

Jingzhi XU
The School of Foreign Language and Literature
Yunnan Normal University
Kunming, China
E-mail: 1808907813@qq.com

Abstract—The changing trend of CPI to a certain extent reflects the degree of inflation, which has a great significance on macro-control and research on national economic. ARMA model is one of the simple and practical models in financial time series analysis with relatively high forecast accuracy. The paper utilizing Eviews software, through the statistical analysis of CPI from the year of 1995 to 2008 monthly in China, through the ADF unit root test [1], by dint of the autocorrelation function ACF diagram[2] and partial autocorrelation function PACF diagram[3] to identify the model consequently establish the model, through the residual serial correlation test of the residuals of the model to select the correct model[5]. The predications of the model showed that the ARMA model is valid and forecast accuracy is relatively high

Keywords- CPI; time series; ARMA; ACF; PACF

I. INTRODUCTION

The time series data is a set of data that a variable is arranged in accordance with the time sequence and the same time interval. Time series analysis is to make full use of these data to explore the law that things change with time going on. At present, there are a lot of theories and methods of time series forecast, such as exponential smoothing method, fitting trend method, seasonal adjustment method and so on, which are relatively accurate for the long-term forecast trend of the series, but for the short-term forecast, the effect is not very satisfactory. The establishment of the ARMA model applies the weighed of past value, the current value and lagging random disturbance model, which has taken not only the dependence on the time series but also the interference of random fluctuations into considerations, which is relatively practical to the short-term trend forecast of the time series, and also the forecasting method is relatively simple. In summary, ARMA model is simple, efficient, and with high accuracy in the short-term forecast which makes it very important in the time series forecast.

CPI (Consumer Price Index) means to reflect the trends and the extent of changes of the price level of consumer goods and services of residents in a certain period. The formation of the CPI, is to understand the basic conditions of price changes across the country, to analyze the impact of price changes on the socio-economic and residents living, to meet all the needs of all levels of governments on the formulations of policy and program and regulations of macroeconomic, as well as to offer references and bases to

national economic accounts. The paper establishes ARMA model based on CPI data from January 1995 to May 2008 in our country, and then makes a forecast of data from April to September in 2008. The rest of this paper is organized as follows, section 2 gives relevant background knowledge of ARMA model, section 3 stimulates the principle of ARMA model, section 4 takes the ARMA model which established based on CPI data as practical example to make a forecast on CPI, and section 5 concludes the paper.

II. BACKGROUND

A. Sequence Stationary

Stationary of the random time series typically refers to weak stationary. That is, for a random time series t , if it is expected value, variance, and auto-covariance mean value does not vary with time t changes; y is to be called as weak-stationary random variable. As follows Equation (1):

For all the t

$$\begin{aligned} E(y_t) &= u \\ \text{Var}(y_t) &= \sigma^2 \\ \text{Cov}(y_t, y_{t-1}) &= c \end{aligned} \quad (1)$$

B. Autocorrelation Function

To assume $\{y_t, t = 1, 2, 3, \dots\}$ is a time series, to

$$\text{call } \rho(t, s) = \frac{\text{cov}(y_t, y_s)}{\sqrt{v(y_t), v(y_s)}} \quad (t \neq s) \quad \text{as}$$

autocorrelation function of time series $\{y_t, t = 1, 2, 3, \dots\}$. Autocorrelation Function is a measure of the degree of correlation between any two elements in measure series $\{y_t, t = 1, 2, 3, \dots\}$.

C. Partial Autocorrelation Function

Partial autocorrelation function of y_t and y_{t-k} is a simple autocorrelation function after remove the indirect impact of $y_{t-1}, y_{t-2}, \dots, y_{t-k}$

Expressed as follows Equation (2):

$$\rho_k^* = \text{Corr}(y_t - E^*(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-k+1}), y_{t-k}) \quad (2)$$

D. ADF Test

The full name of ADF test is Augmented Dickey-Fuller test. The paper uses it to detect stationary of series. 3 kinds of regression models are as follows Equation (3):

$$\begin{aligned} \Delta y_t &= \gamma y_{t-1} + \sum_{i=0}^l \beta_i \Delta y_{t-i} + \varepsilon_t \\ \Delta y_t &= a + \gamma y_{t-1} + \sum_{i=0}^l \beta_i \Delta y_{t-i} + \varepsilon_t \quad (3) \\ \Delta y_t &= a + \delta t + \gamma y_{t-1} + \sum_{i=0}^l \beta_i \Delta y_{t-i} + \varepsilon_t \end{aligned}$$

Among it, L is the lag order of the dependent variable. ADF test uses T statistic for test, but the T statistic doesn't obey T distribution under the original hypothesis. Mackinnon has done a lot of simulations, which gives three different kinds of test regression test model as well as critical values of t statistic at 1%, 5%, 10% test level under different sample. If test t statistic value is less than the critical value, then reject the original hypothesis; otherwise, accept the original hypothesis.

