Journal of Quantitative Finance and Economics. **2019**, 1, 2

ARF INDIA
Academic Open Access Publishing
www. arfjournals. com

A Comparative Study of Different Ordered Criterions in Time Series

Tianzhen Wang, Linfan Lin and Guangbao Guo

School of Mathematics and Statistics, Shandong University of Technology, Zibo 255049, China

Received: 15 March 2019; Revised: 25 May 2019; Accepted: 15 July 2019; Publication: 15 October 2019

Abstract: Model order determination is a common problem in time series modeling. Due to the randomness of the observed values and the variability of the parameters under different orders, the problem of order determination becomes complicated. The effective order determination must consider both the goodness of model fitting and the degree of model simplification. This paper mainly introduces several commonly used order criterion. By using Python software data simulation is used to analysis and compare the order accuracy of different criterion under different sample sizes. The simulation results show that (1) When the sample size is small, the accuracy of an improved AIC criterion is obviously better than AIC criterion and BIC criterion. (2) With the sample increasing, BIC criterion's order accuracy converges to 1 gradually according to probability, while AIC criterion and new AIC criterion tend to fit high-order model, which cannot give the consistent estimation of the real order of the model.

Keywords: time series; model order determination; Python simulation; order accuracy

1. Introduction

After determining the appropriate model type for the time series, it is necessary to determine the order of the model. Although modern computing tools are very developed, when the model order is too large, it will not only make the calculation of model parameters tedious, but also increase the error of the model due to too many iterations of model parameters. Therefore, correct and effective order determination is the key to time series modeling. Scholars at home and abroad have conducted different researches on the model order determination. In 1967, Professor k.j. Astrom[1], a Swedish cybernetic expert, used the F test criterion to determine the order of the time series model, but the selection of the confidence of the F test was subjective to some extent. In 1973, Akaike[2], a Japanese scholar, put forward Akaike information criterion based on the maximum information amount of sample observation values. Akaike Information Criterion provides a standard to weigh the goodness of model fitting against the complexity of model, but Akaike Information Criterion cannot give a consistent estimate. To make up for the deficiency of Akaike Information Criterion, Akaike proposed Bayesian Information Criterion in 1978[3]. In 2002, Chinese scholar Lou feng put forward an improved T-F test criterion [4] by combining the F test and T test method, and carried out example analysis and comparative study.

AR model (auto regression model) is one of the simplest and most commonly used models in time series. Kolmogorov once pointed out that any ARMA or MA process can be represented by an infinite order AR process[5]. This paper will mainly discuss the order determination of AR(p) in auto regressive model.

2. Preliminaries

2.1. AR (p) model

AR model is the most commonly used model for fitting stationary series, in which AR (p) model is the regression model with the dependent variable lagging p period. We call that the mathematical model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

is the AR (p) model.

Where the model assumptions are

1) $\phi_p \neq 0$

2) $\{\varepsilon_t\} = NID\ (0,\ \sigma_\varepsilon^2)$

3) $Cov(X_t, \varepsilon_s) = 0,\ \forall s > t$

Since any time series subject to stationarity and normality can be fitted with AR(p) model, the following discussion will focus on the order determination of the AR(p) model.

2.2. The order determination criterion and parameter estimation method

2.2.1 AIC criterion

In 1973, Japanese scholar Akaike put forward Akaike Information Criterion(written briefly as AIC), which comprehensively considered the likelihood function to measure goodness of fit and the number of parameters to measure model simplicity [6] . A good fitting model should be a combination of goodness of fit and model simplicity.

