J. R. Statist. Soc. B (1993)
55, *No.* 2, *pp.* 559–567

On Outlier Detection in Time Series

By GRETA M. LJUNG†

Massachusetts Institute of Technology, Cambridge, USA

[Received January 1990. Revised March 1992]

SUMMARY

The estimation and detection of outliers in a time series generated by a Gaussian autoregressive moving average process is considered. It is shown that the estimation of additive outliers is directly related to the estimation of missing or deleted observations. A recursive procedure for computing the estimates is given. Likelihood ratio and score criteria for detecting additive outliers are examined and are shown to be closely related to the leave-k-out diagnostics studied by Bruce and Martin. The procedures are contrasted with those appropriate for innovational outliers.

Keywords: AUTOREGRESSIVE MOVING AVERAGE PROCESS; DELETION DIAGNOSTICS; INTERPOLATION; LIKELIHOOD FUNCTION; MISSING VALUES; OUTLIER DETECTION; STUDENTIZED RESIDUALS; TIME SERIES

1. INTRODUCTION

Let $\{z_t\}$ denote an outlier-free time series generated by an autoregressive moving average (ARMA(p, q)) model

$$\phi(B)z_t = \theta(B)a_t \tag{1.1}$$

where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$, $B^k z_t = z_{t-k}$, $E(z_t) = 0$ and $\{a_t\}$ is a sequence of independent and identically distributed $N(0, \sigma^2)$ random variables. The process $\{z_t\}$ is stationary if the 0s of the polynomial $\phi(B)$ lie outside the unit circle. A similar condition on the 0s of $\theta(B)$ ensures that the model is invertible. Under invertibility, the model can be written as $\pi(B)z_t = a_t$, where $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = \phi(B)/\theta(B)$.

Fox (1972) proposed two types of parametric models, additive and innovational, for outliers in time series. The additive outlier model is

$$y_t = z_t + \delta x_t$$

where y_t is the observed value, δ is the magnitude of the outlier and $x_t = 1$ if $t = T$ and $x_t = 0$ otherwise. The model for an innovational outlier is

$$y_t = \frac{\theta(B)}{\phi(B)}(a_t + \delta x_t).$$

In the additive outlier model, the ARMA(p, q) model (1.1) represents most of the data. However, at time T, z_t is contaminated with a gross error δ. In contrast, an innovational outlier represents an unusual shock at time T influencing $z_T, z_{T+1}, \ldots$

†*Address for correspondence*: Department of Mathematics, Room 2-332, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

 0035–9246/93/55559 $2.00

through $\psi(B) = \theta(B)/\phi(B)$. The existence of additive outliers can seriously bias the estimates of the ARMA coefficients and σ^2, whereas innovational outliers in general have a much smaller effect (Chang and Tiao, 1983).

Fox (1972) developed a likelihood ratio test for detecting outliers in a pure autoregressive model. Chang (1982) and Chang and Tiao (1983) extended this test to autoregressive integrated moving average (ARIMA) models and proposed an iterative procedure for detecting multiple outliers. For a discussion of this procedure and some further developments, see also Hillmer *et al.* (1983), Tsay (1986, 1988) and Chang *et al.* (1988). An outlier test based on the score or Lagrange multiplier method was proposed by Abraham and Yatawara (1988), whereas Pena (1987), Abraham and Chuang (1989) and Bruce and Martin (1989) considered the use of deletion diagnostics to detect outliers and influential observations in time series.

The present paper will primarily focus on the estimation and detection of additive outliers in ARMA models. Likelihood-based methods are studied and compared with methods based on observation deletion. Section 2 examines the likelihood function for time series with additive outliers and missing observations. It is shown that the problem of estimating additive outliers is directly related to that of estimating missing or deleted observations. A recursive procedure for estimating additive outliers and smoothing the time series is given in Section 3. Section 4 examines the likelihood ratio and score statistics for detecting additive outliers. A large sample approximation to the probability distribution of these statistics is studied with Monte Carlo simulation. The relationship between the likelihood ratio statistic and deletion diagnostics based on the innovations variance is discussed. Some consideration is also given to the performance of deletion-based methods for innovational outliers.

