Pak. J. Statist.
2009, Vol. 25(4), 647-662

INFLUENCE DIAGNOSTICS IN THE LINEAR REGRESSION MODEL WITH STOCHASTIC LINEAR RESTRICTIONS

S. Liu[1], **S.E. Ahmed**[2], and **L.Y. Ma**[3]

[1] Faculty of Information Sciences & Engineering, University of Canberra, Canberra, Australia
Email: shungzhe.liu@canberra.edu.au
[2] Department of of Mathematics & Statistics, University of Windsor, Windsor, Canada
Email: seahmed@uwindsor.ca
[3] College of Economics & Management, Dalian Nationalities University, Dalian, China
Email: maliying01@163.com

ABSTRACT

In this paper, we use the local influence method introduced by Cook (1986) to study for regression diagnostics in the linear model with stochastic linear restrictions. We assume normality and use a likelihood-based approach. We establish the normal curvatures for identifying influencial observations under several schemes.

KEYWORDS

Local influence. Maximum likelihood estimation. Mixed regression estimation. Prior information.

2000 Mathematics Subject Classification: 62J05.

1 INTRODUCTION

When there is a set of stochastic linear equations available as prior information on the regression coefficients in a linear regression model, the method of mixed regression estimation proposed by Theil and Goldberger (1961) and Theil (1963) can be used. This method basically assumes that the prior information in the form of stochastic linear equations is as important as the sample information in the form of observations on the dependent variable and explanatory variables; therefore, both the sample information and the prior information are treated as two equations equally valued for the model and are simultaneously dealt in the estimation procedure. During the last few decades, the mixed estimation method and its variations have been applied extensively in various situations in statistics and econometrics; for recent work in the literature, see Gross (2003), Rao, Toutenburg, Shalabh and Heumann (2008) and Yang and Xu (2009). On the other hand, in practice, some extremal or influencial observations or outliers, for example, those studied by Cook (1977, 1986), may be involved in the data being modelled by the mixed estimation method. To our knowledge, however, no study on the influence diagnostics in the linear regression model with stochastic linear restrictions has been reported.

In this paper, we fill this gap by considering a (pseudo) likelihood-based approach to use the local influence method originated by Cook (1986) and studied by many other researchers. This is because the local influence method has played a significant role on such studies as regression diagnostics in several areas in statistics and econometrics. After Cook (1986), who studied an application in the linear regression model under the normality assumption, researchers have studied influence diagnostics for the linear model with some variations or additional specific structures. To name a few, Paula (1993) investigated influence diagnostics for the linear model with inequality constraints; Billor and Loynes (1999) applied the local influence approach to ridge regression; Leiva, Barros, Paula and Galea (2007) studied a regression model with censored data; and Osorio, Paula and Galea (2007) made an assessment of local influence in elliptical linear models with longitudinal structure.

It is worth pointing out that the presented estimation procedure for the model to be given (see (3.3) to follow) could be seen as a special case of the one discussed by Osorio et al. (2007) under elliptical distributions. However, their model is a general model that includes several other ones. It is equally interesting and useful to analyze the particular properties that arise in this special case, including those presented in this paper. On the other hand, our study could be complementary or linked to the mixed estimation used in the collinearity problem in linear regression; see, for example, Belsley et al. (1980, Section 4.1).

The present paper is proceeded as follows: in section 2, the local influence method is outlined; in section 3, the postulated linear regression model and estimation results are presented; in section 4, several matrices for the normal curvatures and vectors for the slopes of local influence under the perturbed models are derived; and in section 5, concluding remarks are included.

2 LOCAL INFLUENCE

The local influence method was first proposed by Cook (1986) for assessing the influence of small perturbations in a general statistical model. It was then studied and used by many other researchers. We consider two approaches for the local influence method: one proposed by Cook (1986), and the other by Tsai (1986) and Billor and Loynes (1993).

