African Journal of Mathematics and Computer Science Research Vol. 5(5), pp. 102-106, 9 March, 2012
Available online at http://www.academicjournals.org/AJMCSR
DOI: 10.5897/AJMCSR12.012
ISSN 2006-9731©2012 Academic Journals

Full Length Research Paper

Improving the order selection of moving average time series model

Ali Hussein Al-Marshadi

Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.
E-mail: ALMarshadi-a@hotmail.com.

Accepted 6 February, 2012

We propose an approach could be used to select the right order of moving average model. We used simulation study to compare four model selection criteria with and without the help of our approach. The comparison of the four model selection criteria was in terms of their percentage of number of times that they identify the right order of moving average model with and without the help of our approach. The simulation results indicate that overall, our approach showed improving in the performance of the four model selection criteria comparing to their performance without the help of our approach, where the Schwarz's Bayes information criteria (SBC) criterion provided the best performance for all the cases considered in the study. The main result of our article is that we recommend using our approach with SBC criterion as a standard procedure to identify the right order of moving average model.

Key words: Time series, moving average process, information criteria, MCB Procedure.

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:

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

In this model the time series depends on p past values of itself and on q past random error terms ε that have $E(\varepsilon_t)=0,\ Var(\varepsilon_t)=\sigma^2$ and $Cov(\varepsilon_t,\varepsilon_{t-k})=0,$ for all t, the parameters $\phi_1,\phi_2,...,\phi_p$ are the autoregressive parameters associated with the time series values, the parameters $\theta_1,\theta_2,...,\theta_q$ 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 paper we are concerned with originally stationary moving average model {MA(q)} which is a special case of the autoregressive moving average {ARMA (p,q)} model. The selection of the suitable order of autoregressive moving average process is critical step in the analysis of time series since inappropriate order selection may result into inconsistent estimate of parameters and it increase the variance of the model when the order greater than the true value (Shibata, 1976). In practice statisticians recommend using information criterion to select the true model order among the class of candidate model orders (Hurvich and Tsai, 1991; Kadilar and Erdemir, 2002; Sen, and Shitan, 2002; Andr´es et al., 2004; Sak et al., 2005; Nakamura et al., 2006; Aladag et al., 2010). Statisticians often use information criteria such as Akaike's information criterion (AIC) by Akaike (1969), Schwarz's Bayes information criteria (SBC) by Schwarz (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 select the true model order. Lately, many studies have proposed and evaluated either new or modified criteria that are used to select the true model order (Padmanabhan and Rao, 1982; Wong and Li, 1998; Wu and Sepulveda, 1998; Jun-ichiro, 1999; Broersen and de Waele, 2002; Kadilar and Erdemir, 2002; Sen, and Shitan, 2002; Andr´es et al., 2004; Sak et al., 2005; Bengtsson and Cavanaugh, 2006; Nakamura et al., 2006;

Guoqi, and Xindong, 2007; Aladaget al., 2010; AL-Marshadi, 2011; Neelabh et al., 2011). Unfortunately, these criteria sometimes have low percentage of selecting the true model order.

Our research objective is to evaluate an approach which could be used to select the true moving average model order. Also, our research objective involves comparing four model selection criteria in terms of their ability to identify the right model order with and without the help of our approach.

METHODOLOGY

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 data. In ARMA procedure, users find the following two model selection information criteria available, which can be used as tools to select the true model order. The two model selection information criteria are (SAS Institute Inc., 2008):

1) Akaike's Information Criterion (AIC) by Akaike (1969).
2) 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 moving average model order with and without the help of our approach.

Our approach involves using 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 moving average model order (AL-Marshadi, 2011). The idea of our approach can be justified and applied in a very general context, one which includes the selection of the true moving average model order.

