Journal of the Nigerian Association of Mathematical Physics
***Volume* 22 (November, 2012), pp 219 – 224**

An Interval forecast for stationary Autoregressive process using Bootstrap method

[1]Ekhosuehi N. and [2]Omosigho S. E.

[1,2]Department of Mathematics,
University of Benin, Benin City, Nigeria.

Abstract

In this paper, we propose a new interval forecast for stationary Autoregressive, AR(p), process using bootstrap method. The bootstrap method which we call Akaike information criterion (AIC) bootstrap employs resampling of the residuals generated from the fitted AR(p) model to compute various AIC. The standard error obtained from the AIC statistic is then used to construct forecast interval for an AR(p) model. In our study, this AIC bootstrap forecast interval compares favourably with the theoretical forecast interval in an out of sample forecast performance. A simulation study was used to demonstrate the procedure.

Keywords: AR(p); AIC; Interval forecast

1.0 Introduction

Modelling and forecasting time series data using AR(p) model has been studied excessively. An important aspect of AR(p) process is its usefulness for forecasting future observations. Forecasting these future values may take the form of point forecast or an interval forecast. Using this AR(p) process, future observations can be computed using simple recursion equation formulated from the estimated model while the interval forecast is based on a theoretical function derived from the estimated model. Many types of interval forecast abound in literature, amongst is the one called Sieve Bootstrap method. Notable authors in this area of interval forecast using Sieve bootstrap, are Bühlmann [1], Alonso et al [2, 3] and a host of others. Chatfield [4], also presented a method of constructing interval forecast for stationary and non-stationary models while Christofferson [5] gave a general method for evaluating the interval forecast. Chatfield [6] also asserted that there is no complete theory in evaluating interval forecast and that the theoretical prediction interval sometimes appears too wide or too narrow and therefore identified possible causes of these problems. As a matter of fact, an interval forecast may be judged a good one if it gives more than 95% coverage to the expected future values. A purpose of this study therefore, is to construct a simple forecasting interval for stationary, AR(p) process that is capable of achieving more than 95% coverage to the expected future value. Our method makes use of the standard error obtained from bootstrap re-sampling of the residuals generated from the estimated AR(p) process. In Ekhosuehi and Omosigho [7], we show the close relationship that exists between the theoretical forecast mean square error (FMSE) and the AIC function.

The rest of this work is organized as follows: In Section 2, the methodology is presented. Section 3 is on simulation results while section 4 concludes.

2.0 Methodology

The methods of forecasting with linear AR(p) model discussed below are based on the work of Box and Jenkins [8] and Bowerman and O'Connell [9].

Let the general AR(p) process be written as:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t \qquad (1)$$

[1]Corresponding author: ***Ekhosuehi N.,*** E-mail: -, Tel. +234 7037819078

for some integer $p \geq 1$, $\alpha_1, \alpha_2, \cdots, \alpha_p$ are real parameters such that the zeros of $\phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ lie outside the unit disk and ε_t constitute a sequence of independent random variables with the same normal distribution $N(0, \sigma_\varepsilon{}^2)$. After estimation of the parameters by the least squares method and the model test proved adequate, then the estimated model can be used for forecasting purpose. The forecasting equation for h-step ahead forecast can be constructed as follows:

$$\hat{y}_{t+h} = \hat{\alpha}_1 y_{t+h-1} + \hat{\alpha}_2 y_{t+h-2} + \cdots + \hat{\alpha}_p y_{t+h-p} + \varepsilon_{t+h} \tag{2}$$

Taking the conditional expectation of (2), we have the following results:

$$E(y_{t+h} / \Omega_t) = \hat{y}_t(h) = \hat{\alpha}_1 E(y_{t+h-1}) + \cdots + \hat{\alpha}_p E(y_{t+h-p}) \tag{3}$$

where Ω_t denotes the past history of the time series up to and including the observation at time t, and the following expectations holds:

$$E(y_{t-j}) = y_{t-j}, \quad j = 0,1,2,\ldots$$

$$E(y_{t+j}) = \hat{y}_t(j), \quad j = 0,1,2,\ldots$$

$$E(\varepsilon_{t-j}) = \varepsilon_{t-j} = y_{t-j} - \hat{y}_{t-j-1}(1), \quad j = 0,1,2,\ldots$$

Such that $E(\varepsilon_{t+j}) = 0$ for $j = 1,2,\ldots$; and y_{t-j} for $j = 0,1,2,\ldots$ are values that have already happened at time origin t and y_{t+j} for $j = 1,2,\ldots$ which have not yet happened are replaced by their forecast $\hat{y}_t(j)$ at origin time t for $j = 1,2,\ldots$ so that the equation (3) allows a recursive estimation of forecast into the future.

