

Influence diagnostic analysis in the possibly heteroskedastic linear model with exact restrictions

S. Liu · V. Leiva · T. Ma · A. Welsh

Abstract The local influence method has proven to be a useful and powerful tool for detecting influential observations on the estimation of model parameters. This method has been widely applied in different studies related to econometric and statistical modelling. We propose a methodology based on the Lagrange multiplier method with a linear penalty function to assess local influence in the possibly heteroskedastic linear regression model with exact restrictions. The restricted maximum likelihood estimators and information matrices are presented for the postulated model. Several perturbation schemes for the local influence method are investigated to identify potentially influential observations. Three real-world examples are included to illustrate and validate our methodology.

Keywords Information matrix; Local influence; Restricted least-squares estimator; Restricted maximum likelihood estimator

Mathematics Subject Classification (2010) 62J05; 62J20

1 Introduction

The linear regression model with exact linear restrictions is widely used in applied econometrics and statistics. An example in financial econometrics is the simple linear model for the excess return of a security regressed on the excess return of the market portfolio. When the intercept is set equal to zero, this becomes the capital asset pricing model.

The method of restricted least squares (RLS) provides natural estimators of the regression coefficients in a linear model with exact restrictions. Basic distributional properties of the RLS estimator, efficiency comparisons, hypothesis tests and real-world applications can be found in Chipman & Rao (1964), Trenkler (1987), Ramanathan (1993), Greene (2007) and Wooldridge (2013). It is well known that the RLS estimator can be expressed in terms of the ordinary least squares (OLS) estimator. In particular, Magnus & Neudecker (1999) studied the RLS estimator in various specific situations. Gross (2003) and Rao et al (2008) explored its relationship to a ridge estimator. However, it seems that few publications so far have treated the RLS estimator from a viewpoint of influence diagnostic or sensitivity analyses, even though such analyses are needed and have increasingly been used; see, for example, the monographs Cook & Weisberg (1982) for early studies on residuals and influence in regression, Chatterjee & Hadi (1988) and Rao et al (2008) for sensitivity analysis in linear regression, Atkinson & Riani (2000) and Atkinson et al (2004) for robust diagnostic regression analysis, and Kleiber & Zeileis (2008) for regression diagnostics incorporated in a recent textbook on applied

Shuangzhe Liu
Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra ACT 2601, Australia
E-mail: shuangzhe.liu@canberra.edu.au

Víctor Leiva
Faculty of Engineering and Sciences, Adolfo Ibáñez University, Valparaíso, Chile
Institute of Statistics, University of Valparaíso, Valparaíso, Chile
E-mail: victorleivasanchez@gmail.com

Tiefeng Ma
School of Statistics, Southwestern University of Finance and Economics, Chengdu 610072, China
E-mail: matiefeng@swufe.edu.cn

Alan Welsh
Mathematical Sciences Institute, Australian National University, Canberra ACT 0200, Australia
E-mail: alan.welsh@anu.edu.au

econometrics. To our knowledge, only Liu & Neudecker (2007) provided a local sensitivity result for the RLS estimator. Two alternatives to the RLS method are the model reduction and Lagrange multiplier methods. As pointed out by Hocking (2003, p. 622), in the model reduction method it is difficult to assess the effect of adjoining restrictions due to the lack of a specific expression for the reduced model. The Lagrange multiplier method is more amenable to such theoretical developments. Therefore, in the present paper, our approach is based on this last method. We deal with the estimation and diagnostic issues in a systematic manner, which links the unique solution of the restricted maximum likelihood (RML) and RLS estimators to the local influence method both proven to be very useful in practice.

