Stochastic Hydrol. Hydraul. 3 (1989) 241-260

Stochastic Hydrology and Hydraulics

A practical method for outlier detection in autoregressive time series modelling

M. C. Hau

Dept. of Statistics, University of Wisconsin, Madison, Wisconsin, USA

H. Tong

Inst. of Mathematics, Cornwallis Building, University Canterbury, Kent CT2 7NF, U. K.

Abstract: A practical method is developed for outlier detection in autoregressive modelling. It has the interpretation of a Mahalanobis distance function and requires minimal additional computation once a model is fitted. It can be of use to detect both innovation outliers and additive outliers. Both simulated data and real data are used for illustration, including one data set from water resources.

Key words: Hat matrix, Mahalanobis distance, Additive outliers, Innovation outliers, Influential data, Autoregressive models, Threshold autoregression, Lake Huron.

1 Introduction

Autoregressive models hae been widely used in Stochastic Hydrology for many years. Thomas and Fiering (1962) originally proposed the use of first-order periodic autoregressive models for modeling mean monthly unregulated riverflow. Also, Yevjevich (1963) proposed the use of first order autoregressions for modeling mean annual unregulated river flow. Furthermore, Hipel and McLeod (1978) found that many types of annual geophysical time series could be modeled by simple autoregressive or autoregressive - moving average time series models. In this paper, methods of outlier detection in autoregressive models are discussed.

Although the word "*outlier*" appears frequently in the statistical literature, there does not seem to exist a generally accepted definition. In the estimation of a single parameter, e.g., the location parameter, from independent identically distributed (i.i.d.) observations, outliers are often understood to be "extreme" points in some sense. In this case, outliers can be very often detected by "eyes". Unfortunately, the situation with time series data seems to be more complex. "Outliers" in time series are not necessarily "outliers". Although not always fully justified, a widely used method for outlier detection in the context of time series is based on the residuals. This might stem from the fact that in, e.g., the Box and Jenkins' approach (1970), residuals are used for diagnostic check on the assumptions about the error term. In the following sections, we point out some limitation of using residuals for this purpose, especially for outlier detection. A *new approach*, which is not based on the residuals, is explored. Its emphasis is on the influence of data and our presentation is heuristic. The approach is based on the hat matrix technique, which in the present context admits an interesting interpretation as the Mahalanobis distance. The new method is illustrated with both real and simulated data, including one set of real data pertinent to hydrology.

2 Linear autoregressive model

We consider the linear Gaussian autoregressive model of order p ($AR(p)$):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \tag{2.1}$$

where ε_t's are i.i.d. and $\varepsilon_t \sim N(0,\sigma^2)$. Following Martin's notation (Martin, 1980), we let

$$z_t^T = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}),\ \phi^T = (\phi_1, \phi_2, \ldots, \phi_p).$$

Then (2.1) can be rewritten as

$$Y_t = z_t^T \phi + \varepsilon_t. \tag{2.2}$$

Suppose now we have n observations $Y_1, Y_2, \ldots, Y_n$. Then we have the following n equations

$$\mathbf{Y} = \Gamma\phi + \varepsilon \tag{2.3}$$

where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ is an $n\times 1$ observation vector, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is an $n\times 1$ vector of random errors, and

$$\Gamma = \begin{bmatrix} Y_0 & Y_{-1} & \ldots & Y_{-p+1} \\ Y_1 & Y_0 & \cdots & Y_{-p+2} \\ \cdot & & \cdots & \cdot \\ \cdot & & \cdots & \cdot \\ Y_{n-1} & Y_{n-2} & \cdots & Y_{n-p} \end{bmatrix} = \begin{bmatrix} z_1^T \\ z_1^T \\ \cdot \\ \cdot \\ z_n^T \end{bmatrix}$$

an $n\times p$ "design matrix".

$Y_i, i\le 0$, are assumed fixed, say at zero. The usual root condition is assumed throughout so that $\{Y_t\}$ is stationary. For the estimation of ϕ and σ^2, the conventional least squares procedure can be conveniently applied with well-known large sample properties due to Mann and Wald (1943).

We summarize first some well-known results as follows:

(i) The least square (L.S.) estimate $\hat{\phi}$ of ϕ is given by

$$\hat{\phi} = (\Gamma^T\Gamma)^{-1}\Gamma^T\mathbf{Y} \tag{2.5}$$

(ii) Let $\hat{\mathbf{Y}} = \Gamma\hat{\phi}$ be the fitted values. The least square fitted residuals (LSR), $\mathbf{r} = (r_1, r_2, \ldots, r_n)^T$, are defined by

$$\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}}. \tag{2.6}$$

Note that $\hat{\mathbf{Y}} = \Gamma\hat{\phi} = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T\mathbf{Y}$, i.e.

$$\hat{\mathbf{Y}} = H\mathbf{Y}, \tag{2.7}$$

where $H = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T$.

Here, $H = [h_{ij}]$ is known as the hat matrix in the context of outliers. Since

$$(I-H)\Gamma\phi = \Gamma\phi - \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T\Gamma\phi = 0$$

we have from (2.6), (2.7) that

$$\mathbf{r} = (I-H)\mathbf{Y} = (I-H)(\Gamma\phi+\varepsilon) = (I-H)\varepsilon \tag{2.8}$$

In addition, (2.8) and (2.9) can be rewritten in scalar forms viz.

$$r_t = (l-h_{tt})Y_t - \sum_{j\neq t} h_{tj}Y_j = (1-h_{tt})\varepsilon_t - \sum_{j\neq t} h_{tj}\varepsilon_j. \tag{2.9}$$

(iii) The L.S. estimate of σ^2 is given by

$$\hat{\sigma}_2 = \mathbf{r}^T\mathbf{r}/n \tag{2.10}$$

From now on, we denote the "Diagonal Element(s) of the Hat matrix" by DEH, and the t-the DEH, h_{tt}, by h_t, i.e.

$$h_t = z_t^T(\Gamma^T\Gamma)^{-1}z_t.$$

(iv) By stationarity of $\{Y_t\}$, it holds that as $n\to\infty$, $n^{-1}(\Gamma^T\Gamma)\xrightarrow{p}\Sigma$, where

$$\Sigma = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_{p-1} \\ Y_1 & Y_0 & \dots & Y_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{p-1} & Y_{p-2} & \dots & Y_0 \end{bmatrix}, \quad Y_s = cov(Y_t, Y_{t+s}), \quad a \ \ Toeplitz \ \ matrix.$$

Now, conditional on H being fixed, it holds that

$$var(\mathbf{r}) = (I-H)var(\varepsilon) = (I-H)\sigma^2. \tag{2.11}$$

In scalar form, equation (2.11) becomes

$$var(r_t) = (1-h_t)\sigma^2, \quad cov(r_i, r_j) = -h_{ij}\sigma^2.$$

3 Hat Matrix in linear autoregressive time series modeling

The hat matrix $H = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T$ has been receiving increasing attention in conventional regression analysis for a considerable length of time. See, e.g., Hoaglin and Welsch (1978), Cool and Weisberg (1979), Belsley, Kuh and Welsch (1980) and Cook and Weisberg (1982). Huber (1981) has discussed some particular aspects of the hat matrix in regression analysis. Research on the role of hat matrix in the context of time series seems to be lacking. We will see that the hat matrix has in fact particularly interesting properties in the case of autoregressive modeling. We summarize first some well-known properties of H as follows:

(*i*) H *is idempotent, i.e.,* $H^2 = H$, $H^T = H$.

(*ii*) $$trace(H) = rank(H) = p = AR \ \ order \tag{3.1}$$

(*iii*) $$H^2 = H \Rightarrow \forall t, j = 1, ..., n, \ \sum_j h_{tj}^2 = h_t \tag{3.2}$$

(*iv*) *From* (3.2), *we have* $$h_t^2 + \sum_{j\neq t} h_{tj}^2 = h_t. \tag{3.3}$$

$$h_t \geq h_t^2, \quad 0 \leq h_t \leq 1.$$

(*v*) *From* (3.3) *we have* $$\sum_{j\neq t} h_{tj}^2 \to 0 \ \ as \ \ h_t \to 1. \tag{3.5}$$

It follows that $h_{tj}\to 0 \, \forall_j \neq t$ *as* $h_t \to 1$.

3.1 Interpretations of hat matrix in linear autoregressive time series models

(1) Recall that the fitted values (sometimes also called the predictions) may be expressed as

$$\hat{\mathbf{Y}} = \Gamma\hat{\phi} = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T\mathbf{Y} = H\mathbf{Y} \tag{3.6}$$

In scalar form, we may write

$$\hat{Y}_t = h_t Y_t + \sum_{j \neq t} h_{tj} Y_j. \tag{3.7}$$

From (3.7) and Section 3(v), we know that if h_t is large (i.e. near 1), $\hat{Y}_t$ will be dominated by the term $h_t Y_t$. Therefore h_t may be interpreted as the amount of leverage or influence exerted on $\hat{Y}_t$ by Y_t.