2. LIKELIHOOD FUNCTION FOR MISSING VALUE AND ADDITIVE OUTLIER MODELS

2.1. *Missing Values*

Given an outlier-free series with no missing values, the likelihood function for the parameters $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$ and σ^2 in model (1.1) is given by

$$p(\mathbf{z}|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2\sigma^2}\mathbf{z}'\Sigma^{-1}\mathbf{z}\right), \qquad (2.1)$$

where $\mathbf{z} = (z_1, \ldots, z_n)'$ and $\Sigma = \sigma^{-2}\operatorname{cov}(\mathbf{z})$. The form of the likelihood function when some observations are missing has been studied by Jones (1980), Ljung (1982, 1989), Harvey and Pierse (1984), Kohn and Ansley (1986), Wincek and Reinsel (1986) and others. For the present purposes it will be convenient to use an expression given by Ljung (1989). Suppose that m successive observations $\mathbf{z}_T = (z_T, \ldots, z_{T+m-1})'$ are missing from the time series and let $\mathbf{z}_{(T)}$ denote the $(n-m)\times 1$ vector of observed values. Further, let $\mathbf{z}_0$ denote the observation vector with 0s substituted for $\mathbf{z}_T$ so that $\mathbf{z} = \mathbf{z}_0 + X\mathbf{z}_T$, where $X_{ij}=1$ for $i = T + j - 1$ and $X_{ij}=0$ otherwise. The likelihood function is then given by

$$p(\mathbf{z}_{(T)}|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-(n-m)/2}|\Sigma|^{-1/2}|X'\Sigma^{-1}X|^{-1/2}\exp\left(-\frac{1}{2\sigma^2}\hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}}\right), \qquad (2.2)$$

where $\hat{\mathbf{z}} = \mathbf{z}_0 + X\hat{\mathbf{z}}_T$ and

$$\hat{\mathbf{z}}_T = -(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{z}_0. \tag{2.3}$$

Expression (2.2) is obtained from the identity $p(\mathbf{z}|\boldsymbol{\beta}, \sigma^2) = p(\mathbf{z}_{(T)}|\boldsymbol{\beta}, \sigma^2)\, p(\mathbf{z}_T|\mathbf{z}_{(T)}, \boldsymbol{\beta}, \sigma^2)$ after partitioning $\mathbf{z}'\Sigma^{-1}\mathbf{z}$ as

$$\mathbf{z}'\Sigma^{-1}\mathbf{z} = \hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}} + (\mathbf{z}_T - \hat{\mathbf{z}}_T)'X'\Sigma^{-1}X(\mathbf{z}_T - \hat{\mathbf{z}}). \tag{2.4}$$

We note that the quadratic form $\hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}}$ in the exponent of function (2.2) has the same form as in function (2.1) with estimates $\hat{\mathbf{z}}_T$ substituted for the missing values $\mathbf{z}_T$. Also, $\hat{\mathbf{z}}_T$ is the conditional expectation of $\mathbf{z}_T$, given $\mathbf{z}_{(T)}$ and $\boldsymbol{\beta}$. A procedure for evaluating $\hat{\mathbf{z}}_T$ was discussed by Ljung (1989).

The likelihood calculations implicit in equations (2.2)–(2.3) resemble the calculations performed at the expectation step of the EM algorithm (Dempster *et al.*, 1977). However, some extra terms are added at the expectation step of the EM algorithm to offset the fact that the parameters $\boldsymbol{\beta}$ in $\hat{\mathbf{z}}_T$ are replaced by estimates from the previous iteration, and thus treated as fixed, at the maximization step. Expression (2.2) assumes the same values for $\boldsymbol{\beta}$ in $\hat{\mathbf{z}}_T$ and Σ^{-1}.

Maximization of functions (2.1) and (2.2) with respect to σ^2 for fixed $\boldsymbol{\beta}$ yields

$$\hat{\sigma}^2 = \frac{1}{n}\mathbf{z}'\Sigma^{-1}\mathbf{z}$$

and

$$\hat{\sigma}^2_{(T)} = \frac{1}{n-m}\hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}}$$

respectively. An expression for the difference between these estimators is readily obtained from equation (2.4). These results will be used in examining outlier diagnostics based on observation deletion in Section 4.

2.2. *Additive Outliers*

We assume now that the series has no missing values but that additive outliers occur at $T, \ldots, T+m-1$. The observed series is then $\mathbf{y} = \mathbf{z} + X\boldsymbol{\delta}_T$, where $\boldsymbol{\delta}_T = (\delta_1, \ldots, \delta_m)'$ represents the outlier effects.

