Diagnostics of a Multiresponse Regression Model with Autocorrelated Errors

Sibnarayan Guria

Department of Statistics,
West Bengal State University, Kolkata, India

Sugata Sen Roy*

Department of Statistics,
University of Calcutta, Kolkata, India

In this paper we study the diagnostics of a multiresponse regression model with a first-order autoregressive error sequence. The deletion technique is used to identify the outliers taking account of the dependence structure of the errors. Besides the usual measures, some scalar measures to gauge the impact of the ouliers on the regression are suggested.

Keywords: multiple responses, autoregressive errors, regression diagnostics.

1. Introduction

Ever since the early work of Cook (1977), extensive studies have been made on the diagnostics of multiple regression models with a single response using both the perturbation and the deletion techniques. Most of these studies were, however, applied to models with identically and independently distributed errors (refer Belsley, Kuh and Welsch (1980), Cook and Weisberg (1982) and Chatterjee and Hadi (1988) for a comprehensive discussion).

The studies were later extended to models with correlated error structures. In the generalized least-squares context, Putterman (1988) studied the influence of the first transformed observation on the parameter estimates. Schall and Dunne (1988) distinguished three types of outliers for the general linear model with arbitrary known variance-covariance structure and developed a unified approach to test for the presence of these outliers and a method of adjusting the parameter estimates when such outliers are present. Martin (1992) considered a linear regression model with completely specified error covariance structure and generalized the usual measures based on leverages and residuals. Cerioli and Riani (2002) proposed a new robust technique for the analysis of spatial data through simultaneous autoregressive models. Sen Roy and Guria (2004) applied the deletion technique to a regression model with first-order autoregressive errors and obtained explicit expressions of the diagnostic measures. Zewotir and Galpin (2007) extended the standard diagnostic tools to a linear mixed model.

---

*Corresponding author. E-mail: ssrstat@caluniv.ac.in

However, only a few sporadic studies have been made for multivariate models involving several responses. Such models can be of special importance especially in econometric and biostatistical studies where it is often necessary to model several related responses. Univariate modelling of each response in such cases would fail to capture the co-relationship between these variables and would hence lose vital information on the system as a whole. For such models, Barrett and Ling (1992) and Barrett (2003) suggested two general classes of influence measures based on the trace and determinant of a matrix which are functions of the hat and residual matrices. They, however, did not take account of the influence of the particular observation(s). A detailed study of the distribution of the Cook's Distance for elliptically distributed responses was made by Garcia and Farias (2004) and Garcia, Farias and Castro (2007). Rousseeuw, Aelst, Driessen and Agullo (2004) introduced the minimum covariance determinant estimator as a robust method for estimating the multivariate regression parameters and investigated the finite-sample performances of the estimator. All these studies, however, have assumed that the response vectors are independent.

Usually the deletion technique involves the deletion of a single observation and studying the impact it has on the regression. However, in many situations, as in a depression or a boom, consecutive time points may be outliers. In such cases, instead of deleting one observation at a time, it is desirable to see the impact that the deletion of this sequence of outliers has on the regression. Although studies on multiple outliers have been carried out by several authors like Rosner (1975) and Davies and Gather (1993), the deletion technique in the presence of autocorrelated errors has never been looked into.

In this paper we extend the results of the univariate autocorrelated models to a multi-response model with first-order autoregressive errors. The single observation deletion technique as applied by Sen Roy and Guria (2004) to the univariate model is further extended to the case where a sequence of consecutive observations are deleted. Expressions for *DFBETA*, *DFFIT* and Cook's Distance are then derived. Since these do not come out as simple numbers and hence are difficult to judge from, some scalar measures for identifying outliers are suggested.

The model is discussed in Section 2 while the main results are derived in Section 3. Section 4 contains some remarks and discussions and a numerical example is given in Section 5.

2. The model

Consider a model with q responses, each response being explained by the same set of p predictor variables. Given n observations, the jth response can then be modeled as

$$\mathbf{y}_j = \mathbf{X}\beta_j + \mathbf{u}_j, \qquad j = 1, \ldots, q \qquad (2.1)$$

where $\mathbf{X}$ is the $n \times p$ design matrix of rank p $(< n)$ and the elements of the error vector $\mathbf{u}_j$ follow a first-order autoregressive scheme with $E(\mathbf{u}_j) = \mathbf{0}$ and dispersion matrix

$$\mathbf{V} = \frac{\sigma_{jj}}{(1 - \rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ & & \cdots & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \qquad (2.2)$$

with $|\rho| < 1$.

