Hacettepe Journal of Mathematics and Statistics
Volume 40 (4) (2011), 601 – 607

INFORMATION COMPLEXITY CRITERION FOR ORDER DETERMINATION IN AUTOREGRESSIVE MODELS

Barış Aşıkgil*

Received 29 : 04 : 2010 : Accepted 21 : 12 : 2010

Abstract

Autoregressive models which include lags of the dependent variable are usually used in time series analysis. The corelogram of the series and some information criteria can be used in order to determine the order of these models. In this paper the information complexity criterion is considered for autoregressive time series models. A simulation study is performed in order to examine the performance of information complexity and compare it with some ordinary information criteria.

Keywords: Autoregressive models, Information complexity, Order determination, Simulation.

2000 AMS Classification: 62 M 10.

1. Introduction

The autoregressive (AR) models have been widely used as a valuable tool to fit a variety of practical data in several different areas, such as statistical time series, geophysics and signal processing. For a given set of observations $\{y_t; t = 1, 2, \dots, n\}$ of a zero-mean stationary stochastic process, the AR model is defined by

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{1}$$

where ϕ_i $(i = 1, 2, \dots, p)$ denotes the unknown parameters, ε_t denotes the white noise error term with mean zero and variance σ^2, p is the number of parameters and also the model order [10].

There are various methods used in order to estimate the parameters of the model given in (1). The best known are estimating the parameters by using the well-known Yule-Walker equations, and Burg's [6] method. The steps are given below to estimate the parameters of the AR model by using the Yule-Walker equations [8,9]:

*Department of Statistics, Faculty of Science and Letters, Mimar Sinan Fine Arts University, Çırağan Caddesi, Çidem Sokak No: 1, Beşiktaş 34349, Istanbul, Turkey.
E-mail: basikgil@msgsu.edu.tr

- By multiplying both sides of the model given in (1) by $y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-p}$ separately and recursively, then taking the expectance of each new model, a system of equations is obtained and written in matrix form,

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} c_0 & c_1 & \cdots & c_{p-1} \\ c_1 & c_0 & \cdots & c_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p-1} & c_{p-2} & \cdots & c_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} \quad \text{or} \quad \mathbf{c} = \mathbf{C}\boldsymbol{\phi}, \tag{2}$$

where the autocovariances in the first row of $\mathbf{C}$ are given as

$$\begin{aligned} &E\left(y_{t-1}y_{t-1}\right) = c_0, E\left(y_{t-1}y_{t-2}\right) = c_1, \ldots, E\left(y_{t-1}y_{t-p}\right) = c_{p-1}, \\ &E\left(y_{t-1}\varepsilon_t\right) = 0. \end{aligned} \tag{3}$$

- Equations (2) are known as the Yule-Walker equations, and the vector of estimated parameters is obtained by

$$\hat{\boldsymbol{\phi}} = \mathbf{C}^{-1}\mathbf{c}, \tag{4}$$

where $\mathbf{C}$ is a full-ranked symmetric matrix. Moreover, the autocorrelation values can also be used instead of the autocovariances in (2) in order to obtain the parameter estimates [7].

By considering the first step giving (2) only for y_t and evaluating (4), the variance estimator can be defined by

$$\hat{\sigma}^2 = c_0 - \hat{\boldsymbol{\phi}}'\mathbf{c}. \tag{5}$$

In this study some information criteria are considered for autoregressive time series models by using the findings obtained from the Yule-Walker equations. The rest of this article is organized as follows. In section 2 some ordinary and recently popular information criteria are described and viewed in terms of order determination for autoregressive models. In section 3 a simulation study is performed in order to examine and compare the performance of the information criteria. Finally, a conclusion is given containing a short discussion.

2. Order determination for AR models

The problems of model selection in regression and order determination for AR models are closely related. Therefore, besides ordinary corelograms, model selection criteria can also be used to determine the model order in AR processes. Two of the leading criteria are the Akaike information criterion and the information complexity criterion, which is the primary focus of this paper.