We note that the relationship between the fitted values and the estimates $\hat{\phi}$ is given by

$$\hat{\mathbf{Y}} = \Gamma\hat{\phi} \text{ and } \hat{\phi} = (\Gamma^T\Gamma)^{-1}\Gamma^T\hat{\mathbf{Y}}$$

Hence, knowing $\hat{\mathbf{Y}}$ is equivalent to knowing $\hat{\phi}$. In our case, it is easier to interpret $\hat{\mathbf{Y}}$ than $\hat{\phi}$. Therefore we refer more often to $\hat{\mathbf{Y}}$.

(2) Differentiating (3.7) w.r.t. Y_t, we have

$$\frac{\partial \hat{Y}_t}{\partial Y_t} \doteq h_t, \tag{3.8}$$

which shows that h_t measures approximately the *relative change* of the fitted value $\hat{Y}_t$ when there is a small change in the observation value Y_t.

(3) Define $d_t = z_t^T \Sigma^{-1} z_t$, $t=1, 2, \ldots, n$. Then d_t is just the Mahalanobis distance between z_t and the zero vector (or the mean vector of z_t's in the general case). Note also that

$$nh_t = z_t^T \left(\frac{\Gamma^T\Gamma}{n}\right)^{-1} z_t \xrightarrow{p} d_t \quad as \quad n \to \infty. \tag{3.9}$$

Therefore nh_t has the interpretation as the Mahalanobis distance between z_t and the mean vector (zero vector in our case). Note that in conventional linear regression, h_t's do not have this explicit interpretation.

4 Limitations of LSR for diagnostic check and outlier detection

Since ε_t's are unobservable, LSR are often used as substitute. Although fitted residuals have been routinely used in the time series context and they possess many useful properties, it is our view that they also suffer from a number of limitations which we list as follows:

(1) By assumption, the errors, ε_t's, are independent random variables with zero mean and a common variance σ^2. For large n, the residuals r_t's also have zero mean, but, from (2.11), they tend to be correlated.

(2) Equation (2.8) shows clearly that $\mathbf{r}$ and ε are different, and the difference depends only on H. If all the h_{ij}'s are sufficiently small, $\mathbf{r}$ will serve as a reasonable substitute for ε. Otherwise, the usefulness of $\mathbf{r}$ may be limited. For example, supposing now that Y_t is an outlier, then at first sight ε_t could be expected to be large (in

absolute value). However, if h_t is also large, then by reference to (2.9) it is conceivable that r_t may be reduced to a small value. As a result, the outlier Y_t may go unnoticed if we examine only the LSR.

(3) When the AR order is roughly specified, then examination of LSR is not always as informative as we would like it to be. Specifically, let the true $AR(p)$ model be of the form

$$\mathbf{Y} = \Gamma_p \phi_p + \varepsilon^{(1)} \tag{4.1}$$

where $\Gamma_p = \Gamma$ as in (2.4), $\phi_p = (\phi_1, \phi_2, ..., \phi_p)^T$, and $\varepsilon^{(1)} = (\varepsilon_1^{(1)}, \varepsilon_2^{(1)}, \ldots, \varepsilon_n^{(1)})^T$.

Let Γ_{p-1} be the matrix of the first $p-1$ columns of Γ_p. Let $\mathbf{Y}_p = (Y_{-p+1}, Y_{-p+2}, ..., Y_{n-p})^T$, the p-the column of Γ_p, $\mathbf{b} = \mathbf{Y}_p \phi_p$, an $n \times 1$ vector $\phi_{p-1} = (\phi_1, \phi_2, \ldots, \phi_{p-1})^T$. Then we have

$$\mathbf{Y} = \Gamma_p \phi_p + \varepsilon^{(1)} = (\Gamma_{p-1} | \mathbf{Y}) \phi_p + \varepsilon^{(1)} = \Gamma_{p-1} \phi_{p-1} + \mathbf{Y}_p \phi_p + \varepsilon^{(1)} = \Gamma_{p-1} \phi_{p-1} + \mathbf{b} + \varepsilon^{(1)}. \tag{4.2}$$

Suppose, in fact, we have fitted an AR(p-1), i.e.,

$$\mathbf{Y} = \Gamma_{p-1} \phi_{p-1} + \varepsilon^{(2)} \quad \textit{where} \;\; \varepsilon^{(2)} = (\varepsilon_1^{(2)},\ \varepsilon_1^{(2)},\ ...,\ \varepsilon_n^{(2)})^T. \tag{4.3}$$

Then, the LSR of model (4.3) is given by

$\mathbf{r}^{(2)} = \mathbf{Y} - \hat{\mathbf{Y}}^{(2)} = \mathbf{Y} - \tilde{H}\mathbf{Y}$, *where* $\tilde{H} = \Gamma_{p-1}(\Gamma_{p-1}^T \Gamma_{p-1})^{-1} \Gamma_{p-1}^T$. *Therefore, we have*

$\mathbf{r}^{(2)} = (I - \tilde{H})\mathbf{Y} = (I - \tilde{H})(\Gamma_{p-1}\phi_{p-1} + \mathbf{b} + \varepsilon^{(1)}) = (I - \tilde{H})(\mathbf{b} + \varepsilon^{(1)})$, (*because* $(I - \tilde{H})\Gamma_{p-1}\phi_{p-1} = 0$)

which implies that $\mathbf{r}^{(2)}$ differs from $(I - \tilde{H})\varepsilon^{(1)}$ substantially unless $\mathbf{b} = 0$. (C.f. eqn. (2.9).)

5 Outliers in time series

What are time series outliers? To motivate discussion, we consider first the simple $AR(1)$ case. Suppose that we have n observations, $Y_1, Y_2, \cdots Y_n$, tentatively identified as coming from the following $AR(1)$ model:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \tag{5.1}$$

Let $\hat{\phi}_{LS}$ denote the least square estimate of ϕ. Let $\hat{\phi}_R$ denote a robust estimate of ϕ, e.g., the GM-estimate as described in Kleiner, Martin and Thomson (1979). Let us write

$$rr_t = Y_t - \hat{\phi}_R Y_{t-1}, \quad t = 1, 2, ..., n$$

the *robust fitted residuals* (RFR). Recall that the least square fitted residuals

$$r_t = Y_t - \hat{\phi}_{LS} Y_{t-1} t = 1, 2, ..., n$$

are abbreviated by LSR.

We consider now the scatter plot Y_t versus Y_{t-1}. There are three possible positions where potential outliers may occur. In Fig. 1(a), the outlier is not a *globally extreme point* and hence may be hidden in a marginal view of the data. The outlier may not be revealed by the examination of LSR since the estimated regression line may be 'pushed' towards the outlier. On the other hand, RFR may expose this outlier. In Fig. 1(b), the outlier is quite clear and could be detected by "eyes" from the marginal view and the examination of RFR. Note that in this case, the examination of LSR could fail to detect the outlier. In Fig. 1(c), the outlier is 'consistent' with the regression line of slope ϕ. Outliers of this kind do not have serious effect on the estimate of ϕ and may go unnoticed in the following situations. Suppose that ε_{t_o} is an innovation outlier. Then Y_{t_o} can also be an outlier by reference to (5.1). Its effect could be carried to the next data point, i.e. Y_{t_o+1} could be another outlier. The situation is illustrated by Fig. 1(d). Then even if we know ϕ, the residual $r_{t_o+1} = Y_{t_o+1} - \phi Y_{t_o}$ could be small but the residual $r_{t_o+2} = Y_{t_o+2} - \phi Y_{t_o+1}$ could be large. As a result, the residuals can give misleading information about the positions of the outliers and may result in misspecifying the types of outliers, e.g., innovation outliers may be misspecified as additive outliers. This situation could have a serious repercussion on the robust estimation of the spectrum. Similar problems may occur when consecutive additive outliers are present. A practical example will be given later in example 2. For the case of Gaussian AR(1) the usefulness of marginal view, LSR and RFR for detecting outliers is thus questionable. Fortunately, for the simple cases illustrated in Fig. 1 the outliers all lie outside the area of concentration of the bivariate normal distribution $F(y_t, y_{t-1})$. For marginal (univariate) and bivariate distributions, we have various graphical techniques to examine the normality of the data. For example, the normal QQ-plot in the marginal case and the lag-1 scatter plot in the bivariate case. However, for higher dimensions, it is difficult to picture the multivariate distribution of the data. Therefore, a 1-dimensional measure seems to be desirable.

