

Working Paper 91-06
June 1990

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

MEASURING INFLUENCE IN DYNAMIC
REGRESSION MODELS

Daniel Peña*

Abstract

This article presents a methodology to build measures of influence in regression models with time series data. We introduce statistics that measure the influence of each observation on the parameter estimates and on the forecasts. These statistics take into account the autocorrelation of the sample. The first statistic can be decomposed to measure the change in the univariate ARIMA parameters, the transfer function parameters and the interaction between both. For independent data they reduce to the D statistics considered by Cook in the standard regression model. These statistics can be easily computed using standard time series software. Their performance is analyzed in an example in which they seem to be useful to identify important events, such as additive outliers and trend shifts, in time series data.

Key words

Missing observations; Outliers, Intervention analysis; ARIMA models; Inverse autocorrelation function.

* Departamento de Economía, Universidad Carlos III de Madrid and Laboratorio de Estadística, ETSII, Universidad Politécnica de Madrid.

1. INTRODUCTION

The distinction between additive and innovational outliers in time series was introduced by Fox (1972). Since then, the study of outliers has been an active area of research in time series. Abraham and Box (1979), Martin (1980), Chang, Tiao and Chen (1988), Tsay (1986, 1988) and Abraham and Chuang (1989) are some of the relevant references. The study of influential observations in time series, however, has received little attention in the statistical literature. Peña (1987, 1990) showed that the study of influential observations can be carried out for univariate time series using a missing value approach. We can use any suitable measure of distance, such as the likelihood distance or the change in the predictive or posterior distribution, to measure the model-change when it is assumed that one observation, or a subset of observations, is missing.

In this paper, the missing value approach is applied to build measures of influence for dynamic (transfer function) regression models. When the residuals are not autocorrelated, and therefore we are in the particular case of the standard regression model, the statistic suggested reduces to Cook's D. This paper is organized as follows: section 2 analyzes the problem of building a measure of change in the ARMA parameters; section 3 presents a global measure for the change of the whole model. It is shown that this measure can be decomposed to study the effect on the transfer function parameters and the ARIMA parameters. Section 4 applies these statistics to a dynamic system.

2. INFLUENCE ANALYSIS IN UNIVARIATE ARIMA MODELS

2.1 A measure of influence in univariate ARIMA models

Suppose the ARIMA model

$$\phi(B)\nabla^d z_t = \theta(B)a_t \quad (2.1)$$

where $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ are

polynomial operators in the backshift operator B with roots outside the unit circle, and $\nabla = 1 - B$ is the difference operator. It is also assumed here that the stationary series $\nabla^d z_t$ has zero mean, and a_t is a white noise sequence of iid $N(0, \sigma_a^2)$ variables. Let $k = p+q$ be the number of parameters. Let us call $\hat{\phi}$ and $\hat{\theta}$ the maximum likelihood estimators of the parameters with the complete data set and $\hat{\phi}_{(i)}$ and $\hat{\theta}_{(i)}$ those when observation z_i is missing (these computations will be discussed in the next section, 5). Then, the parameter estimates $\hat{\pi}$ and $\hat{\pi}_{(i)}$ of the autoregressive representation

$$\pi(B) z_t = a_t, \quad (2.2)$$

can be computed from

$$\nabla^d \phi(B) = \theta(B) \pi(B), \quad (2.3)$$

and the change in the model structure can be measured by a Mahalanobis type distance

$$P_i(\pi) = \frac{(\hat{\pi} - \hat{\pi}_{(i)})' \hat{\Sigma}_\pi^{-1} (\hat{\pi} - \hat{\pi}_{(i)})}{k \sigma_a^2}, \quad (2.4)$$

where $\hat{\Sigma}_\pi \hat{\sigma}_a^2$ is the variance covariance matrix of the maximum likelihood estimator $\hat{\pi}$. As shown in Peña (1990), by building the measure of change (2.4) using the π parameters, we avoid the problems linked to near cancellation between AR and MA structures. For large samples, approximately

$$\hat{\Sigma}_\pi = (Z_{t-1}' Z_{t-1})^{-1},$$

where Z_{t-1} is the matrix

$$Z_{t-1} = \begin{vmatrix} z_k & \cdots & z_1 \\ \vdots & & \vdots \\ z_{t-1} & \cdots & z_{t-k} \end{vmatrix}.$$

Now, let $\hat{Z}_t = Z_{t-1}\hat{\pi}$ be the vector of forecasts and $\hat{Z}_t^{(i)} = Z_{t-1}\hat{\pi}_{(i)}$ be the vector of forecasts using $\hat{\pi}_{(i)}$ instead of $\hat{\pi}$. Then

$$(\hat{Z}_t - \hat{Z}_t^{(i)})'(\hat{Z}_t - \hat{Z}_t^{(i)}) = (\hat{\pi} - \hat{\pi}_{(i)})' \hat{\Sigma}_{\pi}^{-1} (\hat{\pi} - \hat{\pi}_{(i)}). \quad (2.5)$$

Therefore, (2.4) can be written as

$$P_i(\pi) = \frac{(\hat{Z}_t - \hat{Z}_t^{(i)})'(\hat{Z}_t - \hat{Z}_t^{(i)})}{k \hat{\sigma}_a^2}. \quad (2.6)$$

Equation (2.6) shows that, as in the regression model, the Mahalanobis distance that measures the change in the parameters can be expressed as the Euclidean distance of the change in some forecast vectors. However, in the regression model $\hat{Z}_t^{(i)}$ is computed by replacing the i th observation by its mean, whereas here $\hat{Z}_t^{(i)}$ is computed by replacing the i th observation with its conditional expectation given the rest of the data. Peña (1990) has shown that the statistic (2.6) can be written as a function of the likelihood-ratio test to check for additive outliers.

2.2 Computing univariate influence

The computation of (2.6) requires the estimation of the parameters when one observation is missing. Jones (1980), Harvey and Pierce (1984), and Kohn and Ansley (1986) have shown how to solve this problem by setting up the model in state space form and applying the Kalman filter. An alternative method is the following. The conditional likelihood function of the parameters when observation z_i is missing can be written (see Ljung 1982) as

$$L(\beta, \sigma^2 | Z_{(i)}, u) = -\frac{(n-1)}{2} \ln \sigma^2 - \frac{1}{2} \ln f_i - \frac{1}{2\sigma^2} S(\beta, u, \hat{Z}_{(i)} z_i/n), \quad (2.7)$$

where β is the vector of ϕ and θ parameters, $Z_{(i)}$ is the observed

data set without z_i , u is the vector of starting values, $f_i \sigma^2$ is the variance of the sample distribution of the estimator of the missing value given the rest of the data, $\hat{z}_{i/n}$ is the expected value of z_i given $Z_{(i)}$ and $S(\beta, u, Z_{(i)}, \hat{z}_{i/n})$ is the residual sum of squares given the vector of parameters and starting values, β, u , in a data set in which observation z_i has been substituted by $\hat{z}_{i/n}$.

