Comparison of New Approach Criteria for Estimating the Order of Autoregressive Process

Salie Ayalew[1] M.Chitti Babu[2], L.K.Mohana Rao[3],

[1]*(Research Scholar): Department of Statistics, Andhra University, India, Visakhapatnam-530003)*
[2]*(Prof): Department of Statistics, Andhra University, India)*
[3]*(Prof): Department of Applied Economics, Andhra University, India)*

Abstract: *This paper compares the performance of the former model order determination criteria with new Approach in terms of selecting the correct order of an autoregressive time series model in small and large samples. The considered criteria are the Akaike information criterion (AIC); Scwarz information criterion (SIC), the Hannan − Quinn criterion (HQ) and the three new median criteria. Our results show that the median between AIC and Scwarz (MAS) criterion performs best in terms of selecting the correct order of an Autoregressive model for small and large samples sizes. Scwarz, Hannan Quinn, the median of Akaike and Hannan Quinn (MAH) criteria and median of Scwarz and Hannan Quinn can be said to perform best in large samples. The comparison of the six model selection criteria was in terms of their percentage of number of times that they identify the "true" model. The simulation results indicate that overall, the proposed Median of Akaike and Scwarz (MAS) information criterion showed very good performance in this simulation study. So, we recommend median of Akaike & Schwarz (MAS), Median of Schwarz & Hannan Quinn (MSH), and median of Akaike &Hannan Quinn information criterion as reliable criteria to identify the "true" model.*
Keywords*: Information criteria, best performance, new median information criteria.*

I. Introduction

It is well known that most economic and financial series follow the Autoregressive Moving average [ARMA (p,q)] model, and more often than not the autoregressive [AR(p)] model where p and/or q are the order of the model. However, determination of the correct order p, q has been a source of serious concern to analysts over the years, since inappropriate order selection may result in inconsistent estimate of parameters if p< true value or may not be consistent and increase in the variance of the model if p>true, Shibata [1].

In recent years, there has been a substantial literature on this problem and different criteria have been proposed to aid in choosing the order of the ARMA (p,q) process. These criteria are based on theoretical considerations that provide only asymptotic properties of the resulting estimators. The practitioner, however, usually faces the problem of making a choice on the basis of a limited data set.

Among the criteria considered are the Akaike information criteria (AIC); Scwarz Information criteria (SIC), Hannan Quinn criteria (HQ); Carlos Information criteria (CIC) and others. It is rarely the case that the 'true' order of a process is known. One of the most difficult and dedicate part of the time series analysis is the selection of the order of the process, based on a finite set of observations, since further analysis of that series is based on it. To overcome this difficulty several order selection criteria had been proposed in the literature but we don't have any criterion which could be considered as the best criterion in all situations

This paper therefore compares the performance of the commonly used information criteria such as Akaike Information Criteria (AIC), Scwarz Information Criteria (BIC), and Hannan – Quinn criteria with other new introduced median information criteria.

II. Review of Literature

The information criterion has been widely used in time series analysis to determine the appropriate order of a model. The information criteria are often used as a guide in model selection. The goal of any order selection procedure is to estimate the order p for an AR model on the basis of n successive observations from a time series X(t) while the notion of an information criterion is to provide a measure of information in terms of the order of the model, that strikes a balance between these measures of goodness of fit and parsimonious specification of the model. The Kullback–Leibler quantity of information contained in a model is the distance from the "true" model and is measured by the log likelihood function.

Models, by true models definition, are only approximations to unknown reality or truth; there are no that true models perfectly reflect full reality. George Box [2] made the famous statement "All models are wrong but some are useful." Further, a "best model," for analysis of data, depends on sample size; smaller effects can often only be revealed as sample size increases.

Several criteria are used for this purpose, in particular, we discuss the AIC (Akaike, 1974), SIC (Scwarz, 1978) and HQ (Hannan and Quinn, 1979) amongst others to compare with our new criteria. All these criteria's objective is minimizing the residual sum of squares and impose a penalty for including an increasingly large number of regressors (lag values).

2.1. Akaike Information Criterion

The Akaike Information Criterion (AIC), Akaike [3] is an objective measure of model suitability which balances model fit and model complexity. Considering a stationary time series {Xt}, t = 1, 2, …, T, the Akaike information criteria consist of minimizing the function:

$$K(p) = T \log \sigma^2 + pc(n) \qquad (1)$$

$$P = \{0, 1, 2, \dots m\}$$

where σ^2 is the estimated residual variance for a fitted AR (p) model, c(n) is a penalty term, n is the number of observations and m is a pre – determined upper autoregressive order.

To obtain:

$$AIC = \ln \hat{\sigma}^2_p + \frac{2m}{T} \qquad (2)$$

$$p \in P = \{0, 1, 2, \dots m\}$$

$$\hat{\sigma}^2_p = \frac{1}{T} \sum_{i=1}^{N} \left[X_i - \sum_{j=1}^{p} \phi_j X_{t-j} \right]^2 \qquad \text{or more compactly as} \qquad \hat{\sigma}^2_p = \frac{1}{T} \sum_{t=1}^{N} \varepsilon_t^2$$

Selection of the chosen model is then made by considering the minimum AIC = min {AIC (p)}, that is the model with smallest AIC is chosen. Details of the proof can be found in Akaike (1974) and Shibata (1976). One advantage of AIC is that it is useful for not only in sample data but also in out of sample forecasting performance of a model. This criterion also designed for minimizing the forecast error variance. In sample forecasting essentially tells us how the chosen model fits the data in a given sample while the out–of–sample forecasting is concerned with determining how a fitted model forecasts future values of the regressed, given the values of the regressors. It is also useful for both nested and non-nested models. The outstanding demerit of this criterion is that of inconsistency. The procedure has been criticized because it is inconsistent and tends to over fit a model, Shibata (1976) showed this for autoregressive model and Hannan [4] for ARMA models.