AIC function is a weighted function for fitting goodness and number of parameters, which is defined as:

$$AIC(p) = \ln[\hat{\sigma}_a^2] + \frac{2p}{N}$$

Where N is the number of observed values of time series samples, and $\hat{\sigma}_a^2$ is the maximum likelihood estimation of residual variance. If

$AIC(p_0) = \min_{1 \le p \le Max(N)} AIC(p)$, we take p_0 as the optimal model order. Max (N) is usually an integer between [N/3] and [2N/3]

2.2.2. BIC criterion

In 1978, Schwarz proved on the basis of Bayes theory that Akaike Information Criterion could not give the model's real order consistent estimation [7], that is, when the sample size tends to infinity, the model fitted by AIC criterion does not converge to the real model. In order to make up for the deficiency of AIC Criterion, Akaike put forward Bayesian Information Criterion (written briefly as BIC)in 1978. The BIC function is defined as:

$$BUC(p) = \ln[\hat{\sigma}_a^2] + \frac{p \ln N}{N}$$

Where N is the number of observed values of time series samples, and $\hat{\sigma}_a^2$ is the maximum likelihood estimation of residual variance. If $BIC(p_0) = \min_{1 \le p \le Max(N)} BIC(p)$, we take p_0' as the optimal model order. Max (N) is usually an integer between [N/3] and [2N/3].

Compared with AIC function, the second term of BIC function replaces the coefficient 2 with lnN. The Bayesian Information Criterion increased the penalty term, which can effectively avoid the disaster phenomenon of dimensionality.

2.2.3. A new AIC criterion

When the sample size is small, there will be a large deviation in order determination using AIC and other criterion. Hurvich and Tsai proposed a new AIC criterion with less error than AIC criterion[8], but it is not completely unbiased. The AIC function is

$$AICc(p) = \ln[\hat{\sigma}_a^2] + \frac{N + p}{N - p - 2}$$

Where N is the number of observed values of time series samples, and $\hat{\sigma}_a^2$ is the maximum likelihood estimation of residual variance. If $AICc(p_0^*) = \min_{1 \le p \le Max(N)} AICc(p)$, we take $p^*{}_0$ as the optimal model order. Max (N) is usually an integer between [N/3] and [2N/3].

The penalty term of the new AIC function is stronger than AIC function to avoid overfitting. Burnham and Anderson pointed out that with the increase of sample size , The new function will converge to AIC function, and the new criterion is obviously superior to AIC criterion when model order determination is conducted in the case of small samples [9].

2.2.4. Parameter estimation method

When calculating the criterion function values of AIC, BIC and the new AIC (AICc), unknown parameters $\left(\phi_1, \phi_2 ... \phi_p, \sigma_a^2\right)$ in the model need to be estimated. The main parameter estimation methods are moment estimation, least square estimation and maximum likelihood function estimation. In this paper, the conditional maximum likelihood function estimation method is used in data simulation. The maximum likelihood function estimation principle of AR (p) is briefly introduced below.

Set X_1, X_2, ... X_T is a sample observation value of AR (p) sequence, and the first p observations in the sample are synthesized into a (p × 1) vector, which can be regarded as an implementation of normal random vector of p-dimension, so $(X_1, X_2, ... X_p) < N(0, \sigma^2 V_p)$. where $\sigma^2 V_p$ is the $p \times p$ covariance matrix of 1 2 $(X_1, X_2, ... X_p)$

$$\sigma^2 V_p = \begin{pmatrix} EX_1^2 & EX_1X_2 & \cdots & EX_1X_p \\ EX_2X_1 & EX_2^2 & \cdots & EX_2X_p \\ \cdots & \cdots & \cdots & \cdots \\ EX_pX_1 & EX_pX_2 & \cdots & EX_p^2 \end{pmatrix}$$

Then the probability density function of $(X_1, X_2, ... X_p)$ is

$$f(X_1, X_2, ... X_p; \theta) = (2\pi)^{-p/2} \left|\sigma^{-2} V_p^{-1}\right|^{1/2} \exp\left[-\frac{1}{2\sigma^2} X_p' V_p^{-1} X_p\right]$$

where $X_p = (X_1, X_2, ... X_p)$, $\theta = (\phi_1, \phi_2 ... \phi_p, \sigma^2)$

