Available online at www.sciencedirect.com
ScienceDirect
Journal of the Nigerian Mathematical Society 34 (2015) 106–114
Journal of the Nigerian Mathematical Society
www.elsevier.com/locate/jnnms

Bootstrap method for minimum message length autoregressive model order selection

T.O. Olatayo*, K.K. Adesanya

Mathematical Sciences Department, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria

Received 7 October 2012; received in revised form 5 September 2014; accepted 24 September 2014
Available online 4 December 2014

Abstract

Minimum Message Length MML87 is an information theoretical criterion for model selection and point estimation. In principle, it is a method of inductive inference, and is used in a wide range of approximations and algorithm to determine the ideal model for any given data. In this study, MML87 model selection criterion was investigated and compared with other notably model selection criteria such as Akaike information criterion (AIC), Bayesian information criterion (BIC), Corrected Akaike information criterion (AICc), and Hannan–Quinn (HQ), using Bootstrap Simulation Technique to simulate autoregressive model of order P. We specified three different counts systems as under inferred, correctly inferred and over inferred. Based on the candidate model explored with autoregressive model and the aggregate true model explored, with the estimated parameters. MML87 performed better than all other model selection criteria through the negative log likelihood function and the mean square prediction error estimated. It is more efficient and correctly inferred.

Keywords: Minimum message length; Bayesian information criterion; Akaike information criterion; Hannan–Quinn and Bootstrap

1. Introduction

The minimum message length (MML87) principle is a method of inductive inference, and has been used in a wide range of approximations and algorithms to determine the ideal model for any given data. MML87 has been successfully applied to areas such as mixture modeling [1–3], univariate polynomial inference [4], decision tree induction [5], autoregressive time series model inference [6] and stock market simulation [7].

The general idea behind MML87 is to minimize the combined length of a two-part message. The first part of the message contains the optimally coded description of the model, or the hypothesis. The second part of the message contains the actual data, encoded using that hypothesis. On one extreme, it is possible to select no model. The data, then, would be encoded without knowledge of the global behavior of the data. The other extreme would be to select any overly-complex model and have very little, or no, data for the second part. We would expect that neither of these models would minimize the total length of the two-part message. The minimized length would typically be found

Peer review under responsibility of Nigerian Mathematical Society.
* Corresponding author.
E-mail address: otimtoy@yahoo.com (T.O. Olatayo).

http://dx.doi.org/10.1016/j.jnnms.2014.10.010

somewhere between these extremes as a trade-off between complexity and fit, and would be more likely to be the correct model. As such, the MML87 principle can be seen as a formal information theory quantitative restatement of Occam's Razor. Entities should not be multiplied beyond necessity [8].

The estimation of order P for Autoregressive (AR) process is also a widely studied problem with many proposed solutions. Amongst the most common methods for order selection are those based on model parsimony that is, methods that punish complexity and attempt to find a trade-off between model complexity and capability in an attempt to improve the generalization capabilities of the selected model. Some of the earliest of these methods was the work by [9–13] on mixture modeling, which later led to the development of the MML87 criterion and its subsequent family of approximations. Other pioneering work on model selection by parsimony includes the AIC criterion, developed for automatic order selection of AR processes. Further refinements of the AIC scheme for time series led to the development of the AICc (Corrected AIC) estimator. Similar work has led to the Bayesian Information Criteria (BIC), and more recently, the symmetric Kullback–Leibler distance information criteria (KIC) and the corrected variant (KICc) [14].

In this study, effort would be geared towards comparison of MML87 as a model selection criterion with other notable model selection criteria. We will explore the use of Bootstrap simulation method to re-sample Autoregressive model of order p.

Various authors have adopted the use of Monte Carlo simulations to simulate autoregressive model, (AR), moving average model, (MA), autoregressive moving average (ARMA) model and many other time series models.

The authors uniformly sampled from the Autoregressive stationary region, and MML87 estimators model selection performance was empirically compared with AIC, corrected AIC [15], BIC, MDL and HQ [16]. While MML87 is geared towards inference rather than prediction. It was found that choosing the autoregressive model order having the MML87 gives a prediction error that is minimal to the other model selection criteria in the Monte Carlo experiment conducted.

