Influential Observations in Time Series
Author(s): Daniel Peña
Source: *Journal of Business & Economic Statistics*, Vol. 8, No. 2 (Apr., 1990), pp. 235-241
Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association
Stable URL: http://www.jstor.org/stable/1391986
Accessed: 24-11-2015 09:39 UTC

Journal of Business & Economic Statistics, April 1990, Vol. 8, No. 2

Influential Observations in Time Series

Daniel Peña
Laboratory of Statistics, Escuela Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid 28006, Spain

This article studies how to identify influential observations in univariate autoregressive integrated moving average time series models and how to measure their effects on the estimated parameters of the model. The sensitivity on the parameters to the presence of either additive or innovational outliers is analyzed, and influence statistics based on the Mahalanobis distance are presented. The statistic linked to additive outliers is shown to be very useful for indicating the robustness of the fitted model to the given data set. Its application is illustrated using a relevant set of historical data.

KEY WORDS: Diagnostic checks; Missing data; Outliers; Robust methods.

1. INTRODUCTION AND SUMMARY

Observed time series almost always have atypical points. These anomalous values can be produced by nonsystematic changes in the variables that are driving the series or affecting them. Since the forecasts from any time series model are based on the extrapolation of the historical patterns, if the parameters of the model are very dependent on a few atypical observations resulting from isolated or nonrepeatable events, then the quality of the forecasts can be expected to be poor. Moreover, when these parameters have physical or economic interpretations, the presence of undetected influential observations can mislead the scientist about the properties of the model. Finally, the study of these observations provides meaningful information about the robustness of the fitted model to the given data set.

This problem is related to, though different from, the study of outliers, because, as is well known, the fact that an observation is an outlier does not imply that this observation substantially affects the parameter estimates of the assumed model, although in general it will affect the variance of the estimates.

Cook and Weisberg (1982) and Belsley, Kuh, and Welsch (1980) presented an overview of influential observations in the regression model. This study has been extended to some other members of the generalized-linear-model family (Pregibon 1981). The main idea of this approach is to delete suspicious observations and measure the change that this deletion produces in relevant features of the model such as the estimated parameter values or the forecasts.

This article attempts to extend these ideas to dependent observations in the context of time series analysis and is organized as follows. Section 2 summarizes the literature on outliers in time series models and discusses the two basic types of outliers that can occur in a dynamic situation. Sections 3 and 4 show how to build measures of influence for additive outliers and for innovational outliers. Section 5 discusses the computation of these statistics, which are compared in Section 6. Section 7 presents an example.

The main result in this article is a statistic that seems to be useful in indicating which observations have strong influence on the estimated parameter values. This statistic is based on substituting the anomalous data by its interpolated value, using all of the sample information, and can be considered a natural generalization of statistics suggested for independent data. Furthermore, this statistic is linked to additive outliers, known to be the most influential. Second, it has been shown that in time series models the deleting approach is linked to innovational outliers, and the statistic resulting from this procedure is unstable and does not seem to be useful in the time series context. Third, simple expressions are found to relate the parameter values estimated with and without outliers in autoregressive integrated moving average (ARIMA) models, allowing the study of the influence problem within a unified framework.

2. OUTLIERS IN TIME SERIES

Fox (1972) defined two types of outliers that may occur in time series data. The first, called the Type I outlier by Fox, corresponds to a modification of the value of the series due to some external cause such as a gross recording error. Assuming that the observed series z_t follows an ARIMA process, the model for a Type I outlier at time T is

$$\phi(B)\nabla^d y_t = \vartheta(B)a_t, \qquad y_t = z_t, \qquad t \neq T$$
$$= z_t - w_T, \quad t = T, \tag{2.1}$$

where B is the backshift operator, $B^k y_t = y_{t-k}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\vartheta(B) = 1 - \vartheta_1 B - \cdots - \vartheta_q B^q$ are the autoregressive (AR) and moving average (MA) polynomials that will have roots outside the unit circle, and $\nabla^d = (1 - B)^d$. This model can also be written as

$$\pi(B)(z_t - w_T I_t^{(T)}) = a_t, \qquad I_t^{(T)} = 0, \quad t \neq T$$
$$= 1, \quad t = T, \tag{2.2}$$

where $\pi(B) = \nabla^d \phi(B) \vartheta^{-1}(B)$ is the autoregressive representation of the process. Therefore, it is a special case of the intervention analysis model (Box and Tiao 1975) with an instantaneous response function, w_T. Model (2.2) was called the additive-outlier model by Denby and Martin (1979) and Chang and Tiao (1983) and the aberrant observation model by Abraham and Box (1979).

The Type I outlier can be interpreted as the effect on the series of some external error or exogenous change. On the other hand, a Type II outlier can be considered as the effect of some internal change or endogenous effect. If we think of a univariate time series model as an aggregate representation of the pattern of behavior of a vector x_t of explicative time series that are causing the observed series z_t, the noise of the univariate model represents the aggregate of the nonsystematic variations of the components x_t, and, therefore, an exogenous intervention outlier in any of the components will produce an anomalous value on the noise of the univariate process. The model for this type of effect will be

$$\phi(B)\nabla^d z_t = \vartheta(B)(a_t + w_T I_t^{(T)}), \tag{2.3}$$

where the atypical behavior appears in the innovation. This model has been called the innovational outlier (Chang and Tiao 1983) or the aberrant innovation model (Abraham and Box 1979).