2.1 Cook's approach

Let $L(\theta)$ represent the log-likelihood for the postulated (i.e. unperturbed) model and observed data, where θ is a $p \times 1$ vector of unknown parameters with its maximum likelihood (ML) estimator $\hat{\theta}$. Let $w = (w_1, ..., w_m)^\top$ denote an $m \times 1$ vector of the (small) perturbations in the model, where Ω represents the open set of relevant perturbations such that $w \in \Omega$, and then $L(\theta|w)$ be the log-likelihood of the perturbed model and $\hat{\theta}_w$ denote the corresponding ML estimator of θ. Let $w_0 \in \Omega$ denote an $m \times 1$ no-perturbation vector with $w_0 = (0, ..., 0)^\top$, or $w_0 = (1, ..., 1)^\top$, or a third choice, depending on the context, such that $L(\theta) = L(\theta|w_0)$. Suppose that $L(\theta|w)$ is twice continuously differentiable in a neighbourhood of $(\hat{\theta}, w_0)$. We are interested in comparing the parameter estimates $\hat{\theta}$ and $\hat{\theta}_w$ by using the idea of local influence. To implement the idea is to investigate the extent to which the inference is affected by the corresponding perturbation. The key concepts and the method are as follows.

In Cook's approach the likelihood displacement is chosen to be

$$LD(w) = 2[L(\hat{\theta}) - 2L(\hat{\theta}_w)] \tag{2.1}$$

which can be used to assess the influence of the perturbation w. Here, large values of $LD(w)$ indicate that $\hat{\theta}$ and $\hat{\theta}_w$ differ considerably relative to the contours of the unperturbed log-likelihood $L(\theta)$. This method is based on studying the local behaviour of an influence graph $a(w) = (w^\top, LD(w))^\top$ around w_0. Cook (1986) suggests investigating the direction in which this influence measure changes most rapidly locally, i.e. the maximum curvature of the surface $a(w)$. Upon $LD(w)$ the maximum curvature C_{max} is given by $C_{max} = max_{||l||=1} C_l$, where $C_l = 2|l^\top F l|$. To find C_{max} and the corresponding direction l_{max}, we need to calculate the $m \times m$ matrix F, which is defined by

$$F = \Delta^\top H^{-1} \Delta, \tag{2.2}$$

where Δ is a $p \times m$ matrix for the perturbed model

$$\Delta = \frac{\partial^2 L(\theta|w)}{\partial\theta\partial w^\top}|_{\theta=\hat{\theta},w=w_0}$$

evaluated at $\hat{\theta}$ and w_0, and $-H$ is the $m \times m$ observed information matrix for the postulated model

$$H = \frac{\partial^2 L(\theta)}{\partial\theta\partial\theta^\top}|_{\theta=\hat{\theta}}$$

evaluated at $\hat{\theta}$; then l_{max} is the eigenvector that corresponds to the largest absolute eigenvalue C_{max} of F, and large values of those elements of l_{max} indicate the corresponding observations may be influential.

2.2 An alternative approach

In addition to Cook's (1986) approach, Tsai (1986) and Billor and Loynes (1993) propose a different measure. The following likelihood displacement is considered:

$$LD^*(w) = -2[L(\hat{\theta}) - L(\hat{\theta}_w|w)]. \tag{2.3}$$

We have a surface $a^*(w) = (w^\top, LD^*(w))^\top$. In this case, the first derivative S_l, in the direction of l, at w_0 does not vanish, except in trivial cases, and therefore provides valuable information about the local behaviour of LD^*. In particular, the maximum slope S_{max} and the corresponding direction l_{max} are useful summaries. Because this approach only involves first derivatives rather than second derivatives, it is easier to deal with than Cook's local influence. In fact, $S_l = 2l^\top h$, the maximum slope is

$$S_{max} = 2|\frac{\partial L(\hat{\theta}|w)}{\partial w}|$$

and thus for finding S_{max} in this approach we simply need to calculate

$$h = \frac{\partial L(\theta|w)}{\partial w}|_{\theta=\hat{\theta},w=w_0}.$$

3 MODELLING AND ESTIMATION

We consider the linear regression model

$$y = X\beta + u, \tag{3.1}$$

where y is an $n \times 1$ observable random rector, u is the error vector with expectation $E(u) = 0$ and covariance matrix $\mathrm{var}(u) = \sigma^2 I$, X is an $n \times p$ known design matrix of rank p, and β is a $p \times 1$ vector of parameters. In addition, β is under stochastic linear restrictions such that

$$r = R\beta + v, \tag{3.2}$$

where r is a $k \times 1$ observable random rector, v is a $k \times 1$ error vector with expectation $E(v) = 0$ and covariance matrix $\mathrm{var}(v) = \sigma^2 V$, R is a $k \times p$ matrix of known constants of rank k ($k \leq p$), and V is a $k \times k$ positive definite matrix of known constants. Furthermore, it is assumed that the elements of v are stochastically independent of the elements of u.