For any forecast made using (3) there exists a forecast error or what is called prediction error ε_{t+h}, and it is given by:

$$\varepsilon_{t+h} = y_{t+h} - \hat{y}_{t+h} \tag{4}$$

where y_{t+h} is the expected future values and $\hat{y}_{t+h}$ is the forecast made at time t.

Franses and Van Dijk [10] stated that it is desirable to choose the forecast $\hat{y}_{t+h}$ that minimizes the forecast mean square error (FMSE). The FMSE can be expressed as:

$$FMSE(h) \equiv E(\varepsilon_{t+h}^2) = E[(y_{t+h} - \hat{y}_{t+h})^2] \tag{5}$$

Assuming normality, a $100(1-\alpha)\%$ forecasting interval for $\hat{y}_{t+h}$ in AR(p) is given by the interval below:

$$\hat{y}_{t+h} \pm Z_{\alpha/2}.(FMSE(h))^{1/2} \tag{6}$$

The available theoretical forecast interval, is constructed [8] by setting

$$(FMSE(h))^{1/2} = \sigma_\varepsilon \left(\sum_{j=0}^{h-1} \psi_j^2 \right)^{1/2} \tag{7}$$

where $\psi_0 = 1$, ψ_j for $j = 1,2,...,h-1$ are the forecast weights which can be computed recursively from infinite representation of (1) and σ_ε^2 is the residual error variance computed from the generated error sequence $\{\hat{\varepsilon}_t\}$ of the fitted model in (1).

The order 'p' of the AR(p) process given by (1) can be determined using certain information criteria such as the Akaike information criterion (AIC) [11]; Schwarz information criterion (SIC) [12]; Hannan and Quinn information criterion (HQC) [13]. In this study, we are interested in AIC of the following forms:

$$AIC(k) = n \log(\hat{\sigma}_\varepsilon^2) + 2k \tag{8}$$

$$AIC(k) = -n \log(\hat{\sigma}_\varepsilon^2) + 2k \tag{9}$$

where k denote the number of parameters in the model, n is the sample size and $\hat{\sigma}_\varepsilon^2 = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2$, with $\hat{\varepsilon}_t$ the residuals generated from estimated model in (1).

The expressions for AIC in (8) and (9) can either be positive or negative. These values are used indiscriminately. Another variant of AIC that ensure positive value is given by following:

$$AIC(k) = \begin{cases} n\log(\hat{\sigma}_{\varepsilon}^{2}) + 2k, & \log(\hat{\sigma}_{\varepsilon}^{2}) > 0 \\ -n\log(\hat{\sigma}_{\varepsilon}^{2}) + 2k, & \log(\hat{\sigma}_{\varepsilon}^{2}) < 0 \end{cases} \tag{10}$$

In Ekhosuehi and Omosigho [14], we show that the sampling distribution of the AIC function represented by (8) and (9) is that of the normal distribution, while the one given by (10) has the Chi-square distribution. As a of rule of thumb, Ekhosuehi and Omosigho [7] used $AIC/\sqrt{n}$ as an approximation to the theoretical FMSE given in (7) to construct interval forecast. Our new method of constructing AR(p) forecast interval to estimate the FMSE in (7) is based on the bootstrap method. This method is achieved by constructing B number of AIC function by re-sampling from the residual sequence obtained from the estimated model, and then computes the standard error. The idea is similar to the Sieve bootstrap for times series proposed by Bühlmann [1] and Sieve bootstrap interval by Alonso *et.al* [2, 3]. The AIC function that is used in this regard is the function defined by (8) or (9). The procedure for constructing the AIC bootstrap forecast interval is given as follows: Given the original sequence of error generated from a fitted model, $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n\}$, where $\hat{\varepsilon}_i = y_i - \hat{y}_i$, then we compute AIC bootstrap statistic as $AIC^{*b} = f\{\hat{\varepsilon}_1^{*b}, \hat{\varepsilon}_2^{*b}, \cdots, \hat{\varepsilon}_n^{*b}\}$ and thus estimate its standard error as :

$$\widehat{SE}_{B(AIC)}^{2} = \frac{1}{(B-1)}\sum_{b=1}^{B}(AIC^{*b} - \overline{AIC}^{*})^{2} \text{ where } \overline{AIC}^{*} = \frac{1}{B}\sum_{b=1}^{B} AIC^{*b} \tag{11}$$

We summarize the procedure using the following steps:

Step1. Given the time series observations $\{y_t\}$, select the order p of the autoregressive approximation using the AIC function.