This estimation can be carried out easily using intervention analysis (Box and Tiao 1975) as follows. Let

$$\pi(\beta) (z_t - w_i I_t^{(i)}) = a_t \quad (2.8)$$

be an intervention model where $I_t^{(i)}$ is an impulse variable that takes the value 1 at $t = i$. Then, it can be shown (Peña 1987) that the likelihood function for the parameters can be written as

$$L(\beta, \sigma^2 | Z, u) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} S(\beta, u, Z_{(i)}, \hat{z}_{i/n}). \quad (2.9)$$

Therefore, if the sample size is large, and the term $\ln f_i$ can be disregarded, the estimation of the parameters using (2.7) or (2.9) will be practically identical.

In summary, $\hat{\pi}(i)$ can be easily computed with any computer package that includes intervention analysis. The computation of the vector of forecasts in (2.6) depends on the starting values, but if the sample is large its effect will be negligible.

In the measure (2.6) the vector $\hat{z}_t^{(i)}$ does have z_i as one of its components, because $\hat{z}_t^{(i)} = Z_{t-1} \hat{\pi}(i)$ and Z_{t-1} includes it. If we were interested in the influence of z_i on the forecast generated from the model, we could monitor the change with a vector of forecasts that does not depend on the i th observation at all. In the vector of forecasts generated by the intervention model (2.8), observation z_i is always replaced by $\hat{z}_{i/n}$, and therefore, we will suggest as a measure of the change in forecast the statistic:

$$D_i(\hat{z}_t) = \frac{(\hat{z}_t - \hat{z}_t^{\text{INT}})'(\hat{z}_t - \hat{z}_t^{\text{INT}})}{k \sigma_a^2}, \quad (2.10)$$

where \hat{z}_t^{INT} is the vector of forecasts from the intervention model (2.8).

3. INFLUENCE IN TRANSFER FUNCTION MODELS

3.1 Statistics of influence.

Suppose now that we have an explanatory variable x_t for the time series y_t . The variable x_t can be either deterministic or stochastic, but we consider the standard case in which the inference is done conditional on the given values of x_t . Then, one can write the model as

$$y_t = \frac{m(B)}{\delta(B)} x_t + \frac{\theta(B)}{\phi(B)} a_t, \quad (3.1)$$

where $m(B) = m_0 + m_1 B + \dots + m_m B^m$ and $\delta(B) = (1 - \delta_1 B - \dots - \delta_a B^a)$ have roots outside the unit circle, or also as the standard lag regression equation

$$y_t = \sum \pi_j y_{t-j} + \sum \alpha_i x_{t-i} + a_t, \quad (3.2)$$

where $\alpha(B) = v(B)\pi(B)$. Now, model (3.2) can be written as

$$Y = Z \pi + T X v + U, \quad (3.3)$$

where Y is the $n \times 1$ vector of observations of the output series, Z is an $n \times h$ matrix of past values of the output, π is a vector $h \times 1$ of coefficients, T is a triangular matrix with coefficients

$$T = \begin{vmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ -\pi_1 & 1 & & & & & \\ -\pi_2 & -\pi_1 & 1 & & & & \\ \vdots & & & \cdot & & & \\ -\pi_p & & & & \cdot & & \\ \cdot & & & & & \cdot & \\ 0 & -\pi_p & \cdot & \cdot & \cdot & \cdot & -\pi_1 & 1 \end{vmatrix} \quad (3.4)$$

X is the $n \times s$ matrix of current and past values of the explanatory variable and s is the dimension of the v vector. Let $\hat{\pi}$ and \hat{v} be the maximum likelihood estimators in model (3.3). The forecast vector is

$$\hat{Y} = Z \hat{\pi} + \hat{T} X \hat{v}, \quad (3.5)$$

where \hat{T} is the estimated T matrix using $\hat{\pi}$. Suppose now that observation y_i is missing, and let $\hat{\pi}_{(i)}$, $\hat{v}_{(i)}$ be the maximum likelihood estimators of the parameters. (We will discuss their computation in the next section.) Then, the vector of forecasts using the new parameters is

$$\hat{Y}_{(i)} = Z \hat{\pi}_{(i)} + \hat{T}_{(i)} X \hat{v}_{(i)}. \quad (3.6)$$

Following the same principles used in the univariate case, we define as the global measure of influence

$$P_i(\beta) = \frac{(\hat{Y} - \hat{Y}_{(i)})' (\hat{Y} - \hat{Y}_{(i)})}{C \hat{\sigma}_a^2}, \quad (3.7)$$

Equation (3.7) shows that to compute $P_i(\beta)$ it is necessary to obtain only the vector of forecasts from the standard transfer function formulation (3.1). Therefore, the order h does not need to be specified, and C will be equal to $p + q + m + a$, the number of parameters used to compute the vector of forecasts.

3.2 A Decomposition of the Statistic.

An observation can be influential because it changes the parameters in the noise model, in the transfer function or in both. As it may be useful to identify these cases separately, we can decompose the statistic to show these effects. A possible decomposition is to use

$$(\hat{Y} - \hat{Y}_{(i)}) = Z(\hat{\pi} - \hat{\pi}_{(i)}) + \hat{T} X (\hat{v} - \hat{v}_{(i)}) + (\hat{T} - \hat{T}_{(i)}) X \hat{v}_{(i)}, \quad (3.8)$$

if we let $c_1 = (p+q)/C$, $c_2 = (m+a)/C$ be the proportion of parameters in the noise and transfer function structure, we may write

$$P_i(\beta) = c_1 P_i(\pi) + c_2 P_i(v|\pi) + INT_i(\pi, v), \quad (3.9)$$

where the term

$$P_i(\pi) = (\hat{\pi} - \hat{\pi}_{(i)})' Z' Z (\hat{\pi} - \hat{\pi}_{(i)}) / \hat{\sigma}_a^2 (p+q), \quad (3.10)$$