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

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

$$O_1 = \left[X_{(1)} \quad X_{(2)} \quad X_{(3)} \quad X_{(4)} \quad X_{(5)} \quad \ldots \quad X_{(n)} \right]^{`},$$

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

$$O_1 = \left[X_{(1)} \quad X_{(2)} \quad X_{(3)} \quad X_{(4)} \quad X_{(5)} \quad \ldots \quad X_{(n)} \right]^{`},$$
$$O_2 = \left[X_{(2)} \quad X_{(3)} \quad X_{(4)} \quad X_{(5)} \quad \ldots \quad X_{(n)} \quad X_{(1)} \right]^{`},$$
$$O_3 = \left[X_{(3)} \quad X_{(4)} \quad X_{(5)} \quad \ldots \quad X_{(n)} \quad X_{(1)} \quad X_{(2)} \right]^{`},$$
$$O_4 = \left[X_{(4)} \quad X_{(5)} \quad \ldots \quad X_{(n)} \quad X_{(1)} \quad X_{(2)} \quad X_{(3)} \right]^{`},$$
$$O_n = \left[X_{(n)} \quad X_{(1)} \quad X_{(2)} \quad . \quad . \quad . \quad X_{(n-2)} \quad X_{(n-1)} \right]^{`}$$

2. Fit all the class of candidate model orders of moving average 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.

3. Repeat step (2) for each data sequence, $(O_2, ..., O_n)$.

4. Statisticians often use the previous collection of information criteria to select the right 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 (that is, 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 (AL-Marshadi, 2011).

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 our approach in terms of their percentage of number of times that they identify the right model order.

Normal data were generated according to stationary moving average model with first and second orders. There were 24 scenarios to generate data involving four settings of the first order moving average, and four settings of the second order moving average, with three different sample sizes ($n = 25$, 50, and 100 observations). The four settings of parameter values for first order moving average model and the four settings of parameter values for second order moving average 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 (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6), total of 6 models) for each one of the four information criteria in order to compare their performance with and without our approach. The objective of implanting MCB procedure by Hsu (1984) in our approach is to select models into a subset with a probability of correct selection $p(\text{correct selection}) = (1\text{-}\alpha)$ that the "best" model is included in the subset where the subset could be single model if possible.

RESULTS AND DISCUSSION

The simulation results indicated that our 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 right model order alone from the class of candidate model orders (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6)) that is, out of 6 models for the four criteria with our approach and also, the percentage of number of times without our approach,

Table 1. The four settings of parameters for the two simulated moving average models used in the simulations.

Setting number	MA(1) model	MA(2) model	
	θ_1	θ_1	θ_2
1	0.8	0.5	0.4
2	-0.9	0.4	-0.5
3	0.5	0.4	-0.3
4	-0.6	0.7	0.28

Table 2. The Percentage of number of times that the procedure selects the right 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).

Sample size	The right model	The four criteria							
		With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	MA(1)	79.00	84.00	92.00	95.00	51.50	56.00	72.00	80.50
	MA(2)	84.00	89.50	96.00	94.50	27.00	33.00	42.50	37.00
50	MA(1)	77.00	84.00	88.00	98.00	63.00	67.00	72.00	92.00
	MA(2)	72.00	81.00	88.00	91.00	45.00	47.00	51.00	68.00
100	MA(1)	50.00	60.00	60.00	98.00	64.00	66.00	66.00	88.00
	MA(2)	68.00	74.00	74.00	96.00	56.00	66.00	66.00	78.00

Table 3. The Percentage of number of times that the procedure selects the right 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).

Sample size	The right model	The four criteria							
		With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	MA(1)	94.50	96.50	98.50	99.00	70.00	71.00	81.50	87.50
	MA(2)	96.50	97.00	99.00	99.50	52.00	56.00	63.00	59.50
50	MA(1)	96.00	97.00	99.00	100.0	68.00	71.00	75.00	92.00
	MA(2)	82.00	87.00	91.00	97.00	52.00	58.00	61.00	71.00
100	MA(1)	78.00	80.00	80.00	100.0	60.00	62.00	62.00	98.00
	MA(2)	70.00	76.00	76.00	98.00	62.00	68.00	66.00	90.00

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 right model order alone from the class of candidate model orders (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6)) that is, out of 6 models for the four criteria with our approach and also, the percentage of number of times without our 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 right model order alone from the class of candidate model orders (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6)) that is, out of 6 models for the four criteria with our approach and also, the percentage of number of times without our 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 right model order alone from the class of candidate model orders (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6)) that is, out of 6 models for the four criteria with our approach and also, the percentage of number of times without our approach, using the fourth parameters setting when n=25, 50, and 100. Table 6 summarizes results of the average percentage of number of times that the procedure selects the right model order alone from the class of candidate model orders (MA(1), MA(2), MA(3), MA(4), MA(5), and MA(6)) that is, out of 6 models for the four criteria with our approach and also, the average percentage of number of times without our approach, averaging over

Table 4. The Percentage of number of times that the procedure selects the right 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).

