Development of New Information Criterion For Model Order Determination in Time Series Modeling

O. I. Shittu[1]; E. A. Adeyemo[2]; O. S. Yaya[3]; F. J. Ayoola[3]; T. I. Shittu[4]

Department of Statistics,
University of Ibadan, Ibadan, Nigeria.
e-mail: oi.shittu@hotmail.com[1]; o.s.olaoluwa@gmail.com[3], lanreshittu535@gmail.com[4]

***Abstract*—Many researchers have used conventional model order selection criteria with considerable amount of success over the years. However, each of them has been identified with one limitation or the order with respect to their ability to correctly identify the correct order of dynamic models. In this paper, a new information criterion (NIC) that is capable of selecting (almost surely) the correct order of an Autoregressive model is proposed. The proposed technique is a hybrid of the well-known order selection criteria. The considered criteria are the Akaike information criterion (AIC); Bayesian information criterion (BIC) and the Hannan-Quinn criterion (HQ). Six series of sizes 20, 30 100, 200, 500, 1000, were simulated and two real data on money supply of size 30 and income data of size 467 were used to test the performance of the proposed criterion. The comparison of the four model selection criteria was in terms of the number of times that they identify the "true" lag length of a model. The results showed the new information criteria (NIC) perform best for small and large sample sizes**

Keywords- *Selection criterion, Dynamic model, Model order and Weight of evidence.*

I. INTRODUCTION

Generally, information criteria provide a means for measuring the relative quality of statistical model for a given set of data. It is founded on information theory that offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. It deals with the trade-off between the goodness of fit of the model and the complexity of the model. Thus given a collection of models for the data, preferred information criteria is used to estimates the quality of each model, relative to each of the other models.

The considered information criteria are the Akaike information criterion (AIC); Bayesian information criterion (BIC) and the Hannan-Quinn criterion (HQ).

Authors that have worked on information criterion include [3], [6], [10] and [11]. They discovered that the BIC perform best in choosing the best order of the model follow by HQ and lastly, the AIC perform least because BIC and HQ have a better order bias correction term of its penalty.

The BIC is an increasing function of σ_p^2 and an p. That is, unexplained variations in the dependent variable and the number of explanatory variables incase the value of BIC. Hence, lower BIC implies fewer explanatory variables, better fit, or both. The BIC generally penalizes free parameters more strongly than does the AIC, though it depends on the size of n and relative magnitude of *n* and *p*.

II MATERIALS AND METHODS

A Akaike Information Criteria (AIC)

The AIC is a measure of the relative quality of statistical model for a given set of data. It was developed by [1] and first published by in 1974. AIC does not provide any information about the quality of the model if all the candidate models are poorly fit. It takes the form of a penalized likelihood (a negative log likelihood plus a penalty term given by

$$AIC = -2\log L(\theta) + 2p$$

where $L(\theta)$ is the maximized likelihood function, and p is the number of estimated parameters in the model. The log likelihood $-2\log L(\theta)$ indicate the lack of fit component and k the penalty component that measures the level of complexity of the model or the compensation for the bias in the lack of fit when the maximum likelihood estimators are used.

B Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike Information Criterion (AIC). It is defined as:

$$BIC = -2\log L(\theta) + p\log p$$

where $L(\theta)$ is the maximized likelihood function. BIC is an increasing function of and an increasing function of p. It is noted that unexplained variation in the dependent variable and the number of explanatory variables increases the value of BIC.

C Hannan-Quinn Information Criterion (HQC)

In statistics, the HQC is a criterion for model selection. It is an alternative to AIC and BIC. It is given as:

$$HQC = -2\log L(\theta) + 2p\log p$$

where p is the number of parameters, n is the number of observations. [2] claimed that HQC, “while often cited, seems to have seen little use in practice” (p. 287). They also noted that HQC, like BIC, but unlike AIC, is not an estimator of Kullback-Leibler divergence. [3] note that HQC, like BIC, but unlike AIC, is not asymptotically efficient and further pointed out that whatever method is being used for fine-tuning the criterion will be more important in practice than the term log n, since this latter number is small even for very large n.

