

New Procedure to Improve the Order Selection of Autoregressive Time Series Model

Ali Hussein Al-Marshadi
Department of Statistics, Faculty of Science, King Abdulaziz University,
Jeddah, Saudi Arabia

Abstract: Problem statement: We propose new approach could be used to guide the selection of the “true” order of autoregressive model for different sample size. **Approach:** We used simulation study to compare four model selection criteria with and without the help of the new approach. The comparison of the four model selection criteria was in terms of their percentage of number of times that they identify the “true” order of autoregressive model with and without the help of the new approach. **Results:** The simulation results indicate that overall, the new proposed approach showed very good performance with all the four model selection criteria comparing to their performance without the help of the new approach, where the SBC, AICC and HQIC criteria provided the best performance for all the cases. **Conclusion:** The main result of our article is that we recommend using the new proposed approach with SBC, AICC and HQIC criteria as a standard procedure to identify the “true” order of autoregressive model.

Key words: Time series, autoregressive process, information criteria, proposed approach, originally stationary, research objective, information criterion

INTRODUCTION

An Autoregressive Moving average, {ARMA (p,q)}, model is a model for a time series that is originally stationary of order p,q with the form Eq. 1:

$$X_t = C + \sum_{k=1}^p \phi_k X_{t-k} + \sum_{m=1}^q \theta_m \varepsilon_{t-m} + \varepsilon_t, \quad (1)$$

In this model the time series depends on p past values of itself and on q past random error terms ε that have $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma^2$ and $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all t, the parameters $\phi_1, \phi_2, \dots, \phi_p$ are the autoregressive parameters associated with the time series values, the parameters $\theta_1, \theta_2, \dots, \theta_q$ are moving average parameters associated with the error terms, p is the order of the autoregressive component of the time series process and q is the order of the moving average component of the time series process (Box and Jenkins, 1976; Pankratz, 1983).

In this study we are concern with originally stationary Autoregressive model {AR (p)} which is a special case of the Autoregressive Moving average {ARMA (p, q)} model. The selection of the suitable order of the Autoregressive process is critical step in the analysis of time series since inappropriate order selection may result into inconsistent estimate of parameters and it increase in the variance of the model

when p greater than the true value (Shibata, 1976). In practice many researchers recommend using some information criterion to guide the selection of the true model order among the class of candidate model orders (Hurvich and Tsai, 1991; Kadilar and Erdemir, 2002; Sen and Shitan, 2002; Nakamura *et al.*, 2006 and Aladag *et al.*, 2010). Statisticians often use information criteria such as Akaike’s Information Criterion (AIC) by Akaike (1974), Schwarz’s Bayes Information Criteria (SBC) by Schwartz (1978), Hannan’s and Quinn’s Information Criterion (HQIC) by Hannan and Quinn (1979) and Bias-Corrected Akaike’s Information Criterion (AICC) by Hurvich and Tsai (1989) to guide the selection of the true model order. Lately, many studies have proposed and evaluated either new or modified criteria that are used to select the true Autoregressive model order (Padmanabhan and Rao, 1982; Hurvich and Tsai, 1989; Wong and Li, 1998; Tjee-Jian and Sepulveda, 1998; Kadilar and Erdemir, 2002; Sen and Shitan, 2002; Bengtsson and Cavanaugh, 2006; Nakamura *et al.*, 2006; Aladag *et al.*, 2010). Unfortunately, these criteria sometimes have low percentage of selecting the true model order.

Our research objective is evaluating a new approach could be used to guide the selection of the true Autoregressive model order. Also, our research objective involves comparing four model selection criteria in terms of their ability to identify the true model order with and without the help of the new approach.

