Journal of Economic and Social Research 2 (1) 2000, 85-97

Weighting and Deletion Approaches to Regression Diagnostics: A Comparison and an Extension

M. Mercedes Suárez Rancel and Miguel A. Gonzáles Sierra[1]

Abstract. A practical approach to influence analysis in statistical modelling is based on case weighting; Cook (1986) pioneered the idea and Billor and Loynes (1993) modified it. Cook's local influence was motivated by Cook's measure (1977), which is based on the local influence of observations on regression coefficients. These measures, however, are not resistant to masking and swamping effects. In this article we extend some influence measures to locally influential measures in order to mitigate these difficulties and compare weighting and deletion diagnostics for detecting influential observations. Similarities between the two approaches are stressed.

JEL Classification Codes: C10, C40.

Keywords: Cook's Distance; Deletion Diagnostic; Hadi's Influence Measure; Weighting Diagnostic; Welsch's Distance; Welsch-Kuh's Distance.

1. Introduction

Statistical models are a simplification of reality; we rarely expect the model to be exactly true. Nevertheless, when we select a statistical technique and perform statistical inference, we often act as if the model were true. This is often justified by claiming that "small" deviations from the theoretical properties of the selected inferential techniques will cause only minor changes in the results produced by the inference. Unfortunately, this argument may not be true. In many applications apparently small changes in a model, a model assumption, or a data point can have very large effects on the results.

[1] Department of Statistics, Faculty of Mathematics, University of La Laguna, SPAIN

Cook (1986) gives a general method for assessing the local influence of departures from assumptions on likelihood-based models (not necessarily regression models). The starting point is that if a minor perturbation in the model leads to a major change in essential parts of the results of the analysis, then there is evidence of a difficulty. This suggests measuring the sensitivity of the analysis to change in the model by some kind of derivative. In fact, Cook (1986) suggests using the normal curvature of the likelihood displacement surface. Billor and Loynes (1993) point out the computational difficulties associated with practical applications of the maximum curvature approach and propose a new measure of local influence which is simpler to compute.

If a data set contains a single local influential observation, the problem of identifying such an observation is relatively simple from both the analytical and computational points of view; however, if a data set contains more than one local influential observation, which is likely to be the case in most data sets, the problem of identifying such observations becomes more difficult. This is due to masking and swamping effects. Masking occurs when an outlying observation goes undetected because of the presence of other, usually adjacent, observations. Swamping occurs when "good" observations are incorrectly identified as outliers because of the presence of another, usually remote, subset of observations.

Recent studies on local influence have been motivated by the statistic D_i proposed by Cook (1977), see (2) below so they have considered only the local influence on regression coefficients, which are not resistant to masking and swamping effects. In this article we propose measures based on three different influence measures in linear regression to mitigate these difficulties and compare the weighting and deletion diagnostics.

In Section 2 we give the general idea of local influence, inspired by Cook's (1986) work. Section 3 defines three of the best known measures of deletion influence. In Section 4 we extend Welsch-Kuh's distance, Welsch's distance and Hadi's influence measure to locally influential measures. Finally, Section 5 compares the weighting and deletion diagnostics.

2. Local Influence

Consider the standard linear regression model:

$$Y = X\beta + \varepsilon, \tag{1}$$

where ε is an $n \times 1$ vector whose elements are assumed to be independent, normal random variables with mean zero and known variance σ^2, X is a known $n \times k$ matrix with full column rank, β is a $k \times 1$ vector of parameters and Y is an $n \times 1$ vector of response variables. Collectively, the *ith* observation y_i on the response variable in combination with the associated values for the explanatory variables will be referred to as the '*ith case*'.

Many measures have been suggested to assess the influence of observations in regression modelling. In fact, Chatterjee and Hadi (1986) have done an excellent review of this subject. Cook (1986) considers a general version of Cook's distance

$$D_i = \frac{\left\| \hat{Y} - \hat{Y}_{(i)} \right\|^2}{k\sigma^2}, \tag{2}$$

where $\hat{Y}$, $\hat{Y}_{(i)}$ are the $n \times 1$ vectors of fitted valued based on the full data and the data without the ith case, respectively, and k is the dimension of β. He investigated

$$D_i(w) = \frac{\left\| \hat{Y} - \hat{Y}_{(w)} \right\|^2}{k\sigma^2}, \tag{3}$$

where $\hat{Y}_{(w)}$ is the vector of fitted values obtained when the ith case has weight w and the remaining cases have weight 1.

