Journal of the Korean
Statistical Society
Vol. 22, No. 2, 1993

Outlier Detection Diagnostic based on Interpolation Method in Autoregressive Models†

Sinsup Cho [1], Gui Yeol Ryu [2], Byeong Uk Park[1], and Jae June Lee [3]

ABSTRACT

An outlier detection diagnostic for the detection of k-consecutive atypical observations is considered. The proposed diagnostic is based on the innovational variance estimate utilizing both the interpolated and the predicted residuals. We adopt the interpolation method to construct the proposed diagnostic by replacing atypical observations. The performance of the proposed diagnostic is investigated by simulation. A real example is presented.

KEYWORDS: Outlier Detection, Autoregressive Model, Consecutive Additive Outlier, Interpolated Residual, Predicted Residual

1. INTRODUCTION

It is not unusual that time series have atypical observations which would affect the parameter estimates and model identification. In the presence of atypical observations, the conventional time series analysis procedure can easily lead to errorneous conclusions. Thus, it is an important problem to develop a procedure that can detect the possible outliers and remove their effects. In this paper, we propose an outlier detection procedure for k-consecutive atypical observations.

---

[1] Department of Computer Science and Statistics, Seoul National University, Seoul 151-742, Korea.

[2] Management Research Lab, Korea Telecom Research Center, Seoul 137-792, Korea.

[3] Department of Statistics, Inha University, Inchon 402-752, Korea

†This research was supported by S.N.U. Daewoo Research Fund 90-91.

Fox(1972) introduced two different types of outliers in time series. One is an Additive Outlier(AO) which may be induced by some external causes such as a gross recording error or measurement error. The other is an Innovational Outlier(IO) which may be caused by some internal changes or endogeneous effects. Other types of outliers have been introduced by Tsay(1988) and Martin and Yohai(1986) to describe more complicated effect of outliers in a parametric form, which include Patches of Outliers(PO), Level Change(LC), and Transient Change(TC). Bruce and Martin(1989) suggested the use of consecutive AO's asserting that outliers and other influential observations typically come in patches. Lee(1990) also advocated the use of consecutive AO's since PO and TC can be approximated by consecutive AO's. We will consider the consecutive AO's model in this paper.

Chang(1982) and Tsay(1986b, 1988) proposed methods based on the intervention analysis. They used the maximum of several tests to identify outliers, but it may have the danger of detecting false types of outliers and it turns out not suitable for consecutive AO's. Peña(1990) proposed an influential statistic $D_A(\cdot)$ which mearsures the effect of the influential observation on parameter estimates. But it is known that detection procedures based on parameter estimates often fail because of smearing effects (see Bruce(1989)).

The detection procedure proposed here is based on the innovational variance estimate, suggested by Bruce and Martin(1989) and Ledolter(1990), utilizing both the interpolated and the predicted residuals. For the construction of the diagnostic, we use interpolated series where atypical observations are replaced by the interpolators of Pourahmadi(1989). The innovational variance estimate is known to be more useful than parameter estimates in the presence of smearing effects. Furthermore, Ledolter(1989) and others observed that the interpolated residual is effective for a single AO while the predicted one for a single IO. Since a single IO may approximate a finite sequence of consecutive AO's with enough accuracy, our procedure should detect both a single and consecutive AO's effectively.

The paper is organized as follows. Section 2 proposes a detection diagnostic and discusses its properties. In Section 3, an iterative procedure for a detection of outliers is given. The performance of the proposed diagnostic is investigated by simulations in Section 4. The proposed procedure is applied to the United Kingdom Spirit data. Proofs are contained in the Appendix.

2. INTERPOLATION DIAGNOSTIC

In this section an outlier detection diagnostic based on the interpolated series is proposed and its properties are discussed. For the details of interpolation method, see Pourahmadi(1989), among others.

Let $x_t, t = 1, 2, \ldots, n$, be a stochastic process following AR(h), then

$$\pi(B)x_t = \varepsilon_t,$$

where B is the backshift operator such that $Bx_t = x_{t-1}$ and $\pi(B) = 1 - \pi_1 B - \cdots - \pi_h B^h$ is a polynomial in B with all zeros lying outside the unit circle. The ε_t's are independent normal random variables with mean 0 and variance σ_ε^2.

A k-consecutive AO's model is :

$$y_t = x_t + \sum_{j=0}^{k-1} \omega_j \xi_t^{T_0+j},$$

where y_t, $t = 1, 2, \ldots, n$, is the observed time series, T_0 the starting time of k-consecutive outliers, ω_j the magnitude of outlier effect in y_{T_0+j}, and $\xi_t^{T_0+j}$ the indicator function identifying the outlier occurrence time, which is given by

$$\xi_t^{T_0+j} = \begin{cases} 1, & \text{if } t = T_0 + j \\ 0, & \text{otherwise.} \end{cases}$$

From now on, $\Pi' = (\pi_1, \ldots, \pi_h)$ is assumed to be known unless we state otherwise. We consider a detection procedure based on the innovational variance computed from the interpolated series.

In the presence of outliers, if we use the observed series to estimate the parameters, a measure based on the parameter estimates will be influenced by the smearing effects. Therefore, it cannot detect the outliers effectively. On the other hand, if we replace the outliers by the interpolators, then a measure will not be affected by the smearing effects, and the outliers would be effectively detected.

An 1-Interpolation Diagnostic(DI_1) for the detection of a single outlier is

$$DI_1(T) = \sum_{t=h+1}^{n-h} (y_t^* - \hat{y}_t^*)^2,$$

where

$$y_t^* = \begin{cases} \tilde{y}_T(t), & \text{t=T} \\ y_t, & \text{otherwise.} \end{cases}$$

Here $\tilde{y}_T(t) = E(y_T | y_1, \ldots, y_{T-1}, y_{T+1}, \ldots, y_n)$ is an interpolated estimator of y_T, T is the time point of a possible outlier, and $\hat{y}_t^* = \Pi'(y_{t-1}^*, \ldots, y_{t-h}^*)'$ which is the one-step ahead prediction based on the interpolated series, where the outlier is replaced by the interpolated estimator. In the presence of k consecutive outliers beginning with T, if we let $Y_M' = (y_T, \ldots, y_{T+k-1})$, then the interpolated estimator following Pourahmadi(1989) is the orthogonal projection of Y_M onto the space spanned by $\{y_t, t < T \text{ and } t > T + k\}$. For more details, see Pourahmadi(1989). We may decompose $DI_1(T)$ as follows :

$$\begin{aligned} DI_1(T) &= \sum_{t=h+1}^{n-h} (y_t^* - y_t + y_t - \hat{y}_t^*)^2 \\ &= (y_T - y_T^*)^2 + \sum_{t=h+1}^{n-h} (y_t - \hat{y}_t^*)^2 - 2(y_T - y_t^*)(y_T - \hat{y}_t^*) \\ &= \tilde{\omega}_T^2 - \tilde{\omega}_T\hat{\omega}_T + \sum_{t=h+1}^{n-h} \hat{\omega}_t{}^2 - \tilde{\omega}_T\hat{\omega}_T, \end{aligned}$$

where $\tilde{\omega}_t = y_t - y_t^*$ is the interpolated residual and $\hat{\omega}_t = y_t - \hat{y}_t^*$ is the predicted residual based on the interpolated series. The properties of $\tilde{\omega}_T$ and $\hat{\omega}_T$ are summarized in Ryu (1991). Since $DI_1(T)$ consists of $\tilde{\omega}_T$ and $\hat{\omega}_T$'s, it is expected to detect both a single and consecutive AO's as we discussed in the introduction. If there exists a single AO at time T_0, since $DI_1(T_0)$ is computed with T_0^{th} observation being replaced by the interpolator, the effects of an outlier on $DI_1(T_0)$ are eliminated. But $DI_1(T)$'s for $T \neq T_0$ are contaminated by an outlier and are larger than $DI_1(T_0)$. The properties of $DI_1(T)$ are summarized in Theorem 2.1 and Theorem 2.2.

Theorem 2.1 If a single AO occurs at T_0 for AR(h), then $E(DI_1(T))$ is minimized at $T = T_0$.

From the proof of Theorem 2.1, we see that $E(DI_1(T) - DI_1(T_0)) = \sum_{i=0}^{h} \pi_i^2 \omega_{T_0}^2 + K$, for some $K(\geq 0)$ which is zero except for $T = T_0 - h, \cdots, T_0 + h$. Hence we can detect a single AO effectively if we choose a proper criterion.

Theorem 2.2 If k-consecutive AO's occur at $T = T_0, T_0 + 1, \ldots, T_0 + k - 1$ for AR(1), then the $E(DI_1(T))$ for $T = T_0 - 1, \ldots, T_0 + k$ are smaller than those for $T \leq T_0 - 2$ or $T \geq T_0 + k + 2$.