The purpose of this research is to propose a new information criteria by intrinsically convoluting the penalty between BIC and HQC; between AIC and HQC; and finally, between BIC, AIC and HQC.

Four information criteria shall be derived (NIC1, NIC2, NIC3 and NIC4) with four new order bias correction terms i.e. (Nip1, Nip2, Nip3 and Nip4) in which the one with the highest bias correction term shall be adopted as the new information penalty term. Base on the above equations we have the following derivations:

D Derivations of new information criterion one:

The penalties for different information criteria are given below:

$$BIC = \ln\sigma_p^2 + \frac{p\ln N}{N}$$

$$HQC = \ln\sigma_p^2 + \frac{2p\ln(\ln N)}{N}$$

$$AIC = \ln\sigma_p^2 + \frac{2p}{N}$$

Convoluting the penalties of BIC and HQC

This is achieved as follows:

$$NIC1 = \frac{1}{2}\left[BIC + HQ\right] = \frac{1}{2}\left[\ln\sigma_p^2 + \frac{p\ln N}{N} + \ln\sigma_p^2 + \frac{2p\ln(\ln N)}{N}\right]$$

$$NIC1 = \left[-2\left(\frac{l}{N}\right) + \frac{\frac{p}{2}\ln N}{N}\right]$$

Penalty for NIC1 is NIP1 = $\frac{\frac{p}{2}\ln N}{N}$

where

l is the log-likelihood function

p is the number of parameters estimated

N is the number of observation.

σ_p^2 is the maximum likelihood estimate of residual variance (after fitting $AR_{(p)}$)

Convoluting the penalty of BIC and AIC to have:

$$NIC2 = \frac{1}{2}\left[BIC + AIC\right] = \frac{1}{2}\left[\ln\sigma_p^2 + \frac{p\ln N}{N} + \ln\sigma_p^2 + \frac{2p}{N}\right]$$

$$NIC2 = \left[-2\left(\frac{l}{N}\right) + \frac{\frac{p}{2}(\ln N + 2)}{N}\right];$$

the penalty of (NIC 2) is NIP2 = $\frac{\frac{p}{2}(\ln N + 2)}{N}$

where the parameters are as defined before.

Convoluting the penalties of AIC and HQ

This is achieved as follows:

$$NIC3 = \frac{1}{2}\left[AIC + HQ\right] = \frac{1}{2}\left[\ln\sigma_p^2 + \frac{2p}{N} + \ln\sigma_p^2 + \frac{2p\ln(\ln N)}{N}\right]$$

to get

$$NIC3 = \left[-2\left(\frac{l}{N}\right) + \frac{p\left(1 + \ln(\ln N)\right)}{N}\right]$$

Hence,

Penalty for (NIC3) is NIP3 = $\frac{p\left(1 + \ln(\ln N)\right)}{N}$

Convoluting the penalties between BIC, AIC and HQC

This is achieved as follows:

$$NIC4 = \frac{1}{3}\left[BIC + HQ + AIC\right]$$

$$= \frac{1}{3}\left[\ln\sigma_p^2 + \frac{p\ln N}{N} + \ln\sigma_p^2 + \frac{2p\ln(\ln N)}{N} + \ln\sigma_p^2 + \frac{2p}{N}\right]$$

$$NIC4 = \left[-2\left(\frac{l}{N}\right) + \frac{\frac{2}{3} p \ln N \left(\frac{1}{2} + \frac{1}{\ln N}\right)}{N} \right];$$

The penalty for NIC4 is NIP4 = $\dfrac{\frac{2}{3} p \ln N \left(\frac{1}{2} + \frac{1}{\ln N}\right)}{N}$

Where the parameters are as defined above.

NIC 1= $\dfrac{\frac{p}{2} \ln N}{N}$

NIC 2= $\dfrac{\frac{p}{2} (\ln N + 2)}{N}$

NIC 3= $\dfrac{p\left(1 + \ln(\ln N)\right)}{N}$

NIC 4= $\dfrac{\frac{2}{3} p \ln N \left(\frac{1}{2} + \frac{1}{\ln N}\right)}{N}$

E Assessing the Best Performing Criterion

The performance of the information criterion is measured by the highest number of cases of selecting the correct order of a model, for example the autoregressive (AR(p)). To achieve this we compute the probability of the correctly identifying the true model by each criterion. This probability could be any number between zero and one .With the following decision rule:

(i) If this probability is 1, then the criterion correctly identified the true lag length in all the cases and therefore is an excellent criterion.

