TECHNOMETRICS, MAY 1989, VOL. 31, NO. 2

Outlier Detection and Time Series Modeling

Bovas Abraham

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Alice Chuang

Department of Applied Statistics and Operations Research
Bowling Green State University
Bowling Green, OH 43403

Some statistics used in regression analysis are considered for detection of outliers in time series. Approximations and asymptotic distributions of these statistics are considered. A method is proposed for distinguishing an observational outlier from an innovational one. A four-step procedure for modeling time series in the presence of outliers is also proposed, and an example is presented to illustrate the methodology.

KEY WORDS: ARMA process; Autoregressive process; Diagnostic checking.

1. INTRODUCTION

Suppose that $z_t (t = 0, \pm 1, \pm 2, \ldots)$ is a stationary time series that can be modeled by the autoregressive moving average (ARMA) process

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

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, and B is a backward shift operator such that $Bz_t = z_{t-1}$. The polynomials $\phi(B)$ and $\theta(B)$ are assumed to have all roots outside the unit circle and to have no factors in common; $\{a_t\}$ is a sequence of iid normal random variables with mean 0 and variance σ^2. Box and Jenkins (1976) suggested a three-stage strategy of model selection, estimation, and diagnostic checking for building models of the form (1.1).

Observed time series are sometimes influenced by outliers that may be the result of gross errors, special promotions, strikes, certain system changes, and so forth. Some suspected outliers may have large residuals but may not affect the parameter estimates, whereas others may not only have large residuals but also may affect model specification and parameter estimation. This may lead to misspecified models, biased estimates, and biased forecasts. Thus it is very important in practice to detect outliers and to have a procedure for model building (specification, estimation, and checking) in the presence of outliers.

Characterization of Outliers

Fox (1972) and Abraham and Box (1979) discussed two characterizations of outliers that are found in time series data. We briefly outline these now. Consider the model

$$z_t = y_t + \delta \xi_t(T), \qquad \phi(B) y_t = \theta(B) a_t, \tag{1.2}$$

where δ is a constant,

$$\begin{aligned} \xi_t(T) &= 0, \qquad t \neq T \\ &= 1, \qquad t = T, \end{aligned} \tag{1.3}$$

and z_t is what we really observe. This model is usually referred to as an aberrant observation (AO) model. Alternatively one also can consider the model

$$z_t = \phi^{-1}(B)\theta(B)(a_t + \delta \xi_t(T)), \tag{1.4}$$

where $\xi_t(T)$ is as defined in (1.3) and the aberration is in the innovation at $t = T$. This is referred to as an aberrant innovation (AI) model.

Several authors have considered outlier-related issues in time series. Chang and Tiao (1983) suggested an iterative maximum likelihood procedure, and Martin (1980) proposed a robust approach for parameter estimation. Martin and Yohai (1986) studied some influence functionals in the context of time series. Muirhead (1986) discussed a method of distinguishing outlier types in an autoregressive (AR) process. Pena (1987) considered some measures to identify outliers that have large influence on parameter estimates. All of these approaches are based on the assumption that the correct model form is known. Tsay (1986) suggested an iterative procedure consisting of specification, detection, and removing cycles to reduce the outlier effects on specification. The computations in all of these approaches are cumbersome.

Several authors (Andrews and Pregibon 1980; Cook and Weisberg 1982; Draper and John 1981) discussed detection of outliers and influential points in regression models. Their approach was basically to delete suspicious observations and to build a measure of the resulting change in features of the model, such as the estimated parameter values and residuals. This article adapts the schemes used in regression analysis to time series situations. In the following sections, several statistics are described for detecting outliers in time series data; approximations are given for those in which the numerical effort is quite involved. In Section 2, we introduce some statistics for outlier detection in an AR process of order p [AR(p)] and discuss the behavior of these statistics when the models are overfitted. Section 3 presents a four-step procedure for modeling time series data in the presence of outliers. In this procedure, an ARMA(p, q) process given in (1.1) is approximated by an AR($p + q$) process in which the outlier detection and adjustment for an AR process is carried out. The adjusted series is then subjected to the usual model-building strategy of Box and Jenkins (1976).

2. SOME DIAGNOSTIC CHECKING MEASURES

Let z_t be a stationary AR process of order p [AR(p)],

$$\phi(B)z_t = a_t, \tag{2.1}$$

where $\phi(B)$ is as defined in (1.1). Given a set of observations $z_1, z_2, \ldots, z_n$, we can write

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\phi} + \mathbf{a}, \tag{2.2}$$

where $\mathbf{Z}' = (z_{p+1}, \ldots, z_n)$, $\boldsymbol{\phi}' = (\phi_1, \ldots, \phi_p)$, $\mathbf{a}' = (a_{p+1}, \ldots, a_n)$, and

$$\mathbf{X} = \begin{bmatrix} z_p & z_{p-1} & \cdots & z_1 \\ z_{p+1} & z_p & \cdots & z_2 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ z_{n-1} & z_{n-2} & \cdots & z_{n-p} \end{bmatrix}.$$

Then the conditional least squares (CLS) estimate of $\boldsymbol{\phi}$ is given by

$$\hat{\boldsymbol{\phi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \tag{2.3}$$

the fitted values are

$$\hat{\mathbf{Z}} = \mathbf{X}\hat{\boldsymbol{\phi}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} = H\mathbf{Z}, \tag{2.4}$$

where $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and the residuals are $\mathbf{e} = (I - H)\mathbf{Z}$. In a linear regression model, $\mathbf{X}$ is assumed to be a constant matrix, which is no longer true in the present situation. Moreover, the z_t's are assumed independent in the linear regression model. An observation could be deleted without affecting the consecutive ones, and the deletion of an equation in (2.2) is equivalent to the deletion of an observation. In the time series context, however, this is no longer true either. A suspect observation, z_T, is involved not only in one equation but in $p + 1$ consecutive equations of (2.2). Thus it may be necessary to delete not only one equation but $p + 1$ equations from (2.2).