Martin (1983) has given a general definition for outliers in time series in this direction:

> "Y_t is an outlier if and only if the prediction residual $r_t = Y_t - \hat{Y}_t^{t-1}$ is large relative to S_M^2 for some M". Here, $\hat{Y}_t^{t-1} = E(Y_t | Y_{t-1}, \dots, Y_{t-M})$ is the conditional-mean-prediction of Y_t given $Y_{t-1}, \dots, Y_{t-M}$, and S_M^2 is the corresponding conditional mean-square error (MSE) of prediction.

Let us discuss this definition.

Following this definition, if the data come from an AR(p) model, then we may need to fit AR models from order 1 up to order p and examine the resulting residuals. Martin (1983, p.195) suggested 4 as a practical choice of p. In addition, when there are outliers, the conditional-mean- predictor $\hat{Y}_t^M$ must be robust-resistant (see, e.g., Martin (1983) p.196), which is more expensive to obtain. Therefore, using this method to detect outliers seems to be computationally quite expensive. On the other hand, note that for a Gaussian time series,

$$f(y_t, y_{t-1}, \dots, y_{t-p}) = f(y_t | y_{t-1}, \dots, y_{t-p}) f(y_{t-1} | y_{t-2}, \dots, y_{t-p}) \cdots f(y_{t-p+1} | y_{t-p}) f(y_{t-p}) \quad (5.2)$$
$$= N(y_t; \hat{y}_t^{t-1}, S_p^2) N(y_{t-1}; \hat{y}_{t-1}^{t-2}, S_{p-1}^2) \cdots N(y_{t-p}; 0, S_o^2)$$

where $S_o^2 = E(Y_t^2)$ and, typically, $N(y_t; \hat{y}_t^{t-1}, S_p^2)$ denotes the Gaussian density function with mean $\hat{y}_t^{t-1}$ and variance S_p^2. Suppose now that Y_{t_o} is a transparent outlier in the multivariate "view" $f(y_t, y_{t-1}, \dots, y_{t-p})$. We may not be aware of its presence if we only examine the individual terms $f(y_t\, y_{t-1}, \dots, y_{t-p}), \dots, f(y_{t-p})$. (Note that an examination of these terms corresponds to an examination of the marginal distribution and the residuals from AR(p) , ..., AR(1).) For, suppose that the outlier Y_{t_o} is large relative to S_p. Now, for $M \le p, S_M^2$, which is always greater than S_p^2, consists of two parts of errors, namely, the pure error and the error due to the lack of fit. Therefore, if the effect of Y_{t_o} enters into the conditional term $f(y_t | Y_{t-1}, \dots, Y_{t-M})$ and S_M^2 is much greater than S_p^2, then Y_{t_o} may not be large relative to S_M^2 and hence disguised in the "conditional" view.

We propose an alternative 1-dimensional measure.

Now, for a fixed M, $z_t^M = (Y_{t-1}, Y_{t-2}, Y_{t-M})^T, t=1,2,\dots,N$, is assumed to have a multivariate normal distribution. Let $\sum_M$ be the variance-covariance of matrix of z_t^M. The Gaussian feature of z_t^M is reflected by the X^2 distribution of the 1-dimensional measure:

$$d_t^M = z_t^{MT} \sum_m^{-1} z_t^M$$

This leads to a new approach for outlier detection to be discussed in the next section.

6 A new approach o outliers detection in linear autoregressive process

We write the linear autoregressive process AR(p), $\{Y_t\}$, in the following state space form:

$$Z_{t+1} = B\ \ z_t + \tilde{\varepsilon}_t, \quad Y_t = (1,0,\ldots,0)z_{t+1}, \tag{6.1}$$

where

$$\tilde{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)^T, \qquad B = \begin{bmatrix} \phi_1 \phi_2 & \ldots & \phi_{p-1} \phi_p \\ 1\ 0 & \ldots & 0\ \ 0 \\ 0\ 1 & \ldots & 0\ \ 0 \\ \cdot\ \ \cdot & \ldots & \cdot\ \ \ \cdot \\ \cdot\ \ \cdot & \ldots & \cdot\ \ \ \cdot \\ \cdot\ \ \cdot & \ldots & \cdot\ \ \ \cdot \\ 0\ 0 & & 1\ \ 0 \end{bmatrix}$$

Suppose now that we have a realization of size $n, Y_1, Y_2, \ldots, Y_n$ say, from the AR(p) process. From the state space point of view, at time t, it is the relative position of $Y_t, Y_{t-1}, \ldots, Y_{t-p+1}$ in the p dimensional space that we are interested in, not just the Y_t itself. Therefore, it is more reasonable to refer to the state vector, i.e. "remote" state vectors. Geometrically speaking, we look for remote points in the p-dimensional space spanned by the columns of Γ.

To measure the "remoteness" of z_t, an attractive metric seems to be the *Mahalanobis distance*, d_t, where

$$d_t = z_t^T \sum^{-1} z_t.$$

We note that $z_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^T$ is the state vector *at time* $t-1$, and that d_t measures the Mahalanobis distance between the state vector at time $t-1$ and the zero vector (or the mean vector in the general case).

Now, under the hypothesis that the autoregressive process is Gaussian and there is no outlier, it holds that

$$\forall t = p+1, \ldots, n, \quad d_t \sim \chi_p^2$$

Therefore, by (3.9) nh_t *leads itself as a useful measure for outlier detection within the linear Gaussian context.* This result is not valid in conventional regression analysis. The discussion in Section 3 suggests that it is practicable to use DEH to detect outlying state vectors. If h_t is sufficiently large, we may say z_t is an outlying state vector.

Recall that

$$z_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^T, \ t=1,2,\ldots,n. \quad h_t = z_t^T(\Gamma^T\Gamma)^{-1}z_t.$$

Suppose that Y_{t-1} is large; then its effect will enter into $z_t, z_{t+1}, \ldots, z_{t+p-1}$ and $h_t, h_{t+1}, \ldots, h_{t+p-1}$ will be large. Therefore, if h_{t-1} is small (i.e. $Y_{t-2}, Y_{t-3}, \ldots, Y_{t-p-1}$ are not outliers) and h_t is large, we may identify Y_{t-1} as an outlier. We have noted earlier that when some of the h_t's are large, there are problems associated with using LSR for diagnostic check and outlier detection. Now we see that when there are outliers, then some of

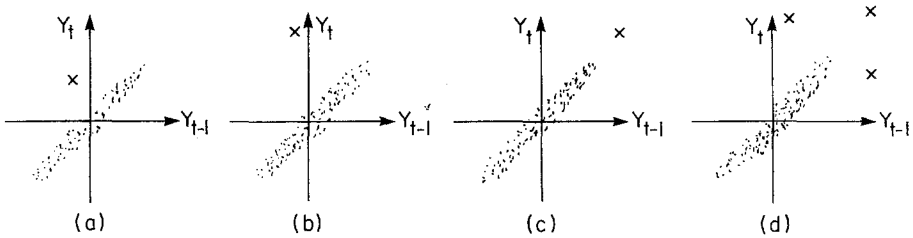


Figure 1. (a) Outliers hidden in marginal view of data; (b) "Clear" outlier, which can be detected by both marginal view and "robust" fitted residuals"; (c) Outliers consistent with the regression line; (d) Scatter plot when one innovation outlier ε_{t_o} is present

h_t's will be large. In fact, a simple examination of the LSR is not a good tool for outliers detection. Suppose Y_{t-1} is an outlier. Then $h_t, h_{t-1}, \ldots, h_{t+p-1}$ will be large and tend to reduce the residuals $r_t, \cdots r_{t+p-1}$ by reference to (2.9). As a result, even if $Y_t, Y_{t+1}, \ldots, Y_{t+p-1}$ are also outliers, $r_t, r_{t+1}, \ldots, r_{t+p-1}$ may not be sufficiently large and hence $Y_t, Y_{t+1}, \ldots, Y_{t+p-1}$ may go unnoticed as outliers if only the LSR are examined.

We may conclude that for time series data adequately described by an autoregressive model, the presence of an outlier may lead to a situation in which outliers immediately following it may go unnoticed if we rely exclusively on LSR for outlier detection., However, in the context of time series, outliers can and do often occur in batches, for example, when innovation outliers are present (see, e.g., Fox (1972) and Kleiner, Martin and Thomson (1979)).

7 Computational aspects

Computational method for a single hat matrix has been discussed by many authors. See, for example, Velleman and Welsch (1981) and Belsley, Kuh and Welsch (1980). In our case, we sometimes need to compute hat matrices for different AR orders. In the following, we give a recursive formula for computing DEH for different orders, which is efficient and quite well suited for computer programming.