Observing that the q responses for a given observation are correlated, (2.1) can be pooled as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \tag{2.3}$$

where $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_q)$, $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_q)$ and $\mathbf{B} = (\beta_1, \ldots, \beta_q)$. Thus

$$D(\mathbf{U}) = \Sigma \otimes \mathbf{V}, \tag{2.4}$$

where $\Sigma = ((\sigma_{ij}))$ is a positive definite scale matrix of dimension $q \times q$.
Observing that

$$Vec(Y) = (I_p \otimes \mathbf{X})Vec(\mathbf{B}) + Vec(U)$$

the best linear unbiased estimator of $\mathbf{B}$ is given by the generalized least-squares estimator

$$Vec(\hat{\mathbf{B}}) = [(I_p \otimes \mathbf{X})'(\Sigma \otimes \mathbf{V})^{-1}(I_p \otimes \mathbf{X})]^{-1}[(I_p \otimes \mathbf{X})'(\Sigma \otimes \mathbf{V})^{-1}Vec(Y)]$$

$$\text{or} \quad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{2.5}$$

$$\text{with} \quad D(\hat{\mathbf{B}}) = \Sigma \otimes (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{2.6}$$

It is interesting to note that $\hat{\mathbf{B}}$ depends on $\mathbf{V}$ but not on Σ which can thus be estimated as

$$\hat{\Sigma} = n^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}). \tag{2.7}$$

The regression diagnostics for this model, like the univariate models, will be based on the residual matrix $\mathbf{E} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}$ and the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$.

Since $\mathbf{V}$ is a positive definite matrix, there exists a nonsingular matrix $\mathbf{P} \ni \mathbf{V}^{-1} = \mathbf{P}'\mathbf{P}$. Therefore, defining $\mathbf{X}^* = \mathbf{PX}$, $\mathbf{Y}^* = \mathbf{PY}$ and $\mathbf{U}^* = P\mathbf{U}$,

$$\hat{\mathbf{B}} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^* \quad \text{and} \quad \hat{\Sigma} = n^{-1}(\mathbf{Y}^* - \mathbf{X}^*\hat{\mathbf{B}})'(\mathbf{Y}^* - \mathbf{X}^*\hat{\mathbf{B}}).$$

Thus when the diagnostic relates to the deletion of one or several observation(s) at a time and observing the effect on the estimators, the corresponding row(s) and column(s) of $\mathbf{V}$ must also be deleted. This would require that the corresponding column(s) of $\mathbf{P}$ is also deleted. Here we study the effect of this on the model (2.3)-(2.4).

3. Main results

Let $I = I(i)$ be the set of k consecutive observations indexed as $\{i, i+1, \ldots, i+k-1\}$, $i = 2, \ldots, n-k$. To see the effect of deleting these k observations on the estimated parameters, let $\mathbf{X}_I, \mathbf{Y}_I$ and $\mathbf{U}_I$ denote the design, response and disturbance matrices obtained from $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{U}$ after deleting these observations. The deletion of these observations will also induce a deletion of the corresponding rows and columns of $\mathbf{V}$ to yield the $(n-k) \times (n-k)$ matrix $\mathbf{V}_I$. We first observe the effect of this deletion on $\mathbf{V}^{-1}$ and $\mathbf{P}$.

Let $\phi_k = 1 + \rho^2 + \rho^4 + \cdots + \rho^{2k}$. Then

Result 3.1. For $i = 2, 3, \cdots, n-k-1$, $\mathbf{V}_I^{-1}$ is obtained from $\mathbf{V}^{-1}$ by deleting the rows and columns corresponding to I and replacing the (i-1, i-1) and (i+k, i+k) elements by $1 + \rho^{2k+2}/\phi_k$, and the (i-1, i+k) and (i+k, i-1) elements by $-\rho^{k+1}/\phi_k$. The corresponding $\mathbf{P}_I$ is obtained from $\mathbf{P}$ by deleting the rows and columns corresponding to I and replacing the (i+k, i-1) element by $-\rho^{(k+1)}/\sqrt{\phi_k}$ and the (i+k, i+k) element by $1/\sqrt{\phi_k}$.