is a measure of the change in the univariate parameters. Letting $A^{-1} = T'T$, the second term can be written

$$P_i(v|\pi) = (\hat{v} - \hat{v}_{(i)})' (X'A^{-1}X) (\hat{v} - \hat{v}_{(i)}) / \hat{\sigma}_a^2 (m+a), \quad (3.11)$$

and is proportional to the Mahalanobis distance between \hat{v} and $\hat{v}_{(i)}$ (we will see in the next section that $\sigma_a^{-2} X' A^{-1} X$ is the inverse of the variance covariance matrix for the generalized maximum likelihood estimator \hat{v}), and represents the change in the transfer function parameters given the π parameters. Finally, the third term is

$$\begin{aligned} C \hat{\sigma}_a^2 INT_i(\pi v) &= 2(\hat{\pi} - \hat{\pi}_{(i)})' Z' (\hat{T} X \hat{v} - \hat{T}_{(i)} X \hat{v}_{(i)}) \\ &\quad + 2(\hat{v} - \hat{v}_{(i)})' X' \hat{T}' (\hat{T} - \hat{T}_{(i)}) X \hat{v}_{(i)}, \end{aligned} \quad (3.12)$$

and represents the interaction between the change in the transfer function and the noise parameters. This interaction term will in many cases be negative, because a decrease in the transfer function parameters will be linked to a change in the opposite direction of the noise model. We will see an example of this situation in section 4.

The breakdown of statistic (3.7) into its components shows clearly that: (1) if there are no explanatory variables in the model, (3.7) reduces to the univariate statistic previously defined, and (2) if the noise is white and therefore we have a regression model, (3.7) reduces to the D statistic introduced by Cook (1977) for the regression model.

3.3 Computing Diagnostics

To compute statistic (3.7) we need to estimate the parameters of the model when one observation is missing. To describe the nature of this estimation, let us write the model as

$$Y = X v + R, \quad (3.13)$$

where R is a vector of noise that follows a multivariate normal distribution with mean zero and positive definite covariance matrix $\sigma^2 A$, such that $A^{-1} = T'T$.

Then, the log likelihood corresponding to (3.13) is

$$-\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln |A| - \frac{1}{2\sigma^2} (Y - Xv)' A^{-1} (Y - Xv), \quad (3.14)$$

and, conditional on A , the generalized least squares estimator of v is

$$\hat{v} = (X'A^{-1}X)^{-1}X'A^{-1}Y, \quad (3.15)$$

with covariance matrix

$$\hat{\Sigma}(v) = (X'^{-1} A X) \sigma^2. \quad (3.16)$$

As the π parameters, and therefore the A matrix, are usually unknown, maximum likelihood estimation requires an iterative algorithm in which, given an initial π value, the matrix A is determined and an initial estimator of v computed with (3.15). Then, the error process R is generated using (3.13), and the usual univariate time series estimation method applied to R to produce a new value for π . The procedure is iterated until convergence.

Now, it is well known that when $A = I$, the identity matrix, and only one observation is missing, the estimation of the parameters can be obtained by including a dummy variable in the regression model. We have shown in section 2.2 that, for large samples, this procedure works with univariate time series data. It will also work for model (3.13) if we consider the X values as fixed. Then, if we estimate the model

$$Y = w_i I_{(i)} + Xv_{(i)} + R, \quad (3.17)$$

where $I_{(i)}$ has a one in the i th position and zeros elsewhere, $v_{(i)}$ will provide the estimator of the v parameters when y_i is missing. As the A matrix is, in general, unknown, the estimation of (3.17) will require an iterative algorithm. To describe its structure, let us first analyze the case in which the parameters π , and therefore the matrix A, are given. Then, using the Cholesky factorization

$$A^{-1} = \hat{T}'\hat{T},$$

and letting $\tilde{Y} = \hat{T}Y$, $\tilde{X} = \hat{T}X$, $\tilde{I}_{(i)} = \hat{T}I_{(i)}$, $U = \hat{T}R$, the model is

$$\tilde{Y} = w_i \tilde{I}_{(i)} + \tilde{X} v_{(i)} + U, \quad (3.18)$$

where now $E[UU'] = \sigma^2 I$. Then, it is shown in the Appendix that the least squares estimators of the parameters are

$$\hat{w}_i = b \bar{I}'_{(i)} (\bar{Y} - \bar{X} \hat{v}), \quad (3.19)$$

and

$$\hat{v}_{(i)} = \hat{v} - (\bar{X}'\bar{X})^{-1}\bar{X}' \bar{I}_{(i)} \hat{w}_i, \quad (3.20)$$

where

$$b = [\bar{I}'_{(i)} (\mathbf{I} - \bar{H}) \bar{I}_{(i)}]^{-1}, \quad (3.21)$$

and

$$\bar{H} = \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'. \quad (3.22)$$

To understand the meaning of these estimators, let us consider the simple dynamic regression model

$$Y_t = x_t \beta + \frac{a_t}{(1 - \phi B)}.$$

Then, the estimator $\beta_{(i)}$ for β given ϕ will be obtained from the model

$$Y_t = w_i I(i) + x_t \beta_{(i)} + \frac{a_t}{(1 - \phi B)}, \quad (3.23)$$

and it can be shown that it is given by

$$\hat{\beta}_{(i)} = \frac{S_{xy}(i|n)}{S_x(i|n)}, \quad (3.24)$$

where

$$S_{xy}(i|n) = \sum (\hat{Y}_t - \phi \hat{Y}_{t-1})(\hat{x}_t - \phi \hat{x}_{t-1}),$$

$$S_x(i|n) = \Sigma(\hat{x}_t - \phi \hat{x}_{t-1})^2,$$

and

$$\hat{y}_t = \begin{cases} y_t & t \neq i \\ \hat{y}_{i|n} = \frac{\phi}{(1 + \phi^2)} (y_{i+1} + y_{i-1}) & t = i, \end{cases}$$

and

$$\hat{x}_t = \begin{cases} x_t & t \neq i \\ \hat{x}_{i|n} = \frac{\phi}{(1 + \phi^2)} (x_{i+1} + x_{i-1}) & t = i. \end{cases}$$

Note that $\hat{y}_{i|n}$ and $\hat{x}_{i|n}$ are the optimal estimators (in the minimum mean square error sense) of the missing values y_i and x_i . Both are computed using the inverse autocorrelation function of the output (noise) model (see Peña 1990). Thus, (3.24) is the estimator of the parameter β using these interpolators instead of the observed values. Besides,

$$\hat{w}_i = y_i - (\hat{y}_{i|n} + \hat{\beta}_{(i)}(x_i - \hat{x}_{i|n})) \quad (3.25)$$

is the difference between the observed value y_i and its estimator using all the information contained in the sample.