The following notation will be adopted:

Γ_m: $n \times m$ design matrix for the maximum AR order m

H_p: $n \times n$ hat matrix for order p, $p=1,2,\ldots,m$

h_t^p: t-th DEH of H_p, $t=1,2,\ldots,n$

R: $m \times m$ upper triangular matrix obtained from QR decomposition of Γ_m

M: $M=[m_{ij}]=\Gamma_m R^{-1}$, an $n \times m$ matrix.

Algorithm

STEP 1: Compute R from QR decomposition (see, e.g., Lawson and Hanson (1974)). $\Gamma_m = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = \tilde{Q}R$ where $\tilde{Q}$ is the first m columns of Q and $\tilde{Q}^T\tilde{Q} = I_m$. R may be obtained by the Householder transformation. Note that this is integral to the autoregressive model building anyway and therefore represents no additional computation.

STEP 2: Compute R^{-1}. Note that R is an upper triangular matrix and can be stably inverted by backward substitution. R^{-1} is also upper triangular.

STEP 3: Compute $M = \Gamma_m R^{-1}$.

STEP 4: Let $\mathbf{m}_t^T = (m_{tl}, m_{t2}, \ldots, m_{tm})$ be the t-the row of M. Then

$$h_t^{p-1} = \sum_{j=1}^{p-1} m_{tj}^2, \quad t=1,2,\ldots,n \qquad (3.6.1)$$

and

$$h_t^p = h_t^{p-1} + m_{tp}^2 \quad \forall p = 2,3,\ldots,m. \qquad (3.6.2)$$

Proof Let M_p be the $p \times p$ block consistuted by the first p columns of M. Then

$$H_p = M_p M_p^T \quad \forall p=1, 2, \ldots, m$$

and (3.6.1), (3.6.2) follows immediately.

Formula (3.6.1) and (3.6.2) provide us a very convenient and efficient method for computing DEH for the AR orders from one to m recursively. Note that the number of operations for computing DEH for all the orders from one to m is the same as that for the single order m.

8 Comparison between using DEH and LSR for outlier detection

(1) Untile now we have assumed that the order of the AR model is known. Even then, LSR still suffer from many limitations as described previously. In practice, the order is rarely known. If there are some outlying data, we do not as yet have a reliable method to choose the order. Besides, LRS depend heavily on estimates of $\hat{\phi}$(*because* $\mathbf{r} = \mathbf{Y} - \Gamma\hat{\phi}$). As a result, we may be looking at something quite irrelevant when LSR are examined. Fortunately, the calculation of DEH does not depend on the estimates $\hat{\phi}$. Much of the conclusion on outlier detection using DEH remains valid even when the order unknown. If we do want to examine DEH for different orders, the convenient recursive formula given in Section 7 is available.

(2) LSR are underlined by a loss function (such as square loss function in least square procedure) while DEH are loss function free. This is important since, e.g., the square loss function

$$L(\mathbf{Y}, \phi) = \sum_i (Y_i - z_i \phi)^2$$

is of questionable value when some of the observations are outliers.

(3) It holds that trace $(H) = \sum_t h_t = p =$ constant. It follows that if some of h_t's are extraordinarily large, then the other h_t's must be small because of the constant sum. Therefore, outlying h_t's will stand out clearly.

9 Examples

Example 1.

Two series NIOAR(2) and IOAR(2) of size 100 are simulated from the following AR(2) model:

$Y_t = 0.9Y_{t-1} - 0.6Y_{t-2} + \varepsilon_t$.

In NIOAR(2), $\varepsilon_t \sim N(0, 0.03)$. In IOAR(2), $\varepsilon_t \sim N(0, 0.03)$ but ε_{50} is set to 1, i.e., it is an innovation outlier. Figs. 2a, 2b are the time series plots of the two series. In IOAR(2), Y_{50} and Y_{51} appear to be large. Fig. 3 shows the time series plot of the LSR from IOAR(2). As expected, only r_{50} appears to be large and hence we could be misled into concluding that Y_{50} is the only outlier, i.e., an "additive outlier". To examine DEH from the two series, we compare three kinds of plots: histogram (Stem-and-leaf plots), time series plot and scatter plot (Figs. 4a, b; 5a, b; 6a,b). From these plots, we see that there are three clear outlying data for the series IOAR(2), namely. h_{51}, h_{52}, and h_{53}. Therefore $z_{51} = (Y_{50}Y_{49})^T, z_{52} = (Y_{51}Y_{50})^T$ and $z_{53} = (Y_{52}Y_{51})^T$ are remote (outlying) state vectors. Since h_t's, $t \le 50$ and ≥ 54 are small, the "remoteness" of z_{51}, z_{52} and z_{53} must be due to Y_{50} and Y_{51}. Therefore Y_{50} and Y_{51} are identified as two *consecutive* outliers by examining DEH. The identification of consecutive outliers tends to suggest the possibility of underlying innovation outlier.

Example 2.

The second example is taken from Marin, Damarov and Vandaele (1982). The series RESEX is monthly series of Bell Canada inward movement of residential telephone extensions in a fixed geographic area from January 1966 to May 1973, a total of 89 data. The time series plot of RESEX (Fig. 7a) shows clearly two extremely large values in November and December 1972 (correspond to the 83th and 84th observations) due to a November "bargain month", i.e., free installation of residence extensions, and a spillover effect in December because not all November's orders could be fitted in the same month. Brubacher (1974) identified an ARIMA(2,0,0)×(0,1,0)$_{12}$ models, i.e., RESEX data is represented by an AR(2) model after seasonal differencing. L.S. estimates of the AR(2) model are $\hat{\phi}_1 = 0.537, \hat{\phi}_2 = -0.106$ and the resulting residuals are plotted in Fig. 8. As expected, only r_{83} is large. (h_{84} is as large as 0.91!). Let X_n denote the RESEX data.

Martin and Zeh (1977) have obtained the robust GM-estimates ($\hat{\phi}_1 = 0.50, \hat{\phi}_2 = 0.38$) and discussed the resulting RFR for this series. The lag-1 scatter plot (Figure 9.b) of these RFR shows five outlying points. At first sight, it might be expected that these large values correspond to the outliers X_{83}, X_{84} in the data. In fact, from the time series plot of RFR (Figure 9.a) we see that the outlying points in the residuals are r_{83}, r_{85} and r_{86} (instead of r_{83} and r_{84}). As a result, the RFR provide misleading information about the positions of the outliers, a situation already discussed in Section 5. In fact, the above authors have noticed the problem of using this kind of residuals for determining the types of outliers. By contrast, an examination of the DEH along the same line as in example 1 has enabled us to identify the observations in November and December 1972 as outliers. (See Figs 10a, 10b and 10c)

10 Generalization to non-linear threshold autoregressive models

Threshold autoregressive models (TAR) were proposed by Tong (1978) and it seems to be generally agreed that they from one useful class of non-linear time series models. (See, e.g., discussion of Tong and Lim, 1980). For a comprehensive account of this class of models, see Tong (1983). It may be remarked that these models have helped to bring about a rapid development of non-linear time series analysis. See, e.g., recent review paper by Tong (1987).

We discuss now the self-exciting threshold autoregressive model of order (2; k_1, k_2), or SETAR (2; k_1, k_2) model, which is of the following form:

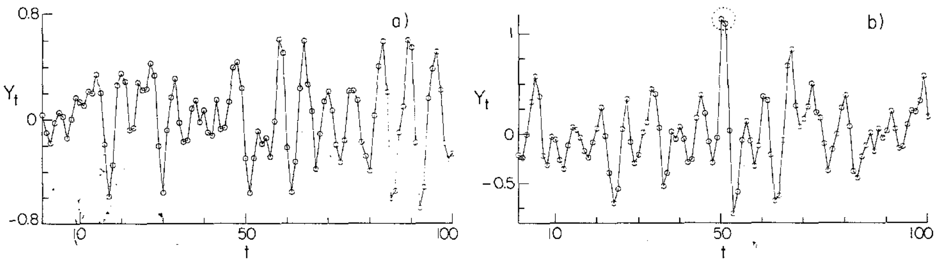
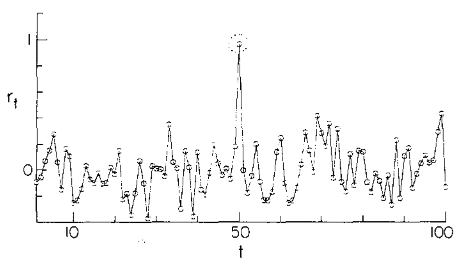