Both types of outlier can be written as $\pi(B)(z_t - w_T v(B) I_t^{(T)}) = a_t$, showing that, as indicated by Chang, Tiao, and Chen (1988), both can be modeled as particular cases of the intervention-analysis model. For an additive outlier, $v(B) = 1$, whereas for an innovational outlier, $v(B)\pi(B) = 1$.

Fox (1972) derived the maximum-likelihood-ratio test for both types of outliers for autoregressive processes. Abraham and Box (1979) used the normal contaminated model as the basic setup for making inferences in these models. Denby and Martin (1979) developed generalized M estimators for the first-order autoregressive process and showed that the loss of efficiency in estimating the parameter by least squares is expected to be much larger for additive than for innovational outliers. Martin (1980) presented a class of bounded-influence estimates for pth-order autoregressions, and Martin, Samarov, and Vandaele (1983) proposed robust M estimators for ARIMA models using the state-space representation. Peña and Guttman (1988, 1989) developed a robust Kalman filter to estimate the time series parameters in the presence of outliers. Chang and Tiao (1983) and Chang et al. (1988) extended Fox's results to general ARIMA models and suggested a useful iterative procedure for outlier detection and parameter estimation. They recommended computing the likelihood-ratio statistics $\lambda_{I,T}$ and $\lambda_{A,T}$ to check if the observation T is either an innovational ($\lambda_{I,T}$) or an additive outlier ($\lambda_{A,T}$). These statistics are given by

$$\lambda_{I,T} = \frac{\hat{w}_{I,T}}{\hat{\sigma}_a}$$

and

$$\lambda_{\mathbf{A,T}} = \frac{\hat{w}_{A,T}}{\hat{\sigma}_a(\sum_0^\infty \pi_i^2)^{-1/2}},$$

where $\hat{w}_{I,T}$ and $\hat{w}_{A,T}$ are the estimated values of the outlier w_T, assuming that it belongs to the innovational type or additive type, and π_i are the parameters of the autoregressive representation of the process. In the same spirit, Tsay (1986, 1988) proposed an iterative procedure to specify a tentative model for a time series that accounts for outliers, level shifts, and variance changes. Finally Martin and Yohai (1986) defined influence functionals for time series and related them to the influence curve due to Hampel (1974), and Brillinger (1986) applied a missing-value technique, together with principal-components analysis, to identify influential observations.

3. A MEASURE OF INFLUENCE FOR ADDITIVE OUTLIERS

3.1 The Change in the Parameter Estimates

Suppose we have a stochastic process y_t that follows a univariate ARIMA(p, d, q). It is assumed in what follows that y_t represents deviations from some origin μ that will be the mean if the series is stationary and that the moving average part has a characteristic equation with roots outside the unit circle so that the process is invertible. Then the process can be represented as

$$y_t = \sum_{l=1}^{h} \pi_l y_{t-l} + a_t$$

for some lag h. If the process is purely autoregressive, $h = p + d$. Otherwise, the π coefficients are obtained from $\pi(B) = \phi(B)(1 - B)^d \theta(B)^{-1}$ and, because of the invertibility of $\theta(B)$, these coefficients will decrease and will become practically 0 for some lag h.

Let us now assume that an additive outlier happens at time T; that is, instead of observing y_t we observe z_t, where $z_t = y_t$ $(t \neq T)$ and $z_T = y_T + w_T$. Then, as the Jacobian of the transformation from y_t to z_t is 1, the likelihood function for the observed series z_t conditional on the first h values is

$$l(\pi, \sigma^2, w) = -\left(\frac{n-h}{2}\right) \ln \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{I_2} (z_t - \pi_T' x_t)^2$$
$$- \frac{1}{2\sigma_a^2} \sum_{l=0}^{h} (z_{T+l} + \pi_{l,T} w_T - \pi_T' x_{T+l})^2,$$

where σ_a^2 is the variance of the noise process a_t, $\mathbf{x}_t' = (z_{t-1}, \ldots, z_{t-h})$, $\boldsymbol{\pi}_T = (\pi_{1,T}, \ldots, \pi_{h,T})$, and $\pi_{0,T} = -1$. The set of indexes I_1 is $(h + 1, \ldots, T - 1, T + h + 1, \ldots, n)$. The conditional maximum likelihood estimate of $\boldsymbol{\pi}_T$ is

$$\hat{\boldsymbol{\pi}}_T = (\hat{\mathbf{X}}_y'\hat{\mathbf{X}}_y)^{-1}\hat{\mathbf{X}}_y'\hat{\mathbf{Y}} \tag{3.1}$$

when we have called $\hat{\mathbf{X}}_y$ the matrix of estimated values for the real process y_t that is unobserved at T and $\hat{\mathbf{Y}}$ the vector of responses:

$$\hat{\mathbf{X}}_y = \begin{bmatrix} \hat{y}_h & \hat{y}_{h-1} & \cdots & \hat{y}_1 \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{n-1} & \hat{y}_{n-2} & & \hat{y}_{n-h} \end{bmatrix} \hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_{h+1} \\ \vdots \\ \hat{y}_n \end{bmatrix},$$

where $\hat{y}_j = z_j$ for $j \neq T$ and $\hat{y}_T = z_T - \hat{w}_T$ is obtained by

$$\hat{w}_T = z_T - \hat{z}_{T/n}$$

$$\hat{y}_T = \hat{z}_{T/n} = \sum_{i=1}^{h} \delta_i(z_{T+i} + z_{T-i}), \tag{3.2}$$

where

$$\delta_i = \left(\hat{\pi}_{i,T} - \sum_{l=1}^{h} \pi_{l,T}\pi_{l+i,T}\right) \bigg/ \left(\sum_{l=0}^{h} \hat{\pi}_{l,T}^2\right), \qquad \hat{\pi}_{0,T} = -1.$$