The local influence method was introduced by Cook (1986). Subsequently, alternative approaches were suggested by, for example, Billor & Loynes (1993, 1999), Poon & Poon (1999) and Shi & Huang (2011). Unlike case deletion methods, which can suffer from masking effects, the local influence method is a powerful tool for detecting observations that can be influential in the estimation of model parameters. It now becomes a general method for assessing the influence of local departures from model assumptions on maximum likelihood (ML) estimates. The local influence method is employed in several areas of applied econometrics and statistics. For example, there are a number of applications and studies in regression modelling and time series analysis; see Cook (1986), Galea et al. (1997), Liu (2000, 2002, 2004), Díaz-García et al (2003), Galea et al. (2008) and Shi & Chen (2008) for studies in linear regression and time series models, de Castro et al. (2007) and Galea & de Castro (2012) for heteroskedastic errors-in-variables models, Leiva et al (2007, 2014) for influence diagnostics with censored and uncensored data, Barros et al (2010) for a Tobit model and Paula et al (2012) for robust modelling applied to insurance data. In particular, the local influence method can play an important role in regression models involving restrictions. Paula (1993) used this method to handle the linear model with inequality restrictions; Cysneiros & Paula (2005) and Paula & Cysneiros (2010) considered parameter constraints in univariate symmetrical linear regression models; and Liu et al (2009) studied a normal linear model with stochastic linear restrictions.

The objective of this paper is to provide a methodology to assess local influence in the possibly heteroskedastic linear regression model with exact linear restrictions. While influence diagnostics were studied for the spherical linear model by Cook (1986) and the capital asset pricing model by Galea et al. (2008), no such studies have been carried out for the RLS estimator. Therefore, we fill this gap. In addition, heteroskedasticity is a classic issue that is widely encountered in practical situations; see, for example, Greene (2007) and Wooldridge (2013). It was investigated by de Castro et al. (2007) and Galea & de Castro (2012) in models with errors-in-variables, but not with linear restrictions. Furthermore, we present results which are relevant to but different from those considered by Cysneiros & Paula (2005), Liu & Neudecker (2007), Liu et al (2009) and Paula & Cysneiros (2010). We use the ML method under normality of the model errors to estimate the corresponding parameters with exact restrictions. We deal with not only the general linear model for spherical disturbances, corresponding to the univariate models studied by Cysneiros & Paula (2005) and Paula & Cysneiros (2010), but also the general linear model for non-spherical or heteroskedastic disturbances as an extension. Our approach is distinguished from the quadratic penalty function approach considered in Cysneiros & Paula (2005) and Paula & Cysneiros (2010) by the fact that we employ the Lagrange multiplier method with a linear penalty function, and that the RML estimators are closely related to the RLS or generalized least-squares (GLS) estimators. In addition, we introduce three global influence statistics and compare our local influence statistics with them.

Note that exact restrictions in the model often arise from past experience, economic or finance theory, or the area under study, which must be treated as prior information. We treat such prior information as equivalent to the sample data, although we also allow the possibility that the linear restrictions on the coefficients are not really prior information, but are just a null hypothesis, proposed to simplify the relationship between the response variable (called response hereafter) and explanatory variables (called covariables hereafter).

The rest of the paper is organized as follows. In Sections 2 and 3, we propose methodology to assess local influence in the normal linear regression model with restrictions. Specifically, in Section 2, we formulate and discuss the model, RML estimation, the information matrix and inference. In Section 3, we describe the local influence method with different perturbation matrices, and two global influence statistics for comparison purposes. In Section 4, we illustrate and validate the proposed methodology with real-world data. In Section 5, we make some concluding remarks about this work. The technical proofs of our results are included in the Appendix.

2 Modelling and estimation

In this section, we formulate the general linear regression model with restrictions in two cases: the first with spherical disturbances and the second with non-spherical disturbances. We then provide several estimators for the corresponding parameters, together with the information matrix for our formulation.

2.1 Formulation of the model

Spherical disturbance Consider the general linear regression model given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

where $\mathbf{y} = (Y_1, \dots, Y_n)^\top$ is an $n \times 1$ response vector, \mathbf{X} is an $n \times p$ known design matrix of rank p containing the values of the covariables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters to be estimated, and \mathbf{u} is an $n \times 1$ error vector with expectation $E[\mathbf{u}] = \mathbf{0}_{n \times 1}$ and variance-covariance matrix $D[\mathbf{u}] = \sigma^2 \mathbf{I}_n$. Here, $\mathbf{0}_{n \times 1}$ is an $n \times 1$ zero vector, $\sigma^2 > 0$ is an unknown parameter to be estimated and \mathbf{I}_n is the $n \times n$ identity matrix.