Figure 2. (a) Time series plot of NIOAR(2); (b) Time series plot of IOAR(2)



Midpoint	Count	Midpoint	Count
0.00	40	0.00	19
0.02	40	0.01	28
0.04	14	0.02	18
0.06	1	0.03	17
0.08	0	0.04	10
0.10	1	0.05	3
0.12	0	0.06	0
0.14	0	0.07	0
0.16	0	0.08	3
0.18	0		
0.20	2		
a)		b)	

Figure 3

Figure 4a,b

Figure 3. Time series plot of LSR of IOAR(2)

Figure 4. (a) Frequency counts of DEH of NIOAR(2); (b) Frequency counts of DEH of IOAR(2)

$$X_t = \begin{cases} a_0^{(1)} + \sum_{j=1}^{k_1} a_j^{(1)} X_{t-j} + \varepsilon_t^{(1)} & \text{if } X_{t-d} \le r \\ a_0^{(2)} + \sum_{j=1}^{k_2} a_j^{(j)} X_{t-j} + \varepsilon_t^{(2)} & \text{if } X_{t-d} > r \end{cases} \tag{10.1}$$

where d and r are delay and threshold parameters respectively. For each i=1, 2, $\varepsilon_t^{(i)}$'s are assumed to be i.i.d. and normally distributed, say, $\varepsilon_t^{(i)} \sim N(0, \sigma_i^2)$ i=1,2. For the estimation of $a_j^{(i)}$, σ_i^2; i=1,2, j=0, 1,..., k_i, we can extend the conventional least square procedure since model (10.1) consists of 2 piecewise linear models (Tong (1983) p.133). Suppose now we have n observations $X_1, X_2, \ldots, X_n$ from SETAR (2; k_1, k_2). Let the delay parameter d and the threshold parameter r be fixed. Let k=max(k_1, k_2, d). The (effective) data $\{X_{k+1}, \ldots, X_n\}$ may be divided into two sets by the rule:

$$X_j \in \text{first set if and only if } X_{j-d} \le r,$$
$$X_j \in \text{second set if and only if } X_{j-d} > r.$$

Let $\{X_{j1}^{(1)}, X_{j2}^{(1)}, \ldots, X_{jn_1}^{(1)}, \ldots, X_{jn_1}^{(1)}\}$ and $\{X_{j1}^{(2)}, X_{j2}^{(2)}, \ldots, X_{jn_2}^{(2)}\}$, $(n_1+n_2=n)$, denote the data in the first and second sets respectively, after the division. We have the following "piecewise linear model" formalism:

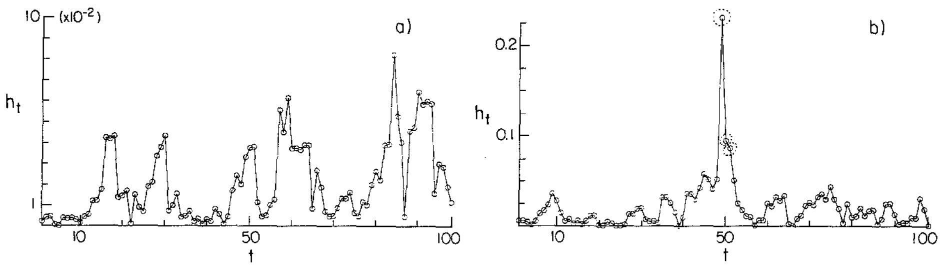


Figure 5. (a) Time series plot of DEH of NIOAR(2); (b) Time series plot of DEH of IOAR(2)

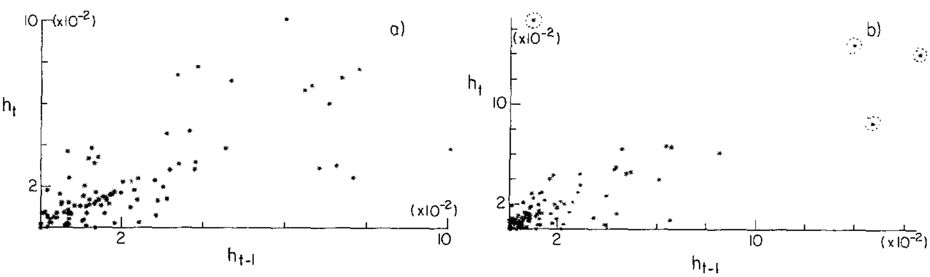


Figure 6. (a) Scatter plot of DEH of NIOAR(2); (b) Scatter plot of DEH of IOAR(2)

$$\begin{cases} \mathbf{X}_1 = \mathbf{A}_1\ \theta_1 + \varepsilon_1 \\ \mathbf{X}_2 = \mathbf{A}_2\ \theta_2 + \varepsilon_2 \end{cases} \tag{10.2}$$

where, for i=1,2, $\mathbf{X}_i = (X_{j1}^{(i)},..., X_{jn_i}^{(i)})^T$, $\varepsilon_i = (\varepsilon_{j1}^{(i)},..., \varepsilon_{jn_i}^{(i)})^T$, $\theta_i = (a_o^{(i)},..., a_{k_i}^{(i)})$, and

$$\mathbf{A}_i = \begin{bmatrix} 1 & X_{j1-1}^{(i)} & X_{j1-2}^{(i)} & \cdots & X_{j1-k_i}^{(i)} \\ 1 & X_{j2-1}^{(i)} & X_{j2-2}^{(i)} & \cdots & X_{j2-k_i}^{(i)} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1 & X_{jn_i-1}^{(i)} & X_{jn_i-1}^{(i)} & \cdots & X_{jn_i-k_i}^{(i)} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_i^{(i)^T} \\ \mathbf{z}_2^{(i)^T} \\ . \\ . \\ . \\ \mathbf{Z}_{n_i}^{(i)^T} \end{bmatrix}$$

where $\mathbf{Z}_l^{(i)^T} = (1\ \ X_{jl-1}^{(i)}\ \ X_{jl-2}^{(i)}\ \ \cdots\ \ X_{jl-k_i}^{(i)})$.

Because of the piecewise linearity, the procedure for outlier detection in the linear AR case can be generalized straightforwardly to the case of threshold autoregressive model. In the case of SETAR model (10.2), we define hat matrices TAR H_i, i=1,2, as follows:

$$TARH_i = \mathbf{A}_i(\mathbf{A}_i^T\mathbf{A}_i)^{-1}\mathbf{A}_i^T, i=1,2.$$

For outlier detection in model (10.2), the diagonal elements $h_t^{(i)}$ of $TARH_i$ (denote by DEH_i) are examined. Large values of $h_t^{(i)}$'s correspond to outlying row vectors $\mathbf{z}_t^{(i)^T}$'s of $\mathbf{A}_i$. In the linear AR case, a large value, h_{t_o} say, is due to the fact that some of the X_{t_o-1}, ..., and $X_{t_o-k_i}$ are large. In the case of SETAR modeling, of course, $X_{t_o-1}^{(i)}$ is not

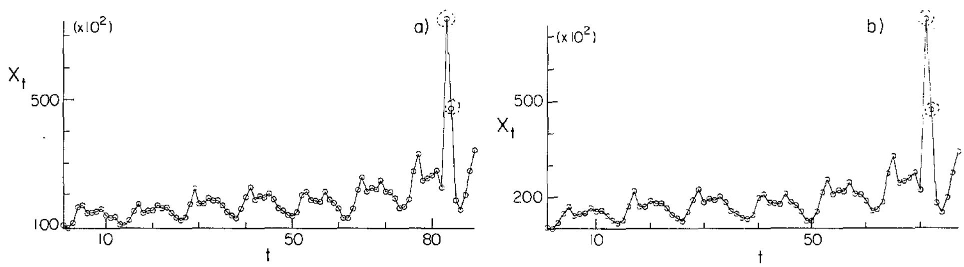


Figure 7. (a) Time series plot of RESEX; (b) Time series plot of differenced RESEX

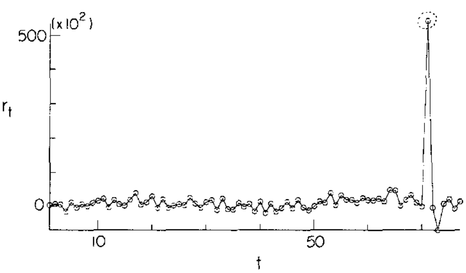


Figure 8. Time series plot of LSR of RESEX

necessarily X_{t_o-1}. Therefore, unlike the linear AR case, even if $h_{t_o-1}^{(i)}$ is small and $h_{t_0}^{(i)}$ is large, we cannot conclude that X_{t_0-1} is an outlier. To identify outliers, it is suggested that we print out the design matrix A_i with the vector $\mathbf{h}_i$ augmented in the last column:

$$I_i = (\mathbf{A}_i | \mathbf{h}_i), \quad where \ \mathbf{h}_i^T = (h_1^{(i)}, h_2^{(i)}, \ldots, h_{n_i}^{(i)}) \ \text{and} \ h_j^{(i)} = z_j^{(i)T}(A_i^T A_i)^{-1} z_j^{(i)}$$

Examining DEH_i using this format, we can see clearly the relationship between DEH_i and the original observations and can therefore identify outlying data if any exist.

