Pakistan Journal of Applied Sciences 2(7): 783-788, 2002
© Copyright by the Science Publications, 2002

Performance of Order Selection Criteria for Short Time Series

Zazli Chik
Pusat Asasi Sains, University of Malaya, , 50603 Kuala Lumpur, Malaysia

Abstract: The order of fitted time series models is unknown and constitutes, in effect, additional unknown parameters for which suitable values have to be estimated from the observed data. The approached pioneered by Akaike and Parzen involving the use of an order selection criterion provides a remarkable breakthrough which transforms the order selection problem from one of hypothesis testing to that of estimation. Different authors use different methods of determining the order of their fitted time series models. Various order selection criteria will be used in a simulation study on fitted short time series models and the performance of each of the order selection criteria in estimating the correct order are investigated.

Key Words: Order Selection, Akaike's FPE, AIC, BIC, Schwarz's Criterion, Hannan and Quinn's Criterion

Introduction

A number of approaches to determining the value of (p, q) for the ARMA(p, q) model from the observed data have been suggested. In the case of AR model fitting the analogy with multiple regression analysis may be followed and the problem is approached via hypothesis testing, that is test the null hypothesis that the model is AR(p) against the alternative hypothesis that it is AR(p+1), and continue increasing the value of p until the test gives a non significant result. However, the hypothesis-testing procedures are only suitable when the models under the null and alternative hypothesis are prescribed *a priori* (Bhansali, 1993).

The approached pioneered by Akaike (1970) and Parzen (1974) involving the use of an order selection criterion provides a remarkable breakthrough which transforms the order selection problem from one of hypothesis testing to that of estimation.

Whittle (1963) introduced an order selection technique based on residual variance plots. In this technique, it is assumed that the true model is autoregressive of finite (but unknown) order and an AR(p) model is fitted to the data. If a value of p smaller than the true order is choosen then the (unbiased) estimated residual variance, $\hat{\sigma}_\varepsilon^2$ is expected to be larger than the true residual variance, σ_ε^2, since the additional terms omitted from the model would explain a further part of the variance of X_t. On the other hand, once the value of p reaches the true order any further increase in p will not significantly reduce the residual variance. Hence, if a sequence of models of increasing order are fitted, $\hat{\sigma}_\varepsilon^2$ evaluated in each case, and then plotted against p, the graph is expected to decrease at first and then level out at the point where p approaches the true order. Jenkins and Watts (1968) also used this technique and suggested that the same technique can be applied to order selection for MA models and ARMA models.

A more refined version of the residual variance plots was developed by Akaike (1969) for AR order selection. In Akaike's procedure, autoregressive models of increasing order were fitted. For each order p, p =0,1, . . . , m, where m is a preassigned upper bound, the value of an *order selection criterion* (based on the estimated residual variance and the order p is calculated and the estimated order is the value of p that minimises the order selection criterion. As well as Akaike's original order selection criterion, other order selection criterion have been proposed by Akaike, Schwarz, Hannan and Quinn, and Parzen and these are described below.

Order Selection Criteria

Akaike's FPE and AIC: Akaike (1969) proposed the order selection criterion defined by

$$FPE(p) = \frac{N+p}{N-p}\hat{\sigma}_p^2$$

where N is the number of observations to which the model is fitted, and $\hat{\sigma}_p^2$ is the maximum likelihood estimate of the variance of the residuals based on the pth order model. The value of p at which the FPE attains its minimum is the estimated order of the model. The expression FPE stands for "final prediction error". The motivation for this terminology can be found in Priestley (1981).

Akaike (1974) introduced a new expression called AIC (Akaike's information criterion). This general criterion can be used for statistical model identification in a wide range of situations and is not restricted to the time series context. When a model involving p independently adjusted parameters is fitted to data, the AIC is defined by,

$$AIC(p) = -2 \log \text{(maximum likelihood)} + 2p$$

For an AR model, this criterion becomes

$$AIC(p) = \mathrm{N}\log(\hat{\sigma}_p^2) + 2p$$

Again, increasing order, p, of autoregressive models were fitted. The procedure suggests the choice of the p that minimises $AIC(p)$ for p =0,1, ... ,m, where m is a preassigned upper bound to the order.