The likelihood function for the parameters $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\delta}_T$ is given by function (2.1) with $\mathbf{z}'\Sigma^{-1}\mathbf{z} = (\mathbf{y} - X\boldsymbol{\delta}_T)'\Sigma^{-1}(\mathbf{y} - X\boldsymbol{\delta}_T)$. For given $\boldsymbol{\beta}$, the likelihood is maximized with respect to $\boldsymbol{\delta}_T$ by

$$\hat{\boldsymbol{\delta}}_T = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{y}, \tag{2.5}$$

which is also the generalized least squares estimator of $\boldsymbol{\delta}_T$. We note that equation (2.5) is an exact finite sample expression for $\hat{\boldsymbol{\delta}}_T$, whereas expressions given by earlier researchers such as Chang and Tiao (1983) and Hillmer *et al.* (1983) for $m=1$ assume an infinite sequence of observations.

It is interesting to compare $\hat{\boldsymbol{\delta}}_T$ with $\hat{\mathbf{z}}_T$ in equation (2.3). Since $\mathbf{y} = \mathbf{z}_0 + X\mathbf{y}_T$, where $\mathbf{y}_T = (y_T, \ldots, y_{T+m-1})'$, we see that

$$\hat{\boldsymbol{\delta}}_T = \mathbf{y}_T - \hat{\mathbf{z}}_T. \tag{2.6}$$

The outliers are thus estimated by the difference between the observed values and the interpolated, or smoothed, values $\hat{\mathbf{z}}_T = E(\mathbf{z}_T|\mathbf{z}_{(T)})$. Alternatively, writing $\hat{\mathbf{z}}_T = \mathbf{y}_T - \hat{\boldsymbol{\delta}}_T$, we see that missing values can be estimated by inserting arbitrary values $\mathbf{y}_T$ into the series and treating the inserted values as additive outliers. Since equations (2.3) and (2.5) are valid for an arbitrary set of indicator variables in X, result (2.6) extends to an arbitrary pattern of additive outliers and missing values.

For given $\boldsymbol{\beta}$, the maximum likelihood estimate of the innovations variance σ^2 in the additive outlier model is $\hat{\sigma}^2_{(\mathrm{AO})} = (1/n)\hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}}$. Ignoring the outliers, σ^2 would be estimated by $\hat{\sigma}^2 = (1/n)\mathbf{y}'\Sigma^{-1}\mathbf{y}$. The relationship between these estimators is given by $\hat{\sigma}^2_{(\mathrm{AO})} = \hat{\sigma}^2 - (1/n)\hat{\boldsymbol{\delta}}'X'\Sigma^{-1}X\hat{\boldsymbol{\delta}}$, where the second term represents the bias in $\hat{\sigma}^2$ due to the outlier effects.

Whereas an additive outlier at time T is estimated by the interpolation residual $\hat{\delta}_T$, an innovational outlier is estimated by the one-step-ahead prediction residual $\hat{a}_T$. The variances of these estimators are $\sigma^2(X'\Sigma^{-1}X)^{-1}$ and σ^2 under their respective models. An incorrect assumption about outlier type results in some loss of efficiency, and the outlier estimation should thus be accompanied by tests for outlier type as in the iterative outlier detection procedure discussed by Chang *et al.* (1988).

3. ESTIMATION OF ADDITIVE OUTLIERS

If T is known, the additive outliers can be estimated from result (2.6) by first computing the interpolated values $\hat{\mathbf{z}}_T$. This approach is implicit in the work of Bruce and Martin (1989). However, direct use of equation (2.5) results in more efficient calculations if the location of the outlier is unknown and estimates are needed for all possible values of T.

To illustrate the procedure, we consider first the case when $m=1$ and z_t follows an AR(p) model $\phi(B)z_t = a_t$. For this model,