The proof of Theorem 2.2 is analogous to that of Thoerem 2.1 hence is omitted. Theorem 2.2 implies that when k-consecutive AO's exist, $DI_1(T)$ tends to have a patch of small values around the time points of outliers, hence it is not useful to detect k-consecutive AO's. To eliminate this effect of k-consecutive AO's we propose $DI_k(T)$, where the k-consecutve observations are replaced by interpolators, as follows :

$$DI_k(T) = \sum_{t=h+1}^{n-h} (y_t^* - \hat{y}_t^*)^2$$

where

$$y_t^* = \begin{cases} \tilde{y}_T(t) & t = T, T+1, \ldots, T+k-1 \\ y_t & t \neq T, T+1, \ldots, T+k-1, \end{cases}$$

$\hat{y}_T^* = \Pi'(y_{T-1}^*, \ldots, y_{T-h}^*)'$, and $T, \ldots, T+k-1$ are the time point of possible k-consecutive outliers occurrence and $\tilde{y}_T(t)$ are obtained following Pourahmadi(1989) as before. When there exist k-consecutive AO's, the properties of $DI_k(T)$ in AR(1) case is described in Theorem 2.3.

Theorem 2.3 If k-consecutive AO's occur at $T = T_0, T_0+1, \ldots, T_0+k-1$ for AR(1), then $E(DI_k(T))$ is minimized at $T = T_0$.

The above theorem tells us that if there exist k-consecutive AO's at $T = T_0, \ldots, T_0 + k - 1$, $DI_k(T_0)$ tends to be smaller than $DI_k(T)$, for $T \neq T_0$. Thus, if we choose a proper cutoff using the distribution of $DI_k(T)$ under H_0 (no outlier), then, in the case of k-consecutive AO's, $DI_k(T_0)$ tends to be smaller than the cutoff while $DI_k(T)$, $T \neq T_0$, would be larger than the cutoff. If there is no outlier, all the $DI_k(T)$ would be smaller than the cutoff. If we compute $DI_1(T)$ in the presence of k-consecutive AO's, $DI_1(T)$ would have a patch of smaller values at $T = T_0 - 1, \ldots, T_0 + k$. But they are not considerably smaller than the other $DI_1(T)$'s, since the effect of any single AO in k-consecutive AO case can be overwhelmed by the effects of the other outliers, which can fail to give clear evidence. Thus, if we have smaller values in a patched form, we have to increase the number of interpolated observations and recompute $DI_k(T)$ sequentially. The cutoff value can be obtained using the following Theorem.

Theorem 2.4 If there exists no outlier for AR(1), $DI_k(T)$ has the same distribution as the sum of the two independent scaled chi-squares,

$$\sigma_\varepsilon^2 \chi^2(n-2-(k+1)) + \frac{(\sum_{i=0}^{k} \pi_1^i)^2}{\sum_{i=0}^{k} \pi_1^{2i}} \sigma_\varepsilon^2 \chi^2(1).$$

For large n, the above distribution may be approximated as $\sigma_\varepsilon^2 \chi^2(n-2-k)$ and we use this as the reference distribution under the null hypothesis of no outlier.

So far we have considered only AR(1) case. We extend it to AR(2) case.

Theorem 2.5 If 2-consecutive AO's at T_0 and T_0+1 for AR(2), then $E(DI_2(T))$ is minimized at $T = T_0$.

From Theorem 2.5, we see that $E(DI_2(T_0))$ attains the minimum when there are 2-consecutive AO's for AR(2), because $E(DI_2(T_0))$ is free of the contamination of outliers while the others are not. For AR(h), when there are k-consecutive AO's with starting point T_0, we conjecture that $E(DI_k(T))$ will attain the minimum at

T_0, since $E(DI_k(T_0))$ is computed under k-consecutive outliers being all eliminated while others are computed under the effects of k-consecutive AO's being remained in squared forms.

We cannot derive the distribution of $DI_2(T_0)$ since coefficients of ε_t's are too complicated. But if we extend the inference about AR(1) to AR(2), we conjecture $DI_2(T)$ would follow approximately $\sigma_\varepsilon^2\chi^2(n-2\times 2-2)$. In AR(h), if we interpolate k-consecutive observations, we get $DI_k(T_0)$ which is the squared sum of the interpolated residuals. Extending the argument of Theorem 2.3 to AR(h), $\sigma_\varepsilon^2\chi^2(n-2h-k)$ may be adopted as the reference distribution.

3. ITERATIVE PROCEDURE FOR DETECTING OUTLIERS

After investigating the properties of $DI_k(T)$, we are led to propose an iterative procedure for the detection of outliers. Because we test whether there exist outliers or not, we consider cutoff values using the reference distribution. For cutoff values, 75%, 90% and 95% quantiles of the χ^2-distribution are commonly used by many authors. Based on our simulation experience, we suggest the use of 85% quantile as a cutoff. The procedure begins by setting $k=1$ as follows:

Step 1.

Compute $DI_k(T)$ for $T=h+1,\ldots,n-h$ and choose the time point $T=T_0$ where $DI_k(T)$ is minimized. Compute a cutoff using the quantile of $\hat{\sigma}_{T_0}^2\chi^2(\nu)$ where $\nu=n-2h-k$ is the degree of freedom and $\hat{\sigma}_{T_0}^2$ is the variance estimator with $T_0^{th},(T_0+1)^{th},\ldots,(T_0+k-1)^{th}$ observations being replaced by the interpolated values. If all $DI_k(T)$'s are smaller than the cutoff, determine there is no further outliers. Otherwise, go to Step 2.

Step 2.

If $DI_k(T)$'s are larger than the cutoff except $DI_k(T_0)$, then $T_0^{th},(T_0+1)^{th},\ldots,$ $(T_0+k-1)^{th}$ observations are determined as outliers. After replacing the outliers by the interpolated values go to Step 1 with $k=1$. If $DI_k(T)$'s appear in a patched form near the minimum point while other $DI_k(T)$'s are larger than the cutoff, increase k by 1 and go to Step 1.

In Step 2, the effect of outlier is estimated as the difference between the observed and interpolated values. For the properties of the estimates, see Ryu(1991). In practice, we seldom know the values of parameters and have to estimate the parameters.

For the estimation of parameters we use the Yule-Walker estimate since it is much easier to obtain and is as efficient as the maximum likelihood estimate in the Autoregressive models. Also it was shown by Dunsmuir and Robinson(1981) that the Yule-Walker type estimate has the asymptotic normality in the case of irregularly observed time series. We use the Yule-Walker estimate to obtain the interpolators and use the interpolated series to estimate parameters, so that the effect of missing observations on parameter estimates are reduced. In this case the degree of freedom of a reference distribution is reduced by as many as the number of estimated parameters.

4. NUMERICAL EXAMPLES

To see the performance of the proposed iterative procedures, time series are generated from various models and our procedure is applied to these data. Besides the simulated data, United Kingdom spirit data, which was used by Tsay(1986a) and Lee(1990), is also used to compare the performance of our procedure with other methods. For the simulation, we use the Box-Muller method to generate the normal random numbers with mean 0 and variance 1 using SUN3/280. In the following examples, $n = 200$ observations are generated from the specified models with the FORTRAN77 and the first 100 data points are discarded to eliminate the effects of the starting value. We impose the outlier effect at time point 30.

In a single AO case, we examine the empirical powers(percentage of correct detection) of our procedure by varying the value of π with $\sigma_\varepsilon^2 = 1$. Table 4.1 and 4.2 summarize the simulation results of 1000 replications for each combination of ω_T and π using 75% and 85% cutoff values, respectively. In general, the empirical powers of our procedure increases as the cutoff value decreases and the value of the parameter increases in the absolute value. For a large-size outlier, $\omega_T = 5.0\sigma_\varepsilon$, the percentage of correct detection ranges from 98.8% to 99.7% when 75% cutoff is used and from 90.5% to 99.4% when 85% cutoff is used. It is also observed that percentages are nearly symmetric about $\pi = 0$. For $\omega = 4.0\sigma_\varepsilon$ and $3.0\sigma_\varepsilon$,though the percentage of correct detection fluctuates depending on the value of π, it works fairly well. We recommend the use of 75% cutoff for high sensitivity and 85% cutoff for moderate sensitivity. In the following examples, we use 85% cutoff.

Table 4.1 Number of correct detection in 1000 replications using 75% quantile

ω_T / π	0.9	0.6	0.3	-0.3	-0.6	-0.9
$5.0\sigma_\varepsilon$	996	995	989	988	993	997
$4.0\sigma_\varepsilon$	975	936	905	890	941	967
$3.0\sigma_\varepsilon$	728	615	459	471	621	751

Table 4.2 Number of correct detection in 1000 replications using 85% quantile

ω_T / π	0.9	0.6	0.3	-0.3	-0.6	-0.9
$5.0\sigma_\varepsilon$	994	965	905	911	971	991
$4.0\sigma_\varepsilon$	892	783	692	702	771	874
$3.0\sigma_\varepsilon$	502	417	326	352	422	513

For the consecutive AO's case, the following examples explain the performance of our procedure.

Example 4.1 2-consecutive AO's for AR(2), $\phi_1 = 1.1$, $\phi_2 = -0.4$, $n = 100$ and $\omega = 5.0$

In this examlpe we generate data from AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.4$ and impose 2 consecutive AO's at T =15 and 16 with the magnitude of $5\sigma_\varepsilon$. The time series plot is given in Figure 4.1(a). The dot line represents pure series following AR(2) and the solid line represents outlier contaminated series. From the plot of $DI_1(T)$, Figure 4.1(b), we see patched small values instead of a single negative peak around $T = 14$, while others are larger than the cutoff. This is the phenomena which we explained in Theorem 2.2. Due to the smearing effects, if there exist k-consecutive AO's, $DI_1(T)$ is not guaranteed to achieve the minimum at the point where the outlier exists. But as we explained before, after observing the patched form around the minimum, we increase the number of interpolated observations by 1, i.e., k=2. Figure 4.1 (c) shows that $DI_2(T)$ has the minimum at $T = 15$, while the other $DI_2(T)$'s are larger than the cutoff. Since we don't observe any further patched form, we determine 15th and 16th observations as outliers. After eliminting the effects of outliers and employing $DI_1(T)$ again, Figure 4.1(d), we conclude that no more outlier exists, since all the $DI_1(T)$'s are smaller than the cutoff.

