Impact Analysis for Regression Models with Heteroscedastic Errors

Muhammad Aman Ullah  
Assistant Professor, Department of Statistics, 
Bahauddin Zakariya University, Multan.  
Email: aman_stat@yahoo.com

Muhammad Imdadullah  
M.Phil Scholar, Department of Statistics, 
Bahauddin Zakariya University, Multan.  
Email: mimdadasad@gmail.com

Abstract
This article concentrates on the development of the impact measures for the perturbation strategy of simultaneous omission of an individual variable and an observation in a regression model with heteroscedastic errors. The effect of particular perturbation strategy on residual sum of squares, fitted and predicted values is evaluated. Moreover, the interpretation of change in the significance of the model and or the significance of the measured variables due to specific perturbation scheme is highlighted in this article. A particular data example furnishes analysts with the over all images of the proposed measures.

Keywords: Impact measures; Influential observations; Perturbation; Weighted residuals
Mathematics Subject Classification: 62J20; 62J12.

I. Introduction
The weighted least squares (WLS) method is considered as one of the most significant and widely applicable technique in the least square estimation community to mitigate the effect of heteroscedasticity in a set of data that is hard to elucidate by ordinary least squares (OLS) procedure. It has got great attention in regression analysis see, e.g., Nelder and Wedderburn (1972), Cleveland (1979), Carroll and Ruppert (1988), Cleveland and Devlin (1988), McCullagh and Nelder (1989), Graybill and Iyer (1994), Ryan (1997), Montgomery et al. (2001) and Draper and Smith (2005). The technique has also been used in many other situations such as physics (Weisberg et al. 1978), market structures (Novak and Stangor, 1987), computer algorithms (Doshita, 1993), categorical data analysis (Agresti, 2002) and business and economics (Wooldridge, 2009).

The omission of an individual data element and a variable has an unusually large impact on the results of the analysis. An important aspect of this movement began with Cochran (1938). Numerous influence measures has deeply been studied in the literature to examine the appropriateness of assumptions underlying the modeling process and to investigate unusual character of the data that may influence subject matter based decisions since the seminal work of Cook (1977). Comprehensive treatments can be found in Behnken and Draper (1972), Andrew and Pregibon (1978), Cook (1979), Draper and John (1981), Velleman and Welsch (1981), Atkinson (1982), Chatterjee and Hadi
All of these past articles, with one notable exception, have studied the influence measures using single case or multiple cases omission approach within the least square regression equation-context. This is not a condemnation; it was requirement in order to attain the grasp of influence analysis in linear least square regression model. The one notable exception is an article by Chatterjee and Hadi (1988a) on assessing the joint effect of simultaneous deletion of a variable and an observation on an OLS regression equation. Yet for a linear model with the assumption of heteroscedastic error variances, there has been almost no major breakthrough for assessing the joint impact of simultaneous deletion of a variable and a data point on WLS key results. Therefore, the chief objective of this article is the assessment of effects of simultaneous deletion of a variable and an observation on key results in the WLS regression equation.

II. Preliminaries

A. Background

Consider the standard regression model:

\[ Y = X\beta + \varepsilon \]  

In this model, \( X \) is an \( n \times p \) full rank matrix of known constants, \( Y \) is an \( n \times 1 \) vector of observable responses, \( \beta \) is a \( p \times 1 \) vector of unknown parameters and \( \varepsilon \) is a vector of random errors with \( E(\varepsilon) = 0 \), \( Var(\varepsilon) = \sigma^2W^{-1} \) and \( W \) is known \( n \times n \) diagonal matrix with \( w_{ii} > 0 \).

If \( Var(\varepsilon) = \sigma^2I \), then the OLS estimator of \( \beta \) is \( \hat{\beta}_{OLS} = (X'X)^{-1}X'y \) and the vector of fitted responses is \( \hat{y} = X\hat{\beta}_{OLS} \). The unbiased estimator of \( \sigma^2 \) is \( s^2 = \left\| e \right\|^2 / (n-p)^{-1} \), where \( e = y - \hat{y} \) is the vector of OLS residuals.

This paper assumes that the reader is familiar with the primary ideas of leverage, residual sum of square, fitted value, partial F-test, and influence in least square (LS) regression as given, for example, in the studies of Chatterjee and Hadi (1988a,b).