For the model given in (1), suppose we have prior information about $\boldsymbol{\beta}$ in the form of a set of q independent exact linear restrictions expressed as

$$\mathbf{r} = \mathbf{R}\boldsymbol{\beta}, \quad (2)$$

where \mathbf{R} is a $q \times p$ known matrix of rank $q \leq p$ and \mathbf{r} is a $q \times 1$ vector of known elements. The $k \times 1$ parameter vector in the formulation given in (1) and (2) is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$, where $k = p + 1$.

Non-spherical disturbance To extend the spherical disturbance assumption for the model given in (1), we consider a groupwise heteroskedastic case. Without loss of generality, we split the data into two groups to correspond to the non-spherical disturbance by writing $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ and $\mathbf{u} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top)^\top$, where $n_1 + n_2 = n$. We extend the model given in (1) by assuming the variance-covariance matrix $D[\mathbf{u}] = \text{diag}(\sigma_1^2 \mathbf{I}_{n_1}, \sigma_2^2 \mathbf{I}_{n_2}) = \mathbf{V}$ (say), with $\sigma_g^2 > 0$, for $g = 1, 2$, and $\sigma_1^2 \neq \sigma_2^2$ (non-spherical or heteroskedastic case). We assume the same prior information about $\boldsymbol{\beta}$ as in (2). The $k \times 1$ parameter vector is now $\boldsymbol{\theta}_6 = (\boldsymbol{\beta}^\top, \sigma_1^2, \sigma_2^2)^\top$, where $k = p + 2$.

2.2 Estimation

Spherical disturbance The RLS estimators of the elements of the parameter vector $\boldsymbol{\theta}$ in the formulation (1) and (2) are well-known to be

$$\hat{\boldsymbol{\beta}}_{\text{RLS}} = \mathbf{b} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\mathbf{b} - \mathbf{r}), \quad (3)$$

$$\hat{\sigma}_{\text{RLS}}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{RLS}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{RLS}})}{n - p + q}, \quad (4)$$

where $\mathbf{b} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the (unrestricted) OLS estimator of $\boldsymbol{\beta}$.

If we add the normality assumption $\mathbf{u} \sim N_n(\mathbf{0}_{n \times 1}, \sigma^2 \mathbf{I}_n)$ to the formulation given in (1) and (2), we can use the ML method to estimate $\boldsymbol{\theta}$. In this case, the log-likelihood function for $\boldsymbol{\theta}$ to be optimised subject to the restriction $\mathbf{r} = \mathbf{R}\boldsymbol{\beta}$ (called the restricted log-likelihood function) is given by

$$\ell = \ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}), \quad (5)$$

where $\boldsymbol{\lambda}$ is the $q \times 1$ vector of Lagrange multipliers. In order to obtain the RML estimator, we use matrix calculus to take the differential of ℓ given in (5) with respect to $\boldsymbol{\beta}$ and σ^2 and equate it to $\mathbf{0}_{k \times 1}$. We get

$$\hat{\boldsymbol{\lambda}} = \frac{[\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\mathbf{b} - \mathbf{r})}{\hat{\sigma}^2}$$

and then the corresponding RML estimators are

$$\hat{\boldsymbol{\beta}} = \mathbf{b} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\mathbf{b} - \mathbf{r}), \quad (6)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}, \quad (7)$$

where \mathbf{b} is defined in (3).

Rao et al (2008) noted that the RML and RLS estimators of $\boldsymbol{\beta}$ given in (3) and (6) are the same, but those for σ^2 given in (4) and (7) are not the same. They considered a special case with exact knowledge of a subvector and stepwise inclusion of exact linear restrictions. From (6) we obtain

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad (8)$$

$$D(\hat{\boldsymbol{\beta}}) = \sigma^2 [(\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1}]. \quad (9)$$

Note from (8) and (9) that $\hat{\boldsymbol{\beta}}$ is unbiased and more efficient than \mathbf{b} , which has variance $\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.