When the π parameters are unknown, the ML estimator of all the parameters of the model when y_i is assumed to be missing can be computed as follows:

- (1) Assuming initial values for the parameters $\beta^{(0)} = (\hat{\pi}^{(0)}, \hat{\varphi}^{(0)})$ compute the estimator $\hat{\pi}_{(i)} | \hat{\varphi}^{(0)}$ for the univariate parameters of the noise, as shown in section 2.2.
- (2) Using $\hat{\pi}_{(i)} | \hat{\varphi}^{(0)}$ compute the ML estimator of the transfer function parameters $\hat{\varphi}_{(i)}$ using (3.19) and (3.20).
- (3) Iterate until convergence.

In every step (1) or (2), using the initial value for the parameters, the minimum mean square error interpolator for the missing value is computed using

$$\hat{z}_{i|n} = - \sum \hat{\rho}_{ik} (z_{i+1} + z_{i-1}) ,$$

where $\hat{z}_{i|n}$ can be $\hat{y}_{i|n}$ or $\hat{x}_{i|n}$ and in both cases $\hat{\rho}_{ik}$ is the inverse autocorrelation function of the output process computed using the current estimate of the parameters. Then, new series are constructed using $\hat{y}_{i|n}$ and $\hat{x}_{i|n}$ instead of (y_i, x_i) , and the maximum likelihood estimates are computed. These values provide a new estimator of ρ_{ik} , that is used to build new estimators for the values $\hat{y}_{i|n}$, $\hat{x}_{i|n}$. The iterations are repeated until convergence.

Note that although the nature of the estimators has been analyzed using the linearized or structural form of the model (3.2), all the computation should be carried out in the parsimonious representation (3.1). Hence, the orders h and s in (3.3) need not be specified. Thus, in practice we only need to estimate the model

$$Y_t = w_i I(i) + \frac{m(i)(B)}{\delta(i)(B)} x_t + \frac{\theta(i)(B)}{\phi(i)(B)} a_t , \quad (3.26)$$

and the vector $\hat{Y}_{(i)}$ needed to compute (3.7) can be obtained from the vector $\hat{Y}_{(i)}^{INT}$ of forecasts generated from model (3.26) using

$$\hat{Y}_{(i)} = \hat{Y}_{(i)}^{INT} - \hat{w}_i \hat{\phi}(i)(B) \hat{\delta}(i)(B) I(i) , \quad (3.27)$$

where \hat{w}_i , $\hat{\phi}(i)(B)$ and $\hat{\delta}(i)(B)$ come from the estimation of (3.26). The values $\hat{\pi}(i)$ and $\hat{\nu}(i)$ needed in the computation of statistics (3.10) and (3.11) can be obtained using the relations

$$\hat{\delta}(i)(B) \hat{\nu}(i)(B) = \hat{m}(i)(B) ,$$

and

$$\hat{\theta}_{(i)}(B) \hat{\pi}_{(i)}(B) = \hat{\phi}_{(i)}(B) .$$

The statistic (3.7) is designed to measure the change in the parameters of the model, and is based on the vector of forecasts using the vector of parameters $\hat{\pi}_{(i)}$, $\hat{\varphi}_{(i)}$ with the sample data. An alternative measure that takes into account the change in the forecast is

$$D_i(\hat{Y}) = \frac{(\hat{Y} - \hat{Y}_{(i)}^{INT})' (\hat{Y} - \hat{Y}_{(i)}^{INT})}{k \hat{\sigma}_a^2} , \quad (3.28)$$

where $\hat{Y}_{(i)}^{INT}$ is the vector of forecast generated by the intervention model (3.26). This measure reduces to (2.10) if there were no explanatory variables in the model.

In summary, the computation of (3.7) and (3.28) can be carried out with a program that computes ML estimators for the parameters of a transfer function model. It is only necessary to introduce a dummy variable (an intervention impulse variable) at every point, estimate the model and compute the vector of forecasts. Then, to obtain the statistic (3.7) we need the vector of forecasts (3.27), whereas to obtain the statistic (3.28) we use directly the vector of forecasts from the intervention model.

3.4 Finding Influential Points

We can say that a point is influential at level α if the parameters estimated using a modified sample in which this point is assumed to be missing are not included in the $1 - \alpha$ joint confidence region for the parameters estimated with the complete sample. For a time series model, the joint $1 - \alpha$ confidence interval for the vector of parameters β is given by (Priestley, 1981, p.369)

$$\frac{S(\beta) - S(\hat{\beta})}{p \hat{\sigma}_a^2} \leq F(p, n-p; 1-\alpha) ,$$

where $S(\beta)$ is the residual sum of squares for the parameter vector β , $\hat{\beta}$ the ML estimator, p the number of parameters, $\hat{\sigma}_a^2$ the residual variance and $F(p, n-p; 1-\alpha)$ the $1-\alpha$ percentile of the F distribution with p and $n-p$ degrees of freedom.

For large samples, statistics (2.6) and (3.7) can be compared with a X^2 distribution with k and C degrees of freedom. Then, if, for example, $P_i(\pi)$ equals the 0,25 value of the corresponding X^2 distribution, assuming that the i -th point is missing would move the estimate of π to the edge of the 0,25 joint confidence region for $\hat{\pi}$.

This reference distribution is only an approximation. However, we believe that the main usefulness of these statistics is as exploratory tools, and a plot of the values $P_i(\beta)$ over time will indicate if there are influential points and will suggest possible hypothesis to be tested.

4. AN EXAMPLE

To illustrate the previous procedures, we analyze a dynamic system represented by two series. The input series is the gas feed rate in a gas furnace, and the output is the CO_2 concentration. Box and Jenkins (1976) included 296 pairs of data points with a sampling interval of 9 seconds. To simplify the computations, we have selected a sample starting with the first observation and assuming that the system was sampled every 27 seconds and so taken one observation out of every three in the whole data set. From now on we will use this sample of 99 observations.