B. Measures of Impact in Least Squares

The understanding of the change in the significance of a model and/or the significance of measured variables after deleting simultaneous omission of an observation and a variable has made very little considerable contribution see, for example, Chatterjee and Hadi (1988a,b). Various statistical quantities for investigating the joint impact of simultaneous omission of a variable and an observation on the outcomes of a LS regression equation are given in Chatterjee and Hadi (1988a, b). They called the use of such statistical quantities “impact analysis.”

The usual role of impact analysis is to examine the changes made in a given aspect of the study when the variables and data are perturbed simultaneously. An especially interesting perturbation strategy is simultaneous deletion of an observation and a variable.
This strategy will be applied throughout this paper. See Weisberg (1983) for a good
debate on other perturbation strategies.

Usually, the impact of such perturbation scheme can be viewed as a function of
three elements, the first a function of the $i$-$k$th position of residual sum of squares in the
regression of $Y$ and $X_{i}$ on all other ingredients of $X$, the second a function of the $i$-$k$th
position of the element in the leverage space and finally estimated coefficient of $k$th
parameter in $X$ space.

The most common single perturbation scheme impact measure is to compare the
residuals sum of squares of full data model and a sub-model without the $k$th variable and
ith observation ($DSSE_{ik}$, Chatterjee and Hadi, 1988a, b), which measured at the $i$-kth
position is contributed by:

$$ DSSE_{ik} = SSE_{r(\{i\}k)} - SSE_{r,X} $$

where $SSE_{r(\{i\}k)}$ is the LS residual sum of squares
without the $i$th observation and $k$th variable and $SSE_{r,X}$ is LS residual sum of squares
for complete data. $DSSE_{ik}$ is the change in the residual sum of squares of a single
perturbation scheme when it is removed. Thus it can be taken as a technique of impact on
individual residual sum of squares at $ik$th position.

Another convenient measure of impact is $DFIT_{ik}$ (Chatterjee and Hadi, 1988a, b),
which measure at the $ik$th position is given by

$$ DFIT_{ik} = \left\| \tilde{Y} - \tilde{Y}_{(i)k} \right\| $$

The quantity $DFIT_{ik}$ is a measure of the change in the significance of $k$th variable. It can also
be used as a measure to declare either ith observation is an outlier (Chatterjee and Hadi,
1988a, b). It is obvious that $X_{i[k]}$ is not contributed in the fitting procedure of a sub model
without $i$th observation and $k$th variable. Chatterjee and Hadi’s name it the $i$th predicted
value. The measure of impact ($DPRD_{ik}$), on the prediction at $i$th position can be seen in
Chatterjee and Hadi (1988a, b). Even though $DPRD_{ik}$ is found on different theoretical
considerations, it is closely linked to the standardized cross-validated prediction error

Finally, a measure of impact related with significance tests is $F_{i(i)}$ (see,
Chatterjee and Hadi, 1988a, b), evaluated at the $i$th position is given by

$$ F_{i(i)} = \frac{\tilde{r}_{\cdot i}^{2} + t_{i}^{2}}{\hat{\sigma}^{2}_{i} - \hat{\sigma}^{2}_{i} \left( 1 - h_{i} \right) } $$

where, the statistics $F_{i}, \hat{\sigma}^{2}_{i}$ and $t_{i}$ can be found in
Chatterjee and Hadi (1988a, b). Cook and Weisberg (1982) deduced a similar but various
form for $F_{i(i)}$.

It is considerable to observe that these impact measures are useful for measuring
the change in the significance of the model or the significance of the measured variables
after simultaneous omission of a variable and an observation. Its generalization for multiple observations and/or multiple variables is anticipated to be computationally difficult and yet to be explored. Further, expansion in various settings of linear, non-linear models and in different offshoots of LS method are also yet to be studied.

C. Heteroscedasticity and Impact Analysis

In general, heteroscedasticity is defined as the presence of non-constant error variance. The potential effects of heteroscedasticity on various LS statistical quantities are well known (see, e.g., Montgomery and Peck, 2001; Draper and Smith, 2005).