Minimum Message Length (MML87) based model selection criterion performed better than the other criteria in correct model order selection in a moving average time series, Mean square prediction error (MSPE) and log likelihood for the small sample size, $T = 30$ and $T = 50$. For larger sample sizes, MML87 performance was still good, coming an overall second best in many of the experiments. For the larger sample sizes, it was found that the MML87 was not consistent across their assessment criteria. The HQ criterion scores the best for correct model order selection and MSPE for $T = 150$ and $T = 200$ but does not score best in log likelihood.

This study will therefore explore the use of bootstrap simulation method to simulate autoregressive model of order p and therefore compare MML87 with other notable model selection criteria to determine which of them is more efficient than others empirically. Basically Monte Carlo and Bootstrap method are both simulation method, but while Monte Carlo rely on initial value to randomly simulate, the Bootstrap make use of original real data for resampling with replacement.

2. Bootstrap method

Bootstrap resampling methods have emerged as a powerful tool for constructing inferential procedures in modern statistical data analysis. Bootstrapping is not perfectly synonymous with Monte Carlo. Bootstrap method rely on using an original sample or some part of it, such as residuals as an artificial population from which to randomly resampled [17]. The Bootstrap approach, as proposed by [18], avoids having to derive formulas via different analytical arguments by taking advantage of fast computers. The bootstrap methods have and will continue to have a profound influence throughout science, as the availability of fast, inexpensive computing has enhanced the ability to make valid statistical inferences about the world without the need for using unrealistic or unverifiable assumptions [19].

An excellent introduction to the bootstrap may be found in the work of [20], [21] and [22] who have independently introduced non-parametric versions of the bootstrap that are applicable to weakly dependent stationary observations. Their sampling procedure have been generalized by [23]. Some authors have worked on the stationary bootstrap [8] and the moving block bootstrap [21] and [24]. Recently [25] introduced bootstrap method for dependent data structure. The approach of bootstrap method in modeling was achieved through the following algorithm with the aid of computer.

Considered the model

$$y = \mu + \epsilon.$$

The bootstrap algorithm is as follows:

1. Fit the model to the original data, obtain the estimates and the residuals from the fitted model $\hat{\epsilon}$, $i = 1, 2, \ldots, n$.
2. Compute a bootstrap sample $\epsilon_n^* = (\epsilon_1^*, \epsilon_2^*, \ldots, \epsilon_n^*)$ from the residuals by sampling with replacement.
3. Using the design matrix, create the bootstrap values for the response using: $Y^* = \mu + \epsilon^*$.
4. Fit the model using as response Y^* and the design matrix X.
5. Repeat steps 2, 3 and 4 B times, where B is a large number, in other to create B resample. The practical size of B depends on the tests to be run on the data.
6. Estimate parameter of interest in model order selection.

3. Methodology

In MML87 message length expressions for an autoregression model, the first decision that must be made is the form of MML87 message. There are two seemingly obvious message formats that can be used based on the conditional or exact likelihood function. We assumed these data come from a stationary process. The process of regression and autoregressive model selection are closely related. Indeed many of the proposed solutions can be applied equally well to both problems. Suppose data are generated by the operating model, that is true model,

$$Y = \mu + \epsilon$$

where

$$y = (y_1, y_2, \ldots, y_n)^T, \qquad \mu = (\mu_1, \mu_2, \ldots, \mu_n)^T \quad \text{and} \quad \epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$$

where ϵ_n are independent and identically distributed in normal random variables with mean zero and variance σ^2. Additional assumptions about the form of the operating model will be made below. Consider the approximating, or candidate, family of models.

$$y = h(\theta) + u$$

where θ is an $m * 1$ matrix

$$u = (u_1, u_2, \ldots, u_n)^T, \qquad h(\theta) = (h_1(\theta), h_2(\theta), \ldots, h_n(\theta))^T$$

h is assumed to be twice continuously differentiable in θ, and u_i are independent identically distributed normal with mean zero and variance σ^2.

$y = h(\theta) + u$ is referred to as a model, or alternatively as a family of models, one model for each particular value of (θ, σ^2).