Consequently, the residual at time T is $\hat{z}_{T/n} - \sum_{i=1}^{h} \hat{\pi}_i z_{t-i}$, and $\hat{z}_{T/n}$ can be interpreted as the best estimate of the unobserved process y_T using all of the sample information. Neither this residual nor $\hat{z}_{T/n}$ depends on the value z_T.

The estimate of the additive outlier is the difference between the observed data and its interpolation given the rest of the sample. Note that the computation of $\hat{z}_{T/n}$ is done applying some weighting coefficients to the new series

$$s(j) = z_{T+j} + z_{T-j}. \tag{3.3}$$

These weights are such that $-\delta_i$ is the ith coefficient in the generating function $(\Sigma\pi_i^2)^{-1}\pi(B)\pi(F)$ and, therefore, can be interpreted as the coefficients of the inverse autocorrelation function of the process. It is well known (see Grenander and Rosenblatt 1957, p. 83) that the expected value of a missing value given the rest of the data, $Z_{(T)} = (z_1, \ldots, z_{T-1}, z_{T+1}, \ldots, z_n)'$, is

$$E[z_T \mid Z_{(T)}] = -\sum_{j=1}^{\infty} \rho i_j s(j),$$

where ρi_j is the jth inverse autocorrelation coefficient and $s(j)$ is the process defined by (3.3). Therefore, it can be concluded that, given the parameters, the estimation of the additive outlier is the difference between the observed data and its optimum interpolation (in the mean squared error sense).

The system of equations given by (3.1) and (3.2) has to be resolved iteratively. Starting with an initial value $\hat{\boldsymbol{\pi}}_T(0)$ for $\hat{\boldsymbol{\pi}}_T$, the weights δ_i can be calculated and a value $\hat{w}_T(0)$ computed. This value is then used to compute $\hat{y}_T(0) = z_T - \hat{w}_T(0)$, which provides a new estimate $\hat{\boldsymbol{\pi}}_T(1)$. The process is repeated until convergence.

To study the effect of the outlying value on the estimated parameters, let us call $\hat{\boldsymbol{\pi}}$ the conditional maximum likelihood (ML) estimator of $\boldsymbol{\pi}$, assuming no outliers. Then, $\hat{\boldsymbol{\pi}} = (\mathbf{X}_z'\mathbf{X}_z)^{-1}\mathbf{X}_z'\mathbf{Z}$, where the matrix $\mathbf{X}_z$ and the vector $\mathbf{Z}$ correspond to the observed data and have the same structure as $\hat{\mathbf{X}}_y$ and $\hat{\mathbf{Y}}$ and the same data values except at time T. Of course, the two sets will be identical if $\hat{y}_T = z_T$. Therefore,

$$\mathbf{X}_z = \hat{\mathbf{X}}_y + \hat{w}_T\mathbf{M}, \tag{3.4}$$

where the matrix $\mathbf{M}$ is given by

$$\mathbf{M}' = [\mathbf{0}_{h\times T-k}; \mathbf{I}_{h\times h}; \mathbf{0}_{h\times(n-h-\mathbf{T})}], \tag{3.5}$$

$\mathbf{0}_{a\times b}$ is a null rectangular matrix, and $\mathbf{I}_{h\times h}$ is the identity matrix. Moreover,

$$\mathbf{Z} = \hat{\mathbf{Y}} + \hat{w}_T\mathbf{V}, \tag{3.6}$$

where $\mathbf{V}$ can be partitioned as

$$\mathbf{V}' = [0, \ldots, 01; \mathbf{0}_{h\times 1}'; \mathbf{0}_{(n-h-T)\times 1}']. \tag{3.7}$$

To relate $\hat{\boldsymbol{\pi}}_\mathbf{T}$ and $\hat{\boldsymbol{\pi}}$, let us partition the matrices $\mathbf{X}_z$ and $\hat{\mathbf{X}}_\mathbf{y}$ and the vectors $\mathbf{Z}$, $\hat{\mathbf{Y}}$ in the same way as (3.5) and (3.7). If we state that $\mathbf{X}_z' = [\mathbf{X}_z'(1)\mathbf{X}_z'(2)\mathbf{X}_z'(3)]$, where

$$\mathbf{X}_z(1) = \begin{bmatrix} z_h & \cdots\cdots & z_1 \\ \vdots & & \vdots \\ z_{T-1} & \cdots\cdots & z_{T-h} \end{bmatrix},$$

$$\mathbf{X}_z(2) = \begin{bmatrix} z_T & \cdots\cdots & z_{T-h+1} \\ \vdots & & \vdots \\ z_{T+h-1} & \cdot\cdot & z_T \end{bmatrix},$$