Table 4.1 gives the univariate models for these input (x_t) and output series (y_t). The transfer function modelling procedure in Box and Jenkins (1976) leads to

$$y_t = 53.38 - \begin{matrix} (1.27B + 1.76B^2) \\ (.16) \quad (.09) \quad (.09) \end{matrix} x_t + N_t \quad (4.1)$$

$$(1 - \begin{matrix} .78B + .20B^2 \\ (.10) \quad (.12) \end{matrix}) N_t = a_t \quad \hat{\sigma}_a = .68 .$$

MODEL	σ_a	Q(36)
$(1 - .93B + .48B^2 - .24B^3)x_t = a_t$.745	17.8
$(1 - 1.28B + .82B^2 - .36B^3)y_t = 9.21 + a_t$	1.641	24.2

Table 4.1

Models for the feed rate and CO_2 concentration (27 sec. sampling). Q(36) is the Ljung-Box statistic computed with 36 residual correlation coefficients.

Table 4.2 shows the largest values of the global influence measure (3.7) and its breakdown according to (3.9). These statistics have been computed using intervention analysis as follows. To compute $P_i(\beta)$ the following model was fitted:

$$y_t = w_0 I_t^{(i)} + c + (w_1 B + w_2 B^2)x_t + N_t ,$$

and

$$(1 - \phi_1 B - \phi_2 B^2)N_t = a_t .$$

Let $\hat{\beta}'_{(i)} = (c^{(i)}, w_1^{(i)}, w_2^{(i)}, \phi_1^{(i)}, \phi_2^{(i)})$ be the parameters estimated with this model and $\hat{Y}_{(i)}^{\text{INT}}$ the vector of forecasts generated with it, we compute

$$\hat{Y}_{(i)} = \hat{Y}_{(i)}^{\text{INT}} - \hat{w}_0 (1 - \hat{\phi}_1^{(i)} B - \hat{\phi}_2^{(i)} B^2) I_t^{(i)} ,$$

which is the vector of forecasts using the parameters $\hat{\beta}_t^{(i)}$ in the transfer function model. Then, $P_i(\beta)$ is computed using (3.7).

To compute the components of $P_i(\beta)$ we have ignored the covariance between the estimators that were small, and hence

$$P_i(\pi) = \sum_{j=1}^2 \frac{(\hat{\phi}_j - \hat{\phi}_j^{(i)})^2}{\hat{\sigma}(\hat{\phi}_j)},$$

and

$$P_i(v|\pi) = \sum_{j=1}^2 \frac{(\hat{w}_j - \hat{w}_j^{(i)})^2}{\hat{\sigma}(\hat{w}_j)}.$$

	Mean	Interq. Range	90	96	91
$P_i(\beta)$.004	.014	.93	.18	.14
$P_i(v \pi)$.001	.013	.99	.08	.12
$P_i(\pi)$.003	.011	.49	.22	.24
$D_i(Y)$.131	.315	5.05	1.03	1.16

Table 4.2

Distribution of the influence statistics in model (4.1). The three important points (90, 96, 91) are shown, as well as the mean and Interquartile range of the distribution of each influence statistic for the whole data set.

Table 4.2 shows the diagnostic statistics $P_i(\beta)$ and $D_i(\hat{Y})$ that are plotted in figures 1 and 2. It can be seen from them that y_{90} is the most influential point both on the parameters of the model

and on the forecasts. The Table also shows that in this case $P_i(\beta)$ and $D_i(\hat{Y})$ pinpoint the same observations, and although there is a difference in the scale in both measures, observation 90 is roughly five times the value of next large ones. However, the small value of $P_{90}(\beta)$ compared with a X^2 distribution with 5 degrees of freedom indicates that its effect on the parameter estimates is small, that is, it moves the estimator slightly within the joint confidence region for the parameters. This is confirmed by the breakdown of the statistic in the table. These results mean that the global structure of the model is robust to the sample. To illustrate this fact, Table 4.3 shows the parameters estimated assuming that y_{90} was missing, and it can be seen that their change is small.

The largest residuals from model (4.1) are displayed in Table (4.4) where it can be seen that all the larger values are concentrated at the end of the sample and after y_{90} . The plot of the residuals (Figure 3) shows some evidence of a change after y_{90} , and if we apply the outlier detection techniques developed by Chang, Tiao and Chen (1988) and Tsay (1986), observation 90 is identified as an additive outlier. (The size of the outlier is estimated as 2.12 with a t value of 4.13). Therefore, we conclude that observation 90 is an additive outlier, that it is the most influential point both on the parameters and on the forecast (figures 1 and 2), but also that its effect on the estimated parameters is small, as shown by tables 4.2 and 4.3.

	c	w_1	w_2	ϕ_1	ϕ_2	σ
whole sample	53	-1.27	-1.76	.77	-.20	.679
value 90 missing	53	-1.33	-1.65	.88	-.17	.608

Table 4.3

Parameters of the model with the whole sample and assuming that observation 90 is missing. The value of y_{90} is 54.5 and its optimal estimate is 52.15.

t	91	98	97	90	99
a_t/σ_a	-3.53	3.43	3.14	2.57	2.34

Table 4.4

Residuals greater than $2\sigma_a$ in the transfer function model.

The large residuals after observation 90 seem to indicate an important change after this point. Two hypotheses may be considered: the first is a level shift after this point, the second a trend shift. These effects may be modelled by a step function or a linear trend function after $t=90$. Therefore, we will estimate the models

$$\nabla y_t = \underset{(.06)}{.24} S_t^{(90)} - \underset{(.09)}{(1.29B + 1.78B^2)} \nabla x_t + \frac{\nabla a_t}{\underset{(.10)}{(1 - .65B + .22B^2)} \underset{(.12)}}{,} \quad (4.2)$$

and

$$y_t = \underset{(.10)}{53.29} + \underset{(.78)}{2.49} S_t^{(90)} - \underset{(.09)}{(1.31B + 1.82B^2)} x_t + \frac{a_t}{\underset{(.11)}{(1 - .60B + .28B^2)} \underset{(.12)}}{,} \quad (4.3)$$

where $S_t^{(90)}$ is a step function that takes the value one for $t \geq 90$ and zero before. Model (4.2) includes a linear trend after $t=90$, and model (4.3) a step at this point. The residual standard error of a_t in (4.2) is .629, and .638 in (4.3), and the residuals of model (4.2) after $t=90$ show a better behavior than those from model (4.3). Note again the robustness of the parameters of the model comparing (4.2) with the results of Table 4.3. As the

statistics of influence for model (4.2) keep showing an influential effect at $t = 90$ but nothing afterwards, we will add an impulse at this point in model (4.2), and we finally estimate the model

$$\nabla y_t = \frac{(.24 + 2.27\nabla^2)}{(.07)(.47)} S_t^{(90)} - \frac{(1.40B + 1.66B^2)}{(.07)(.08)} \nabla x_t + \frac{(.03)(1 - .96B)a_t}{(1 - .76B + .25B^2)}, \quad (4.4)$$

(.10)(.12)

that has variance .588. This model shows that an unusual event (represented by an impulse) occurred at $t = 90$ and produced a big effect at this point, and from then on a linear trend of .24 units every period was added to the series. Note again that the parameters for the noise and transfer function in model (4.1), and (4.4) are similar. Although the model for the noise in (4.4) seems very different from the one in (4.1), the π weights for these models are very similar.