It is not strange to have heteroscedasticity and unusual observations simultaneously in a data set (see, e.g., Cook and Weisberg, 1983, pp. 7). The issue of modeling variances has greatly been discussed specifically in the econometric field (see, for example, Park, 1966; Harvey, 1976). Carroll and Ruppert (1988) presented local influence diagnostics for the variance parameter estimates in several nonlinear models for the mean, whereas Verbyla (1993) equates residual maximum likelihood and complete estimates using case omission and likelihood displacement under normal error. Yet for WLS regression, there has been almost no development for evaluating the effect that heteroscedasticity can have on the joint impact of simultaneous omission of an observation and a variable. We temporarily conclude that diminution in heteroscedasticity should be first step for the effective impact analysis of specific perturbation scheme in WLS regression.

In the following sections, it is proven empirically how the joint impact of simultaneous omission of an observation and a variable can be modified when WLS is used to mitigate the effect of heteroscedasticity. In Section 3, the WLS estimators are introduced. Moreover, the updated related statistical models and the roles of joint impact of simultaneous omission of a variable and an observation are also discussed. An illustrative example is considered in Section 4. Finally, concluding remarks are given in Section 5.

III. Measuring the Impact of Specific Perturbation in Regression Model with Heteroscedastic Errors

A. WLS Regression

WLS regression is an estimation method used when heteroscedasticity is present in the data. In practical terms, it consists of transforming $\text{Var}(\epsilon) = \sigma^2 W^{-1}$ to $\text{Var}(\epsilon) = \sigma^2 I$. The resulting statistical quantities are more stable than the OLS statistical quantities. WLS can also be understood as a generalized least squares estimator in certain conditions. For an excellent debate on this point of view and other related information see, e.g., Gross et al. (2001), Tian and Weins (2006) and Luati and Proietti (2009).

The WLS estimator of $\beta$ can be written as: $\hat{\beta}_{\text{WLS}} = (SS')^{-1} (S'Z)$, where

$$S = W^{1/2} X$$ and $$Z = W^{1/2} Y$$

Using this transformation the WLS residuals are: $e_{\text{WLS}} = W^{1/2} (Y - X \hat{\beta}_{\text{WLS}})$
B. Updated Notation

Following the notations of Chatterjee and Hadi (1988a, b), we reserve the letters \( z_i \) and \( s_i \) to refer the \( i \)th row of \( Z \) and \( S \) respectively, and \( S_k \) to denote the \( k \)th column of \( S \). In the matrix \( (S : Z) \), \( (s_i : z_i) \) represents the \( i \)th data point or the \( i \)th row. We also apply the notation “(i)” or “[k]” as a subscript to a measure to show the deletion of the \( i \)th data point or the \( k \)th variable respectively. Thus, for instance, \( S_{(i)} \) is the matrix \( S \) with the \( i \)th row deleted, \( S_{[k]} \) is the matrix \( S \) with the \( k \)th column deleted. Let \( \hat{\beta}_{WLS_{(i)}} \) is the weighted least square estimated parameter vector when the \( i \)th data point is deleted… etc. We apply \( A^{-1}, A', (A^{-1})' \) to represent the inverse, transpose, and transpose of the inverse of the matrix \( A \) respectively.

Without loss of generality, we will assume that the \( k \)th column and the \( i \)th row are the last column and row of \( S \), i.e., \( i = n \) and \( k = p' \) respectively. Then

\[
S_{[i]} = \begin{bmatrix} S_{(i)[k]} \\ S_{(i)[k]} \end{bmatrix}, \quad S_i = \begin{bmatrix} S_{(i)k} \\ s_{ik} \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{(i)} \\ z_i \end{bmatrix}
\]

C. Updating Related Statistical Models

To study the joint impact of dropping the \( i \)th observation and \( k \)th variable simultaneously from the data set in a cost-effective way needs the use of simple, cheap updating related statistical models. In this sub-section such related statistical models are provided. When properly expressed, the updating related statistical models are unusually similar to the related statistical models used in linear least square regression see, for instance Chatterjee and Hadi (1988a, b). These models are:

Model (1): Complete data model: 

\[
Z = S_{[k]} \beta + S_k \delta_k + \epsilon
\]

Model (2): \( i \)th observation deleted: 

\[
Z = S_{[i]} \beta + S_k \delta_k + m \eta + \epsilon
\]