One problem of using DEH for outlier detection in SETAR modeling is that we do not know as yet the distribution of $h_t^{(i)}$, i=1,2, t=1,2,..., n_i. Experience suggests that $h_t^{(i)}$'s are very sensitive to outliers and outlying data will cause extreme values in $h_t^{(i)}$'s which will stand out clearly. In any event, we can identify those data which are the *most influential*. Therefore, examining DEH still seems to be worthwhile in the case of SETAR modeling.

Example 3.

TAR1 is a series of 140 observations simulated from the following SETAR (2; 1, 1) model:

$$X_t = \begin{cases} -2.0X_{t-1}+\varepsilon_t & \varepsilon_t \sim N(0, 0.2^2) \\ -0.4X_{t-1}+\varepsilon_t \end{cases}$$

The 53th observation X_{53}=0.322 of TAR1 is changed to X'_{53}=1.2, i.e., an additive outlier. We denote the resulting series by AOTAR1. The first 100 data of TAR1 and AOTAR1 are shown in Fig. 11. Note that X'_{53} is still within the dynamic range. Let us

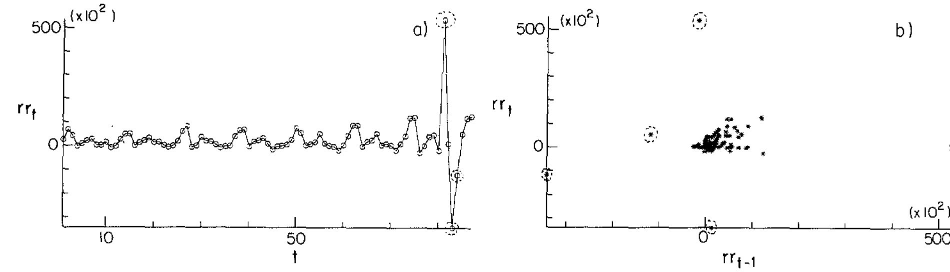


Figure 9. (a) Time series plot of RFR of RESEX; (b) Scatter plot of RFR of RESEX

Midpoint	Count
0.0	72
0.1	0
0.2	1
0.3	0
0.4	0
0.5	0
0.6	1
0.7	0
0.8	1

Figure 10. (a) Frequency count of DEH of RESEX

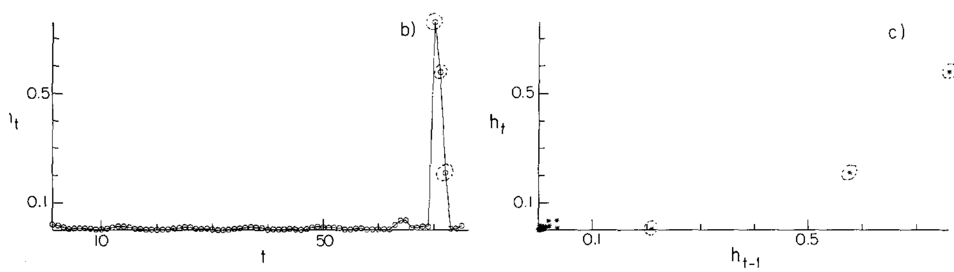


Figure 10. (b) Time series plot of DEH of RESEX; (c) Scatter plot of DEH of RESEX

now examine the DEH_i,i=1,2, for these two series. Since the outlier X_{53}' affects only DEH_2, we compare only DEH_2 of TAR1 and AOTAR1. From the histograms, time series plots, and scatter plots (Fig. 12a, 12b and 12c respectively), we are able to identify $h_{29}^{(2)}$, which corresponds to X_{53}', as an influential data.

Example 4.

Table 1 is the well-known data set (log-transformed) of Canadian LYNX trapped in the Mackenzie River district of North-west Canada for the period 1821-1934. Fig. 13 shows the time series plot of these 114 data. This data set appeared originally in a paper by Elton and Nicholson (1942). This is a "noisy" data set and outliers may be suspected. (See Elton and Nicholson (1942) for an account of the collection of data). A routine application of the method developed in the previous sections for *linear* autoregressive models reveals that more than one third of the data should be considered outliers! This clearly demonstrates the inadequacy of linear autoregressive models for the data. A *non-linear* model is therefore entertained. Tong and Lim (1980) have fitted a a threshold

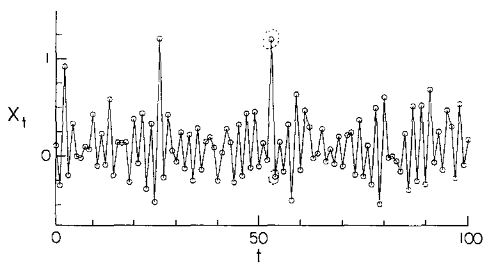


Figure 11

Midpoint	Count
0.00	55
0.04	14
0.08	3
0.12	0
0.16	1
0.20	0
0.24	0
0.28	0
0.32	1

a i)

Midpoint	Count
0.00	19
0.01	16
0.02	7
0.03	6
0.04	9
0.05	6
0.06	5
0.07	2
0.08	1
0.09	0
0.10	3

a ii)

Figure 12a

Figure 11. Time series plot of AOTAR1

Figure 12a. i) Frequency counts of DEH in the 2nd piece of TAR1; ii) Frequency counts of DEH in the 2nd piece of AOTAR1

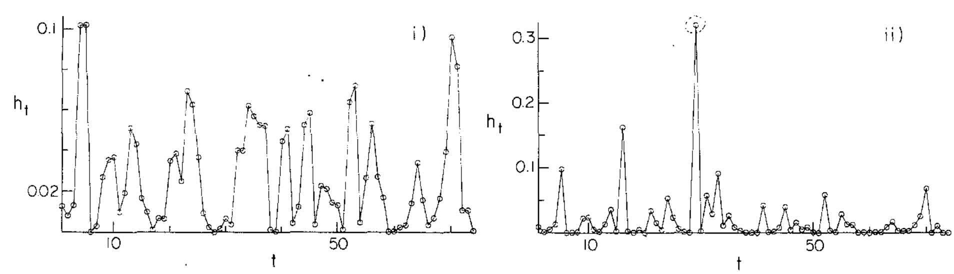


Figure 12b. i) Time series plot of DEH in the 2nd piece of TAR1; ii) Time series plot of DEH in the 2nd piece of AOTAR1

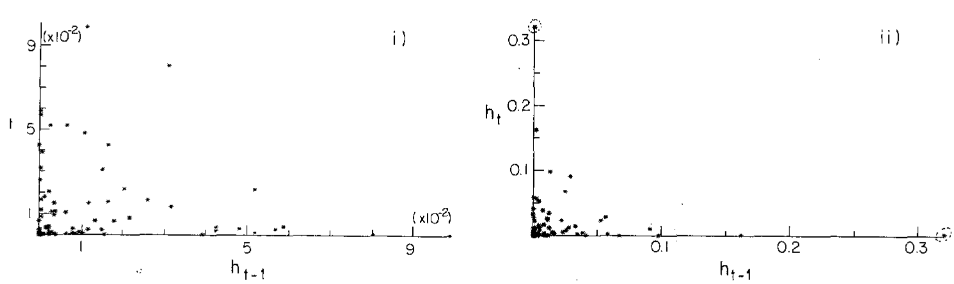


Figure 12c. i) Scatter plot of DEH in the 2nd piece of TAR1; ii) Scatter plot of DEH in the 2nd piece of AOTAR1

model to the first 100 data and we examine the DEH_i, i=1,2, for this model. The time series plots, stem-and-leaf plot are shown in Figs. 14a and 14b respectively. From these plots we do not see any clear outlying points, but we can nevertheless identify the most influential data. In the first piece, $h_{38}^{(1)}, h_{39}^{(1)}, h_{40}^{(1)}$ and $h_{50}^{(1)}$ have the largest values. Therefore the corresponding row vectors of the design matrix A_1 in the first piece are the most influential vectors. To identify which observations correspond to these vectors, we