To compare the performance of our procedure with Chang's, same data set is analyzed by SCA [4] package which uses Chang's method for the detection of outliers. Chang's method does not identify 15th and 16th observations correctly. Instead, 15th observation is identified as an IO and 17th observation is as a TC in Table 4.4. In other words, Chang's method cannot detect consecutive AO's correctly, as we explained in Section 1.

[4] SCA : Scientific Computing Associates

Table 4.3 Summary Table of Example 4.1

	Step	1	2	3
2-5	True Value	Estimate	Estimate	Estimate
π_1	1.1	1.050	1.034	1.082
π_2	-0.4	-0.432	-0.400	-0.410
σ_ε^2	1.0	1.234	1.110	0.993
ω	5.0		3.879	
	5.0		4.545	

Table 4.4 The Results of Chang's Method applied to the data of Example 4.1

Time	Estimate	T–value	Type
15	4.833	4.71	IO
17	-3.661	-3.83	TC

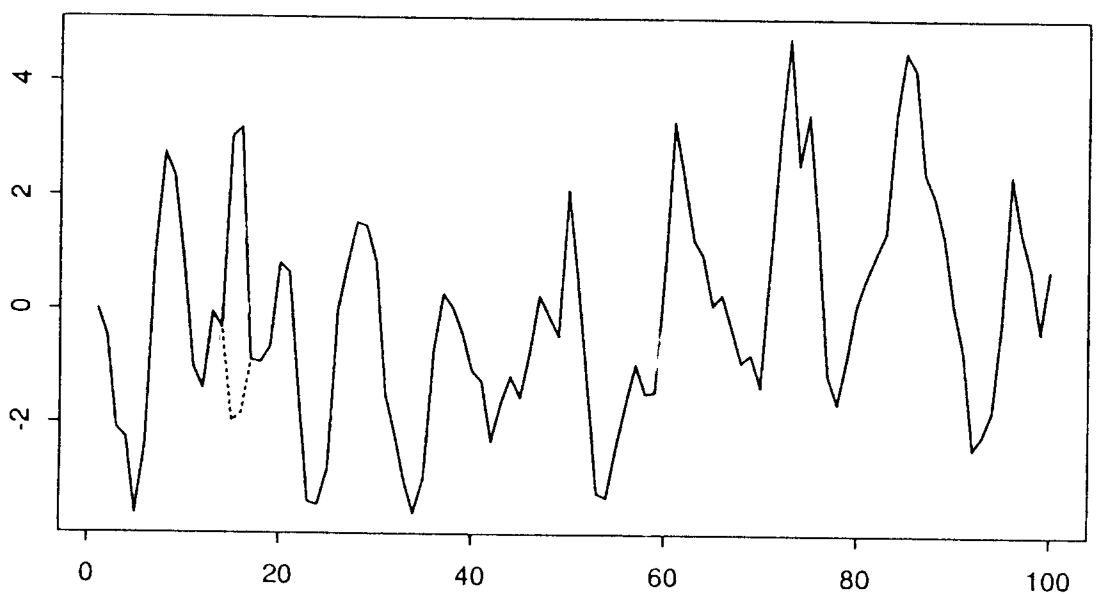

Figure 4.1 (a) Time series Plot for AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.4$ with 2-consecutive AO's at $T = 15, 16$

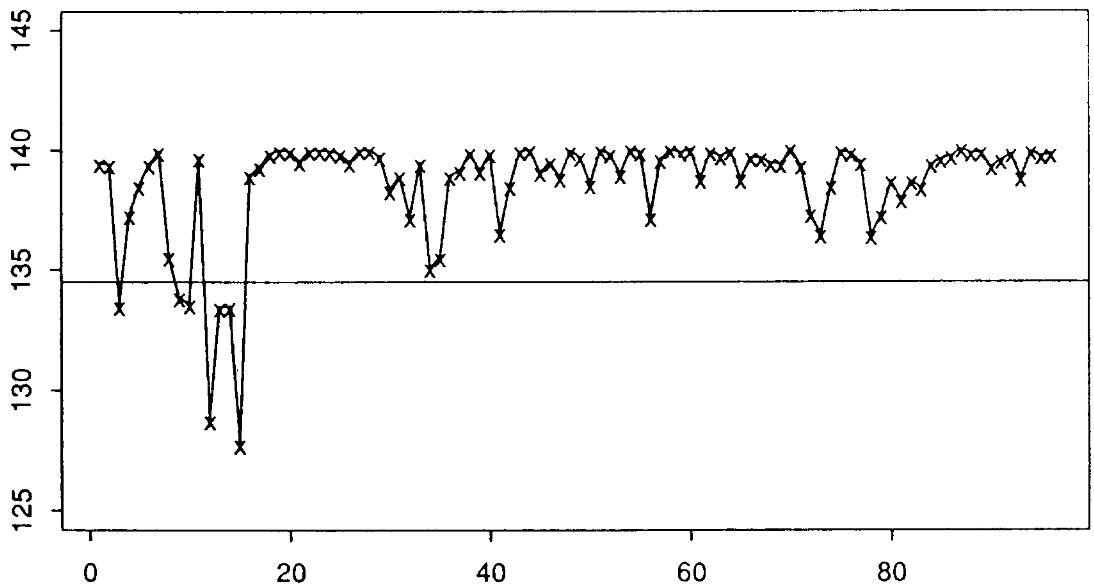

Figure 4.1 (b) Estimated Test Statistics of $DI_1(T)$

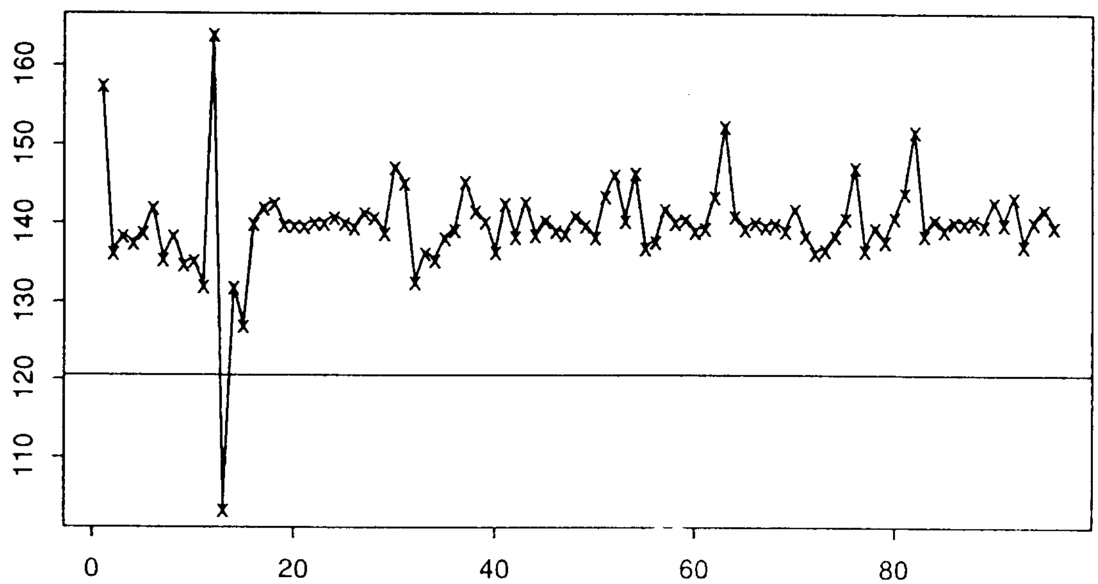

Figure 4.1 (c) Estimated Test Statistics of $DI_2(T)$

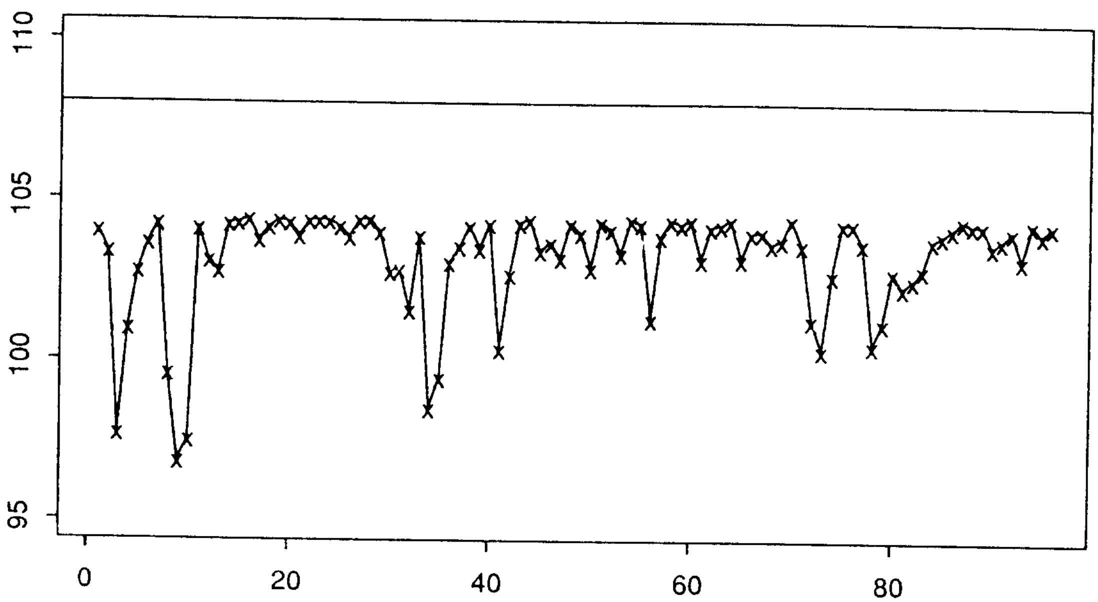

Figure 4.1 (d) Estimated Test Statistics of $DI_1(T)$ after eliminating outlier effect at $T = 15, 16$