Suppose that there is one suspected observation at $t = T$. The matrix $\mathbf{X}$ and vectors $\mathbf{Z}$ and $\mathbf{e}$ can be partitioned as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \begin{matrix} (T-p) \times p \\ k \times p \\ (n-T-k) \times \mathrm{p} \end{matrix},$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \mathbf{Z}_3 \end{bmatrix} \begin{matrix} (T-p) \times 1 \\ k \times 1 \\ (n-T-k) \times 1 \end{matrix},$$

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix} \begin{matrix} (T-p) \times 1 \\ k \times 1 \\ (n-T-k) \times 1 \end{matrix},$$

where k is the number of equations that are to be deleted. The residuals, $\mathbf{e}$, can be expressed in the partitioned form as

$$\mathbf{e} = \begin{bmatrix} I - H_{11} & -H_{12} & -H_{13} \\ -H_{21} & I - H_{22} & -H_{23} \\ -H_{31} & -H_{32} & I - H_{33} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \mathbf{Z}_3 \end{bmatrix}, \tag{2.5}$$

where

$$H_{ij} = \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_j', \qquad i, j = 1, 2, 3. \tag{2.6}$$

Following the suggestion of Draper and John (1981) for regression situations, we consider the statistics

$$Q_{k(T)} = \mathbf{e}_2'(I - H_{22})^{-1}\mathbf{e}_2 \tag{2.7}$$

and

$$AP_{k(T)} = (1 - Q_{k(T)}/\mathrm{RSS})|I - H_{22}|, \tag{2.8}$$

where RSS is the residual sum of squares. When $k = 1$, $\mathbf{e}_2' = e_T$, and when $k = p + 1$, $\mathbf{e}_2' = (e_T, \ldots, e_{T-p})$. Now $Q_{k(T)}$ could be decomposed into two terms:

$$\begin{aligned} Q_{k(T)} &= \mathbf{e}_2'\mathbf{e}_2 + (\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}_*)' \\ &\qquad \times (\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_3'\mathbf{X}_3)(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}_*) \\ &= Q_{k1(T)} + Q_{k2(T)}, \end{aligned} \tag{2.9}$$

where $\hat{\boldsymbol{\phi}}_* = (\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_3'\mathbf{X}_3)^{-1}(\mathbf{X}_1'\mathbf{Z}_1 + \mathbf{X}_3'\mathbf{Z}_3)$ is the estimate of $\boldsymbol{\phi}$ after deleting k equations from (2.2). Note here that $Q_{k2(T)}$ and Cook's statistic with k observations deleted,

$$C_{k(T)} = \mathbf{e}_2'(I - H_{22})^{-1}H_{22}(I - H_{22})^{-1}\mathbf{e}_2/(p\hat{\sigma}^2), \tag{2.10}$$

are related as follows:

$$\frac{1}{1 - \lambda_1} \le p\hat{\sigma}^2 \frac{C_{k(T)}}{Q_{k2(T)}} \le \frac{1}{1 - \lambda_k}, \tag{2.11}$$

where $\hat{\sigma}^2 = \text{RSS}/(n - p - 1)$ and λ_1 and λ_k are the smallest and the largest eigenvalues of H_{22}.

It is found in simulations that the statistics Q_k, Q_{k1}, and Q_{k2} are useful indicators for outliers. They further suggest that the sampling behavior of the AP statistic is difficult to interpret, and hence we will consider only the Q's hereafter.

2.1 Patterns of the Q Statistics

The statistics defined in (2.7) and (2.9) are functions of e_t's and h_{ij} $[i, j = t, \ldots, (t + k - 1)]$. Their behavior is different for AO and AI outliers. Hence they may be used not only to detect an outlier but also to distinguish an AO outlier from an AI one.

A suspected AO outlier at $t = T$ will affect z_T by δ and, consequently, e_{T+i} by $\phi_i\delta$ $(i = 0, 1, \ldots, p;$ $\phi_0 = 1)$. An AI outlier will affect e_T by δ and hence z_{T+i} by $\psi_i\delta$ $(i = 0, 1, \ldots)$, where ψ_i is the coefficient of B^i in $\psi(B) = \phi^{-1}(B) = 1 - \psi_1 B - \psi_2 B^2 - \cdots$ $- (\psi_0 = 0)$.

For illustration of the patterns, consider an AR(1) process. Suppose that $k = 1$; then $H_{22} = h_{TT} = z_{T-1}^2/\sum_{t=2}^n z_{t-1}^2$, $Q_{1(T)} = e_T^2(1 - z_{T-1}^2/\sum_{t=2}^n z_{t-1}^2)^{-1}$, $Q_{11(T)} = e_T^2$, and $Q_{12(T)} = e_T^2 z_{T-1}^2/\sum_{t \neq T} z_{t-1}^2 = e_T^2(h_{TT}/(1 - h_{TT}))$. Q_{11} depends only on e_T, whereas Q_1 and Q_{12} depend on e_T and h_{TT}; however, h_{TT} is relatively small compared with 1, and the behavior of Q_1 is dominated by e_T. Now $h_{TT}/(1 - h_{TT})$ is a monotone function of h_{TT}, and it is a measure of the distance of X_2 to the center of the ellipsoid formed by $(\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_3'\mathbf{X}_3)$. Thus the behavior of Q_{12} depends on $e_T^2 z_{T-1}^2$.

