[1.]Abass I. TAIWO, [1.]Timothy O. OLATAYO

# OPTIMAL AND EFFICIENT MODEL SELECTION CRITERIA FOR PARAMETRIC SPECTRAL ESTIMATION

[1.]Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, NIGERIA

**Abstract:** Traditional Parametric spectral estimation methods have been widely used to obtain spectral estimate, resolution and variance distribution in signals or time series data across several fields. But the challenges of how to select an optimal and efficient model order were often encountered. In this research work, modified forms of Final Prediction Error, Akaike Information Criteria, Bayesian Information Criteria and Minimum Description Length which involved the replacement of variance error with sample autocorrelation function that has capabilities of detecting non-randomness in time series data were used to select the optimal model order. In order to determine the efficiency of the modified criteria, Autoregressive model, AR(11) was selected based on all the modified information criteria while AR(7) was selected based on all the traditional information criteria as the optimal model orders. Using the AR spectral estimation method, AR(11) and AR(7) models were used to analyse 1000 points Heartbeat readings. The results indicated that the spectral estimate and resolution of AR(11) were better than that of AR(7). Conclusively, the optimal model order selected using the modified information criteria gives a better result when compared to the traditional information criteria based on the spectral estimate and resolution of the Autoregressive spectral estimation method.

**Keywords:** Parametric Spectral estimation, Information Criteria, Heart-beat readings, Optimal model order, Autocorrelation function

## INTRODUCTION

There are many criteria that has been used to determine model order selection in parametric spectral estimation. The final prediction error (FPE) criterion was the first of two tools proposed by Akaike for the purpose of model order selection (Akaike, 1969; Akaike, 1970; Akaike, 1971). Since, the second one, presented several years later and known as the Akaike's information criterion (AIC) (Akaike, 1974) has deeper statistical justification and wider range of applicability than FPE, it is much more frequently used and referred to. Both criteria were derived for time-invariant systems/signals operated under stationary conditions and whenever both can be applied they asymptotically yield the same results. Schwartz, 1974 went ahead to propose Akaike's Bayesian Information criteria (BIC) and this was an improvement over AIC. The penalties in BIC are there to reduce the effects of overfitting and it is of note that the penalty is stronger in BIC than AIC for any reasonable sample size. BIC as well generally comes across only true models and penalizes free parameters more strongly with more accuracy but in practice it often overfit the data (Cruz-Ramírez *et al.*, 2006).

In another vain, minimum description length (MDL) proposed in (Rissanen, 1978) and further discussed in (Rissasen, 183) is a formalization of Occam's razor in which the best hypothesis (a model and its parameters) for a given set of data is the one that leads to the best compression of the data. Despite the abilities of this criteria it is computationally difficult and in practice it often leads to overfit of data (Cruz-Ramírez *et al.*, 2006).

Over the years, several other criteria and algorithms has been developed and these included Criterion Autoregressive Transfer (CAT) (Parzen, 1974), Residual Variance (RV) (Box and Jenkins, 1970), Hannan and Quinn (HQ) (Hannan and Quinn, 1979), Generic algorithm (Palanippan, 2006), Particle swarm optimization (PSO) (Bijaya, 2010) and many more. But these model selection methods are not completely outlined here since vast amount of techniques for solving the problem of selecting model order may have not being mentioned. The main goal of this research article is to improve the method of selecting the optimal and efficient model order in parametric spectral estimation. This will be done by modifying some existing traditional information criteria (Final prediction error, Akaike, Schwartz Bayesian and Minimum Description length information criteria) for selecting optimal order. The modification will involve the replacement of variance of error with sample autocorrelation function which has the capabilities of detecting non-randomness, help in identifying an appropriate model for non-randomness time series data and instead of estimating the error variance that required minimization of the log-likelihood function of the given model.

## MATERIALS AND METHODS

### — Final prediction error (FPE)

This is the first criteria proposed by (Akaike, 1969) and it is based on minimizing one step ahead predictor error. It is denoted by

$$FPE_{(k)} = \left(1 + \frac{k}{N}\right)\hat{\sigma}_k^2 \qquad (1)$$

where $\hat{\sigma}_k^2$ is the unbiased estimate of $\sigma_k^2$ after fitting the $k^{th}$ order model.

### — Akaike information criteria

This is the most well-known and mostly used criteria and it was proposed by (Akaike, 1970). It is denoted by

$$AIC(k) = N\ln\hat{\sigma}_k^2 + 2k \qquad (2)$$

where N is the number of observation and $\hat{\sigma}_k^2$ is the maximum likelihood estimation of the residual after fitting the $k^{th}$ order model.