2.2. Schwarz Information Criteria

The Schwarz information criterion is obtained by replacing the non-negative function c(n) in (2) by $\log_e(N)$. Hence, we have

$$SC = \ln \hat{\sigma}^2_p + \frac{m}{T} \ln T \qquad (3)$$

$$p \in P = \{0, 1, 2, \dots m\}$$

where σ_p^2 is obtained as above and the appropriate model is obtained as that which minimizes the SIC (p) above, that is min (SIC (p). Details of the discussion can be found in Rissanen [5]; Schwarz[6] and Stone[7].

The SIC imposes a harsher penalty than AIC, as its obvious from comparing (2) and (3). An important advantage of SIC is that for a wide range of statistical problems, it is order consistent (i.e. when the sample size grows to infinity, the probability of choosing the right model converges to unity) leading to more parsimonious models. Like the AIC, the lower the value of SIC, the better the model. The difference between AIC and SC is due to the way they were designed. AIC is designed to select the model that will predict best and is less concerned with having a few too many parameters. SIC is designed to select the true values of p and q exactly Like AIC, SIC can be used to compare in–sample or out–of–sample forecasting performance of a model.

2.3. Hannan – Quinn Criterion

The Hannan-Quinn criterion for identifying an autoregressive model denoted by HQ(p) was introduced by Hannan and Quinn (1979). The adjusted version of it can also be applied to regression models, Al-Subaihi [8]. It is obtained by replacing the non – negative penalty function c (n) in equation (2) by log(logT).

Thus, we have :

$$HQ = \ln \hat{\sigma}^2_p + 2 \frac{m}{T} \ln(\ln T) \qquad (4)$$

$$p \in P = \{0, 1, \dots m\}.$$

The best model is the model that corresponds to minimum HQ i.e. min (HQ (p)). The order selection procedure presented above have the advantage of being objective and automatic, but it over-fit when the sample size is small. Detailed discussion on this can be found in Hannan-Quinn (1979); McQuarrie and Tsai [9].

Having noticed the limitations of these criteria, the aim of this study is to examine the veracity of these claims by introducing other new median criteria and compare their performance using the simulated data.

III. Methodology

3.1. Identification of optimal orders:

If the order of the AR process is unknown, it can be estimated with the help of information criteria (Helmute Lutkepohl [10]. For this purpose, AR processes with successively increasing orders $p = 1, 2..., p_{max}$ are estimated. Finally, the order p* is chosen which minimizes $\hat{\varepsilon}_t(p)$, the estimated residuals of the AR (p) process with the respective criterion, while m is the number of estimated parameters. If the constant term is estimated too, m is p plus one for an AR (p) process. These criteria are always based on the same principle: They consist of the sum of squared residuals (or its logarithm), which decreases when the number of estimated parameters increases, and of a 'punishment term', which increases when the number of estimated parameters increases. This is also reflected in the following relations, which, because of the different punishing terms, hold for these criteria:

$$\hat{P}(SIC) \leq \hat{P}(HQ) \tag{5}$$

$$\hat{P}(SIC) \leq \hat{P}(AIC) \text{ for } T \geq 8 \tag{6}$$

$$\tag{7}$$

$$\hat{P}(HQ) \leq \hat{P}(AIC) \text{ for } T \geq 16$$

These proofs are based on Quinn (1980) and Paulsen (1984) and the basic idea is to show that, for P> m, the quantity ln(SSE(p)/T)/ln(SSE(m)/T) is less than one in large samples, where p is the optimality of orders and m is the order of autoregressive and indicate minimum log likelihood function. This implies that ln (SSE (p)/T) is less than ln (SSE (m)/T).

The relation can be proved in the following ways:

From lemma stated in Lütkepohl (1991),let $a_0, a_1,...,a_M$, $b_0,b_1,...,b_M$ and $c_0,c_1,...,c_M$ be real numbers. If $b_{m+1} - b_m < a_{m+1} - a_m$, m =0, 1,...,M − 1, holds and if nonnegative integers n and k are chosen such that $c_n + a_n = \min\{c_m + a_m | m = 0, 1,...,M\}$ and $c_k + b_k = \min\{c_m + b_m | m = 0, 1,...,M\}$, then $k \geq n$. and let $c_m = \ln |SSE(m)/T|$, $b_m = 2mK^2/T$ and $a_m = mK^2 \ln T/T$. Then AIC (m) $= c_m + b_m$ and SIC (m) $= c_m + a_m$. The sequences a_m, b_m, and c_m satisfy the conditions of the lemma if $2K^2/T = 2(m+1)K^2/T − 2mK^2/T = b_{m+1} − b_m < a_{m+1} − a_m = (m+1)K^2 \ln T/T − mK^2 \ln T/T = K^2 \ln T/T$ or, equivalently, if $\ln T > 2$ or $T > e^2 = 7.39$. Hence, choosing $k = \hat{p}$ (AIC) and $n = \hat{p}$ (SC) gives $\hat{p}$ (SC) $\leq \hat{p}$ (AIC) if $T \geq 8$. Alternatively, it can be proved like as follows:

$$AIC(m) - SIC(M) = \ln(\frac{SSE}{T}) + 2(\frac{p+1}{T}) - \left[\ln(\frac{SSE}{T}) + (p+1)(\frac{LnT}{T})\right] = 2(\frac{p+1}{T}) - (p+1)\frac{\ln T}{T}$$

$$= (p+1)\left(\frac{2}{T} - \frac{\ln T}{T}\right) = (p+1)\left(\frac{2}{T} - \frac{\ln T}{T}\right) = (p+1)\left(\frac{2 - \ln T}{T}\right) \tag{8}$$

