Detecting over-influential observations in time series

BY BERNARD LEFRANÇOIS

Industry Measures and Analysis Division, Statistics Canada, Ottawa, Ontario K1A 0T6, Canada

SUMMARY

The purpose of this paper is to develop a tool for identifying over-influential observations in time series when they are viewed in the time domain. We present a method for obtaining various measures of influence for the autocorrelation function, as well as thresholds for declaring an observation over-influential. An example of the use of these thresholds is also presented.

Some key words: Bessel function; Influential observation; Time series.

1. INTRODUCTION

For a weakly stationary and regular discrete time process \{X_t\}, the autocorrelation function, defined as \( \rho_k = \gamma_k / \gamma_0 \), where \( \gamma_k = E{(X_t - \mu)(X_{t+k} - \mu)} \) is the \( k \)th autocovariance and \( \mu = E(X_t) \) is the mean of the process, is a fundamental quantity in time series analysis. Usually, each observation in a realization contributes to the evaluation of the sample autocorrelation function. The contribution of one observation, however, may be discordant to the point of sensibly determining the value of an autocorrelation coefficient, or even the whole function. Such an observation clearly needs to be detected, and is said to be over-influential. It is important to specify for which statistic an observation is over-influential, for it may be so for some, but not for others.

Specific measures of influence for the autocorrelation function have already been suggested by Chernick, Downing & Pike (1982), Lattin (1983) and Li & Hui (1987). However, almost all were developed without recourse to a general appropriate theory, and without critical values for declaring an observation as over influential or not. Only Chernick et al. (1982) mentioned a threshold, but it does not correspond to the distribution theory of their suggested measure.

In the present paper we develop a tool for identifying over-influential observations for the sample autocorrelation function of a regular, stationary Gaussian time series. We consider only the case of at most one over-influential observation. In § 2, we obtain three measures of the influence of one observation on a specific autocorrelation. A comparison of these measures with those previously proposed leads us to adopt the sample measure of influence. In § 3, it is extended to measure the influence of one observation on the whole sample autocorrelation function. Some of its distributional properties are derived to obtain thresholds for declaring an observation as over-influential.

In § 4, an example is provided comparing the application of this measure of influence and its thresholds with an outlier detection technique. The example shows its usefulness in early detection of over-influential observations, which may also be outliers. Finally, we conclude with recommendations for further research.
2. INFLUENTIAL OBSERVATIONS IN TIME SERIES

For samples of independent identically distributed random variables, Hampel's (1974) influence function measures the effect on the value of a statistic of an infinitesimal change in the weight given to an observation by the theoretical distribution function.

A finite-sample version of the influence function is called a measure of influence. The three most frequently used are the empirical added influence, \( \text{EIC}_i \), the empirical deleted influence, \( \text{EIC}(i) \), and the sample measure of influence, \( \text{sic}_i \). These measures are usually obtained as approximations to the limit defining Hampel's influence function; see for example Cook & Weisberg (1982, §3.4). There is however an interesting relationship linking these three measures which can be shown to hold for any statistic expressible as a functional of the empirical distribution function. If \( w_i/n \) is the weight given to the \( i \)th observation by the empirical distribution function and \( T \) is the statistic of interest, then

\[
(n-1)\text{EIC}_i = \frac{\partial T}{\partial w_i}(1), \quad n\text{EIC}(i) = \frac{\partial T}{\partial w_i}(0), \quad \text{sic}_i = (n-1) \int_0^1 \frac{\partial T}{\partial w_i} \, dw_i. \tag{2.1}
\]

This provides both an interpretation and a means for computing these measures of influence in more complex situations. For instance, they cannot be obtained for time series using either Hampel's (1974) influence function or Martin & Yohai's (1986) generalization without making special assumptions on the nature of the dependency between the observations, as given by Li & Hui (1987) for example.

