Equal-Weighted Information Criterion for Model Selection in Univariate Time Series Analysis

S. Ankrah*, B.L. Peiris[1] and R.O. Thattil[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Sri Lanka

ABSTRACT: *In model building, the model with appropriate number of parameters needs to be identified. Thus, a variety of information criteria have already been developed, each with a different background to handle this challenge. The mostly used information criteria are the Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), and Hannan and Quinn Information Criterion (HQ). However, available literature and the preliminary analysis done by authors indicated that when selecting the appropriate model, these information criteria usually lacked uniformity. Thus, in this study, an information criterion that serves as a unifier to the three commonly used criteria; AIC, SIC and HQ is proposed. The penalties of these three information criteria are considered as a linear function. Simulations were conducted on the performance of the proposed information criterion (PIC) together with the three conventional information criteria using nine models and seven different sample sizes. The results revealed that the proposed information criterion (PIC) performed better than the AIC, SIC and HQ with respect to the overall performance in choosing the true model. The performance of PIC increased as sample size increased. However, PIC turns to under fit, when the true model is not selected. When sample size is large, PIC is asymptotically robust with respect to single processes, Autoregressive (AR) and Moving Average (MA). Thus, the proposed information criterion is recommended when selecting the order of a univariate time series.*

Keywords: *Information criteria, model selection, robust and model size, sample size*

INTRODUCTION

In model building, the focus is that, there is information in the observed data, and we want to express this information in a compact form through a "model", (Burnham and Anderson, 2002). Thus, the goal of model selection is to attain a perfect 1-to-1 translation such that no information is lost in going from the data to a model of the information in the data. However, such models (true models) do not exist in the real world. Thus, we can say that models are only approximations. However, we can attempt to find a model for the data that is "best" or close to the true model (the model loses as little information as possible). This thinking leads directly to Kullback–Leibler information (K-L). Thus, we wish then to select a model that minimizes K-L information loss as the best model for inference.

[1] Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka
* Corresponding author: sampson.ankrah@yahoo.com

According to Mees (1993) and Burnham *et al.,* (2002), information criteria are used to identify the model that minimizes K-L information loss. These information criteria are based on the K-L discrepancy, which is a class of loss function. The K-L discrepancy between two models *f* and *g*, *I (f, g),* can be written equivalently as:

$$I(f,g) = \int f(x)\log(f(x))dx - \int f(x)\log(g(x/\theta))dx, \quad 1$$

The term on the right is a statistical expectation with respect to *f* (the true model). Thus, the K-L distance in Equation 1 can be expressed as a difference between two statistical expectations;

$$I(f,g) = E_f\left[\log(f(x))\right] - E_f\left[\log(g(x/\theta))\right], \quad 2$$

each with respect to the distribution *f.* The first expectation of Equation 2 is a constant that depends only on the unknown true distribution (Burnham and Anderson, 2002). When the second expectation of Equation 2 is computed, this $I(f,g)$ could be estimated up to a constant C;

$$I(f,g) = C - E_f\left[\log(g(x/\theta))\right], \quad 3$$

The term $\left[I(f,g) - C\right]$ is a relative directed distance between *f* and *g.* Thus, $E_f\left[\log(g(x/\theta))\right]$, becomes the quantity of interest for selecting a best model. This is to say, the model with the minimum *I (f, g),* is considered as the best model.
The information criteria in this study uses log-likelihood function as the K-L information. Generally, information criteria have two main parts; it is of the form;

$$IC = Deviance + p(n,k) \quad 4$$

Here, the deviance is the loss function, which assesses the quality of the model to the data. ***P (n, k)*** is the model complexity, which penalize the model order and grows as the number of parameters increase and it is dependent on sample size (***n***) and the number of fitted parameters (***k***).

A variety of information criteria have been developed for this purpose. The information criteria that are of interest to this study are Akaike Information Criterion, AIC (Akaike, 1974), Schwarz Information Criterion, SIC (Schwarz, 1978) and Hannan and Quinn Information Criterion, HQ (Hannan and Quinn, 1979).