2.1. Akaike information criterion. The Akaike information criterion (AIC) [1] was designed to be an approximately unbiased estimator of the expected Kullback-Leibler (KL) information of a fitted model. KL information can be used as a measure of goodness of fit. The rationale of Akaike's concept of choosing the best approximating model from finite samples can be formulated as maximizing entropy, or equivalently as minimizing KL information [4].

Considering a probability density function $f(\mathbf{y}|\boldsymbol{\phi}^*)$ of the continuous random variable $\mathbf{y}$ of interest, and $f\left(\mathbf{y}|\boldsymbol{\phi}\right) \equiv g\left(\mathbf{y}|\boldsymbol{\phi}\right)$ is the density function that specifies the model or is an approximation to $f\left(\mathbf{y}|\boldsymbol{\phi}^*\right)$ with a given parameter vector $\boldsymbol{\phi} \equiv \phi_p = (\phi_1, \phi_2, \ldots, \phi_p) \in R^p$. The goodness of fit of $f\left(\mathbf{y}|\boldsymbol{\phi}^*\right)$ with respect to the model $f\left(\mathbf{y}|\boldsymbol{\phi}\right) \equiv g\left(\mathbf{y}|\boldsymbol{\phi}\right)$ is measured by the generalized entropy (B) or KL information (I) given in [4].

The original AIC takes the form of a penalized likelihood (a negative log likelihood plus a penalty term),

$$\text{AIC} = -2\log L\big(\hat{\boldsymbol{\phi}}\big) + 2p, \tag{6}$$

where $L\left(\hat{\boldsymbol{\phi}}\right)$ is the maximized likelihood function and p is the number of free parameters in the model. The model with a minimum AIC value is chosen as the best model to fit the data [2,3]. In a similar manner, under the assumption of a Gaussian model, the AIC statistic for the p^{th} order autoregressive model can be defined by

$$\text{AIC} = n\big(\log\hat{\sigma}^2 + 1\big) + 2\,(p+1)\,, \tag{7}$$

where $\hat{\sigma}^2$ is given in (5).

Moreover, a corrected version of AIC (CAIC) is defined by

$$\text{CAIC} = \text{AIC} + \frac{2\,(p+1)\,(p+2)}{n-p-2}. \tag{8}$$

As given in (8), CAIC is the sum of AIC and an additional nonstochastic penalty term [11].

2.2. Information complexity criterion. The information complexity criterion (ICOMP) is related to AIC. However, ICOMP is based on the structural complexity of an element or set of random vectors via a generalization of the information-based covariance complexity index [3].

ICOMP is designed to estimate a loss function (lack of fit plus lack of parsimony plus profusion of complexity). ICOMP, penalizing the covariance complexity of the model, is defined by

$$\text{ICOMP} = -2\log \mathrm{L}\big(\hat{\boldsymbol{\phi}}\big) + 2\mathsf{C}\big(\widehat{\boldsymbol{\Sigma}}_{\text{model}}\big), \tag{9}$$

where $\mathrm{L}\big(\hat{\boldsymbol{\phi}}\big)$ is the maximized likelihood function, C denotes a real-valued complexity measure and $\mathrm{Cov}\big(\hat{\boldsymbol{\phi}}\big) = \widehat{\boldsymbol{\Sigma}}_{\text{model}}$ denotes the estimated covariance matrix of the parameter vector of the model [5].