Considering an order p autoregressive $(\mathrm{AR}(p))$ model parameters: $\theta = (\phi_1, \phi_2, \ldots, \phi_n, \sigma^2)$. Thus a linear autoregressive with unconditional mean of zero related to the expected value of the time series linearly to p.

The conditional likelihood of θ is

$$\begin{aligned} &-\log f(y_{p+1}, \ldots, y_T | \theta, y_1, \ldots, y_p) \\ &\quad = \frac{T-p}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\left(\sum_{i=p+1}^{T}(y_i - \phi_i y_{i-1} - \cdots - \phi_p y_{i-p})\right) \quad \text{with } E(y_t) = 0 \forall t \\ &E(y_t y_{t-j}) = E(y_t y_{t+j}) = \gamma_j \forall t\ \forall j \end{aligned}$$

where Fisher information is defined as

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ & & \ldots & & \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

The stationary assumption is that the first p values are distributed as a multivariate normal distribution.

The negative log-likelihood function, when stationary is assumed, is:

$$-\log f\big(y|\theta\big) = -\log f\big(y_1, \dots, y_p|\theta\big) - \log f\big(y_{p+1}, \dots, y_T|\theta, y_1, \dots, y_p\big).$$

3.1. Model order selections criteria

3.1.1. Akaike's information criterion (AIC)

Originally proposed by Akaike in 1974, Akaike's Information Criteria, abbreviated to AIC, is the most popular and widely used criterion for model selections.

It is given by $AIC = -2\log\big(\sigma^2\big) + 2r$

where σ^2 is the maximum likelihood estimate of the variance and r denotes the number of independent parameters, including the variance σ^2.

3.1.2. Corrected Akaike's information criterion ($AICc$)

Hurvich and Tsai derived a bias correcting version for AIC in 1989 called corrected AIC (abbreviated in the literature as AICC, or AICc). It is given by:

$$AIC_c = -2\log\big(\sigma^2\big) + 2\frac{rT}{(T - r - 1)}$$

where T is the sample size with r is the parameter and σ^2 is the variance.

3.1.3. Bayesian information criterion (BIC)

Schwarz's derived his criterion in a Bayesian context and subsequently became known as Bayesian Information Criterion (BIC). It is defined as:

$$BIC = -2\log\big(\sigma^2\big) + 2r\log T$$

where T, r and σ^2 are as above.

3.1.4. Hannan–Quinn criterion (HQ)

Hannan–Quinn introduced another consistent criteria, known as HQ, in 1979 and is given by: $HQ = -2\log\big(\sigma^2\big) + 2r\log\log T$

where T, r and σ^2 are as above.

The bootstrap, when it was first proposed by [18], was an entirely parametric procedure. The idea was to draw bootstrap samples from the empirical distribution function (EDF) of the data, a procedure that he called resampling.

Since the EDF assigns probability $\frac{1}{n}$ to each point in the sample, this procedure amounts to drawing each observation of a bootstrap sample randomly, with replacement, from the original sample. Each bootstrap sample thus contains some of the original data points once, some of them more than once, and some of them not at all. Resampling evidently requires the assumption that the data are Identically and Independent Distributed (IID).

3.1.5. Minimum mean squared error

An estimator $\hat{\theta}$ in a class of estimators is said to have a minimum mean squared error if the mean squared error of $\hat{\theta}$, MSE$\big(\hat{\theta}\big)$, is not greater than the mean squared error of any other estimator in the class. The mean squared error of an estimator $\hat{\theta}$ of θ is defined as

$$\begin{aligned} MSE\big(\hat{\theta}\big) &= E\big(\hat{\theta} - \theta\big)^2 \\ &= E\big(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\big) \\ &= E\big(\hat{\theta}^2\big) - \big(E\big(\hat{\theta}\big)\big)^2 - 2\theta E\big(\hat{\theta}\big) + \theta^2 \\ &= E\big(\hat{\theta} - \theta\big)^2 + E\big(\hat{\theta} - E\big(\hat{\theta}\big)\big)^2 \\ &= \big(E\big(\hat{\theta}\big) - \theta\big)^2 + E\big(\hat{\theta} - E\big(\hat{\theta}\big)\big)^2 \\ &= squared\ bias\ of\ \hat{\theta} + V\big(\hat{\theta}\big). \end{aligned}$$

Table 1
Model order selection accuracy (candidate models).