From the last term we can observe that (p+1)/T is always positive and AIC (m)-SIC (m) is less than zero from Lütkepohl (1991) then the value of 2-lnT must be negative. That means T is greater than or equal to eight and consequently the optimal order of AIC is greater than SC. Similarly for the case of (7), if we subtract HQ from AIC, AIC (p)-HQ (P) we also obtain

$$AIC(m) - HQ(M) = \ln(\frac{SSE}{T}) + 2(\frac{p+1}{T}) - \left[\ln(\frac{SSE}{T}) + 2(p+1)(\frac{\ln(LnT)}{T})\right] = 2(\frac{p+1}{T}) - 2(p+1)\frac{\ln(\ln T)}{T}$$

$$= 2((p+1)/T)(1 - \ln(\ln(T))) \tag{9}$$

The right side of equation (9) is less than zero and it implies that 1-ln (lnT)≤0, that is T≥e^e ≅16.

The purpose of this paper is to introduce a new information criteria between AIC and SC or AIC and HQ. We consider the median of AIC and SIC, HQ and SIC or AIC and HQ as a criteria.

3.2. Median of SC and HQ denoted as MSH

From equation (5), it is also possible to take the median of SIC and HQ as a new information criterion:

$$MSH(m) = \left(\frac{SIC + HQ}{2}\right) = \frac{1}{2}\left[\ln(\frac{SSE}{T}) + 2\left(\frac{p+1}{T}\right)\ln(\ln T) + \ln\left(\frac{SSE}{T}\right) + (p+1)\frac{\ln T}{T}\right]$$

$$= \ln\left(\frac{SSE}{T}\right) + \left(\frac{p+1}{2T}\right)\left(2\ln(\ln T) + \ln T\right) \tag{10}$$

Proposition 1: if $\hat{p}$ (SC) $\leq \hat{p}$ (HQ), then $\hat{p}$ (SIC) $\leq \hat{p}$ (MSH) for T ≥ 1

Proof:

Take the difference between MSH and SIC

$$MSH(m) - SIC(m) = \left[\ln(\frac{SSE}{T}) + \left(\frac{p+1}{2T}\right)(2\ln(\ln T) + \ln T)\right] - \ln\left(\frac{SSE}{T}\right) + \left(\frac{p+1}{T}\right)\frac{\ln T}{T}$$

$$= \frac{(p+1)}{T}\left[\frac{2\ln(\ln T) + \ln T}{2} - \ln T\right] \leq 0 \tag{11}$$

$2\ln(\ln T) - \ln T \leq 0$, this implies that $\ln\left[\frac{(\ln T)^2}{T}\right] \leq 0$, and $\left[\frac{(\ln T)^2}{T}\right] \leq 1$ *which is* $(\ln T)^2 \leq T$, finally $T \leq e^{\sqrt{T}}$.

This is true for all T ≥ 1, hence MSH is superior than SIC for model selection.

Proposition 2: if $\hat{p}$ (SC) $\leq \hat{p}$ (HQ), then $\hat{p}$ (MSH) $\leq \hat{p}$ (HQ) for all T

$$HQ(m) - MSH(m) = \left[\ln(\frac{SSE}{T}) + 2\left(\frac{p+1}{T}\right)\ln(\ln T)\right] - \ln\left(\frac{SSE}{T}\right) + (p+1)\frac{2\ln(\ln T) + \ln T}{2T}$$

$$= \frac{(p+1)}{2T}(-\ln T) \tag{12}$$

From this equation we get $-\ln T \leq 0$ and which shows that T is greater than one. MSH is more optimal criterion than HQ for any value of T.

3.3. Median of AIC and SIC and denote as MAS

From equation (6), it is possible to take the median of AIC and SC as a new information criterion:

$$MAS(m) = \left(\frac{AIC + SIC}{2}\right) = \frac{1}{2}\left[\ln(\frac{SSE}{T}) + 2\left(\frac{p+1}{T}\right) + \ln\left(\frac{SSE}{T}\right) + (p+1)\frac{\ln T}{T}\right]$$

$$= \ln\left(\frac{SSE}{T}\right) + \left(\frac{p+1}{2T}\right)(2 + \ln T) \tag{13}$$

If we compare MAS with AIC and SC, does it have a contribution for identification of model selection as equivalent as AIC and SIC? Does it have superiority in model selection with respect to number of observation?