ICOMP can be considered as an approximation to the sum of two KL distances. The new model selection criterion is called ICOMP(IFIM) and it resembles a penalized likelihood method similar to AIC except that the penalty depends on the curvature of the log likelihood function via the scalar complexity value $\mathsf{C}_1(\cdot)$ of the estimated inverse Fisher information matrix (IFIM). ICOMP(IFIM) is defined by

$$\text{ICOMP(IFIM)} = -2\log \mathrm{L}\big(\hat{\boldsymbol{\phi}}\big) + 2\mathsf{C}_1\big(\hat{F}^{-1}\big(\hat{\boldsymbol{\phi}}\big)\big), \tag{10}$$

where C_1 denotes the maximal informational complexity of $\hat{F}^{-1}$. The first component of ICOMP(IFIM) in (10) measures the lack of fit of the model and the second component measures the complexity of the estimated IFIM [5].

Considering ICOMP given in (9), ICOMP for multiple linear regression models is defined in [5]. In a similar manner, ICOMP for the AR model defined in (1) can be obtained as

$$\begin{aligned}\text{ICOMP}_{\text{AR}} &= n\log\left(2\pi\right) + n\log\left(\hat{\sigma}^2\right) + n \\ &\quad + 2\left[\frac{p}{2}\log\left(\frac{\text{trace}\left(\hat{\sigma}^2\mathbf{C}^{-1}\right)}{p}\right) - \frac{1}{2}\log\left|\hat{\sigma}^2\mathbf{C}^{-1}\right|\right],\end{aligned} \tag{11}$$

where $\hat{\sigma}^2$ is given in (5) and $\mathbf{C}$ is defined by the Yule-Walker equations in (2).

Considering ICOMP(IFIM) given by (10), ICOMP(IFIM) for multiple linear regression models is defined in [5]. In a similar manner, ICOMP(IFIM) for the AR model defined in (1) can be obtained as

$$\text{ICOMP(IFIM)}_{\text{AR}} = n \log(2\pi) + n \log(\hat{\sigma}^2) + n + \mathtt{C}_1(\hat{F}^{-1}(\hat{\phi})), \tag{12}$$

where

$$\mathtt{C}_1(\hat{F}^{-1}(\hat{\phi})) = (p+1)\log\left[\frac{\operatorname{trace}(\hat{\sigma}^2\mathbf{C}^{-1}) + \frac{2\hat{\sigma}^4}{n}}{p+1}\right] - \log\left|\hat{\sigma}^2\mathbf{C}^{-1}\right| - \log\left(\frac{2\hat{\sigma}^4}{n}\right), \tag{13}$$

$\hat{\sigma}^2$ is given by (5) and $\mathbf{C}$ is defined by the Yule-Walker equations in (2).

In this paper it is considered as a contribution that a penalty term 2(p+1) given in (7) can be added to the information complexity criteria in order to indicate the usual differences among model orders and determine the true model order. These modified criteria are called penalized ICOMP (ICOMP_PN) and penalized ICOMP(IFIM) (ICOMP(IFIM)_PN), respectively.

3. Monte Carlo simulation

In this section a simulation study is considered in order to observe and compare the performances of the various information criteria by determining the autoregressive model order. The conditions under which the simulation was performed are as follows:

(1) Four different wide-sense stationary models are considered in this study:
- AR(1) Model: $y_t = 0.9y_{t-1} + \varepsilon_t$,
- AR(2) Model: $y_t = 0.99y_{t-1} - 0.8y_{t-2} + \varepsilon_t$,
- AR(3) Model: $y_t = 1.42y_{t-1} - 1.31y_{t-2} + 0.56y_{t-3} + \varepsilon_t$,
- AR(4) Model: $y_t = 1.82y_{t-1} - 1.32y_{t-2} + 0.46y_{t-3} - 0.06y_{t-4} + \varepsilon_t$,

(2) The error terms (ε_t) are independent and identically distributed from $\mathrm{N}(0, 1)$,

(3) The sample size is determined as $n = 25, 50, 100, 1000, 5000$ and 10000,

(4) An ideal maximum model order cut-off is used as 8(AR(1) - AR(8)),

(5) This simulation study is repeated 1000 times for each model and sample size.

Under these conditions, the simulation results were obtained by using the program MATLAB 7.1 and are given in Table 1. This table shows the percentage of true order determination by using each information criterion for each model and sample size.