All these criteria have relative advantages depending upon the situation in which they are used. However, in the available literatures and our preliminary analysis, a simple comparison of the AIC, SIC and HQ shows that, they often disagree and when they disagree, the AIC would choose the largest, the SIC the smallest and the HQ the model in between. Thus, there is little uniformity in deciding the model size of a time series analysis. Here, we propose an information criterion that minimizes this challenge. The basis of this proposed information criterion is that, its takes into consideration the average of penalty terms of the AIC, SIC and HQ. The proposed information criterion is validated with the conventional information criteria by studying the asymptotic and non-asymptotic properties of these criteria.

METHODOLOGY

Proposed information criterion (PIC)

The proposed information criterion, as indicated below combines the strength of AIC, SIC and HQ. The AIC is expressed as:

$$AIC = -2\ln L(\hat{\theta}) + m \quad (5)$$

where $m = 2k$. The SIC can also be written as:

$$SIC = -2\ln L(\hat{\theta}) + y \quad (6)$$

where $y = k\ln(n)$. The HQ is expressed as:

$$HQ = -2\ln L(\hat{\theta}) + x \quad (7)$$

Where $x = 2k\ln\ln(n)$; $L(\hat{\theta})$ is the likelihood function of the fitted model, $-2\ln L(\hat{\theta})$ is the deviance, m, y, x are the penalty for AIC, SIC and HQ, respectively, k = number of free parameters in the model and n = number of observations. (McQuarrie and Tsai, 1998; Box and Jenkins, 1994).

Thus, the proposed criterion, PIC, can be defined as:

$$PIC = -2\ln\mathrm{L}(\hat{\theta}) + (m + y + x)/3 \quad (8)$$

The penalty of the proposed criterion can be expressed in the following form:

$$penalty = k\left[(2 + \ln(n) + 2\ln\ln(n))/3\right] = \varpi \quad (9)$$

Thus, the proposed criterion can be simply written as:

$$PIC = -2\ln\mathrm{L}(\hat{\theta}) + k\left[(2 + \ln(n) + 2\ln\ln(n))/3\right]$$

$$PIC = -2\ln\mathrm{L}(\hat{\theta}) + \varpi \quad (10)$$

Here, the penalty term of the proposed information criterion, ϖ, is the average of the penalty terms of AIC, SIC and HQ. The asymptotic properties among penalty terms are indicated below:

a) As $n \to \infty$, the penalty terms $m \to 2k$ (constant); $y \to \infty$; and $x \to \infty$. Hence, $\varpi \to \infty$. This indicates that the ϖ(penalty of PIC), behaves like SIC and HQ when the sample size, n, is very large.

b) As $n \to 0$, the penalty terms $m \to 2k$ (constant); $y \to 0$; and $x \to 0$. Hence, $\varpi \to 2k$. This means, the ϖ(penalty of PIC), behaves like AIC when the sample size, n, is very small.

Efficiency and consistency

Let $\mu = E\{Y\}$. Let f be the true model and g be the model selected as the best model using an information criterion. The selection criterion is said to be asymptotically loss efficient if convergence in probability (Shao, 1997) is as given below.

$$n \xrightarrow{\lim} \infty \ \frac{L(\mu, \mu(g))}{L(\mu, \mu(f))} = 1 \qquad 11$$

The selection criterion is said to be consistent if $n \xrightarrow{\lim} \infty \ \Pr\{g = f_0) = 1$. The terms over-fitting and under-fitting were defined as two ways based on either consistency or efficiency (McQuarrie and Tsai, 1998). Using efficiency, over-fitting is defined as choosing a model that has more variables than f_o. Under-fitting is defined as choosing a model with too few variables compared to f_o.

Simulation Study for the proposed information criterion

In order to validate the performance of the proposed information criterion with other conventional information criteria, a simulation study was conducted on different model structure of univariate time series processes.

Precisely, nine simulated datasets were generated from AR(*p*), *p = 1, 2,3*; and MA(*q*), *q*= 1, 2, 3; and ARMA (*p, q*), *p, q* = 1, 2, 3; with 10 replicates and seven different sample sizes. The set of these sample sizes is to examine the influence of sample size on the ability of an information criterion to correctly select the true model as sample size increases.