Hence, we used (2.1) directly to obtain the three main measures of the influence of the \( i \)th observation on the \( k \)th sample autocorrelation coefficient. Let \( n \) denote the length of the observed process, and let \( w_1, \ldots, w_n \) be weights such that \( w_t \in [0,1] \) and \( w_t = 1 \) for \( t \neq i \). Then the sample estimators of \( \mu \), \( \gamma_k \) and \( \rho_k \) can be respectively written as

\[
\bar{X}(w_i) = (n-1+w_i)^{-1} \sum_{t=1}^n w_t X_t, \quad c_k(w_i) = (n-1+w_i)^{-1} \sum_{t=1}^{n-k} w_t w_{t+k} \{X_t - \bar{X}(w_i)\} \{X_{t+k} - \bar{X}(w_i)\}, \quad r_k(w_i) = c_k(w_i)/c_0(w_i).
\]

Notice that \( \bar{X}(1) = \bar{X} \), \( c_k(1) = c_k \) and \( r_k(1) = r_k \) are the usual full-sample estimators of \( \mu \), \( \gamma_k \) and \( \rho_k \) respectively, while \( \bar{X}(0) = \bar{X}(i) \), \( c_k(0) = c_k(i) \) and \( r_k(0) = r_k(i) \) are their most natural estimators when the \( i \)th observation is missing at random (Dunsmuir, 1984).

Then, writing \( Y_t = \{X_t - \bar{X}(i)\}/c_0(i)^3 \) and \( Z_t = \{X_t - \bar{X}\}/c_0(i)^3 \) for \( t = 1, \ldots, n \), and 0 otherwise, and using (2.1) and keeping only the terms of highest order relative to the sample size \( n \), the three measures are

\[
\text{EIC}_{i,k} = n^{-1}(n-1)^{-1}(Z_i Z_{i+k} + Z_i Z_{i-k} - 2r_k Z_i^2) + o_p(n^{-3}), \tag{2.2}
\]
\[
\text{EIC}_k(i) = (n-1)^2(Y_i Y_{i+k} + Y_i Y_{i-k}) + o_p(n^{-3}), \tag{2.3}
\]
\[
\text{sic}_{i,k} = (1 - Z_i^2/(n-1))^{-1}(Z_i Z_{i+k} + Z_i Z_{i-k} - r_k Z_i^2) + o_p(1). \tag{2.4}
\]

Exact expressions are easily obtained, but the above are much easier to work with. Their derivation from (2.1) is essentially new, and the empirical measure of deleted influence \( \text{EIC}_k(i) \) does not appear to have been derived before.

Comparing now (2.2) to (2.4) with previous work, Chernick et al. (1982) suggested using \( C_{i,k} = Z_i Z_{i+k} - \frac{1}{2}(Z_i^2 + Z_{i+k}^2) r_k \), which was obtained by analogy with the influence function for the correlation coefficient in a bivariate normal distribution. Its major shortcoming is that it does not consider the interaction of \( Z \) with \( Z_{i-k} \) in the computation of \( r_k \), unlike any of the above measures.
Another suggestion is that of Li & Hui (1987), who obtained a measure of influence by assuming that the observed time series is a \( p \)th order autoregressive process. The structure of the resulting measure is, however, identical to that of \( C_k \). Finally, Lattin (1983) suggested using the measure of sample influence, but for \( r_k' = (n - k)^{-1} n r_k \), which is not the commonly used estimator of \( p_k \).

Now we follow Cook (1986) in selecting the measure of sample influence \( \text{sic}_{ik} \) in preference to the others as it is a more appropriate summary of the whole curve \( \partial r_k / \partial w_i \) than \( \text{EIC}_{ik} \) or \( \text{EIC}_k(i) \). Furthermore, \( \text{sic}_{ik} \) has a 'leave-one-out' interpretation, as well as being the product of a measure of 'outlyingness' \( (\bar{Z}_{i+k} + \bar{Z}_{i-k} - r_k \bar{Z}_i) \) with a measure of 'leverage' \( (\bar{Z}_i) \), where \( \bar{Z}_i = n^i(X_i - \bar{X}) / \{(n - 1)c_0(i)\}^{1/2} \) is a multiple of the externally standardized observed time series. The structure of \( \text{sic}_{ik} \) is thus similar to that of Cook's (1977) distance.

3. Identifying Influential Observations in Time Series

3-1. General

In practice, an observation's influence can be measured both on a specific autocorrelation coefficient, and on the whole autocorrelation function. In previous work by Chernick et al. (1982), Lattin (1983) and Li & Hui (1987), the reduction used to measure the influence of one observation on the whole function was obtained by taking the sum of the square of the elements of a 'clothes-pin' in the table formed by the influence of each observation on each coefficient, as shown in Table 1. This reduction leads to an unduly complicated distributional theory, and mixes the influence theoretically attributable to one observation with that of others. We suggest a different reduction in § 3-1.