and

$$\mathbf{X}_z(3) = \begin{bmatrix} z_{T+h} & \cdots & z_{T+1} \\ \vdots & & \vdots \\ z_{n-1} & \cdots & z_{n-h} \end{bmatrix},$$

then

$$\begin{aligned}(\hat{\mathbf{X}}_y'\hat{\mathbf{X}}_y) &= \mathbf{X}_z'\mathbf{X}_z + \hat{w}_T^2\mathbf{I} - \hat{w}_T(\mathbf{X}_z(2) + \mathbf{X}_z'(2)) \\ &= \mathbf{X}_z'\mathbf{X}_z - \hat{w}_T\mathbf{A}_T, \end{aligned} \tag{3.8}$$

where $\mathbf{A}_T = \mathbf{X}_z(2) + \mathbf{X}_z'(2) - \hat{w}_T\mathbf{I}$ is a symmetric matrix with $a_{ij} = a_{ji} = s(i) = z_{T+i} + z_{T-i}$ and $a_{ii} = z_T + \hat{z}_{T/n}$. Moreover, let us partition the vector $\mathbf{Z}$ accordingly as

$$\begin{aligned}\mathbf{Z}' &= [z_{h+1}, \ldots, z_T; z_{T+1}, \ldots, z_{T+h}; \\ &\qquad z_{T+h+1}, \ldots, z_n] \\ &= [\mathbf{Z}'(1)\mathbf{Z}'(2)\mathbf{Z}'(3)].\end{aligned}$$

Then, from (3.4) and (3.6), $\hat{\mathbf{X}}_y'\hat{\mathbf{Y}} = (\mathbf{X}_z - \hat{w}_T\mathbf{M})'(\mathbf{Z} - \hat{w}_T\mathbf{V})$ and, since $\mathbf{M}'\mathbf{V} = 0$, $\mathbf{M}'\mathbf{Z} = \mathbf{Z}(2)$, and $\mathbf{V}'\mathbf{X}_z = [z_{t-1}, \ldots, z_{T-h}]$,

$$\hat{\mathbf{X}}_y'\hat{\mathbf{Y}} = \mathbf{X}_z'\mathbf{Z} - \hat{w}_T\mathbf{S}_T, \tag{3.9}$$

where $\mathbf{S}_T' = (s(1), \ldots, s(h))$ and $s(j)$ is given by (3.3). Expressing the estimated parameters $\hat{\boldsymbol{\pi}}_T$ as a function of the observed data, for (3.8), (3.9), and (3.1), $(\mathbf{X}_z'\mathbf{X}_z - \hat{w}_T\mathbf{A}_T)\hat{\boldsymbol{\pi}}_T = \mathbf{X}_z'\mathbf{Z} - \hat{w}_T\mathbf{S}_T$, which leads to

$$\hat{\boldsymbol{\pi}}_T = \hat{\boldsymbol{\pi}} - \hat{w}_T(\mathbf{X}_z'\mathbf{X}_z)^{-1}(\mathbf{S}_T - \mathbf{A}_T\hat{\boldsymbol{\pi}}_T), \qquad (3.10)$$

calling $\hat{a}_{T+l}$ the residuals from Estimation (3.1),

$$\hat{a}_{T+l} = z_{T+l} - \hat{\pi}_{1,T}z_{T+l-1} - \cdots - \hat{\pi}_{l,T}(z_T - \hat{w}_T) - \cdots - \hat{\pi}_h z_{T+l-h},$$

and

$$\hat{b}_{T-l} = z_{T-l} - \hat{\pi}_{1,T}z_{T-l+1} - \cdots - \hat{\pi}_{l,T}z_T - \cdots - \hat{\pi}_{h,T}z_{T-l+h},$$

the backward residuals, then stating that $\hat{\mathbf{E}}_T' = [\hat{a}_{T+1} + \hat{b}_{T-1}, \ldots, \hat{a}_{T+h} + \hat{b}_{T-h}]$. It is clear that $\hat{\mathbf{E}}_T = \mathbf{S}_T - \mathbf{A}_T\hat{\boldsymbol{\pi}}_T$ is a vector of pseudoresiduals and that (3.10) can be written as

$$\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}_T + \hat{w}_T(\mathbf{X}_z'\mathbf{X}_z)^{-1}\hat{\mathbf{E}}_T. \qquad (3.11)$$

For instance, for the AR(1) model,

$$\hat{E}_T = y_{T+1} - \hat{\phi}_T\hat{y}_T + y_{T-1} - \hat{\phi}_T(\hat{y}_T + \hat{w}_T)$$
$$= s(1) - 2\hat{\phi}_T\hat{y}_T - \hat{\phi}_T\hat{w}_T$$

and, using (3.8), $\Sigma z_t^2 = \Sigma\hat{y}_t^2 + \hat{w}_T(2\hat{y}_T + \hat{w}_T)$; then $\hat{\phi} = (1 - \alpha)\hat{\phi}_T + \alpha(s(1) - 2\hat{\phi}_T\hat{y}_T)\hat{w}_T^{-1}$, where $\alpha = \hat{w}_T^2(\Sigma\hat{y}_T^2 + 2\hat{w}_T\hat{y}_T + \hat{w}_T^2)^{-1}$. Clearly, as $\hat{w}_T \to \infty$, $\alpha \to 1$ and $\hat{\phi} \to 0$. This result, that gross errors pull all of the autocorrelation coefficients and estimated parameters toward 0, was indicated by Martin and Jong (1977), Guttman and Tiao (1978), and Treadway (1978). See Peña and Sánchez-Albornoz (1983) for examples with economic data.