E. In-sample Forecasts and Out-of-Sample Forecasts

Assuming the current time is T3, the known data range is T1-T3. Estimated range is from T1 to T2. The value of the dependent variable in this range is calculated based on the results of the model estimation is historical simulation, it is called In-sample Forecasts. Based on the model, calculated the dependent variable values from T2+1 to T3 is out of sample forecast.

III. ARMA MODEL

A. Principle of ARMA Model

For stationary time series, Autoregressive Moving Average Models, namely, ARMA model, with which the trend of change of variables can be well studied. In General, the ARMA (p, q) model includes an auto-regressive process AR (p) and a moving average MA (q). Showing as follows Equation (4):

$$\begin{aligned} \varepsilon_t &= c + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \phi_3 \varepsilon_{t-3} + \dots + \\ &\phi_p \varepsilon_{t-p} + \mu_t + \theta_1 \mu_{t-1} + \theta_2 \mu_{t-2} + \theta_3 \mu_{t-3} \\ &+ \dots + \theta_q \mu_{t-q} \end{aligned} \quad (4)$$

Among them, the p, q represent the lag order respectively, μ_t is white noise process.

Using lag operator to express Equation (5):

$$\begin{aligned} &(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p) \varepsilon_t \\ &= c + (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \dots \\ &+ \theta_q L^q) \mu_t \end{aligned} \quad (5)$$

From deformation we get Equation (6):

$$\begin{aligned} \varepsilon_t &= \frac{c}{1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p} \\ &+ \frac{1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \dots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p} \mu_t \end{aligned} \quad (6)$$

ARMA model can also be seen as a regression model, explanatory variables of the regression model is the lag variable of be explained variable, at the same time, the disturbance term of regression model exists q order autocorrelation. The characteristic equation of ARMA model is Equation (7):

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \dots - \phi_p z^p = 0 \quad (7)$$

Stationary condition is: the root of characteristic equation is out of unit circle, namely, the reciprocal of its root is in the unit circle. Therefore, we can see that the ARMA model is a linear combination of the white noise series, stationary of ARMA model only relate to autoregressive coefficient $(\phi_1, \phi_2, \phi_3, \dots, \phi_p)$ to make reversible

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \theta_3 z^3 + \dots + \theta_q z^q = 0$$

Condition is the root of $\theta(z)$ out of unit circle.

B. Identification of ARMA model

To build ARMA(p,q) model for stationary time series, we can get help from diagram of autocorrelation function and partial autocorrelation function of series. Along with the lag order K increases decays exponentially or sine wave attenuation AR (p) of the autocorrelation function tends to 0, showing the performance of the trailing; after lag order K>p, its partial autocorrelation function becomes 0, showing the performance of Censoring. Process MA (q) is completely different from process AR(p).

IV. EMPIRICAL ANALYSIS

The article chooses the data from January 1995 to September 2008 in our country as the observation samples, a total of 165 data, using the front 161 data to establish the ARMA model, the rest to out-of sample forecasts.

A. Series Stationary Test

The ARMA model is only applicable to stationary sequence, so we have to do stationary test on CPI series before we actually establish the model. This paper uses ADF unit root to test.

- (1) From the observation of CPI graphics, we can define the form of the unit root test regression equations, as shown in the CPI line graph in the Fig 1.

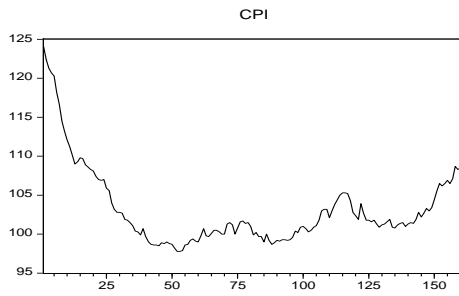


Figure 1. CPI line graph.

From Fig 1, it is noted that series CPI showed no change over time trends, Series deviation from the zero value and random fluctuations, So the test equation contains the intercept term, no time trend[6].