The FPE and AIC are asymptotically equivalent procedures as far as the determination of p is concerned since asymptotically $AIC(p) \approx$ Nlog($FPE(p)$) (Akaike,1979,Priestley, 1981). FPE criterion was originally introduced for autoregressive order determination only, the AIC criterion is quite general and can be applied to all the standard models including MA and ARMA models. For example, for the ARMA model, the AIC criterion is

$$AIC(p,q) = (\mathrm{N}-p)\log(\hat{\sigma}_p^2) + 2(p+q+1)$$

The orders of the AR and MA operators are determined by computing the AIC criterion over a selected grid of values of p and q and choosing those values of p and q at which AIC attains its minimum. The AIC criterion has largely superseded the FPE procedure and is generally accepted as one of the most reliable procedures for order determination.

Shibata (1976) has investigated the asymptotic properties of Akaike's AIC and his investigation shows that the estimate is not consistent but overestimates the order asymptotically, with a non-zero probability. In spite of this result Akaike's AIC criterion is widely used. There are two reasons why Shibata's result is not an important result in practice. Firstly the result is an asymptotic result. Secondly the underlying process is unlikely to be an exact autoregressive model in practice and the fitted autoregressive model is being used as an approximation.

As discussed by Bhansali (1993), the FPE criterion may be viewed as a special case, with α=2, of an extended criterion,

$$FPE_\alpha(p) = \hat{\sigma}_p^2(1+\alpha p/\mathrm{N}),$$

in which α is a positive constant. The AIC criterion may also be generalised to the following criterion (Akaike, 1979).

$$AIC_\alpha(p) = \mathrm{N}\,log(\hat{\sigma}_p^2) + \alpha p,$$

in which α is a positive constant. For a fixed α the AIC α and FPEα are closely related. AIC and FPE do not estimate p consistently, but asymptotically the probability of selecting p correctly increases as α increases (Bhansali, 1988). However, Hurvich and Tsai (1993) observe that if N is small but the ratio m/N is negligibly small where m is the upper bound, AIC may select a highly non-parsimonious model.

Akaike's BIC: Akaike (1979) has developed a Bayesian extension of the minimum AIC procedure called BIC and the order selection criterion is defined by

$$BIC(p) = \mathrm{N}\,log\,\hat{\sigma}_p^2 - (\mathrm{N}-p)log(1-\frac{p}{\mathrm{N}}) + p\,log\,\mathrm{N} + p\,log\left\{p^{-1}\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_p^2}-1\right)\right\}$$

where, as before, $\hat{\sigma}_p^2$ is the estimate of the variance of the residuals based on the p th parameter model and $\hat{\sigma}_X^2$ is the raw sample variance of the observations. When p is small relative to N, the approximation $\{-(\mathrm{N}-p)\log[1-(p/\mathrm{N})]\} \approx p$ may be used, so that

$$BIC(p) \approx \mathrm{N}\,log\,\hat{\sigma}_p^2 + (p + p\,log\,\mathrm{N}) + p\,log\left\{p^{-1}\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_p^2}-1\right)\right\}$$

The last term on the left is independent of N and an approximate expression for BIC that will be used in the simulation study is

$$BIC(p) \approx \mathrm{N}\,log(\hat{\sigma}_p^2) + p\{1 + log\,\mathrm{N}\}$$

The term $p\{1+log\,\mathrm{N}\}$ has the effect of increasing the weight attached to the "penalty term" (which takes account of the number of parameters in the model), and consequently the minimisation of BIC leads, in general, to lower model orders than those obtained by minimising AIC. Shibata (1976) has shown that the AIC criterion tends to overestimate the true order of an autoregressive model, but that the estimated order obtained using the BIC criterion may well underestimate the true order.