3.2 A Statistic to Measure Influential Outliers

A natural way to measure the influence of observation z_T is to relate it to the change in the parameter estimates when this observation is assumed to be an outlier. Since $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\pi}}_T$ are vectors, the usual way to measure their distance is to build a metric using some relevant positive semidefinite matrix. A natural selection is to use the variance–covariance matrix of either of these two estimated vectors and to build a Mahalanobis distance. To have a common ground to compare all of the observations, it seems more useful to choose the covariance matrix of the parameters, assuming no outliers (see Cook, Peña, and Weisberg 1987). Then

$$D_2(T) = \frac{(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_T)'(\mathbf{X}_z'\mathbf{X}_z)(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_T)}{h\hat{\sigma}_a^2}, \qquad (3.12)$$

where we have divided the distance by the dimension of the vectors involved, h, to have a proper standardization.

The statistic (3.12) can also be interpreted as measuring the change in the vector of one-step-ahead forecasts. Using the estimated parameters, assuming no outliers, this vector is $\hat{\mathbf{Z}} = \mathbf{X}_z\hat{\boldsymbol{\pi}}$, and using the parameters estimated assuming an additive outlier at T, $\hat{\mathbf{Z}}_T = \mathbf{X}_z\hat{\boldsymbol{\pi}}_T$. The Euclidean distance between both vectors of forecasts is

$$(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T)'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T) = (\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_T)'\mathbf{X}_z'\mathbf{X}_z(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_T), \qquad (3.13)$$

so $D_2(T)$ can also be interpreted as a standardized measure of the Euclidean distance between the vectors of one-step-ahead forecasts built with $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\pi}}_T$.

Using (3.11), the statistic can be written as

$$D_2(T) = \frac{\hat{w}_T^2}{\hat{\sigma}_a^2} \cdot \frac{\hat{\mathbf{E}}_T'(\mathbf{X}_z'\mathbf{X}_z)^{-1}\mathbf{E}_T}{h}. \qquad (3.14)$$

That shows clearly that the influence statistic depends on two factors; the first measures the size of the outlier relative to the innovation standard deviation, the second the relative size of the observations before and after the outlier.

The likelihood-ratio test to check for additive outliers is asymptotically equivalent (see Chang et al. 1988) to

$$\lambda_{A,T} = \hat{w}_T^2/(\hat{\sigma}_a^2(\hat{\Sigma}\pi_i^2)^{-1}),$$

so $D_2(T)$ can be written as a function of this statistic,

$$D_2(T) = \frac{\lambda_{A,T}^2}{(\Sigma\hat{\pi}_i^2)} \frac{\mathbf{E}_T'(\mathbf{X}_z'\mathbf{X}_z)^{-1}\hat{\mathbf{E}}_T}{h}. \qquad (3.15)$$

4. A MEASURE OF INFLUENCE FOR INNOVATIONAL OUTLIERS

The model (2.3) for innovational outliers can be written using the autoregressive approximation of Section 3.1,

$$z_t = \boldsymbol{\pi}_{(I)}'\mathbf{x}_t + w_{(I)}I_t^{(T)} + a_t, \qquad (4.1)$$

where $\boldsymbol{\pi}_{(I)}$ represents the vector of parameters assuming an innovational outlier $w_{(I)}$ at time T and, as before, $\mathbf{x}_t' = (z_{t-1}, \ldots, z_{t-h})$. This is a linear model with a dummy variable and, therefore, calling $\hat{\boldsymbol{\pi}}_{(I)}$ and $\hat{w}_{(I)}$ the ML estimators and $\hat{\boldsymbol{\pi}}$ the usual estimator if $w_{(I)} = 0$,

$$\hat{\boldsymbol{\pi}}_{(I)} = \hat{\boldsymbol{\pi}} + \hat{w}_{(I)}(\mathbf{X}_z'\mathbf{X}_z)^{-1}\mathbf{x}_T \qquad (4.2)$$

and

$$\hat{w}_{(I)} = (1 - d_T)^{-1}e_T, \qquad (4.3)$$

where $e_T = z_T - \hat{\boldsymbol{\pi}}'\mathbf{x}_T$ and $d_T = \mathbf{x}_T'(\mathbf{X}_z'\mathbf{X}_z)^{-1}\mathbf{x}_T$ is the distance between the vector of regressors at the time of the intervention, $\mathbf{x}_T$, and the origin. It is well known (Cook and Weisberg 1982) that $\hat{\boldsymbol{\pi}}_{(I)}$ is computed deleting the Tth row of the regressor matrix, $\mathbf{X}_z$, and the data vector, $\mathbf{Z}$. Therefore, the observation is not completely disregarded in the computation of the parameters as in the additive outlier case, because it appears as one of the regressors in the rows $T + 1$ to $T + h$, which do not change. This result shows the difference from the standard regression setup in which the deleting procedure leads to estimators that do not depend on the response variable.