![Table 1. The elements, shown by crosses, of a 'clothes-pin'](image)

Whether we measure the influence of one observation on a single autocorrelation coefficient or on the whole function, in each case we wish to determine if this influence is discordant compared to the influence of the other observations, and also determine whether the largest observed influence is discordant. In §§3-1 and 3-2, we derive the distributional theory required for obtaining asymptotic thresholds for each case.

3-2. Testing the Influence of a Specific Observation

From (2-4), it is seen that \( \text{sic}_{ik} \) depends on \( r_k \) as well as being a ratio of dependent random variables. Hence, we can expect its finite-sample density to be quite complex.
We thus used its asymptotic equivalent whose properties are more readily derivable:

\[ S_{ik} = \lim_{n \to \infty} \text{sic}_{ik} = \tilde{X}_i (\tilde{X}_{i+k} + \tilde{X}_{i-k} - \rho_k \tilde{X}_i), \]

where \( \tilde{X}_i = (X_i - \mu) / \gamma_i^2 \) for \( t = 1, \ldots, n \), and 0 otherwise.

Under the hypothesis that the observed time series is Gaussian, the density of \( S_{ik} \) is that of the product of two Gaussian random variables, which was obtained by Craig (1936). Thus, asymptotic critical values for testing the discordancy of \( \text{sic}_{ik} \) can be obtained, but they depend on the autocorrelation function. However, when the transformation suggested below for measuring the influence of one observation on the whole function is applied to \( \text{sic}_{ik} \) singly, it yields a statistic whose asymptotic distribution is independent of this function. Thus, asymptotic critical values for both tests can be readily tabulated.

For fixed \( L \), the length of the sample autocorrelation function on which the influence of the \( i \)th observation is measured, and \( n \), we define

\[ h(i) = \min \{i, n - i, L\}, \quad J_{h(i)} = \text{diag} (1_{h(i)}, 0_{L-h(i)}), \]

where 1\(_m\) denotes the \( m \times 1 \) column vector of ones, and 0\(_m\) denotes the \( m \times 1 \) column vector of zeros.

Let \( T_{k,m} = (\rho_k, \rho_{k+1}, \ldots, \rho_m) \), \( T_{k,m} \) be a \( m \times m \) Toeplitz matrix with \((p,q)\)th element \( \rho_{k+p+q} \) for \( p, q = 1, \ldots, m \), and \( H_{k,m} \) be a \( m \times m \) Hankel matrix with \((p,q)\)th element \( \rho_{k+p+q} \) for \( p, q = 1, \ldots, m \). Finally, we also define

\[ \tilde{T}_{k,m} = T_{k,m} - \rho_k R_{1m-k+1} R_{1m-k+1}^t, \quad \tilde{H}_{k,m} = H_{k,m} - \rho_k R_{1m-k+1} R_{1m-k+1}^t. \]

Now, the asymptotic influence \( S_{ix} = (S_{1x}, \ldots, S_{Lx})' \) of the \( i \)th observation on the autocorrelation function is the product of \( \tilde{X}_i \) with

\[ \xi_i = \{\tilde{X}_{i+k} + \tilde{X}_{i-k} - \rho_k \tilde{X}_i; k = 1, \ldots, L\}. \]

Under the hypothesis that the observed time series is Gaussian, the distribution of \( \xi_i | \tilde{X}_i \) is \( N(\mu_i, \psi_i) \), where

\[ \mu_i = J_{h(i)} R_{1m-L} X_{i}, \quad \psi_i = \tilde{T}_{0m-L} + J_{h(i)} T_{0m-L} J_{h(i)} + \tilde{H}_{2m-L} J_{h(i)} + J_{h(i)} \tilde{H}_{2m-L}. \]

Let \( Q_i = \tilde{X}_i^2 (\xi_i - \mu_i)' \psi_i^{-1} (\xi_i - \mu_i) \). Then, the distribution of \( Q_i | \tilde{X}_i \) is Gamma \((\frac{1}{2} L, 2 \tilde{X}_i^2)\)
where the density of a Gamma \((\alpha, \beta)\) distribution is \( x^{\alpha-1} e^{-x} \beta^{-\alpha} \{\Gamma(\alpha)\}^{-1} \) for \( x > 0 \).