(2)The results of ADF unit root test on the series CP, As shown in the following Fig2

Augmented Dickey-Fuller Unit Root Test on CPI		
Null Hypothesis: CPI has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=13)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.363639	0.0000
Test critical values: 1% level	-3.471454	
5% level	-2.879494	
10% level	-2.576422	

Figure 2. Unit root test.

It is seen from the Fig2, t Statistics of ADF test equals to -6.363639, the absolute value of t Statistics is greater than the absolute value of 1%, 5% level and 10% level, so rejecting the original hypothesis, that is, there is no unit root of CPI series, it is a stationary series.

B. Order Determination of ARMA Model

Inspection from Fig 2 CPI series is stationary, so we can make use of the autocorrelation function and the partial autocorrelation function to determine the order of the model. AC and PAC of CPI series as shown in Fig 3:

Correlogram of CPI					
Date: 11/17/12 Time: 16:41					
Sample: 1 161					
Included observations: 161					
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.935	0.935	143.30	0.000	
2	0.871	0.019	298.61	0.000	
3	0.810	-0.018	377.57	0.000	
4	0.748	0.037	471.13	0.000	
5	0.685	-0.047	550.04	0.000	
6	0.625	0.014	616.85	0.000	
7	0.575	0.008	673.13	0.000	
8	0.529	0.033	721.22	0.000	
9	0.488	-0.002	752.34	0.000	
10	0.449	0.009	797.37	0.000	
11	0.413	0.004	827.16	0.000	
12	0.380	0.000	852.54	0.000	
13	0.357	0.001	876.11	0.000	
14	0.332	-0.027	894.81	0.000	
15	0.301	0.063	911.14	0.000	
16	0.268	-0.044	924.14	0.000	
17	0.235	0.001	934.43	0.000	
18	0.207	0.012	942.31	0.000	
19	0.177	-0.013	948.12	0.000	
20	0.147	0.022	952.15	0.000	
21	0.120	-0.007	954.83	0.000	
22	0.093	-0.021	956.46	0.000	
23	0.065	-0.033	957.26	0.000	
24	0.035	0.035	957.59	0.000	
25	0.012	0.014	957.53	0.000	

Figure 3. AC and PAC.

From Fig 3, it is seen that autocorrelation function of series CPI begins to small after lag 21 orders, which illustrates the moving average process MA should be low-level, partial autocorrelation function censored after lag 1 order. Therefore, we initially established ARMA(1,1)model.

C. ARMA Model Estimations

This paper attempts to use 161 data of sample to do the estimation. The model estimation results are shown in Fig 4:

Dependent Variable: CPI				
Method: Least Squares				
Date: 11/17/12 Time: 16:58				
Sample (adjusted): 2 161				
Included observations: 160 after adjustments				
Convergence achieved after 8 iterations				
MA Backcast: 1				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	101.5392	0.927782	109.4429	0.0000
AR(1)	0.936362	0.011136	84.08568	0.0000
MA(1)	0.189124	0.079198	2.387993	0.0181
R-squared	0.984585	Mean dependent var	103.0606	
Adjusted R-squared	0.984389	S.D. dependent var	4.783450	
S.E. of regression	0.597671	Akaike info criterion	1.827020	
Sum squared resid	56.08210	Schwarz criterion	1.884680	
Log likelihood	-143.1616	Hannan-Quinn criter.	1.850434	
F-statistic	5013.929	Durbin-Watson stat	1.963858	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.94			
Inverted MA Roots	-.19			

Figure 4. Estimation results.

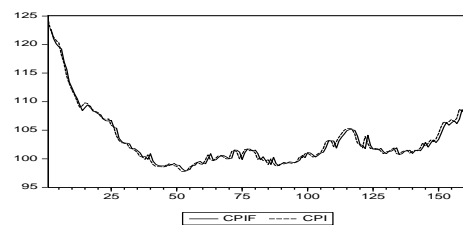


Figure 5. In-sample Forecasts.

From Fig 4, the goodness of fit of the model estimation results $R^2 = 0.984585$, $\bar{R}^2 = 0.984389$ F statistic = 5013.929, the corresponding probability value is 0. All the previous data shows that the model is salient on the whole and its fitting effect is pretty good. AR part of ARMA(1,1) modal has a Inverted AR Roots, that is 0.94, and its modal is less than 1; MA part also has a Inverted MA Roots, that is -0.19 and whose modal is less than 1. Thus, it is considered that ARMA(1,1)modal is stable and reversible. The estimated result of ARMA(1,1)model achieved by Lag polynomial:

$$(1 - 0.936362L)CPI_t = 101.5392 + (1 + 0.189124L)\epsilon_t$$

$$R^2 = 0.984585$$

$$\bar{R}^2 = 0.984389$$

$$AIC \text{ principle} = 1.827020$$

$$SC \text{ principle} = 1.884680$$

The forecast of 161 data In-sample forecast as the Fig 5, among which CPI refers to original data and CPIF to the data in-sample forecast. The effects of fit in ARMA(1,1)model is pretty good.