If the outlier at $t = T$ is an AO, then e_T and e_{T+1} are affected, and hence $Q_{11(T)}$, $Q_{11(T+1)}$, and $Q_{1(T)}$, $Q_{1(T+1)}$ are large compared with the rest. Although $Q_{12(T)}$ and $Q_{12(T+1)}$ are influenced by the outlier at $t = T$, often the latter is the larger, because e_{T+1} and z_T are affected by the outlier.

If the outlier is AI, then only e_T is affected, which implies that $Q_{1(T)}$ and $Q_{11(T)}$ are large compared with others. The behavior of $Q_{12(T)}$ is less reliable, since observations $z_T, \ldots, z_n$ are all affected.

The patterns of these statistics are illustrated by the following two time series, generated using the following models:

Model 1 (AO). $y_t = z_t + 4.5\xi_t(80)$ and $z_t = .5z_{t-1} + a_t$.

Model 2 (AI). $z_t = .5z_{t-1} + a_t + 4.5\xi_t(80)$ and $y_t = z_t$.

In Models 1 and 2, $t = 1, 2, \ldots, 100$;

$$\xi_t(80) = 0, \qquad t \neq 80$$
$$= 1, \qquad t = 80;$$

and $\{a_t\}$ is a white noise sequence with mean 0 and variance $\sigma^2 = 1$.

For Model (1), the plot of the statistic $Q_{1(t)}$ versus t is given in Figure 1a. It is seen that $Q_{1(80)}$ and $Q_{1(81)}$ are large compared with the rest. The plot of $Q_{12(t)}$ versus t indicates that $Q_{12(80)}$ is small and $Q_{12(81)}$ is very large (see Fig. 1b). These are typical patterns for the Q_1 and Q_{12} statistics in the presence of an AO outlier at $t = 80$ in an AR(1) process. Note also that in this example $Q_{2(79)}$, $Q_{2(80)}$, and $Q_{2(81)}$ are large (when $k = 2$) compared with others; $Q_{22(80)}$ and $Q_{22(81)}$ are also large. These are not shown here to save space.

For Model 2, Figure 2 gives the plot of (a) $Q_{1(t)}$ versus t and (b) $Q_{12(t)}$ versus t, respectively. We find that $Q_{1(80)}$ and $Q_{12(80)}$ are large compared with the rest. This is the characteristic of an AI outlier at $t = 80$ for an AR(1) process. When $k = 2$, it can also be seen that $Q_{2(79)}$ and $Q_{2(80)}$ are large compared with the other $Q_{2(t)}$'s. The plot of Q_2 is not shown here to save space.

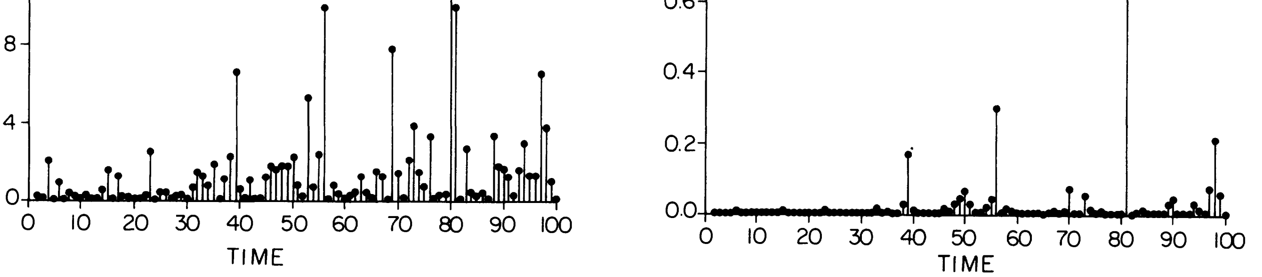


Figure 1. $Q_{1(t)}$ (a) and $Q_{12(t)}$ (b) Versus Time Data From an AR(1) Process ($\phi = .5$, $\sigma^2 = 1$) With AO at $t = 80$.

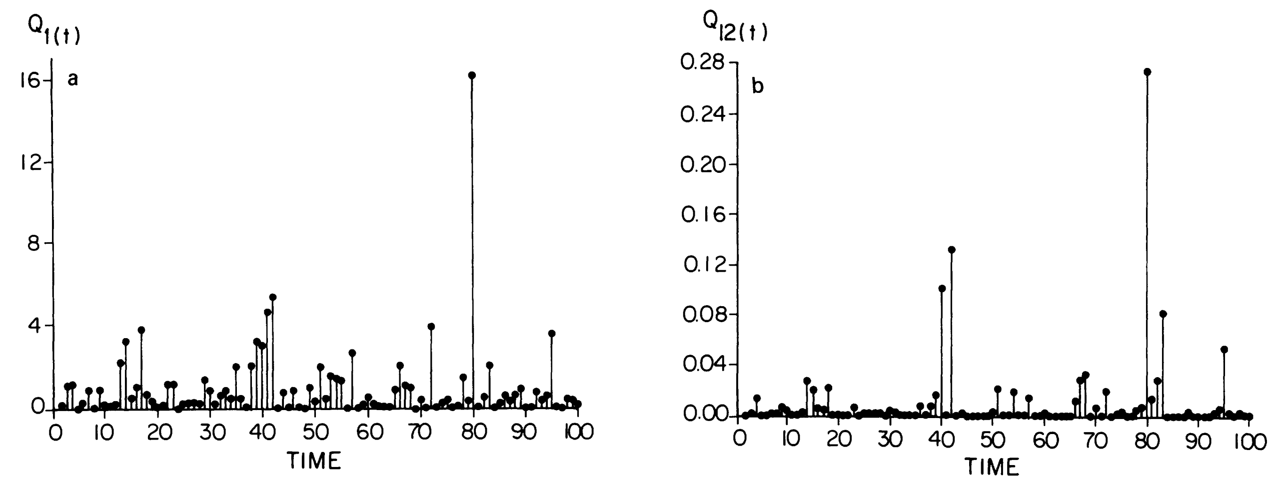


Figure 2. $Q_{1(t)}$ (a) and $Q_{12(t)}$ (b) Versus Time Data From an AR(1) Process (ϕ = .5, σ^2 = 1) With AI at t = 80.