Here, the underlined models or true models of the datasets are known, since the datasets were generated from true models. Thus, the generated datasets in each case are fitted to four (4) different models including the true model of which the datasets were generated. In our study, for the AR(*p*) dataset, we fitted four models i.e., AR(1), AR(2), AR(3) and AR(4); and for the MA(*q*) datasets; MA(1), MA(2), MA(3) and MA(4) were fitted; while ARMA(1,1), ARMA(2,2), ARMA(3,3) and ARMA(4,4) were fitted to the ARMA(*p, q*) datasets. These fitted models were selected so that, we can study the non-asymptotic properties of under-fitting and over-fitting. For each dataset, the best fitted model was identified by the minimum criterion estimate of the four models. The model selected for each criterion per dataset was then recorded and the frequency of a criterion selecting the correct model or true model was tallied, since we know the model that generated the dataset. The criterion which selected the correct model most often was then considered the best criterion for the analysis.

RESULTS AND DISCUSSION

Measuring the stability of information criteria

The results of the simulation study indicating the models, different sample sizes and IC performances are reported in the respective tables. The tabulated results are the percentage or probability of correctly selecting the true model by the criterion. For clarity of interpretations, we define five subjective performance capability categories. These categories show the ability of a criterion to correctly identify the true model from which the specific data was generated.

The first performance rating is defined as “very good” and the assigned percentage is [90 – 100] of times a criterion selects the correct model. The second rating is “good” and the assigned percentage is [75% - 90%]. The third performance rating is “acceptable” and the assigned percentage is [60 – 75%]. The fourth rate is “poor” with the assigned percentage as [45% - 60]. The fifth rate is “unacceptable” with assigned percentage as [0 – 45%].

Non-asymptotic properties

Here, the focus is to examine; (1) the performance of the individual information criterion, with respect to the probability that the information criterion selects the true model; (2) the overall performance of the four information criteria considered in this study. This is done by ranking the score of the frequency of an information criterion selecting the true model with respect to the weights assign to the performance capabilities categories; (3) the performance of the information criteria as model order increases; (4) under-fitting and over-fitting of information criteria.

Performance of individual information criterion

Here, the performance of AIC, SIC, HQ and the proposed information criterion, (PIC), according to the simulation study are evaluated. Table 1 gives the percentage of selecting the true model using AIC. In the autoregressive processes, AIC performance gets better as sample size increases. However, it's performance in the moving average processes is not consistent as sample size increases. It is obvious that, AIC performed relatively well in both the autoregressive and moving average processes. However, its performance with respect to the mixed processes was not recommendable except when the sample size, n = 1200.

Table 1. Percentage of selecting the true model using AIC

Model	Sample Size						
	5	**15**	**25**	**45**	**100**	**500**	**1200**
AR(1)	80	80	80	80	90	90	100
AR(2)	30	30	50	60	80	80	90
AR(3)	50	50	60	60	90	90	100
MA(1)	30	40	40	60	40	60	90
MA(2)	30	40	40	30	70	70	90
MA(3)	10	10	40	30	50	90	100
ARMA(1,1)	40	60	80	80	50	60	90
ARMA(2,2)	10	20	20	20	50	70	90
ARMA(3,3)	0	0	10	20	0	60	80

Table 2 gives the percentage of selecting the true model using SIC.

Table 2. Percentage of selecting the true model using SIC

Model	Sample Size						
	5	**15**	**25**	**45**	**100**	**500**	**1200**
AR(1)	80	90	100	100	100	100	100
AR(2)	10	20	40	80	100	100	100
AR(3)	20	30	30	50	80	100	100
MA(1)	50	60	70	100	90	100	100
MA(2)	10	30	40	30	70	100	100
MA(3)	10	10	20	10	20	100	100
ARMA(1,1)	60	70	90	100	90	100	100
ARMA(2,2)	10	10	0	20	40	100	100
ARMA(3,3)	0	0	0	0	0	60	100

The performance of SIC was excellent in AR(1) for all sample sizes but performed poorly in AR(2) and AR(3) with small sample sizes. However, SIC performance was better as sample size increased. Thus, SIC performance was consistent as sample size increased; and better with large sample sizes. It is obvious that SIC performance decreased as the order of both single and mixed processes increases. The percentage of selecting the true model using HQ criteria is given in Table 3.