Step2. Obtain the estimates of the autoregressive coefficients $(\hat{\alpha}_1, \alpha_2, \cdots, \hat{\alpha}_p)'$ using the least squares method.

Step3. Compute the residuals $\hat{\varepsilon}_t$ of fitted model in (1) for $t = p+1, \cdots, n$

Step4. Compute the empirical distribution function of the centered residuals $\{\tilde{\varepsilon}_t\}$

$$\hat{F}_{\tilde{\varepsilon}}(x) = (n-p)^{-1}\sum_{t=p+1}^{n} 1_{\{\tilde{\varepsilon} \le x\}}, \text{ where } \tilde{\varepsilon}_t = \hat{\varepsilon}_t - (n-p)^{-1}\sum_{t=p+1}^{n}\hat{\varepsilon}_t .$$

Step5. Obtain forecast of length h from the estimated model using (3).

Step6. Generate AIC bootstrap statistic based on $\tilde{\varepsilon}_t$ of independent and identically distributed observations from $\hat{F}_{\tilde{\varepsilon}}$.

Step7. Construct the prediction interval using (6) where $Z_{\alpha/2}(FMSE)^{1/2}$ is replaced by $2\widehat{SE}_{B(AIC)}$

3 Simulation Results

We consider the following data generating process (DGP) using AR(1) and AR(2) models respectively.

Model (I) $\quad y_t = 0.5y_{t-1} + \varepsilon_t, \quad y_0 = 0$

Model (II) $\quad y_t = 0.75y_{t-1} - 0.5y_{t-2} + \varepsilon_t, \quad y_0 = 0$

The sequence of errors $\{\varepsilon_t\}$ which is normally distributed with mean 0 and variance 1 is generated using the random number generator in MATLAB 7.5.0. These error sequences were in turn used in generating $(300+n)$ sample sizes using the DGP in models (I) and (II) respectively. We keep the last n observations and discard the first 300 so as to minimize initialization effect. Among the n generated artificial time series observations, the first $(n-s)$ observations are used for modelling and the remaining s observations are kept for out-of-sample performance. In this study, we set $n =$ 50 and 100 sample sizes and $s = 5$ as the forecast horizon. The modelling cycle which involves test of stationarity, model identification, parameter estimation and diagnostic test is written and executed in MATLAB 7.5.0 using MATLAB codes. After estimation of the model, it is then used for forecasting and for constructing forecast interval based on AIC bootstrap method and the theoretical method for the purpose of comparison. The bootstrap sample B taken is 500 where we employed the idea of Efron and Tibshirani [15], Martinez and Martinez [16]. The forecasting horizon (h) is taken from h = 1 to s, so as to agree with the number of observations kept for the out-of-sample performance. This process is replicated 100 times. We however, present few of the results in Figures 1, 2 3, and 4.

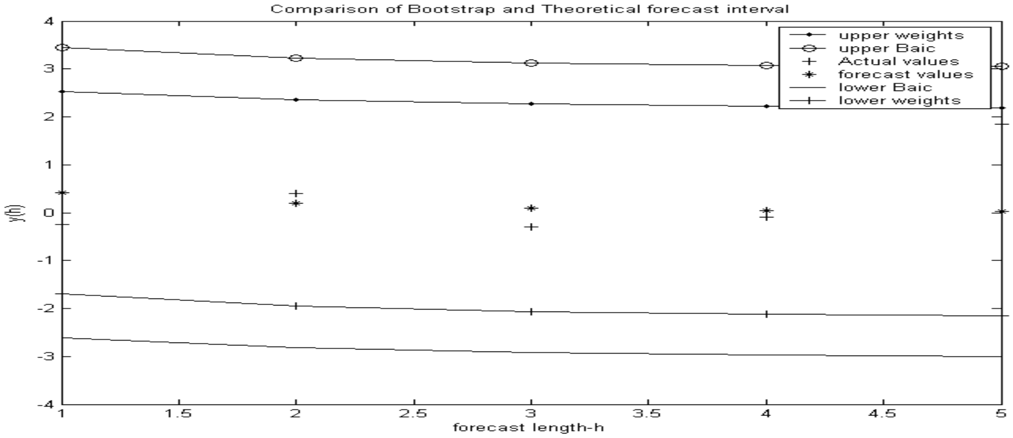

Figure 1: 95% forecast interval using the theoretical forecast interval and the proposed AIC Bootstrap forecast interval using twice standard error for AR(1) with sample size n- S = 50, S = 5.