Therefore, we can conclude that the relationship described for model (4.1) is well defined, that an intervention happened at time $t = 90$ producing an increasing trend in the output and making the process non-stationary, and that model (4.4) represents an adequate approximation to describe the effect of the intervention.

To study the effect of an anomalous event in the middle of the sample, let us assume now that in the original sample at points 40 and 41 an error of measurement is made, and instead of the value 59.4, (the same at both points) we observed 49.4. Let us call this series \tilde{y}_t . Then the estimated model is

$$\tilde{y}_t = 53.16 - \frac{(1.15B + 1.48B^2)}{(.22)(.18)} x_t + (1 - .66B + .28B^2)^{-1} a_t, \quad (4.5)$$

(.10) (.10)

and some relevant statistics for diagnosis are displayed in Table 4.5. It can be seen that the most influential point, as far as affecting either the parameters or the forecasts, is $t = 41$. However, although the effect on both the parameters of the transfer function ($P_i(v|\pi) = 8.25$) and the noise ($P_i(\pi) = 3.51$) is strong, (compared with a X^2 with 2 degrees of freedom) the global change is

not very large, because of compensation effects between both parts.

Note that the breakdown of the statistic $P_i(\beta)$ allows us to say that if the objective of the experimentation is to estimate the transfer function parameters we should conclude that model (4.5) is not robust, because observation 41 is able, by itself, to modify the transfer function parameters significantly.

	38	39	40	41	42	43
$P_i(\beta)$.49	.84	.71	2.39	1.1	.41
$P_i(v_i \pi_i)$.32	.10	.76	8.25	.26	.20
$P_i(\pi)$.64	1.34	1.34	3.51	2.05	.57
$\hat{D}_i(Y)$	2.18	4.77	4.27	5.28	4.20	-1.8
$a_t \sigma_a$	-.22	.55	-6.6	-2.54	3.33	.89

Table 4.5

Statistics of influence for model (4.5). All the other values are small.

5. CONCLUDING REMARKS

The identification of influential observations complements the study of outliers. As is well known from the standard regression set up, when the model includes explanatory variables it is possible to have highly influential points that are not identified as outliers. The importance of this analysis depends on the objectives of the study. When the model is built to interpret the parameters or to test the gain of the transfer function, we may want to know whether or not the conclusions we draw from the data

are very much affected by some small number of observations which may or not be outliers.

When a point is identified as influential in a given model, we should first check if this point is also an outlier. If it is, the usual procedure is to incorporate it into the model using dummy variables (see Tsay 1986, 1988). If the point is not an outlier, we face essentially the same problem that has been studied in standard regression with high leverage points. There is not enough information in which to verify or deny the given point based just on the data. The a priori knowledge of the problem under investigation must be used to choose the appropriate model. A wise strategy may be to keep both models -the one that includes the suspicious data and the one that assumes this point is missing- and to check them with new data as a means of finding a better model. In this latter case the study of influential observations may indicate the shadowy regions of present knowledge and suggest possible hypotheses to be explored.

ACKNOWLEDGMENTS

This research has been supported by the Grant PB87-0808 Direccion General de Investigación Científica y Técnica del Ministerio de Educación y Ciencia, Spain and by the Graduate School of Business at the University of Chicago. The author is grateful to an anonymous referee for helpful comments.

APPENDIX

The estimation of model (3.18) will be

$$\begin{vmatrix} \hat{w}_i \\ \hat{v}_{(i)} \end{vmatrix} = \begin{vmatrix} \tilde{I}'_{(i)} & \tilde{I}_{(i)} & \tilde{I}'_{(i)} & \tilde{X} \\ \tilde{X}' & \tilde{I}_{(i)} & \tilde{X}'\tilde{X} & \end{vmatrix}^{-1} \begin{vmatrix} \tilde{I}' & \tilde{Y} \\ \tilde{X}' & \tilde{Y} \end{vmatrix},$$

and using the expression for the inverse of a partitioned matrix

$$\begin{vmatrix} \hat{w}_i \\ \hat{v}_{(i)} \end{vmatrix} = \begin{vmatrix} b & -b\tilde{I}'_{(i)}\tilde{X}(\tilde{X}'\tilde{X})^{-1} \\ -(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{I}_{(i)}b & (\tilde{X}'\tilde{X})^{-1}A \end{vmatrix} \begin{vmatrix} \tilde{I}'_{(i)} & \tilde{Y} \\ \tilde{X}' & \tilde{Y} \end{vmatrix},$$

where

$$b = [\tilde{I}'_{(i)}(I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')\tilde{I}_{(i)}]$$

$$A = I + b\tilde{X}'\tilde{I}_{(i)}\tilde{I}'_{(i)}\tilde{X}(\tilde{X}'\tilde{X})^{-1},$$

and after some straightforward operations results (3.19) and (3.20) are obtained.