Proposition 3: if $\hat{p}$ (SIC) $\leq \hat{p}$ (AIC) for T ≥ 8, then $\hat{p}$ (MAS) $\leq \hat{p}$ (AIC) for T ≥ 8

Proof: Take the difference between of AIC and MAS,

$$AIC(m) - MAS(m) = \left[\ln(\frac{SSE}{T}) + 2\left(\frac{p+1}{T}\right)\right] - \left[\ln\left(\frac{SSE}{T}\right) + (p+1)\frac{2 + \ln T}{2T}\right]$$

$$= \frac{(p+1)}{2T}(2 - \ln T) \leq 0 \tag{14}$$

This implies that $(1 - \ln T/2) \leq 0$ and T ≥ 8

Proposition 4: if $\hat{p}$ (SIC) $\leq \hat{p}$ (AIC) for T ≥ 8, then $\hat{p}$ (SC) $\leq \hat{p}$ (MAS) for T ≥ 8

Proof: Take the difference between MAS and SC

$$MAS(m) - SIC(m) = \left[\ln(\frac{SSE}{T}) + \left(\frac{p+1}{2T}\right)(2 + \ln T)\right] - \left[\ln\left(\frac{SSE}{T}\right) + (p+1)\frac{\ln T}{T}\right]$$

$$= \frac{(p+1)}{2T}(2 + \ln T) - (p+1)\frac{\ln T}{T} = \frac{(p+1)}{2T}(2 - \ln T) \leq 0 \tag{15}$$

Then $-\ln T \leq -2$ and this implies that $\ln T$ is greater than or equal to two this implies to T ≥ 8

3.4. Median of AIC and HQ and denote as MAH

From equation (7), it is also possible to take the median of AIC and HQ as a new information criterion.

$$MAH(m) = \left(\frac{AIC+HQ}{2}\right) = \frac{1}{2}\left[\ln(\frac{SSE}{T}) + 2\left(\frac{p+1}{T}\right) + \ln\left(\frac{SSE}{T}\right) + 2(p+1)\frac{\ln(\ln T)}{T}\right]$$

$$= \ln\left(\frac{SSE}{T}\right) + \left(\frac{p+1}{T}\right)(1+\ln(\ln T)) \qquad 16$$

Proposition 5: if $\hat{p}$ (HQ) $\leq \hat{p}$ (AIC) for T$\geq$ 16, then $\hat{p}$ (MAH) $\leq \hat{p}$ (AIC) for

T$\geq$ 16

Proof: Take the difference between AIC and MAH

$$AIC(m) - MAH(m) = \left[\ln(\frac{SSE}{T}) + \left(2\frac{p+1}{T}\right)\right] - \left[\ln\left(\frac{SSE}{T}\right) + (p+1)\frac{(1+\ln T)}{T}\right]$$

$$= \frac{(p+1)}{T}(1\text{-}\ln(\ln T) \leq 0 \qquad (17)$$

$$1\text{-}\ln(\ln T) \leq 0 \Rightarrow \ln(\ln T) \geq 1 \Rightarrow T \geq e^e \approx 16$$

Proposition 6: if $\hat{p}$ (HQ) $\leq \hat{p}$ (AIC) for T$\geq$ 16, then $\hat{p}$ (HQ) $\leq \hat{p}$ (MAH) for

T$\geq$ 16

Proof: Take the difference between HQ and MAH

$$MAH(m) - HQ(m) = \left[\ln(\frac{SSE}{T}) + \left(\frac{p+1}{T}\right)(1+\ln(\ln T)\right] - \left[\ln\left(\frac{SSE}{T}\right) + 2(p+1)\frac{\ln(\ln T)}{T}\right]$$

$$= \frac{(p+1)}{T}(1+\ln(\ln T)\text{-}2\frac{p+1}{T}\ln(\ln T) = \frac{(p+1)}{T}(1\text{-}\ln(\ln T) \leq 0 \qquad (18)$$

$$1\text{-}\ln(\ln T) \leq 0 \Rightarrow \ln(\ln T) \geq 1 \Rightarrow T \geq e^e \approx 16$$

3.5. Model Selection

A large effort has been spent on a coherent theory of model selection over the past 40 years. We will not re-view this material in detail as it is covered in a number of books (e.g., Linhart and Zucchini1986 [11], McQuarrie and Tsai 1998; Burnhamand and Anderson2002), research monographs, Sakamtoel [12], and hundreds of journals papers, deLeeuw [13].

The starting point for effective model selection theory is Kullback-Leibler (K-L) information, I(f, g) Kullback and Leibler[14]. This is interpreted as the information, I, lost when full truth, f, is approximated by a model, g. Given a set of candidate models g$_i$, one might compute K-L information for each of the R models and select the one that minimizes information loss that is, minimize I(f, g) across models. The starting point for effective model selection theory   Akaike (1973, 1974) provided a simple way to estimate expected K-L information, based on a bias corrected maximized log-likelihood value. This was a major breakthrough Parzen et al.,[15]. Soon thereafter, better approximations to the bias were derived, Sugiura, Hurvich and Tsai1[16] and the result, of relevance here, is an estimator Akaike Information Criterion (AICc) of twice the expected K-L information loss

$$AICc = n\log(\sigma^2) + 2k + \left(\frac{2k(k+1)}{n-k-1}\right) \qquad (19)$$

where σ^2 is the estimated residual variance, n is the number of observations, and k is the number of estimated parameters for the model. Here, the estimator of σ^2 =WSSR/n, where WSSR is the weighted sum of squared residuals. The second term accounts for first-order bias, and the third term accounts for second-order bias resulting from a small number of observations. This is a precise mathematical derivation, with the third term depending on the assumed distribution of residuals, in this case, normally distributed error. Accounting for second-order bias is important when n/k <40. The aforementioned expression applies to analyses undertaken by a least squares approach; similar expressions are available for those using maximum likelihood procedures (Akaike1973). AICc is computed for each of the models; the model with the lowest AICc value is the best model, and the remaining models are ranked from best to poorest, with increasing AICc values.