Table 3. Percentage of selecting the true model using HQ criteria

Model	Sample Size						
	5	**15**	**25**	**45**	**100**	**500**	**1200**
AR(1)	70	80	80	90	100	100	100
AR(2)	10	30	50	80	90	100	100
AR(3)	50	60	60	60	80	100	100
MA(1)	30	40	40	80	60	80	100
MA(2)	30	40	40	30	70	80	100
MA(3)	10	10	30	20	40	100	100
ARMA(1,1)	50	60	80	90	90	100	100
ARMA(2,2)	10	20	20	50	40	100	100
ARMA(3,3)	0	0	0	0	10	70	100

HQ performed very well with very large sample size in all models. It is obvious that HQ performed well when the order of a process is small. The performance of HQ is consistent with respect to the autoregressive processes. The performance of HQ is unacceptable for the ARMA(3,3) with small sample size. However, HQ performed relatively well, overall. The percentage of selecting the true model using PIC is given in Table 4. The performance of PIC was better in AR(1) for all sample sizes but performed poorly in AR(2), AR(3), MA(3), ARMA(2,2) and ARMA(3,3) with small sample sizes, n = 5 and 15. Generally, PIC performance was better as sample size increases. Thus, PIC performance was consistent as sample size increases in AR(1), AR(2), AR(3), MA(1), MA(2) and ARMA(1,1). In other words, PIC does very well with large sample sizes, but is not consistent as model or process order increases.

Table 4. Percentage of selecting the true model using PIC

Model	Sample Size						
	5	15	25	45	100	500	1200
AR(1)	90	90	90	100	100	100	100
AR(2)	10	30	50	80	90	100	100
AR(3)	20	40	60	70	80	100	100
MA(1)	30	50	50	90	60	90	100
MA(2)	40	50	40	30	70	90	100
MA(3)	10	10	20	30	40	100	100
ARMA(1,1)	60	70	80	90	90	100	100
ARMA(2,2)	10	20	20	50	40	100	100
ARMA(3,3)	0	0	0	0	10	60	100

Overall performance of information criterion

A weighted ranking scale based on the performance categories was proposed in order to compare the overall performance of each criterion with respect to other criteria. The overall performance rating summary of Information criteria is given in Table 5. A decreasing weight was assigned to a decreasing performance category. At each criterion, the number of times that the criterion occurred with respect to the five categories is multiplied by their respective weights. The resultant is summed across categories as the overall performance score of the criterion. Then these scores are ranked with a rank of 1 given to the highest score.

In terms of overall relative performance of criteria, the PIC is ranked as the highest performed information criterion with respect to the probability of selecting the true model. The second performed information criterion is the SIC; followed by HQ and lastly AIC. Thus, we can say that the proposed information criteria (PIC) can consistently select the true model than most of the conventional methods. Therefore, in terms of correctly selecting the true model of an observed data, we recommend the use of the PIC.

Table 5. Overall performance rating summary of information criteria

Weight (w_i)	4	3	2	1	0	Score (#)	Rank
	Performance Capabilities Categories						
Criterion	V. Good	Good	Acceptable	Poor	Unacceptable		
AIC	13	9	11	6	24	107	4
SIC	27	3	6	2	25	131	2
HQ	20	8	8	4	23	124	3
PIC	26	3	7	5	22	132	1

$score = \sum w_i \xi_i$, where ξ_i is the frequency of a criterion with respective to performance capabilities categories.

Performance of information criteria as the order of the processes increases

In this section, we assess the performance of information criteria as the order of processes increases. The performance of criteria as model order increases is given in Table 6. We consider two sample sizes, n=25 (small sample) and n=45 (large sample) for each model, the highest performing criterion is shaded.

It is obvious that, the performances of information criteria are not consistent as the order increase in all the models. However, when sample size n=25, the AIC has the highest number of best performance (given by the 6 shaded values) across the 9 models. When sample size n=45, PIC has the highest number of high performance (the 6 shaded values) across the 9 models, which was followed by the SIC (recording 5 shaded values). Fig. 1.Shows the performance of different criteria as model order increases for n=25.