$$\Sigma^{-1} = L'_\phi L_\phi - V_\phi V'_\phi, \tag{3.1}$$

where L_ϕ is an $n \times n$ lower triangular band matrix with 1s on the main diagonal, $-\phi_1$ on the first subdiagonal, $-\phi_2$ on the second subdiagonal, and so on, and V_ϕ is an $n \times p$ matrix with (i, j)th element $\phi_{p-(j-i)}$ for $i \leqslant j$, and 0s elsewhere (Ljung and Box, 1979). Now let $e_T = X'\Sigma^{-1}\mathbf{y}$ and $d_T = X'\Sigma^{-1}X$, so that $\hat{\delta}_T = e_T/d_T$. For $T > p$, we then have $e_T = \phi(F)\hat{a}_T$, where $F = B^{-1}$, $\hat{a}_T = \phi(B)y_T$ for $T \leqslant n$ and $\hat{a}_T = 0$ for $T > n$. To obtain e_T for $T \leqslant p$, we note that Σ^{-1} also equals $L_\phi L'_\phi$ except for a $p \times p$ correction term in the lower right-hand corner. Hence $e_T = \phi(B)\tilde{a}_T$, where $\tilde{a}_T = \phi(F)y_T$ for $T \geqslant 1$ and $\tilde{a}_T = 0$ for $T < 1$. The d_Ts are the diagonal elements of Σ^{-1} so that $d_T = d_{n+1-T} = \Sigma_{i=1}^T \phi_{i-1}^2$, with $\phi_0 = -1$ and $\phi_i = 0$, for $i > p$.

The computations can be generalized to the ARMA(p, q) model by writing $\Sigma^{-1} = L'_\pi L_\pi - V_\pi V'_\pi$, where $L_\pi = L_\theta^{-1}L_\phi = L_\phi L_\theta^{-1}$ and L_θ is a matrix similar to L_ϕ. Assuming that n is sufficiently large that the second term $V_\pi V'_\pi$ can be ignored for $T \geqslant n/2$, we have $d_T = d_{n+1-T} = \Sigma_{i=1}^T \pi_{i-1}^2$ with $\pi_0 = -1$. The e_Ts could be computed by using two sets of recursions as above, one for each half of the series. However, the values can be obtained by a single backward recursion $\theta(F)e_T = \phi(F)\hat{a}_T$ if the recursion for the residuals $\hat{a}_T$ is properly initialized by a prior estimation of the unknown presample values. This follows from expressions for Σ^{-1} and the presample estimates given by Ljung and Box (1979) which show that $\mathbf{e} = \Sigma^{-1}\mathbf{y} = L'_\pi\hat{\mathbf{a}}$. Writing $L_\pi = L_\phi L_\theta^{-1}$, we have $L'_\theta\mathbf{e} = L'_\phi\hat{\mathbf{a}}$, so that $\theta(F)e_T = \phi(F)\hat{a}_T$ with $e_T = \hat{a}_T = 0$ for $T > n$. These recursions yield exact estimates for all possible locations T. The calculations described by Chang *et al.* (1988) and earlier researchers are approximate for small T.

Procedures for estimating the unknown presample values have been given by Box and Jenkins (1976), Hillmer and Tiao (1979) and Ljung and Box (1979), where the first two procedures do not require stationarity of $\{z_t\}$. If $\{z_t\}$ is a non-stationary ARIMA(p, d, q) process, the calculations described apply after replacing $\phi(B)$ by $\Phi(B) = \phi(B)(1-B)^d$.

Patches of additive outliers are estimated by $\hat{\boldsymbol{\delta}}_T = D_T^{-1}\mathbf{e}_T$, where $\mathbf{e}_T = (e_T, \ldots, e_{T+m-1})'$ and the elements of $D_T = X'\Sigma^{-1}X$ are sums of squares and cross-products of the π-weights.

Bruce and Martin (1989) and others have noted that a single outlier at time T will affect the outlier estimate not only at time T but also at adjacent time points. The reason for this smearing effect becomes apparent by examining the form of $\hat{\delta}_T$. For example, if z_t follows at AR(p) model, then, for $p < T \leqslant n-p$,

$$\hat{\delta}_T = y_T + c_0^{-1} \sum_{k=1}^{p} c_k(y_{T-k}+y_{T+k}), \tag{3.2}$$

where $c_k = \Sigma_{i=k}^{p} \phi_i \phi_{i-k}$ and the coefficients c_k/c_0 are the inverse autocorrelations of the process. Now, if $y_T = z_T + \delta_T$ and $y_t = z_t$ for $t \neq T$, then for $2p < T \leqslant n-2p$, $E(\hat{\delta}_T) = \delta_T$ and $E(\hat{\delta}_{T-k}) = E(\hat{\delta}_{T+k}) = c_0^{-1}c_k\delta_T$. The effect of the outlier will therefore extend over $2p+1$ time periods according to the values of c_k/c_0. For MA(q) and mixed ARMA(p, q) models, the coefficients c_k/c_0 and the effects are infinite in extent but decay to 0 as k increases.