Moreover, Figure 1, Figure 2, Figure 3 and Figure 4 are given in order to visualize the performance of each information criterion for each model and sample size.

Table 1. Simulation result for determination of model order

		MODELS			
		AR(1)	**AR(2)**	**AR(3)**	**AR(4)**
n = 25	**AIC**	82	85	51	4
	CAIC	91	92	46	2
	ICOMP	39	67	55	5
	ICOMP_PN	93	94	40	1
	ICOMP(IFIM)	59	78	52	2
	ICOMP(IFIM)_PN	96	95	34	1
n = 50	**AIC**	79	77	75	9
	CAIC	84	85	80	7
	ICOMP	34	62	77	7
	ICOMP_PN	91	94	83	2
	ICOMP(IFIM)	61	76	81	4
	ICOMP(IFIM)_PN	96	96	79	1
n = 100	**AIC**	74	73	73	11
	CAIC	77	78	77	10
	ICOMP	31	58	78	9
	ICOMP_PN	91	94	95	3
	ICOMP(IFIM)	60	77	86	7
	ICOMP(IFIM)_PN	96	97	96	2
n = 1000	**AIC**	73	73	71	45
	CAIC	73	74	72	45
	ICOMP	32	63	83	40
	ICOMP_PN	92	95	96	28
	ICOMP(IFIM)	63	79	89	36
	ICOMP(IFIM)_PN	97	98	98	25
n = 5000	**AIC**	72	72	73	74
	CAIC	72	72	73	74
	ICOMP	29	61	85	76
	ICOMP_PN	92	95	97	92
	ICOMP(IFIM)	58	79	91	80
	ICOMP(IFIM)_PN	97	97	98	93
n = 10000	**AIC**	71	73	73	74
	CAIC	71	73	73	74
	ICOMP	29	63	85	76
	ICOMP_PN	91	95	97	94
	ICOMP(IFIM)	59	78	91	81
	ICOMP(IFIM)_PN	97	97	98	96