Table 6. Performance of criteria as model order increases

	Model								
	AR(1)	**AR(2)**	**AR(3)**	**MA(1)**	**MA(2)**	**MA(3)**	**ARMA(1,1)**	**(2,2)**	**(3,3)**
	N=25								
AIC	80	50	60	40	40	40	80	20	10
SIC	100	40	30	70	40	20	90	0	0
HQ	80	50	60	40	40	30	80	20	0
PIC	90	50	60	50	40	20	80	20	0
	N=45								
AIC	80	60	60	60	30	30	80	20	20
SIC	100	80	50	100	30	10	100	20	0
HQ	90	80	60	80	30	20	90	50	0
PIC	100	80	70	90	30	30	90	50	0

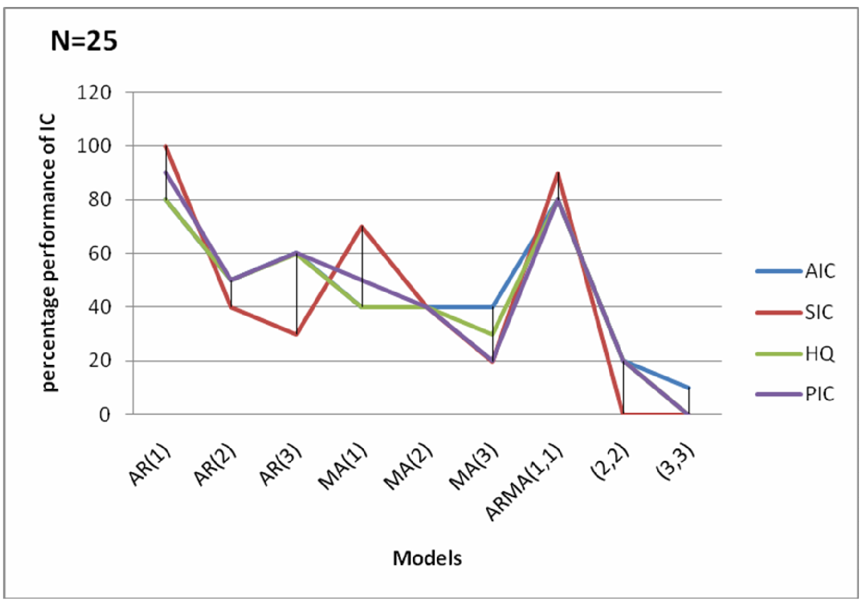

Fig. 1. Performance of criteria as model order increases for n=25

When sample size, n=25, all information criteria performed well at both AR (1) and ARMA (1,1). It is obvious that, SIC performed poorly at high order of mixed models or processes; and AR (2) and AR (3) processes. However, SIC performed well at AR (1), MA (1) and ARMA (1,1). Thus, with small sample sizes, SIC does well at smaller order of processes. PIC did not perform well at MA (2), MA (3) and ARMA (3,3). However, PIC performed well at AR (2), AR (3), ARMA (1,1) and ARMA (2,2). Figure 2 gives the performance of different criteria when model order increases for n=45.