Model probabilities: Remember from the derivation of the BIC

$$P(M_j/Y) = \frac{P(M_j)\exp(-\frac{1}{2}BIC_{n,j})}{\sum_{j=1}^{k} P(M_j)\exp(-\frac{1}{2}BIC_{n,j})} \qquad (20)$$

where, M= {$M_1...M_k$} are the specified list of models used and P (M_j) prior probabilities**.**
Using Schwarz's approximation:

$$P\left(M_j/Y\right) \approx \exp\left\{-\frac{1}{2}BIC(M_j)\right\} \qquad (21)$$

Using this as weights and rescaled of normalized such that the weights sum to one.
This motivates the smooth BIC weights in frequents model averaging and easy to compute approximation to posterior probabilities. Smoothed weights from model selection, for instance AIC, BIC, FIC, Akaike weights – Burnham Anderson, 2002[17]

$$C_{aic}\left(S\right) = \frac{\exp(-\frac{1}{2}AIC_s)}{\sum_{j=1}^{k}\exp(-\frac{1}{2}AIC_s)} \qquad (22)$$

$$C_{bic}\left(S\right) = \frac{\exp(-\frac{1}{2}BIC_s)}{\sum_{j=1}^{k}\exp(-\frac{1}{2}BIC_s)} \qquad (23)$$

Calculation of the AICc values can be posed to retain or omit values that are constant across models (e.g., multinomial coefficients) and are affected by the number of observations; thus, it is essential to compute and use simple differences

$$\Delta_i = AIC_{ci} - AIC_{c\min} \qquad (24)$$

For each model, i, in the set of R models, where AIC_{cmin} is the minimum AICc value of all the models in the set. These values are on an information scale (-log [probability]), free from constants and sample size issues. A Δ_i represents the information loss of model i relative to the best model. As discussed by Burnham and Anderson (2002), models with $\Delta_i < 2$ are very good models, while models with $4 < \Delta_i < 7$ have less empirical support. In most cases, models with Δ_i greater than 10 can be discarded from further consideration.
Simple transformation yields model probabilities or Akaike weights (also referred to as posterior model probabilities)

$$W_i = \frac{\exp^{-.5\Delta_i}}{\sum_{j=1}^{R}\exp^{-.5\Delta_j}} \qquad (25)$$

where w_i is the weight of evidence in favour of model i being the best model in the sense of minimum K-L information loss. These weights are also useful in multimodel inference.
As parameters are added to a model, accuracy and variance increase, (fit improves, while uncertainty increases). Use of AICc selects models with a balance between accuracy and variance; this is the principle of parsimony. Prediction can be further improved by basing inference on all the models in the set.

IV.   Data Illustrations and discussion

In order to contrast the performance of the order identification criteria, simulation study was conducted on a wide range of autoregressive (AR) processes with different characteristics.
The following assumptions were made. The random numbers follow a standard normal distribution with mean zero and variance unity. Data were generated for samples of T = 2-10, 50, 100, 1000, and 2000 with 100 replications each. Autoregressive models of orders p is equal to zero was generated and compared with 1, 2, 3, 4 and 5 lags. For each of the model structure, an autoregressive model was fit. The model order were examined using the criteria AIC, BIC, HQ and the 3 other new criteria using both the Eviews software package and Microsoft Excel.
The performance criterion is that, the information criterion with the highest number of cases (or weight of evidence) of selecting the correct order of the given autoregressive (AR), process is considered the best. To achieve our objectives we compute the probability of the correct estimation for each of these criteria. This probability (weight of evidence) could be any number between zero and one. Possible results are as follows:
1.   If this probability is 1, then it means that the criterion picks up the true lag length in all the cases and therefore is an excellent criterion.
2.   If the probability is close to 1 or greater than 0.5, then it implies that the criterion manages to pick up the true lag length in most of the cases and hence is a good criterion.
3.   If the probability is close to zero or less than 0.5 then it mean that the criterion fails to select the true lag length in most of the cases therefore it is not a good criterion.

4. If this probability is zero, it implies that criterion fails to pick up the true lag length in all the cases and hence is poor criteria.
5. A criterion under estimate the true lag length if it picks up a lag length, which is lower than the true lag length and if it selects a lag length, which is greater than the true lag length then it over estimates the lag length.

Since we want to study the behavior of all these criteria, therefore, along with the cases of correct estimation we also observe the selected lag length of all these criteria in all the cases to compute the probability of under estimation and over estimation.

After fitting the models to the generated data, the number of times each identification criteria (AIC, BIC, HQ, MAS, MAH and MSH) was able to accurately identify the correct order of a given AR structure and frequencies obtained.

In measuring the performance of the generated models, the criterion that has the highest number of cases (or percentage frequency) of selecting the correct order of the given AR structure is considered to be the best criterion. The results of the analysis were tabulated for different criterion of T = 2-10, 50, 100, 1000 and 2000 with replication R=100 in Table 1, Table 2 and Table 3.The new approach median information criterion (MAS) show outstanding performance for large samples and small samples of data. Among the criteria, MAS, SIC and MSH for large sample series are consistent in the sense that if the set of candidate models contains the true" model, then these criteria select the true model with probability 1 asymptotically. We found that MAS performed better for small sample and for large samples. Other criteria like AIC, HQ, MSH and MAH are inconsistent for small sample size of the series. AIC, SC, HQ, MSH and MAH over estimate for T=7 and AIC over estimate for T=2000 the true lag.