The calculations described can be extended to estimating the signal in a Gaussian signal-plus-noise model $y_t = z_t + \epsilon_t$, where $\{z_t\}$ is an ARMA(p, q) process and $\{\epsilon_t\}$ is a sequence of $N(0, \sigma_\epsilon^2)$ random variables. Given the parameters of the model, the estimate of the signal is $\hat{z}_t = E(z_t|\mathbf{y}) = y_t - (\sigma^2/\sigma_\epsilon^2)e_t$ with e_t computed as before. A state space approach to estimating z_t is described in Kohn and Ansley (1989).

4. DETECTION OF ADDITIVE OUTLIERS

4.1. *Likelihood Ratio and Score Criteria*

4.1.1. *Autoregressive moving average parameters known*

Given the parameters of the ARMA process, the likelihood ratio statistic for testing the presence of an additive outlier at a known time point T is

$$u_T = d_T^{1/2}\hat{\delta}_T/\sigma. \tag{4.1}$$

Under the null hypothesis of no outlier, u_T is distributed as $N(0, 1)$. The statistic u_T is also the score statistic for testing H_0: $\delta_T = 0$. The score, which is obtained by differentiating the log-likelihood with respect to δ_T, equals e_T/σ^2 and has variance d_T/σ^2.

If the location of the outlier is unknown, we may proceed by calculating $u = \max_{1\leqslant T\leqslant n}|u_T|$. However, the finite sample distribution of u is complicated because of correlation between the u_Ts. Chang *et al.* (1988) used Monte Carlo simulation to determine suitable critical values C. In the present study, we consider choosing C by using an asymptotic result given by Berman (1964).

Let $\rho_k = E(u_T u_{T-k})$ denote the correlation between u_T and u_{T-k} and let $u' = \max_{1\leqslant T\leqslant n}(u_T)$. Berman (1964) showed that if either $\Sigma_{k=1}^{\infty} \rho_k^2 < \infty$ or $\rho_k \log k \rightarrow 0$ as $k \rightarrow \infty$ then

$$\lim_{n\rightarrow\infty} [P\{a_n(u'-b_n) \leqslant x\}] = \exp\{-\exp(-x)\} \tag{4.2}$$

where $a_n = (2 \log n)^{1/2}$ and

$$b_n = (2 \log n)^{1/2} - \tfrac{1}{2}(2 \log n)^{-1/2}\{\log(\log n) + \log(4\pi)\}.$$

This result was generalized from the independent case where, for fixed n, $P\{u' < x\} = \Phi(x)^n$. The correlation among the u_ts is determined by the e_ts which follow the process $\theta(F)e_t = \phi(F)a_t$. This process has the same autocorrelation function as a forward ARMA(q, p) process $\theta(B)e_t = \phi(B)\epsilon_t$, which is stationary assuming that model (1.1) is invertible. For $k > q$, the ρ_ks therefore satisfy the difference equation $\theta(B)\rho_k = 0$. By expressing the solution to this equation in terms of the roots of $\theta(B) = 0$, it is then easy to see that $\Sigma_{k=1}^{\infty} \rho_k^2 < \infty$, establishing equation (4.2). The percentiles of $u = \max_{1 \leqslant T \leqslant n} |u_T|$ can now be determined by observing that the largest and the smallest values have the same distribution.

4.1.2. *Autoregressive moving average parameters unknown*

In practice, the ARMA parameters are unknown and have to be estimated from the data. The score test replaces the parameters in equation (4.1) by estimates computed under the hypothesis of no outliers. This statistic is used at the first step of the iterative outlier detection procedure discussed by Chang *et al.* (1988). As mentioned earlier, however, an additive outlier will bias the parameter estimates. In particular, σ^2 tends to be overestimated, thus lowering the power of $\hat{u}_T$.

If σ^2 is unknown but $\boldsymbol{\beta}$ is given, the likelihood ratio criterion for testing H_0: $\delta_T = 0$ against H_1: $\delta_T \neq 0$ is

$$\lambda = (\mathbf{y} - X\hat{\delta}_T)'\Sigma^{-1}(\mathbf{y} - X\hat{\delta}_T)/\mathbf{z}'\Sigma^{-1}\mathbf{z} = \hat{\mathbf{z}}'\Sigma^{-1}\hat{\mathbf{z}}/\mathbf{z}'\Sigma^{-1}\mathbf{z}.$$

Equivalently, using equation (2.4), we have $\lambda = \{1 + (n-1)u_T^{*2}\}^{-1}$ where

$$u_T^* = d_T^{1/2}\hat{\delta}_T/\hat{\sigma}_{(T)} \tag{4.3}$$

with $\hat{\sigma}_{(T)} = \{(n\hat{\sigma}^2 - d_T\hat{\delta}_T^2)/(n-1)\}^{1/2}$. If H_0 is true, u_T^{*2} is distributed as $F(1, n-1)$ and u_T^* as t_{n-1}.