Non-spherical disturbance In this case, the restricted log-likelihood function for θ_G is given by

$$\ell(\theta_G) = -\frac{n}{2} \log(2\pi) - \sum_{g=1}^2 \frac{n_g}{2} \log(\sigma_g^2) - \sum_{g=1}^2 \frac{1}{2\sigma_g^2} (\mathbf{y}_g - \mathbf{X}_g \boldsymbol{\beta})^\top (\mathbf{y}_g - \mathbf{X}_g \boldsymbol{\beta}) - \boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}), \quad (10)$$

where $\boldsymbol{\lambda}$ is again the $q \times 1$ vector of Lagrange multipliers. We use the same matrix calculus to find the RML estimators under groupwise heteroskedasticity, obtaining

$$\hat{\boldsymbol{\beta}}_G = \mathbf{b}_G - (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\mathbf{b}_G - \mathbf{r}), \quad (11)$$

$$\hat{\sigma}_g^2 = \frac{(\mathbf{y}_g - \mathbf{X}_g \hat{\boldsymbol{\beta}}_G)^\top (\mathbf{y}_g - \mathbf{X}_g \hat{\boldsymbol{\beta}}_G)}{n_g}, \quad g = 1, 2, \quad (12)$$

where $\mathbf{b}_G = (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{y}$ is the GLS estimator of $\boldsymbol{\beta}$ with $\hat{\mathbf{V}} = \text{diag}(\hat{\sigma}_1^2 \mathbf{I}_{n_1}, \hat{\sigma}_2^2 \mathbf{I}_{n_2})$.

From the formulas given in (11) and (12), we need to calculate the RML estimates iteratively. We can employ a procedure starting with the OLS estimate \mathbf{b} and then iterating between $\hat{\mathbf{V}}$ and \mathbf{b}_G and $\hat{\boldsymbol{\beta}}_G$. Under the usual assumptions, when the restrictions $\mathbf{R}\mathbf{b}_G = \mathbf{r}$ are true, in large samples the distribution of $\hat{\boldsymbol{\beta}}_G$ can be approximated by a normal distribution with mean $\boldsymbol{\beta}$ and a variance matrix which is consistently estimated by

$$(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} - (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} \mathbf{R}(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}.$$

For further ideas with examples and relevant asymptotic results, see, for example, Efron & Hinkley (1978), Judge et al (1988) and Greene (2007). By assuming $\sigma_1^2 = \sigma_2^2$, we note that the results given in (11) and (12) reduce to the RML estimators for $\boldsymbol{\beta}$ given in (6) and for σ^2 given in (7). When \mathbf{V} in (11) is known, we obtain

$$\mathbb{E}(\hat{\boldsymbol{\beta}}_G) = \boldsymbol{\beta},$$

$$\mathbb{D}(\hat{\boldsymbol{\beta}}_G) = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} \mathbf{R}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

2.3 Information matrices

Spherical disturbance Using the Hessian matrix derived in Appendix A and given by

$$\mathbf{H}(\boldsymbol{\theta}) = - \begin{pmatrix} \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} & \frac{1}{\sigma^4} \mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{1}{\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{X} & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{pmatrix}, \quad (13)$$

we obtain the expected Fisher information matrix given by

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

The observed Fisher information matrix is $\mathbf{J}(\hat{\boldsymbol{\theta}}) = -\mathbf{H}(\hat{\boldsymbol{\theta}}) = \mathbf{I}(\hat{\boldsymbol{\theta}})$.