Example 4.2 3-consecutive AO's for AR(1), $\phi_1 = -0.4$, $n = 100$ and $\omega = 5.0$

We impose 3-consecutive AO's at $T = 30$, 31 and 32 with the magnitude of $5\sigma_\varepsilon$. We observe patched small values around T =31, while other $DI_1(T)$'s are larger than the cutoff, Figure 4.2(a). After we increase k to 2 and employ $DI_2(T)$, patched pattern is again observed, Figure 4.2(b). After $DI_3(T)$ is employed, Figure 4.2(c), no more patched pattern appears and all $DI_3(T)$'s are larger than the cutoff except at $T = 30$, which indicates that 3-consecutive AO's exist at T =30, 31 and 32. After the effects of AO's are eliminated, the plot of $DI_1(T)$ shows no abnormal patterns and all $DI_1(T)$'s are smaller than the cutoff. Thus we conclude that 3-consecutive AO's exist at T =30, 31 and 32. The application result of our procedure and Chang's method are summarized in Table 4.5 and Table 4.6, respectively. Again, Chang's method does not detect 3 consecutive AO's effectively. Instead it identifies a TC at 30th observation together with two more AO's at T =29 and 34, which are not outliers in our example. From the above simulation results we see that our procedure can detect not only a single AO but also consecutive AO's effectively.

Table 4.5 Summary Table of Example 4.2

	Step	1	2	3	4
2-6	True Value	Estimate	Estimate	Estimate	Estimate
π_1	-0.4	-0.409	-0.409	-0.372	-0.381
σ_ε^2	1.0	1.472	1.214	1.060	0.996
ω	5.0			5.013	
	5.0			7.023	
	5.0			4.827	

Table 4.6 The Results of Chang's Method applied to the data of Example 4.2

Time	Estimate	T–value	Type
29	3.753	3.92	AO
30	4.974	8.65	TC
34	-3.913	-4.05	AO

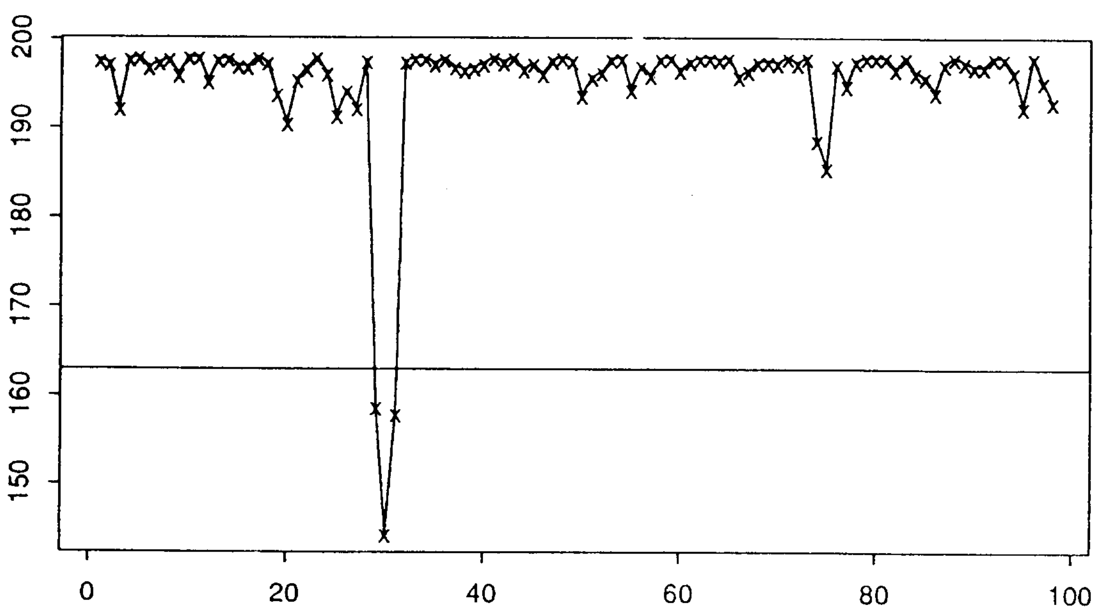

Figure 4.2 (a) Estimated Test Statistics for AR(1) when there is 3-consecutive AO's at $T = 30, 31$ and 32 $DI_1(T)$

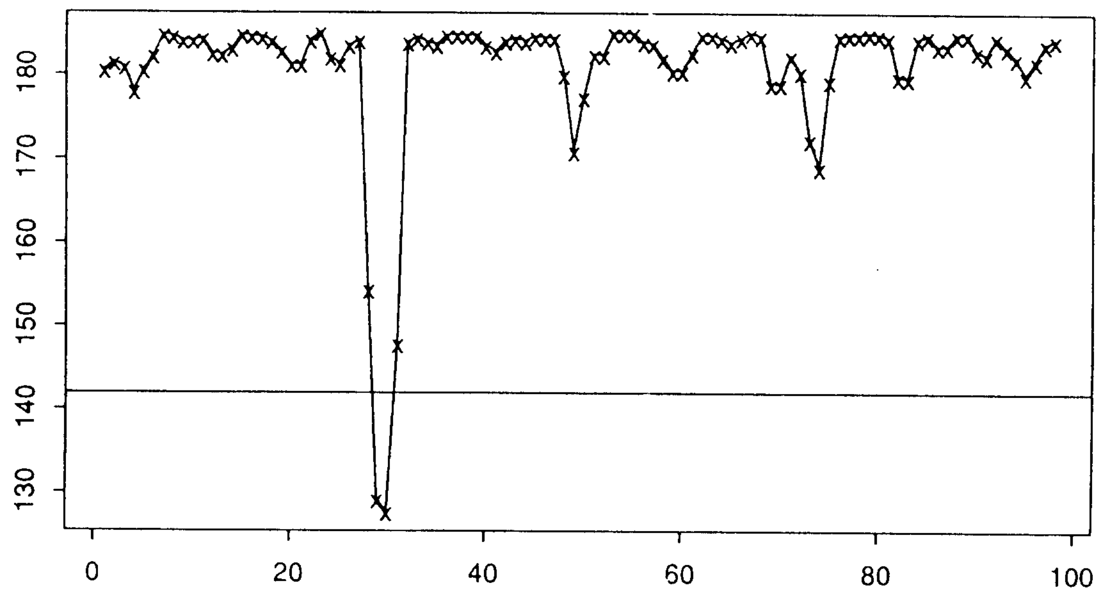

Figure 4.2 (b) Estimated Test Statistics for AR(1) when there is 3-consecutive AO's at $T = 30, 31$ and 32 $DI_2(T)$