These ideas have been extended to general models. Their extension is partially motivated by the following relationship between $D_i(w)$ and the log-likelihood $L(\beta)$ for model (1):

$$kD_i(w) = \frac{\left[\left\| Y - \hat{Y}_w \right\|^2 - \left\| Y - \hat{Y} \right\|^2 \right]}{\sigma^2} = 2\left[L(\hat{\beta}) - L(\hat{\beta}_w) \right]$$

where $\hat{\beta} = \hat{\beta}_{w=1}$ and $\hat{\beta}_w$ is the maximum likelihood estimator of β when the ith case has weight *w*. The form of this relationship is a consequence of the statistical structure assumed for the errors in model (1).

The log-likelihood for the unperturbed and perturbed models are denoted by $L(\theta)$ and $L(\theta\,|w)$, respectively. Then the likelihood displacement *LD(w)* is defined by

$$LD(w) = 2\left[L(\hat{\theta}) - L(\hat{\theta}_w) \right], \tag{4}$$

where $\hat{\theta}$ and $\hat{\theta}_w$ are the maximum likelihood estimators of θ under the unperturbed and perturbed models, respectively. The vector of the values *w* and *LD(w)* forms the surface of interest as *w* varies over certain space. The direction $h_{máx}$ of maximum curvature of the likelihood displacement surface in the postulated model (where $w = w_0$) indicates the greatest local sensitivity against perturbations. The direction of maximum curvature is used as the main diagnostic tool in the local influence approach.

Billor and Loynes (1993) show some of the practical and theoretical difficulties that arise in Cook's approach. For example, the restriction of the computability of the maximum curvature to the linear regression model, the lack of invariance of the curvature under reparametrisation of the perturbation scheme; and the lack of definition of the parameters. To avoid these difficulties, Billor and Loynes (1993) suggest an alternative likelihood displacement:

$$LD^*(w) = -2\left[L(\hat{\theta}) - L(\hat{\theta}_w \mid w) \right] , \tag{5}$$

where $L(\hat{\theta}_w|w)$ is the log-likelihood of the perturbed model, while Cook (1986) uses only the perturbation in the estimation of the parameters. Billor and Loynes (1993) suggest that the first derivative of LD^* provides valuable information about the local behavior of LD^*, so they use the direction which produces the maximum increment in LD^* with the following slope:

$$l_{\max} = \left\| \nabla LD^*(w_0) \right\| = 2 \left\| \nabla L\left(\hat{\theta} \mid w\right) \right\| .$$

If we take the (perturbed) model

$$Y = X\beta + \varepsilon \tag{1a}$$

where var$(\varepsilon) = \sigma^2 W^{-1}$ with W = diag (1, 1, ..., 1 + w_i , 1, ..., 1), then

$$\ell_i = \ell_{max,i} = \left(1 - \frac{e_i^2}{\sigma^2}\right) . \tag{6}$$

3. Deletion Influence Measures

The main local influence results are based on Cook's measure, so they consider only the influence on the estimators of the regression coefficients. But there are other measures of influence which are more sensitive to the estimation of the variance of the errors, give more emphasis to high-leverage points, and are more resistant to swamping and masking effects. In this Section we define three of the best known measures of influence with these properties.

1. Welsch-Kuh's Distance (1977): The influence of the ith observation on the predicted value $\hat{y}_i$ can be measured by the change in the prediction at x_i when the *ith* observation is omitted, relative to the standard error of $\hat{y}_i$; that is,

$$\frac{\left|\hat{y}_i - \hat{y}_{i(i)}\right|}{\sigma\sqrt{p_{ii}}} = \frac{\left|x_i^t(\hat{\beta} - \hat{\beta}_{(i)})\right|}{\sigma\sqrt{p_{ii}}} \tag{7}$$

where p_{ii} is the ith diagonal element of $P = X(X'X)^{-1}X'$, and $\hat{\beta}$ and $\hat{\beta}_{(i)}$ are the estimators of β based on the full data and the data without case *i*, respectively. Welsch and Kuh (1977), Welsch and Peters (1978), and Besley *et al*. (1980) suggest using $\hat{\sigma}_{(i)}$ as an estimator of σ in (7) to study the influence not only on the regression coefficients but also on σ. Therefore, (7) can be written as

$$WK_i = \left|r_i^*\right|\sqrt{\frac{p_{ii}}{1-p_{ii}}} \quad , \qquad (8)$$

where $r_i^* = \dfrac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-p_{ii}}}$ is the *ith* externally Studentized residual.