As an example, consider the AR(1) case. Then, calling $\phi_{(I)}$ the parameter estimated assuming an innovational outlier at T,

$$\hat{\phi}_{(I)} = \left(\sum_{t\neq T} z_t z_{t-1}\right) \Big/ \left(\sum_{t\neq T} z_{t-i}^2\right),$$

and note that z_T appears in the numerator (in the term $z_{T+1}z_T$) as well as in the denominator. This estimator according to (4.2) also can be written as

$$\hat{\phi}_{(I)} = \hat{\phi} + \hat{w}_{(I)} \frac{z_{T-1}}{\Sigma z_{t-1}^2},$$

and $\hat{\phi}_{(I)}$ can be greater or smaller than $\hat{\phi}$, depending on the sign of z_{T-1}. When $n \to \infty$, $\hat{\phi}_{(I)} \to \hat{\phi}$, so $\hat{\phi}_{(I)}$ is a consistent estimator of ϕ. This result was first obtained by Mann and Wald (1943), and Martin and Jong (1977) showed that, although consistent, this estimator can be quite inefficient.

The change in the parameters can be measured by

$$D_1(T) = \frac{(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(I)})'\mathbf{X}_z'\mathbf{X}_z(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(I)})}{h\sigma_a^2},$$

and it is easy to see that this statistic can be written as

$$D_1(T) = \frac{1}{h} \frac{e_T^2}{\sigma_a^2(1 - d_T)} \left(\frac{d_T}{1 - d_T}\right),$$

which is identical to the statistic suggested by Cook (1977) to measure the effect of an observation on the parameters of a regression model. This statistic can be interpreted as the product of two terms; the first, $e_T^2\sigma_a^{-2}(1 - d_T)^{-1}$, is the standardized residual at the point of the intervention; the second, $d_T(1 - d_T)^{-1}$, represents the distance of x_T from the origin but with relation to a metric built without taking into account x_T. $D_1(T)$ can also be expressed as a function of the likelihood criteria advocated by Fox (1972) and Chang and Tiao (1983) to test for innovational outliers:

$$D_1(T) = \frac{\lambda_{I,T}^2}{h} \frac{d_T}{(1 - d_T)^2},$$

where $\lambda_{I,T} = \hat{w}_{(I)}/\hat{\sigma}_a$ is the likelihood-ratio test for observation Tth being an innovational outlier. Note that $D_1(T)$ now depends only on the relative values of the h previous observations before the intervention (the regressors at $t = T$), in contrast with $D_2(T)$ that depends on the observations after the intervention as well.

5. COMPUTATION OF THE STATISTICS

The two statistics for influence that we have derived can be written as

$$D_i(T) = \frac{(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_i)'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_i)}{h\sigma_a^2}, \qquad i = 1, 2, \qquad (5.1)$$

where $\hat{\mathbf{Z}}_1 = \mathbf{X}_z\hat{\boldsymbol{\pi}}_{(I)}$, $\hat{\mathbf{Z}}_2 = \mathbf{X}_z\hat{\boldsymbol{\pi}}_{(T)}$, and, therefore, to compute the statistics we only need the vector of forecasts using parameters computed assuming an innovational outlier ($\hat{\boldsymbol{\pi}}_{(I)}$) or additive outlier ($\hat{\boldsymbol{\pi}}_T$).

These forecasting vectors can be easily obtained with the ARIMA representation of the model. The autoregressive approximation that has been used throughout the article to develop the procedure and to show the nature of the statistics is not needed to compute them. Therefore, the order h need not be specified, and in (5.1) h should always be taken as the number of parameters estimated to compute the forecast $(p + q)$.

Note that the autoregressive approximation is needed to define in a meaningful way the change in the parameters in an ARIMA model. The parameters ϕ and ϑ of the ARIMA representation (2.1) cannot be used directly because of the possible near cancellation between the AR and MA structures. For instance, if we consider the models

$$M_1 : (1 - .70B)Z_t = (1 - .69B)a_t,$$

$$M_2 : (1 - .75B)Z_t = (1 - .10B)a_t,$$

$$M_3 : (1 - .11B)Z_t = (1 - .10B)a_t,$$

and metrics for the vector (ϕ, ϑ) like $D(ij) = (\phi_i - \phi_j)^2 + (\vartheta_i - \vartheta_j)^2$, model M_1 will be closer to M_2 than to M_3, although both M_1 and M_3 are almost white noise and M_2 has a strong autoregressive structure.

One might consider using the $\psi(B)$ weights given by $\Delta^d\phi(B)\psi(B) = \vartheta(B)$, but the ψ weights do not converge for nonstationary models, whereas the π weights are always well defined and allow a clear comparison with the regression setup.

In summary, these statistics can be computed in a routine way for any ARIMA model with any statistical package that has an intervention-analysis option. We only need to estimate the parameters, assuming an innovational or additive outlier at every point and afterward use these parameters to compute a forecast vector for the whole sample.

6. DISCUSSION

It has been shown that building a measure of influence for additive outliers leads to procedures based on interpolating the suspicious observation given the rest of the data, whereas the innovational outlier leads to a deleting approach similar to the regression setup.

The first approach seems much more relevant for dependent data, and, furthermore, computer simulations (Peña 1984) have shown that the statistic D_1 is very unstable, whereas the statistic D_2, introduced in Section 3, seems to be always able to identify points with strong influence on the estimated parameters. The superiority of D_2 to detect heterogeneity can be foreseen, because it is known that time series models are much more robust to innovational than to additive outliers.