Non-spherical disturbance In this case, the Hessian matrix is obtained as

$$\mathbf{H}(\theta_G) = - \begin{pmatrix} \mathbf{h}_{\beta\beta} & \mathbf{h}_{\beta\sigma_1^2} & \mathbf{h}_{\beta\sigma_2^2} \\ \mathbf{h}_{\beta\sigma_1^2}^\top & \mathbf{h}_{\sigma_1^2\sigma_1^2} & \mathbf{h}_{\sigma_1^2\sigma_2^2} \\ \mathbf{h}_{\beta\sigma_2^2}^\top & \mathbf{h}_{\sigma_1^2\sigma_2^2}^\top & \mathbf{h}_{\sigma_2^2\sigma_2^2} \end{pmatrix}, \quad (14)$$

where

$$\begin{aligned}
\mathbf{h}_{\beta\beta} &= \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} = \frac{1}{\sigma_1^2} \mathbf{X}_1^\top \mathbf{X}_1 + \frac{1}{\sigma_2^2} \mathbf{X}_2^\top \mathbf{X}_2, \\
\mathbf{h}_{\beta\sigma_1^2} &= \frac{1}{\sigma_1^4} \mathbf{X}_1^\top (\mathbf{y}_1 - \mathbf{X}_1 \beta), \\
\mathbf{h}_{\beta\sigma_2^2} &= \frac{1}{\sigma_2^4} \mathbf{X}_2^\top (\mathbf{y}_2 - \mathbf{X}_2 \beta), \\
\mathbf{h}_{\sigma_1^2 \sigma_1^2} &= -\frac{n_1}{2\sigma_1^4} + \frac{1}{\sigma_1^6} (\mathbf{y}_1 - \mathbf{X}_1 \beta)^\top (\mathbf{y}_1 - \mathbf{X}_1 \beta), \\
\mathbf{h}_{\sigma_1^2 \sigma_2^2} &= 0, \\
\mathbf{h}_{\sigma_2^2 \sigma_2^2} &= -\frac{n_2}{2\sigma_2^4} + \frac{1}{\sigma_2^6} (\mathbf{y}_2 - \mathbf{X}_2 \beta)^\top (\mathbf{y}_2 - \mathbf{X}_2 \beta).
\end{aligned}$$

Then, the corresponding expected Fisher information matrix is given by

$$\mathbf{I}(\theta_G) = \begin{pmatrix} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} & 0 & 0 \\ 0 & \frac{n_1}{2\sigma_1^4} & 0 \\ 0 & 0 & \frac{n_2}{2\sigma_2^4} \end{pmatrix}$$

and we have $\mathbf{J}(\hat{\theta}_G) = -\mathbf{H}(\hat{\theta}_G) = \mathbf{I}(\hat{\theta}_G)$.

3 Influence diagnostics

In this section, we present the local influence method, the perturbation matrices for some different schemes, and then three global influence measures.

3.1 Local influence

Let $\ell(\theta)$ be the log-likelihood function for the model given in (1) and (2), which we call the postulated or non-perturbed model. Here θ is the $k \times 1$ unknown parameter vector to be estimated and we denote its ML estimator by $\hat{\theta}$. Let $\mathbf{w} = (w_1, \dots, w_m)^\top$ denote an $m \times 1$ perturbed vector and Ω an open set of relevant perturbations such that $\mathbf{w} \in \Omega$. Then, let $\ell(\theta|\mathbf{w}) = \ell_{\mathbf{w}}$ be the log-likelihood function for the perturbed model and $\hat{\theta}_{\mathbf{w}}$ be the corresponding ML estimator of θ . Let $\mathbf{w}_0 \in \Omega$ denote an $m \times 1$ non-perturbed vector with $\mathbf{w}_0 = (0, \dots, 0)^\top$, or $\mathbf{w}_0 = (1, \dots, 1)^\top$, or even a third choice, depending on the context, such that $\ell(\theta) = \ell(\theta|\mathbf{w}_0)$. Suppose that $\ell(\theta|\mathbf{w})$ is twice continuously differentiable in a neighborhood of $(\hat{\theta}, \mathbf{w}_0)$. We are interested in comparing the parameter estimates $\hat{\theta}$ and $\hat{\theta}_{\mathbf{w}}$ using the idea of local influence to detect how the inference is affected by the perturbation. Specifically, in the Cook local influence method, the likelihood displacement (LD) is given by

$$\text{LD}(\mathbf{w}) = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_{\mathbf{w}})),$$