In the likelihood ratio test, unknown βs are estimated under hypothesis H_1 as well as under hypothesis H_0. However, if the location of the outlier is unknown, re-estimation of $\boldsymbol{\beta}$ for each T requires lengthy computations. As an approximation we therefore use $\hat{u}_T^* = \hat{d}_T^{1/2}\hat{\delta}_T/\hat{\sigma}_{(T)}$ with unknown βs estimated under hypothesis H_0 only. This statistic agrees with the score statistic $\hat{u}_T$ but uses a 'delete-one' estimate of σ. Since $\hat{\sigma}_{(T)}$ decreases at the location of an outlier, this simple modification results in increased power. We note that $\hat{u}_T^*$ resembles an externally Studentized prediction residual in linear regression.

For $m > 1$, the approximate likelihood ratio statistic is $\hat{\boldsymbol{\delta}}_T'\hat{D}_T\hat{\boldsymbol{\delta}}_T/\hat{\sigma}_{(T)}^2$.

4.2. *Some Numerical Results*

Table 1 shows estimated significance levels of $\hat{u} = \max_{1 \leqslant T \leqslant n} |\hat{u}_T|$ and $\hat{u}^* = \max_{1 \leqslant T \leqslant n} |\hat{u}_T^*|$ based on 5000 realizations of a first-order autoregressive process with ϕ equal to 0.0, 0.5 and 0.9. The entries show the proportion of times that the two statistics exceeded the 10%, 5% and 1% significance points given by equation (4.2) for n equal to 50, 100 and 200. The results for $\hat{u}^*$ are generally close to the nominal values, suggesting that equation (4.2) is a useful approximation for $\hat{u}^*$. The empirical

TABLE 1
Empirical significance levels of the statistics $\hat{u} = max_{1\leqslant T\leqslant n}|\hat{u}_T|$ and $\hat{u}^* = max_{1\leqslant T\leqslant n}|\hat{u}_T^*|$ based on 5000 replications of an AR(1) process

n	ϕ	$\hat{u}$ 0.10	$\hat{u}$ 0.05	$\hat{u}$ 0.01	$\hat{u}^*$ 0.10	$\hat{u}^*$ 0.05	$\hat{u}^*$ 0.01
50	0.0	0.051	0.018	0.001	0.129	0.064	0.011
	0.5	0.043	0.015	0.001	0.110	0.053	0.009
	0.9	0.034	0.010	0.001	0.091	0.044	0.006
100	0.0	0.056	0.021	0.002	0.105	0.045	0.007
	0.5	0.054	0.021	0.001	0.101	0.047	0.006
	0.9	0.055	0.018	0.001	0.097	0.047	0.007
200	0.0	0.068	0.028	0.003	0.097	0.045	0.008
	0.5	0.068	0.026	0.003	0.101	0.043	0.005
	0.9	0.062	0.024	0.003	0.091	0.041	0.006

Significance levels for the following statistics and nominal levels.

levels of $\hat{u}$ are smaller and converge only slowly to the values given by equation (4.2).

Table 2 shows the power of $\hat{u}$ and $\hat{u}^*$ for series of length $n = 50$ and $n = 100$ generated from the AR(1) model with an additive outlier inserted at $T = n/2$. The computations were performed for two outlier sizes and with $C = 3.5$, which is one of the critical values recommended by Chang *et al.* (1988). The value is close to the 5% significance points given by equation (4.2) which are 3.42 for $n = 50$ and 3.58 for $n = 100$. We note that $\hat{u}^*$ outperforms $\hat{u}$ for the smaller values of δ_T and n with some of the differences attributable to differences in level. Both statistics are more powerful for larger values of ϕ. This is expected since $E(u_T^2) = 1 + (1 + \phi^2)\delta_T^2/\sigma^2$ so that the non-centrality increases with $|\phi|$.