Based on information criteria differences in Table 5, the constant intercept parameters (order 0) is substantial supported by all criteria regardless of the sample size. Order one or 1st lag is considerably less support by all samples, except MAS for n=50, and except SC and MSH for n=1000 and 2000. Furthermore, the 5th lag is essentially discarded by some criteria for small size but for large samples of T=100, 1000 and 2000 all criteria discards lag five.

V. Summary and conclusion

The output from Table 1-5 show that the performance of the MAS criteria approach is better than the others approach. All criteria estimate the true lag length for large samples except AIC.

The output of the study shows that for a series having more than 10 observations, there is an improvement in performance for each of these five criteria except AIC as the sample size grows. A criterion under estimate the true lag length if it picks up a lag length, which is lower than the true lag length, and if it selects a lag length, which is greater than the true lag length then it over estimates the lag length. This study also shows that AIC is inconsistent to select the correct lag length and over estimate the true lag length for large observation. Then AIC is not a good criteria to select the true lag.
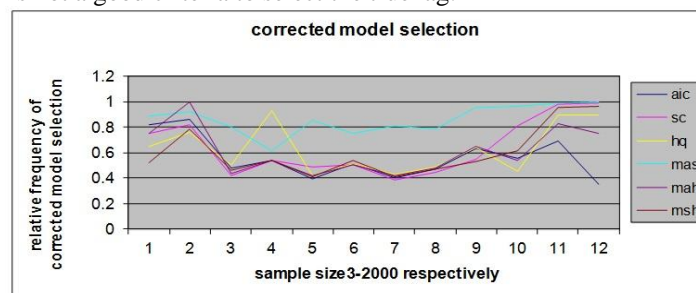

Figure 1.Relative frequency of correct model order selection for the considered information criteria.

Table 1: Frequency and percentage distribution from simulated distribution data for different sample size with replication R=100

sample size	criteria	order					
		0	1	2	3	4	5
T=2	AIC						
	Sc						
	HQ						
	MAS						
	MAH						
	MSH						
T=3	AIC	0.823938	0.176062				
	Sc	0.748864	0.251136				
	HQ	0.654144	0.345856				
	MAS	0.890178	0.109822				
	MAH	0.748435	0.251565				
	MSH	0.522823	0.477177				
T=4	AIC	0.859472	0.140525				
	Sc	0.818179	0.181817				
	HQ	0.757227	0.242766				
	MAS	0.924427	0.075573				
	MAH	1.00000	0.0000				
	MSH	0.789314	0.210681				
T=5	AIC	0.481211	0.086237	0.354237	0.078315		
	Sc	0.384897	0.083852	0.418718	0.112533		
	HQ	0.238853	0.072295	0.501568	0.187283		
	MAS	0.805199	0.064532	0.118548	0.011721		
	MAH	0.352929	0.082197	0.4388	0.126074		
	MSH	0.307975	0.079084	0.465483	0.147458		
T=6	AIC	0.190194	0.035824	0.539077	0.177832	0.057073	
	Sc	0.15546	0.03249	0.54264	0.19865	0.07075	
	HQ	0.010651	0.003043	0.934603	0.034772	0.01693	
	MAS	0.616811	0.04743	0.291377	0.039241	0.005141	
	MAH	0.125473	0.029109	0.539535	0.219224	0.086658	
	MSH	0.112242	0.027431	0.535604	0.229256	0.095467	
T=7	AIC	0.39879	0.077502	0.25884	0.25884	0.005836	0.000192
	Sc	0.488915	0.097623	0.334975	0.070246	0.007972	0.00027
	HQ	0.419325	0.096318	0.380198	0.091719	0.011974	0.000466
	MAS	0.852054	0.062588	0.079005	0.006095	0.000254	
	MAH	0.419325	0.096318	0.380198	0.091719	0.011974	0.000466
	MSH	0.409411	0.095321	0.381386	0.100111	0.013247	0.00052
T=8	AIC	0.277649	0.055133	0.513691	0.132628	0.019445	0.001453
	Sc	0.295181	0.056332	0.504421	0.125163	0.017636	0.001266
	HQ	0.162514	0.042185	0.513799	0.17341	0.098474	0.009618
	MAS	0.755454	0.053037	0.174712	0.015948	0.000827	
	MAH	0.208352	0.047303	0.503908	0.148751	0.08447	0.007216
	MSH	0.230448	0.051291	0.535644	0.15501	0.025473	0.00213
T=9	AIC	0.310858	0.062705	0.403941	0.193343	0.026331	0.002821
	Sc	0.360368	0.065866	0.384457	0.166737	0.020575	0.001997
	HQ	0.215481	0.053774	0.428554	0.253767	0.042756	0.005667
	MAS	0.809232	0.054412	0.116839	0.018641	0.000846	
	MAH	0.260899	0.058536	0.41942	0.22329	0.033824	0.00403
	MSH	0.283537	0.060555	0.413008	0.209298	0.030179	0.00342
T=10	AIC	0.261627	0.05341	0.473008	0.181694	0.02696	0.0033
	Sc	0.33171	0.05821	0.443134	0.14632	0.018663	0.001964
	HQ	0.194842	0.046958	0.490939	0.222628	0.038998	0.005635
	MAS	0.788405	0.050897	0.14254	0.017314	0.000812	3.14E-05
	MAH	0.226774	0.050301	0.484016	0.20201	0.032568	0.004331