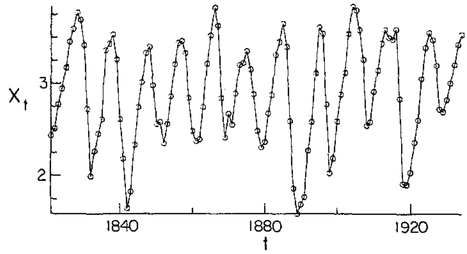


Figure 13. Time series plot of (LOG) LYNX Data (1821-1934)

Stem-and-leaf
Leaf Unit = 0.010

5	0	23333
8	0	445
13	0	66677
20	0	8888899
25	1	01111
(6)	1	222333
20	1	444555555
11	1	66666
6	1	889
3	2	011

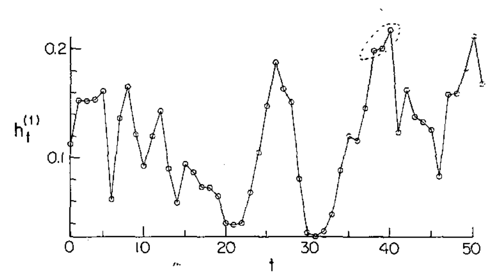


Figure 14a

Figure 14b

Figure 14. (a) Time series plot of $h_t^{(1)}$; (b) Stem-and-leaf display of $h_t^{(1)}$

examine the design matrix A_1 with vector h_1 of DEH, augmented in the last column, which is shown in Table 2. From Table 2 we see that $h_{38}^{(1)}$, $h_{39}^{(1)}$ and $h_{40}^{(1)}$ correspond to the observations X_{1891}, X_{1892}, and X_{1893}. $h_{50}^{(1)}$ corresponds to the vector of observations $(X_{1918}\, X_{1917}\, X_{1916}\, X_{1915}\, X_{1914})$. Note that $h_{51}^{(1)}$, which corresponds to the vector $(X_{1919}\, X_{1918}\, X_{1917}\, X_{1916}\, X_{1915})$, is small when compared with $h_{50}^{(1)}$. Altogether, it would seem that $X_{1891}, X_{1892}, X_{1893}$ and X_{1914} are the most influential data. Similarly, from the examination of DEH_2 (see Figs. 15a and 15b) the state vector $(X_{1904}\, X_{1905})$ turns out to be the most influential.

On referring to the history of the data set (Elton and Nicholson (1942)), it transpires that these 114 data came from three different sources. Briefly, records from 1812 to 1891 and from 1897 to 1913 were obtained from the London archives of Hudson's Bay Company and those for 1815-34 were from the Company's fur Trade Department in Winnipeg. Those for 1892-1896 and 1914 were supplied to Elton in 1928 by Mr. Charles French, then Fur Trade Commissioner of the Company in Canada, who said that the figures were obtained from private records kept by some of the older fur-trade factors. It is intriguing that our analysis of DEH_1 (but not of DEH_2) seems to be reasonably compatible with the historical background. Of course, we cannot rule out the possibility that the apparent compatibility is just a coincidence. It is also interesting to report that an examination of the DEH corresponding to linear AR models does not lead to similar results. (Details are given in an unpublished M.Phil. thesis, Chinese University of Hong Kong, 1984).

```
Stem-and-leaf
Leaf Unit = 0.0010
  1    2   6
  7    3   023447
 13    4
 16    5   027
      18   6
(3)    7   013
 18    8   0116
 14    9   68
 12   10   016
  9   11   7
  8   12   357
  5   13   09
  3   14   07
  1   15
  1   16   6
```

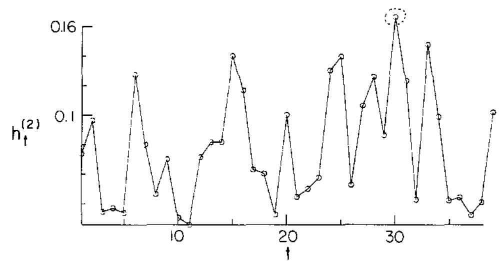


Figure 15a

Figure 15b

Figure 15. (a) Time series plot of $h_t^{(2)}$; (b) Stem-and-leaf display of $h_t^{(2)}$

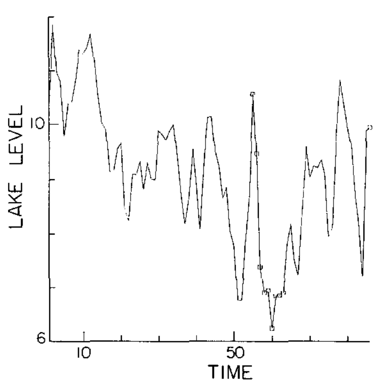


Figure 16. Level of Lake Huron in feet (reduced by 570), 1875-1972: identified outliers are marked by □

Table 1. Log transformed LYNX data (1821 - 1934)

Year	1	2	3	4	5	6	7	8	9	10
1821 - 1830	2.4298	2.5065	2.7672	2.9400	3.1688	3.4504	3.5942	3.7740	3.6946	3.4111
1831 - 1840	2.7185	1.9912	2.2648	2.4456	2.6117	3.3589	3.4289	3.5326	3.2610	2.6117
1841 - 1850	2.1790	1.6532	1.8325	2.3284	2.7372	3.0141	3.3282	3.4041	2.9809	2.5575
1851 - 1860	2.5763	2.3522	2.5563	2.8639	3.2143	3.4354	3.4580	3.3261	2.8351	2.4757
1861 - 1870	2.3729	2.3892	2.7419	3.2103	3.5200	3.8274	3.6288	2.8370	2.4065	2.6749
1871 - 1880	2.5539	2.8943	3.2025	3.2243	3.3524	3.1541	2.8785	2.4757	2.3032	2.3598
1881 - 1890	2.6712	2.8669	3.3101	3.4489	3.6465	3.3998	2.5899	1.8633	1.5911	1.6902
1891 - 1900	1.7709	2.2742	2.5763	3.1113	3.6054	3.5434	2.7686	2.0212	2.1847	2.5877
1901 - 1910	2.8797	3.1163	3.5397	3.8445	3.8002	3.5791	3.2639	2.5378	3.5821	2.9077
1911 - 1920	3.1424	3.4334	3.5798	3.4901	3.4749	3.5786	2.8287	1.9085	1.9031	2.0334
1921 - 1930	2.3598	2.6010	3.0538	3.3860	3.5532	3.4676	3.1867	2.7235	2.6857	2.8209
1931 - 1934	3.0000	3.2014	3.4244	3.5310						

Table 2. Design matrices of LYNX data

1st piece

	z_t^T					h_t
1	1.9912	2.7185	3.4111	3.6949	3.7740	0.1123
	2.2648	1.9912	2.7185	3.4111	3.6946	0.1529
	2.4456	2.2648	1.9912	2.7185	3.4111	0.1523
	2.6117	2.4456	2.2648	1.9912	2.7185	0.1536
	3.3589	2.6117	2.4456	2.2648	1.9912	0.1615
	2.1790	2.6117	3.2610	3.5326	3.4289	0.0618
	1.6532	2.1790	2.6117	3.2610	3.5326	0.1361
	1.8325	1.6532	2.1790	2.6117	3.2610	0.1652
	2.3284	1.8325	1.6532	2.1790	2.6117	0.1214
	2.7372	2.3284	1.8325	1.6532	2.1790	0.0924
	3.0141	2.7372	2.3284	1.8325	1.6532	0.1199
	3.3282	3.0141	2.7372	2.3284	1.8325	0.1431
•	2.5575	2.9809	3.4041	3.3282	3.0141	0.0895
•	2.5763	2.5575	2.9809	3.4041	3.3282	0.0585
•	2.3522	2.5763	2.5575	2.9809	3.4041	0.0937
	2.5563	2.3522	2.5763	3.5575	2.9809	0.0860
	2.8639	2.5563	2.3522	2.5763	2.5575	0.0729
	3.2143	2.8639	2.5563	2.3522	2.5763	0.0721
	2.4757	2.8351	3.3261	3.4580	3.4354	0.0645
	2.3729	2.4575	2.8351	3.3261	3.4580	0.0400
	2.3892	2.3729	2.4757	2.8351	3.3261	0.0383
	2.7419	2.3892	2.3729	2.4757	2.8351	0.0398
	3.2103	2.7419	2.3892	2.3729	2.4757	0.0674
	2.4065	2.8370	3.6288	3.8274	3.5200	0.1041
	2.6749	2.4065	2.8370	3.6288	3.8274	0.1478
	2.5539	2.6749	2.4065	2.8370	3.6288	0.1879
1	2.8943	2.5539	2.6749	2.4065	2.8370	0.1635
	3.2025	2.8943	2.5539	2.6749	2.4065	0.1511
	2.4757	2.8785	3.1541	3.3524	3.2243	0.0803
	3.3032	2.4757	2.8785	3.1541	3.3524	0.0312
	2.3598	2.3032	2.4757	2.8785	3.1541	0.0276
	2.6712	2.3598	2.3032	2.4757	2.8785	0.0328
	2.8669	2.6712	2.3598	2.3032	2.4757	0.0481
	3.3101	2.8669	2.6712	2.3598	2.3032	0.0885
	1.3633	2.5899	3.3998	3.6465	3.4489	0.1198
	1.5911	1.3633	2.5899	3.3998	3.6465	0.1153
	1.6902	1.5911	1.8633	2.5899	3.3998	0.1453
•	1.7709	1.6902	1.5911	1.3633	2.5899	0.1983
•	2.2742	1.7709	1.6902	1.5911	1.8633	0.2004
•	2.5763	2.2742	1.7709	1.6902	1.5911	0.2178
	3.1113	2.5763	2.2742	1.7709	1.6902	0.1235
	2.0212	2.7686	3.5434	3.6054	3.1113	0.1623
	2.1847	2.0212	2.7686	3.5434	3.6054	0.1377
	2.5877	2.1847	2.0212	2.7686	3.5434	0.1329
	2.8797	2.5877	2.1847	2.0212	2.7686	0.1256
	3.1163	2.8797	2.5877	2.1847	2.0212	0.0826
	2.5821	2.5378	3.2639	3.5791	3.8002	0.1586
	2.9074	2.5821	2.5378	3.2639	3.5791	0.1595
	3.1424	2.9074	2.5821	2.5378	3.2639	0.1823
	1.9085	2.8287	3.5786	3.4749	3.4901	0.2119
1	1.9031	1.9085	2.8287	3.5786	3.4749	0.1681