REFERENCES

- Abraham, B., and Box, G.E.P. (1979), "Bayesian analysis of some outlier problems in time series", Biometrika, 66, 2, 229-36.
- Abraham, B., and Chung, A. (1989), "Outlier detection and Time Series Modeling", Technometrics, 31, 2, 241-248.
- Box, G.E.P., and Jenkins, G.M. (1976), Time Series Analysis Forecasting and Control, Holden-Day.
- 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.
- Chang, I., Tiao, G.C., and Chen Ch. (1988), "Estimation of Time Series Parameters in the presence of outliers", Technometrics, 30, 193-204.
- Cleveland, W.S. (1972), "The inverse autocorrelation of a time series and their applications", Technometrics, 14, 277-293.
- Cook, D.R. (1977), "Detection of influential observations in linear regression", Technometrics, 19, 15-18.
- Fox, A.J. (1972), "Outliers in Time Series", Journal of the Royal Statistical Society, B, 34, 350-363.
- Harvey, A.C., and Pierce, R.G. (1984), "Estimating Missing observations in Economic time Series", Journal of the American Statistical Association, 76, 125-131.
- Jones, R.H. (1980), "Maximum likelihood Fitting of ARMA models to time series with missing observations", Technometrics, 22, 389-395.
- Kohn, R., and Ansley, C.F. (1986), "Estimation, Prediction and Interpolation for ARIMA models with missing data", Journal of the American Statistical Association, 81, 395, 751-761.
- Ljung, G.M. (1982), "The likelihood function for a stationary Gaussian autoregressive-moving average process with missing observations", Biometrika, 69, 1, 265-8.
- Martin, R.D. (1980), "Robust Estimation of Autoregressive Models", in Direction in Time Series, eds. D.R. Brillinger and G.C. Tiao. Hayward, CA: Institute of Mathematical Statistics.

Forecasting Influence

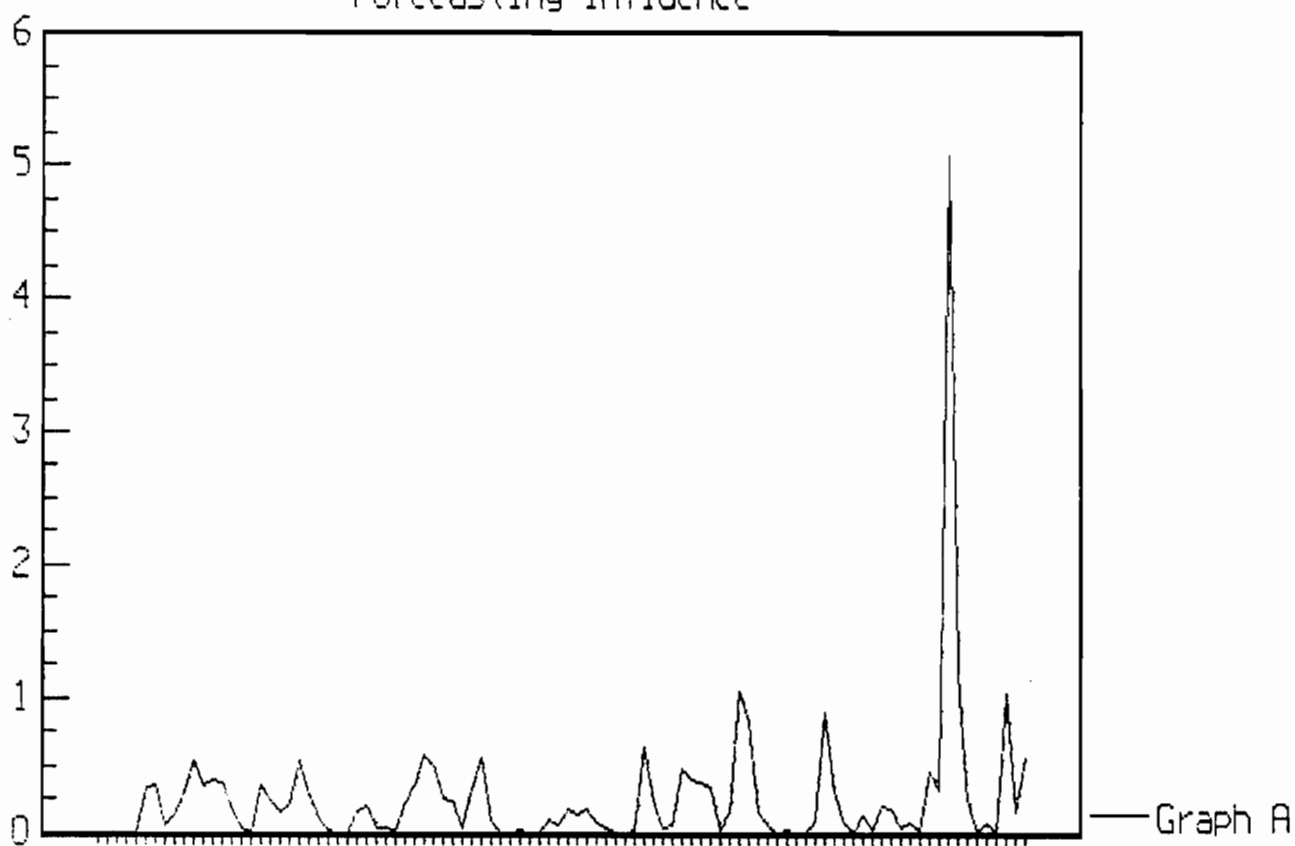


Figure 2

Plot of the statistic $D_i(Y)$ in model (4.1)

- Peña, D. (1987), "Measuring the importance of outliers in ARIMA models", in New Perspectives in Theoretical and Applied Statistics, eds. M. Puri et al, John Wiley, pp. 109-118.
- Peña, D. (1990), "Influential observations in time series", Journal of Business and Economic Statistics, 8, 2, 235-241.
- Priestley, M.B. (1981), Spectral Analysis and Time Series, Academic Press.
- Tsay, R.S. (1986), "Time series model specifications in the presence of outliers", Journal of the American Statistical Association, 31, 393, 132-141.
- Tsay, R.S. (1988), "Outliers, level shift and variance changes in time series", Journal of Forecasting, 7, 1-20.