Order	AIC	AICc	BIC	HQ	MDL78	MML87	Negloglike	MSPE
AR1	9.84813	9.705272	5.045735	3.47581	6.037438	5.180016	373.9286	8.96
AR2	9.835657	9.6928	5.033262	3.469573	6.031202	5.330008	371.6039	7.64
AR3	9.82742	9.684563	5.025025	3.465455	6.027083	5.029656	370.0765	6.87
AR4	9.8292	9.686342	5.026805	3.466345	6.027973	4.811131	370.406	7.03
AR5	9.80521	9.662352	5.002815	3.45435	6.015978	4.628587	365.9895	5.18
AR6	9.797302	9.654445	4.994907	3.450396	6.012024	5.65114	364.5454	4.69
AR7	9.792621	9.649764	4.990226	3.448055	6.009684	6.281158	363.6931	4.42
AR8	9.796008	9.653151	4.993614	3.449749	6.011377	7.856812	364.3096	4.62

3.1.6. Minimum message length $(MML87)$

This is used to approximate the message length for a model consisting of several continuous parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ by

$$MsgLen(\theta, y) = -\log \frac{h(\theta)h(n)}{\sqrt{k_n^n \left|I(\theta)\right|}} - \log f(y|\theta) \in^n + \frac{n}{2}$$

where $h(\theta)$ is a prior distribution over the parameters value, $h(\nu)$ is the prior on the number of parameters, k_n is a lattice constant $\left(K_1 = \frac{1}{2}, K_2 = \frac{5}{36\sqrt{3}}, K_n \to \frac{1}{2\pi s} \text{ as } n \to \infty\right)$ which accounts for the expected error in the log-likelihood function due to quantization of the n-dimensional space, $|\theta|$ is the expected fisher information matrix, and ϵ is the measurement accuracy of the data.

3.1.7. Likelihood function

Given the assumption that the target are truncated with multiplicative heteroscedastic error structure the negative log-likelihood of the linear regression is

$$-\log f(y|X, \beta, \sigma, \delta) = \frac{N}{2}\log(2\pi) + N\log\sigma + \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta) - N\log\in$$

where σ^2 is the variance of the multiplicative heteroscedastic error structure and is the measurement accuracy of the data. The accuracy in which the data is measured is often used in estimating a lower bound on the accuracy of the model.

4. Results and discussion

The result of our investigation, through empirical bootstrap simulation of thirty years real data of crimes in Nigeria, collected from Nigerian Prisons Service, were simulated [26]. Our candidate models run from $AR(1)$ to $AR(8)$, our true models run from $AR(1)$ to $AR(4)$ and bootstrap simulation was set at 100 for each of the true models of $p = (1, 2, 3, 4)$. The summary of the models selection criteria and bootstrap simulation are shown in Table 1.

Specifically, model selection criteria AIC, BIC, AICc, HQ, MDL78 and MML87 are reported in the above table with the respective negative log likelihood and mean square prediction error. In all, MML87 performed better than all other model selections used in our study, with minimum order of 5. Our candidate model runs from order $AR(1)$ to $AR(8)$. All model selection criteria selected at order 7, while MML87 selected at order 5. This implies the efficiency of MML87 above all other model selection criteria adopted in our research. Based on our candidate model we explore with our p maximum of 8, we found out that MML87 performed better and more efficient with minimum order selection criteria than all other selection criteria of AIC, BIC, AICc, HQ, MDL78.

Table 2 depicts the results for sample of size $t = 30$ and true model $p = 1$ in a 100 time series bootstrap simulation. We simulated our true model of AR of order $p = 1$ in 100 time using time series bootstrap simulation. We carried out the study based on the comparison of model order selection criteria with our well known MML87. We specify three different counts systems as under inferred, correctly inferred and over inferred. AIC outperformed all other criteria including our celebrated MML87 Model selection criteria. MML87 over inferred for $t = 30$ and $p = 1$.