Model (3): \( k \)th variable deleted: 

\[
Z = S_{[i]} \beta + \epsilon
\]

Model (4): \( k \)th variable and \( i \)th observation deleted: 

\[
Z = S_{[i]} \beta + m \eta + \epsilon
\]

In a short way, the symbols \( \hat{Z}, \epsilon \) and WSSE to represent the vector of weighted fitted values, weighted residuals and the weighted residual sum of squares, respectively from the complete data model \( Q_k = e_{ZS_{(i)}} \) to represent the vector of weighted residuals in the regression of \( Z \) on \( S_{[i]} \) and \( T_k = e_{S_i S_{[i]}} \) to represent the vector of weighted residuals obtained when \( S_k \) is regressed on \( S_{[i]} \)

D. Impact Measures

Chatterjee and Hadi (1988a, b) deduced the impact measures under the perturbation of simultaneous omission of a variable and an observation in LS regression
equation. In this subsection we extend these results to the linear regression model with WLS estimate.

D(i). \( DSSE_{ak} \) for WLS

A version of \( DSSE_{ak} \) for WLS regression is

\[
WDSSE_{ak} = WSSE_{z(i\backslash k)} - WSSE_{z S},
\]

\[
WDSSE_{z(i\backslash k)} = WSSE_{z S[i]} - \frac{q_a^2}{1 - p_{[i]}}
\]

\[
WDSSE_{ak} = T^t_k T_k \delta_k^2 - \frac{q_a^2}{1 - p_{[i]}},
\]

Where \( WSSE_{z(i\backslash k)} \) and \( WSSE_{z S} \) are the residual sums of squares computed with \( i \)-th observation and \( k \)-th variable deleted and for full data respectively in WLS regression. Eq. (5) is simplified using the facts that can be seen in Belsley et al. (1980), while (6) is simplified using (4) and (5). Formula (6) have a similar form as in Chatterjee and Hadi (1988a, b), but in this case, \( T_k \), \( \delta_k \), \( q_a \) and \( p_{[i]} \) are the statistical quantities, which depends on the WLS estimator. Moreover, the interpretation of \( WDSSE_{ak} \) is also similar to the one for the OLS case (see, for instance, Chatterjee and Hadi (1988a, b).

D(ii). \( DFIT_{ak} \) for WLS

The \( DFIT_{ak} \) measures for WLS regression is

\[
WDFIT_{ak} = \hat{Z} - \hat{Z}_{(i\backslash k)}. \text{ It may be determined as:}
\]

\[
\hat{Z} - \hat{Z}_{(i\backslash k)} = S \left( \hat{\beta} : \hat{\delta}_k \right)' - S_{[i]} \hat{\beta}_{(i\backslash k)}
\]

\[
\hat{Z} - \hat{Z}_{(i\backslash k)} = T_k \delta_k + \frac{q_a}{1 - p_{[i]}} S_{[i]} \left( S_{[i]}' S_{[i]} \right)^{-1} S_{[i]}
\]

and that

\[
WDFIT_{ak} = \left\| \hat{Z} - \hat{Z}_{(i\backslash k)} \right\|^2 = \frac{q_a^2 p_{[i]}}{1 - p_{[i]}} + T_k T_k \delta_k^2
\]

Where, \( q_a^2 \) is the square of the \( i \)-th element of \( Q_z \), the residuals when \( Z \) is regressed on \( S_{[i]} \). Formula (7) is simplified using a useful identity given in Bingham (1977). The \( WDFIT_{ak} \) have the same form as in Chatterjee and Hadi (1988a, b), but in current case statistical quantities which are used to obtained it depends on the WLS estimator. The interpretation of \( WDFIT_{ak} \) is same to the one for the OLS case (see, for example, Chatterjee and Hadi (1988a, b).

D(iii). \( DPRD_{ak} \) for WLS

A version of \( DPRD_{ak} \) for WLS regression is
A little algebra will verify that

\[
WDPRD_a = \left[ f_i \left(Z, S\right) - f_i \left(Z_{(i)} s_{(i)}\right) \right]^2
\]  

(9)

As observed, the quantities (10) and (11) found so far are in the similar form as was in LS (see, e.g., Chatterjee and Hadi (1988a, b). The only difference, the statistical quantities, \(q_a\) and \(p_{(i)}\), which depends on the WLS estimator \(\hat{\beta}_{WLS}\), are used instead of LS estimator. Large values of (10) show that deleting the \(i\)th data point and \(k\)th variable will have see, for example, Chatterjee and Hadi (1988a, b).