Note. For $i = 1$ and $(n - k + 1)$, $\mathbf{V}_I$ will be of the same form as $\mathbf{V}$ except for a reduction in the dimension by k. Hence the corresponding $\mathbf{V}_I^{-1}$ and $\mathbf{P}_I$ will be the same as $\mathbf{V}^{-1}$ and $\mathbf{P}$ with k dimension less.

Proof. The proof follows by partitioning $\mathbf{V}_I$ as

$$\mathbf{V}_I = (1 - \rho^2)^{-1} \begin{bmatrix} 1 & \cdots & \rho^{i-2} & \rho^{i+k-1} & \cdots & \rho^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ \rho^{i-2} & \cdots & 1 & \rho^{k+1} & \cdots & \rho^{n-i-1} \\ \rho^{i+k-1} & \cdots & \rho^{k+1} & 1 & \cdots & \rho^{n-i+3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ \rho^{n-1} & \cdots & \rho^{n-i+1} & \rho^{n-i-k} & \cdots & 1 \end{bmatrix} \tag{3.1}$$

where the two block diagonal square matrices are of order $(i - 1)$ and $(n - k - i + 1)$ respectively, and then inverting the partitioned matrix using results from Rao (1973). $\qquad\square$

Defining $\mathbf{X}_I^* = \mathbf{P}_I \mathbf{X}_I$ and $\mathbf{Y}_I^* = \mathbf{P}_I \mathbf{Y}_I$, the estimator of $\mathbf{B}$ after the deletion of the observations corresponding to I can be obtained as

$$\hat{\mathbf{B}}_I = (\mathbf{X}_I' \mathbf{V}_I^{-1} \mathbf{X}_I)^{-1} \mathbf{X}_I' \mathbf{V}_I^{-1} \mathbf{Y}_I = (\mathbf{X}_I^{*'} \mathbf{X}_I^*)^{-1} \mathbf{X}_I^{*'} \mathbf{Y}_I^* \tag{3.2}$$

$$\text{and} \quad \hat{\Sigma}_I = (n - p - k)^{-1} (\mathbf{Y}_I^* - \mathbf{X}_I^* \hat{\mathbf{B}}_I)' (\mathbf{Y}_I^* - \mathbf{X}_I^* \hat{\mathbf{B}}_I). \tag{3.3}$$

Result 3.2. Under the above set-up,
  (i) $\hat{\mathbf{B}}_I$ is the best linear unbiased estimator of $\mathbf{B}$.
  (ii) Also if $n^{-1}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) \to \mathbf{M}$, where $\mathbf{M}$ is a p.d. matrix of constants,
$\hat{\mathbf{B}}_I$ and $\hat{\Sigma}_I$ are consistent for $\mathbf{B}$ and Σ respectively.

Note. $n^{-1}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) \to \mathbf{M}$ implies that $\hat{\mathbf{B}}$ is consistent for $\mathbf{B}$.

Proof. For the model, $\mathbf{Y}_I = \mathbf{X}_I \mathbf{B} + \mathbf{U}_I$, since $E(\mathbf{P}_I \mathbf{U}_I) = 0$ and $D(\mathbf{P}_I \mathbf{U}_I) = \sigma^2 I_n$, (i) follows from the Gauss-Markov Theorem.
Under $n^{-1}(\mathbf{X}_I^{*'} \mathbf{X}_I^*)^{-1} \to \mathbf{M}$ and since the second moments of $\mathbf{U}_I^*$ are finite, $n^{-1}(\mathbf{X}_I^{*'} \mathbf{U}_I^*) \to 0$. Hence

$$\hat{\mathbf{B}}_I = \hat{\mathbf{B}} + (\mathbf{X}_I^{*'} \mathbf{X}_I^*)^{-1}(\mathbf{X}_I^{*'} \mathbf{U}_I^*) \xrightarrow{P} \mathbf{B}$$

Again using Khinchine's Theorem, $n^{-1}(\mathbf{U}_I^{*'} \mathbf{U}_I^*)^{-1} \to \Sigma$, so that

$$\hat{\Sigma}_I = (\mathbf{U}_I^{*'} \mathbf{U}_I^*) - \mathbf{U}_I^{*'} \mathbf{X}_I^* (\mathbf{X}_I^{*'} \mathbf{X}_I^*)^{-1} \mathbf{X}_I^{*'} \mathbf{U}_I^* \xrightarrow{P} \Sigma. \quad \square$$