2.Welsch's Distance: Welsch (1982) suggests the following modification of the measure (8):

$$W_i = WK_i\sqrt{\frac{n-1}{1-p_{ii}}} \quad , \qquad (9)$$

which is a measure more sensitive to high-leverage points.

3. Hadi's Influence Measure: Hadi (1992) proposes a measure that is based on the simple fact that potentially influential observations are outliers in either the *X*-space, the *Y*-space, or both. This yields the overall influence measure:

$$H_i^2 = \frac{k}{(1-p_{ii})}\frac{d_i^2}{(1-d_i^2)} + \frac{p_{ii}}{1-p_{ii}} \quad , \qquad i = 1, 2, ..., n , \qquad (10)$$

where $d_i^2 = e_i^2 / e^t e$ is the square of the *ith* normalized residual and p_{ii} is defined in (7). The diagnostic measure (10) is the sum of two components each of which has a nice interpretation. A large value for the first term on the right hand side of (10) indicates that the model has a poor fit (a large prediction error), and a large value for the second term indicates the presence of an outlier in the *X* space.

4. Local Influence Measures

In the previous section we defined three of the most commonly known measures of influence. In this section, we extend these measures to local influence.

4.1 Local Influence Based on Welsch-Kuh's Distance

Considering the relation between the internally and externally Studentized residuals, Welsch Kuh's distance (8) can be obtained from Cook's distance (2)

$$WK_i^2 = \frac{k(n-k-1)}{n-k-r_i^2} D_i \,,$$

where $r_i = \dfrac{e_i}{\hat{\sigma}\sqrt{1-p_{ii}}}$ is the internally Studentized residual. For the perturbed model (1a), the likelihood displacement of this measure is

$$LD_i^*(WK^2) = \frac{k(n-k-1)}{\left[n-k-\dfrac{e_i^2(1+w_i)}{\hat{\sigma}^2(1-p_{ii})} \right]} LD_i^*(w) \,.$$

So, the slope of the maximum increment direction of $LD^*(WK^2)$ is

$$\ell_i(WK^2) = \frac{(1-p_{ii})k(n-k-1)\left(1-e_i^2 \big/ \hat{\sigma}^2\right)}{(n-k)(1-p_{ii}) - e_i^2 \big/ \hat{\sigma}^2} \quad . \tag{11}$$

If we replace the residuals in this expression by the errors ε, which are assumed to be independent, normally distributed random variables with means 0 and variance σ^2, and we consider $\hat{\sigma}^2 \cong \sigma^2$ and $p_{ii} = k / n$, we recommend using

$$A\left(z^2_{\frac{1-\alpha}{2}} - 1\right) / \, n z^2_{\frac{1-\alpha}{2}} - B$$

as a cut-off point for (11), where $A = k\,(n-k)\,(n-k-1)$, $B = (n-k)^2$, and $z_{1-\alpha}$ is a critical value of the N(0,1).

4.2 Local Influence Based on Welsch's Distance

If we consider the relation (9), we can write

$$W_i^2 = WK_i^2 \, \frac{n-1}{1-p_{ii}} \quad . \tag{12}$$

Therefore, if we take the perturbed model (1a), the slope of the direction of maximum increment of $LD^*(WK)$ is

$$\ell_i(W^2) = \ell_i(WK^2)\frac{n-1}{1-p_{ii}},$$

and with a development similar to that of Section 4.1, the cut-off point for this measure would be

$$A'\left(z^{2}_{\frac{1-\alpha}{2}} - 1 \right) / n z^{2}_{\frac{1-\alpha}{2}} - B \,,$$

where $A' = n\,k\,(n - 1)\,(n - k - 1)$ and $B = (n - k)^2$.

