International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 2, 2003, pp 11–17

A Conditional Approach to the Analysis of Residuals in Linear Regression

A. H. M. Rahmatullah Imon
Department of Statistics,
University of Rajshahi, Bangladesh
Email: imon_ru@yahoo.com

[Received June 14, 2002; Revised October 23, 2003; Accepted December 1, 2003]

Abstract

The analysis of residuals forms the basis of diagnostics in linear regression. The residuals are used as substitutes of the true random disturbances since the latter ones are unobserved. The most commonly used least squares residuals can be severely affected in the presence of outliers. This problem with the least squares residuals is not clearly understood from the expectation of the residuals. In this paper we show that a conditional expectation approach clearly exhibits this problem when conditioning a group of outlying observations. We also notice that residuals become even worse when mean shift outlier errors are used in the model. We observe, however, that the deletion residuals possess better conditional properties for both the standard model and the outlier model. Deletion residuals represent the true disturbances accurately irrespective of whether they correspond to the outlying observations or not.

Keywords and Phrases: Regression Diagnostics, Deletion Residuals, Mean Shift Outliers, Conditional Analysis.

1 Introduction

Let us consider a standard linear regression model

$$Y = X\beta + \in \tag{1}$$

where Y is an n-vector of observed responses, X is an $n \times k$ matrix representing k explanatory variables with full column rank, β is a k vector of unknown finite parameters and $\in$ is an n-vector of uncorrelated random disturbances with $E(\in) = 0$

and $V(\in) = \sigma^2 I$, where σ^2 is an unknown parameter and I is an identity matrix of order n. In fitting a model like (1), ordinary least squares (OLS) technique has been generally adopted because of tradition and ease of computation. Under usual assumptions, the OLS estimator $\hat{\beta} = (X^T X)^{-1} X^T Y$ possesses some nice and useful properties. The vector of fitted values can be expressed as $\hat{Y} = X\hat{\beta} = WY$, where the matrix $W = X(X^T X)^{-1} X^T$ is known as weight or leverage matrix. The OLS residual vector, denoted by $\hat{\in}$, can be expressed in terms of the vector of the true disturbance vector $\in$ as

$$\hat{\in} = Y - \hat{Y} = (I - W) \in \tag{2}$$

In fitting a linear regression model by the OLS method we often observe that a variety of estimates and residuals can be substantially affected in the presence of outliers. Observations corresponding to excessively large $\in$ values are termed as outliers. Rousseeuw and Leroy (1987), Barnett and Lewis (1994) and many others study different aspects of outliers in linear regression. Looking at the right hand side of (2), Weisberg (1980) pointed out that each of the OLS residuals could be expressed as the weighted average of the true disturbances. Therefore the presence of a single outlier not only effects the residual corresponding to that particular observation, the entire residual set can be distorted in the presence of that outlier. In the least squares theory we assume $E(\in) = 0$ which also leads to $E(\hat{\in}) = 0$ and thus we observe that the problem with the OLS residuals in the presence of outliers is not properly addressed while considering the unconditional expectation of the residuals. In section 2, we introduce deletion residuals that have extensive use in regression diagnostics [see Cook and Weisberg (1982), Chatterjee and Hadi (1988), Imon (2002)]. We also discuss mean shift outlier (MSO) model in this section which is frequently used [see Barnett and Lewis (1994)] in diagnostics. We observe in section 3 that under the conditional expectation approach, the performance of neither of the OLS residuals or the residuals obtained from the MSO model is satisfactory in the presence of outliers. However, the deletion residuals corresponding to all of the observations possess better properties conditioning on a group of observations.

2 Deletion Residuals and the Mean Shift Outlier Model

Here we define deletion residuals which are based on a fit to the data with a group of observations deleted. The full set of residuals is reestimated from the rest after deleting a group of observations. This type of residuals is also known as predictive residuals [Cook and Weisberg (1982)] in the literature. Let us denote a set of cases 'remaining' in the analysis by R and a set of cases 'deleted' by D. Let us also suppose that R contains $(n-d)$ cases after $d < (n-k)$ cases in D are deleted. Without loss of generality, assume that these observations are the last of d rows of X and Y so that they can be partitioned as

$$X = \begin{bmatrix} X_R \\ X_D \end{bmatrix}, \quad Y = \begin{bmatrix} Y_R \\ Y_D \end{bmatrix}$$