For each of the $(n - k)$ intermediate observations it is however not necessary to calculate $\hat{\mathbf{B}}_I$ afresh. This can more easily be obtained from the relationship between $\hat{\mathbf{B}}$ and $\hat{\mathbf{B}}_I$.
Let $\mathbf{w}_{i+j} = (\rho \mathbf{x}_{i+j}^* - \mathbf{x}_{i+j-1}^*)$ and $\mathbf{z}_{i+j} = (\rho \mathbf{y}_{i+j}^* - \mathbf{y}_{i+j-1}^*)$ for $j = 1, 2, \ldots, k$, and define $\mathbf{W}_I' = (\mathbf{w}_{i+1}, \ldots, \mathbf{w}_{i+k})'$ and $\mathbf{Z}_I = (\mathbf{z}_{i+1}, \ldots, \mathbf{z}_{i+k})'$. Also let $\mathbf{A}_I = ((a_{rs}))$ be a $k \times k$ matrix $\ni a_{rs} = a_{sr}$, and

$a_{k-s+1\,k-r+1} = a_{rs}$ so that only the the terms $a_{rs}, r \leq s$ and $r + s \leq k + 1$ are distinct. For these terms $a_{rs} = \phi_k^{-1} \sum_{j=|r-s|}^{k-1} a_{rsj} \rho^{2j-|r-s|}$ where

$$a_{rsj} = \begin{cases} min(j - |r-s| + 1, r) & if \;\; j \leq \left[\frac{k-|r-s|+1}{2}\right] \\ min(k - j, r) & if \;\; j > \left[\frac{k-|r-s|+1}{2}\right] \end{cases} \tag{3.4}$$

Result 3.3. For $I = \{i, \ldots, i + k - 1\}$, $i = 1, \ldots, (n - k - 1)$ and $\mathbf{E}_I = \mathbf{Z}_I - \mathbf{W}_I' \mathbf{B}$,

$$\hat{\mathbf{B}}_I = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}_I'(\mathbf{A}_I^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}_I')^{-1}\mathbf{E}_I \tag{3.5}$$

$$\text{and} \quad \hat{\Sigma}_I = (n - p - k)^{-1}[(n - p)\hat{\Sigma} - \mathbf{E}_I'(\mathbf{A}_I^{-1} - \mathbf{W}_I'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}_I)^{-1}\mathbf{E}_I]. \tag{3.6}$$

Proof. Let $\mathbf{x}_t^{*'}$ and $\mathbf{x}_{It}^{*'}$ denote respectively the tth rows of $\mathbf{X}^*$ and $\mathbf{X}_I^*$. Note that the $t = i, \ldots, (i + k - 1)$ rows of $\mathbf{X}_I^*$ are missing and that $\mathbf{x}_t^{*'} = \mathbf{x}_{It}^{*'}$ for the remaining rows except the $(i + k)^{th}$, which is

$$\mathbf{x}_{I(i+k)}^{*'} = \phi_k^{-1/2}(\mathbf{x}_{(i+k)}' - \rho^{(k+1)}\mathbf{x}_{(i-1)}')$$

$$= \phi_k^{-1/2} \sum_{t=i}^{i+k} \rho^{k+i-t}\mathbf{x}_t^{*'}. \tag{3.7}$$

Using (3.7) and simplifying

$$\mathbf{X}_I^{*'}\mathbf{X}_I^* = \sum_{t=1, \notin I}^{n} \mathbf{x}_{It}^*\mathbf{x}_{It}^{*'} = \mathbf{X}^*\mathbf{X}^{*'} - \sum_{t \in I} \mathbf{x}_t^*\mathbf{x}_t^{*'} - \mathbf{x}_{(i+k)}^*\mathbf{x}_{(i+k)}^{*'} + \mathbf{x}_{I(i+k)}^*\mathbf{x}_{I(i+k)}^{*'}$$

$$= \mathbf{X}^{*'}\mathbf{X}^* - \mathbf{W}_I\mathbf{A}_I\mathbf{W}_I'. \tag{3.8}$$

Similarly, $\qquad \mathbf{X}_I^{*'}\mathbf{Y}_I^* = \mathbf{X}^{*'}\mathbf{Y}^* - \mathbf{W}_I'\mathbf{A}_I\mathbf{Z}_I. \tag{3.9}$