IV. Partial F test for WLS

The major concern of the impact quantities put in thus for is to measure the effect of the omission of the \(k\)th variable and the \(i\)th data point on different statistical quantities in linear regression models fitted by WLS. Therefore, these quantities can be equated relative to each other. However, these quantities are not helpful in significance tests. The partial F-tests are usually used to simplify the original model. In these tests we made the hypothesis that the individual coefficients of \(\hat{\beta}_{WLS}\) are zero.

In this study, we are concerned with one complete and three sub models. We can make six possible pair-wise comparisons in order to examine their significance. Since numerous diagnostics can be deduced under these six pair-wise comparisons in terms of the basic statistics, studentized residual and leverage measure, so they provide similar influence knowledge see, for example, Chatterjee and Hadi (1988a, b). For brevity, among these only the common and more informative diagnostic is deduced below.

A. \(F_{i(k)}\) for WLS

Equating model (2) versus model (4) is similar to testing the hypothesis NH:

\(H_0: \delta_k = 0\) versus AH: \(\delta_k \neq 0\) applying the data set with the \(i\)th data point deleted. In this scenario the usual normal theory of F-statistic is most suitable for the development of a new statistic that is

\[
F_{i(k)} = \frac{WSSE_z z_{(i)\hat{\delta}_k} - WSSE z_{(i)\hat{\beta}_{WLS}}}{WSSE z_{(i)\hat{\delta}_k} / (n - p' - 1)}
\]  

(15)

Substituting (16) and (11) in (19) and simplifying, we find
\[ F_{k(i)} = F_{1}, \frac{\sigma^2}{\hat{\sigma}_{(i)}^2} + r_i \cdot \frac{q_{\alpha}^2}{\hat{\sigma}_{(i)}^2 \left(1 - p_{A(i)}\right)} \]  

(16)

Which have the similar form as in Chatterjee and Hadi (1988a, b), but in eq. (16) the statistical quantities on its right hand side depends on WLS estimator. The critical region is reject NH if \( F_{k(i)} \geq F_{\alpha(1, N-K-1)} \). Another similar but more complicated from can be seen in Cook and Weisberg (1980). For remarkable comments on the relationship between \( F_{k(i)} \) and \( F_k \) and other related information see, for example, Chatterjee and Hadi (1988a, b).

V. Data Description and Impact Analysis

As a numerical illustration, we consider the Rat data (Weisberg, 1982, pp.110-113). This data set consists of three explanatory variables and 19 cases. The response variable is percentage of dose in the liver, and explanatory variables are body weight, liver weight and relative dose. Cook and Weisberg (1982), Cook (1986) and Loynes (2001) have also studied this data set. Hadi (1984) used this data set to disclose the influential point in OLS regression and discovered that cases 1, 2, 3, 5, 13 and 19 have the most influential observations using \( DSSE_{\alpha}, DFIT_{\alpha}, DPRD_{\alpha} \) and \( F_{k(i)} \).

In present section we also consider the same data set to discover the influential observations in WLS by using \( WDSSE_{\alpha}, WDFIT_{\alpha}, WDPRD_{ik} \) and \( F_{k(i)} \) and to compare these results with the setting of LS (see, Hadi, 1984 and Loynes, 2001). The scatter plot given in Cook (1986) hinted strong evidence of heteroscedasticity. Cook and Weisberg (1983) also declared same argument about heteroscedasticity using score test. The weight vector of WLS is used as in Carroll (1982). The WLS regression summary is presented in Table (1). We are concerned with the following linear transformed model.

\[ Z = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + e \]

Where \( Z \) and \( S_1, S_2 \) and \( S_3 \) are defined in (2).