MATERIALS AND METHODS

The ARMA procedure of the SAS system is a standard tool for fitting time series data. One of the main reasons that the ARMA procedure of the SAS system is very popular is the fact that it is a general-purpose procedure for time series. In ARMA procedure, users find the following two model selection criteria available, which give users tools can be used to select an appropriate model order. The two model selection criteria are (SAS Institute Inc, 2008):

- Akaike’s Information Criterion (AIC) by Akaike (1974)
- Schwarz’s Bayes Information Criteria (SBC) by Schwarz (1978)

Two more model selection criteria will be considered in this study that are bias-corrected Akaike’s Information Criterion (AICC) by Hurvich and Tsai (1989) and Hannan and Quinn Information Criterion (HQIC) by Hannan and Quinn (1979). Our study concerns with comparing the four information criteria in terms of their ability to identify the true Autoregressive model order with and without the help of the new approach.

The new approach involves using new sequence sampling technique and the Multiple Comparisons with the Best (MCB) procedure by Hsu (1984) as tools to help the four information criterion in identifying the right Autoregressive model order. The idea of the new approach can be justified and applied in a very general context, one which includes the selection of the true Autoregressive model order.

In the context of the Autoregressive models, the algorithm for using the new sequence sampling technique in our new approach can be outlined as follows:

Let the observed order vector of data O_1 is defined as follows:

$$O_1 = [X_{(1)} \ X_{(2)} \ X_{(3)} \ X_{(4)} \ X_{(5)} \ \dots \ X_{(n)}]$$

Table 1: The four settings of parameters for the two simulated Autoregressive models used in the Simulations

Setting number	AR (1) model		AR (2) model	
	ϕ_1		ϕ_1	ϕ_2
1	0.8		0.5	0.40
2	-0.9		0.4	-0.50
3	0.5		0.4	-0.30
4	-0.6		0.7	0.28

Generate the new sequence samples (O_1, O_2, \dots, O_n) using the new sequence sampling technique according to the order of the observed data (original sample) as follow:

$$O_1 = [X_{(1)} \ X_{(2)} \ X_{(3)} \ X_{(4)} \ X_{(5)} \ \dots \ X_{(n)}]$$

$$O_2 = [X_{(2)} \ X_{(3)} \ X_{(4)} \ X_{(5)} \ \dots \ X_{(n)} \ X_{(1)}]$$

$$O_3 = [X_{(3)} \ X_{(4)} \ X_{(5)} \ \dots \ X_{(n)} \ X_{(1)} \ X_{(2)}]$$

$$O_4 = [X_{(4)} \ X_{(5)} \ \dots \ X_{(n)} \ X_{(1)} \ X_{(2)} \ X_{(3)}]$$

$$O_n = [X_{(n)} \ X_{(1)} \ X_{(2)} \ \dots \ X_{(n-2)} \ X_{(n-1)}]$$

Fit all the class of candidate model orders of Autoregressive model, which we would like to select the true model order among them, to the observed data, (O_1), thereby obtaining the AIC*, HQIC*, AICC* and SBC* for each model order of the class of candidate model.

Repeat step (2) for each data sequence, (O_2, O_3, \dots, O_n).

Statisticians often use the previous collection of information criteria to guide the selection of the true model order such as selecting the model with the smallest value of the information criteria (Pankratz, 1983). We will follow the same rule in our approach, but we have the advantage that each information criteria has (n) replication values result of fitting the different sequences of the observed data (from step 1, 2 and 3).

To make use of this advantage, we propose using MCB procedure by Hsu (1984) to pick the winners (i.e., selecting the best set of models or single model if possible), when we consider the replicates of the information criteria, that is produced by each of the candidate model, as group.

The simulation study: A simulation study of PROC ARMA’s time series model analysis of data was conducted to compare the four model selection criteria with and without the new approach in terms of their percentage of number of times that they identify the true model order.