2nd piece

	z_t^T		h_t
1	3.4111	3.6946	0.0738
	2.7185	3.4111	0.0964
	3.4289	3.3589	0.0349
	3.5326	3.4289	0.0372
	3.2610	3.5326	0.0341
	2.6117	3.2610	0.1276
	3.4041	3.3282	0.0380
	2.9809	3.4041	0.0466
•	2.4354	3.2143	0.0702
•	3.4580	3.4354	0.0308
•	3.3261	3.4580	0.0261
	2.8351	3.3261	0.0715
	3.5200	3.2103	0.0814
	3.8274	3.5200	0.0819
	3.6288	3.8274	0.1404
	2.8370	3.6288	0.1170
	3.2243	3.2025	0.0630
1	3.3524	3.2243	0.0604
	3.5141	3.3524	0.0330
	2.8785	3.1541	0.1000
	3.4489	3.3101	0.0447
	3.6465	3.4489	0.0502
	3.3998	3.6465	0.0577
	2.5899	3.3998	0.1305
	3.6054	3.1113	0.1399
	3.5434	3.6054	0.0526
•	2.7686	3.5434	0.1063
	3.5397	3.1163	0.1259
•	3.8445	3.5397	0.0864
•	3.8002	3.8445	0.1668
	3.5791	3.8002	0.1235
	3.2639	3.5791	0.0429
	2.5378	3.2639	0.1479
	3.4334	3.1424	0.0989
	3.5798	3.4334	0.0421
	3.4901	3.5798	0.0442
	3.4749	3.4901	0.0327
	3.5786	3.4749	0.0412
1	2.8287	3.5786	0.1019

11 A hydrological example

The final example is concerned with the level of Lake Huron for July of each year from 1875 to 1972 inclusively as listed in the recent book by Brockwell and Davis (1987, p.499). Preliminary data analysis reveals that a first differencing or possibly a second differencing of the data is advisable so as to reduce the data to stationarity in the mean (c.f. Box and Jenkins, 1970). The hat matrix technique then suggests the patch of the points numbered approximately 55 to 63 and the last datum as outliers. On assuming that the listing is error-free up to data point numbered say 65, then the suggested outlying patch corresponds to the period of the 1930's. It would then seem interesting to explore the connection between this and the famous *dust-bowl* period of the 1930's. As for the suggested outlying singleton, it seems that it may well be due to the omission of a column of twelve data taking place in the vicinity of the singleton. The book has inadvertently listed only 86 of the announced 98 data!

We have included the above example merely to illustrate the potential of our methodology in analysising hydrological time series. We do not pretend that the analysis is complete. However, a complete analysis can only be possible with access to information not at present available to us, e.g. the omitted column of data, the amount of precipitation, the temperature and the wind speed, ect. in the Lake Huron catchment area, and similar data on Lake Huron's neighboring great lakes.

12 Concluding remarks

Examination of the DEH gives a direct, efficient and conceptually appealing method for outlier detection in autoregressive modeling. We have demonstrated that it has definite practical advantages over simple examination of LSR and RFR. However, we have throughout this paper avoided rigorous limit arguments, i.e., subsequent rigorization of our heuristics may be deemed necessary by a more mathematical audience. We plan to supply the rigor elsewhere along the lines developed by Kunsch (1984).

Acknowledgements

We are very grateful to Dr. R. Moeanaddin for computational assistance with the hydrological example.

References

Belsley, D.A.; Kuh, E,; Welsch, R.E. 1980: Regression diagnostics. New York: Wiley

Box, G.E.P; Jenkins, G.M. 1970: Time series analysis: Forecasting and control. San Franciso: Holden-Day

Brockwell, P.J; Davis, R.A. 1987: Time series: Theory and methods. New York: Springer-Verlag

Brubacher, S.R. 1974: Time series outlier detection and modeling with interpolation. Bell Laboratories Technical Memo

Cook, R.D.; Weisberg, S. 1982: Residuals and influence in regression. London: Chapman and Hall

Elton, C.; Nicholson, M. 1942: The ten-year cycle in numbers of the LYNX in Canada. J. Anim. Ecol. 11, 215-244.

Fox, A.J. 1972: Outliers in time series. J. Roy. Statist. Soc., B 34, 3, 350-363

Hipel, K.W.; Mcleod, A.I. 1978: Preservation of the rescaled adjusted range 2. Simulation studies using Box-Jenkins models. Water Resources Research 14, 509-516

Hoaglin, D.C.; Welsch, R. 1978: The hat matrix in regression and ANOVA. Amer. Statistician 32, 17-22

Huber, P. 1981: Robust statistics. New York: Wiley

Kleiner, B.; Martin, R.D.; Thomson, D.J. 1979: Robust estimation of power spectra. J. Roy. Statist. Soc., B 41, 313-351

Kunsch, H. 1984: Infinitesimal robustness for autoregressive processes. Annals of Statistics 12, 843-855

Lawson, C.R.; R.J. Hanson 1974: Solving least-square problems. Englewood Cliffs, N.J.: Prentice-Hall

Mann, H.B.; Wald, A. 1943: On the statistical treatment of linear stochastic difference equations. Econometrica 11, 173-221

Martin, R.D. 1980: Robust methods for time series. In: Findley, D.E.(Ed.) Applied time series. New York: Academic Press

Martin, R.D. 1983: Robust-resistant spectral analysis. In: Brillinger, P.R.; Krishnaiah, P.R. (Eds.) Time series in the frequency domain, handbook of statistics 3. North-Holland

Martin, R.D.; Samarov, A.; Vandaele, W. 1982: Robust method for ARMIA models. Tech. Report. No. 21, Department of Statistics, University of Washington, Seattle

Martin, R.D.; Zeh, J.E. 1977: Determining the character of time series outliers. Proceedings of the Amer. Statist. Assoc., Business and Economics Section

Thomas, H.A.; Fiering, M.B. 1962: Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation. In: Maass et al. (Eds.) Design of Water Resources. Harvard University Press

Tong, H. 1978: On a threshold model. In: Chan, C.H. (Ed.) Pattern recognition and signal processing. Netherlands: Sythoff and Noordhoff

Tong, H. 1983: Threshold models in non-linear time series analysis. New York: Springer-Verlag

Tong, H. 1987: Non-linear time series models of regularly sampled data: A review in proceedings of the first world congress of the Bernoulli society of mathematical statistics and probability, held in Sept. 1986 at Tashkent, USSR. Vol. 2, pp. 355-367, VNU science press, Netherlands

Tong, H.; Lim, K.S. 1980: Threshold autoregression, limit cycles and cyclical data (with discussion). J. Roy. Stat. Soc. 13, 42, 245-292

Velleman, P.; Welsch, R. 1981: Efficient computing of regression diagnostics. Amer. Statistician 35, 234-42

Yevjevich, V.M. 1963: Fluctuation of wet and dry years 1, research data assembly and mathematical modesl. Hydrological Papers, Colorado State University, Fort Collins, Colorado

Accepted June 15, 1989