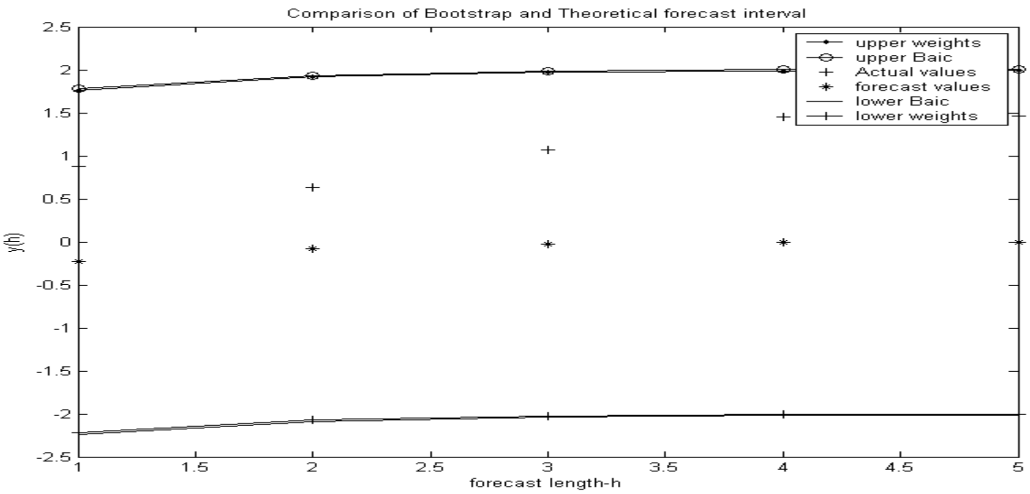

Figure 2: 95% forecast interval using the theoretical forecast interval and the proposed AIC Bootstrap forecast interval using twice standard error for AR(2) using sample size n- S = 50, S = 5.

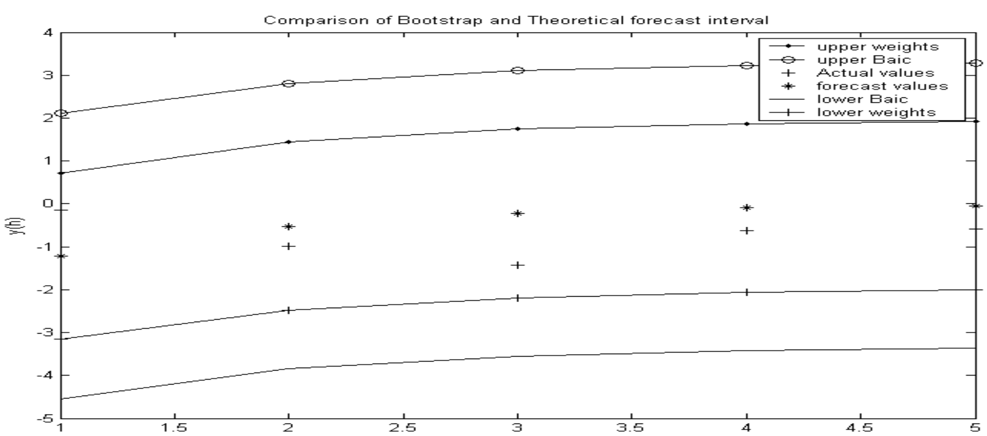

Figure 3: 95% forecast interval using the theoretical forecast interval and the proposed AIC Bootstrap forecast interval using twice standard error for AR(1) with sample size n- S = 100, S = 5.

