Pakistan Journal of Applied Sciences 2(9): 912-915, 2002


# The Effect of Outliers on the Performance of Order Selection Criteria for Short Time Series

Zazli Chik
Centre for Foundation Studies in Science, University of Malaya,
50603 Kuala Lumpur, Malaysia

**Abstract:** The performance of various order selection criteria on short time series with the presence of outliers will be investigated. In order to reduce the influence of outliers, robust order selection criteria were introduced and their performance investigated. The results for order estimation for short time series with innovation outliers using least squares were identical to the results obtained for the case without outliers. Overall, the Bayesian modified of the AIC criterion, BIC gave the best results. The robust order selection criteria did not perform well.

**Key Words:** Innovation Outliers, Additive Outliers, Order Selection

## Introduction

Time series are sometimes influenced by outliers that may be the results of gross errors, certain system changes, special promotions or strikes. The two main types of outliers are Innovation Outliers (IO) and Additive Outliers (AO). The IO and AO are of interest since they are the most common types of outliers to be found in practice. The concept of IO and AO was introduced by Fox (1972). The AO corresponds to the situation in which a gross error of observation or recording error affects a single observation and the IO corresponds to the situation in which a single innovation is extreme. In the IO case, the extreme innovations will affect not only the particular observation but also subsequent observations. The IO and AO have also been discussed for example in Abraham and Chuang (1989, 1993), Chang, Tiao and Chen (1988), Chen and Liu (1993), Chuang and Abraham (1989), Tsay (1986), Martin (1979, 1980) and Denby and Martin (1979).

However, other changes to an underlying stationary stochastic process are possible. Outliers may not only occur in isolation but may also occur in a continous block. For example a block of unusually high and low values may occur in such a way that the sum of the observations within the block is the same as might have been expected for an undisturbed series. Wu, Hosking and Ravishanker (1993) called such a block a 'reallocation' and the individual observations within the block 'reallocation outliers'. Recent studies of changes to an underlying stationary stochastic process such as level shift and temporary change include Balke (1993) and Chen and Liu (1993).

Temporary change produces an initial effect at a particular time to the series and this effect dies out gradually with time while level shift produces an abrupt and permanent step change in the series.

The presence of outliers can seriously distort model specification, affect parameter estimation, innovation variances and impede forecasts. For example, it is recognised that outliers have a large influence on least squares estimators, pulling the least squares "fit" towards them too much. A resulting examination of the residuals is misleading because they look more like normal ones (Hogg, 1979). Accordingly, robust estimation methods have been created which modify least squares estimators so that the outliers have much less influence on the final estimates.

The impact of outliers on the estimates of time series model parameters and innovation variance has been extensively studied since Fox (1972) proposed the concepts of AO and IO in time series modeling. In general, there are two methods of reducing the influence of outliers. The 1st, follows a two-step operation of identifying the locations and types of outliers and then adjusting for the effects of the outliers for the purpose of model parameter estimation. The 2nd method of reducing the influence of outliers is to consider robust estimation methods such as those suggested by Martin (1979, 1980). The innovation variance estimate based on the sum-of-squared residuals will be effected by the presence of outliers. It is well known (for example, Martin, 1980) that the innovation variance estimate is non-robust toward heavy-tailed distributions, and can result in both serious bias and inflated variances. Most order selection criteria such as AIC, the Bayesian modification of the AIC criterion (BIC), Schwarz's criterion and Hannan and Quinn's criterion described in Chik (2002) depend on the innovation variance estimate. Thus, it is necessary to use robust methods that can select an appropriate order for the fitted autoregressive model in the presence of these outliers. In this paper the effect of outliers on the performance of order selection criteria, FPE, AIC, BIC, Hannan and Quinn's criterion, Schwarz's criterion and robust order selection will be investigated.

## Materials and Methods

The IO and AO can be modelled as follows.

**Model for Innovation Outliers:** Let $\{x_k, k=1, \ldots, n\}$ denote a set of data generated by an autoregressive process,

$$x_k = \sum_{i=1}^{p} \alpha_i x_{k-i} + \varepsilon_k \qquad (1)$$

where the innovations sequence $\{\varepsilon_k\}$ is independent and identically distributed with a symmetric distribution G which has mean zero. Let μ be a location parameter. Then the observations are

$$y_k \equiv \mu + x_k$$

and stationarity is assumed. In this model the autoregression is observed perfectly and innovation outliers occur when G is heavy-tailed compared to the Gaussian distribution.