Using (3.8) and (3.9) in (3.2) and simplifying, (3.5) follows.
Again since $\mathbf{Y}_I^{*'}\mathbf{Y}_I^* = \mathbf{Y}^{*'}\mathbf{Y}^* - \mathbf{Z}_I'\mathbf{A}_I\mathbf{Z}_I$, similar methods lead to (3.6). $\qquad \square$

4. Some remarks

Remark 1: *A single observation is deleted.* In most diagnostic studies observations are deleted one at a time to see their impact on the regression. This comes out as a special case of the above results with $k = 1$. So if $\mathbf{X}_i, \mathbf{Y}_i$ and $\mathbf{U}_i$ are $\mathbf{X}, \mathbf{Y}$ and $\mathbf{U}$ with the ith observation deleted,

$$\hat{\mathbf{B}}_i = (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i \tag{4.1}$$

$$\text{and} \quad \hat{\Sigma}_i = (n - p - 1)^{-1}\mathbf{E}_i'\mathbf{V}_i^{-1}\mathbf{E}_i, \tag{4.2}$$

where $\mathbf{E}_i = \mathbf{Y}_i - \mathbf{X}_i\hat{\mathbf{B}}_i$. Result 3.3 will then simplify as

$$\hat{\mathbf{B}}_i = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{w}_i\mathbf{E}_i'/[(1 + \rho^2)^{1/2}(1 + \mathbf{w}_i'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{w}_i)]. \tag{4.3}$$

Remark 2: *Unknown ρ.* Observe that the expressions in (3.5) and (3.6) depend on the parameter ρ, so that ρ needs to be known to calculate these. If $\hat{\mathbf{B}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the ordinary least squares estimator of $\mathbf{B}$, then defining the residuals as $e_{tl} = (y_{tl} - \mathbf{x}'_{tl}\hat{\mathbf{B}}_{OLS})$, an estimator of ρ is

$$r = (\sum_{t=2}^{n}\sum_{l=1}^{q} e_{tl}e_{\overline{t-1}\,l})/(\sum_{t=2}^{n}\sum_{l=1}^{q} e^2_{\overline{t-1}\,l}). \tag{4.4}$$

Of course, when the k observations are deleted ρ should be strictly estimated from the deleted set. However, as Result 3.1 suggests, this would introduce nonlinearity in the structure of u_{i+k} and would require weighted least squares for estimation. Moreover, no simple relation of the type (3.5) or (3.6) of r_I as obtained after the deletion can be found with r so that the regressions need to be run separately after each deletion thus defeating the whole purpose of computational reduction.

Remark 3: *The DFFIT and the Cook's Distance.* The expression for *DFBETA* is given by the second term on the right-hand-side of (3.5), which avoids re-doing the estimation afresh for each deletion. Denoting by $\mathbf{X}_{(I)}$ the $k \times p$ deleted set of explanatory variables, the corresponding expressions for *DFFIT* and Cook's Distance, for known ρ and Σ, can be easily derived as

$$\begin{aligned}
DFFIT_I &= \mathbf{X}_{(I)}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I) \\
&= \mathbf{X}_{(I)}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I(\mathbf{A}_I^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)^{-1}\mathbf{E}_I \tag{4.5}
\end{aligned}$$

$$\begin{aligned}
\text{and} \quad CD_I &= p^{-1}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)'[\Sigma^{-1} \otimes (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})](\hat{\mathbf{B}} - \hat{\mathbf{B}}_I) \\
&= p^{-1}\mathbf{E}'_I(\mathbf{A}_I^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)^{-1}\mathbf{W}'_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I \\
&\quad (\mathbf{A}_I^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)^{-1}\mathbf{E}_I. \tag{4.6}
\end{aligned}$$

Remark 4: *Standardized measures.* It is often appropriate to standardize the *DFBETA* and *DFFIT* values by the corresponding standard deviations (Neter, Wasserman & Kutner, 1985).