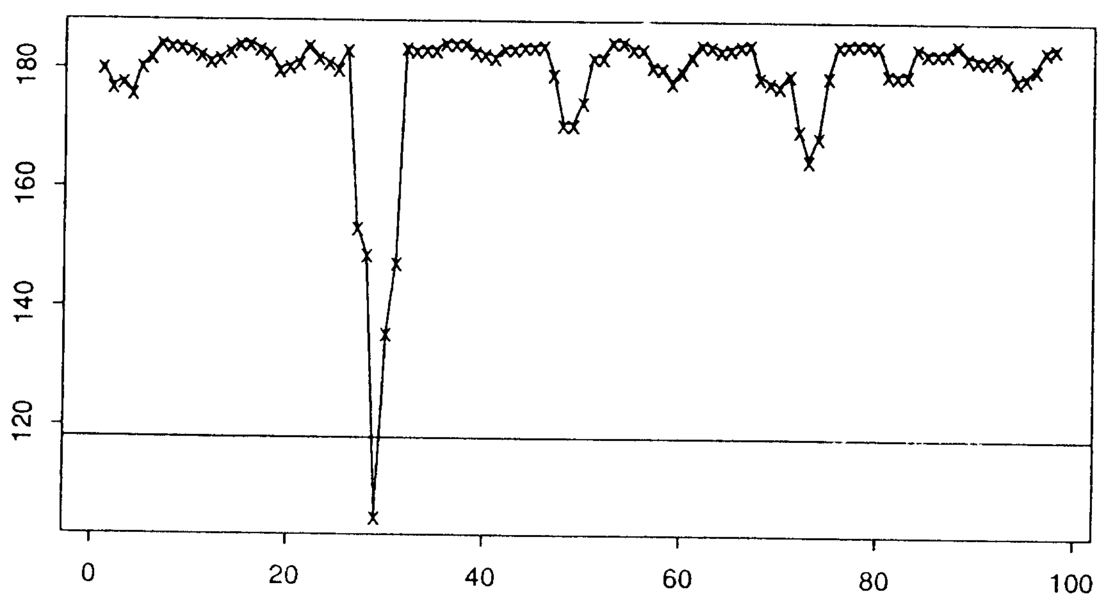

Figure 4.2 (c) $DI_3(T)$ after eliminating 3-consecutive AO effect

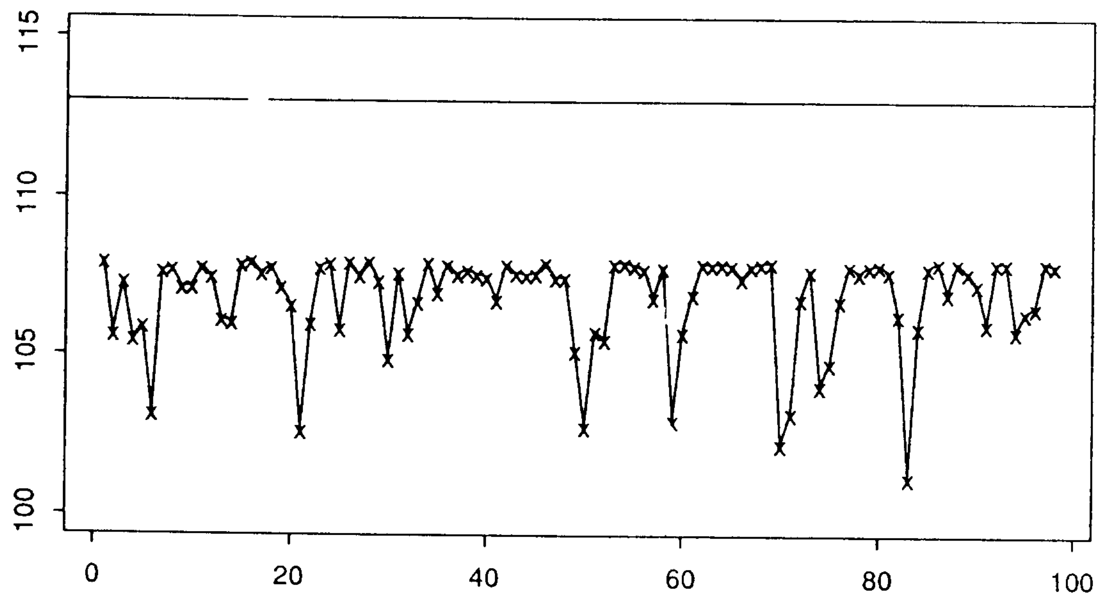

Figure 4.2 (d) $DI_1(T)$ after eliminating 3-consecutive AO effect

Up to this point we apply our detection procedure to the simulated data. In the next example we apply our procedure to the real data.

Example 4.3 United Kingdom Spirit Data

We analyze the data of annual consumption of spirits in the United Kingdom from 1870 to 1938, which was analyzed by Fuller (1976) and Tsay (1986b). They treated the residual series after fitting the time series regression model as an observed series Z_t. A time series plot of Z_t is given in Figure 4.3(a). By the iterative procedure of Tsay (1986b), three outliers at $T = 40$ (IO), 46 (AO) and 49 (AO) were detected. And Lee (1990) detected 40, 41, 46 and 49^{th} observations as outliers. We use AR(2) as an appropriate model to this series.

We employ $DI_1(T)$ and observe a sharp negative peak at $T = 49$. Since we do not see any patched pattern around $T = 49$ in Fig 4.3(b) we identify 49th observation as an outlier and eliminate the effect, $\tilde{\omega}_{49} = -0.0737$. When $DI_1(T)$ is again employed to the data with the effect of 49th observation eliminated, we obtain the same pattern at $T = 46$. This observation is again identified as an outlier and the effect is eliminated by $\tilde{\omega}_{46} = 0.0363$. When we employ $DI_1(T)$ to the adjusted data as given in Fig 4.3(d), we observe a patched pattern around $T = 40$ and we increase k from 1 to 2, i.e., $DI_2(T)$ is employed. From Fig 4.3(e), we determine 2 consecutive

AO's at $T = 40$ and 41. After the effects of outliers are eliminated at T = 40 and 41 with $\tilde{\omega}_{40} = 0.0997$ and $\tilde{\omega}_{41} = -0.0618$, we do not see any further abnormal pattern in Fig 4.3(f). From the above example we conclude that our iterative procedure performs relatively well in the presence of consecutive AO's as well as a single AO. As we see in Example 4.3, when there exist consecutive AO's, we detect all the AO's eventually.

The result applied to SCA is summarized in Table 4.7. Similar to Tsay's result, it identifies 3 outliers at T =40, 46 and 49.

Table 4.7 The Results of Chang's Method applied to U. K. Spirit data

Time	Estimate	T–value	Type
40	-0.087	-8.56	TC
46	0.046	6.54	AO
49	-0.061	-8.55	AO