D. Test for ARMA Mode in the Residual Serial Correlation

After the estimation of ARMA model, it is necessary to test whether the model is correct. Here is a LM test of ARMA(1,1)in residual series.

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	1.022615	Prob. F(4,153)	0.3975	
Obs*R-squared	4.166188	Prob. Chi-Square(4)	0.3840	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 11/19/12 Time: 16:02				
Sample: 2 161				
Included observations: 160				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.102398	0.930949	0.109993	0.9126
AR(1)	-0.001422	0.012091	-0.117635	0.9065
MA(1)	2.865570	5.490504	0.521914	0.6025
RESID(-1)	-2.948914	5.495159	-0.534411	0.6049

Figure 6. LM test.

Fig 6 shows the results of ARMA (1,1) model in Residual serial correlation LM test, which clearly shows that the test statistics is Obs*R-squared=4.166188, and its corresponding probability value is P=0.3840. Thus, the original hypothesis that reception of the residual sequence correlation holds that ARMA(1,1) model is effective.

E. ARMA Model Forecast

We make out-of sample forecasts of the CPI based on the established ARMA (1, 1) model from June to September in 2008. As shown in table I, from the table, we can see that the relative error is relatively small between the forecast value and the real value of the series CPI, the forecast effect of the model is quite well, but with the increase in the forecast period, the relative error of model forecast is also increasing, the short-term forecast of the ARMA model is still relatively ideal.

TABLE I. OUT-OF SAMPLE FORECASTS

	June 2008	July 2008	August 2008	September 2008
True value	107.1000	106.3000	104.9000	104.6000
Predictive Value	107.2160	106.7242	105.9168	104.4938
Relative error	0.1083%	0.3991%	0.9693%	0.1015%

V. CONCLUSION

In this paper, for CPI time series we do identification, estimation, residual correlation test and make forecast with the method of Box-Jenkins. Based on the ARMA model of

previous 161, and data forecast of the four from 162nd to 165th. Conclude that an appropriate model should meet the following conditions: concise, its coefficient shows to be stable and reversible can be better fitted to the data, there is no residual autocorrelation, relatively good out-of sample forecast and so on. ARMA model, established in this paper, the effects are relatively satisfactory and the relative error of short-term forecasts is relatively small. But we know that the ARMA model is only suitable for stable series, for unstable series, we can use the different approach and other methods to stabilize the series, but we also know that it cannot be over differentiated, otherwise it will lose a lot of important information of the data itself, then there is no point in establishing the ARMA model; sometimes it is also more difficult to determine the order of model, and some model's goodness of fit may be not good, but the effect of out-of sample forecasts is better, so for the selection of the order of the ARMA model we should grope our way in practice. The short-term forecast of ARAM model is still relative ideal, to obtain good results from the long-term forecast still needs further studies.

REFERENCES

- [1] Hongyu Pan. Financial Time Series Model. Beijing: University of International Business and Economics Press, 2008.
- [2] Zhongjie Xie. Analysis of Time Series. Beijing: Peking University Press, 1990.9.
- [3] Xuejiang Chang. Analysis of Time Series. Beijing: Higher Education Press, 1993.11.
- [4] Desi En, Jiang Du, Zhichao Xie. Applied Econometrics Analysis of Time Series. Beijing: Higher Education Press, 2006.6.
- [5] Danhui Yi. Analysis of Time Series : Method and Apply. Beijing: China Renmin University Press, 2011.3.
- [6] Yinhua Ren. EVIEWS Applied Experimental Tutorial. Changsha: Hunan University Press, 2008.7.
- [7] Jisi Zhan, Chenzhang Wang. Analysis of Time Series Forecasting and Control 4 edition. Beijing: China Machine Press, 2011.4.
- [8] Cryer Chan, Hongyu Pan. Analysis and apply of Time Series. Beijing: China Machine Press, 2011.1.
- [9] Chensi Zhang. Time Series Analysis Perspective. Beijing: China Renmin University Press, 2012.1.
- [10] Mingyue Guo, Zhihong Xiao. Analysis of Time Series and SAS Apply. Wuhan: Wuhan University Press, 2012.2.