which can be used to assess the influence of the perturbation \mathbf{w} . Here, large values of $\text{LD}(\mathbf{w})$ indicate that $\hat{\theta}$ and $\hat{\theta}_{\mathbf{w}}$ differ considerably in relation to the contours of the non-perturbed log-likelihood function $\ell(\theta)$. The idea is based on studying the local behaviour of $\text{LD}(\mathbf{w})$ and the normal curvature C_l in a unit-length direction vector l , where $\|l\| = 1$. According to Cook (1986), the normal curvature is given by $C_l = 2|l^\top \mathbf{F}(\theta)l|$. The maximum curvature C_{\max} and the corresponding direction vector l_{\max} may reveal those observations that exercise more influence on $\text{LD}(\mathbf{w})$, where $C_{\max} = \max_{\|l\|=1} C_l$. To examine the magnitude of influence, it is useful to have a benchmark value for C_{\max} and for the elements of l_{\max} . For C_{\max} , 2 or a q -value measure suggested by Shi & Huang (2011) can be used. For example, when the q -value is greater than 2, we can say the associated direction vector is significant for detecting influential observations. For the elements of l_{\max} , Poon & Poon (1999) suggested $1/\sqrt{n}$ and Shi & Huang (2011) noted $2/\sqrt{n}$ can be more reasonable, where n is the sample size. In our case, we consider as influential those observations with absolute values of l_{\max} exceeding $2/\sqrt{n}$. To find C_{\max} and l_{\max} , we need to calculate the $m \times m$ matrix $\mathbf{F}(\theta)$ defined by

$$\mathbf{F}(\theta) = -\Delta(\theta)^\top \mathbf{H}(\theta)^{-1} \Delta(\theta), \tag{15}$$

where $\Delta(\theta)$ is a $p \times m$ matrix for the perturbed model (perturbation matrix), which must be obtained from $d_{\theta_w}^2 \ell_w$ and evaluated at $\theta = \hat{\theta}$ and $w = w_0$. Here, $H(\theta)$ is obtained from (13) or (14). In order to detect local influence, one or both of the following two options can be considered:

- (i) The vector $f = (F_{11}, \dots, F_{nn})^\top$, where F_{ii} is the i th diagonal element of $F(\theta)$ given in (15), for $i = 1, \dots, n$. Clearly, F_{ii} indicates the possible impact of the perturbation of the i th observation on the RML estimates of the model parameters. We consider the i th case as outstanding if $F_{ii} \geq 2\bar{F}$, similar to $C_i \geq 2\bar{C}$, where C_i is the i th total local influence corresponding to F_{ii} , and $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$ is as given in, for example, Leiva et al (2007) and Paula & Cysneiros (2010).
- (ii) The vector l_{\max} , which is the eigenvector associated with the largest eigenvalue C_{\max} of $F(\theta)$; see Cook (1986), Leiva et al (2007), Liu et al (2009) and Paula et al (2012). Large absolute values of those elements of l_{\max} , that is, those greater than $2/\sqrt{n}$, indicate the corresponding observations to be influential.

3.2 Perturbation matrices

3.2.1 Model perturbation

Spherical disturbance We replace the normal distribution in (1) and (2) by $u_w \sim N_n(\mathbf{0}_{n \times 1}, \sigma^2 \mathbf{W}^{-1})$, where $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ is an $n \times n$ diagonal matrix with its elements w_i being the perturbations or weights and $\mathbf{W}_0 = \text{diag}(1, \dots, 1)$ is the $n \times n$ identity matrix for non-perturbed values, with $i = 1, \dots, n$. We also use $\text{vec} \mathbf{W} = \mathbf{S} \mathbf{w}$, where \mathbf{S} is the $n^2 \times n$ selection matrix, $\mathbf{w} = (w_1, \dots, w_n)^\top$ and $\text{vec} \mathbf{W}$ is the vectorization of \mathbf{W} ; see Neudecker et al (1995) and Liu et al (2014). In this perturbation scheme, the relevant part of the log-likelihood function subject to the restriction $r = \mathbf{R}\beta$ is given by

$$\ell_{w_1}(\theta) = -\frac{n}{2} \log(\sigma^2) + \frac{1}{2} \log(|\mathbf{W}|) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top \mathbf{W} (\mathbf{y} - \mathbf{X}\beta) - \lambda^\top (\mathbf{R}\beta - r). \quad (16)$$