Schwarz's Criterion: Schwarz (1978) suggested the order selection criterion

$$S(p) = \mathrm{N}\,log\,\hat{\sigma}_p^2 + p\,log\,\mathrm{N},$$

which is similar to Akaike's BIC in terms of its dependence on logN. In fact, if the approximate expression for BIC is used, this relationship can be written

BIC(p) $\approx$ S(p) + O(p),

where O(p) denotes a term which is functionally independent of N.

Hannan-Quinn's Criterion: The order selection criterion proposed by Hannan and Quinn, $HQ(p)$, is of the same type as that proposed by Akaike, namely based on the minimisation of $\log\hat{\sigma}_p^2 + p\,C_\mathrm{N}$, where C_N is a quantity dependent on N. However, it was proposed so that it would be strongly consistent for the estimated order and so that C_N decreases as fast as possible. The Hannan-Quinn criterion is

$$HQ(p) = \mathrm{N}\,log\,\hat{\sigma}_p^2 + 2pc\,log\,log\,\mathrm{N}, \quad c > 1$$

Hannan and Quinn (1979) commented that it is not reasonable to expect any definite conclusion on which criterion is the best. If N is large and an autoregression is thought to be a good approximation then the use of $HQ(p)$ would have something to recommend it. This might not be true in other circumstances. The method provides some compromise between procedures such as Akaike's BIC with $C_N = N^{-1} \log N$ designed, in a Bayesian analysis, for a true autoregressive situation and procedures such as Akaike's AIC designed for fitting an autoregression where the true structure may be more general.

Parzen's CAT: Parzen (1974) proposed a different order selection criterion which he calls CAT (criterion for autoregressive transfer function). This criterion takes the following form

$$CAT(p) = \begin{cases} \left(\frac{1}{N}\sum_{j=1}^{p}\frac{1}{\hat{\hat{\sigma}}_j^2}\right) - \frac{1}{\hat{\hat{\sigma}}_p^2}, & p = 1,2,3,\ldots, \\ -(1+(1/N)), & p = 0, \end{cases}$$

where $\hat{\hat{\sigma}}_j^2$ is the unbiased estimate of the innovation variance when an AR model of order j is fitted to the data. The value p is chosen as that value for which $CAT(p)$ attains its minimum value. However, calculation of $CAT(p)$ is slightly different from the other criteria and it is not used in the simulation study. Also, Parzen (1974) pointed out that CAT and AIC often give identical results.

Results and Discussion

Simulation: To compare the performance of the various criteria discussed previously on small and moderate size sample, realisations of the processes listed below were generated. These are autoregressive processes from Coates and Diggle (1986), Stoica (1990), Swanapoel and van Wyk, (1986) and Newton and Pagano, (1984).

The following types of stationary autoregressive processes were used

(a)AR(1), α_1 =0.2,0.4, 0.6 and 0.8 (Coates and Diggle);

(b) AR (2), α_1 =0 α_2 =0.2,0.4,0.6 and 0.8(Coates and Diggle);

(c)AR (3), α_1 =0.6, α_2= 0.65 and α_3 = -0.63, (Stoica);

(d) AR(5), α_1 =-1.7, α_2= -2.4, α_3 = -1.63, α_4 = -0.872 and α_5 = -0.168,

(Swanapoel and van Wyk, and Newton and Pagano)

The performance of the following order selection criteria on the above types of stationary autoregressive processes were compared.

1. FPE
2. AIC
3. BIC
4. Schwarz's criterion (Schwarz)
5. Hannan and Quinn's criterion (HQ)

One hundred replicates of each type of processes, (a) - (d), were generated and in each case samples size of N=64 and N=256 were considered. The upper bound for the order, m, was set at eleven. For each criterion the frequency distribution was obtained. To reduce the impact of starting-up values, 100 presample observations were generated for each replicate of each of the autoregressive processes. Least squares estimates were computed for each of the sets of data using the last p of the 100 presample values in addition to the N sample values when estimating the parameters for an AR(p) process. This was done because this is equivalent to maximum likelihood estimation if the last p of the 100 presample values are treated as fixed initial values.