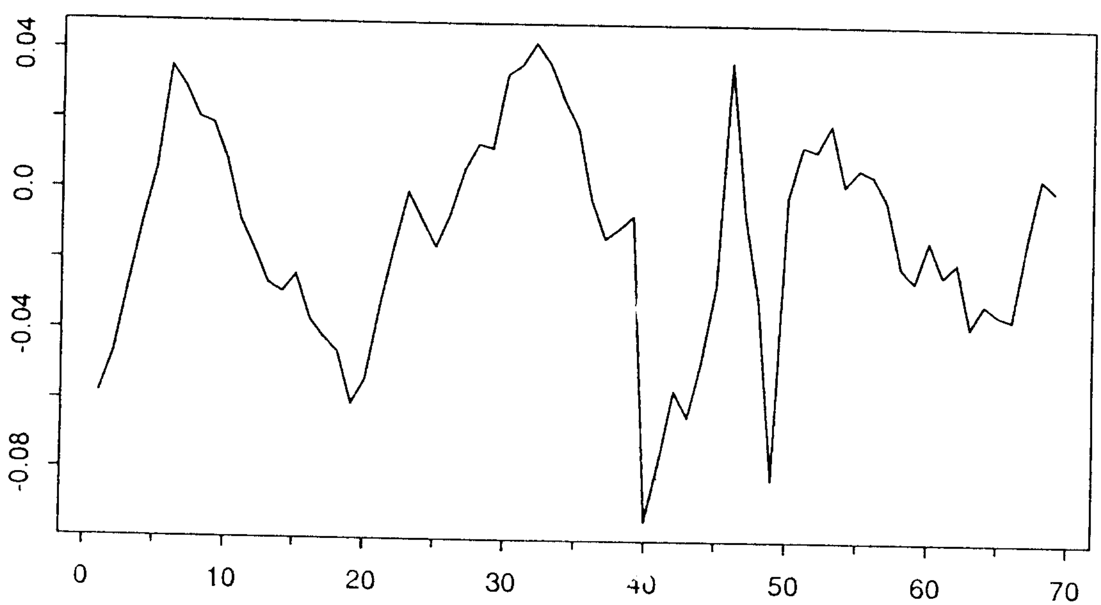

Figure 4.3 (a) Time Series Plot of U.K. Spirit Data

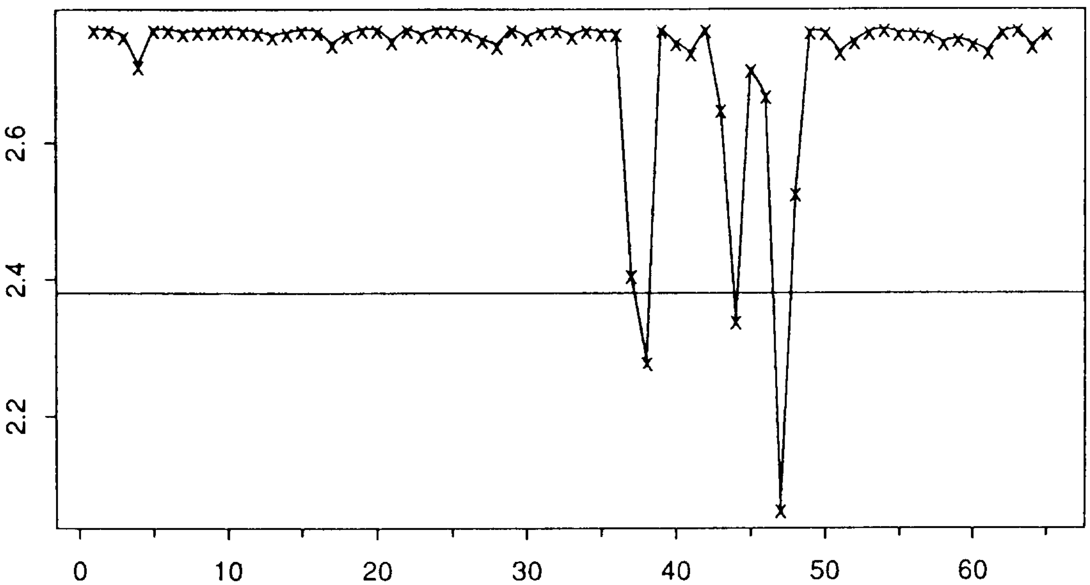

Figure 4.3 (b) Estimate Test Statistics of $DI_1(T)$

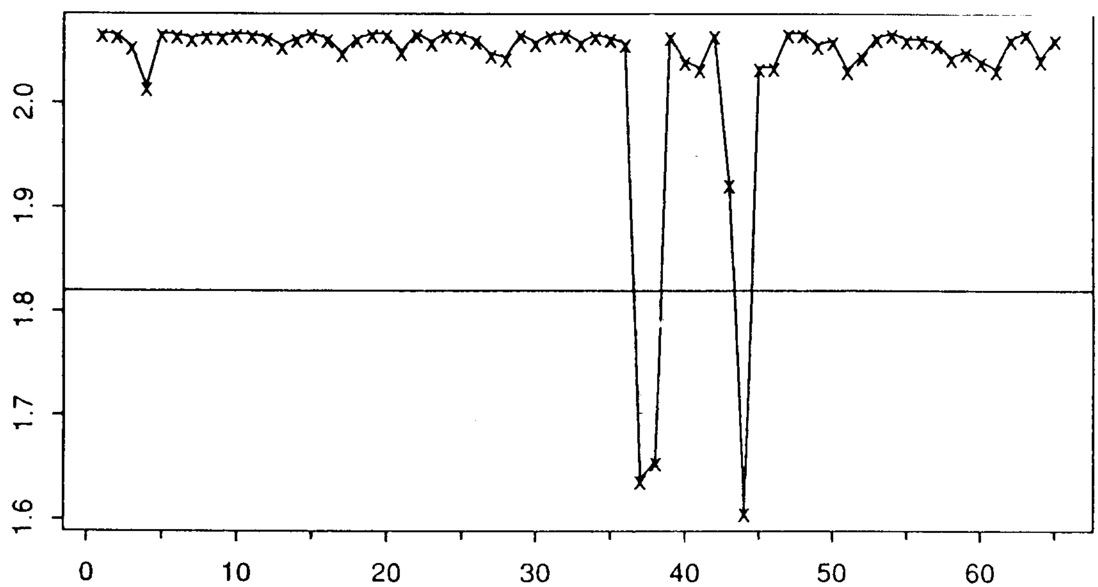

Figure 4.3 (c) Estimated Test Statistics of $DI_1(T)$ after eliminating outlier effect at 49

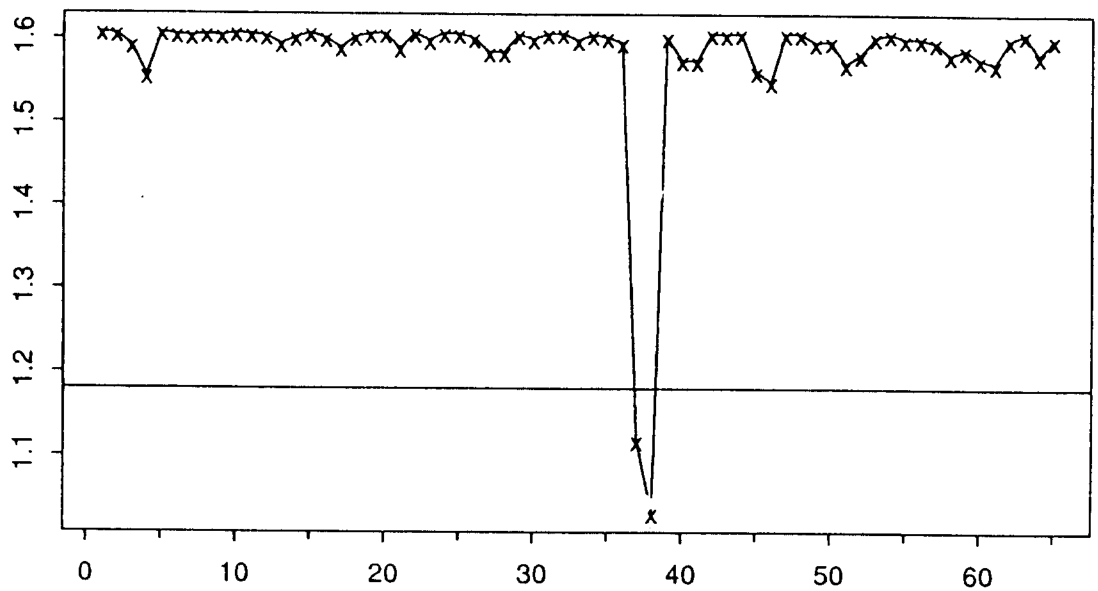

Figure 4.3 (d) Estimated Test Statistics of $DI_1(T)$ after eliminating outliers effect at 46 and 49

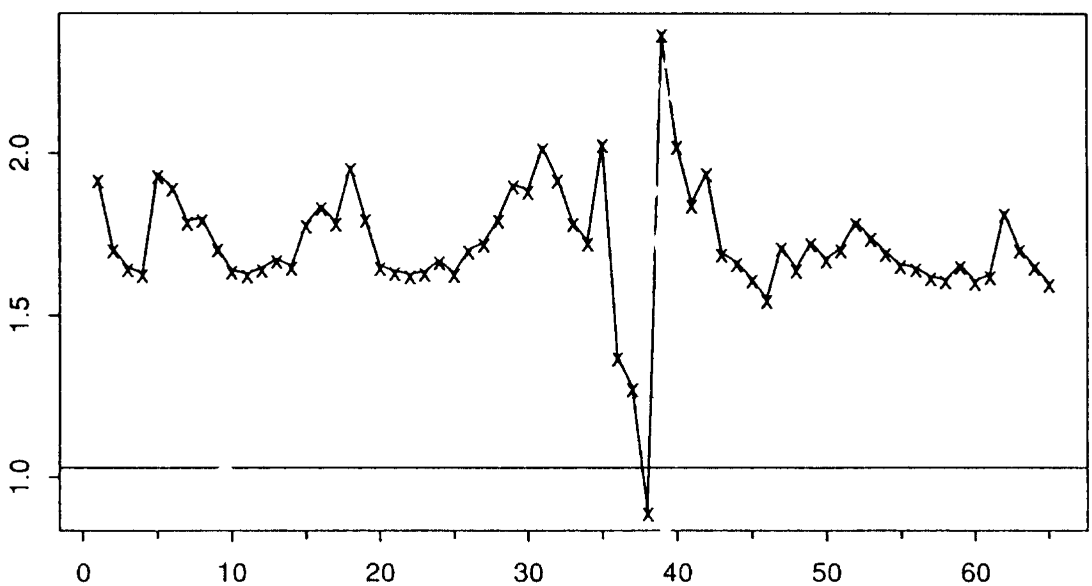

Figure 4.3 (e) Estimated Test Statistics of $DI_2(T)$ after eliminating outliers effect at 46 and 49

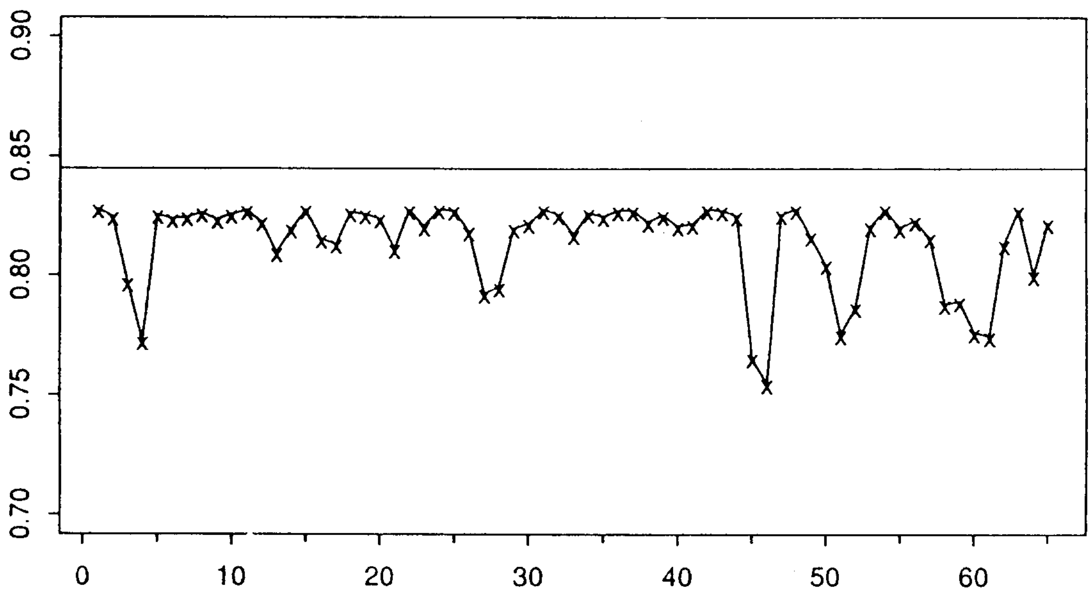

Figure 4.3 (f) Estimated Test Statistics of $DI_1(T)$ after eliminating all outlier effect

APPENDIX

Proof of theorem 2.1

For $T \leq T_0 - h - 1$ or $T \geq T_0 + h + 1$

$$\begin{aligned} DI_1(T) = & \sum_{t=h+1}^{T-1} \varepsilon_t^2 + (y_T^* - \sum_{i=1}^{h} \pi_i y_{T-i})^2 + \{y_{T+1} - (\pi_1 y_T^* + \sum_{i=2}^{h} \pi_i y_{T+1-i})\}^2 \\ & + \cdots + \{y_{T+h} - (\sum_{i=1}^{h-1} \pi_i y_{T+h-i} + \pi_h y_T^*)\}^2 + \sum_{t=h+1}^{T_0-1} \varepsilon_t^2 \\ & + (y_{T_0} - \sum_{i=1}^{h} \pi_i y_{T_0-i})^2 + (y_{T_0+1} - \sum_{i=1}^{h} \pi_i y_{T_0+1-i})^2 + \cdots \\ & + (y_{T_0+h} - \sum_{i=1}^{h} \pi_i y_{T_0+h-i})^2 + \sum_{t=T_0+h+1}^{n-h} \varepsilon_t^2, \end{aligned}$$

with expectation

$$E(DI_1(T)) = [n - 2h + K_1^{-1}\{-1 + \sum_{i=1}^{h}(\pi_i^2 + 2\pi_i)\}]\sigma_\varepsilon^2 + \sum_{i=0}^{h} \pi_i^2 \omega_{T_0}^2 \ , \quad \text{(A.1)}$$

where $K_1 = \sum_{i=0}^{h} \pi_i^2$.