4.3 Local Influence Using Hadi's Influence Measure

To assess the influence of varying w throughout Ω, Billor and Loynes (1993) use the likelihood displacement (5). As we shall see in Example 2, this approach is not resistant to swamping and masking effects. Therefore, we propose an alternative likelihood displacement:

$$LD_{(i)}(w_i) = -2 \left[L(\hat{\theta}) - L_{(i)}(\hat{\theta}_{w_i} \mid w_i) \right] ,$$

where $L_{(i)}(\hat{\theta}_{w_i})$ is the log-likelihood displacement under the perturbed model when the ith observation is deleted.

If we apply this likelihood displacement to the perturbed model (1a), we have

$$L_{(i)}(\hat{\theta}_{w_i} | w_i) = -\frac{1}{2}\ln 2\Pi - \frac{1}{2}\ln\left[\frac{\hat{\sigma}^2_{(i)}}{1+w_i} \right] - \frac{1}{2\hat{\sigma}^2_{(i)}}\left(y_i - x_i^t \hat{\beta}_{(i)} \right)^2 \left(1+w_i\right)^2 .$$

Then the slope of the maximum rate of increase in $LD_{(i)}(w_i)$ is given by

$$\ell_{i(i)} = \ell_{\max,i(i)} = 1 - \frac{e^2_{i(i)}}{e^t_{(i)} e_{(i)}} = 1 - \frac{d_i^2}{(1-p_{ii})(1-d_i^2)}(n-k-1) \,, \qquad (13)$$

where d_i is defined in (10). This expression is very similar to the first term on the right-hand side of (10). Therefore, to try to control the influence of the high-leverage observations in (13), we propose a *quasi likelihood displacement*:

$$LD_{(i)}(w_i) = -2\left[L(\hat{\theta}) - L_{(i)}(\hat{\theta}_{w_i} \mid w_i)\right] + \left[\operatorname{var}\left(\hat{Y}\right) - \operatorname{var}\left(\hat{Y}_w\right)\right].$$

Thus, the slope of the maximum increment direction of $LD_{(i)}(w_i)$ is

$$\ell^*_{i(i)} = \ell^*_{\max,i(i)} = 1 - \frac{d_i^2}{(1-p_{ii})(1-d_i^2)}(n-k-1) - \frac{\hat{\sigma}^2 p_{ii}}{(1-p_{ii})^2}.$$

Since $\Sigma p_{ii} = k$ while $\Sigma d_i^2 = 1$, multiplying the second term by $k\hat{\sigma}^2 / (n-k-1)$ prevents $\ell^*_{i(i)}$ from being dominated by its third term,

$$\ell^*_{i(i)} = \ell^*_{\max,i(i)} = 1 - \hat{\sigma}^2\left[\frac{kd_i^2}{(1-p_{ii})(1-d_i^2)} - \frac{p_{ii}}{(1-p_{ii})^2}\right].$$

We can consider $\ell^*_{i(i)}$ to be large if it exceeds mean($\ell^*_{i(i)}$) + c(var($\ell^*_{i(i)}$))$^{1/2}$. The problem with these cut-off points, however, is that both the mean and variance are non-robust. Extreme values change the mean and inflate variance, yielding a high cut-off point. This problem can be avoided by replacing the mean and variance by more robust estimators such as the median and the median of the absolute deviations, respectively.

5. Comparisons Between Weighting and Deletion Diagnostics Using Some Examples

We now compare the local and deletion influence measures in the context of an artificial and a real-life data set.

Example 1: This example illustrates that $\ell_i(WK^2)$ and $\ell_i(W^2)$ are more sensitive than ℓ_i and $\ell^*_{i\,(i)}$ for measuring the local influence on $\hat{\beta}$ and $\hat{\sigma}^2$. Table 1 shows a data set taken from Chaterjee and Hadi (1988), where all the observations except observation number 6 (the most influential observation on $\hat{\sigma}^2$) lie on the line y = 2 + x. As can be seen from Table 1, observation number 6 is more locally influential according to $\ell_i(WK^2)$ and $\ell_i(W^2)$ than the others. On the other hand, $\ell_i(W^2)$ gives more emphasis to the high-leverage observation.