(ii) If the probability is close to 1 but greater than or equal to 0.5, then the criterion correctly identified the true lag length in most of the cases and hence is a good criterion.

(iii) If the probability is close to zero or less than 0.5, then the criterion fails to select the true lag length in most of the cases and therefore, not a good criterion.

F Selecting the Best of the proposed criterion

The four penalties NIP1, NIP2, NIP3, NIP4 is examined to see the penalty that compensated most for the bias or model inaccuracy (badness of fit or lack of fit). For instance, we consider sample sizes of 30 and 100 with parameter (p) where p = 1, 2…,20.

It was found out that penalty of NIC2 compensated most for the bias or model inaccuracy (badness of fit or lack of fit) for both small and large samples sizes as $p \to 1$ and $N \to \infty$. NIC2 is then adopted as the new information criterion {now labeled NIC) penalty which is believe to cater more for the biasedness in the maximum Likelihood estimate and penalize underfitting and overfitting association with AIC.

G Monte Carlo Simulation

The random deviates are generated from a unit Gaussian using R statistical package. The routine generates realization for a given AR structure.

In order to measure the performance of the order identification criteria, simulation and real data study was conducted using a wide range of autoregressive (AR) processes with different characteristics. Data were generated for samples of sizes N=20, 30, 100, 200, 500 and 1000 alongside with real data on money in circulation and income data.

Autoregressive model of order 1, 2, 3, 4, 5, 6 were generated with maximum lag of six. Dickey Fuller unit root test was performed on all generated and the available series to ensure that the series are stationary. The model order were examined using the criteria AIC, BIC, and HQ using EViews 7 software package and Microsoft Excel for the graphs. The performance criterion is that, the information criterion with the highest number of cases (or probability) of selecting the correct order of the given autoregressive (AR), process is considered the best.

III RESULTS AND DISCUSSIONS

In this section the performance of the new information criteria with the existing ones was carried out using real data and simulated data. This comparison is based on the statistical technique discussed in earlier sections. Stationarity were ensued in all the simulated data.

Simulation 1

Table 1: Probability of correctly estimation the true Lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
20	0.3796186	0.4027936	0.3672783	0.46819089

Simulation 2

After ensuring stationarity, autoregressive models of orders one to six was fitted and the summary is given in tables below:

Table 2: Probability of correctly estimating the true lag length

Sample Size	**Criteria**			
	AIC	BIC	HQ	NIC
30	0.4420543	0.4419659	0.44143	0.4420188

Simulation 3
After ensuring stationarity, autoregressive models of orders one to six was fitted and the summary is given in tables below:

Table 3: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
100	0.5003858	0.5354654	0.5256117	0.61242153

From the table 3 NIC and BIC has the highest probability of correctly estimating the true lag among the four criterions when the sample size is hundred follow by HQ and lastly by AIC.

Simulation 4
After ensuring stationarity, autoregressive models of orders one to six was fitted and the summary is given in tables below:

Table 4: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
200	0.519873	0.569222	0.537164	0.614878

Table 5: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
500	0.53805	0.56680150	0.54845998	0.5514901

From Table 5, BIC and NIC has the highest probability of correctly estimating the true lag among the four criterions when the sample size is five hundred follow by HQ and AIC

Table 6: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
1000	0.529509	0.534113	0.534755	0.54649

From Table 6, NIC and HQ has the highest probability of correctly estimating the true lag among the four criterions when the sample size is one thousand by BIC and AIC.

Comparison with real life data
Data on money in circulation of size n=30. A unit root test with Augmented Dickey Fuller test indicated that the series was non stationary but was made stationary after the first difference at 0.05 percent level. Then the autoregressive model of order one to six was fitted and the summary is given in tables below:

Table 7: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
30	0.4853961	0.5039052	0.5147478	0.521675

From table 7 above, NIC and HQ has the highest probability of correctly estimating the true lag among the four criterions when the sample size is five hundred follow by BIC and AIC.