The number of times (out of 100 replications) the orders were estimated correctly for an AR(1) process are given in Table 1 (a - d) for a sample size of N=64 and in Table 2 (a - d) for a sample size of N=256. The BIC gave the best results, the correct order was chosen 96% - 100% of the time for the sample size of N=64. The Schwarz's criterion gave similar results (94% - 98% correct) followed by Hannan and Quinn's criterion, FPE and AIC. The BIC and the Schwarz's criterion gave the best results (almost identical) for a sample size of N=256 (99% - 100% correct for the BIC and 96% - 100% correct for the Schwarz's criterion). This is followed by the results for Hannan and Quinn's criterion, FPE and AIC.

The results for an AR(2) process are given in Table 1(e –h) for a sample size of N=64 and in Table 2 (e – h) for a sample size of N=256. The BIC and the Schwarz's criterion gave almost identical results (18% - 99% correct for the BIC and 22% - 97% correct for the Schwarz's criterion) for a sample size of N=64. The Hannan and Quinn's criterion gave better results than the BIC and the Schwarz's criterion for AR(2) with parameters α_1 =0, α_2 =0.2 and α_1 =0, α_2 =0.4 whereas the BIC gave better results than the Schwarz's criterion and the Hannan and Quinn's criterion for AR(2) with parameters α_1 =0, α_2 =0.6 and α_1 =0, α_2 =0.8. Again, this is followed by the results for FPE and AIC. Both the BIC and the Schwarz's criterion gave the best results (almost identical) for a sample size of N=256 (69% - 98% correct for the BIC and 73% - 98% correct for the Schwarz's criterion), followed by Hannan and Quinn's criterion, FPE and AIC. In general, the results for a sample size of N=256 are better than the results for a sample size of N=64.

Again, the BIC and the Schwarz's criterion gave the best and almost identical results (94% correct for the BIC and 91% correct for the Schwarz's criterion) for the AR(3) process of a sample size of N=64 as shown in Table 1(i). Similar and better results were obtained for a sample size of N=256 as shown in Table 2(i) (96% correct for the BIC and 95% correct for the Schwarz's criterion).

None of the criteria performed well for the case of AR(5); the results are given in Table 1(j) for a sample size of N=64 and in Table 2(j) for a sample size of N=256. However, both the FPE and the AIC gave the best results (both 28% correct) in the case of the sample size of N=64 and Hannan and Quinn's criterion (80% correct) gave the best results in the case of the sample size of N=256.

Table 1: Frequency Table of Orders Estimated for AR Series Using Various Criteria (100 Replications) for Sample Size of 64

	Criterion Estimated order	FPE	AIC	BIC	Schwarz	HQ
(a) AR(1), $\alpha_1 = 0.2$	1	65	65	96	94	85
	2	15	15	3	4	9
	3	7	6	1	2	3
	4 - 11	13	14	0	0	3
(b) AR(1), $\alpha_1 = 0.4$	1	73	72	98	96	87
	2	12	13	2	2	6
	3	8	8	0	1	4
	4 - 11	7	7	0	1	3
(c) AR(1), $\alpha_1 = 0.6$	1	73	73	100	98	84
	2	6	6	0	1	7
	3	7	7	0	1	4
	4 - 11	14	14	0	0	5
(d) AR(1), $\alpha_1 = 0.8$	1	68	68	98	94	85
	2	11	11	2	6	8
	3	4	4	0	0	1
	4 - 11	17	17	0	0	6
(e) AR(2), $\alpha_1 = 0$, $\alpha_2 = 0.2$	1	42	42	82	78	60
	2	28	28	18	22	28
	3	11	11	0	0	8
	4 - 11	19	19	0	0	4
(f) AR(2), $\alpha_1 = 0, \alpha_2 = 0.4$	1	4	4	24	18	7
	2	64	64	73	78	81
	3	12	12	3	3	6
	4 - 11	20	20	0	1	6
(g) AR(2), $\alpha_1 = 0, \alpha_2 = 0.6$	1	0	0	0	0	0
	2	78	78	98	97	90
	3	10	10	2	2	6
	4 - 11	12	12	0	1	4
(h) AR(2), $\alpha_1 = 0$, $\alpha_2 = 0.8$	1	0	0	0	0	0
	2	78	78	99	97	91
	3	7	7	1	3	6
	4 - 11	15	15	0	0	3
(i) AR(3) $\alpha_1 = 0.6$, $\alpha_2 = 0.65$ and $\alpha_3 = -0.63$,	1-2	1	1	3	2	2
	3	71	71	94	91	82
	4	13	13	2	4	8
	5	6	6	1	3	5
	6 -11	9	9	0	0	3
(j) AR(5) $\alpha_1 = -1.7, \alpha_2 = -2.4$, $\alpha_3 = -1.63$, $\alpha_4 = -0.872$ and $\alpha_5 = -0.168$,	1-3	0	0	0	0	0
	4	50	50	83	73	65
	5	28	28	16	24	25
	6	10	10	1	2	6
	7 -11	12	12	0	1	4