As an example for this type of outlier, consider an AR (1) model,

$$x_k = \alpha x_{k-1} + \varepsilon_k$$

$$y_k = x_k$$

Where $|\alpha| < 1$ and the innovations $\varepsilon_k$ are independent and identically distributed. The IO occur when the density of $\varepsilon_k$ is a heavy-tailed non-Gaussian density. For example, if the density is a contaminated normal density

$$CN(\gamma, \sigma^2) = (1-\gamma)N(0,1) + \gamma N(0, \sigma^2)$$

With $\gamma > 0$ and $\sigma^2 > 1$, then $\varepsilon_k$ and hence $x_k$, are said to contain outliers (Denby and Martin, 1979).

**Model for Additive Outliers:** Let $\{x_k, k=1, \ldots, n\}$ denote a set of data generated by an autoregressive process, where the innovations sequence $\{\varepsilon_k\}$ is independent and identically distributed with a symmetric distribution G which has mean zero. Let μ be a location parameter. Then the observations are

$$y_k \equiv \mu + x_k + v_k$$

Where $x_k$ is as defined in (1) and $\{v_k, k=1, \ldots, n\}$ is a set of data generated by a stationary process and has a symmetric distribution with $Pr(v_k=0) = 1-\gamma$ where $\gamma$ is not too large. Both independence and correlation specifications for the stationary process need to be considered.

As an example for this type of outlier again consider an AR (1) model,

$$x_k = \alpha x_{k-1} + \varepsilon_k$$

$$y_k = x_k + v_k$$

Where the innovations $\varepsilon_k$ are zero-mean Gaussian with variance $\sigma_\varepsilon^2$. The $v_k$ are also assumed to be independently identically distributed with a Gaussian mixture density

$$CND(\gamma, \sigma^2) = (1-\gamma)\delta(.) + \gamma N(0, \sigma^2)$$

containing a Gaussian component and a degenerate central component $\delta(.)$ such that $Pr(v_k = 0) = 1-\gamma$ and $Pr(v_k \neq 0) = \gamma$ (Denby and Martin, 1979).

In the AO model, the observations consist of the Gaussian 'core' $x_k$ plus additive 'effects' $v_k$. The $v_k$ are called effects rather than errors since effects seems to be a more appropriate descriptor of the diverse sources for $v_k$ (such as keypunch errors and local movement of an economic time series due to a strike or a bargain month). It is assumed that $\gamma$ is small (typically ($\gamma \leq 0.2$)) because it appears that outliers in time series are present only a small fraction of the time.

**Robust Order Selection Criteria:** Depending on the model specification (or identification) tool employed, the existence of IO's and AO's can lead to misspecification of the model form (Chang, Tiao and Chen, 1988). This is because the innovation variance estimate based on the sum-of-squared residuals is affected by IO and AO and most order selection criteria (such as AIC, BIC, Schwarz's criterion and Hannan and Quinn's criterion) depend on the innovation variance estimate. Thus, such an estimate would not be very suitable for either IO or AO situations and robust alternatives to these order criteria are needed.

Using the idea of M-estimates, Martin (1980) modified Akaike's information criterion, AIC, for order selection in modeling AR processes. If only IO were of concern, a robust maximum likelihood type order selection criterion would be obtained by choosing $p$ to minimize

$$M(p) = \frac{1}{n-1} \sum_{k=2}^{n} \rho\left(\frac{y_k - z_k^T \hat{\alpha}_M}{\hat{s}_M}\right) + 2\log(\hat{s}_M) + \frac{2p}{n-1}$$

where $\rho$ is a suitable M-estimate loss function and $\hat{\alpha}_M$, $\hat{s}_M$ are M-estimates obtained from Martin (1980) by setting $W(z_i) \equiv 1$.

However, such M-estimates are not robust toward AO situations and so instead of minimizing $M(p)$ with respect to $p$, Martin proposed estimating $p$ by minimising the GM-estimate-based function

$$GM(p) = \frac{1}{n-1} \sum_{k=2}^{n} W(z_k) \rho\left(\frac{y_k - z_k^T \hat{\alpha}_{GM}}{\hat{s}_{GM}}\right) + 2\log(\hat{s}_{GM}) + \frac{2p}{n-1}$$

where $\hat{\alpha}_{GM}$ and $\hat{s}_{GM}$ are GM-estimates obtained from Martin (1980).