Application To Real life Data
Data on money in circulation of size N=467 were use used to illustrate the performance of the new information criterion. The data were found stationary after the first difference using the Dicky-Fuller unit root test. Autoregressive process of order one to six was fitted and the summary is given in Table 8.

Table 8: Probability of correctly estimating the true lag length

Sample Size	Criteria			
	AIC	BIC	HQ	NIC
467	0.2568689	0.2257349	0.2470586	0.5039573

From Table 8, NIC has the highest probability of correctly estimating the true lag among the four criterions when the sample size is four hundred and sixty seven (467) follow by AIC, HQ, and BIC.

Table 9: Frequency and Probability from the Available Data for Different Sample Size.

Sample Size	Lag Length Selection Criteria			
	AIC	BIC	HQ	NIC
20	0.3796186	0.402793	0.367278	0.468190*
30	0.4420543	0.441965	0.44143	0.442018
100	0.5003858	0.535465	0.525611	0.612421*
200	0.5198730	0.569221	0.537163	0.614878*
500	0.5380573	0.566801	0.548459	0.55149
1000	0.5295099	0.534113	0.534755	0.54649*
Probability from the Available Data Set				
30	0.4853961	0.503905	0.514747	0.521675*
467	0.4863962	0.504905	0.514750	0.521775*

Table 9 shows that the performance of the new information criteria approach is better than BIC, HQ, and AIC approach. NIC selected the true lag with the highest weight of evidence (i.e. probability) five times out of the eight series used in this study, BIC selected the true lag with the highest weight of evidence two times, AIC selected the true lag with highest weight of evidence once while HQ do not select any.

IV CONCLUSION

It was observed that AIC is not consistent while the weight of evidence of NIC is consistent and considering the analysis of the available data in table 1a to 8a, NIC is better than BIC in terms of the closeness to the true value.
From the study, the new information criterion NIC and BIC performed better for small and large sample size. It was also discovered that AIC performed best when the sample size is less than or equal to thirty and HQ performs as the second best only in extremely large (i.e. greater than of equal to 1000) samples. This result is in agreement with [10] and [11].

The study also revealed that for a series having more than 30 observations, there is an improvement in the weight of evidence performance for each of these four criteria.

REFERENCES

[1] Akaike, H (1973): "Information Theory and an Extension of the Maximum Likelihood Principle" In B.N Petro and F. Csaki, ed. 2nd, International symposium on Inference Theory, Academia Kiado, Budapest, pp. 267-281.

[4] Burham K. P Anderson D. R (2002): Model selection and multimodel inference : a practical information-theoretic approach / Kenneth P. Burnham, David R. Anderson.—2nd ed. p. cm

[5] Cleasken G and Hjort N. L (2011): Model selection and model Averaging,http://www.math.rug.nl/stat/models/files/claeskens.pdf

[6] Hannan, E.J and Quinn B.G (1979): "The Determination of the Order of an Autoregression", JRSS, B41, 190-195.

[7] 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.

[12] Shibata (1976): "Some Properties of the Order of an Autoregressive model selected by the Generalization of Akaike's EPF Criterion" Biometrika 1977 64(3):547-551; doi:10.1093/biomet/64.3.547.

[13] Shibata (1976): "Selection of the Order of an Autoregressive Mode by Akaike Information Criterion", Biometrica, 63, pp.117-126.

[14] Shwartz, (1978): "Estimating the Dimension of a Model" Ann. Statist., 6, pp. 461-464.

[15] Stone, M (1979): "Comments on Model Selection Criteria of Akaike and Swartz" Journal of Royal Statistical Society B41, No.2, pp. 276-278.

[16] Shittu O.I. and Asemota M.J. (2009): Comparison of Criteria for Estimating the Order of Autoregressive Process: A Monte Carlo Approach. Euro Journals Publishing, Inc. pp. 409-416.

[17] Henry de-Graft Acquah (2009): Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in selection of an asymmetric price relationship, Journal of Development and Agricultural Economics Vol. 2(1) pp.001-006.