These ideas can be extended in a straightforward way to deal with multiple cases. For instance, (3.12) is still valid if $\hat{\boldsymbol{\pi}}_T$ represents the parameter estimated when it is assumed that additive outliers occur at the set of

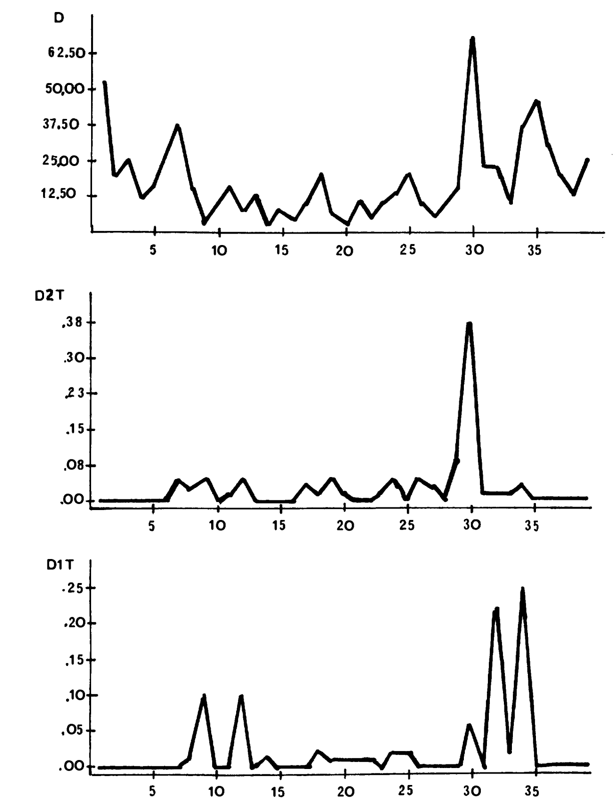

Figure 1. Plots of the Extinction-Rate Series, of the Diagnostic Statistics Recommended in the Article (D_{2T}), and of the Diagnostic Statistics for Innovational Outliers (D_{1T}).

points indicated by $T = (t_1, \ldots, t_n)$. On the other hand, the computation of the statistic becomes more difficult. A possible strategy is to first compute $D_2(T)$ for all of the individual points and all of the $\binom{n}{2}$ combinations. Then select only the $\alpha\%$ points (between 10% and 25%) with highest influence either individually or in pairs and compute the statistics for all of the 3, 4, . . . combinations of this small set of points. This strategy seems to be useful given our limited experience on the problem.

7. AN APPLICATION

The data that will be used to illustrate the previous statistics are the series of extinctions of marine animals over the past 250 million years displayed in Figure 1. These data were studied by Raup and Sepkoski (1984) and show periodicity in the peaks of extinctions that they attribute to deterministic extraterrestrial causes. Kitchell and Peña (1984) showed that the observed pseudoperiodicity can be explained by a fifth-order nonstationary autoregressive process with one root equal to 1 and four others complex.

Table 1 displays the original death-rate series, the residuals of the best estimated model, and the values of the D_1 and D_2 statistics. Both of those influence statistics are plotted in Figure 1. It can be seen that D_1 is more unstable than D_2; besides, D_1 fails to indicate the influential points clearly and shows peaks in the 32th and 34th observations. D_2 pinpoints observation 30 without any doubt, however. The atypical value of this observation is clear from Figure 1, and the residual at this point is outstanding and bigger than three standard deviations. The small value of D_2 for this point (.389) suggests, however, that this observation is not very influential as far as the parameter values are concerned. So, although there are only 39 observations, the autoregressive model is very robust to the effect of a single outlier.

Model A of Table 2 presents the estimated autoregressive model with and without outliers. As the data are proportions, different transformations have been used to test the sensitivity to the metric of the data. Model B, Table 2, presents the estimated models for the raw data and for transformation $y_t = \ln z_t/(1 - z_t)$. The results are very similar, and the same holds for other possible transformations that have been applied.

ACKNOWLEDGMENTS

I am grateful to George E. P. Box and George C. Tiao for their comments and encouragement, and I am

Table 1. Original Series (Z_t), Estimated Residual From the AR(5) Model (a_t), and Statistics $D_1(T)$ and $D_2(T)$

Observation	Z_t	a_t	$D_1(T)$	$D_2(T)$
1	52.500	—	.000	.000
2	21.000	—	.000	.000
3	24.000	—	.000	.000
4	12.800	—	.000	.000
5	15.900	—	.000	.000
6	26.400	−.142	.002	.005
7	38.600	.703	.005	.051
8	15.900	−.643	.011	.030
9	2.600	−1.867	.104	.044
10	10.100	.140	.001	.000
11	15.200	−.306	.004	.021
12	7.100	−1.429	.097	.047
13	11.600	.361	.003	.004
14	3.500	.567	.006	.003
15	7.600	.014	.000	.000
16	6.000	−.527	.005	.006
17	9.800	.289	.002	.028
18	19.500	.949	.016	.012
19	3.900	−.823	.014	.047
20	3.600	−.479	.005	.010
21	9.500	.650	.011	.000
22	6.000	−.679	.010	.003
23	10.200	.087	.000	.008
24	12.000	.859	.016	.051
25	18.900	1.111	.021	.000
26	9.900	−.167	.000	.047
27	5.800	−.374	.001	.031
28	9.200	.107	.000	.001
29	14.700	.257	.001	.087
30	66.300	2.384	.056	.381
31	22.200	−.054	.001	.011
32	21.900	.760	.216	.022
33	11.100	−.230	.020	.019
34	36.700	.851	.250	.033
35	45.800	.031	.000	.000
36	29.400	.080	.000	.000
37	20.000	−.114	.000	.000
38	12.500	−.417	.000	.000
39	25.000	−.076	.000	.000