3.1 Mixed estimation

By virtue of the mixed estimation we have the model rewritten as

$$z = Z\beta + e, \tag{3.3}$$

where

$$z = \begin{pmatrix} y \\ r \end{pmatrix}, \quad Z = \begin{pmatrix} X \\ R \end{pmatrix}, \quad e = \begin{pmatrix} u \\ v \end{pmatrix} \tag{3.4}$$

are of order $m \times 1$, $m \times p$ and $m \times 1$ ($m = n + k$), respectively, and e has mean $E(e)$ of order $m \times 1$ and variance matrix $\mathrm{var}(e)$ of order $m \times m$

$$E(e) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathrm{var}(e) = \sigma^2 Q \quad \text{with} \quad Q = \begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix}. \tag{3.5}$$

The mixed estimator of β is actually based on the generalised least squares method

$$\begin{aligned} b &= (Z^\top Q^{-1} Z)^{-1} Z^\top Q^{-1} z \\ &= (X^\top X + R^\top V^{-1} R)^{-1}(X^\top y + R^\top V^{-1} r). \end{aligned} \tag{3.6}$$

Note that σ^2 could be estimated as follows, which is also based on generalised least squares and can be seen in Rao et al. (2008, section 4.2), for example.

$$\begin{aligned} s^2 &= \frac{1}{m-p} \hat{e}^\top Q^{-1} \hat{e} \\ &= \frac{1}{m-p}[(y - X\hat{\beta})^\top (y - X\hat{\beta}) + (r - R\hat{\beta})^\top V^{-1}(r - R\hat{\beta})], \end{aligned} \tag{3.7}$$

where $\hat{e} = z - Z\hat{\beta}$.

3.2 Maximum likelihood estimation

In this paper, we consider a (pseudo) likelihood-based approach along the line as Billor and Loynes (1999), among others, have done for ridge regression to use the local influence method. That is, we assume a distribution for the errors in the model in order to consider ML estimation, and then use the local influence method to find the normal curvatures and slopes. We need to derive the ML estimators and the normality assumption for it is made as follows

$$e \sim N(0, \sigma^2 Q), \quad e = z - Z\beta \tag{3.8}$$

The ML estimators are then obtained as follows.

Theorem 1. *For model (3.1) with stochastic linear restrictions (3.2), we have*

$$\hat{\beta} = (X^\top X + R^\top V^{-1} R)^{-1}(X^\top y + R^\top V^{-1} r), \tag{3.9}$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{m}\hat{e}^\top Q^{-1}\hat{e} \\ &= \frac{1}{m}[(y - X\hat{\beta})^\top (y - X\hat{\beta}) + (r - R\hat{\beta})^\top V^{-1}(r - R\hat{\beta})], \end{aligned} \tag{3.10}$$

where $\hat{e} = z - Z\hat{\beta}$.

Proof. We use the following relevant part of the log-likelihood $L(\theta)$

$$L = -\frac{m}{2} ln\sigma^2 - \frac{1}{2} ln|Q| - \frac{1}{2\sigma^2} e^\top Q^{-1} e, \tag{3.11}$$

where $\theta = (\beta^\top, \sigma^2)^\top$. Using the matricx calculus developed by Magnus and Neudecker (1999) and taking the differential of L with respect to β, we have

$$d_\beta L = \frac{1}{\sigma^2} e^\top Q^{-1} Z d\beta. \tag{3.12}$$

Taking the differential of L with respect to σ^2, we have

$$d_{\sigma^2} L = (-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} e^\top Q^{-1} e) d\sigma^2. \tag{3.13}$$

By setting the equations equal to 0 and solving the solutions for θ, we get the ML estimators $\hat{\beta}$ and $\hat{\sigma}^2$.