Table 1(a): Frequency Table of Orders Estimated for an AR (1) Series With IO Using Various Criteria (using least squares estimation) for Sample Size of 64

| | Criterion | FPE | AIC | BIC | Schwarz | HQ | M | GM |
|---|---|---|---|---|---|---|---|---|
| | Estimated order | | | | | | | |
| AR(1), $\alpha_1 = 0.2$ | 1 | 69 | 68 | 97 | 97 | 89 | 55 | 51 |
| | 2 | 9 | 9 | 3 | 3 | 5 | 6 | 8 |
| | 3 | 12 | 12 | 0 | 0 | 4 | 15 | 13 |
| | 4 – 11 | 10 | 11 | 0 | 0 | 2 | 24 | 28 |
| AR(1), $\alpha_1 = 0.4$ | 1 | 70 | 70 | 98 | 95 | 87 | 55 | 49 |
| | 2 | 8 | 8 | 2 | 4 | 5 | 7 | 8 |
| | 3 | 13 | 13 | 0 | 1 | 6 | 17 | 17 |
| | 4 – 11 | 9 | 9 | 0 | 0 | 2 | 21 | 26 |
| AR(1), $\alpha_1 = 0.6$ | 1 | 67 | 67 | 97 | 96 | 86 | 53 | 47 |
| | 2 | 14 | 14 | 2 | 3 | 8 | 11 | 12 |
| | 3 | 9 | 9 | 1 | 1 | 5 | 12 | 12 |
| | 4 – 11 | 10 | 10 | 0 | 0 | 1 | 24 | 29 |
| AR(1), $\alpha_1 = 0.8$ | 1 | 69 | 69 | 95 | 95 | 86 | 56 | 50 |
| | 2 | 16 | 16 | 5 | 5 | 11 | 12 | 13 |
| | 3 | 7 | 7 | 0 | 0 | 2 | 9 | 11 |
| | 4 – 11 | 8 | 8 | 0 | 0 | 1 | 23 | 26 |

Table 1 (b): Frequency Table of Orders Estimated for an AR (1) Series With AO Using Various Criteria (Using Least Squares Estimation) for Sample Size of 64.

| | Criterion | FPE | AIC | BIC | Schwarz | HQ | M | GM |
|---|---|---|---|---|---|---|---|---|
| | Estimated order | | | | | | | |
| AR(1), $\alpha_1 = 0.2$ | 1 | 74 | 74 | 98 | 96 | 84 | 69 | 62 |
| | 2 | 12 | 12 | 2 | 3 | 11 | 8 | 10 |
| | 3 | 4 | 4 | 0 | 1 | 3 | 7 | 5 |
| | 4 - 11 | 10 | 11 | 0 | 0 | 2 | 16 | 23 |
| AR(1), $\alpha_1 = 0.4$ | 1 | 72 | 72 | 95 | 89 | 82 | 60 | 58 |
| | 2 | 13 | 13 | 5 | 11 | 13 | 13 | 11 |
| | 3 | 5 | 5 | 0 | 0 | 3 | 7 | 7 |
| | 4 - 11 | 10 | 10 | 0 | 0 | 2 | 20 | 24 |
| AR(1), $\alpha_1 = 0.6$ | 1 | 62 | 62 | 91 | 88 | 80 | 51 | 46 |
| | 2 | 21 | 21 | 9 | 12 | 15 | 22 | 20 |
| | 3 | 4 | 3 | 0 | 0 | 2 | 6 | 7 |
| | 4 - 11 | 13 | 14 | 0 | 0 | 3 | 21 | 27 |
| AR(1), $\alpha_1 = 0.8$ | 1 | 36 | 36 | 74 | 72 | 54 | 31 | 28 |
| | 2 | 42 | 42 | 26 | 27 | 36 | 40 | 33 |
| | 3 | 9 | 9 | 0 | 0 | 4 | 10 | 9 |
| | 4 - 11 | 13 | 13 | 0 | 1 | 6 | 19 | 30 |

## Results and Discussion

The performance of the various order selection will be compared using realisations of AR (1) processes with IO and AO. The order criteria which will be compared are

1. FPE
2. AIC
3. BIC
4. Schwarz's criterion
5. Hannan and Quinn's criterion
6. Modified AIC using M-estimates, M
7. Modified AIC using GM-estimates, GM

The IO were obtained by generating the innovations, $\varepsilon_k$ from a contaminated normal density,

$$CN(0.25,9) = (1\text{-}0.25)N(0,1) + 0.25N(0,9)$$

and the AO were obtained by adding to the data, at random,

$v_k = 3.0N(0,1)$ with $Pr(v_k=0)=1-\gamma=0.9$

A sample size of N=64 was used. However an initial block of at least 60 values was generated and the final number from this initial block was used as the starting value for the sample. This procedure was taken to reduce the effect of initial values and the final value from the block of 60 is assumed to be a good approximation to a value from the desired stationary initial distribution (Denby and Martin, 1979). However, Tsay (1986) generated 400 values and used the last 100 values for a sample size of 100.