The patterns of these statistics for higher-order ($p > 1$) processes are similar and are summarized in Table 1. In general, our simulation experience indicates that Q_k (or Q_{k1}) is more useful in detecting outliers than Q_{k2}.

2.2 Approximations

In practical situations the identity of the outlier may not be known. Hence the procedure suggested in Section 2.1 requires that $Q_{k(t)}$, $Q_{k1(t)}$, and $Q_{k2(t)}$ be computed for all $t = p + 1, p + 2, \ldots, (n - k + 1)$, and this requires $(n - k - p + 1)$ inversions of the matrix $(I - H_{22})$, which may invite numerical problems. Recursive computational procedures to avoid this inversion can be time-consuming and may result in large rounding errors. If off-diagonal elements, $-h_{ij}$, of $(I - H_{22})$ are small in absolute value, the following approximation, in which no matrix inversion is performed, can be made:

$$Q_{k(t)} \approx \sum_{i=t}^{t+k-1} e_i^2/(1 - h_{ii}). \tag{2.12}$$

This approximation is usually adequate in large samples. Once $Q_{k(t)}$ is obtained, $Q_{k2(t)}$ can be obtained by subtracting $Q_{k1(t)} = \mathbf{e}_2'\mathbf{e}_2$ from $Q_{k(t)}$. These approximations are also attractive because the exact forms require the computation and the storage of the symmetric band matrix H (i.e., $h_{t,t-k}, \ldots, h_{t,t+k}$ for all t), whereas the approximations require these only for the diagonal elements of H. Note that when $k = 1$ (deleting one equation) the exact values and the approximations are the same.

2.3 Asymptotic Distributions

We consider the statistics $\max_t Q_{k(t)}$, $\max_t Q_{k1(t)}$, and $\max_t Q_{k2(t)}$ for identifying outlier locations for which the sampling properties of these statistics are required. The exact sampling distributions are difficult to come by, however, and hence we appeal to large-sample theory.

If there are no outliers, it is well known that $\hat{\boldsymbol{\phi}}$ converge in probability to $\boldsymbol{\phi}$ ($\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}$) and $\hat{\sigma}_z^2 \xrightarrow{p} \sigma_z^2$, where $\hat{\sigma}_z^2 = \sum_{t=p+1}^n (z_t - \bar{z})^2/(n - p)$ (e.g., see Anderson 1971). It is also easy to show that the re-

Table 1. Patterns of Q Statistics Assuming an Outlier at t = T

Statistic	*AI outlier*	*AO outlier*
Q_{11}, Q_1, deleting one equation ($k = 1$)	Large value at $t = T$ and small values elsewhere.	The values at $t = T, T+1, \ldots, T + p$ are affected.
Q_{12}, deleting one equation ($k = 1$)	The values at $t = T, T + 1, \ldots,$ are affected (less reliable).	The values at $t = T, T + 1, \ldots, T + p$ are affected.
$Q_{(p+1)1}$, $Q_{(p+1)}$, deleting $p + 1$ equations ($k = p + 1$)	Large values at $t = T - p, T - p + 1, \ldots, T$, and small values elsewhere.	The values at $t = T - p, T - p + 1, \ldots, T + p$ are affected, with the largest value at $t = T$.
$Q_{(p+1)2}$, deleting $p + 1$ equations ($k = p + 1$)	The values at $t = T - p, \ldots, T, \ldots,$ are affected (less reliable).	The values at $t = T - p, T - p + 1, \ldots, T + p$ are affected, with the largest value at $t = T$.

sidual, e_t, converges in probability to a_t and the elements of H converge to 0 as n increases. Then it can be seen that

$$Q_{k1(t)} \xrightarrow{p} Q^*_{k(t)} = \sum_{i=t}^{t+k-1} a_i^2 \sim \sigma^2\chi^2_{(k)}, \quad (2.13)$$

and

$$Q_{k(t)} \xrightarrow{p} Q^*_{k(t)}, \qquad Q_{k2(t)} \xrightarrow{p} 0,$$

$$\max_t Q_{k1(t)} \xrightarrow{p} \max_t Q^*_{k(t)},$$

$$\max_t Q_{k(t)} \xrightarrow{p} \max_t Q^*_{k(t)}, \quad (2.14)$$

where $\chi^2_{(k)}$ denote a chi-squared distribution with k df. If $k = 1$, then $\{Q^*_{k(t)}\}$ is a sequence of iid $\chi^2_{(1)}$ variables, whereas it is a dependent sequence of $\chi^2_{(k)}$ variables for $k \geq 2$.

Case 1: $k = 1$. Let $F_1(\cdot)$ denote the cdf of $\sigma^2\chi^2_{(1)}$ and $\tau = m[1 - F_1(C_m(\tau))]$, where $m = n - p$ and $C_m(\tau)$ is a critical value. Then Chuang (1987) showed that $\Pr(\max_t Q^*_{1(t)} \leq C_m(\tau)) \to \exp(-\tau)$ as $m \to \infty$. Therefore, given a significance level α, the critical value $C_m(\tau)$ can be obtained by solving $C_m(\tau) = F_1^{-1}(1 + (\ln(1 - \alpha))/m)$. The same asymptotic distribution holds for $\max_t Q_{11(t)}$ and $\max_t Q_{1(t)}$.