Table 2
Results for $T = 30$ and true model $P = 1$.

Criteria	$p < \hat{p}$	$p = \hat{p}$	$p > \hat{p}$
AIC	3	40	57
BIC	100	–	–
AICc	–	–	100
HQ	7	4	89
MDL78	6	3	91
MML87	–	–	100

Table 3
Results for $T = 30$ and true model $P = 2$.

Criteria	$p < \hat{p}$	$p = \hat{p}$	$p > \hat{p}$
AIC	7	1	92
BIC	–	–	100
AICc	–	–	100
HQ	7	4	89
MDL78	6	3	91
MML87	–	–	100

Table 4
Results for $T = 30$ and true model $P = 3$.

Criteria	$p < \hat{p}$	$p = \hat{p}$	$p > \hat{p}$
AIC	10	1	89
BIC	11	–	89
AICc	11	–	89
HQ	10	3	87
MDL78	10	3	87
MML87	32	68	–

Table 3 depicts the results for $t = 30$ and true model $p = 2$ 100 time series bootstrap simulation. We simulate our true model of AR of order $p = 2$ 100 time using time series bootstrap simulation. We carried out the study based on the comparison of model order selection criteria with our well known MML87. We specify three different counts systems as under inferred, correctly inferred and over inferred. All the model selection criteria including MML87 over inferred for $t = 30$ and $p = 2$.

Table 4 depicts the results for $t = 30$ and true model $p = 3$ 100 time series bootstrap simulation. We simulate our true model of AR of order $p = 3$ 100 time using time series bootstrap simulation. We carried out the study based on the comparison of model order selection criteria with our well known MML87. We specify three different counts systems as under inferred, correctly inferred and over inferred. MML87 performed better than all other criteria. Vividly, MML87 is correctly inferred for $t = 30$ and $p = 3$.

Table 5 depicts the results for $t = 30$ and true model $p = 4$ 100 time series bootstrap simulation. We simulate our true model of AR of order $p = 4$ 100 time using time series bootstrap simulation. We carried out the study based on the comparison of model order selection criteria with our well known MML87. We specify three different counts systems as under inferred, correctly inferred and over inferred. MML87 performed better than all other criteria and correctly inferred.

Table 6 depicts the results for criteria under/correctly/over selecting the true models for $p = 1, 2, 3, 4$ 100 time series bootstrap simulation. We simulated our true model of AR of order $p = 1$ to 4 100 time using time series bootstrap simulation. We carried out the study based on the comparison of model order selection criteria with our well known MML87. We specify three different counts systems as under inferred, correctly inferred and over inferred. MML87 performed better than all other criterion selecting $68(17\%)$ followed by AIC $44(11\%)$ of all other criteria. The figures visualizing the under/correctly/over inferred of the six model selection criteria are depicted in Figs. 1–3.

Table 5
Results for $T = 30$ and true model $P = 4$.

Criteria	$p < \hat{p}$	$p = \hat{p}$	$p > \hat{p}$
AIC	11	2	87
BIC	11	3	86
AICc	11	1	88
HQ	11	3	86
MDL78	11	2	87
MML87	100	–	–

Table 6
Results for criteria under/correctly/over selecting the true models for $p(1, 2, 3, 4)$.

Criteria	$p < \hat{p}$	$p = \hat{p}$	$p > \hat{p}$
AIC	31	44	325
BIC	122	3	275
AICc	22	1	377
HQ	35	14	351
MDL78	33	11	356
MML87	132	68	200

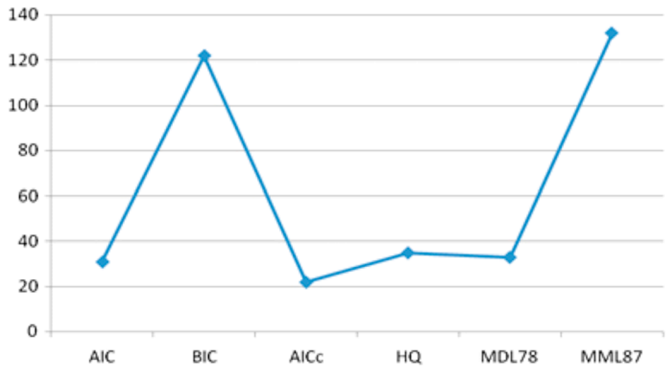

Fig. 1. Showing under inferred model selection.