4.3. *Deletion Diagnostics*

The use of delete-one diagnostics to detect atypical observations is a well-established tool in regression analysis. In time series analysis, observation deletion has been considered by Pena (1987), Abraham and Chuang (1989) and Bruce and Martin (1989) for ARMA models, while some related state space developments are given by

TABLE 2
Frequency of correct detection of an additive outlier at $T = n/2$ in 1000 replications of an AR(1) process†

n	ϕ	$\delta_T = 3\sigma$: $\hat{u}$	$\delta_T = 3\sigma$: $\hat{u}^*$	$\delta_T = 5\sigma$: $\hat{u}$	$\delta_T = 5\sigma$: $\hat{u}^*$
50	0.3	170	316	826	907
	0.7	261	456	927	976
	0.9	400	599	976	995
100	0.3	273	352	912	940
	0.7	460	540	987	994
	0.9	610	692	998	998

Frequency for the following statistics.

†Critical value $C = 3.5$.

Kohn and Ansley (1989). In the ARMA context, Bruce and Martin (1989) found that diagnostics based on the innovations variance σ^2 give a clearer indication of the location of an outlier than diagnostics based on the coefficients $\boldsymbol{\beta}$.

Using equation (2.4), it is easy to see that the likelihood ratio statistic for additive outliers measures the change in the estimate of σ^2 when the observation at time T is deleted from the series. For $m = 1$ and $\boldsymbol{\beta}$ given, we have $d_T \hat{\delta}_T^2 = n\hat{\sigma}^2 - (n-1)\hat{\sigma}_{(T)}^2$, so that

$$u_T^{*2} \doteq n\left(\frac{\hat{\sigma}^2 - \hat{\sigma}_{(T)}^2}{\hat{\sigma}_{(T)}^2}\right). \tag{4.4}$$

For $m > 1$, the statistic $\hat{\boldsymbol{\delta}}_T' D_T \hat{\boldsymbol{\delta}}_T / \hat{\sigma}_{(T)}^2$ similarly measures the change in the variance estimate when m observations are deleted from the series. It is interesting that approximation (4.4) can be computed for all T without removing any observations from the series.

When $\boldsymbol{\beta}$ is unknown, the likelihood ratio criterion is

$$\hat{\lambda} = (n-1)\hat{\sigma}_{(T)}^2 / n\hat{\sigma}^2 \doteq \hat{\sigma}_{(T)}^2 / \hat{\sigma}^2,$$

where, in $\hat{\sigma}_{(T)}^2$, $\boldsymbol{\beta}$ is estimated by deleting the value at time T. The quantity $n(\hat{\lambda}^{-1} - 1)$ is now of the form (4.4). Bruce and Martin (1989) studied the behaviour of DV $= (n/2)(\hat{\sigma}^2/\hat{\sigma}_{(T)}^2 - 1)^2$ where, for each T, $\boldsymbol{\beta}$ was re-estimated by using the likelihood procedure described by Harvey and Pierse (1984). Bruce and Martin (1989) argued that DV outperforms the Chang–Tiao statistic $\hat{u}_T = \hat{d}_T^{1/2}\hat{\delta}_T/\hat{\sigma}$ which uses estimates of $\boldsymbol{\beta}$ and σ^2 based on all the data. The results given here show, however, that the difference in power can be reduced by simply replacing $\hat{\sigma}^2$ by $\hat{\sigma}_{(T)}^2$. Although a re-estimation of $\boldsymbol{\beta}$ for each T may result in further improvements, we believe that the gain in power will in many cases not be sufficiently large to justify the increased computational burden.

Because of their near equivalence to likelihood ratio tests, deletion diagnostics based on $\hat{\sigma}^2$ are expected to perform well for additive outliers. However, if there are innovational outliers or other types of disturbances in the series, procedures specifically designed to detect such disturbances (see, for example, Tsay (1988)) will in general be more powerful.

To illustrate this point, suppose that an innovational outlier occurs at some time point T. The score statistic for detecting this outlier is $\hat{v}_T = \hat{a}_T/\hat{\sigma}$. The likelihood ratio criterion may be approximated by $\hat{v}_T^* = \hat{a}_T/\hat{\sigma}_{(T)}$, where $\hat{\sigma}_{(T)} = \{\Sigma_{t \neq T} \hat{a}_t^2/(n-1)\}^{1/2}$. Some simulations performed for the AR(1) model with an innovational outlier δ_T inserted at $T = n/2$ showed that $\hat{v}^* = \max_{1 \leqslant T \leqslant n} |\hat{v}_T^*|$ outperformed the additive outlier criterion $\hat{u}^* = \max_{1 \leqslant T \leqslant n} |\hat{u}_T^*|$ for large $|\phi|$. For example, for $n = 150$, $\delta_T = 5$ and $\phi = 0.95$, $\hat{v}^*$ detected the outlier correctly in 912 out of 1000 series whereas the corresponding figure for $\hat{u}^*$ was only 360. For $\phi = 0.3$, the detection frequencies were 914 and 859 respectively. The difference in power is suggested by the expected values $E(v_T) = \delta_T/\sigma$ and