— **Schwartz's SBC criteria**

Similar to Akaike's, Bayesian criteria, (Akaike, 1971) suggest the Bayesian criteria defined as

$$SBIC(k) = N\ln\hat{\sigma}_k^2 + M\ln N \quad (3)$$

where $\hat{\sigma}_k^2$ is the maximum likelihood estimate of $\sigma_k^2$, M is the number of parameters in the model and N is the number of observations.

— **Minimum description length criteria**

Based on the work of (Akaike, 1974), it was proved that the Akaike information criteria is inconsistent and it tends to overestimate the order. Schwartz, 1974, proposed a modified Akaike information criteria by replacing the term 2k by a term which increases more rapidly. This criterion was named minimum description length (MDL) and is of the form

$$MDL = N\ln|\hat{\sigma}_k^2| + k\ln(N) \quad (4)$$

— **Modified final prediction error (MFPE)**

This is the first criteria proposed by (Akaike, 1969) and it is modified by replacing the variance of error by sample autocorrelation function. It is denoted by

$$FPE_{(k)} = \frac{|N+k|}{N-k}|\hat{\rho}_k| \quad (5)$$

where $\hat{\rho}_k = \hat{\gamma}_k(0) + \sum_{i=1}^{k}\hat{a}_j\,\hat{\gamma}_{kk}(1)$ is the power of the prediction error that decreases with k while the term $\frac{N+K}{N-k}$ increases with k.

— **Modified Akaike information criteria**

This is the most well-known criteria and it was proposed by (Akaike, 1970) and was modified by replacing the variance of error by sample autocorrelation function. It is denoted by

$$AIC(k) = N\ln|\hat{\rho}_k| + 2k \quad (6)$$

This criterion is more general than final prediction error and it can be applied to determine the order of the moving average part of an autoregressive moving average model.

— **Schwartz's SBC criteria**

Similar to Akaike's, Bayesian criteria, Akaike, 1971 suggest the Bayesian criteria model that is modified by replacing variance of error with sample autocorrelation function and it is defined as

$$SBC(k) = k\ln|\hat{\rho}_k| + N\ln k \quad (7)$$

where $\hat{\rho}_k$ is an estimate of $\rho_k$, k is the number of parameters in the model and N is the number of observations.

— **Minimum description length criteria**

Based on the work of Kashyap (Akaike, 1974), it was proved that the Akaike information criteria is inconsistent and it tends to overestimate the order. Schwartz, 1974 proposed a modified Akaike information criteria by replacing the term 2k by a term which increases more rapidly. This criterion is named minimum description length (MDL) and is of the form

$$MDL = N\ln|\hat{\rho}_k| + k\ln(N) \quad (8)$$

— **Spectral density function of AR(1) Process**

Given an autoregressive model AR(1),

$$y_t = \emptyset_0 + \emptyset_1 y_{t-1} + \varepsilon_t \quad (9)$$

Using the backshift operator gives

$$\Phi(B)y_t = \varepsilon_t, \text{ where } \Phi(B) = (1 - \emptyset_1 B)$$

since $y_t = \Phi^{-1}(B)\varepsilon_t$, then

$$\gamma_k = E(y_t y_{t-1}) = \sigma_\varepsilon^2\left(\frac{1}{(1-\phi B)(1-\phi B^{-1})}\right) \quad (10)$$

Equation (10) is written in spectral representation as

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{|\Phi(e^{-i\omega})|^2}\right] \quad (11)$$

Since $|1-\Phi(e^{-i\omega})|^2 = 1 - \phi_1 e^{i\omega} - \phi_1 e^{-i\omega} + \phi_1^2$

$$= 1 - \phi_1(e^{i\omega} + e^{-i\omega}) + \phi_1^2$$

and using a standard trigonometric form given as

$$\cos\omega = \frac{e^{i\omega} + e^{-i\omega}}{2}$$

$$|1-\Phi(e^{-i\omega})|^2 = 1 - 2\phi_1\cos\omega + \phi_1^2$$

Then, equation (11) becomes

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{1 - 2\phi_1\cos\omega + \phi_1^2}\right] \quad (12)$$

— **Spectral density function of AR(2) Process**

The Yule Walker process is given by

$$y_t = \emptyset_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \varepsilon_t \quad (13)$$