Since \( \tilde{X}_i^2 \sim \Gamma(\frac{1}{2}, 2) \), the unconditional density of \( Q_i \) is therefore

\[ f(q; L) = q^{(L-3)} K_{(L-1)} (q^4)/(2^{(L-1)} \pi^{1/2} \{\Gamma(\frac{1}{2} L)\}, \]

where \( K_v(x) \) is the modified Bessel function of the second kind of order \( v \):

\[ K_v(x) = \int_0^\infty \exp \{-x \cosh (t)\} \cosh (vt) \, dt. \]

We designate these distributions by \( W_i \); that is \( X \sim W_i \) if and only if the density of \( X \) is (3.1).

But \( Q_i \) is not a statistic as it depends on the unknown autocorrelation function. In practice, \( S_{ix} \) will be replaced by \( \text{sic}_{ix} \), \( \tilde{X}_i \) by \( Z_i \), defined in § 2, while \( \mu_i \) and \( \psi_i \) will be replaced by the consistent estimators \( \hat{\mu}_i \) and \( \hat{\psi}_i \) obtained by estimating \( \rho_k \) with \( r_k \). Note,
however, that $L$ must be less than $\frac{1}{2}n$ to ensure the consistency of $\hat{\psi}_i$. Thus for fixed $L < \frac{1}{2}n$, $W_L$ is the asymptotic distribution of

$$QIC_L = (\text{sic}_i - Z_i \hat{\mu}_i) / (\text{sic}_i - Z_i \hat{\mu}_i).$$

A major advantage of QIC, as a measure of influence is that the asymptotic distribution of QIC only depends on the length, $L$, of the autocorrelation function; asymptotic critical values can thus be readily tabulated as in Table 2. Another is that it is easily modified to measure the influence on any subset of coefficients, in particular that on a single one.

### Table 2. Some percentage points of the $W_L$ distribution

<table>
<thead>
<tr>
<th>$L$</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.544</td>
<td>4.761</td>
<td>7.764</td>
<td>12.990</td>
</tr>
<tr>
<td>3</td>
<td>8.211</td>
<td>12.984</td>
<td>19.646</td>
<td>28.695</td>
</tr>
<tr>
<td>4</td>
<td>10.705</td>
<td>16.917</td>
<td>24.323</td>
<td>35.883</td>
</tr>
<tr>
<td>5</td>
<td>13.394</td>
<td>20.812</td>
<td>29.517</td>
<td>42.911</td>
</tr>
<tr>
<td>6</td>
<td>16.083</td>
<td>24.687</td>
<td>34.663</td>
<td>49.841</td>
</tr>
<tr>
<td>7</td>
<td>18.773</td>
<td>28.550</td>
<td>39.777</td>
<td>56.707</td>
</tr>
<tr>
<td>8</td>
<td>21.464</td>
<td>32.405</td>
<td>44.870</td>
<td>63.527</td>
</tr>
<tr>
<td>9</td>
<td>24.156</td>
<td>36.256</td>
<td>49.949</td>
<td>70.314</td>
</tr>
<tr>
<td>10</td>
<td>26.849</td>
<td>40.103</td>
<td>55.017</td>
<td>77.076</td>
</tr>
<tr>
<td>11</td>
<td>29.544</td>
<td>43.949</td>
<td>60.076</td>
<td>83.818</td>
</tr>
</tbody>
</table>

### 3.3. Testing the most influential observation

The largest observed influence, $QIC(n)$, is an appropriate statistic for testing that no single observation has a discordant influence. Since we believe its exact distribution to be intractable, we will obtain bounds on its asymptotic critical values using inequalities of the Bonferroni type. Note that testing the largest influence on a single autocorrelation coefficient is a particular case of the following.

Let $A_i$ denote the event $\{Q_i > c\}$. Then the first Bonferroni lower bound and Hunter’s (1976) upper bound yield:

$$\text{LB}(c) \leq \text{pr}(Q(n) > c) = \text{UB}(c),$$

where

$$\text{LB}(c) = \sum_{i=1}^{n} \text{pr}(A_i) - \sum_{i<j} \text{pr}(A_i \cap A_j), \quad \text{UB}(c) = \sum_{i=1}^{n} \text{pr}(A_i) - \sum_{i<j} \text{pr}(A_i \cap A_j),$$

with $G$ being a maximum spanning tree among the events $A_i$, for $i = 1, \ldots, n$, which are connected if and only if $\text{pr}(A_i \cap A_j) > 0$.