	MSH	0.258434	0.053147	0.474142	0.183471	0.027424	0.00338
T=50	AIC	0.023396	0.00514	0.004772	0.003347	0.633822	0.329523
	Sc	0.545665	0.046086	0.016448	0.004435	0.322843	0.064524
	HQ	0.101051	0.015427	0.009952	0.00485	0.638177	0.230543
	MAS	0.954939	0.02967	0.003896	0.000386	0.010348	0.000761
	MAH	0.049612	0.009086	0.007031	0.004111	0.64893	0.28123
	MSH	0.274711	0.031193	0.014967	0.005426	0.531017	0.000126
T=100	AIC	0.086581	0.019172	0.036778	0.554139	0.128102	0.175229
	Sc	0.812206	0.048889	0.025493	0.104411	0.006561	0.00244
	HQ	0.342512	0.044769	0.050692	0.450837	0.061518	0.049671
	MAS	0.968092	0.021437	0.004112	0.006196	0.000143	
	MAH	0.185852	0.031619	0.0466	0.539434	0.095808	0.100687
	MSH	0.614564	0.054512	0.041887	0.252802	0.023409	0.01283
T=1000	AIC	0.691688	0.154221	0.077645	0.057921	0.012946	0.005579
	Sc	0.980345	0.018789	0.000813			
	HQ	0.900294	0.07899	0.01565	0.004594	0.000404	
	MAS	0.992885	0.007001	0.000111	0	0	0
	MAH	0.827546	0.115745	0.036555	0.017106	0.002399	0.000648
	MSH	0.956619	0.039228	0.003632	0.000498	2.05E-05	1.62E-06
T=2000	AIC	0.354081	0.078976	0.030043	0.024968	0.291331	0.2206
	Sc	0.986292	0.013372	0.000309			
	HQ	0.900132	0.071801	0.009768	0.002903	0.012115	0.003281
	MAS	0.994994	0.004963				
	MAH	0.750924	0.100163	0.022786	0.011325	0.07902	0.035783
	MSH	0.965822	0.031761	0.001781	0.000218	0.000375	0.00004

Table 2: Probability of correctly estimated the true lag length of AR process ($\hat{p} \leq 5$)

Sample size	Lag length selection criteria					
	AIC	SIC	HQ	MAS	MAH	MSH
3	0.823938	0.748864	0.654144	0.890178	0.748435	0.522823
4	0.859472	0.818179	0.757227	0.924427	1.00000	0.789314
5	0.481211	0.418718	0.501568	0.805199	0.43880	0.465483
6	0.539077	0.54264	0.934603	0.54264	0.539535	0.535604
7	0.39879	0.48890	0.419325	0.852054	0.419325	0.409411
8	0.513691	0.504421	0.513799	0.755454	0.503908	0.535644
9	0.403941	0.384457	0.428554	0.809232	0.41942	0.413008
10	0.473008	0.443134	0.490939	0.788405	0.484016	0.474142
50	0.633822	0.545665	0.638177	0.954939	0.64893	0.531017
100	0.554139	0.812206	0.450837	0.968092	0.539434	0.614564
1000	0.691688	0.980345	0.900294	0.992885	0.827546	0.956619
2000	0.354081	0.986292	0.900132	0.994994	0.750924	0.965822

Table 3. Probability of over and under estimated the true lag length of AR process ($\hat{p} \leq 5$)

Sample size and estimate		Order selection criteria					
		AIC	SIC	HQ	MAS	MAH	MSH
3	**over**	0.176062	0.251136	0.345856	0.109822	0.251565	0.477177
	under						
4	**over**	0.140525	0.181817	0.242766	0.075573	0.0000	0.210681
	under						
5	**over**	0.518789	0.112533	0.187283	0.311148	0.126074	0.324723
	under	-	0.468748	0.238853		0.435126	0.139673

6	**over**	0.234905	0.269402	0.051703	0.383189	0.305882	0.324723
	under	0.226018	0.187956	0.013694	-	0.154582	0.139673
7	**over**	0.60121	0.511085	0.580675	0.147943	0.580675	0.590585
	under						
8	**over**	0.153526	0.144066	0.281502	0.244524	0.240437	0.240437
	under	0.332783	0.351513	0.204699	-	0.255655	0.281739
9	**over**	0.222495	0.189309	0.30219	0.190768	0.261144	0.242897
	under	0.373564	0.426233	0.269255	-	0.319436	0.344092
10	**over**	0.211955	0.166946	0.267261	0.211595	0.238909	0.214275
	under	0.315037	0.38992	0.2418	-	0.277075	0.311581
50	**over**	0.329523	0.454335	0.230543	0.045061	0.28123	0.000126
	under	0.023396	-	0.13128	-	0.06984	0.326297
100	**over**	0.303331	0.187794	0.11119	0.031889	0.196495	0.38544
	under	0.142531	-	0.437973	-	0.26407	-
1000	**over**	0.308312	0.019602	0.099638	0.007112	0.172454	0.04338
	under	-	-	-	-	-	-
2000	**over**	0.645919	0.013681	0.099868	0.004963	0.034175	0.034175
	under	-	-	-	-	-	-

Table 4: Estimated lag length of AR process, ($\hat{p}$ <=5).