Table 2. The Estimated Autoregressive Model With and Without Outliers

Model	Q(11)	$\hat{\sigma}_a^2$
A		
$(1 + .66B + .56B^2 + .71B^3 + .38B^4)\nabla Z_t = a_t$	6.2	122.8
$Z_t = \underset{(6.71)}{42.68} I_{30} + \dfrac{a_t}{\nabla(1 + \underset{(1.94)}{.32B} + \underset{(3.75)}{.56B^2} + \underset{(2.71)}{.45B^3} + \underset{(1.55)}{.22B^4})}$	4.1	63.65
B		
$(1 + \underset{(4.0)}{.62B} + \underset{(3.8)}{.59B^2} + \underset{(3.9)}{.62B^3} + \underset{(2.7)}{.42B^4})\nabla y_t = a_t$	7.9	.574
$y_t = \underset{(3.68)}{2.05} I_{30} + \dfrac{a_t}{\nabla(1 + \underset{(3.1)}{.49B} + \underset{(4.2)}{.65B^2} + \underset{(2.5)}{.43B^3} + \underset{(2.7)}{.40B^4})}$	5.0	.440

NOTE: Z_t is the extinction-rate series and Y_t is its logit transformation $[Y_t = \ln Z_t/(100 - Z_t)]$, B is the backshift operator, $\nabla = 1 - B$, I_{30} is an indicator variable with $I(30) = 1$ and $I(i) = 0$ for all $i \neq 30$, $Q(g)$ is the Ljung–Box statistic with g df, and $\hat{\sigma}_a^2$ is the residual variance of the model.

also indebted to Arthur B. Treadway, who taught me the importance of outliers in time series models. This research was accomplished when I was a visiting professor at the Mathematics Research Center, University of Wisconsin–Madison.

[*Received May 1987. Revised October 1989.*]

REFERENCES

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

Belsley, D. A., Kuh, W., and Welsch, R. E. (1980), *Regression Diagnostics*, New York: John Wiley.

Box, G. E. P., and Tiao, G. C. (1975), "Intervention Analysis With Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70–79.

Brillinger, D. (1986), Discussion of "Influence Functional for Time Series," by R. D. Martin and V. J. Yohai, *The Annals of Statistics*, 14, 781–788.

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

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.

Cook, R. D. (1977), "Detection of Influential Observations in Linear Regression," *Technometrics*, 19, 15–18.

Cook, R. D., Peña, D., and Weisberg, S. (1987), "The Likelihood Displacement: A Unifying Principle for Influence Measures," *Communications in Statistics—Theory and Methods*, 17, 623–640.

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

Denby, L., and Martin, R. D. (1979), "Robust Estimation of the First-Order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140–146.

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

Grenander, U., and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, New York: John Wiley.

Guttman, I., and Tiao, G. G. (1978), "Effect of Correlation on the Estimation of a Mean in the Presence of Spurious Observation," *The Canadian Journal of Statistics*, 6, 229–247.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 62, 1179–1186.

Kitchell, J. A., and Peña, D. (1984), "Periodicity of Extinctions in the Geological Past: Deterministic Versus Stochastic Explanations," *Science*, 226, 689–692.

Mann, H. B., and Wald, A. (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, 11, 173–220.

Martin, R. D. (1980), "Robust Estimation of Autoregressive Models," in *Directions in Time Series*, eds. D. Brillinger and G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics, pp. 228–254.

Martin, R. D., and Jong, J. (1977), "Asymptotic Properties of Robust Generalized *M*-Estimates for the First Order Autoregressive Parameter," memorandum, Bell Laboratories, Murray Hill, NJ.

Martin, R. D., Samarov, A., and Vandaele, W. (1983), "Robust Methods for ARIMA Models," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Bureau of the Census, pp. 153–169.

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

Peña, D. (1984), "Influential Observations in Time Series," Technical Summary Report 2718, University of Wisconsin–Madison, Mathematics Research Center.

Peña, D., and Guttman, I. (1988), "Bayesian Approach to Robustifying the Kalman Filter," in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J. C. Spall, New York: Marcel Dekker, pp. 227–254.

——— (1989), "Optimal Collapsing of Mixture Distributions in Robust Recursive Estimation," *Communications in Statistics—Theory and Methods*, 18, 817–833.

Peña, D., and Sánchez-Albornoz, N. (1983), *Dependencia Dinámica Entre Precios Agrícolas: El Trigo en España, 1857–1890*, Madrid: Servicio de Estudios del Banco de España.

Pregibon, D. (1981), "Logistic Regression Diagnostics," *The Annals of Statistics*, 9, 705–724.

Raup, D. M., and Sepkoski, J. J. (1984), "Periodicity of Extinctions in the Geological Part," *Proceedings of the National Academy of Science*, 81, 801.

Treadway, A. B. (1978), *Efectos Sobre la Economía Española de una Devaluación de la Peseta*, Madrid: Fundación Ramón-Areces.

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

——— (1988), "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7, 1–20.