Sample size	The right model	The four criteria							
		With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	MA(1)	74.50	79.50	92.00	94.00	47.50	52.00	70.00	78.00
	MA(2)	93.00	97.00	98.50	96.00	33.50	35.50	40.50	35.00
50	MA(1)	67.00	73.00	86.00	98.00	58.00	62.00	67.00	92.00
	MA(2)	76.00	84.00	86.00	80.00	40.00	47.00	49.00	49.00
100	MA(1)	64.00	74.00	74.00	100.0	62.00	68.00	68.00	94.00
	MA(2)	70.00	74.00	74.00	92.00	66.00	68.00	68.00	74.00

Table 5. The Percentage of number of times that the procedure selects the right 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).

Sample size	The right model	The four criteria							
		With the new approach				Without the new approach			
		AIC (%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	MA(1)	89.00	92.50	97.50	98.00	64.00	67.00	76.00	84.00
	MA(2)	86.50	92.00	96.50	93.50	21.50	24.00	28.00	26.00
50	MA(1)	81.00	89.00	94.00	99.00	59.00	62.00	68.00	91.00
	MA(2)	71.00	81.00	90.00	87.00	32.00	39.00	42.00	35.00
100	MA(1)	70.00	76.00	76.00	100.0	62.00	66.00	66.00	96.00
	MA(2)	60.00	74.00	70.00	82.00	56.00	64.00	64.00	58.00

Table 6. The Average percentage of number of times that the procedure selects the right 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).

Sample size	The right model	The four criteria							
		With the new approach				Without the new approach			
		AIC(%)	HQIC (%)	AICC (%)	SBC (%)	AIC (%)	HQIC (%)	AICC (%)	SBC (%)
25	MA(1)	84.25	88.125	95.00	96.50	58.25	61.500	74.875	82.500
	MA(2)	90.00	93.875	97.50	95.87	33.50	37.125	43.500	39.375
50	MA(1)	80.25	85.750	91.75	98.75	62.00	65.500	70.500	91.750
	MA(2)	75.25	83.250	88.75	88.75	42.25	47.750	50.750	55.750
100	MA(1)	65.50	72.500	72.50	99.50	62.00	65.500	65.500	94.000
	MA(2)	67.00	74.500	73.50	92.00	60.00	66.500	66.000	75.000

the four parameters settings when n=25, 50, and 100.

Although our approach shows very good performance over all with all the criteria in all the cases, it was outstanding with SBC criterion.

Conclusion

In our simulation, we considered moving average process, looking at the performance of our approach for selecting the right moving average model order under different scenarios. Overall, our approach provided the best performance in selecting the right model order. Our approach showed outstanding performance with SBC criterion. Thus, our approach can be recommended to be used with the SBC criterion. Note for users of our approach: if the MCB procedure suggested the best subset of models contains more than one model, we recommend selecting the right 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 SBC criterion is the best criteria in term of their ability to identifying the right model order with the help of our approach.

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 CH, Egrioglu E, Gunay S, Basaran MA (2010). Improving weighted information criterion by using optimization. J. Comput. Appl. Math., 233: 2683-2687. DOI:10.1016/j.cam.2009.11.016.

AL-Marshadi AH (2011). New Procedure to Improve the Order Selection of Autoregressive Time Series Model. J. Math. Stat., 7(4):270-274. http://thescipub.com /stat/10.3844 /jmssp.2011.270.274.

Andr´es MA, Daniel P, Juan R (2004). Introducing Model Uncertainty In Time Series Bootstrap. Statistica Sinica, 14: 155-174. http://www3.stat.sinica.edu.tw/statistica /oldpdf/A14n16.pdf.

Bengtsson T, Cavanaugh JE (2006). An Improved Akaike Information Criterion for State-Space Model Selection. Comput. Stat. Data Anal., 50(10): 2635-2654. DOI:10.1016/j.csda.2005.05.003.

Box GEP, Jenkins GM (1976). Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco, ISBN: 0816211043, p. 575. http://onlinelibrary.wiley.com/doi/10.1002/9780470316566.fmatter/pdf

Broersen PMT. de Waele S (2002). Costs of Order Selection in Time Series Analysis. IEEE Instrumentation and Measurement, Technology Conference, Anchorage, AK, USA, 21-23 May, 2002. DOI: 10.1109/IMTC.2002.1006858 http://ieeexplore.ieee.org/xpl /free abs_all.jsp?arnumber=1006858.