Table 2: Frequency Table of Orders Estimated for AR Series Using Various Criteria (100 Replications) for Sample Size of 256

	Criterion Estimated order	FPE	AIC	BIC	Schwarz	HQ
(a) AR(1), $\alpha_1 = 0.2$	1	73	73	99	96	92
	2	14	14	1	2	5
	3	3	3	0	2	3
	4 - 11	10	10	0	0	0
(b) AR(1), $\alpha_1 = 0.4$	1	81	81	99	99	92
	2	9	9	1	1	6
	3	5	5	0	0	2
	4 - 11	5	5	0	0	0
(c) AR(1), $\alpha_1 = 0.6$	1	75	75	100	100	97
	2	8	8	0	0	2
	3	4	4	0	0	0
	4 - 11	13	13	0	0	1
(d) AR(1), $\alpha_1 = 0.8$	1	84	84	100	98	95
	2	6	6	0	2	3
	3	2	2	0	0	1
	4 - 11	8	8	0	0	1
(e) AR(2), $\alpha_1 = 0, \alpha_2 = 0.2$	1	4	4	30	24	9
	2	64	64	69	73	82
	3	16	16	1	3	9
	4 - 11	16	16	0	0	0
(f) AR(2), $\alpha_1 = 0, \alpha_2 = 0.4$	1	0	0	0	0	0
	2	64	64	97	96	87
	3	17	17	3	4	10
	4 - 11	19	19	0	0	3
(g) AR(2), $\alpha_1 = 0, \alpha_2 = 0.6$	1	0	0	0	0	0
	2	74	74	98	98	89
	3	12	12	2	2	10
	4 - 11	14	14	0	0	1
(h) AR(2), $\alpha_1 = 0$, $\alpha_2 = 0.8$	1	0	0	0	0	0
	2	74	74	98	96	91
	3	12	12	2	4	7
	4 - 11	14	14	0	0	2
(i) AR(3) $\alpha_1 = 0.6$, $\alpha_2 = 0.65$ and $\alpha_3 = -0.63$,	1-2	0	0	0	0	0
	3	72	72	96	95	82
	4	13	13	3	4	8
	5	6	6	1	1	6
	6 -11	9	9	0	0	4
(j) AR(5) $\alpha_1 = -1.7, \alpha_2 = -2.4$, $\alpha_3 = -1.63$, $\alpha_4 = -0.872$ and $\alpha_5 = -0.168$,	1-3	0	0	0	0	0
	4	4	4	48	40	14
	5	68	68	52	58	80
	6	9	9	0	2	5
	7 -11	19	19	0	0	1

Different authors use different methods of determining the order of their fitted time series models. For example, Martin (1980) uses Akaike's order selection criterion, AIC in robust estimation whereas Swanapoel and van Wyk (1986) use Hannan and Quinn's order selection criterion in spectral density function estimation. Newton and Pagano (1984) use Parzen's order selection criterion, CAT. In the application literature other, less formal, methods are sometimes used. For example, Kane and Trivedi (1991) use autoregressive models with the order equal to 33%, 50%, 67% and 80% of the data. They dismiss the use of Akaike's FPE with the comment "this criterion often fails".