Table 1: A small data set illustrating that $\ell_i(WK^2)$ and $\ell_i(W^2)$ are more sensitive than ℓ_i for measuring local influence on both $\hat{\beta}$ and $\hat{\sigma}^2$

n. case	x	y	e_i	p_{ii}	$\lvert\ell_i\rvert$	$\lvert\ell^*_{i(i)}\rvert$	$\lvert\ell_i(WK^2)\rvert$	$\lvert\ell_i(W^2)\rvert$
1	1	3	-0.27	0.84	0.90	21.33	1.60	49.12
2	5	7	0.23	0.19	0.92	0.77	1.42	8.73
3	6	8	0.35	0.17	0.82	0.75	1.30	7.81
4	7	9	0.48	0.21	0.67	0.62	1.12	7.08
5	8	10	0.60	0.30	0.47	0.27	0.88	6.33
6	8	8	-1.40	0.30	1.79	4.07	20859263.80	149300894.00

Example 2: Scottish Hill Races Data. This data set gives the record times (in seconds) of 35 Scottish Hill races in 1984 along with two explanatory variables, the distance of the race (in miles) and the climb (in feet). This data set has been analyzed by Hadi (1992). The following model fits the data:

$$Time = \beta_0 + \beta_1\, Distance + \beta_2\, Climb + \varepsilon. \qquad (14)$$

The data contain two clear outliers, observations 18 with r_i = 4.6 and 7 with r_i = 2.8. For comparison purposes, we remove these two observations and refit model (14) to the remaining 33 races.

The deletion and the new local influence measures are shown in Tables 2 and 3. The values of p_{ii} and e_i show that observation 11 is a high-leverage point and that observation 17 is an outlier.

Table 2: **Scottish Hill Races Data: Various local influence measures**

Case	e_i	Case	p_{ii}	Case	$\|\ell_i\|$	Case	$\|\ell_i(WK^2)\|$	Case	$\|\ell_i(W^2)\|$	Case	$\|\ell^*_{i(i)}\|$
19	623.01	11	0.70	33	8.15	19	6.27	11	302.30	11	320298544.74
33	1098.55	33	0.26	19	1.94	33	40.23	33	1742.48	33	8944770.42

Table 3: **Scottish Hill Races Data: Various deletion influence measures**

Case	D_i	Case	WK_i^2	Case	W_i^2	Case	H_i^2
33	0.12	31	0.36	11	36.12	11	2.40
35	4.78	33	7.21	33	312.14	33	2.13

Examination of Tables 2 and 3 shows that W_i^2, H_i^2, $\ell_i(W^2)$, and $\ell^*_{i(i)}$ succeed in detecting these two observations while WK_i^2 and $\ell_i(WK^2)$ highlight only observation 33. These results indicate the similarities between the two approaches. Note that the results based on the local diagnostic are more relevant than those based on the deletion diagnostic.

6. Summary and Recommendations

On the one hand, the generalization of the Welsch-Kuh and Welsch measures to the local influence context leads us to measures more sensitive to the estimate of the variance and gives more emphasis to high-leverage points. On the other hand, $\ell^*_{i(i)}$ is an additive function of the residuals and of the leverage values which is more resistant to swamping and masking effects. If we compare the weighting and deletion results, we can see that the new local influence measures are more sensitive than the deletion influence measures. This connection between both diagnostics provides a further justification for local-influence analysis.

References

Belsey, D.A., Kuh, E. and Welsch, R.E. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity*. New York: John Wiley.

Billor, N. and Loynes, R.M. (1993) "Local Influence: A New Approach." *Comm. Statist.-Theory Meth*. 22: 1595-1611.

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

Cook, R.D. (1986) "Assessment of Local Influence (with discussion)." *Journal of the Royal Statistical Society,* Ser. B., 48: 133-169.

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

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

Hadi, A.S. (1992) "A New Measure of Overall Potential Influence in Linear Regression." *Computational Statistics and Data Analysis* 14: 1-27.

Welsch, R.E. (1982) "Influence Functions and Regression Diagnostics," In: R.L. Launer and A.F. Siegel, (ed.) *Modern Data Analysis*, New York: Academic Press.

Welsch, R.E., and Kuh, E. (1977) "Linear Regression Diagnostics." *Technical Report* 923-77, Sloan School of Management, Massachusetts Institute of Technology.

Welsch, R.E., and Peters, S.C. (1978) "Finding Influential Subsets of Data in Regression Models", In: A.R. Gallant and T.M. Gerig, (ed.) *Proceedings of the Eleventh Interface Symposium on Computer Science and Statistics*, Raleigh: Institute of Statistics, North Carolina State University.