Using the backshift operator to obtain

$$\Phi(B)y_t = \varepsilon_t,$$

where $\Phi(B) = (1 - \emptyset_1 B - \phi_2 B^2)$

Since $y_t = \Phi^{-1}(B)\varepsilon_t$

$$\gamma_k = \sigma_\varepsilon^2\left(\frac{1}{(1-\emptyset_1 B - \phi_2 B^2)(1-\emptyset_1 B^{-1} - \phi_2 B^{-2})}\right) \quad (14)$$

Equation (14) is written in spectral representation as

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{|1-\phi_1(e^{-i\omega}) - \phi_2(e^{-2i\omega})|^2}\right]$$

$$= \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{|1-\phi_1(\cos\omega - i\sin\omega) - \phi_2(\cos 2\omega - i\sin 2\omega)|^2}\right]$$

$$= \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{(1-\phi_1 e^{-i\omega} - \phi_2 e^{-2i\omega})(1-\phi_1 e^{i\omega} - \phi_2 e^{2i\omega})}\right]$$

$$= \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{\begin{matrix}1-\phi_1 e^{i\omega} - \phi_2 e^{2i\omega} - \phi_1 e^{-i\omega} + \phi_1^2 + \\ +\phi_1\phi_2 e^{i\omega} - \phi_2 e^{-2i\omega} + \phi_2\phi_1 e^{-i\omega} + \phi_2^2\end{matrix}}\right]$$

Using a standard trigonometric form given as

$$\cos\omega = \frac{e^{-i\omega} + e^{i\omega}}{2}$$

Equation (14) can be expressed as

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left[\frac{1}{\begin{matrix}1 + \phi_1^2 + \phi_2^2 - 2\phi_1\cos\omega - \\ - 2\phi_2\cos 2\omega + 2\phi_1\phi_2\cos\omega\end{matrix}}\right] \quad (15)$$

## RESULT AND DISCUSSIONS

In order to identify the optimal model, autocorrelation and partial autocorrelation functions were used to identify tentative models AR(3), AR(5), AR(7), AR(9) and AR(11) for 1000 heart-beat readings observed at equal space and time of 0.05 second. To validate these models, modified and traditional information criteria were obtained.

From table 1 and figure 2, the lowest values of all the modified information criteria occurred at AR(11) while the lowest values for the traditional information criteria occurred at AR(7) in table 2. The performance of both information criteria were determined based on spectral estimate and resolution of Autoregressive spectral estimation using modified covariance autoregressive estimator. Based on figure 3, the spectral estimate of AR(7), AR(9) and AR(11) indicated a relatively fast oscillation. This was explained by the two sinusoidal components in all autoregressive order but AR(11) is better since it has a dominant peak, better spectral estimate and resolution when compared to AR(7) and AR(9). AR(11) depict the general oscillation better than AR(7) and AR(9). Thus the modified information criteria that is, modified final prediction error, Akaike, Schwartz Bayesian and Minimum Description length information criteria gives an optimal and efficiency model selection as against the most frequently used traditional information criteria.

Table 1: Modified information criterion for 1000 heartbeat readings

| NUMBER | FPE | AIC | BIC | MDL |
|---|---|---|---|---|
| 1 | 45.74546 | -1.45679 | -0.06914 | 0.45524 |
| 2 | 19.42644 | -4.61658 | 34.3127O | -0.79254 |
| 3 | 16.46170 | -8.77227 | 54.04428 | -3.03620 |
| 4 | 14.04178 | -12.50380 | 67.67442 | -4.85569 |
| 5 | 11.93420 | -16.31390 | 77.84051 | -6.75377 |
| 6 | 9.63539 | -22.57740 | 85.43868 | -11.10530 |
| 7 | 7.23339 | -32.35490 | 90.80581 | -18.97080 |
| 8 | 5.00028 | -46.11840 | 94.03314 | -30.82220 |
| 9 | 3.18856 | -63.76340 | 95.14381 | -46.55520 |
| 10 | 1.79256 | -87.53610 | 93.62203 | -68.41590 |
| 11 | 0.59493 | -137.46600 | 84.81228 | -116.43400 |
| 12 | 0.50280 | -140.43800 | 84.78019 | -117.49400 |
| 13 | 1.38346 | -84.13700 | 99.61185 | -59.28070 |
| 14 | 1.95248 | -60.92240 | 107.05460 | -34.15410 |
| 15 | 2.22812 | -47.98480 | 112.00710 | -19.30440 |
| 16 | 2.28275 | -40.02700 | 115.58080 | -9.43460 |
| 17 | 8.28499 | -35.61560 | 117.99140 | -3.11123 |
| 18 | 1.89294 | -34.28410 | 119.21630 | 0.13236 |
| 19 | 1.56426 | -35.18040 | 119.41340 | 1.14806 |
| 20 | 1.24642 | -36.87910 | 119.03500 | 1.36134 |