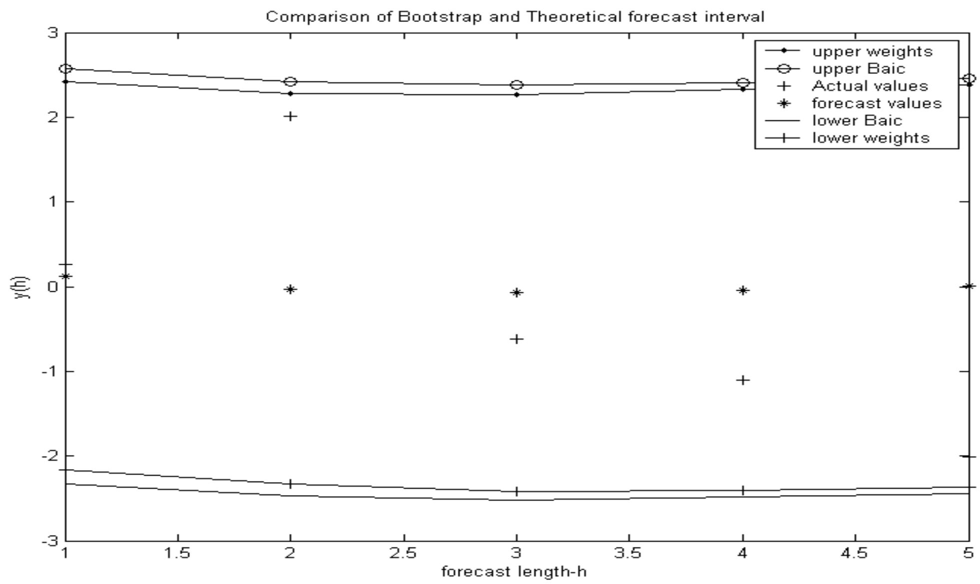

Figure 4: 95% forecast interval using the theoretical forecast interval and the proposed AIC Bootstrap forecast interval using twice standard error for AR(2) using sample size $n\text{-}s = 100$, $s = 5$.

The interval forecast in Figures 1, 2, 3 and 4 are constructed using the theoretical FMSE and the AIC Bootstrap forecast interval for sample sizes n = 50 and 100 respectively. The graphs also display the point forecast made and the actual values that are kept for out of sample performance. It is seen that the theoretical forecast interval is below the AIC Bootstrap forecast interval in Figure 1 while both methods are approximately the same in Figure 2. Similarly, Figure 3 exhibits the same structure as Figure 1, while the two methods are relatively close in Figure 4. It is interesting to see that both the AIC forecast interval and the Bootstrap forecast interval achieves 100% coverage to the expected future values in all cases considered.

4.0 Conclusion

In this paper, we show an alternative but simple method of constructing interval forecast for linear AR(p) time series model using AIC bootstrap method. Results from the simulation results show that the forecast interval predicted by this method closely approximates the traditional method of constructing forecast interval using a theoretical function.

References

[1] Bühlmann, P. (1997) Sieve bootstrap for time series. Bernoulli, 3(2) pp. 232-148.

[2] Alonso, A. M, Peña, D., J. Romo (2002) Forecasting time series with sieve bootstrap. J.statist. Planning inference 100, 1-11

[3] Alonso, A. M, Peña, D., J. Romo (2003) On sieve bootstrap prediction intervals. Statistics & Probability Letters 65, 13-20.

[4] Chatfield, C. (1993) Calculating Interval Forecasts. Journal of Business Statistics, Vol. 11, No. 2 pp. 121-135.

[5] Christofferson, P. F. (1998) Evaluating Interval Forecasts. International Economic Review, Vol. 39, No. 4 pp. 841-862

[6] Chatfield, C. (2000) Time series Forecasting. Taylor & Francis.

[7] Ekhosuehi, N. and Omosigho, S. E. (2010). New interval forecast for stationary Autoregressive models. International Journal of Natural and Applied Sciences, 6(4): pp. 471-477.

[8] Box, G. E. P and G. M. Jenkins (1976). Time series analysis, forecasting and control, Holden-Day, San Francisco.

[9] Bowerman, B. L. and R. T. O'Connell (1993). Forecasting and Time Series: An applied approach. 3rd Edition. Duxbury Press, California.

[10] Franses, P. H. and D. Van Dijk (2000). Non-linear time series models in empirical finance, Cambridge University press.

[11] Akaike, H. (1969) Fitting autoregressive models for prediction. Annals of the institute of statistical Mathematics, 21, 243 - 247.

[12] Schwarz. G. (1978 Estimating the dimension of a model. Annals of Statistics, 6, 461-464

[13] Hannan, E. J and B. G. Quinn (1979) The determination of the order of an autoregression. Journal of Royal Statistics Society, 41, 190 -195

[14] Ekhosuehi, N. and Omosigho, S. E. (2012). On the sampling distribution of Akaike information criterion. Nigerian Journal of Applied Science, vol. 30 pp. 31-37

[15] Efron, B. and R. J. Tibshirani (1998) An introduction to the Bootstrap. Chapman and Hall, London.

[16] Martinez, W. L. and A. R. Martinez (2002) Computational Statistics Handbook with MATLAB.Chapman and Hall, London.