Figure 1. Simulation result for AR(1) model

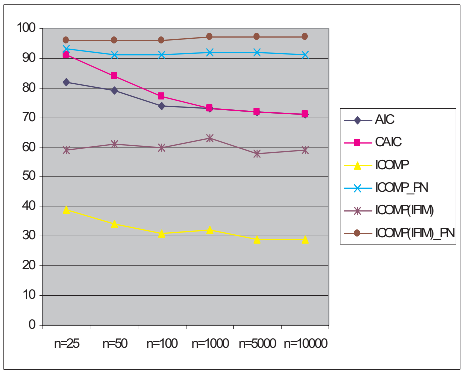


Figure 2. Simulation result for AR(2) model

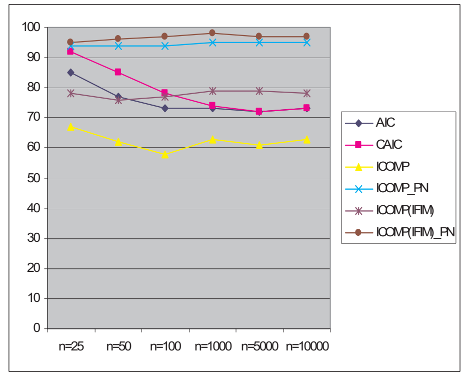


Figure 3. Simulation result for AR(3) model

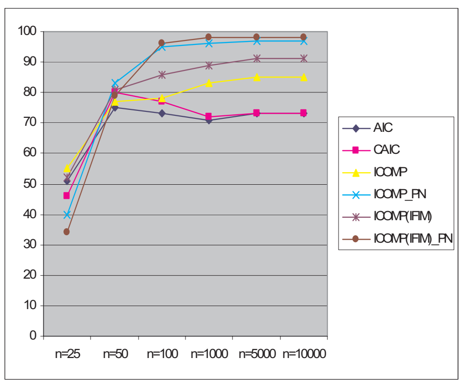
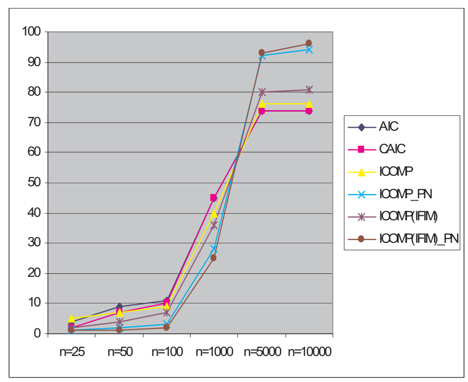


Figure 4. Simulation result for AR(4) model



4. Discussion and conclusion

In the simulation the variance of the error term is also changed, but there is no observable difference in the performance of the information criteria. When the cut-off value is increased, the percentage of true order determinations for each information criterion decreases. On the other hand, ICOMP_PN and ICOMP(IFIM)_PN are affected less in this respect. Moreover, the stationarity of the autoregressive models is important since unexpected results can appear for some nonstationary models.

As stated in [11], it can be seen from the simulation that CAIC performs well for small sample sizes, especially in the AR(1) and AR(2) models. ICOMP(IFIM)_PN, as a modified information criterion, has the best performance in the AR(1) and AR(2) models, also for big sample sizes in the AR(3) and AR(4) models. If the sample size is small, as the model order increases the performance of each information criterion will be worse and there can be problems in analyzing these models in some sense. As the sample size increases in the AR(3) and AR(4) models, the percentage of true order determination for each information criterion increases. Specifically, if the sample size and the autoregressive model order increase together, the performances of ICOMP_PN and ICOMP(IFIM)_PN increase noticeably.

As a conclusion, it is proposed that the information criteria ICOMP_PN and ICOMP (IFIM)_PN can be used to determine the true orders of AR models since they show more consistent behaviour.

Acknowledgements.

The author is grateful to Prof. Dr. Aydın Erar for his support in this work. Also, the helpful comments of the anonymous referee are gratefully acknowledged.

References

[1] Akaike, H. *A new look at the statistical model identification,* IEEE Transactions on Automatic Control, AC-**19**, 716-723, 1974.

[2] Akaike, H. *A Bayesian extension of the minimum AIC procedure of autoregressive model fitting*, Biometrika **66** 237–242, 1979.

[3] Bozdogan, H. *ICOMP: A new model selection criterion,* In H.H. Bock (Ed.), Classification and Related Methods of Data Analysis, North-Holland, CityplaceAmsterdam, 599–608, 1988.

[4] Bozdogan, H. *Akaike's information criterion and recent developments in information complexity,* Journal of Mathematical Psychology **44**, 62–91, 2000.

[5] Bozdogan, H. *Intelligent statistical data mining with information complexity and genetic algorithms,* In H. Bozdogan (Ed.), Statistical Data Mining and Knowledge Discovery, Chapman and Hall, Boca Raton, Florida, 15–56, 2004.

[6] Burg, J. P. *Maximum Entropy Spectral Analysis* (PhD Thesis, Department of Geophysics, Stanford University, Stanford, CA, 1975).

[7] Cho, S. H. and Song, I. *A formula for computing the autocorrelations of the AR process*, The Journal of the Acoustical Society of Korea **15**, 4–7, 1996.

[8] Enders, W. *Applied Econometric Time Series* (John Wiley and Sons, New York, 1995).

[9] Eshel, G. *The Yule Walker equations for the AR coefficients*, Citeulike-article-id: 763363, 2003.

[10] Fuller, W. A. *Introduction to Statistical Time Series* (Second Edition, John Wiley and Sons, New York, 1996).

[11] Hurvich, C. M. and Tsai, C. L. *Regression and time series model selection in small samples*, Biometrika **76**, 297–307, 1989.