Journal of Modern Mathematics and Statistics 2 (1): 7-11, 2008
ISSN: 1994-5388

Identification of Optimal Models in Higher Order Integrated Autoregressive Models and Autoregressive Integrated Moving Average Models in the Presence of 2^k-1 Subsets

J.F. Ojo
Department of Statistics, University of Ibadan, Ibadan, Nigeria

Abstract: Significant efforts have been made in the study of the theory of integrated autoregressive models and autoregressive integrated moving average models, but less concerted effort has been made in the identification of optimal models which are of great importance in the forecasting of future values. Little attention has been focused on higher order integrated autoregressive models and autoregressive integrated moving average models which are always characterized by many parameters and the use of subsetting that eliminate redundant parameters in these higher order models. This study therefore focuses on identification of optimal models in higher order integrated autoregressive models and autoregressive integrated moving average models in the presence of 2^k-1 subsets. The parameters of these models were estimated using Marquardt algorithm and Newton-Raphson iterative method and the statistical properties of the derived estimates were investigated. An algorithm was proposed to eliminate redundant parameters from the full order integrated autoregressive models and autoregressive integrated moving average models. To control the parameters of integrated autoregressive models and autoregressive integrated moving average models in the estimation procedure, the elements of 2^k-1 subsets (when k=3) was used. To determine optimal models, residual variance, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were adopted.

Key words: Marquardt algorithm, newton-raphson, forecasting, redundant parameters and residual variance

INTRODUCTION

It may be said that the era of linear time series models began with such linear models as Yule's Autoregressive (AR) models (Yule, 1927). In the past five decades or so, we have seen remarkable successes in the application of linear time series models in diverse fields for example Box and Jenkins (1970), Hannan (1962, 1970), Chatfield (1980) and Anderson (1971, 1977). Nottingham International Time Series Conference in March 1979. These successes are perhaps rather natural in view of the significant contributions of linear differential equation in all branches of science.

Many empirical time series (for example, stock prices) behave as though they had no fixed mean. Even so, they exhibit homogeneity in the sense that, apart from local level, or perhaps local level and trend, one part of the series behaves much like any other part. Models which describe such homogenous non-stationary behaviour can be obtained by supposing some suitable difference of the process to be stationary. We now consider the properties of the important class of models for which the dth difference is a stationary process. These models are called Integrated Autoregressive (IAR) process and Autoregressive Integrated Moving Average (ARIMA) process.

In time series modeling, subset models are often desirable, especially when the data exhibits some form of periodic behaviour. In such cases, fitting full order models often results in the fitted coefficients of some lags being close to zero. Before considering subset modeling, consider the problem of fitting a full model of order k.

MATERIALS AND METHODS

Integrated autoregressive model: Integrated autoregressive models (p, d, 0,) is given as $\Psi(B) X_t = \phi(B) \nabla^d X_t$ where, $\phi(B) = 1-\phi_1 B-\phi_2 B^2.....-\phi_p B^p$

$$X_t = \Psi_1 X_{t-1}+.....+ \Psi_{p+d} X_{t-p-d} + e_t \quad (1)$$

$\phi(B)$ will be called the autoregressive operator; it is assumed to be stationary, that is the roots of $\phi(B) = 0$ lie outside the unit circle. $\Psi(B) = \nabla^d \phi(B)$ will be called the generalized autoregressive operator; it is a non stationary operator. Ψ_i are the parameters of the integrated autoregressive part of the model and the e_t are error terms. The error terms e_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Autoregressive integrated moving average moodel: In statistics, an Autoregressive Integrated Moving Average (ARIMA) model is a generalisation of an Autoregressive Moving Average or (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series. The model is generally referred to as an ARIMA (p,d,q) model where p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated and moving average parts of the model, respectively.

Given a time series of data X_t where t is an integer index and the X_t are real numbers, then an ARIMA(p,d, q) model isgiven by $\Psi(B) X_t = \phi(B) \nabla^d X_t = \theta(B) e_t$

Where:

$\phi(B) = 1-\phi_1 B-\phi_2 B^2 .....-\phi_p B^p$ and $\theta(B)\ 1-\theta_1 B-\theta_2 B^2.....-\theta_q B^q$

$$X_t = \Psi_1 X_{t-1}+.....+\Psi_{p+d}X_{t-p-d}+e_t-\theta_1 e_{t-1} -...-\theta_q e_{t-q} \quad (2)$$

Ψ_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the e_t are error terms. The error terms e_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Description of an algorithm for fitting of subset integrated autoregressive and autoregressive integrated moving average models: We fit full integrated autoregressive and autoregressive integrated moving average models of various orders and choose that model for which Akaike Information (AIC) is minimum. Let average model be p+d+q and let the models be $X_t = \Psi X_{t-1} +.....+\Psi_{p+d}X_{t-p-d}+e_t$ denoted by IA (p,d) and $X_t = \Psi_1 X_{t-1} +.....+\Psi_{p+d}X_{t-p-d}+e_t-\theta_1 e_{t-1} -...-\theta_q e_{t-q}$ denoted by ARIMA (p,d,q), respectively.

Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(1)}$ and its Akaike Information (AIC) be equal to AIC(1) for integrated autoregressive model and for autoregressive integrated moving average model, let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(1)}$ and its (AIC) be equal to AIC. The estimation of models is done by using Malquardt algorithm and Newton-Raphson Iterative method. Having fitted the full models, we can now fit the best subset models by considering the 2^k-1 subsets using the fitted full models with minimum AIC. We consider the elements of 2^k-1 subsets using the approach of Hagan and Oyetunji (1980) and choose that model for which AIC is minimum. Let the best subset Integrated Autoregressive model be $X_t = \Psi_{m1}X_{t-m1}+.....+\Psi_{ml+d}X_{t-ml-d}+e_t$ where m_1, m_2, $.....m_{l+s}$ are subsets of the integers (1, 2, …., p+d). Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(2)}$ and the AIC value be AIC(2), AIC(2)≤AIC(1). Also let the best subset autoregressive integrated moving average model be $X_t = \Psi_{n1}X_{t-n1}+.....+\Psi_{nl+d}X_{t-nl-d}+e_t-\theta_{k1}e_{t-k1}-...-\theta_{kq}\ e_{t-kq}$ where n_1 n_2 $.....n_{l+d}$; k_1,$.....k_q$ are subsets of the integers (1, 2, …., p+d+q). Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(21)}$ and the AIC value be AIC (22), AIC (22)≤AIC(11). This is our subset integrated autoregressive and subset autoregressive integrated moving average models.

Estimation technique: The estimation of the parameters of the models under consideration is as follows:

$$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d}X_{t-p-d} + e_t \quad (3)$$
$$e_t = X_t - \psi_1 X_{t-1} - ..... - \psi_{p+d}X_{t-p-d}$$

$$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d}X_{t-p-d} + e_t - \theta_1 e_{t-1} - ... - \theta_q e_{t-q}$$
$$e_t = X_t - \psi_1 X_{t-1} - ..... - \psi_{p+d}X_{t-p-d} + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} \quad (4)$$

Minimizing the function

$$Q(G) = \sum_{t=m1}^{N} e_t^2$$

with respect to the parameters $(\psi_{l_1}, \psi_{2_2}, ..., \psi_{k_1}; \theta_1, \theta_2, ..., \theta_3)$. Let $G^T = (\psi_{l_1}, \psi_{2_2}, ..., \psi_{k_1}; \theta_1, \theta_2, ..., \theta_3)$. For convenience, we shall write,

$$G_1 = \psi_{l_1}, G_2 = \psi_{2_2}, ..., G_k = \psi_{k_p},$$
$$G_{k+1} = \theta_{l_1}, ..., G_R = \theta_m$$

Where, R = k + m
The partial derivatives of Q(G)are

$$G_i = \frac{dQ(G)}{dG} = 2\sum_{t=l_1}^{N} e_t \frac{de_t}{dG_i}, \quad (5)$$

$$H(G) = \frac{d^2Q(G)}{dG_i dG_j} = 2\sum_{t=1}^{N}\left(\frac{de_t}{dG_i}\right)\left(\frac{de_t}{dG_j}\right) + 2\sum_{t=1}^{N} e_t \cdot \frac{d^2e_t}{dG_i dG_j}, \quad (6)$$

where the partial derivatives satisfy the recursive equations

$$\frac{de_t}{d\psi_1} = X_{t-l_1} \quad (7)$$

$$\frac{de_t}{d\psi_{k_r}} = X_{t-p_r}, (p = 2,3,...,k) \quad (8)$$

$$\frac{de_t}{d\theta_1} = e_{t-1}$$

$$\frac{de_t}{d\theta_m} = e_{t-q} (q = 2,3,.....,m) \quad (9)$$

In evaluating the second order partial derivatives we approximate

$$H(G) = 2\sum_{t=1_1}^{N}\left(\frac{de_t}{dG_i}\right)\left(\frac{de_t}{dG_i}\right)$$

as is done in Marquardt algorithm.

Now let $V'(G) = \left(\frac{dQ(G)}{dG_1}, \frac{dQ(G)}{dG_2}, ..., \frac{dQ(G)}{dG_R}\right)$

and let $H(G) = \frac{d^2Q(G)}{dG_i dG_j}$,

Expanding V ($\hat{G}$) near $\hat{G}$ = G in a Taylor series, we obtain

$$0 = V(G) + H(G)(\hat{G} - G)$$

Rewriting this equation, we get $(\hat{G} - G) = -H^{-1}(G)V(G)$ and thus obtain the Newton-Raphson iterative equation

$$G^{(k)} = G^{(k+1)} + H^{-1}(G^{(k)})V(G^{(k)})$$
$$G^{(k)} = G^{(k+1)} + H^{-1}(G^{(k)})V(G^{(k)}) \quad (10)$$

Where, $G^{(k)}$ is the set of estimates obtained at kth stage of iteration. The estimates obtained by the above iterative equations usually converge. For starting the iteration, we need to have good sets of initial values of the parameters.