<table>
<thead>
<tr>
<th>Table 1: WLS regression summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical quantities</td>
</tr>
<tr>
<td>( \hat{\beta}_{WLS} )</td>
</tr>
<tr>
<td>s.e. ( \hat{\beta}_{WLS} )</td>
</tr>
<tr>
<td>( t )</td>
</tr>
<tr>
<td>p-values</td>
</tr>
<tr>
<td>( n = 19 )</td>
</tr>
</tbody>
</table>

Table (2) contributes the case statistics for data point’s number 3 and 7 for WLS regression. We observe that data point number 3, the case with the largest potential influence \( (p_{3,3} = 0.746) \) has comparatively large observed influence \( (D_3 \times 10 = 17.57) \). Likewise; the data point number 7 is comparatively influential due to its large value of \( r_7^* = 2.493 \). Numerous traditional influence measures
(results are not given for sake of brevity) fails to report that either data point 3 or 7 is influential on all the parameters or on a few subsets of them.

**Table 2: Case statistics of few data points for WLS regression**

<table>
<thead>
<tr>
<th>Case (i)</th>
<th>$e_i$</th>
<th>$p_i$</th>
<th>$r_i^*$</th>
<th>$10 \times D_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3.403</td>
<td>0.746</td>
<td>1.633</td>
<td>17.57</td>
</tr>
<tr>
<td>7</td>
<td>7.623</td>
<td>0.336</td>
<td>2.493</td>
<td>5.83</td>
</tr>
</tbody>
</table>

Figures (1-3) shows plots of $WDSSE_{ij}$ against case numbers ($i=1,...,19$) for $k=1,...,3$. We have no difficulty in discovering that cases 3 and 7 are influential under the specific perturbation scheme considered in Section 3. A further point from Fig. 2 is that majority of the cases have the negative sign, which may indicate associated variable is insignificant. Of course it agrees with our theoretical findings in Eq. (6).

Figs. (4-6) plots the $WDFIT_{ij}$ values of the data. From these figures we can see that case numbers 3 and 7 has the strong influence on $\hat{\beta}_{kLS}$ when a variable $S_i$ is perturbed, case 3 is influential when $s_i$ is perturbed and also case 7 is influential when $s_i$ is perturbed. The effect of simultaneous omission of a variable and an observation to the prediction of WLS regression equation is investigated by $WDPRD_{ij}$ and displayed in Figs. (7-9). It is clear that case 3 and 7 has a much larger influence than the other cases when either $s_2$ or $s_3$ is perturbed respectively. Also, case 9 is discovered as one of the most influential observation in Fig. 7, this however does not happen in Figs. 3 and 4.
This data set was also analyzed with reference to OLS multiple regression by Hadi (1984, pp. 58-74) and Loynes (2001, pp. 52). Hadi (1984) considered various diagnostic measures such as $DSSE_{ik}$, $DFIT_{ik}$ as well as $DPRD_{ik}$ for variables $X_1$, $X_2$, and $X_3$, while Loynes (2001) applied new measure of local influence. Interestingly, only one of the four most influential observations for WLS were listed as influential for OLS based on one or more of the Hadi’s measures while none of the four most influential observation for WLS were discovered as influential as compared with Loynes’s procedure for OLS see, Table 3. It in turn reveals that method such as WLS can be influenced by observations differently from OLS, for which a large number of diagnostic procedures have been reported in literature.

**Table 3**: Influential cases in Rat data using various methods of diagnostics and estimation

<table>
<thead>
<tr>
<th>Estimation Methods</th>
<th>Notation of Diagnostics</th>
<th>Most Influential Cases (in order)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>$DSSE_{ik}$</td>
<td>3</td>
<td>Hadi (1984)</td>
</tr>
<tr>
<td></td>
<td>$DFIT_{ik}$</td>
<td>3,5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$DPRD_{ik}$</td>
<td>3,5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{k(i)}$</td>
<td>1, 13, 19</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>$l_{max}$</td>
<td>19, 13, 1</td>
<td>Loynes (2001)</td>
</tr>
<tr>
<td></td>
<td>$WDSSE_{ik}$</td>
<td>3,7</td>
<td>Current article</td>
</tr>
<tr>
<td></td>
<td>$WDFIT_{ik}$</td>
<td>3, 7</td>
<td></td>
</tr>
<tr>
<td>WLS</td>
<td>$WDPRD_{ik}$</td>
<td>3, 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{k(i)}$</td>
<td>1, 7</td>
<td></td>
</tr>
</tbody>
</table>