Guoqi Q, Xindong Z (2007). On time series model selection involving many candidate ARMA models. Comput. Stat. Data Anal., 51: 6180 – 6196. http://www.sdasinfo. org.cn/zjk/%E9%87%8D%E7%82%B9%E9%A2%86%E5%9F%9F%E6%95%B0%E6%8D%AE%E5%BA%93/wf3012.pdf.

Hannan EJ, Quinn BG (1979). The Determination of the order of an autoregression. J. Royal Stat. Soc., Ser. B., 41: 190–195. http://www.citeulike.org/user/fabiobayer/article/4647560.

Hsu JC (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best. Ann. Stat., 12: 1136-1144. http://www.jstor.org/pss/2240990.

Hurvich CM, Tsai CL (1989). Regression and time series model selection in small samples. Biometrika, 76: 297-307. DOI: 10.1093/biomet/76.2.297.

Hurvich CM, Tsai CL (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. Biometrika, 78(3): 499-509. DOI:10.1093/biomet/78.3.49.

Jun-ichiro F (1999). Subsampling and Model Selection in Time Series Analysis. Biometrika, 86(3): 591-604. http://www.jstor. org/pss/ 2673656.

Kadilar C, Erdemir C (2002). Comparison of performance among information criteria in var and seasonal var models. Hacettepe J. Math. Stat., 31: 127-137. http://www.mat.hacettepe.edu.tr/hjms/english/issues/vol31/full-text/127-137.pdf.

Nakamura T, Judd K, Mees AI, Small M (2006). A comparative study of information criteria for model selection. Int. J. Bifurcat. Chaos, 16(8): 2153-2175. http://www.eie.polyu.edu.hk/~ensmall/pdf/IJBC16-2.pdf.

Neelabh R, Ramanathan TV (2011). A study on the focused information criterion for order selection in ARMA models. Technical Report 1/2011, Department of Statistics and Centre for Advanced Studies University of Pune, 411 007, INDIA. http://stats.unipune.ernet.in /TVR-NR.pdf.

Padmanabhan G, Rao AR (1982). Order selection of AR models of Hydrologic Time Series. Nordic Hydrol., 13: 93-104. http://www.iwaponline.com/nh/013/0093/0130093.pdf.

Pankratz A (1983). Forcasting with Univariate Box-Jenkins Model: Concepts and Cases. 1st Edn., John Wiley & Sons, Inc., New York, ISBN 0-471-09023, p. 562. DOI: 10.1002/9780470316566.fmatter http://onlinelibrary.wiley.com/doi/10.1002/9780470316566.fmatter/pdf

Sak M, Dowe D, Ray S (2005). Minimum message length moving average time series data mining. Computational Intelligence Methods and Applications, 2005 ICSC Congress on, 6. DOI:10.1109/CIMA. 2005.1662352http://ieeexplore.ieee. org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber% 3D1662352&authDecision=-203.

SAS Institute Inc (2008). SAS OnlineDoc 9.1.3. SAS Institute Inc., Cary, NC. http://support.sas.com/onlinedoc/913/docMainpage.jsp.

Schwartz G (1978). Estimating the Dimension of a Model. Ann. Stat., 6: 461-464. http://projecteuclid.org/DPubS?service=UI&version=1.0 &verb=Display&handle= euclid.aos/1176344136.

Sen LK, Shitan M (2002). The Performance of AICC as an Order Selection Criterion in ARMA Time Series Models. Pertanika J. Sci. Technol., 10(1): 25-33. http://psasir.upm.edu.my/3803/1/The_Performance_of_AICC_as_an_Order_Selection_Criterion_in.pdf.

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.

Wong CS, Li WK (1998). A note on the corrected Akaike Information criterion for threshold Autoregressive models. J. Time Ser. Anal., 19(1): 113-124. DOI: 10.1111/1467-9892.00080.

Wu T-J, Sepulveda A (1998). The weighted average information criterion for order selection in time series and regression models. Stat. Prob. Lett., 39: 1-10. DOI:10.1016/S0167-7152(98)00003-0.