For a strictly stationary process, both bounds simplify in an interesting manner since $\text{pr}(A_i \cap A_j)$ depends on $i$ and $j$ only through $|i-j|$: $\text{pr}(A_i \cap A_j) = p_{|i-j|}(c)$, say. The lower bound becomes

$$\text{LB}(c) = n \left\{ p_0(c) - \sum_{k=1}^{n-1} \left( 1 - \frac{k}{n} \right) p_k(c) \right\}.$$  

To simplify Hunter’s upper bound, let $p_k(c)$ denote the ordered probabilities $p_k(c)$ for $k = 1, \ldots, n-1$, that is $p_{n-1}(c) \geq \ldots \geq p_1(c)$, and let $\sigma$ denote the induced permutation of $\{1, \ldots, n-1\}$, that is $\sigma$ is such that $p_{\sigma(k)}(c) = p_k(c)$. Then, the maximum spanning
tree is found to be the graph formed by taking the $N$ largest probabilities $p_j(c)$ such that

$$1 = \frac{1}{n} + \sum_{i=1}^{N} \left\{ 1 - \frac{\sigma(n-i)}{n} \right\}.$$

Thus, Hunter's upper bound in (3-2) becomes:

$$\text{UB}(c) = n \left[ p_0(c) - \sum_{k=1}^{N} \left\{ 1 - \frac{\sigma(n-k)}{n} \right\} p_{n-k}(c) \right]. \quad (3.4)$$

There are however three difficulties in computing $\text{UB}(c)$ and $\text{LB}(c)$ in our particular context. First, the stochastic process $\{Q_i: i = 1, \ldots, n\}$ is not necessarily strictly stationary, but the segment $\{Q_i: i = L+1, \ldots, n-L\}$ is. Since the effect of the nonstationary terms in the bounds (3.2) will certainly decrease as the observed length of the time series increases, the bounds (3.3) and (3.4) should nonetheless be adequate.

The second difficulty stems from the finite length of the observed times series, which does not permit the computation of $p_k(c)$ for $k$ up to $n-1$. If the time series is a moving average process of order $M$, $Q_{i+k}$ and $Q_{i-L}$ are independent whenever $k > M + 2L$. Since $L$ is usually such that the sample autocorrelation function contains all the salient features of the unobserved one, the bounds (3.3) and (3.4) can be computed under this hypothesis, with $M \geq L$, in particular $M = 4L$.