Note that the estimates calculated by the formulas in the theorem above will be used in the normal curvatures to be established in the following section.

4 NORMAL CURVATURES

For the postulated and perturbed models, we now derive the observaed information matrix $-H$ and derivative matrix Δ for Cook's approach, and derivative vector h for the alternative approach, so that we can establish the normal curvatures.

4.1 Observaed information matrix

For our postulated model, the parameter vector θ is partioned $\theta = (\beta^\top, \sigma^2)^\top$. The observaed information matrix $-H$ is partioned accordingly.

Theorem 2. *For model (3.1) with stochastic linear restrictions (3.2), we have*

$$-H = \begin{pmatrix} \frac{1}{\hat\sigma^2} Z^\top Q^{-1} Z & \frac{1}{\hat\sigma^4} Z^\top Q^{-1} \hat e \\ \frac{1}{\hat\sigma^4} \hat e^\top Q^{-1} Z & \frac{m}{2\hat\sigma^4}, \end{pmatrix} \tag{4.1}$$

where $\hat e = y - Z\hat\beta$.

Proof. Taking the differential of $d_\beta L$ with respect to β, we get $d^2_\beta L$ evaluated at $\beta = \hat\beta$

$$d^2_\beta L = -\frac{1}{\sigma^2} d\beta^\top Z^\top Q^{-1} Z d\beta \tag{4.2}$$

$$d^2_\beta L|_{\theta=\hat\theta} = -\frac{1}{\hat\sigma^2} d\beta^\top Z^\top Q^{-1} Z d\beta. \tag{4.3}$$

Taking the differential of $d^2_{\sigma^2} L$ with respect to σ^2, we have

$$d^2_{\sigma^2} L = (\frac{m}{2\sigma^4} - \frac{1}{\sigma^6} e^\top Q^{-1} e) d^2\sigma^2 \tag{4.4}$$

$$d^2_{\sigma^2} L|_{\theta=\hat\theta} = (-\frac{m}{2\hat\sigma^4}) d^2\sigma^2, \tag{4.5}$$

where $\hat\sigma^2 = \frac{1}{m}\hat e^\top Q^{-1}\hat e$ in (3.10) is used.

Taking the differential of $d_\beta L$ with respect to σ^2, we have

$$d^2_{\beta\sigma^2} L = -\frac{1}{\sigma^4} d\beta^\top Z^\top Q^{-1} e d\sigma^2 \tag{4.6}$$

$$d^2_{\beta\sigma^2} L|_{\theta=\hat\theta} = -\frac{1}{\hat\sigma^4} d\beta^\top Z^\top Q^{-1} \hat e d\sigma^2. \tag{4.7}$$

Hence by using (2.2), (4.3), (4.5) and (4.7) and rearranging the terms, we establish $-H$.

4.2 Perturbation of case weights

Model (3.1) with (3.2) is our postulated model. Let W be an $n \times n$ diagonal matrix of perturbation with positive elements and $W_0 = I$ be the matrix of no-perturbation such that $L(\theta|W_0) = L(\theta)$. Using $\sigma^2 W^{-1}$ instead of $\sigma^2 I$, where I is an $n \times n$ idenetity matrix, we get the relevant part of the log-likelihood $L(\theta|w)$ for our perturbed model in case weights

$$L = -\frac{m}{2} ln\sigma^2 - \frac{1}{2} ln|Q_w| - \frac{1}{2\sigma^2} e^\top Q_w^{-1} e, \tag{4.8}$$

where

$$Q_w = \begin{pmatrix} W^{-1} & 0 \\ 0 & V \end{pmatrix}. \tag{4.9}$$

Theorem 3. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$\Delta = \begin{pmatrix} \frac{1}{\hat{\sigma}^2} X^\top (\hat{u}^\top \otimes I) J \\ \frac{1}{2\hat{\sigma}^4} (\hat{u}^\top \otimes \hat{u}^\top) J \end{pmatrix}, \tag{4.10}$$

where $\hat{u} = y - X\hat{\beta}$, $\otimes$ *is Kronecker matrix product, and* J *is the* $n^2 \times n$ *selection matrix as used in Neudecker, Polasek and Liu (1995), for example.*