Case 2: $k \geq 2$. Let $F_k(\cdot)$ denote the cdf of a $\sigma^2\chi^2_{(k)}$. Then it can be shown that

$$\Pr(\max_t Q^*_{k(t)} \leq C_m(\tau)) \to \exp(-\nu\tau) \quad (2.15)$$

(see Chuang 1987), where, for some ν $(0 < \nu \leq 1)$ and for each $\tau > 0$, $m[1 - F(C_m(\tau))] \to \tau$ as $m = n - p - k + 1 \to \infty$. Given a significance level α, we have $\tau = -\ln(1 - \alpha)/\nu$, and the significance point $C_m(\tau)$ can be obtained from

$$C_m(\tau) = F_k^{-1}[1 + (\ln(1 - \alpha))/(\nu m)]. \quad (2.16)$$

Now $\max_t Q_{k1(t)}$ and $\max_t Q_{k(t)}$ have the same distribution as in (2.15). The computation of ν was discussed by Chuang (1987).

Simulation Study. We conducted a simulation study to gain some understanding of the accuracy of the extreme-value type of approximations made. Samples of size $n = 100$ and 200 were generated from the following models:

Model 3. $z_t = .5z_{t-1} + a_t$ and $a_t \sim N(0, 1)$.

Model 4. $z_t = z_{t-1} - .5z_{t-2} + a_t$ and $a_t \sim N(0, 1)$.

For Model 3, the statistics $\max_t Q_{k(t)}$ and $\max_t Q_{k1(t)}$ were calculated for $k = 1$ and 2, whereas for Model 4 these were calculated for $k = 1$ and 3. This process was repeated 1,000 times. The 1%, 2.5%, 5%, and 10% significance points were estimated. These are shown in Tables 2–5, together with those obtained from the extreme-value approximations. (Since calculations were done in double precision and large storage was required, it was decided to limit the number of simulations to 1,000. The estimated 2.5% significance points from 500 simulations were almost the same as those from 1,000 simulations. We also felt that 1,000 repetitions would yield fairly accurate estimates of the 1% significance points.) The results indicate good agreement between the simulated values and the extreme-value approximations. For example, when $n = 100$ and $k = 2$ (Model 3, Table 4), the 10% significance point from the approximation is 13.34, the simulated value for $\max_t Q_{k1(t)}$ is 13.25, and that for $\max_t Q_{k(t)}$ is 13.53. When $n = 100$, $k = 3$ (Model 4, Table 5), the 10% significance point from the approximation is 15.11, the simulated value for $\max_t Q_{k1(t)}$ is 14.85, and that for $\max_t Q_{k(t)}$ is 15.30. Similar or better agreement is seen for other significance levels and other sample sizes also.

In our experience, the plots of the statistics introduced in Section 2.1 are usually sufficient to spot the

Table 2. Significance Points for Q Statistics When $p = 1$ and $k = 1$—Model 3

		α			
n	*Type*	*.1*	*.05*	*.025*	*.01*
100	EV	10.71	12.05	13.38	15.13
	SQ_{11}	10.73	12.36	13.67	15.72
	SQ_1	10.88	12.44	14.03	15.73
200	EV	12.01	13.36	14.68	16.47
	SQ_{11}	11.85	13.42	14.68	16.12
	SQ_1	11.98	13.43	14.75	16.17

NOTE: EV indicates significance points from an extreme value distribution, SQ_{11} indicates simulated significance points for $\max_t Q_{11(t)}$, and SQ_1 indicates simulated significance points for $\max_t Q_{1(t)}$.

Table 3. Significance Points for Q Statistics When $p = 2$ and $k = 1$—Model 4

		α			
n	*Type*	*.1*	*.05*	*.025*	*.01*
100	EV	10.69	12.04	13.36	15.11
	SQ_{11}	10.70	12.20	13.61	15.51
	SQ_1	10.87	12.36	13.85	15.71
200	EV	12.00	13.35	14.67	16.45
	SQ_{11}	11.80	13.46	14.44	15.99
	SQ_1	11.99	13.56	14.53	16.11

NOTE: EV indicates significance points from an extreme value distribution, SQ_{11} indicates simulated significance points for $\max_t Q_{11(t)}$, and SQ_1 indicates simulated significance points for $\max_t Q_{1(t)}$.

outlier locations and types. The significance tests considered in this section are also helpful.

2.4 Overfitting

The diagnostic statistics in the previous sections were obtained under the assumption that the true order p of the process is known. In practice, however, this may not be the case. Then a common strategy is to fit a process of larger order. In this section, we look at the behavior of the Q statistics when an AR(p^*), $p^* > p$, is fit to the data.