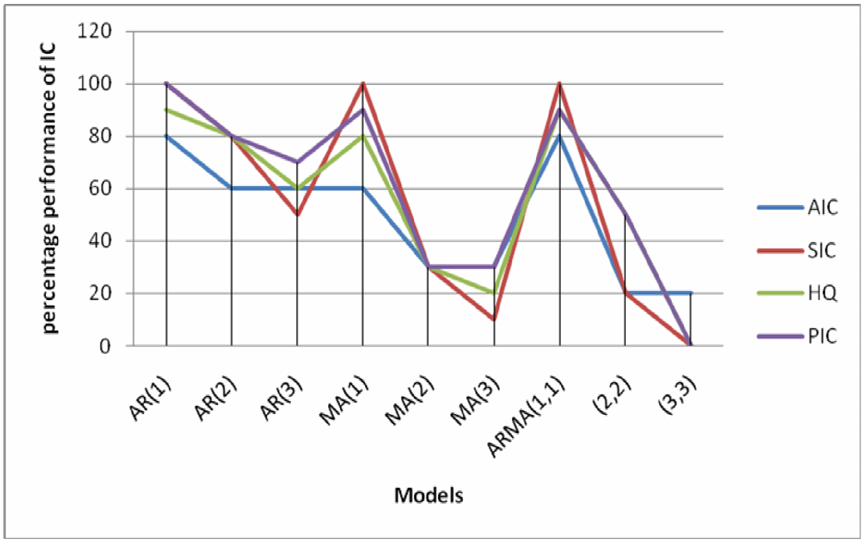

Fig. 2. The performance of criteria as model order increases for n=45

When sample size, n=45, all information criteria performed well at both AR(1) and ARMA(1,1). It is obvious that, SIC performed poorly at high order of mixed models or processes; AR(3) and MA(3) processes. However, SIC performed well at AR(1), MA(1) and ARMA(1,1). Thus, with large sample sizes, SIC does well at smaller order of processes. PIC did not perform well at MA(2), MA(3) and ARMA(3,3). However, PIC performed well at AR(2), AR(3), MA(1), MA(2), MA(3), ARMA(1,1) and ARMA(2,2).

Under-fitting

We examine the percentage or the probability of an information criterion to choose a model with few parameters compared to the true model. Our interest is to identify which information criterion has the possibility of under-fitting under a given situation.

We follow the same weight assigned to the performance capabilities categories. However, we define different performance capability categories this time. The categories are as follow: [0-10%] is very good; [20-30%] is good; [40-50%] is acceptable; [60-69%] is poor; [70-100%] is unacceptable. The results are presented in Table 7.

Table 7. Under-fitting performance rating summary of information criteria

Weight (w_i)	4	3	2	1	0	Score (#)	Rank
	Performance Capabilities Categories						
Criterion	V. Good	Good	Acceptable	Poor	Unacceptable		
AIC	26	5	5	3	6	132	1
SIC	21	4	4	3	13	107	4
HQ	23	9	3	2	9	127	2
PIC	22	8	4	2	8	122	3

AIC is ranked as the criterion which has the lowest probability of under-fitting a true model, followed by HQ. Thus, SIC and PIC turn to underfit more than the AIC and HQ when they do not select the true model.

Over-fitting

We examine the percentage or the probability of an information criterion to choose a model that has more parameters compared to the true model. Our interest is to identify which information criterion has the possibility to over-fit.

We follow the same weight assigned to the performance capabilities categories. However, we defined different performance capabilities categories due to our interest at this time. The categories are as follow: [0-10%] is very good; [20-30%] is good; [40-50%] is acceptable; [60-69%] is poor; [70-100%] is unacceptable. The results are presented in Table 8.

Table 8. Over-fitting performance rating summary of information criteria

Weight (w_i)	4	3	2	1	0	Score (#)	Rank
	Performance Capabilities Categories						
Criterion	V. Good	Good	Acceptable	Poor	Unacceptable		
AIC	16	17	8	4	0	135	4
SIC	39	5	1	0	0	173	1
HQ	29	11	3	2	0	157	3
PIC	34	8	3	0	0	166	2

SIC is ranked as the criterion which has the lowest probability of over-fitting, when it does not select the true model, followed by PIC. Thus, AIC and HQ turn to over-fit more than the SIC and PIC, when they do not select the true model.

Asymptotic properties

Here, the focus is to examine; (1) the performance of the information criteria with respect to increase in the sample sizes; and (2) the asymptotic robustness of the proposed information criterion, (PIC).

Performance of information criteria as sample size increases

In other to examine the information criteria performance as sample size increases, we took the average performance of all the nine (9) models in each case of the four (4) information criteria. We then ranked the average performance of each information criteria with respect to each sample size. In each sample size, the highest performed information criterion is shaded. The results are presented in Table 9. AIC performed better when sample size is small. However, it performed relatively poor when sample size is large. Its individual average performance increases as sample size increases. SIC, relatively performed poorly when the sample size is small, but performed excellently well when sample size is small. HQ relative average performance is not consistent with sample size increase. However, it does well when sample is very large. PIC performed well as the sample sizes increased. Its relative rank performance is second.

Table 9. Average performance of information criteria as sample size increases

Criterion	Sample Size						
	5	15	25	45	100	500	1200
AIC	31.1	36.7	46.7	48.9	57.8	74.4	92.2
SIC	27.8	35.6	43.3	54.4	65.6	95.6	100
HQ	28.9	37.8	44.4	55.6	64.4	92.2	100
PIC	30	40	45.6	60	64.4	93.3	100

We represent this information graphically. The performance of IC as sample size increases is given in Fig. 3.