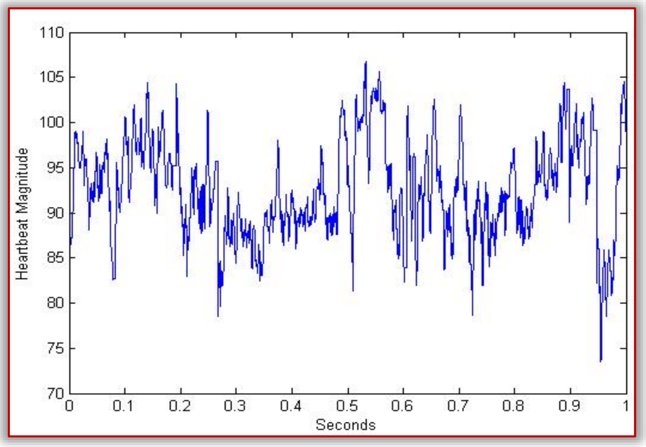


Figure 1: Heartbeat readings in time

Table 2. Traditional information criterion values for 1000 heartbeat readings

| NUMBER | FPE | AIC | BIC | MDL |
|---|---|---|---|---|
| 3 | 1.84691 | 1.99214 | 1.99214 | 1.85246 |
| 5 | 1.43891 | 1.56250 | 1.56250 | 1.43891 |
| 7 | 1.18234 | 1.47838 | 1.47838 | 1.18234 |
| 9 | 1.34821 | 1.29929 | 1.71723 | 1.34821 |
| 11 | 1.41245 | 1.86863 | 1.86863 | 1.41245 |

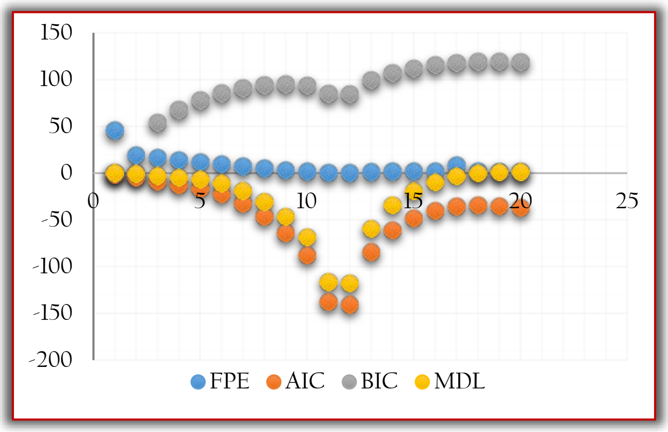


Figure 2: Graphical representation of modified information criteria

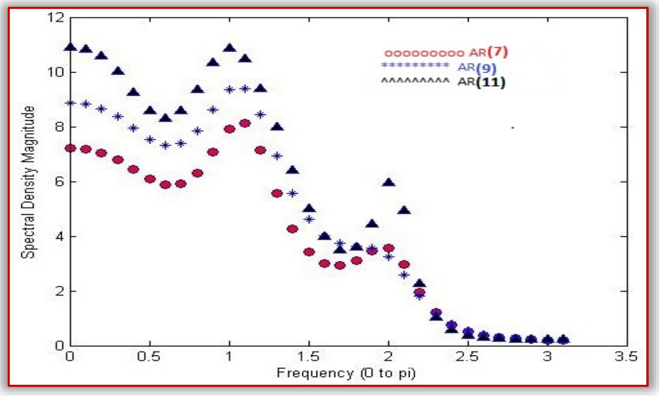


Figure 3: Spectral estimates for AR(7), AR(9) and AR(11)

## CONCLUSION

This paper was used to propose improved and modified methods that can be used to select optimal model order in parametric spectral estimation. This was done by modifying the traditional information criteria and from the results obtained, AR(11) and AR(7) were the optimal model selected using modified and traditional information criteria.

The spectral estimate and resolution of AR(11) were better than AR(7). Conclusively, the modified information criteria outperformed the traditional information criteria when analysing the spectral estimate of 1000 heart beat readings, with lower parameters than the traditional methods.