Proof. Taking the differentials of L with respect to β we have

$$d_\beta L = \frac{1}{\sigma^2} d\beta^\top Z^\top Q_w^{-1} e \tag{4.11}$$

and then with respect to w

$$\begin{aligned} d^2_{\beta w} L &= \frac{1}{\sigma^2} d\beta^\top Z^\top \begin{pmatrix} dW & 0 \\ 0 & 0 \end{pmatrix} e \qquad (4.12) \\ &= \frac{1}{\sigma^2} d\beta^\top X^\top dW u \\ &= \frac{1}{\sigma^2} d\beta^\top X^\top \mathrm{vec}(dW u) \\ &= \frac{1}{\sigma^2} d\beta^\top X^\top (u^\top \otimes I) \mathrm{vec} dW \\ &= \frac{1}{\sigma^2} d\beta^\top X^\top (u^\top \otimes I) J dw \\ d^2_{\beta w} L|_{\theta=\hat{\theta}, w=w_0} &= \frac{1}{\hat{\sigma}^2} d\beta^\top X^\top (\hat{u}^\top \otimes I) J dw, \qquad (4.13) \end{aligned}$$

where vec is the vectorisation operator, w is an $n \times 1$ vector of the diagonal elements of W, $\hat{u} = y - X\hat{\beta}$, and $\hat{\beta}$ and $\hat{\sigma}^2$ are as given in Theorem 1.

Taking the differential of L with respect to σ^2 we have

$$d_{\sigma^2} L = (-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} e^\top Q_w^{-1} e) d\sigma^2 \tag{4.14}$$

and then with respect to w

$$\begin{aligned} d^2_{\sigma^2 w} L &= \frac{1}{2\sigma^4} d\sigma^2 u^\top dW u \\ &= \frac{1}{2\sigma^4} d\sigma^2 \mathrm{vec}(u^\top dW u) \\ &= \frac{1}{2\sigma^4} d\sigma^2 (u^\top \otimes u^\top) \mathrm{vec} dW \\ &= \frac{1}{2\sigma^4} d\sigma^2 (u^\top \otimes u^\top) J dw \\ d^2_{\sigma^2 w} L|_{\theta=\hat{\theta}, w=w_0} &= \frac{1}{2\hat{\sigma}^4} d\sigma^2 (\hat{u}^\top \otimes \hat{u}^\top) J dw \end{aligned} \tag{4.15}$$

Hence by rearranging the terms we establish Δ.

Theorem 4. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$h = J^\top [\frac{1}{2} \mathrm{vec}(I) - \frac{1}{2\hat{\sigma}^2} \mathrm{vec}(\hat{u}\hat{u}^\top)], \tag{4.16}$$

where h is an $n \times 1$ vector.

Proof. Using $P_w = Q_w^{-1}$ and its differential with respect to w

$$d_w P_w = \begin{pmatrix} dW & 0 \\ 0 & 0 \end{pmatrix} \tag{4.17}$$

we can rewrite L to be

$$L = -\frac{m}{2} ln\sigma^2 + \frac{1}{2} ln|P_w| - \frac{1}{2\sigma^2} e^\top P_w e. \tag{4.18}$$

We get, by taking the differentials of L with respect to w

$$\begin{aligned}
d_w L &= \frac{1}{2}\mathrm{tr}P_w^{-1}dP_w - \frac{1}{2\sigma^2}e^\top dP_w e \\
&= \frac{1}{2}\mathrm{tr}W^{-1}dW - \frac{1}{2\sigma^2}u^\top dWu \\
&= \frac{1}{2}\mathrm{vec}^\top(W^{-1})\mathrm{vec}dW - \frac{1}{2\sigma^2}\mathrm{tr}(uu^\top dW) \\
&= \frac{1}{2}\mathrm{vec}^\top(W^{-1})Jdw - \frac{1}{2\sigma^2}\mathrm{vec}^\top(uu^\top)Jdw \qquad (4.19) \\
d_w L|_{\theta=\hat{\theta},w=w_0} &= [\frac{1}{2}\mathrm{vec}^\top(I) - \frac{1}{2\hat{\sigma}^2}\mathrm{vec}^\top(\hat{u}\hat{u}^\top)]Jdw, \qquad (4.20)
\end{aligned}$$

where $\hat{u} = y - X\hat{\beta}$.