For $T_0 - h \le T \le T_0 - 1$ and $j = T_0 - T$,

$$E(DI_1(T)) = [n - 2h + K_1^{-1}\{-1 + \sum_{i=1}^{h}(\pi_i^2 + 2\pi_i)\}]\sigma_\varepsilon^2 + \{\sum_{i=0}^{j-1} \pi_i^2 \delta_j^2 + \sum_{i=j}^{h-j}(\pi_{i-j} - \pi_{j+i}\delta_j)^2 + \sum_{i=h-j+1}^{h} \pi_i^2\}\omega_{T_0}^2, \quad \text{(A.2)}$$

where

$$\delta_j = K_1^{-1}(\pi_j - \sum_{i=1}^{h} \pi_i \pi_{i+j}).$$

For $T = T_0$,

$$E(DI_1(T_0)) = [n - 2h + K_1^{-1}\{-1 + \sum_{i=1}^{h}(\pi_i^2 + 2\pi_i)\}]\sigma_\varepsilon^2 \ . \quad \text{(A.3)}$$

For $T = T_0 + 1, \ldots, T_0 + h$,

$$E(DI_1(T)) = [n - 2h + K_1^{-1}\{-1 + \sum_{i=1}^{h}(\pi_i^2 + 2\pi_i)\}]\sigma_\varepsilon^2 + \{\sum_{i=0}^{T-T_0} \pi_i^2 + (\pi_{T-T_0} + \delta_{T-T_0})^2 + (\pi_1\delta_{T-T_0} + \pi_{T-T_0+1})^2 + \cdots + (\pi_{h+T_0-T}\delta_{T-T_0} + \pi_h)^2 + \sum_{i=h+T_0+1-T}^{h} \pi_i^2 \delta_{T-T_0}^2\}\omega_{T_0}^2. \quad \text{(A.4)}$$

Thus from (A.1), (A.2), (A.3) and (A.4), we can see that $E(DI_1(T))$ attains the minimum at T_0. ∎

Proof of theorem 2.3

For $T \le T_0 - k - 2$ or $T \ge T_0 + k + 1$

$$\begin{aligned} DI_k(T) = & \sum_{\substack{t \neq T, T+1, \ldots, T+k, \\ T_0, T_0+1, \ldots, T_0+k}} \varepsilon_t^2 + (y_T^* - \pi_1 y_{T-1})^2 + (y_{T+1}^* - \pi_1 y_T^*)^2 + \cdots \\ & + (y_{T+k-1}^* - \pi_1 y_{T+k-2}^*)^2 + (y_{T+k} - \pi_1 y_{T+k-1}^*)^2 + (y_{T_0} - \pi_1 y_{T_0-1})^2 \\ & + (y_{T_0+1} - \pi_1 y_{T_0})^2 + \cdots + (y_{T_0+k} - \pi_1 y_{T_0+k-1})^2, \end{aligned}$$

where

$$\begin{pmatrix} y_T^* \\ y_{T+1}^* \\ \vdots \\ y_{T+k-1}^* \end{pmatrix} = \begin{pmatrix} \pi_1 y_{T-1} + \pi_1^k C \\ \pi_1^2 y_{T-1} + \pi_1^{k-1}(1+\pi_1^2) C \\ \vdots \\ \pi_1^k y_{T-1} + \pi_1 \sum_{i=0}^{k-1} \pi_1^{2i} C \end{pmatrix},$$

$$C = K_2^{-1}(y_{T+k} - \pi_1^{k+1} y_{T-1})$$

and $K_2 = \sum_{i=0}^{k} \pi_1^{2i}$ with

$$\begin{aligned} E(DI_k(T)) = & \{n - 2 - (k+1) + K_2^{-1} \sum_{i=0}^{k-1} \pi_1^{2i}\} \sigma_\varepsilon^2 \\ & + \{1 + (k-1)(1-\pi_1)^2 + \pi_1^2\} \omega_{T_0}^2, \end{aligned}$$

To show that $E(DI_k(T))$ attains the minimum at T_0, it would be sufficient to show $E(DI_k(T_0 - 1)) > E(DI_k(T_0))$ and $E(DI_k(T_0 + 1)) > E(DI_k(T_0))$, because for $T = T_0 \pm j$, $j = 1, \ldots, k-1$, $DI_k(T)$ has $(j+1)$ extra squared terms which are contaminated by outliers.

For $T = T_0 - 1$,

$$E(DI_k(T_0 - 1)) = \{n - 2 - (k+1) + K_2^{-1} \sum_{i=0}^{k-1} \pi_1^{2i}\} \sigma_\varepsilon^2 + (\pi_1^2 + K_2^{-1}) \omega_{T_0}^2 .$$

For $T = T_0$,

$$E(DI_k(T_0)) = \{n - 2 - (k+1)\} \sigma_\varepsilon^2 + K_2^{-1} \sum_{i=0}^{k-1} \pi_1^{2i} \sigma_\varepsilon^2 ,$$

which is independent of ω_{T_0} and is the same one in the case of no outlier.

For $T = T_0 + 1$,

$$E(DI_k(T_0+1)) = \{n-2-(k+1)\} \sigma_\varepsilon^2 + K_2^{-1} \sum_{i=0}^{k-1} \pi_1^{2i} \sigma_\varepsilon^2 + (1 + K_2^{-1} \pi_1^{k+1}) \omega_{T_0}^2 .$$

Proof of Theorem 2.4

$$DI_k(T) = \sum_{t=2}^{T-1} \varepsilon_t^2 + (y_T^* - \pi_1 y_{T-1})^2 + (y_{T+1}^* - \pi_1 y_T^*)^2 + \cdots$$
$$+(y_{T+k-1}^* - \pi_1 y_{T+k-2}^*)^2 + (y_{T+k} - \pi_1 y_{T+k-1}^*)^2 + \sum_{t=T+k+1}^{n-1} \varepsilon_t^2.$$

where $(y_T^*, \ldots, y_{T+k-1}^*)'$ is the same as in the proof of Theorem 2.3, when we compute $DI(T_0)$. Therefore

$$DI_k(T) = \sum_{t \neq T, T+1, \ldots, T+k} \varepsilon_t^2 + K_3^{-1} (\sum_{i=0}^{k} \pi_1^i)^2 (\sum_{i=0}^{k} \pi_1^i \varepsilon_{T+k-i})^2,$$

where $K_3 = (\sum_{i=0}^{k} \pi_1^{2i})^2$. Since

$$\sum_{t \neq T, T+1, \ldots, T+k} \varepsilon_t^2 \quad \text{is distributed as} \quad \sigma_\varepsilon^2 \chi^2(n-2-(k+1)),$$

and

$$K_3^{-1} (\sum_{i=0}^{k} \pi_1^i)^2 (\sum_{i=0}^{k} \pi_1^i \varepsilon_{T+k-i})^2 \quad \text{is as} \quad \frac{(\sum_{i=0}^{k} \pi_1^i)^2}{\sum_{i=0}^{k} \pi_1^{2i}} \sigma_\varepsilon^2 \chi^2(1)$$

and $\sum_{t \neq T, T+1, \ldots, T+k} \varepsilon_t^2$ and $K_3^{-1} (\sum_{i=0}^{k} \pi_1^i)^2 (\sum_{i=0}^{k} \pi_1^i \varepsilon_{T+k-i})^2$ are independent. ∎