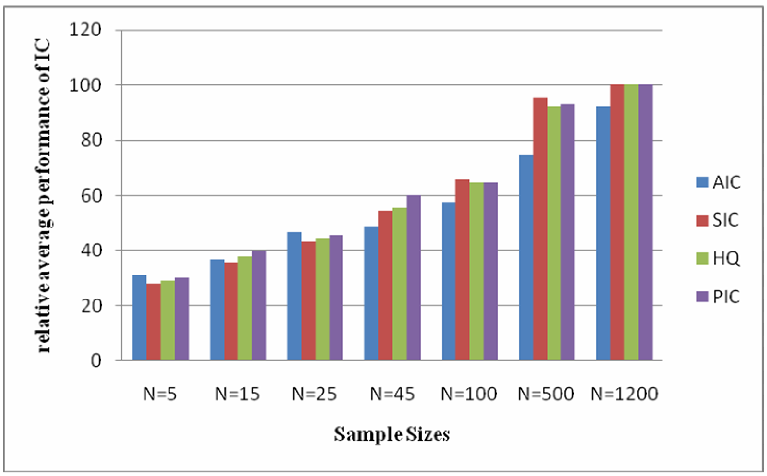

Fig. 3. Performance of IC as sample size increases

It is obvious that, when sample size is less than 45, the AIC performed better but its performance was always poor when the sample size is greater than 25. The performance of

SIC is outstanding when the sample is large. The performance of PIC was good. Figure 4 shows the performance of the information criteria (IC) as sample size increases.

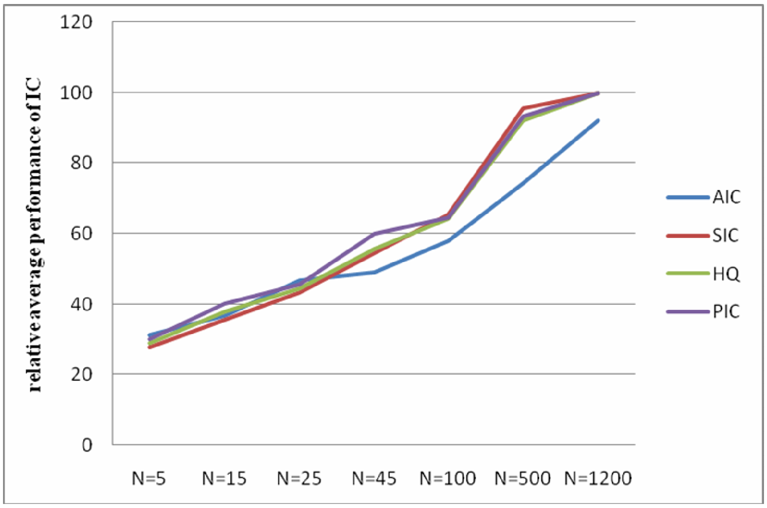

Fig. 4. The performance of IC as sample size increases

The PIC was placed first or second when sample size was less than 45. However, for n=45 the PIC was placed first. The PIC closely followed the SIC when sample size increased from 100 to 1200. The AIC was the least performed criterion when sample size is greater than 25. It is obvious that PIC and SIC performed better as the sample size increased asymptotically.

Robust nature of the proposed information criterion, (PIC)

The standard normal error or innovation of a series is N(0,1), to examine the robustness of the proposed criterion, we simulated two additional datasets under AR(1), MA(1) and ARMA(1,1) by varying the variance of the error or innovation, [N(0,2) and N(0,3)]. The focus is to identify if the proposed information criterion selects the true model in these three scenarios. Table 10 shows the robustness of the proposed IC for different error terms.

Table 10. Robustness of the proposed information criterion for different error terms

Model	**Sample Sizes**			
	15	30	45	1200
AR(1)		$\sigma^2 = 1$		
(1,0,0)	42.34	89.49	136.82	3323.73
(2,0,0)	44.39	91.79	139.45	3328.04
(3,0,0)	46.47	92.40	139.65	3331.54
		$\sigma^2 = 2$		
(1,0,0)	70.64	131.25	199.42	4987.45
(2,0,0)	72.59	133.57	202.05	4991.78
(3,0,0)	70.45	134.20	202.29	4995.27
		$\sigma^2 = 3$		
(1,0,0)	82.88	155.65	235.98	5960.65
(2,0,0)	84.83	157.97	238.63	5964.96