Hence by rearranging the terms we establish h.

4.3 Perturbation of response

Replacing y by $y + w$ with $w_0 = 0$, we get the relevant part of the log-likelihood $L(\theta|w)$ for our perturbed model

$$L = -\frac{m}{2}ln\sigma^2 - \frac{1}{2}ln|Q| - \frac{1}{2\sigma^2}e_w^\top Q^{-1}e_w \qquad (4.21)$$

where

$$e_w = \begin{pmatrix} y+w \\ r \end{pmatrix} - Z\beta. \qquad (4.22)$$

Theorem 5. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$\Delta = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}X^\top \\ \frac{1}{\hat{\sigma}^4}\hat{u}^\top \end{pmatrix} \qquad (4.23)$$

where $\hat{u} = y - X\hat{\beta}$.

Proof. Taking the differentials of L with respect to β we have

$$d_\beta L = \frac{1}{\sigma^2}d\beta^\top Z^\top Q^{-1}e_w \qquad (4.24)$$

and then with respect to w

$$\begin{aligned} d^2_{\beta w}L &= \frac{1}{\sigma^2}d\beta^\top Z^\top Q^{-1}\begin{pmatrix} dw \\ 0 \end{pmatrix} \\ &= \frac{1}{\sigma^2}d\beta^\top X^\top dw \\ d^2_{\beta,w}L|_{\theta=\hat{\theta},w=w_0} &= \frac{1}{\hat{\sigma}^2}d\beta^\top X^\top dw, \end{aligned}$$

where $\hat{u} = y - X\hat{\beta}$.

Taking the differential of L with respect to σ^2 we have

$$d_{\sigma^2}L = (-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}e_w^\top Q^{-1}e_w)d\sigma^2 \tag{4.25}$$

and then with respect to w

$$\begin{aligned} d^2_{\sigma^2 w}L &= \frac{1}{\sigma^4}d\sigma^2 e_w^\top Q^{-1}\begin{pmatrix} dw \\ 0 \end{pmatrix} \\ &= \frac{1}{\sigma^4}d\sigma^2 u_w^\top dw \\ d^2_{\sigma^2 w}L|_{\theta=\hat{\theta},w=w_0} &= \frac{1}{\hat{\sigma}^4}d\sigma^2 \hat{u}^\top dw \end{aligned} \tag{4.26}$$

where $u_w = y + w - X\beta$ and $\hat{u}_w = \hat{u} = y - X\hat{\beta}$.

Hence by rearranging the terms we establish Δ.

Theorem 6. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$h = -\frac{1}{\hat{\sigma}^2}\hat{u}, \tag{4.27}$$

where $\hat{u} = y - X\hat{\beta}$.

Proof. Taking the differentials of L with respect to w we have

$$\begin{aligned} d_w L &= -\frac{1}{\sigma^2} e_w^\top Q^{-1} de_w \\ &= -\frac{1}{\sigma^2} e_w^\top Q^{-1} \begin{pmatrix} dw \\ 0 \end{pmatrix} \\ &= -\frac{1}{\sigma^2} u_w^\top dw \\ d_w L|_{\theta=\hat{\theta}, w=w_0} &= -\frac{1}{\hat{\sigma}^2} \hat{u}^\top dw. \end{aligned} \tag{4.28}$$

Hence by rearranging the terms we establish h.

4.4 Perturbation of explanatory variables

Replacing X by $X + W$ with $W_0 = 0$, we get the relevant part of the log-likelihood $L(\theta|w)$ for our perturbed model

$$L = -\frac{m}{2} ln\sigma^2 - \frac{1}{2} ln|Q| - \frac{1}{2\sigma^2} e_w^\top Q^{-1} e_w, \tag{4.29}$$

where $e_w = z - Z_w \beta$ and

$$Z_w = \begin{pmatrix} X + W \\ R \end{pmatrix}. \tag{4.30}$$

Theorem 7. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$\Delta = \begin{pmatrix} \frac{1}{\hat{\sigma}^2} [(u^\top \otimes I)D - X^\top (\hat{\beta}^\top \otimes I)] \\ -\frac{1}{\hat{\sigma}^4} \hat{u}^\top (\hat{\beta}^\top \otimes I) \end{pmatrix}, \tag{4.31}$$

where $\hat{u} = y - X\hat{\beta}$*, and* D *is the* $np \times np$ *duplication matrix as used in Magnus and Neudecker (1999).*