Suppose that $\mathbf{Z}' = (z_{p+1}, z_{p+2}, \ldots, z_n)$,

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{C} \end{bmatrix},$$

and

$$\mathbf{V} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{C} \end{bmatrix},$$

where

$$\mathbf{B} = \begin{bmatrix} z_{p^*} & z_{p^*-1} & \cdots & z_{p^*-p+1} \\ z_{p^*+1} & z_{p^*} & \cdots & z_{p^*-p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ z_{n-1} & z_{n-2} & \cdots & z_{n-p} \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} z_{p^*-p} & \cdots & z_1 \\ z_{p^*-p+1} & \cdots & z_2 \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ z_{n-p-1} & \cdots & z_{n-p^*} \end{bmatrix},$$

and

$$\mathbf{A} = \begin{bmatrix} z_p & \cdots & z_1 \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ z_{p^*-1} & \cdots & z_{p^*-p} \end{bmatrix}.$$

Table 4. Significance Points for Q Statistics When p = 1 and k = 2—Model 3

		α			
n	*Type*	*.1*	*.05*	*.025*	*.01*
100	EV	13.34	14.83	16.26	18.12
	SQ_{k1}	13.25	14.49	15.98	17.71
	SQ_k	13.53	14.68	16.34	18.04
200	EV	14.80	16.25	17.67	19.51
	SQ_{k1}	14.38	15.79	17.00	18.99
	SQ_k	14.53	15.83	17.22	19.31

NOTE: EV indicates significance points from an extreme value distribution, SQ_{k1} indicates simulated significance points for $\max_t Q_{k1(t)}$, and SQ_k indicates simulated significance points for $\max_t Q_{k1(t)}$.

Table 5. Significance Points for Q Statistics When p = 2 and k = 3—Model 4

		α			
n	*Type*	*.1*	*.05*	*.025*	*.01*
100	EV	15.11	16.73	18.25	20.21
	SQ_{k1}	14.85	16.24	17.95	19.69
	SQ_k	15.30	16.91	18.57	20.67
200	EV	16.70	18.38	19.81	21.85
	SQ_{k1}	16.16	17.73	18.92	21.12
	SQ_k	16.44	18.09	19.26	21.39

NOTE: EV indicates significance points from an extreme value distribution, SQ_{k1} indicates simulated significance points for $\max_t Q_{k1(t)}$, and SQ_k indicates simulated significance points for $\max_t Q_{k(t)}$.

Then the CLS estimate of $\phi = (\phi_1, \ldots, \phi_p)$ is as given in (2.3), and the CLS estimate of $\boldsymbol{\phi}^* = (\phi_1, \ldots, \phi_p \mid \phi_{p+1}, \ldots, \phi_{p^*})' = (\boldsymbol{\phi}', \boldsymbol{\phi}_2')$ is given by

$$\hat{\boldsymbol{\phi}}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Z}. \qquad (2.17)$$

Now let $\hat{\boldsymbol{\phi}}'_{(N)} = (\hat{\boldsymbol{\phi}}' \mid \mathbf{0}')$, where $\mathbf{0}$ is a $(p^* - p) \times 1$ vector of zeros. Then it can be shown that $\hat{\boldsymbol{\phi}}^* = \hat{\boldsymbol{\phi}}_{(N)} - (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{V}'\mathbf{e}$, where $\mathbf{e} = \mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\phi}}$, the residual vector from the true model. Moreover, it follows that $\hat{z}_t^* = \hat{z}_t - \mathbf{x}_t^{*\prime}(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{V}'\mathbf{e}$ and $e_t^* = e_t + \mathbf{x}_t^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{e}$ ($t > p^*$), where $\hat{z}_t^*$ is the fitted value and e_t^* is the residual corresponding to the estimate $\hat{\boldsymbol{\phi}}^*$, and $\mathbf{x}_t^{*\prime} = (z_{t-1}, \ldots, z_{t-p}^*)$. Hence it can be shown (see Chuang 1987) that $e_t^* = e_t + O_p(n^{-1/2})$ for $t > p^*$. Thus we expect the residuals from the true and overfitted models to behave the same way for $t > p^*$.

The patterns of the Q statistics for the overfitted case depends on e_t^* and H_{22}^* (or h_{tt}^* when approximations are used). Hence Q_k, Q_{k1}, and Q_{k2} have basically the same pattern as in Table 1, with p replaced by p^*.

3. MODEL-BUILDING STRATEGY

In Section 2 we discussed the patterns of the diagnostic statistics for an AR(p) model and indicated that these patterns can be used to detect and identify outlier types even when the model is overfitted. In general, the process may not be autoregressive. Suppose that the true process is ARMA(p, q) as given in (1.1). Often such processes can be approximated by an AR($p + q$) model (see Box and Jenkins 1976, p. 282). In practice, for outlier detection, we found this to be a good approximation. Our model-building strategy then starts with the fitting of a sufficiently large AR process. We propose the following model-building procedure, based on the outlier detection methods discussed in the previous sections.

Step 1. Use any model-selection technique to

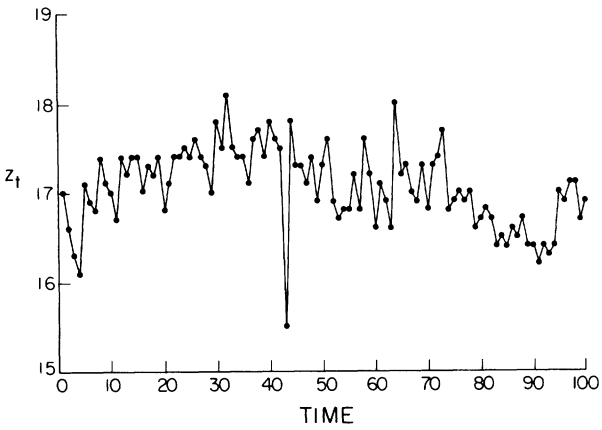


Figure 3. Chemical Process Concentration Reading (Ser. A) With AO Outlier Introduced at t = 43.

identify a tentative order (p', q'), which may not be the true order (p, q). Choose a $p^* > p' + q'$.

Step 2: Outlier Detection. Estimate $\boldsymbol{\phi}^*$ by the CLS method and compute Q_k (and/or Q_{k2}) for $k = 1$ and $k = p^* + 1$. Determine the outlier and its type based on the plots of Q_k (and/or Q_{k1}, Q_{k2}). The significance tests based on the maximum of these statistics may also be used. Go to Step 4 if there are no outliers; otherwise continue to Step 3. (If an outlier is detected, then it is very important to investigate the reasons for its occurrence.)