(3,0,0)	82.73	158.60	238.89	5968.46
MA(1)		$\sigma^2 = 1$		
(0,0,1)	47.90	88.21	120.32	3490.83
(0,0,2)	48.59	90.57	123.15	3492.46
(0,0,3)	47.33	89.40	121.39	3496.74
		$\sigma^2 = 2$		
(0,0,1)	68.78	129.89	182.90	5154.49
(0,0,2)	69.39	132.23	185.67	5156.12
(0,0,3)	68.47	131.32	184.15	5160.40
		$\sigma^2 = 3$		
(0,0,1)	80.98	154.27	219.46	6127.65
(0,0,2)	81.57	156.57	222.25	6129.28
(0,0,3)	80.77	155.74	220.79	6133.56
		$\sigma^2 = 1$		
ARMA(1,1)				
(1,0,1)	49.91	103.51	140.17	3454.50
(2,0,2)	48.50	99.84	145.48	3461.70
(3,0,3)	53.16	103.19	136.25	3468.47
		$\sigma^2 = 2$		
(1,0,1)	66.43	139.11	195.47	5113.02
(2,0,2)	66.78	134.66	201.04	5120.14
(3,0,3)	70.98	138.95	192.23	5127.07
		$\sigma^2 = 3$		
(1,0,1)	77.47	161.71	230.07	6085.12
(2,0,2)	78.50	156.98	235.70	6092.32
(3,0,3)	79.56	161.63	227.19	6098.97

It is obvious that, as sample size approaches infinity, PIC correctly select the true model in the AR (p) and MA (q) processes. In context, as $N_{(30)} \rightarrow \infty$, PIC is asymptotically robust, in selecting the true model, with respect to autoregressive and moving average processes. Thus, PIC is asymptotically robust in selecting the true model under the autoregressive and moving average processes. However, this robustness does not work with respect to the mixed processes (ARMA), except when sample size is very large, say n = 1200.

CONCLUSION

We have investigated four information criteria including the proposed information criterion using 9 different models and 7 different sample sizes.

The results revealed that the proposed information criterion performed better than the SIC, AIC and HQ with respect to the overall performance in choosing the true model. The performance of PIC increased when sample size increased. However, PIC turns to under-fit, when the true model is not selected. When sample size is large, PIC is asymptotically robust with respect to single processes, AR(p) and MA(q). Thus, we recommend the proposed information criterion when selecting the order of a univariate time series.

REFERENCES

Akaike, H. (1974). A new look at the statistical identification model. IEEE Trans. Automat. Contr., *19,* 716 - 723.

Box, G. and Jenkins, G. (1994). Time Series Analysis, Forecasting and Control. John Wiley & Sons: New Jersey, USA.

Burnham, K.P. and Anderson, D.R. (2002). Model selection and multimodel inference: a practical information-theoretic approach, 2nd Ed, Springer-Verlag New York, Inc.

Chatfield, C. (2004). The Analysis of Time Series: An Introduction. Chapman & Hall/CRC: London, UK.

Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an auto regression. *Journal of the Royal Statistical Society*, Series B, *41*, 190 - 195.

McQuarrie, A.D.R. and Tsai, C. (1998). Regression and time series model selection. Singapore: World Scientific.

Mees, A. (2000). Nonlinear Dynamics and Statistics, ed. Mees, A.I. (Birkha¨user, Boston).

Nakamura, T., Mees, A.I. and Judd, K. (2003) "Refinements to model selection for nonlinear time series," ***Int. J. Bifurcation and Chaos, 13,*** 1263 - 1274.

Nakamura, T. and Small, M. (2006) "Modelling nonlinear time series using improved least squares method," ***Int. J. Bifurcation and Chaos 16,*** 445 - 464.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, *6,* 461 - 464.

Shumway, R.H. and Stoffer, D.S. (2011). Time series analysis and its applications: with R examples. 3rd ed. New York: Springer.

Tiao, G.C. and Box, G.E.P. (1981), "Modelling multiple time series with applications," *Journal of the American Statistical Association*, *76,* 802 - 816.

Wei, W. (2006). Time Series Analysis: Univariate and Multivariate Methods. Addison-Wesley: New York, USA.