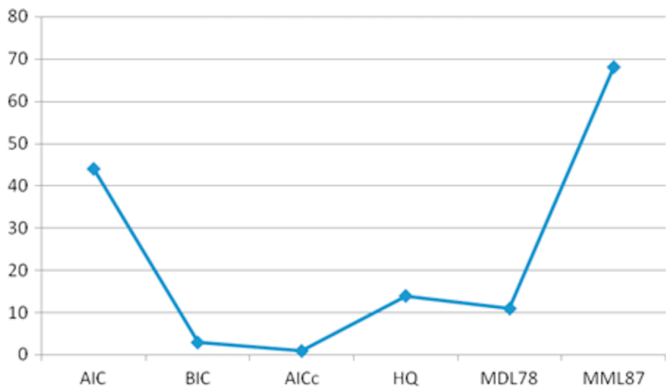

Fig. 2. Showing correctly inferred model selection.

Fig. 1 is a visual representation of under inferred model selection shown in Table 6. From the figure above Minimum Message Length 87 is depicted at 132 followed by BIC, HQ, MDL87, AIC, and AICc at 122, 35, 33, 31 22 respectively.

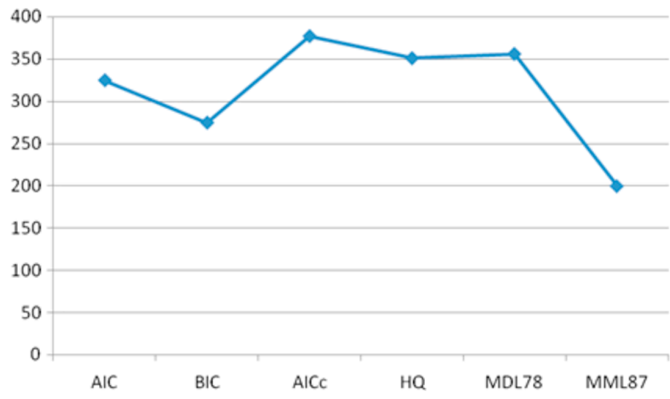

Fig. 3. Showing over inferred model selection.

Fig. 2 is a visual representation of correctly inferred model selection shown in Table 6. From the figure above Minimum Message Length 87 is depicted at 68 followed by AIC, HQ, MDL87, BIC and AICc at 44, 14, 11, 3, 1 respectively.

Fig. 3 is a visual representation of over inferred model selection shown in Table 6. From the figure above MML87 is depicted at 200 followed by BIC, AIC, HQ, MDL87, and AICc at 275, 325, 351, 356, 377 respectively, which revealed the efficiency of MML87.

5. Conclusion

Time series models provide a method for better understanding and an ability to predict future behavior. Model selection is a method of selecting the 'best' model between competing models. The minimum message length was described in detail and its technique was explained. This was compared via empirical bootstrap simulations and quantitative results presented. Our findings shows that MML87 performed faster and better than all other model selection criteria and also more efficient with minimum model order selection. Based on the candidate model explored with autoregressive model of order p and the aggregate true model explored with p maximum of 4, the study shows that MML87 performed better than all other selection criteria with the negative log likelihood function and the mean square prediction error estimated. It was faster, better and efficient in all overall correct model selection, that is, correctly inferred, correctly choosing the true model 68 times (17%) compared to the next best AIC 49 times (11%). These confirm the efficiency of MML87 over other notable model selection criteria in our research.