Residual variance: Residual variance or unexplained variance is part of the variance of any residual. In analysis of variance and regression analysis, residual variance is that part of the variance which cannot be attributed to specific causes. The unexplained variance can be divided into two parts. First, the part related to random, everyday, normal, free will differences in a population or sample. Among any aggregation of data these conditions equal out. Second, the part that comes from some condition that has not been identified, but that is systematic. That part introduces a bias and if not identified can lead to a false conclusion.

Akaike Information Criteria (AIC): The Akaike Information Criterion (AIC) (pronounced ah-kah-ee-keh), developed by Hirotsugu Akaike (1971) and proposed in Akaike (1974) is a measure of the goodness of fit of an estimated statistical model. It is grounded in the concept of entropy. The AIC is an operational way of trading off the complexity of an estimated model against how well the model fits the data.

In the general case, the AIC is

$$AIC = 2k - 2\ln(L)$$

Where, k is the number of parameters and L is the likelihood function.

Over the remainder of this entry, it will be assumed that the model errors are normally and independently distributed. Let n be the number of observations and RSS be the residual sum of squares. Then AIC becomes

$$AIC = 2k + n\ln(RSS/n) \quad (11)$$

Increasing the number of free parameters to be estimated improves the goodness of fit, regardless of the number of free parameters in the data generating process. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. The preferred model is the one with the lowest AIC value. The AIC methodology attempts to find the model that best explains the data with a minimum of free parameters. The AIC penalizes free parameters less strongly than does the Schwartz.

Bayesian information criterion: In statistics, the Bayesian Information Criterion (BIC) is a statistical criterion for model selection. The BIC is sometimes also named the Schwarz criterion, or Schwarz Information Criterion (SIC). It is so named because prof. Gideon E. Schwarz gave a Bayesian argument for adopting it.
Let:

n = The No. of observations, equivalently, the sample size.
k = The number of free parameters to be estimated. If the estimated model is a linear regression, k is the number of regressors, including the constant.
RSS = The residual sum of squares from the estimated model.
L = The maximized value of the likelihood function for the estimated model.

The formula for the BIC is:

$$BIC=\left(\frac{-2.\ln L+k\ln(n)}{n}\right)$$

Under the assumption that the model errors or disturbances are normally distributed, this becomes:

$$BIC=\ln\left(\frac{RSS}{n}\right)+k\left(\frac{\ln n}{n}\right) \quad (12)$$

Given any two estimated models, the model with the lower value of BIC is the one to be preferred. The BIC is a decreasing function of RSS, the goodness of fit and an increasing function of k. The BIC penalizes free parameters more strongly than does the Akaike Information criterion.

RESULTS AND DISCUSSION

The real series used was sunspot numbers which was considered at three levels that is when t = 50, 150 and 250. This series is a non-stationary series and models considered applied. The choice of t was to allow us to have the optimal models. The optimal models identified are discussed in turn.

Fitting of full and subset ARIMA (p, 1, 0): ARIMA (p, 1, 0) of orders 1 to 10 was fitted using the real series. The choice of the best order is made on the basis of AIC and BIC and the minimum AIC and BIC is the best model and this was found when p = 7. The fitted model is:

$$X_t = 0.165220X_{t-1}-0.469042X_{t-2}-0.153955X_{t-3}-0.315054X_{t-4}-0.275924_{t-5}-0.366829X_{t-6}-0.324887X_{t-7}+e_t$$

There are $2^7-1 = 127$ possible subsets. The choice of the order is made on the basis of minimum AIC and BIC and having considered the 127 possible subsets, it was found that AIC and BIC is minimum in the following model

$$X_t = 0.314548X_{t-1}-0.458429X_{t-2}-0.302114X_{t-4}-0.220568X_{t-5}-0.86159X_{t-6}+e_t$$

Fitting of full and subset ARIMA (p, 1, (1, 2, 3)): ARIMA (p, 1, (1, 2, 3)) of orders 1 to 10 were fitted using the real series. The choice of the best order is made on the basis of AIC and BIC and the minimum AIC and BIC is the best model and this was found when p = 6 and q = 1, 2, 3. The fitted model is:

$$X_t = 0.872827X_{t-1}+0.006059X_{t-2}-0.528694X_{t-3}+0.069292X_{t-4}-0.126321X_{t-5}-0.003146X_{t-6}-0.964514e_{t-1}-0.704616e_{t-2}+0.884318e_{t-3}+e_t$$