The likelihood function of $(X_1, X_2, ... X_T)$ is

$$f(X_1, X_2, ... X_T; \theta) = f(X_1, X_2, ... X_p;\, \theta) \prod_{t=p+1}^{T} f_{X_t|X_{t-1} ..., X_{t-p}}(X_t \mid X_{t-1}, ..., X_{t-p}; \theta)$$

$$= (2\pi)^{-T/2} (\sigma^2)^{-T/2} \left|V_p^{-1}\right|^{1/2} \exp\left[-\frac{1}{2\sigma^2} X_p' V_p^{-1} X_p\right]$$

$$\exp\left[-\frac{1}{2\sigma^2} \sum_{t=p+1}^{T} (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_p X_{t-p})^2\right]$$

The logarithmic likelihood function is

$$L(\theta) = [-T/2]\log(2\pi) + [-T/2]\log(\sigma^2) + 1/2\log\left|V_p^{-1}\right| - \frac{1}{2\sigma^2}X_p'V_p^{-1}X_p$$

$$-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T}(X_t - \phi_1X_{t-1} - \phi_2X_{t-2} - \dots - \phi_pX_{t-p})^2$$

Similarly, we can also discuss the conditional maximum likelihood estimation of θ. The logarithmic likelihood function can be written as:

$$f_{X_{p+1},\dots X_T|X_p,\dots X_1}(X_{p+1}\dots X_T \mid X_p, \dots X_1;\theta)$$

$$\prod_{t=p+1}^{T} f_{X_t|X_{t-1},X_{t-p}}(X_i \mid X_{t-1},\dots,X_{t-p};\theta)$$

$$= -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{T}(X_t - \phi_1X_{t-1} - \phi_2X_{t-2} - \dots - \phi_pX_{t-p})^2$$

As can be seen from the above equation, the logarithmic likelihood function is maximized, that is, the value of $\sum_{t=p+1}^{T}(X_t - \phi_1X_{t-1} - \phi_2X_{t-2} - \cdots - \phi_pX_{t-p})^2$ is minimized. Therefore, the least squares estimation of X_t for its p lagged term regression model is the conditional maximum likelihood estimation of parameters $\phi_1, \phi_2 \dots \phi_p$. The conditional maximum likelihood estimation of white noise variance is

$$\hat{\sigma}^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}(X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \cdots - \hat{\phi}_p X_{t-p})^2$$

Where $\hat{\phi}_1, \hat{\phi}_2 \dots \hat{\phi}_p$ is the conditional maximum likelihood estimation of parameter $\phi_1, \phi_2 \dots \phi_p$.

3. Data simulation and analysis

Here, AR (2) model is used to compare the order determination effects of different criterion under different sample sizes. It is assumed that the model is $X_t = -0.6X_{t-1} + 0.4X_{t-2} + \varepsilon_t$. Where $\{\varepsilon_i,\ t = 1, 2, 3, \dots\}$ is the sequence of normal white notise. Python was used to simulate and repeat 1000 times for different sample sizes. The criterion function values of AIC, BIC and

AICc were calculated to determine the order of the fitting model, and then compared with the actual model [10], so as to verify the order accuracy of different criterion. The specific processes are as follows.

1) Give the initial value, let $X_1 = X_2 = 0.5$.
2) Generate $n + 20$ ($n = 15, 30, 60, 120, 200$) normal random Numbers through Python.
3) Through the initial value and the generated $n + 20$ normal random numbers, $n + 20$ X_t are generated according to the known model.
4) In order to ensure the randomness of X_t and ensure that is not affected by the initial value, the last n are taken as the real simulated data of the model.
5) For the obtained X_t, the parameter estimation of $\phi_1, \phi_2 \ldots \phi_p, \sigma^2_a$ is obtained by using the conditional maximum likelihood estimation.
6) For the obtained X_t, use AR(1), AR(2),..., AR(5) fitting, respectively calculate AIC, BIC, AICc criterion function value, find the minimum criterion function value, then determine the best order of the model.
7) Repeat the above steps 1000 times and record the times of positive order of different criterion.