Proof of theorem 2.5

For $T \leq T_0 - 3$ or $T \geq T_0 + 4$

$$DI_2(T) = \sum_{\substack{t \neq T, T+1, T+2, T+3, \\ T_0, T_0+1, T_0+2, T_0+3}} \varepsilon_t^2 + (y_T^* - \pi_1 y_{T-1} - \pi_2 y_{T-2})^2 + (y_{T+1}^* - \pi_1 y_T^* - \pi_2 y_{T-1})^2$$
$$+(y_{T+2} - \pi_1 y_{T+1}^* - \pi_2 y_T^*)^2 + (y_{T+3} - \pi_1 y_{T+2} - \pi_2 y_{T+1}^*)^2$$
$$+(y_{T_0} - \pi_1 y_{T_0-1} - \pi_2 y_{T_0-2})^2 + (y_{T_0+1} - \pi_1 y_{T_0} - \pi_2 y_{T_0-1})^2$$
$$+(y_{T_0+2} - \pi_1 y_{T_0+1} - \pi_2 y_{T_0})^2 + (y_{T_0+3} - \pi_1 y_{T_0+2} - \pi_2 y_{T_0+1})^2$$

where

$$y_T^* = \pi_1 y_{T-1} + \pi_2 y_{T-2} + c_{11}(y_{T+2} - \hat{y}_{T+2}) + c_{21}(y_{T+3} - \hat{y}_{T+3})$$
$$y_{T+1}^* = (\pi_1^2 + \pi_2) y_{T-1} + \pi_1 \pi_2 y_{T-2} + c_{12}(y_{T+2} - \hat{y}_{T+2}) + c_{22}(y_{T+3} - \hat{y}_{T+3})$$
$$C = D(I_2 + A'A)^{-1} T_k$$

and

$$T_r = \begin{pmatrix} 1 & 0 & \dots & 0 \\ b_1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{r-1} & b_{r-2} & \dots & 1 \end{pmatrix}$$

$$B_{r,k} = (b_{k+i-j}) \text{ for } i = 0, 1, \dots, r-1 \text{ and } j = 0, 1, \dots, k-1$$
$$T_r' A = B_{r,k}$$
$$T_r D = A, r = n - T_0 - 1$$

and $b_i's$ are MA parameters of y_t

$$\begin{aligned} y_T^* - \pi_1 y_{T-1} - \pi_2 y_{T-2} &= c_{11}(\pi_1 \varepsilon_{T+1} + (\pi_1^2 + \pi_2)\varepsilon_T + \varepsilon_{T+2}) \\ &\quad + c_{21}(\varepsilon_{T+3} + \pi_1 \varepsilon_{T+2} + (\pi_1^2 + \pi_2)\varepsilon_{T+1} + \pi_1^3 \varepsilon_T) \\ &= f_1(\underline{\varepsilon}) \\ y_{T+1}^* - \pi_1 y_T^* - \pi_2 y_{T-1} &= c_{12}(\pi_1 \varepsilon_{T+1} + (\pi_1^2 + \pi_2)\varepsilon_T + \varepsilon_{T+2}) \\ &\quad + c_{22}(\varepsilon_{T+3} + \pi_1 \varepsilon_{T+2} + (\pi_1^2 + \pi_2)\varepsilon_{T+1} + \pi_1^3 \varepsilon_T) \\ &= f_2(\underline{\varepsilon}) \\ y_{T+2} - \pi_1 y_{T+1}^* - \pi_2 y_T^* &= \varepsilon_{T+2} + \pi_1(\pi_1 \varepsilon_{T+1} + (\pi_1^2 + \pi_2)\varepsilon_T + \varepsilon_{T+2}) \\ &\quad + \pi_2(\varepsilon_{T+3} + \pi_1 \varepsilon_{T+2} + (\pi_1^2 + \pi_2)\varepsilon_{T+1} + \pi_1^2 \varepsilon_T) \\ &= f_3(\underline{\varepsilon}) \\ y_{T+3} - \pi_1 y_{T+2} - \pi_2 y_{T+1}^* &= \varepsilon_{T+3} + \pi_2(\pi_1 \varepsilon_{T+1} + (\pi_1^2 + \pi_2)\varepsilon_T + \varepsilon_{T+2}) \\ &= f_4(\underline{\varepsilon}). \end{aligned}$$

After heavy algebra, it can be shown that for $T \leq T_0 - 3$ or $T \geq T_0 + 4$

$$\begin{aligned} E(DI_2(T)) &= (n-8)\sigma_\varepsilon^2 + E(f_1^2(\underline{\varepsilon}) + f_2^2(\underline{\varepsilon}) + f_3^2(\underline{\varepsilon}) + f_4^2(\underline{\varepsilon})) \\ &\quad + \{1 + (1-\pi_1)^2 + (\pi_1 + \pi_2)^2 + \pi_1^2\}\omega_{T_0}^2. \end{aligned}$$

For $T = T_0 - 1$,

$$\begin{aligned} E(DI_2(T_0 - 1)) &= (n-8)\sigma_\varepsilon^2 + E(f_1^2(\underline{\varepsilon}) + f_2^2(\underline{\varepsilon}) + f_3^2(\underline{\varepsilon}) + f_4^2(\underline{\varepsilon})) \\ &\quad + \{c_{11}^2 + (1 - \pi_1 c_{12} - \pi_2 c_{11})^2 + (\pi_1 + \pi_2 c_{12})^2 + \pi_1^2\}\omega_{T_0}^2. \end{aligned}$$

For $T = T_0$,

$$E(DI_2(T_0)) = (n-8)\sigma_\varepsilon^2 + E(f_1^2(\underline{\varepsilon}) + f_2^2(\underline{\varepsilon}) + f_3^2(\underline{\varepsilon}) + f_4^2(\underline{\varepsilon})).$$

For $T = T_0 + 1$,

$$\begin{aligned} E(DI_2(T_0 + 1)) &= (n-8)\sigma_\varepsilon^2 + E(f_1^2(\underline{\varepsilon}) + f_2^2(\underline{\varepsilon}) + f_3^2(\underline{\varepsilon}) + f_4^2(\underline{\varepsilon})) \\ &\quad + [1 + \{\pi_1 - c_{11}(\pi_1^2 + \pi_2) - c_{21}(\pi_1^3 + 2\pi_1\pi_2)\}^2 \\ &\quad + \{(\pi_1 c_{11} - c_{12})(\pi_1^2 + \pi_2) + (\pi_1 c_{21} - c_{22})(\pi_1^3 + 2\pi_1\pi_2)\}^2 \\ &\quad + \{\pi_1^3 + 2\pi_1\pi_2 - (\pi_1 c_{12} + \pi_2 c_{11})(\pi_1^2 + \pi_2) \\ &\quad - (\pi_1 c_{22} + \pi_2 c_{21})(\pi_1^3 + 2\pi_1\pi_2)\}^2 \end{aligned}$$

$$+\pi_1^2\{\pi_1^2\pi_2 - c_{12}(\pi_1^2 + \pi_2) - c_{22}(\pi_1^3 + 2\pi_1\pi_2)\}^2]\omega_{T_0}^2 .$$

Thus $E(DI_2(T_0+1) - DI_2(T_0))$ and $E(DI_2(T_0-1) - DI_2(T_0))$ are positive. Because of the same reason as Theorem 2.4, it would be sufficient to compare $E(DI_2(T_0))$ with $E(DI_2(T_0 - 1))$ and $E(DI_2(T_0 + 1))$ and we can determine $E(DI_2(T_0))$ is the minimum of $E(DI_2(T))$. ∎

REFERENCES

(1) Bruce, A.G. (1989). *Diagnostics for Time Series Models* . Unpublished Ph.D. dissertation, University of Washington, Department of Statistics.

(2) Bruce, A.G. and Martin, R. D. (1989). Leave-k-out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society* , Ser. B, 51, 375–424.

(3) Chang, I. (1982). *Outliers in time series.* Unpublished Ph.D. dissertation, University of Wisconsin-Madison, Department of Statistics.

(4) Chang, I., Tiao, G. C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30, 193–204.

(5) Dunsmuir, W. and Robinson, P. M. (1981). Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā,* A, 43, 260–281.

(6) Fox, A.J. (1972). Outliers in time series. *Journal of the Royal Statistical Society* , Ser. B, 32, 337–345.

(7) Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: John Wiley.

(8) Harvey, A.C. and Pierse, R. G. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79, 125–131.

(9) Ledolter, J. (1989). Comment on Leave-k-out Diagnostics for Time Series by Bruce. *Journal of the Royal Statistical Society* , Ser. B, 51, 413–414.

(10) Ledolter, J. (1990). Outlier Diagnostics in Time Series Anaysis. *Journal of Time Series Analysis,* 11, 317–324.

(11) Lee, J.J. (1990). *A Study on Influential Observations in Linear Regression and Time Series.* Unpublished Ph.D. dissertation, University of Wisconsin-Madison, Department of Statistics.

(12) Martin, R.D. (1981). Robust methods for time series. In D.F. Findley, editor, *Directions in Time Series*, 683–759, Academic Press, New York.

(13) Martin, R.D. and Yohai, V. J. (1986). Influence functional for time series (with discussion). *The Annals of Statistics*, 14, 781–855.

(14) Peña, D. (1990). Influential observations in time series. *Jounal of Business & Economic Statistics,* 8, 235–241.

(15) Pourahmadi, M.(1989). Estimation and interpolation of missing values of stationary time series. *Journal of Time Series Analysis*, 10, 149–169.

(16) Ryu, G.Y. (1991). *Outlier Detection Diagnostic in Time Series.* Unpublished Ph.D. dissertation, Seoul National University, Department of Computer Science and Statistics.

(17) Tsay, R.S. (1986a). *Effects of outliers in time series.* Technical Report, Department of Statistics, Canegie-Mellon Universty, Pittsburgh.

(18) Tsay, R.S. (1986b). Time series model specification in the presence of outliers. *Journal of the American Statistical Association,* 81, 131–141.

(19) Tsay, R.S. (1988). Outliers, Level Shifts, and Variance Changes in Time Series. *Journal of Forecasting,* 7, 1–20.