The preceding analysis is evident that the impact measures (6), (8) and (11) are not enough because these gives the magnitude and not the direction of influence for the $i$th data point on the $k$th variable. Therefore, a statistical quantity that gives both magnitude and direction of influence is required. The $F_{k(i)}$ measure, contributes both the magnitude and direction of the impact of the $i$th data point on the $k$th coefficients of the given variables see, for example, Chatterjee and Hadi (1988a, b). Table 4 exhibits the measure $F_{k(i)}$ for WLS regression setting. In each column, large elements (reported with ↑) suggests that deletion of the $i$th data point increases the significance of $k$th variable, while small elements (reported with ↓) suggests that deletion of the $i$th data point decreases the significance of $k$th variable.
Moreover, it is obvious that pairs which change the significance of the variables in the framework of OLS (Hadi, 1984) fail to change the significance of the variables in the framework of WLS. Table 5 exhibited the summary of this debate.

Table 5: Effect of deleting ith data point on the significance of kth variable

<table>
<thead>
<tr>
<th>Significance effect</th>
<th>Variables and Methods</th>
<th>OLS*</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X₁</td>
<td>X₂</td>
</tr>
<tr>
<td>Decreased</td>
<td></td>
<td>3,5</td>
<td>13,5</td>
</tr>
<tr>
<td>Increased</td>
<td></td>
<td>1,13</td>
<td>1,2</td>
</tr>
</tbody>
</table>

* Hadi (1984)

Thus, the diagnostic analysis (cf. (6), (8) and (11)) discovered the following five as potentially influential cases: 1, 3 and 7. In order to disclose the effect of these five influential cases on the WLS, we refit the model under eliminating individually each one of these five cases. Tables 6 give the parameter estimates based on full data and deleted cases and relative changes (in %) of the assumed model. For computation of relative changes see for instance, Leiva et al. (2007).

Following the same method given in Lee et al. (2006) and Leiva et al. (2007), we apply the total relative changes: \( TRC = \sum_{j=1}^{3} |RC_{β_j}| \), the maximum relative changes: \( MRC = \max_j |RC_{β_j}| \). We see for case 1 that \( TRC = 167.41 \) and \( MRC = 127.78 \). Similarly for other settings, the results are given in the last two rows of the Table 6. Hence, a glance at results of Table 6 reveals that the results are more sensitive for the influential observation.
Table 6: Impact of Influential Cases on WLS Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Full data</th>
<th>Case Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>$\hat{\beta}_{1,\text{WLS}}$</td>
<td>-0.0204</td>
<td>-0.0250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-22.55%)</td>
</tr>
<tr>
<td>$\hat{\beta}_{2,\text{WLS}}$</td>
<td>0.0126</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-127.78%)</td>
</tr>
<tr>
<td>$\hat{\beta}_{3,\text{WLS}}$</td>
<td>4.010</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-17.08%)</td>
</tr>
<tr>
<td>TRC</td>
<td>167.41</td>
<td>257.02</td>
</tr>
<tr>
<td>MRC</td>
<td>127.78</td>
<td>99.15</td>
</tr>
</tbody>
</table>

(CO)= Coefficients; and C = % of Change

VI. Summary

We have discussed the WLS method and obtained several impact measures based on the particular perturbation scheme. As Chatterjee and Hadi (1988a, b) successfully assessed the joint impact of simultaneous deletion of a variable and an observation in OLS framework, the criteria used in this article can be also applied to WLS framework. By similar derivations from OLS framework given by Chatterjee and Hadi (1988a, b), $DSSE_{ik}$, $DFIT_{ik}$, $DPRD_{ik}$ and $F_{i(i)}$ can be established for WLS.

In this article, we show that, when WLS is applied, the impact of each case on various statistical quantities changes as compared to OLS (see, Hadi, 1984 and Loynes, 2001). This change is not incredible. This is primarily caused by the behavior of residuals (3) as a function of weight. Moreover, we have discovered in the preceding debate that Rat data is extremely sensitive to the omission of its very few elements. Hence, in practice it would be significant to study the joint impact of simultaneous omission of a variable and an observation on various outcomes of regression along with other statistical quantities of concern.

References