Then the weight matrix $W = X(X^T X)^{-1} X^T$ can be partitioned as

$$W = \begin{bmatrix} U_R & V \\ & \\ V^T & U_D \end{bmatrix}$$

where

$$U_R = X_R(X^T X)^{-1} X_R^T, \quad U_D = X_D(X^T X)^{-1} X_D^T \tag{3}$$

are symmetric matrices of order $(n-d)$ and d respectively, and

$$V = X_R(X^T X)^{-1} X_D^T \tag{4}$$

is an $(n-d) \times d$ matrix. Hence using the result of Chatterjee and Hadi (1988), the vector of estimated parameters after the deletion of d observations, denoted by $\hat{\beta}^{(-D)}$, is obtained as

$$\begin{aligned} \hat{\beta}^{(-D)} &= (X_R^T X_R)^{-1} X_R^T Y_R \\ &= \hat{\beta} - (X^T X)^{-1} X_D^T (I_D - U_D)^{-1} \hat{\in}_D \end{aligned} \tag{5}$$

assuming that $(I_D - U_D)$ is invertible and where $\hat{\in}_D = Y_D - X_D \hat{\beta}$. Thus an $n \times 1$ vector of deletion residuals can be defined as

$$\hat{\in}^{(-D)} = Y - X \hat{\beta}^{(-D)} \tag{6}$$

Using (5) and (6), the deletion residual vector can be expressed in terms of the OLS residual vector $\hat{\in}$, and the vector of OLS residuals as

$$\hat{\in}^{(-D)} = \hat{\in} + X(X^T X)^{-1} X_D^T (I_D - U_D)^{-1} \hat{\in}_D \tag{7}$$

We can partition $\hat{\in}^{(-D)}$ as

$$\hat{\in}^{(-D)} = \begin{bmatrix} \hat{\in}_R^{(-D)} \\ & \\ \hat{\in}_D^{(-D)} \end{bmatrix} = \begin{bmatrix} Y_R - X_R \hat{\beta}^{(-D)} \\ & \\ Y_D - X_D \hat{\beta}^{(-D)} \end{bmatrix}$$

that implies

$$\begin{aligned} \hat{\in}_R^{(-D)} &= \hat{\in}_R + V(I_D - U_D)^{-1} \hat{\in}_D \\ \hat{\in}_D^{(-D)} &= (I_D - U_D)^{-1} \hat{\in}_D \end{aligned} \tag{8}$$

It is also possible to express the different components of deletion residual vector $\hat{\in}^{(-D)}$ in terms of the vectors of true disturbances, $\in_R$ and $\in_D$, using the results of Chatterjee and Hadi (1988) and Imon (2002) as

$$\begin{aligned} \hat{\in}_R &= Y_R - X_R\hat{\beta} = (I_R - U_R) \in_R - V \in_D \\ \hat{\in}_D &= Y_D - X_D\hat{\beta} = (I_D - U_D) \in_D - V^T \in_R \end{aligned} \tag{9}$$

and

$$\begin{aligned} \hat{\in}_R^{(-D)} &= [I_R - U_R - V(I_D - U_D)^{-1}V^T] \in_R \\ \hat{\in}_D^{(-D)} &= \in_D - (I_D - U_D)^{-1}V^T \in_R \end{aligned} \tag{10}$$

At this point let us introduce a mean shift outlier (MSO) regression model. Anscombe (1960) first proposed the concept of mean shift outlier in statistical data. Here we assume that $x_1, x_2, \cdots, x_{n-1}$ is a random sample from a population with mean μ and variance σ^2, but an observation x_n comes from a population with mean $\mu + a$ and the same variance σ^2. But in a regression problem observations corresponding to unusual errors are termed as outliers. Therefore we consider a mean shift outlier model

$$Y = X\beta + \in^* \tag{11}$$

where $\in^* = \delta + \in$, δ is an $n \times 1$ vector with $(n-d)$ zeros and d unknown non-zero values corresponding to the d mean shift errors (outliers). For this model we can partition the vector of true errors as

$$\in^* = \begin{pmatrix} \in_R + \delta_R \\ \in_D + \delta_D \end{pmatrix} = \begin{pmatrix} \in_R \\ \in_D + \delta_D \end{pmatrix} \tag{12}$$

so that

$$E(\in^*) = \begin{pmatrix} 0_R \\ \delta_D \end{pmatrix}. \tag{13}$$

Gentleman and Wilk (1975) suggest this type of model that has extensive use in the study of the identification and handling of outliers in linear regression. The OLS residuals obtained from this MSO model for the two sets of observations R and D are denoted by $\hat{\in}_R^*$ and $\hat{\in}_D^*$ respectively. Using (9) and (12) we obtain

$$\begin{aligned} \hat{\in}_R^* &= (I_R - U_R) \in_R - V (\delta_D + \in_D) \\ \hat{\in}_D^* &= (I_D - U_D) (\delta_D + \in_D) - V^T \in_R \end{aligned} \tag{14}$$

In a similar way the deletion residuals obtained from the MSO model can be expressed as

$$\begin{aligned} \hat{\in}_R^{*(-D)} &= [I_R - U_R - V(I_D - U_D)^{-1}V^T] \in_R \\ \hat{\in}_D^{*(-D)} &= (\in_D + \delta_D) - (I_D - U_D)^{-1}V^T \in_R \end{aligned} \tag{15}$$

3 Results Conditional on a Group of Observations

In this section we will show the average effect of the deletion of a group of observations on the analysis of residuals. We would expect that the residuals are unbiased estimators of the true disturbances. To form a deletion set D, we consider a group of suspect outlying observations as candidates for deletion. How to select such a group is another topic that is not discussed in this paper. A conditional expectation approach may be useful as the average effect of deletion on a set of residuals can be investigated while conditioning on the values of the set of residuals that have been deleted.