Normal data were generated according to stationary Autoregressive model with first and second orders. There were 24 scenarios to generate data involving four settings of the first order autoregressive and four settings of the second order Autoregressive, with three different sample sizes ($n = 25, 50$ and 100 observations). The four settings of parameter values for first order Autoregressive model and the four settings of

parameter values for second order Autoregressive model are given in Table 1. For those scenarios with sample size 25, we simulated 200 datasets, for those scenarios with sample size 50, we simulated 100 datasets and for those scenarios with sample size 100, we simulated 50 datasets. SAS code was written to generate the datasets according to the described setup using the SAS®9.1.3 package (SAS Institute Inc, 2008). The algorithm of our approach was applied to each one of the generated data sets with each candidate model (AR(1), AR(2), AR(3), AR(4), AR(5) and AR(6), total of 6 models) for each one of the four information criteria in order to compare their performance with and without the new approach. The objective of implanting MCB procedure by Hsu, (1984) in our new approach is the same objective that was used in my previous studies (AL-Marshadi, 2007; 2009; 2010A; 2010B).

RESULTS

The simulation results indicated that the new procedure selects the right model order as member of the best subset hundred percent of the times from the class of candidate model orders for all the information criteria. Table 2 summarizes results of the percentage of number of times that the procedure selects the true model order alone from the class of candidate model orders (AR(1), AR(2), AR(3), AR(4), AR(5) and

AR(6)) i.e. out of 6 models for the four criteria with the new approach and also, the percentage of number of times without the new approach, using the first parameters setting when n=25, 50 and 100. Table 3 summarizes results of the percentage of number of times that the procedure selects the true model order alone from the class of candidate model orders (AR(1), AR(2), AR(3), AR(4), AR(5) and AR(6)) i.e. out of 6 models for the four criteria with the new approach and also, the percentage of number of times without the new approach, using the second parameters setting when n=25, 50 and 100. Table 4 summarizes results of the percentage of number of times that the procedure selects the true model order alone from the class of candidate model orders (AR(1), AR(2), AR(3), AR(4), AR(5) and AR(6)) i.e., out of 6 models for the four criteria with the new approach and also, the percentage of number of times without the new approach, using the third parameters setting when n=25, 50 and 100. Table 5 summarizes results of the percentage of number of times that the procedure selects the true model order alone from the class of candidate model orders (AR(1), AR(2), AR(3), AR(4), AR(5) and AR(6)) i.e., out of 6 models for the four criteria with the new approach and also, the percentage of number of times without the new approach, using the fourth parameters setting when n=25, 50 and 100.

Table 2: The Percentage of number of times that the procedure selects the true model order alone from the class of candidate model for the four criteria with the first parameters setting, and (nominal Type I error=0.05)

		The four criteria							
Sample size	The right model	With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	AR (1)	99.00	98.50	99.50	100.00	69.50	72.00	82.00	87.00
	AR (2)	90.00	96.00	97.00	94.50	50.50	51.50	54.00	47.50
50	AR (1)	96.00	97.00	99.00	100.00	72.00	72.00	80.00	93.00
	AR (2)	91.00	93.00	93.00	95.00	61.00	64.00	68.00	74.00
100	AR (1)	86.00	90.00	90.00	100.00	66.00	68.00	70.00	96.00
	AR (2)	74.00	76.00	76.00	98.00	62.00	66.00	66.00	84.00

Table 3: The Percentage of number of times that the procedure selects the true model order alone from the class of candidate model for the four criteria with the second parameters setting, and (nominal Type I error=0.05)

		The four criteria							
Sample size	The right model	With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	AR (1)	90.50	91.00	97.00	97.50	60.50	64.50	67.50	84.00
	AR (2)	96.50	96.00	99.00	97.00	45.50	46.50	48.50	46.00
50	AR (1)	98.00	98.00	98.00	100.0	71.00	74.00	78.00	89.00
	AR (2)	98.00	98.00	99.00	99.0	67.00	67.00	70.00	81.00
100	AR (1)	98.00	98.00	100.00	100.0	68.00	68.00	68.00	96.00
	AR (2)	98.00	98.00	98.00	100.0	80.00	84.00	84.00	98.00