References

[1] Politis D, Romano J. The stationary Bootstrap. J Amer Statist Assoc 1994;47:1303–13.
[2] Wallace CS. Statistical and inductive inference by minimum message length. In: Springer series in information science and statistics. New York: Springer Science and Business Media; 2005.
[3] Wallace CS, Boulton DM. An information measure for classification. Comput J 1986;11(2):185–94.
[4] Efron B, Tibshirani R. An introduction to the Bootstrap. London: Chapman and Hall/CRC; 1993.
[5] Wallace CS, Dowe DL. Minimum message length and Kolmogorov complexity. Comput J 1999;42(4):270–83.
[6] Efron B. Bootstrap methods another look at the Jacknife. Ann Statist 1979;7:1–26.
[7] Collie MJ, Dowe DL, Fitzgibbon LJ. Stock market simulation and inference technique, 2005. In: Proceedings of the 5th international conference on Hybrid Intelligent Systems (HIS'05), Rio de Janeiro, Brazil, November 6–9. LNAI, vol. 1160. Springer; 2005. p. 213–27. Sydney, Australia, 23–25 October 1996. [in press].
[8] Leger C, Politis D, Romano J. Bootstrap technology and applications. Technometrics 1992;34:378–98.
[9] Politis D, Romano J. A general resampling scheme for triangular arrays of-mixing random variables with applications to the problem of spectral density estimation. Ann Statist 1992;20:1985–2007.
[10] Politis D, Romano J. A nonparametric resampling procedure for multivariate confidence regions in time series analysis. In: Page C, Lepage R, editors. Computing science and statistics. Proceedings of the 22nd symposium on the interface. New York: Springer-Verlag; 1992. p. 98–103.
[11] Wallace CS, Freeman PR. Estimation and inference by compact encoding. J R Stat Soc Ser B Methodol 1987;49:240–52.
[12] Wallace CS, Dowe DL. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. Stat Comput 2000;10:73–83.
[13] Wallace CS, Patrick JD. Coding decision trees. Mach Learn 1993;11:7–22.
[14] Daniel, Enes. Times series analysis. New Jersey: Princeton University Press; 2005.

[15] Fitzgibbon LJ, Allison L, Dove DL. Minimum message length grouping of ordered data. In: Algorithmic learning theory. LNAl, vol. 1968. Springer-Verlag; 2000. p. 56–70.

[16] Fitzgibbon LJ, Dowe DL, Vahid F. Minimum message length autoregressive model order selection. In: Palanaswami M, Chandra Sekhar C, Kimar Venayagamoorthy G, Mohan S, Ghantasala MK, editors. International conference on intelligent sensing and information processing (ICISP), p. 439–444. Chennai, India, 4–7. (IEEE Catalogue Number: 04EX783), www.case.monash.edu.au/dld/Publications/2004/Fitzgibbon+Dowe+Vahid 2004.ref. 2004.

[17] Barreto, Howland FH, Lagerkvist CJ. Introductory econometrics-using Monte Carlo simulation with Microsoft Excel. Eur Rev Agric Econ 2007;34:286–7;
Comley JW, et al. Minimum message length and generalized bayesian nets with asymnetric languages. In: Advances in minimum description length: theory and applications. M.I.T. Press; 2005. April, 2005. [Chapter 11] Final camera ready copy submitted in Oct. 2003.

[18] Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge University Press; 1997.

[19] Hurvich CM, Tsai C. Regression and Time series model selection in small samples. Biometrika 1989;76:297–307.

[20] Kendall MG. Rank correlation methods. 4th ed. London: Griffin; 1948.

[21] Hannan EJ, Quinn BG. The determination of the order of an autoregression. J R Statist Soc Ser B 1979;41(2):190–5.

[22] Kunsch HR. The Jacknife and the Bootstrap for general stationary observations. The Ann of Statistics 1989;17:1217–41.

[23] Needham SL, Dowe DL. Message length as an effective Ockham's Razor in decision tree induction. In: Proc. 8th international workshop on artificial intelligence and statistics, 2001.

[24] Liu RY, Singh K. Moving blocks Jackknife and Bootstrap capture weak dependence. In: Lepage R, Billard L, editors. Exploring the limits of Bootstrap. New York: John Wiley; 1992.

[25] Olatayo TO, Amahia GN, Obilade TO. Bootstrap method for dependent data structure and measure of Statistical precision. J Math Statist 2010;6(2):84–8.

[26] Adeleye AO. Minimum Message Length Autoregressive Model Order Selection. [M.Sc. Dissertation], submitted in the Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria, 2012 [Unpublished].