There are $2^6-1 = 63$ possible subsets. The choice of the order is made on the basis of minimum AIC and BIC and having considered the 63 possible subsets, it was found that AIC and BIC is minimum in the following model

$$X_t = 0.872827X_{t-1}+0.006059X_{t-2}-0.528694X_{t-3}+0.069292X_{t-4}-0.126321X_{t-5}-0.003146X_{t-6}-0.964514e_{t-1}-0.704616e_{t-2}+0.884318e_{t-3}+e_t$$

Table 1: Full IA and ARIMA models at different levels of t

	T = 50				T =150				T = 250			
Time Model	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)
R^2	0.59	0.62	0.56	0.70	0.52	0.53	0.51	0.56	0.50	0.50	0.52	0.53
$\bar{R}^2$	0.52	0.57	0.51	0.63	0.50	0.51	0.50	0.53	0.49	0.49	0.51	0.52
RV	204.5	200.2	245.3	153.4	219.8	213.9	217.2	201	308.5	308.6	306.4	284.8
S.E	17.09	16.23	17.30	15.02	15.63	15.47	15.53	15.11	18.19	18.22	18.21	17.90
AIC	8.67	8.54	8.64	8.44	8.38	8.37	8.38	8.34	8.67	8.68	8.68	8.65
BIC	8.96	8.78	8.84	8.81	8.53	8.54	8.54	8.54	8.79	8.81	8.83	8.81
F	8.32	12.56	12.83	10.07	24.19	21.67	21.87	18.62	32.72	28.53	25.58	24.73
P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2: Subset IA and ARIMA Models at Different Levels of t

	T = 50				T =150				T = 250			
Time Model	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)	P, 1,0	P, 1,1	P, 1, (1,2)	P, 1, (1, 2, 3)
R^2	0.57	0.62	0.56	0.70	0.51	0.53	0.53	0.57	0.50	0.51	0.52	0.53
$\bar{R}^2$	0.52	0.59	0.51	0.63	0.49	0.51	0.51	0.54	0.48	0.48	0.49	0.50
RV	224.5	202.4	245.3	153.4	225.2	213.9	222.0	197.5	309.5	308.7	306.4	294.8
S.E	17.19	15.91	17.3	15.02	15.70	15.42	15.48	14.93	18.18	18.20	18.21	17.90
AIC	8.64	8.46	8.64	8.44	8.38	8.36	8.35	8.31	8.67	8.67	8.68	8.65
BIC	8.84	8.62	8.84	8.81	8.48	8.5	8.45	8.49	8.77	8.79	8.83	8.81
F	12.38	21.66	12.83	10.07	35.13	25.47	38.6	21.77	38.1	32.69	25.58	24.73
P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This coincides with the full model: Elements of 2^k-1 subset when k = 3 that is 1; 1, 2 and 1, 2, 3 were used in full and subset models in the Table 1 and 2 the optimal model was recorded at t = 50 w ith model ARIMA (p, 1, (1, 2, 3)) with lowest residual variance and highest R^2. For the integrated autoregressive models the optimal model occurred at t = 50 as well with model IA (p, 1, 0) with lowest residual variance. Because of the elements of 2^k-1 subsets, we saw in our result that full ARIMA and subset ARIMA coincide which is not always the case. The study is saying that when these elements of 2^k-1 are used we should stop at the fitting of the full ARIMA and the rigour of fitting the subset will not occur again. Also we have seen from our findings that working with smaller series will not deny us of having optimal model and finally ARIMA models perform better that IA models.

CONCLUSION

We have seen in this study the impact of elements of 2^k-1 subsets in the models considered. Also, we have been able to identify optimal models which are of great importance in the forecasting of future values. We have considered models of higher order and have been able to remove the redundant parameters.

REFERENCES

Anderson, T.W., 1977. Estimation for autoregressive moving average models in the time and frequency domains. Ann. Stat., 5: 842-865.

Anderson, T.W., 1971. The statistical analysis of time series. New York and London Wiley.

Box, G.E.P. and G.M. Jenkins, 1970. Time series, forecasting and control. Holden-Day: San Francisco.

Chatfield, C., 1980. The analysis of time series: An Introduction. Chapman and Hall.

Haggan, V. and O.B. Oyetunji, 1980. On the selection of subset autoregressive time series models. UMIST Technical Report No. 124 (Department of Mathematics).

Hannan, E.J., 1962. Time series analysis. Methuen Monographs.

Hannan, E.J., 1970. Multiple time series. John Wiley: New York.

Yule, G.U., 1927. On the method of investigating periodicities in disturbed series with special reference to wolfer's sunspot numbers. Phil. Trans. Roy. Soc. London (A), pp: 226-267.