Step 3: Cleaning the Series. Let T be the position of the outlier identified in Step 2. If the outlier type is AO, then delete equations $(T - p^*)$ to T from (2.2) to obtain the estimate, $\tilde{\boldsymbol{\phi}}^*$. We now adjust the Tth observation, using the predictive mean of z_T conditional on all other observations, $E(z_T \mid z_t, t \neq T)$; that is, we replace z_t by

$$\begin{aligned} \tilde{z}_t &= z_t, & t &\neq T \\ &= \sum_{j=1}^{p^*} \tilde{\eta}_j(z_{t-j} + z_{t+j}), & t &= T, \end{aligned} \tag{3.1}$$

where $\tilde{\eta}_j = (\tilde{\phi}_j^* - \sum_{i=1}^{p^*-j} \tilde{\phi}_i^* \tilde{\phi}_{i+j}^*)/(1 + \sum_{i=1}^{p^*} \tilde{\phi}_i^{*2})$ $(j = 1, \ldots, p^*)$ (see Abraham and Box 1979).

On the other hand, if the outlier is of AI type, delete the Tth equation from (2.2) to estimate $\boldsymbol{\phi}^*$, and adjust the observations as follows:

$$\begin{aligned} \tilde{z} &= z_t, & t &< T \\ &= z_t - \tilde{e}_t^*, & t &= T \\ &= z_t - \tilde{\psi}_{t-T}^* \tilde{e}_T^*, & t &> T, \end{aligned} \tag{3.2}$$

where $\tilde{e}_t^*$ is the residual corresponding to the estimate $\tilde{\boldsymbol{\phi}}^*$ and $\tilde{\psi}_j^*$ is the coefficient of B^j in $1 - \tilde{\psi}_1^* B - \tilde{\psi}_2^* B^2 - \cdots = (1 - \tilde{\phi}_1^* B - \cdots - \tilde{\phi}_{p^*}^* B^{p^*})^{-1}$.

With the cleaned series, we go to Step 2 and repeat Steps 2 to 3 until no further outliers are found.

Step 4: Specification. Use the cleaned series in the last iteration to specify a tentative model. This model is then estimated using maximum likelihood, and the iterative strategy of model building discussed by Box and Jenkins (1976) is adopted.

Example

We now illustrate the four-step model-building procedure using the first 100 observations from Series A (chemical process concentration readings) of Box and Jenkins (1976), where the model employed by the authors is ARMA(1, 1). We introduce an AO outlier at $t = 43$ by subtracting 1 unit from the original observation. The series is presented in Figure 3 and is analyzed with the mean ($\hat{\mu} = 17$) subtracted out. Tentative model selection suggests $p^* = 3$. Then the estimates, $\hat{\boldsymbol{\phi}}^* = (.228, .263, .140)'$, are computed from Equation (2.17). The plots of $Q_{1(t)}$ and $Q_{4(t)}$ are shown in Figure 4 and they indicate that the observations 43 and 64 may be discrepant. Now $\max_t Q_{1(t)}/\hat{\sigma}^2 = Q_{1(43)}/\hat{\sigma}^2 = 22.72$ and $\Pr(\max_t Q_{1(t)}/\hat{\sigma}^2 > 22.72) < .01$. The patterns of the statistics suggest that the 43rd observation is an AO outlier. Deletion

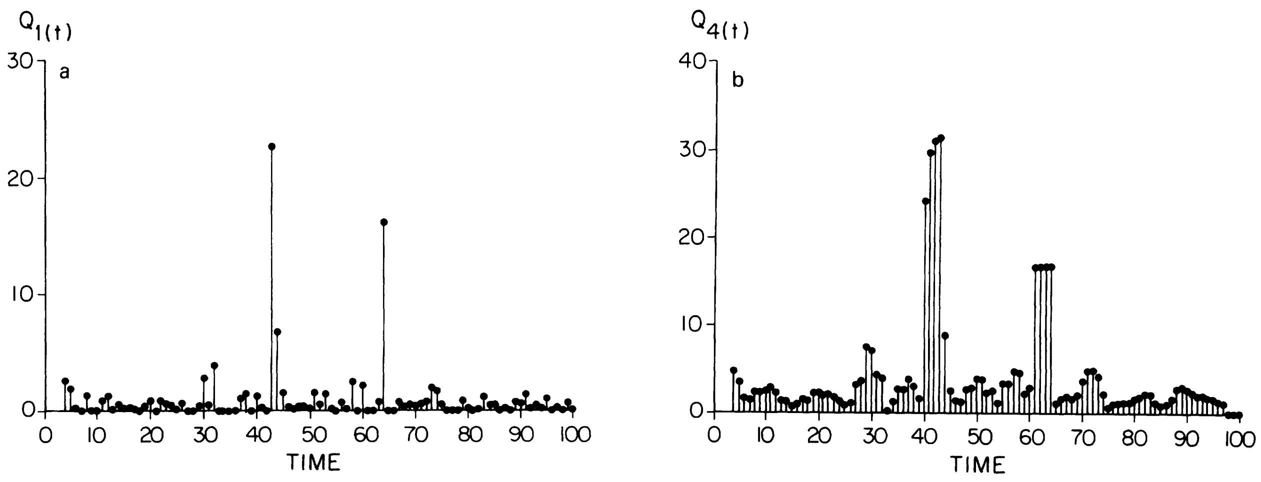
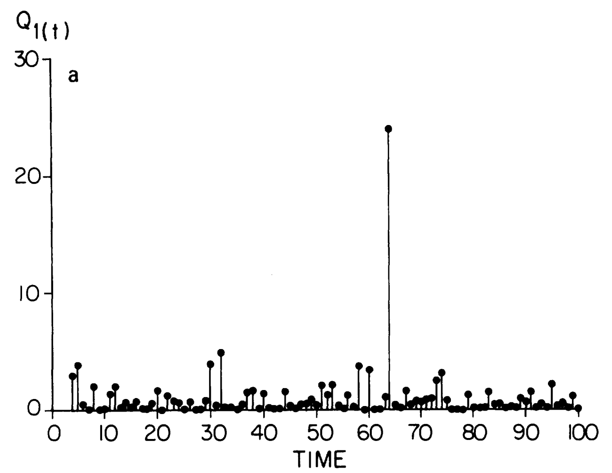