In this simulation study, the BIC and the Schwarz's criterion gave almost identical results and FPE and AIC also gave almost identical results as expected (see Section on order selection criteria). In general, both the BIC and the Schwarz's criterion gave the best results for autoregressive order selection followed by Hannan-Quinn's criterion, FPE and AIC. In conclusion, it is best to use different order selection criterion for estimating the order of fitted short time series models.

Other references to the works in which the order selection techniques are reviewed include Hurvich and Tsai (1989, 1993), Koreisha and Pukkila (1993), Paparoditis and Streitberg (1991), Bhansali (1988) and Hannan and Quinn (1979).

References

Akaike, H., 1969. Fitting Autoregressive Model for Prediction, Annals Institute of Statistical *Mathematics, 21: 243-247.*

Akaike, H., 1970. Statistical Predictor Identification, Annals Institute of Statistical Mathematics. 22: 203-217.

Akaike, H., 1974. A New Look at the Statistical Model Identification, I.E.E.E Trans. Automatic Control. AC-19: 716-722.

Akaike, H., 1979. A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting, Biometrika. 66: 237-242.

Bhansali, R. J., 1988. Consistent Order Determination For Process With Infinite Variance, J. of the Royal Statistical Society Series B, 50:46-60.

Bhansali, R. J., 1993. Order Selection for Time Series Model: A Review, in Developments in Time Series Analysis By Rao, T. S., Chapman and Hall, London.

Coates, D. S. and P. J. Diggle, 1986. Tests for Comparing Two Estimated Spectral Densities, J. of Time Series Analysis. 7:7-20.

Hannan, E. J. and B. G. Quinn, 1979. The Determination of the Order of an Autoregression, J. of the Royal Statistical Society Series B. 41:190-195.

Hurvich, C. M. and L. Tsai, 1989. Regression and Time Series Model Selection in Small Samples, Biometrika.76: 297-307.

Hurvich, C. M. and L. Tsai, 1993. A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection, J. of Time Series Analysis. 14:271-279.

Jenkins, G. M. and D. G. Watts, 1968. Spectral Analysis and Its Application, Holden-Day, San Francisco.

Kane, R. P. and N. B. Trivedi, 1991. Spectral Analysis Of Annual Sunspot Series - An Update, Pure and Applied Geophysics. 135: 463-474.

Koreisha, S. G. and T. Pukkila. 1993. Determining the Order of A Vector Autoregression When the Number of Component Series Is Large, J. of Time Series Analysis. 14: 47-69.

Martin, R. D., 1980. Robust Estimation of Autoregressive Models, in Direction in Time Series Eds Brillinger and Tiao, Institute of Mathematical Statistics.

Newton, H. J. and M. Pagano, 1984. Simultaneous Confidence Bands for Autoregressive Spectrum, Biometrika. 71: 197-202.

Paparoditis, E. and B. Streitberg, 1991. Order Identification Statistics in Stationary Autoregressive Moving-Average Models: Vector Autocorrelations and the Bootstrap, J. of Time Series Analysis. 13: 415-434.

Parzen, E., 1969. Multiple Time Series Modelling, in Multivariate Analysis II Edited By Krishnaiah, P. R., *Academic Press, New York.*

Parzen, E., 1974. Some Recent Advances in Time Series Modelling, IEEE Transactions on Automatic Control. AC-19: 723-730.

Priestley, M. B., 1981. Spectral Analysis and Time Series, Academic.Press, London.

Schwarz, G., 1978. Estimating the Dimension of a Model, Annals of Statistics. 6: 461-464.

Shibata, R., 1976. Selection of The Order of an Autoregressive Model by Akaike's Information Criterion, Biometrika. 63: 117-126.

Stoica, P., 1990. Performance Evaluation of Some Methods for Off-Line Detection of Changes in Autoregressive Signals, Signal Processing. 19, 4: 301-310.

Swanapoel, J. W. H. and J. W. J. van Wyk, 1986. The Bootstrap Applied to Power Spectral Density Function Estimation, Biometrika. 73: 135-141.

Whittle, P., 1963. Prediction and Regulation, English Universities Press, London.