Using the differential of $\ell_{w_1}(\theta)$ given in (16) with respect to θ detailed in Appendix B, we obtain $\Delta(\hat{\theta})$ defined in (15) as

$$\Delta(\hat{\theta}) = \begin{pmatrix} \frac{1}{\sigma^2} ((\mathbf{y} - \mathbf{X}\hat{\beta})^\top \otimes \mathbf{X}^\top) \mathbf{S} \\ \frac{1}{2\sigma^4} ((\mathbf{y} - \mathbf{X}\hat{\beta})^\top \otimes (\mathbf{y} - \mathbf{X}\hat{\beta})^\top) \mathbf{S} \end{pmatrix},$$

where \otimes is the Kronecker product.

Non-spherical disturbance In this case, the relevant part of the log-likelihood function subject to the restriction $r = \mathbf{R}\beta$ is given by

$$\ell(\theta_G) = -\sum_{g=1}^2 \frac{n_g}{2} \log(\sigma_g^2) + \frac{1}{2} \log(|\mathbf{W}|) - \sum_{g=1}^2 \frac{1}{2\sigma_g^2} (\mathbf{y}_g - \mathbf{X}_g \beta)^\top \mathbf{W}_g (\mathbf{y}_g - \mathbf{X}_g \beta) - \lambda^\top (\mathbf{R}\beta - r).$$

We use again matrix calculus to get

$$\Delta(\hat{\theta}_G) = \begin{pmatrix} \frac{1}{\sigma_1^2} ((\mathbf{y}_1 - \mathbf{X}_1 \hat{\beta}_G)^\top \otimes \mathbf{X}_1^\top) \mathbf{S}_1 & \frac{1}{\sigma_2^2} ((\mathbf{y}_2 - \mathbf{X}_2 \hat{\beta}_G)^\top \otimes \mathbf{X}_2^\top) \mathbf{S}_2 \\ \frac{1}{2\sigma_1^4} ((\mathbf{y}_1 - \mathbf{X}_1 \hat{\beta}_G)^\top \otimes (\mathbf{y}_1 - \mathbf{X}_1 \hat{\beta}_G)^\top) \mathbf{S}_1 & 0 \\ 0 & \frac{1}{2\sigma_2^4} ((\mathbf{y}_2 - \mathbf{X}_2 \hat{\beta}_G)^\top \otimes (\mathbf{y}_2 - \mathbf{X}_2 \hat{\beta}_G)^\top) \mathbf{S}_2 \end{pmatrix},$$

where S_g is the $n_g^2 \times n_g$ selection matrix, for $g = 1, 2$.

3.2.2 Response perturbation

Spherical disturbance We replace the assumption of normality by $u_w \sim N_n(\mathbf{0}_{n \times 1}, \sigma^2 \mathbf{I}_n)$, where $u_w = \mathbf{y} + \mathbf{w} - \mathbf{X}\beta$ is based on $\mathbf{y} + \mathbf{w}$ instead of \mathbf{y} given in (1), $\mathbf{w} = (w_1, \dots, w_n)^\top$ is an $n \times 1$ perturbed vector, and $\mathbf{w}_0 = (0, \dots, 0)^\top$ is an $n \times 1$ non-perturbed vector. In this perturbation scheme, the relevant part of the log-likelihood function subject to the restriction $r = \mathbf{R}\beta$ is given by

$$\ell_{w_2}(\theta) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} + \mathbf{w} - \mathbf{X}\beta)^\top (\mathbf{y} + \mathbf{w} - \mathbf{X}\beta) - \lambda^\top (\mathbf{R}\beta - r). \quad (17)$$

Taking the differential of $\ell_{w_2}(\boldsymbol{\theta})$ given in (17) with respect to $\boldsymbol{\theta}$ as detailed in Appendix B, we get

$$\Delta(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{1}{\sigma^2} \mathbf{X}^\top \\ \frac{1}{\sigma^4} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \end{pmatrix}.$$