The ordering results of different criterion are shown in Table 1.

Table 1: The order results of different criterion

Sample size	$\hat{p}$	*1*	*2*	*3*	*4*	*5*	*Order accuracy %*
$n = 15$	AIC	2	930	52	14	2	93.0%
	BIC	3	958	31	6	2	95.8%
	AICc	4	991	5	0	0	99.1%
n=60	AIC	0	886	74	28	12	88.6%
	BIC	0	976	18	5	1	97.6%
	AICc	0	982	14	4	0	98.2%
n=120	AIC	0	888	82	21	9	88.8%
	BIC	0	982	16	1	1	98.2%
	AICc	0	968	26	5	1	96.8%
n=200	AIC	0	879	77	30	14	87.9%
	BIC	0	993	7	0	0	99.3%
	AICc	0	963	34	3	0	96.3%

As can be seen from table 1, when the sample size is 15, the order accuracy of AIC and BIC criterion is 93.0% and 95.8% respectively, while the order accuracy of AICc criterion is 99.1% and there is no serious overfitting phenomenon, which is obviously better than AIC and BIC criterion. Therefore, when the sample size is small, AICc criterion can be selected as a priority. When the sample size was 60, 120 and 200, the order accuracy of BIC criterion reached 97.6%, 98.2% and 99.3% respectively. It can be concluded that if the sample size is large enough, the order accuracy of BIC criterion will converge to 1 according to the probability.

No matter what the sample size is, the AIC criterion always has the phenomenon of overfitting. When the sample size was 15, 60, 120 and 200, 68 times, 114 times, 112 times and 121 times were overfitted respectively. The optimal model order was also relatively dispersed, but the improved AICc criterion avoided overfitting in the case of small sample size. However, with the increase of sample size, the probability of AICc criterion overfitting is also increasing. In contrast to BIC criterion, the probability of over-fitting decreases with the increase of sample size. It is conceivable that as N approaches infinity, the optimal order selected using AIC and AICc criterion is often higher than the real model. Since N plays a smoothing role in BIC criterion function discriminant calculation formula, BIC criterion overcomes the disadvantages of AIC criterion, and the determined optimal order is often consistent with the real order.

4. Conclusion

In this paper, several commonly used time series order determination criterion are introduced, and a large number of data simulation and analysis are carried out to compare the order accuracy of different criterion in different sample sizes. None of the criterion is completely better than the other criterion. Therefore, factors such as the size of the sample size and the significance of the autoregressive coefficient should be taken into account in the model order determination, so as to select the appropriate criterion function and determine the model order correctly and effectively.

References

[1] Wu Huai-yu. Time Series Analysis and Synthesis [M]. Wuhan University Press, 2004:82-96.

[2] AKAIKE H. A new look at statistical model identification [J]. IEEE Trans,1974(19):716-723.

[3] Wang Zhen-long, Hu Yong-hong. Applied Time Series Analysis [M]. Science Press, 2007:92-98.

[4] Lou Feng, Yang Qing etc. An improved method for time series AR ordering criterion [C]. Proceedings of the 7th Academic Conference of the Chinese Society of Rock Mechanics and Engineering. ,2002:354-358.

[5] Lu Guang-hua, Peng Xue-yu. Random Signal Processing [M]. Xi 'an University of Electronic Science and Technology Press, 2002.

[6] Wang Yan. Applied Time Series Analysis [M]. China Renmin University Press, 2005:81-84.

[7] SHIBATA R. Selection of the order of an autoregressive model by Akaike´s information criterion [J]ÿBiometrika,1976,63:117-1260.

[8] HURVICH C M, SHUMWAY R, TSAI C L . Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples [J], Biometrika, 1990(77), 709—719.

[9] ZHAO BOJUAN, WU XIZHI. Another way to see criterion AICc [J]. *Journal of Nankai University*, 1998, 31, 4: 43-49.

[10] Song Tian. Python Language Programming Basics [M]. Higher Education Press, 2014.