Sample size	Lag length selection criteria					
	AIC	SIC	HQ	MAS	MAH	MSH
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	2	2	0	2	2
6	2	2	2	0	2	2
7	0	0	0	0	0	0
8	2	2	2	0	2	2
9	2	2	2	0	2	2
10	4	0	4	0	4	4
50	3	0	3	0	3	0
100	0	0	0	0	0	0
1000	0	0	0	0	0	0
2000						

Table 5. Confidence set for Kullback-Leibler best model based on information criteria differences ($\bigwedge_i$)

Sample size	AR orders	Information Criteria					
		AIC	SC	HQ	MAS	MAH	MSH
T=10	**0**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	4.8622	4.9541	4.3430	6.9541	4.6026	4.6485
	2	3.7451	3.6808	2.4696	7.6808	3.1074	3.0752
	3	8.9637	8.4448	6.6215	14.4448	7.7926	7.5331
	4	17.9854	16.6416	14.1493	24.6416	16.0673	15.3954
	5						
T=50	0	0	0	0	0	0	0
	1	4.413467	6.285084	5.120865	**8.285084**	4.767166	5.702975
	2	6.042664	9.744244	7.435943	13.74424	6.739303	8.590093
	3	8.299879	13.78845	10.35666	19.78845	9.32827	12.07255
	4	0.337637	7.568821	3.034628	15.56882	1.686133	5.301725
	5	3.524631	12.45258	6.837555	22.45258	5.181093	9.645069
T=100	0	0	0	0	0	0	0
	1	4.184122	6.769192	5.229742	4.706932	4.706932	5.999467
	2	4.127142	9.276874	6.209481	5.168312	5.168312	7.743178
	3	0.149258	7.842932	3.259247	1.704253	1.704253	5.55109
	4	4.352086	14.56866	8.480484	6.416285	6.416285	11.52457
	5	5.171234	17.88933	10.30862	7.739928	7.739928	14.09897
T=1000	0	0	0	0	0	0	0
	1	4.022103	8.927857	5.886813	4.954458	4.954458	7.407335
	2	6.417443	16.22695	10.14628	8.281862	8.281862	13.18661
	3	8.03107	22.74232	13.62346	10.82726	10.82726	18.18289
	4	12.04929	31.66027	19.50464	15.77697	15.77697	25.58246
	5	14.76033	39.26903	24.07807	19.4192	19.4192	31.67355
T=2000	0	0	0	0	0	0	0
	1	3.993893	9.593796	6.050164	5.022029	5.022029	7.82198
	2	6.920652	18.11946	11.03293	8.976791	8.976791	14.57619
	3	8.287323	25.08403	14.45534	11.37133	11.37133	19.76969
	4	4.382939	26.77654	12.60644	8.49469	8.49469	19.69149
	5	5.942231	33.93172	16.22095	11.08159	11.08159	25.07634

References

[1]. Shibata. 1976. "Selection of the Order of an Autoregressive Mode by Akaike information Criterion", Biomertica, 63, pp.117-126.
[2]. Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control.1st Edn, Holden-Day, San Francisco, and ISBN: 0816211043, pp: 575.
[3]. Akaike, H .1974. "Information Theory and an Extention of the Maximum Likelihood Principle" In B.N Petror and F. Csaki, ed. 2nd, International symposium on Inference Theory,
[4]. Hannan, E.J and Quinn B.G (1979): "The Determination of the Order of an Auto regression", JRSS, B41, 190-195.
[5]. Rissanen, J .1978. "Modelling by Shortest data Description" Automatica, 14, pp. 445- 71.
[6]. Schwarz.1978. "Estimating the Dimension of a Model" Ann. Statist., 6, pp. 461-464.
[7]. Stone, M.1979. "Comments on Model Selection Criteria of Akaike and Swartz" J.R.Statistics, B41, No.2, pp. 276-278.
[8]. Al-Subaihi, Ali.A. 2007. "Variable Selection in Multivariate Regression using SAS / IML, Saudi Arabia.
[9]. McQuarrie, A.D and Tsai .1998. "Regression and Time Series Model Selection", World Scientific Publishing Co. Plc, LTD, River Edge, N.J
[10]. Lutkepohl, H. 1985. "Comparison of Criteria for estimating the Order of a Vector Autoregressive Process" Journal of Time Series Analysis vol. 6, 35 – 36, 47, 50.
[11]. Linhart, H. and Walter Zucchini. 1986. Model Selection. New York. Wiley.
[12]. Sakamoto, Y., Ishiguro, M. and Kitagawa, G.1986. Akaike Information Criterion Statistics, D.Reidel, Dordrecht.
[13]. DeLeeuw, Jan. 1992. "Introduction to Akaike (1973) Information Theory and an extension of the Maximum Likelihood Principle." Pp. 599-609, in Breakthroughs in Satistics, vol. 1, edited by Samuel Kotz and Norman L. Johnson. London: Springer-Verlag.
[14]. Kullback, Soloman and Richard A. Leibler. 1951. "On Information and Sufficiency."Annals of Mathematical Statistics 22:79-86.
[15]. Parzen, Emmanuel, Kunio Tanabe, and Genshiro Kitagawa. (Eds.). 1998. SelectedPapers of Hirotugu Akaike. New York: Springer-Verlag.
[16]. Sugiura, Nariaki. 1978. "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections." Communications in Statistics, Theory and MethodsA7:13-2626.
[17]. Burnham, Kenneth P. and David R. Anderson. 2002. Model Selection and Multimodal Inference: A Practical Information-Theoretical Approach. 2nd Ed. New York: Springer-Verlag.