## References

[1] Abdul-Majeeb, H, A., Mohammed, Q. A. and Wadhah, S. I.: Estimating the Spectral Power density function of Non-Guassian Second order Autoregressive model.,Journal of Asian Scientific Research, 4, 513 – 521, 2014.

[2] Andrea, F., Arianna, P., Laura, S. and Massimilland, M.: Fourier analysis for Demand Forecasting in a Fashion Company. International Journal of Engineering Business Mnagements, 5, 1 – 10, 2013.

[3] Akaike, H.: Fitting autoregressive models for prediction, Ann. Inst. Statist. Math., 21, 243–247, 1969.

[4] Akaike, H.: Statistical predictor identification, Ann. Inst. Statist. Math., 22, 203–217, 1970.

[5] Akaike, H.: Autoregressive model fitting for control, Ann. Inst. Statist. Math., 23, 163–180, 1971.

[6] Akaike, H.: A new look at the statistical model identification," IEEE Trans. Automat. Contr., 19, 716–723, 1974.
[7] Andrew, J. G. and Barry, G. Q.: Parametric Spectral Discrimination, Journal of Time series Anaysis, 38, 1 – 27, 2017.
[8] Batorova, I.: Spectral Techniques for Economic Time series, Ph.D. Thesis Dissertation, Cornenius University, Bratislava, Slovakia, 2012.
[9] Beheshti, S.: A New Approach to Order Selection and Parametric Spectrum Estimation, Acoustics, Speech and Signal Processing, Proceedings of ICASSP, Toulouse, France, 2006
[10] Bijaya, G.: Spectral Estimation of Electroencephalogram signal using ARMAX model and Particle Swarm Optimization, M.Sc. Dissertation. Lamar University, Texas, 2010.
[11] Box, G. E. P. and Jenkins, G. M.: Time Series Analysis; forecasting and control, Holden Day, San Francisco, California, 125 – 135, 1970.
[12] Cruz-Ramírez, N., Acosta-Mesa, H. G., Barrientos-Martínez, R. E., Nava-Fernández, L. A.: How Good Are the Bayesian Information Criterion and the Minimum Description Length Principle for Model Selection? A Bayesian Network Analysis. In: Gelbukh, A., Reyes-Garcia, C. A. (eds) MICAI 2006: Advances in Artificial Intelligence. MICAI. Lecture Notes in Computer Science, 4293. Springer, Berlin, Heidelberg 494 – 504, 2006.
[13] Hannan, E. J. and Quinn, B. G. 1979, The determination of the order of an Autoregression, Journal of the Royal Statistical Society, 41, 190 – 195, 1979.
[14] Schwartz, G.: Estimating the Dimension of a model. Annal of Statistics, 6, 461 – 464, 1974.
[15] Palanippan, R.: Towards Optimal model order Selection for Autoregressive Spectral Analysis of Mental Tasks using Genetic Algorithm. International Journal of Computer Science and Network Security, 6, 153 – 161, 2006.
[16] Parzen, E.:S ome recent advances in Time series modeling. IEEE Trans. Automat. Contr., 19, 723-730, 1974.
[17] Rissanen, J.: Modeling by shortest data description, Automatica. 14(5), 465 – 658, 1978.
[18] Rissasen, J.: A Universal prior for the Integers and Estimation by Minimum Description Length Annal of Statistics, 11, 417 – 431, 1983.
[19] Rossi, G. B., Crenna, F., Piscopo, V. and Scamardella, A.: Comparison of Spectrum Estimation Methods for the Accurate Evaluation of Sea State Parameters, Sensors, 20, 1 –15, 2020.
[20] Taiwo, A., I.: Spectral and Fourier Parameter Estimation of Periodic Autocorrelated Time Series Data, Department of Mathematical Sciences, Ph.D. Thesis, Olabisi Onabanjo University, Ago-Iwoye, Nigeria, 2017.
[21] Taiwo, A. I. and Olatayo, T. O.: Performance of an Autoregressive Spectral estimation method based on Signal Length, Annals Computer Science Series Tibiscus, 17(1), 252 – 257, 2019.
[22] Thanagasundran, S. and Schlindwein, F. S.: Autoregressive order selection for Rotating Machinery, International Journal of Acoustics and Vibration, 11, 1 – 11, 2006.

ISSN: 2067-3809
copyright © University POLITEHNICA Timisoara,
Faculty of Engineering Hunedoara,
5, Revolutiei, 331128, Hunedoara, ROMANIA
http://acta.fih.upt.ro