For $DFBETA_I$, an estimator of the dispersion of the full-set estimator of the regression parameter is often used as the norming factor in univariate modelling. Replacing ρ in (2.6) by r as obtained in (4.4), this gives for the multivariate case

$$\hat{Var}(\hat{\mathbf{B}}) = \hat{\Sigma} \otimes (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}. \tag{4.7}$$

Alternative estimators could be $\hat{\Sigma}_I \otimes (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$ or $\hat{\Sigma}_I \otimes (\mathbf{X}'_I\hat{\mathbf{V}}_I^{-1}\mathbf{X}_I)^{-1}$. The latter two would help in accentuating the impact of the outliers as the studentized residuals in a univariate regression model does. However, inspite of the minimality of computation that (3.6) would ensure, the recomputation of $\hat{\Sigma}_I$ for the sole purpose of standardization is unnecessary.

Corresponding to (4.7) an estimator of the dispersion for $DFFIT_I$ would be

$$\hat{Var}(\mathbf{X}_{(I)}\hat{\mathbf{B}}) = \hat{\Sigma} \otimes \mathbf{X}_{(I)}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'_{(I)}. \tag{4.8}$$

However, exact expressions for the dispersions of $DFBETA_I$ and $DFFIT_I$ can also be derived. Let $\Omega = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I(\mathbf{A}^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)^{-1}\mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$.

Result 4.1.

$$Var(DFBETA_I) = \Sigma \otimes \Omega \tag{4.9}$$

$$Var(DFFIT_I) = \Sigma \otimes \mathbf{X}_{(I)} \Omega \mathbf{X}'_{(I)}. \tag{4.10}$$

Proof. Define the $k \times (n-k)$ matrix $\mathbf{Q} = (\mathbf{q}_{i+1}, \dots, \mathbf{q}_{i+k})'$ where the n-dimensional vector $\mathbf{q}_{i+j}$ has -1 and ρ as their $(i+j-1)^{th}$ and $(i+j)^{th}$ elements respectively, all the remaining elements being zero, $j = 1, \dots, k$. It is then easy to see that $\mathbf{A}^{-1} = \mathbf{QQ}'$, $\mathbf{Z_I} = \mathbf{QY}^*$ and $\mathbf{W_I} = \mathbf{QX}^*$.

Therefore $\quad D(\mathbf{Z}_I) = (\mathbf{I}_q \otimes \mathbf{Q})(\Sigma \otimes \mathbf{I}_n)(\mathbf{I}_q \otimes \mathbf{Q}') = \Sigma \otimes \mathbf{A}^{-1}$

and $\quad\quad Cov(\mathbf{Z}_I, \mathbf{Y}^{*'}) = (\mathbf{I}_q \otimes \mathbf{Q})(\Sigma \otimes \mathbf{I}_n) = \Sigma \otimes \mathbf{Q}.$

Hence $\quad\quad Cov(\mathbf{Z}_I, \mathbf{Y}^{*'} \mathbf{X}^* (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I) = \Sigma \otimes \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I.$

Combining the above gives

$$D(\mathbf{E}_I) = \Sigma \otimes (\mathbf{A}^{-1} + \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I - 2\mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)$$
$$= \Sigma \otimes (\mathbf{A}^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I).$$

Writing $Vec(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I) = \mathbf{I} \otimes (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I(\mathbf{A}^{-1} - \mathbf{W}_I(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{W}'_I)^{-1}Vec(\mathbf{E}_I)$, the results follow. $\quad\square$

Remark 5: *Scalar measures to judge outliers.* It is easy to get $DFBETA_I$, $DFFIT_I$ and CD_I from (3.5), (4.5) and (4.6). However, these being matrices, it is difficult to infer the impact of the k observations from them. It is therefore necessary to find a scalar measure to judge this impact.

For $k = 1$, where the $DFFIT_i = \mathbf{x}'_i(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)$ is a q-dimensional vector, a natural measure would be the length

$$LDFFIT_i = [\mathbf{x}'_i(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)'\mathbf{x}_i]^{1/2},$$

a large value of which would indicate that the observation is an outlier. Alternatively, noting that for $k = 1$ (4.8) is $[\mathbf{x}'_i\Omega\mathbf{x}_i]\hat{\Sigma}$, the standardized length

$$SLDFFIT_i = [\mathbf{x}'_i\Omega\mathbf{x}_i)]^{-1/2}[\mathbf{x}'_i(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)\hat{\Sigma}^{-1}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)'\mathbf{x}_i]^{1/2}$$

may be used as a measure. Thus, generalizing for any k, the impact of k observations can be judged by the magnitude of

$$TDFFIT_I = [tr\{\mathbf{X}_{(I)}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)'\mathbf{X}'_{(I)}\}]^{1/2}.$$