The last difficulty is the most important one of the three. We do not yet have an easily computable expression for $p_k(c)$ when $k = 1, \ldots, 4L$. As a first approximation we chose a bivariate log normal distribution, since the univariate log normal distribution provides a reasonable approximation to the distribution of $W_L$, the marginals of $p_k(c)$, as discussed below. Recall that a random variable $x$ is log normal $\text{LN}(\mu, \sigma^2)$ whenever $\log x$ is normal $\text{N}(\mu, \sigma^2)$. The logarithmic transformation decomposes $Q_i$ into a sum of independent log $\chi^2$ random variables, each of which can be approximated by Gaussian distributions (Bartlett & Kendall, 1946). Also, the shapes of the density and of the moment generating function of a $W_L$ random variable are similar to those of a log normal random variable. Finally, as shown in Table 3 for various values of $L$, the exact probability contents of the approximate upper percentage points obtained from $\text{LN}(\mu_L, \sigma^2_L)$, where $\mu_L = \frac{1}{2} \log \left( L^2 / (3(L+2)) \right)$, $\sigma^2_L = \log \left( 3(L+2) / L \right)$, are not too distant from the desired ones.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$\alpha = 0.1$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.001$</th>
<th>$\alpha = 0.0005$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0852</td>
<td>0.0364</td>
<td>0.0067</td>
<td>0.0010</td>
<td>0.0006</td>
</tr>
<tr>
<td>2</td>
<td>0.0811</td>
<td>0.0367</td>
<td>0.0075</td>
<td>0.0012</td>
<td>0.0007</td>
</tr>
<tr>
<td>3</td>
<td>0.0795</td>
<td>0.0368</td>
<td>0.0079</td>
<td>0.0013</td>
<td>0.0008</td>
</tr>
<tr>
<td>4</td>
<td>0.0784</td>
<td>0.0368</td>
<td>0.0081</td>
<td>0.0014</td>
<td>0.0009</td>
</tr>
<tr>
<td>5</td>
<td>0.0776</td>
<td>0.0368</td>
<td>0.0083</td>
<td>0.0015</td>
<td>0.0009</td>
</tr>
<tr>
<td>6</td>
<td>0.0769</td>
<td>0.0367</td>
<td>0.0085</td>
<td>0.0016</td>
<td>0.0010</td>
</tr>
<tr>
<td>7</td>
<td>0.0764</td>
<td>0.0367</td>
<td>0.0086</td>
<td>0.0016</td>
<td>0.0010</td>
</tr>
<tr>
<td>8</td>
<td>0.0759</td>
<td>0.0366</td>
<td>0.0087</td>
<td>0.0017</td>
<td>0.0011</td>
</tr>
<tr>
<td>9</td>
<td>0.0755</td>
<td>0.0365</td>
<td>0.0088</td>
<td>0.0017</td>
<td>0.0011</td>
</tr>
<tr>
<td>10</td>
<td>0.0751</td>
<td>0.0365</td>
<td>0.0089</td>
<td>0.0018</td>
<td>0.0011</td>
</tr>
<tr>
<td>12</td>
<td>0.0746</td>
<td>0.0364</td>
<td>0.0090</td>
<td>0.0018</td>
<td>0.0012</td>
</tr>
<tr>
<td>24</td>
<td>0.0728</td>
<td>0.0360</td>
<td>0.0094</td>
<td>0.0021</td>
<td>0.0014</td>
</tr>
<tr>
<td>36</td>
<td>0.0720</td>
<td>0.0358</td>
<td>0.0095</td>
<td>0.0022</td>
<td>0.0015</td>
</tr>
<tr>
<td>48</td>
<td>0.0716</td>
<td>0.0357</td>
<td>0.0096</td>
<td>0.0023</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Table 3. Value of $\text{pr}(W_L > c_a)$, where $c_a$ is such that $\text{pr}(\text{LN}(\mu_L, \sigma^2_L) > c_a) = \alpha$.
Let \((u, v)\)' have a bivariate \(\text{Ln} (\mu, \theta)\) distribution with \(\mu = (\mu_u, \theta_v)\)' and

\[
\theta = \begin{pmatrix} \sigma_u^2 & \eta \sigma_u \sigma_v \\ \eta \sigma_u \sigma_v & \sigma_v^2 \end{pmatrix}.
\]

Equating the first and second moments of the joint distribution of \((Q_i, Q_{i+k})\)' with those of \((u, v)'\), we get

\[
\mu_u = \mu_v = \mu_L, \quad \sigma_u^2 = \sigma_v^2 = \sigma_L^2, \quad \eta_k = \log (1 + \tau_k / L^2) / \sigma_L^2,
\]

where \(\tau_k\) is the asymptotic covariance of \(QIC\), with \(QIC_{i+k}\):

\[
\tau_k = 2(1 + 2 \rho_k^2) \text{tr} (A_k^2 A_k) + 16 \rho_k W_k A_k W_k + 2 \rho_k^2 L^2 + 4 L W_k W_k + 4 (W_k^2 W_k)^2,
\]

with \(\Psi = 2(\tilde{T}_{0,L-1} + \tilde{H}_{2,L})\), \(D\) such that

\[
DD' = \Psi^{-1}, \quad W_k = D' V_k, \quad A_k = D' B_k D, \quad V_k = R_{k+1,k} + R_{k-1,k} - 2 \rho_k R_{1,k},
\]

\[
B_k = \tilde{T}_{k,k+L-1} + \tilde{T}_{k,k-L+1} + \tilde{H}_{k+k+2L} + \tilde{H}_{k-k-2L} - 2 \tau_k R_{1,k} - 2 R_{1,k} V_k.
\]

To test, at level \(1 - \alpha\), the most influential observation on the whole sample autocorrelation function from a time series of length \(n\), we compute \(C^*\) and \(C^\alpha\) such that \(\text{LB}(C^*) = \text{UB}(C^\alpha) = \alpha / n\), with \(\text{LB}\) and \(\text{UB}\) as given in (3-3) and (3-4) respectively, and where \(p_k(c)\) is approximated as described above. If \(QIC_{(n)} \geq C^*\) we would reject the hypothesis that no single observation has a discordant influence; if \(QIC_{(n)} \leq C^\alpha\) we would ‘accept’ the hypothesis; while for \(C^\alpha < QIC_{(n)} < C^*\) the test would be inconclusive. The tighter these bounds are, the more useful they will be. Their tightness may, however, depend on the particular time series.