Proof. Taking the differentials of L with respect to β we have

$$d_\beta L = \frac{1}{\sigma^2} d\beta^\top Z_w^\top Q^{-1} e_w \tag{4.32}$$

and then with respect to w

$$
\begin{aligned}
d^2_{\beta w}L &= \frac{1}{\sigma^2}d\beta^\top \begin{pmatrix} dW^\top \\ 0 \end{pmatrix} Q^{-1}e_w - \frac{1}{\sigma^2}d\beta^\top Z_w^\top Q^{-1}\begin{pmatrix} dW \\ 0 \end{pmatrix}\beta \\
&= \frac{1}{\sigma^2}d\beta^\top dW^\top u_w - \frac{1}{\sigma^2}d\beta^\top X_w^\top dW\beta, \qquad (4.33)
\end{aligned}
$$

where $u_w = y - (X+W)\beta$ and $X_w = X + W$.

Using $W_0 = 0$ we get

$$
\begin{aligned}
d^2_{\beta w}L|_{\theta=\hat{\theta},W=W_0} &= \frac{1}{\hat{\sigma}^2}d\beta^\top dW^\top \hat{u} - \frac{1}{\hat{\sigma}^2}d\beta^\top X^\top dW\hat{\beta} \\
&= \frac{1}{\hat{\sigma}^2}d\beta^\top \mathrm{vec}(dW^\top \hat{u}) - \frac{1}{\hat{\sigma}^2}d\beta^\top X^\top \mathrm{vec}(dW\hat{\beta}) \\
&= \frac{1}{\hat{\sigma}^2}d\beta^\top(\hat{u}^\top \otimes I)D\mathrm{vec}dW - \frac{1}{\hat{\sigma}^2}d\beta^\top X^\top(\hat{\beta}^\top \otimes I)\mathrm{vec}dW \\
&= \frac{1}{\hat{\sigma}^2}d\beta^\top[(\hat{u}^\top \otimes I)D - X^\top(\hat{\beta}^\top \otimes I)]\mathrm{vec}dW. \qquad (4.34)
\end{aligned}
$$

Taking the differential of L with respect to σ^2 we have

$$
d_{\sigma^2}L = (-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}e_w^\top Q^{-1}e_w)d\sigma^2 \qquad (4.35)
$$

and then with respect to w

$$
\begin{aligned}
d^2_{\sigma^2 w}L &= -\frac{1}{\sigma^4}d\sigma^2 e_w^\top Q^{-1}\begin{pmatrix} dW \\ 0 \end{pmatrix}\beta \\
&= -\frac{1}{\sigma^4}d\sigma^2 u_w^\top dW\beta \\
&= -\frac{1}{\sigma^4}d\sigma^2 u_w^\top(\beta^\top \otimes I)\mathrm{vec}dW \\
d^2_{\sigma^2 w}L|_{\theta=\hat{\theta},w=w_0} &= -\frac{1}{\hat{\sigma}^4}d\sigma^2 \hat{u}^\top(\hat{\beta}^\top \otimes I)\mathrm{vec}dW. \qquad (4.36)
\end{aligned}
$$

Hence by rearranging the terms we establish Δ.

Theorem 8. *For the perturbed case of model (3.1) with stochastic linear restrictions (3.2), we have*

$$
h = \frac{1}{\hat{\sigma}^2}(\hat{\beta} \otimes \hat{u}), \qquad (4.37)
$$

where $\hat{u} = y - X\hat{\beta}$.