The standardized version of this, provided $k \leq p$, is

$$STDFFIT_I = [Vec'\{\mathbf{X}_{(I)}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)\}\{\Sigma \otimes (\mathbf{X}_{(I)}\Omega\mathbf{X}'_{(I)})\}^{-1}Vec\{\mathbf{X}_{(I)}(\hat{\mathbf{B}} - \hat{\mathbf{B}}_I)\}]^{1/2}.$$

Remark 6: *More general form of the error dispersion matrix.* In this paper we have assumed in (2.4) a very simple structure of the error dispersion matrix. For one, ρ in (2.2) can vary for $j = 1, \dots, q$.

Secondly, the error sequences may follow more complex structures than the first-order autoregressive model that we have assumed. Such cases can be handled using the same techniques although the results are likely to be more involved with correspondingly greater complexity in the derivations.

Remark 7: *Masking and Swamping effects.* A major concern in multiple outlier models is the problem of masking, where several outliers, owing to their combined effects, may go undetected. Sometimes swamping or identification of non-outliers as outliers can also take place. Such issues have been addressed by several authors particularly while discussing either robust (see Cerioli and Riani, 2002) or distribution-based methods (see Davies and Gather, 1993). Although no rigourous study has been made on this in the present work, it can be conjectured that for consecutive outliers in a first-order autogressive setting, the possibility of masking will be minimal for high positive autocorrelation and maximum when the autocorrelation in negatively large. This, along with the choice of k, can be the subject of further studies in this area.

5. A numerical example

To illustrate our technique we consider the data on monthly gas and electricity usage in a gas-heated single-family residence in Boston, USA for the months September 1990 to May 1997. We take as our response variables the mean natural gas usage (in therms) per day for the month (y_1) and mean kilowatt hours per day for the month (y_2). The total heating degree days for the month (x) is taken as the independent variable. Since the gas readings for July-August and September-October were sometimes taken together, y_1 for 4 July and 6 September months are not given. We thus impute them with the corresponding August or October values. The data-set can be viewed at http://www.amstat.org/publications/jse/datasets/utility.dat.

Regressing y_1 and y_2 separately on x show the Durbin-Watson statistics to be 1.063015 and 1.468427 respectively, so that a sufficient degree of positive autocorrelation can be assumed. Also y_1 and y_2 show a positive correlation (0.519139) implying that a bivariate model is to be preferred over two independent univariate models. We thus use model (2.3) with $q = 2$ and $p = 2$ (including an intecept term) and calculate the $DFBETA$ and $DFFIT$ values for $k = 1, 2 (\leq p)$.
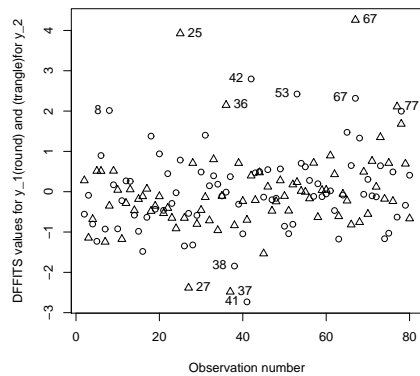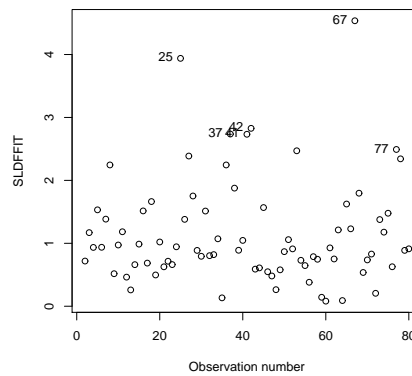