Table 4: The Percentage of number of times that the procedure selects the true model order alone from the class of candidate model for the four criteria with the third parameters setting, and (nominal Type I error=0.05)

The four criteria									
Sample size	The right model	With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	AR (1)	88.00	92.50	97.00	98.00	71.00	73.50	83.50	88.50
	AR (2)	95.50	96.00	97.50	94.00	25.00	24.50	24.50	20.50
50	AR (1)	84.00	91.00	94.00	100.00	73.00	75.00	81.00	96.00
	AR (2)	95.00	97.00	98.00	83.00	47.00	47.00	49.00	42.00
100	AR (1)	78.00	82.00	82.00	100.00	68.00	68.00	70.00	96.00
	AR (2)	88.00	92.00	92.00	86.00	58.00	62.00	62.00	68.00

Table 5: The Percentage of number of times that the procedure selects the true model order alone from the class of candidate model for the four criteria with the fourth parameters setting, and (nominal Type I error=0.05)

The four criteria									
Sample size	The right model	With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	AR (1)	89.50	91.00%	95.50	96.50	64.50	68.50	80.50	86.00
	AR (2)	96.00	97.00	98.50	94.00	35.00	35.50	35.50	31.00
50	AR (1)	91.00	94.00	96.00	100	72.00	76.00	79.00	93.00
	AR (2)	74.00	82.00	80.00	90.00	60.00	64.00	64.00	66.00
100	AR (1)	90.00	94.00	94.00	100	66.00	68.00	68.00	98.00
	AR (2)	74.00	82.00	80.00	90.00	60.00	64.00	64.00	66.00

Table 6: The Average Percentage of number of times that the procedure selects the true model order alone from the class of candidate model for the four criteria averaging over the four parameters settings, and (nominal Type I error=0.05)

The four criteria									
Sample size	The right model	With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	AR (1)	91.75	93.25	97.25	98.00	66.37	69.63	78.37	86.37
	AR (2)	94.50	96.25	98.00	94.88	39.00	39.50	40.63	36.13
50	AR (1)	92.25	95.00	96.75	100	72.00	74.25	92.75	92.75
	AR (2)	89.50	92.50	92.50	91.75	58.75	60.50	62.75	65.75
100	AR (1)	88.00	91.00	91.50	100	67.00	68.00	69.00	96.50
	AR (2)	83.50	87.00	86.50	93.50	65.00	69.00	69.00	79.00

Table 6 summarizes results of the average percentage of number of times that the procedure selects the true model order alone from the class of candidate model orders (AR(1), AR(2), AR(3), AR(4), AR(5) and AR(6)) i.e., out of 6 models for the four criteria with the new approach and also, the average percentage of number of times without the new approach, averaging over the four parameters settings when n = 25, 50 and 100.

DISCUSSION

Table 2-6 show that the performance of all the four criteria with the new approach is better than their performance without the new approach. Although the new approach shows very good performance over all with all the criteria for all the cases, it was outstanding with SBC, AICC and HQIC criteria.

CONCLUSION

In our simulation, we considered Autoregressive process, looking at the performance of the new

proposed approach for selecting the suitable Autoregressive model order with different cases. Overall, the new approach provided the best guide to select the suitable model order. The new approach showed outstanding performance with SBC, AICC and HQIC criteria. Thus, this new approach can be recommended to be used with one of the three mentioned criteria. Note for users of the propose approach: if the MCB procedure suggested the best subset of models contains more than one model, we recommend selecting the true model as the one with a smaller order since the examination of simulation results showed that in this case the other models are over fitted models, i.e., model that contains the right order of the true model and higher order terms. The main result of our article is that the three criteria SBC, AICC and HQIC criteria are competitive in term of their ability to identifying the right model order with the help of the new proposed.