For a standard regression model, under usual assumptions we obtain $E\{\in_R\} = 0_R$ and $E\{\in_D\} = 0_D$. Let $E_R\{\ \in\ \}$ denotes the conditional expectation with respect to all the disturbances $\in$ of the 'remaining' group indexed by R with the $\in$'s in the deleted set D held fixed. In other words we have

$$E_R\{\in_R\} = 0_R, \quad E_R\{\in_D\} = \in_D$$

In a similar fashion we may obtain the conditional and unconditional expectations for the OLS and deletion residuals from (9) and (10) and these results are presented in Table 1.

Table 1: Conditional and unconditional expectations of residuals conditioning on $\in_D$ for the standard model

Expectation	Residual Type	Components	
		$\in_R$	$\in_D$
Unconditional	OLS	0_R	0_D
	Deletion	0_R	0_D
Conditional	OLS	$-V \in_D$	$(I_D - U_D) \in_D$
	Deletion	0_R	$\in_D$

It is clear from Table 1 that the unconditional expectation produces confusing results for both the OLS and deletion residuals. The OLS and the deletion residuals are unbiased but effects of a group of deleted observations on the OLS residuals as anticipated in (2) are not visible here. However, the conditional expectation shows the scenario more clearly. It is observed that the OLS residuals of the R set can be affected in the presence of unusual observations in the D set. They are not able to estimate disturbances corresponding to the D set unbiasedly. The deletion residuals clearly show their merit in the conditional expectation approach. The deletion residuals $\hat{\in}_R^{(-D)}$ are unbiased for $E\{\in_R\} = 0_R$ and $\hat{\in}_D^{(-D)}$ are unbiased estimates of $\in_D$ conditional on the values on D. Thus the deletion residuals represent the true disturbances more accurately than the OLS residuals.

Now we consider the conditional and unconditional expectations of OLS and deletion residuals for the mean shift outlier model. The results of expectations of the OLS and the deletion residuals when conditioning a group of observations are summarized in Table 2 together with their corresponding unconditional results obtained from (14) and (15).

Table 2: Conditional and unconditional expectations of residuals conditioning on $\in_D$ for the MSO model

Expectation	Residual Type	Components	
		$\in_R$	$\in_D$
Unconditional	OLS	$-V\delta_D$	$(I_D - U_D)\,\delta_D$
	Deletion	0_R	δ_D
Conditional	OLS	$-V\,(\in_D + \delta_D)$	$(I_D - U_D)\,(\in_D + \delta_D)$
	Deletion	0_R	$\in_D + \delta_D$

We observe from Table 2 that the unconditional OLS residuals resulting from the MSO model corresponding to the R set and D set are affected in the presence of outliers. Similar remarks may apply with the residuals obtained from the OLS residuals conditioning a group of observations. The unconditional expectation of the vectors of deletion residuals corresponding to the R set and D set are 0_R and δ_D respectively which should be the case in an MSO set up if we used the true errors as it is shown in (13). When conditioning on a group of observations indexed by D, the expectation of the true errors for the R set and D set should be 0_R and $\in_D + \delta_D$ respectively and we observe the same results from the conditional expectation of the deletion residuals of the MSO model.

4 Conclusions

This paper compares usual residuals from the OLS analysis with the deletion residuals, where the calculations are performed after a set of observations is deleted. Comparison is also made with the OLS and deletion residuals from the mean shift outlier model. On the basis of comparison of the expectations of the residuals given the error vector of the deletion set, as well as the unconditional expectations, we conclude that the deletion residuals are the most suitable for analysis, as these have expectations exactly equal to their corresponding true errors.

Acknowledgement

The author gratefully acknowledges valuable comments and suggestions from both of the referees.

References

Anscombe, F.J. (1960). Rejection of outliers, *Technometrics*, 2, 123-147.

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd Ed., Wiley, New York.

Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

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

Gentleman, J.F. and Wilk, M.B. (1975). Detecting outliers in a two-way table: II supplementing the direct analysis of residuals, *Biometrics*, 31, 387-410.

Imon, A.H.M.R. (2002). Deletion residuals and deletion weight matrices: Some of their properties and uses, *Pakistan Journal of Statistics,* 18, 469-484.

Rousseeuw, P.J. and Leroy, A. (1987). *Robust Regression and Outlier Detection,* Wiley, New York.

Weisberg, S. (1980). Comment on 'Some large-sample tests for nonnormality in the linear regression model' by H. White and G.M. MacDonald, *Journal of the American Statistical Association*, 75, 28-31.