Fig. 1. *DFFIT*s for $k = 1$
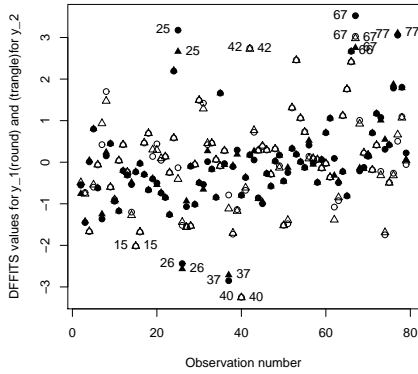

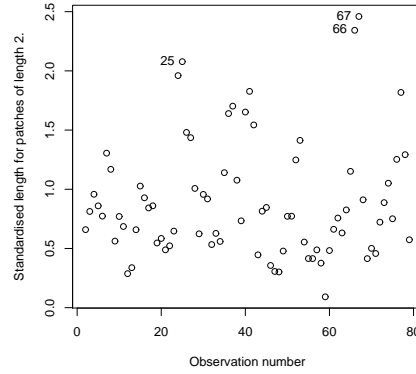Fig. 2. *SLDFFIT*s for $k = 1$

Fig. 3. *DFFIT*s for $k = 2$

Fig. 4. *SLDFFIT*s for $k = 2$

Figures 1 and 3 plot the standardized *DFFIT*s for $k = 1$ and 2. However, since for $k = 1$, the standardized *DFFIT* is a two-dimensional vector while for $k = 2$ it is a 2×2 matrix, plotting all elements (using o for y_1 and Δ for y_2 and *light* for j and *dark* for $(j+1)$), clutter the plots. Hence in Figures 2 and 4 we give the corresponding standardized lengths *SLDFFIT*, which gives a clearer visual of the deletion effects.

From the figures it is clear that the 67th and 25th observations are outliers. Some of the others like the 41st, 42nd and 77th observations are also suspect. However, these are generally isolated outliers and hence come out more strongly for $k = 1$. In fact, for $k = 2$ some of them $(41, 42)$ get masked, while others show up as two consecutive outliers, pulling up their previous and subsequent normal observations in their train.

References

[1] B.E. Barrett, Understanding influence in multivariate regression, *Comm. Statist.: Theory and Methods*, **32**, 3 (2003), 667–680.

[2] B.E .Barrett, and R.F. Ling, General classes of influence measures for multivariate regression, *J. Amer. Statist. Assoc.*, **87** (1992), 182–191.

[3] D.A. Belsley, E. Kuh, and R.E. Welsch, *Regression Diagnostics*, John Wiley, New York, 1980.

[4] A. Cerioli, and M. Riani, Robust methods for the analysis of spatially autocorrelated data, *Statist. Meth. & Appl.*, **11** (1980), 335–358.

[5] S. Chatterjee, and A.S. Hadi, *Sensitivity Analysis in Linear Regression*, John Wiley, New York, 1988.

[6] R.D. Cook, Detection of influential observations in linear regression, *Technometrics*, **19** (1977), 133–169.

[7] R.D. Cook, and S. Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, London, 1982.

[8] L. Davies, and U. Gather, The identification of multiple outliers, *J. Amer. Statist. Assoc.*, **88**, 423 (1993), 782–792.

[9] A.D. Garcia, and G.G. Farias, A note on the Cook Distance, *J. Statist. Plan. Inf.*, **120** (2004), 119–136.

[10] A.D. Garcia, G.G. Farias, and M.A. Castro, Exact distributions for sensitivity analysis in linear regression, *Appl. Math. Sci.*, **1**, 22 (2007), 1083–1100.

[11] R.J. Martin, Leverage, influence and residuals in regression models when observations are correlated, *Comm. Statist.: Theory and Methods*, **21**(1992), 1183–1212.

[12] J.W. Neter, W. Wasserman, and M.H. Kutner, *Applied Linear Statistical Models*, Homewood, Irwin, 1985.

[13] M.L. Putterman, Leverage and influence in autocorrelated regression models, *Appl. Statist.*, **37** (1988), 76–86.

[14] C.R. Rao, *Linear Statistical Inference and its Application*, John Wiley, New York, 1973.

[15] B. Rosner, On the detection of many outliers, *Technometrics*, **17** (1975), 221–227.

[16] P.J. Rousseeuw, S.V. Aelst, K.V. Driessen, and J. Agullo, Robust multivariate regression, *Technometrics*, **46**, 3 (2004), 293–305.

[17] R. Schall and T.T. Dunne, A unified approach to outliers in the general linear model, *Sankhya B*, **50** (1988), 157–167.

[18] S. Sen Roy, and S.N. Guria, Regression diagnostics in an autocorrelated model, *J. Braz. Prob. & Statist.*, **18** (2004), 103–112.

[19] T. Zewotir, and J.S. Galpin, A unified approach on residuals, leverages and outliers in the linear mixed model, *Tests*, **16**, (2007), 58–75.