Figure 4. $Q_{1(t)}$ (a) and $Q_{4(t)}$ (b) Versus t, First Iteration (Ser. A).



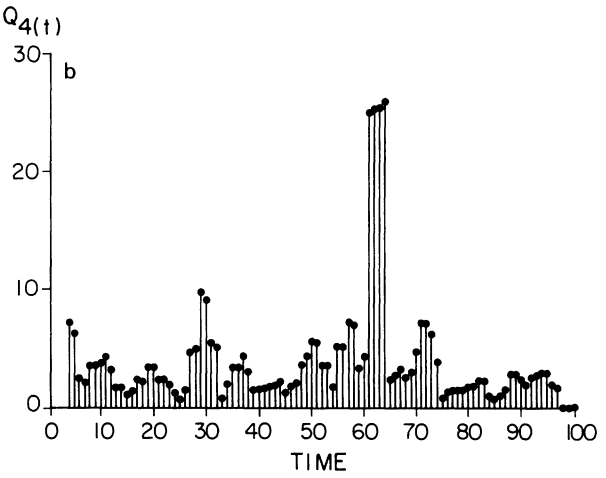


Figure 5. $Q_{1(t)}$ (a) and $Q_{4(t)}$ (b) Versus t, Second Iteration (Ser. A).

of equations corresponding to Z_{43}, Z_{44}, Z_{45}, and Z_{46} from (2.2) leads to $\tilde{\boldsymbol{\phi}}^* = (.326, .271, .145)'$ and $\tilde{y}_{43} = .515$. Then the second iteration with the cleaned series gives $\hat{\phi}^* = (.325, .271, .155)'$. It now appears that Q_1 has a single large value at $t = 64$, $Q_{1(64)}/\hat{\sigma}^2 = 23.90$, indicating the presence of an AI outlier (see Fig. 5). After deleting the equation for Z_{63} from (2.2) and estimating the parameters, the series is cleaned. Now the third iteration leads to small values for $\max_t Q_{1(t)}$ and $\max_t Q_{4(t)}$. Hence no more outliers are suspected and the iteration is terminated. Further model specification suggests an ARMA(1, 1) process and the maximum likelihood estimates are $\hat{\phi} = .94$, $\hat{\theta} = .58$, and $\hat{\sigma}^2 = .087$. These compare favorably with the values reported by Box and Jenkins (1976), $\hat{\phi} = .92$, $\hat{\theta} = .58$, and $\hat{\sigma} = .097$.

4. SUMMARY AND CONCLUSIONS

We have considered some statistics to detect outliers in time series. Moreover, we have considered the question of specifying the outlier type, using deletion of a certain number of observations from Figure 2. We have also found that the pattern of the statistics is valid even when the models are overfitted. Based on these, we have proposed a strategy for model building in the presence of outliers and illustrated this procedure with an example. The proposed strategy is simple and intuitively appealing, and it seems to work reasonably well in a number of examples. It easily can be made a routine part of existing time series softwares.

ACKNOWLEDGMENT

Abraham was supported in part by a grant from the Natural Science and Engineering Research Council of Canada.

[*Received February 1987. Revised May 1988.*]

REFERENCES

Abraham, B., and Box, G. E. P. (1979), "Bayesian Analysis of Some Outlier Problems in Time Series," *Biometrika*, 66, 229–236.

Anderson, T. W. (1971), *The Statistical Analysis in Time Series*, New York: John Wiley.

Andrews, D. F., and Pregibon, D. (1980), "Finding the Outliers That Matter," *Journal of the Royal Statistical Society*, Ser. B, 40, 87–93.

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

Chang, I., and Tiao, G. C. (1983), "*Estimation of Time Series Parameters in the Presence of Outliers*," Technical Report 8, University of Chicago, Statistics Reseach Center, Graduate School of Business.

Chuang, A. (1987), "Outliers in Time Series," unpublished manuscript, University of Waterloo, Dept. of Statistics and Actuarial Science.

Cook, R. D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman & Hall.

Draper, N. R., and John, J. A. (1981), "Influential Observations and Outliers in Regression," *Technometrics*, 23, 21–26.

Fox, A. J. (1972), "Outliers in Time Series," *Journal of the Royal Statistical Society*, Ser. B, 3, 350–363.

Martin, R. D. (1981), "Robust Methods for Time Series," in *Applied Time Series*, ed. D. F. Findley, New York: Academic Press, pp. 683–759.

Martin, R. D., and Yohai, V. J. (1986), "Influence Functionals for Time Series," *The Annals of Statistics*, 14, 781–818.

Muirhead, C. R. (1986), "Distinguishing Outlier Types in Time Series," *Journal of the Royal Statistical Society*, Ser. 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. K. Puri, J. P. Vilaplana, and W. Wertz, New York: John Wiley, pp. 109–118.

Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association*, 81, 132–141.