$$E(u_T) = \frac{\delta_T/\sigma}{(1+\phi^2)^{1/2}},$$

where the latter decreases with increasing $|\phi|$. In contrast, if δ_T is an additive outlier, then $E(u_T) = (\delta_T/\sigma)(1+\phi^2)^{1/2}$ and $E(v_T) = -\delta_T/\sigma$, and $\hat{u}^*$ will be more powerful. Procedures for distinguishing between the two types of outliers have been given by Chang (1982) and Muirhead (1986).

ACKNOWLEDGEMENTS

This work was carried out with partial support of the National Science Foundation under grant DMS-89-14344. The work was also supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

REFERENCES

Abraham, B. and Chuang, A. (1989) Outlier detection and time series modeling. *Technometrics*, **31**, 241–248.

Abraham, B. and Yatawara, N. (1988) A score test for detection of time series outliers. *J. Time Ser. Anal.*, **9**, 109–119.

Berman, S. M. (1964) Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.*, **35**, 502–516.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, sect. 6.4.3. San Francisco: Holden-Day.

Bruce, A. G. and Martin, R. D. (1989) Leave-k-out diagnostics for time series (with discussion). *J. R. Statist. Soc.* B, **51**, 363–424.

Chang, I. (1982) Outliers in time series. *PhD Dissertation*. Department of Statistics, University of Wisconsin, Madison.

Chang, I. and Tiao, G. C. (1983) Estimation of time series parameters in the presence of outliers. *Technical Report 8*. Statistics Research Center, University of Chicago, Chicago.

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

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Fox, A. J. (1972) Outliers in time series. *J. R. Statist. Soc.* B, **34**, 350–363.

Harvey, A. C. and Pierse, R. G. (1984) Estimating missing observations in economic time series. *J. Am. Statist. Ass.*, **79**, 125–131.

Hillmer, S. C., Bell, W. R. and Tiao, G. C. (1983) Modeling considerations in the seasonal adjustment of economic time series. In *Applied Time Series Analysis of Economic Data* (ed. A. Zellner), pp. 74–100. Washington DC: United States Bureau of the Census.

Hillmer, S. C. and Tiao, G. C. (1979) Likelihood function of stationary multiple autoregressive moving average models. *J. Am. Statist. Ass.*, **74**, 652–660.

Jones, R. H. (1980) Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, **22**, 389–395.

Kohn, R. and Ansley, C. F. (1986) Estimation, prediction, and interpolation for ARIMA models with missing data. *J. Am. Statist. Ass.*, **81**, 751–761.

——— (1989) A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika*, **76**, 65–79.

Ljung, G. M. (1982) The likelihood function for a stationary Gaussian autoregressive-moving average process with missing observations. *Biometrika*, **69**, 265–268.

——— (1989) A note on the estimation of missing values in time series. *Communs Statist.*, **18**, 459–465.

Ljung, G. M. and Box, G. E. P. (1979) The likelihood function of stationary autoregressive-moving average models. *Biometrika*, **66**, 265–270.

Muirhead, C. R. (1986) Distinguishing outlier types in time series. *J. R. Statist. Soc.* B, **48**, 39–47.

Pena, D. (1987) Measuring the importance of outliers in ARIMA models. In *New Perspectives in Theoretical and Applied Statistics* (eds M. L. Puri, J. P. Vilaplana and W. Wertz). New York: Wiley.

Tsay, R. S. (1986) Time series model specification in the presence of outliers. *J. Am. Statist. Ass.*, **81**, 132–141.

——— (1988) Outliers, level shifts and variance changes in time series. *J. Forecast.*, **7**, 1–20.

Wincek, M. A. and Reinsel, G. C. (1986) An exact maximum likelihood estimation procedure for regression-ARMA time series models with possibly nonconsecutive data. *J. R. Statist. Soc.* B, **48**, 303–313.