Proof. Using $de_w = -dZ_w\beta$ and taking the differentials of L with respect to w we have

$$\begin{aligned} d_wL &= -\frac{1}{\sigma^2}e_w^\top Q^{-1}de_w \\ &= \frac{1}{\sigma^2}e_w^\top Q^{-1}\begin{pmatrix} dW \\ 0 \end{pmatrix}\beta \\ &= \frac{1}{\sigma^2}u_w^\top dW\beta \\ &= \frac{1}{\sigma^2}\text{vec}(u_w^\top dW\beta) \\ &= \frac{1}{\sigma^2}(\beta^\top \otimes u_w^\top)\text{vec}dW \\ d_wL|_{\theta=\hat{\theta},w=w_0} &= \frac{1}{\hat{\sigma}^2}(\hat{\beta}^\top \otimes \hat{u}^\top)\text{vec}dW. \end{aligned} \quad (4.38)$$

Hence by rearranging the terms we establish h.

5 CONCLUDING REMARKS

In this paper, we have studied the influence diagnostics in the linear model with stochastic linear restrictions. We have derived for the three purturbation schemes the matrices $-H$ and Δ for Cook's approach, and h for the alternative approach. They can be used to calculate matrix F and, therefore, C_{max} and l_{max} for Cook's approach, and S_{max} and l^*_{max} for the alternative approach for the local influence under the purturbation schemes. Local influence assessment can then be made based on the largest normal curvature C_{max} and slope S_{max}, and their corresponding direction l_{max}.

Regarding the different schemes, it can be seen from the results of $-H$, Δ and h that (maybe slightly) different indications on which observations are most influential are expected. However, they are all dependent on the ML estimates of the parameters β and σ^2 in the model in which both sample and prior information are used.

Some future work as suggested by the reviewers of an earlier version of this paper may include issues discussed in Rao et al. (2008, Section 5.10.3). They considered a mean shift outlier model for potential outliers in the equation of stochastic linear restrictions. We may consider perturbation schemes associated with the stochastic restrictions. In addition, the connection of the mixed estimation, influential diagnostics and the collinearity problem in linear regression is worth being further explored.

Acknowledgements

We would like to greatly thank our colleague, Victor Leiva, and the reviewers of an earlier version of this paper for their constructive comments and useful suggestions. The research for the present paper had been discussed and conducted while Shuangzhe Liu met the other two co-authors, and visited Dalian Nationalities University in China. He is grateful to his home institution for supporting his OSP.

REFERENCES

1. Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying influential data ans sources of collinearity.* Wiley, New York.

2. Billor, N. and Loynes, R.M. (1993). Local influence: a new approach. *Commun. Statist. Theor. Meth.*, **22**, 1595-1611.

3. Billor, N. and Loynes, R.M. (1999). An application of the local influence approach to ridge regression. *J. Appl. Statist.*, **2**, 177-183.

4. Cook, D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19**, 15-18.

5. Cook, D. (1986). Assessment of local influence. *J. Roy. Statist. Soc.*, **48B(2)**, 133-169.

6. Gross, J. (2003). *Linear Regression*, Springer, Berlin.

7. Leiva, V., Barros, M., Paula, G.A. and Galea, M. (2007). Influence diagnostics in log-Birnbaum Saunders regression models with censored data. *Comput. Statist. Data Anal.*, **51**, 5694-5707.

8. Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition*, Wiley, Chichester.

9. Neudecker, H., Liu, S. and Polasek, W. (1995) The Hadamard product and some of its applications in statistics, *Statistics*, **26**, 365-373.

10. Osorio, F., Paula, G.A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Comput. Statist. Data Anal.*, **51**, 4354-4368.

11. Paula, G.A. (1993). Assessing local influence in restricted regression models. *Comput. Statist. Data Anal.*, **16**, 63-79.

12. Rao, C.R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear Models and Generalizations*, Springer, Berlin.

13. Tsai, C.L. (1986). Discussion of assessment of local influence by R. D. Cook. *J. Roy. Statist. Soc. Ser. B.*, **48**, 165.

14. Theil, H. and Goldberger, A.S. (1961). On pure and mixed estimation in econometrics, *Int. Econ. Rev.*, **2**, 65–78.

15. Theil, H. (1963). On the use of incomplete prior information in regression analysis *J. Amer. Statist. Assoc.*, **58**, 401–414.

16. Yang, H. and Xu, J. (2009). An alternative stochastic restricted Liu estimator in linear regression. *Statist. Papers*, **50**, 639-647.