Parameter Influence

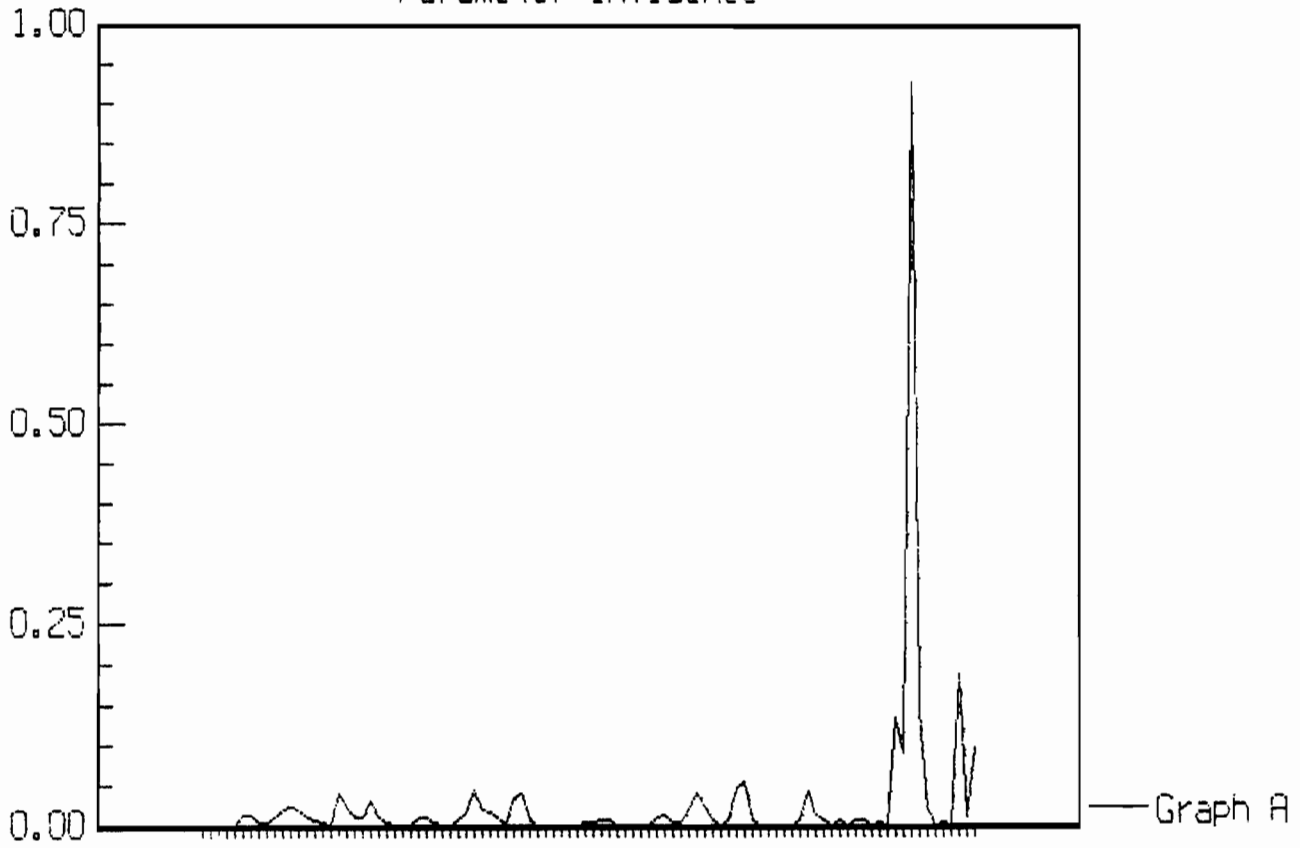


Figure 1

Plot of the statistic $P_i(\beta)$ in model (4.1)

Residuals transfer function model

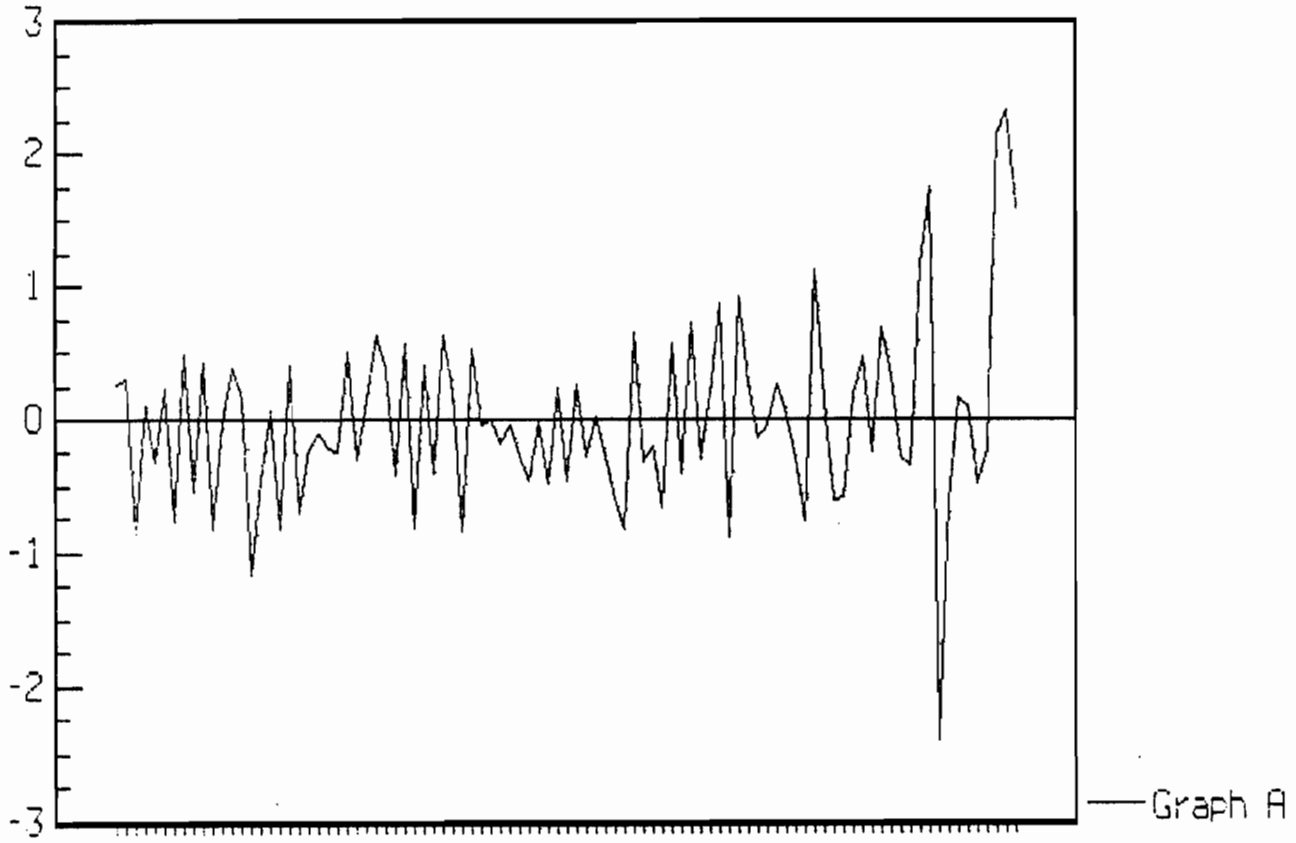


Figure 3

Plot of the residuals in model (4.1)