One hundred replicates of an AR (1) process with outliers for parameter values $\alpha_1$ = 0.2, 0.4, 0.6, 0.8 were generated and a sample size of N=64 was used. The least squares method were used to fit the autoregressive processes. The maximum order $m$ was set at eleven and for each case the frequency distribution was obtained. The simulation uses the subroutines from the NAG library including subroutines for the least squares method.

In Table 1(a), the effect of IO on order estimation is shown using least squares to estimate the parameters. The BIC gave the best results, the correct order was chosen 95% - 98% of the time and this is (almost) identical to the results when there are no outliers (see Chik (2002)). The Schwarz criterion gave similar results (95% - 97% correct) followed by Hannan and Quinn's criterion, FPE and AIC. The modified AIC (M and GM) did not perform well; the correct order was chosen 53% - 56% and 47% - 51% of the time respectively.

The effect of AO on order estimation was considered and the results are given in Tables 1b. Again least squares was used to estimate the parameters. The results shows a similar pattern as before, the BIC gave the best result, the correct order was choosen 74% - 98% of the time. However, for all the criteria, the percentage of choosing the correct order decreases gradually as $\alpha_1$ increases from 0.2 to 0.8. For $\alpha_1$ =0.8, the criteria FPE, AIC, M and GM all estimated the correct order less than 40% of the time.

The results for order estimation for IO using least squares to estimate the parameters were (almost) identical to the results obtained for the case without outliers. However, the results for order estimation for AO using least squares were weaker than the results obtained for the case without outliers. For both IO and AO, BIC gave the best results, Schwarz's criterion gave very similar results followed by Hannan and Quinn's criterion, FPE and AIC. Again for the autoregressive processes considered BIC gave the best results. The overall results using the modified AIC (M and GM) suggested by Martin (1980) were very poor.

## References

Abraham, B. and A. Chuang; 1993. Expectation-Maximization algorithms and the estimation of time series models in the presence of outliers, J. of Time Series Analysis. 14, 3: 221-234.

Abraham, B. and A. Chuang; 1989. Outlier detection and time series modeling, Technometrics. 31,2: 241-248.

Balke; 1993. Detecting level shifts in time series, J. of Business and Economic Statistics. 11: 81-92.

Chang, I., G. C. Tiao and C. Chen; 1988. Estimation of time series parameters in the presence of outliers, Technometrics. 30: 193-204.

Chen, C. and L. M. Liu; 1993. Forecasting time series with outliers, J. of Forecasting. 12: 13-35.

Chik, Z., 2002. Performance of order selection criteria for short time series, Pak. J. of App. Sci. 2(7): 783-788.

Chuang, A. and B. Abraham; 1989. Comparison of parameter estimation methods in time series with outliers: A simulation study, ASA proceedings business and Economic statistics section. 83-92.

Denby, L. and R. D. Martin; 1979. Robust estimation of the first-order autoregressive parameter, J. of the American Statistical Association. 74: 140-146.

Fox, A. J., 1972. Outliers in time series, J. of the Royal Statistical Society Series B. 43: 350-363.

Hogg, R. V., 1979. An introduction to robust estimation, in Robustness in Statistics Eds. Launer and Wilkinson, Academic Press, London.

Martin, R. D., 1979. Robust estimation for time series autoregressions, in Robustness in Statistics eds Launer and Wilkinson, Academic Press, London.

Martin, R. D., 1980. Robust estimation of autoregressive models, in Direction in Time Series Eds. Brillinger and Tiao, Institute of Mathematical Statistics.

Tsay, R. S., 1986. Time series model specification in the presence of outliers, J. of the American Statistical Association. 81: 132-141.

Wu, L. S-Y., J. R. M. Hoskingand and N. Ravishanker; 1993. Reallocation outliers in time series, App. Statistics. 42: 301-313.