4. Example

The example chosen was used by Chang, Tiao & Chen (1988) to demonstrate a method for detecting outlying observations in time series. In the following we will illustrate the identification of over-influential observations prior to the model-building process. We will compare the results of our analysis with those obtained by Chang et al. (1988), but we do not claim that measures of influence can detect outliers, since they were not developed for this purpose. This example shows how useful they can be, however, for the early detection of over-influential observations, which may also be outliers.

The data consist of 197 concentration readings from a chemical process, taken every 2 hours; it is Series A from Box & Jenkins (1976, p. 525). After taking one regular difference, required for identifying a model, we compute the influence of each observation on the first 20 autocorrelation coefficients. The influences are shown in Fig. 1, together with the approximate bounds (3-3) and (3-4) on the 95% critical value for the largest. These bounds are 290-3 and 324-5.

Figure 1 shows that observations numbered 44 and 64 are over-influential. Hence they are temporarily set aside, and the influences of the remaining observations are recomputed, as well as new bounds. These new bounds are 295-0 and 326-0. Again, the observation having the largest influence, observation 43, is almost significantly too large, having an influence of 249-0. No other over-influential observations were found.

Chang et al. (1988) found only two outliers in Series A: observation 43, categorized as an additive outlier, and observation 64, categorized as an innovation outlier. We found that observations 43, 44 and 64 were over-influential for the first difference of the data. Hence if observation 43 is an additive outlier, the first difference will ‘spread’ part of the discordancy to the next, observation 44. Therefore our results are consistent with theirs.
Influence measures could be derived for other statistics used in time series analysis. However, because of its importance, the tools developed herein for the autocorrelation function should usefully complement the outlier detection techniques relying on a prespecified model since their power may be severely reduced when over-influential observations are present.

The thresholds for declaring an observation as over-influential that we developed were obtained using a bivariate log normal approximation, which should be viewed as the first term of an Edgeworth-type expansion. This approximation could be studied further on its own but, since it enters into the computation of the bounds, it may be preferable to study further the coverage probability of the resulting bounds either theoretically or through simulation studies.

Now the lower Bonferroni bound may become negative when the autocorrelation function does not decay rapidly enough to zero. As reported by Schwager (1984), this usually occurs when there is a strong correlation between the events entering into the lower bound. There may be certain conditions, which remain to be found, under which the number of events in the lower bound could be reduced to account for the strong correlation.

**Acknowledgement**

This research was part of my M.Sc. thesis done at Carleton University under the supervision of Professor J. N. K. Rao to whom I am most grateful.

**Appendix**

*Computational details*

To compute the bounds (3·3) and (3·4), and the density (3·1), we need to compute the modified Bessel function of the second kind \( K_\nu(x) \) for integer orders \( \nu = 0, 1, 2, \ldots \), and for fractional
Over-influential observations in time series

orders $\nu = \frac{1}{2}, \frac{3}{2}, \ldots$. We also need integrals of the form

$$J_{\nu}(x) = \int_0^x y^\nu K_{\nu}(y) \, dy.$$  

For the functions $K_0(x)$, $K_1(x)$, $J_0(x)$ and $J_1(x)$ we used the Chebyshev polynomial approximations (Luke, 1969, pp. 338-43), while

$$K_{3/2}(x) = \left( \frac{\pi}{2x} \right)^{3/4} \exp \left( -x \right), \quad K_{5/2}(x) = K_{3/2}(x)(1 + x^{-1}),$$

and $J_{3/2}(x)$, $J_{5/2}(x)$ are obtained by integration. For the functions of higher order, we used the recurrence formula for $K_{\nu}(x)$,

$$xK_{\nu-1}(x) + xK_{\nu-1}(x) = 2xK_{\nu}(x).$$

REFERENCES


[Received September 1989. Revised August 1990]