REFERENCES

- Akaike, H., 1974. A new look at the statistical model identification. *Trans. Autom. Control*, 19: 716-723. DOI: 10.1109/TAC.1974.1100705
- Aladag, C.H., E. Egrioglu, S. Gunay and M.A. Basaran, 2010. Improving weighted information criterion by using optimization. *J. Comp. Applied Math.*, 233: 2683-2687. DOI: 10.1016/j.cam.2009.11.016
- AL-Marshadi, A.H., 2007. The new approach to guide the selection of the covariance structure in mixed model. *Res. J. Med. Med. Sci.*, 2: 88-97.
- AL-Marshadi, A.H., 2009. Bootstrap simulation procedure to guide the selection of the multiple linear regressions. *J. KAU-Sci.*, 21: 197-212.
- AL-Marshadi, A.H., 2010A. Simulation study for evaluating a new three stages procedure for selecting the right multivariate regression model. *Am. J. Biostat.*, 1: 32-41. DOI: 10.3844/amjbsp.2010.32.41
- AL-Marshadi, A.H., 2010B. Simulation study for comparing two three stages procedures for selecting the right multivariate regression model. *Canadian J. Comput. Math., Natural Sci., Eng. Med.*, 1: 112-126.
- Bengtsson, T. and J.E. Cavanaugh, 2006. An improved akaike information criterion for state-space model selection. *Comput. Stat. Data Anal.*, 10: 2635-2654. DOI: 10.1016/j.csda.2005.05.003
- Box, G.E.P. and G.M. Jenkins, 1976. *Time Series Analysis: Forecasting and Control*. 1st Edn., Holden-Day, San Francisco, ISBN: 0816211043, pp: 575.
- Hannan, E.J. and B.G. Quinn, 1979. The determination of the order of an autoregression. *J. Royal Statist. Soc., Ser. B*, 41: 190-195.
- Hsu, J.C., 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals Stat.*, 12: 1136-1144.
- Hurvich, C.M. and C.L. Tsai, 1989. Regression and time series model selection in small samples. *Biometrika*, 76: 297-307. DOI: 10.1093/biomet/76.2.297
- Hurvich, C.M. and C.L. Tsai, 1991. Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, 3: 499-509. DOI: 10.1093/biomet/78.3.499
- Kadilar, C. and C. Erdemir, 2002. Comparison of performance among information criteria in var and seasonal var models. *Hacettepe J. Math. Stat.*, 31: 127-137.
- Nakamura, T., K. Judd, A.I. Mees and M. Small, 2006. A comparative study of information criteria for model selection. *Int. J. Bifurcation Chaos*, 8: 2153-2175.
- Padmanabhan, G. and A.R. Rao, 1982. Order selection of AR models of Hydrologic Time Series. *Nordic Hydrol.*, 13: 93-104.
- Pankratz, A., 1983. *Forecasting with Univariate Box-Jenkins Model: Concepts and Cases*. 1st Edn., John Wiley and Sons, Inc., New York, ISBN: 0-471-09023, pp: 562.
- SAS Institute Inc, 2008. SAS OnlineDoc 9.1.3. SAS Institute Inc., Cary, NC.
- Schwartz, G., 1978. Estimating the dimension of a model. *Annals Stat.*, 6: 461-464.
- Sen, L.K. and M. Shitan, 2002. The performance of AICC as an order selection criterion in ARMA time series models. *Pertanika J. Sci. Technol.*, 1: 25-33.
- Shibata, R., 1976. Selection of the order of an Autoregressive model by Akaike information criterion. *Biometrika*, 63: 117-126. DOI: 10.1093/biomet/63.1.117
- Tiee-Jian, W. and A. Sepulveda, 1998. The weighted average information criterion for order selection in time series and regression models. *Stat. Probability Lett.*, 39: 1-10. DOI: 10.1016/S0167-7152(98)00003-0
- Wong, C.S. and W.K. Li, 1998. A note on the corrected Akaike Information criterion for threshold Autoregressive models. *J. Time Series Anal.*, 1: 113-124. DOI: 10.1111/1467-9892.00080