Non-spherical disturbance In this case, the relevant part of the log-likelihood function subject to the restriction $\mathbf{r} = \mathbf{R}\boldsymbol{\beta}$ is given by

$$\ell(\boldsymbol{\theta}_G) = -\sum_{g=1}^2 \frac{n_g}{2} \log(\sigma_g^2) - \sum_{g=1}^2 \frac{1}{2\sigma_g^2} (\mathbf{y}_g + \mathbf{w}_g - \mathbf{X}_g\boldsymbol{\beta})^\top (\mathbf{y}_g + \mathbf{w}_g - \mathbf{X}_g\boldsymbol{\beta}) - \boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}).$$

By using matrix calculus, we get

$$\Delta(\hat{\boldsymbol{\theta}}_G) = \begin{pmatrix} \frac{1}{\sigma_1^2} \mathbf{X}_1^\top & \frac{1}{\sigma_2^2} \mathbf{X}_2^\top \\ \frac{1}{\sigma_1^4} (\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_G)^\top & 0 \\ 0 & \frac{1}{\sigma_2^4} (\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_G)^\top \end{pmatrix}.$$

3.2.3 Covariable perturbation

Spherical disturbance We now assume $\mathbf{u}_w \sim N_n(\mathbf{0}_{n \times 1}, \sigma^2 \mathbf{I}_n)$, where $\mathbf{u}_w = \mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\boldsymbol{\beta}$ is based on $\mathbf{X} + \mathbf{W}\mathbf{A}$ instead of \mathbf{X} given in (1), with \mathbf{W} being an $n \times p$ perturbed matrix, $\mathbf{W}_0 = \mathbf{0}$ an $n \times p$ non-perturbed matrix, $\mathbf{A} = \text{diag}(a_1, \dots, a_p)$ a $p \times p$ diagonal matrix, and a_j the standard deviation of \mathbf{x}_j corresponding to the j th column of \mathbf{X} , for $j = 1, \dots, p$. In this perturbation scheme, the relevant part of the log-likelihood function subject to the restriction $\mathbf{r} = \mathbf{R}\boldsymbol{\beta}$ is given by

$$\ell_{w_3}(\boldsymbol{\theta}) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\boldsymbol{\beta})^\top (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\boldsymbol{\beta}) - \boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}). \quad (18)$$

Based on the differential of $\ell_{w_3}(\boldsymbol{\theta})$ given in (18) with respect to $\boldsymbol{\theta}$ as detailed in Appendix B, we obtain

$$\Delta(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{1}{\sigma^2} ((\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \otimes \mathbf{A})\mathbf{K} - \frac{1}{\sigma^2} \hat{\boldsymbol{\beta}}^\top \mathbf{A} \otimes \mathbf{X}^\top \\ -\frac{1}{\sigma^4} \hat{\boldsymbol{\beta}}^\top \mathbf{A} \otimes (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \end{pmatrix},$$

where \mathbf{K} is the $np \times np$ commutation matrix such that $\text{vec}\mathbf{W}^\top = \mathbf{K}\text{vec}\mathbf{W}$ for the $n \times p$ matrix \mathbf{W} ; see Magnus & Neudecker (1999) and Liu (2002) for the definition, properties and applications of \mathbf{K} . In particular, if we perturb only \mathbf{x}_j to $\mathbf{x}_j + a_j\mathbf{w}$, where $\mathbf{w} = (w_1, \dots, w_n)^\top$ is the $n \times 1$ perturbed vector and $\mathbf{w}_0 = (0, \dots, 0)^\top$ is the $n \times 1$ non-perturbed vector, we obtain

$$\Delta(\hat{\boldsymbol{\theta}})_j = a_j \begin{pmatrix} \frac{1}{\sigma^2} \mathbf{s}_j (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top - \frac{1}{\sigma^2} \hat{\beta}_j \mathbf{X}^\top \\ -\frac{1}{\sigma^4} \hat{\beta}_j (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \end{pmatrix},$$

where \mathbf{s}_j is the $p \times 1$ vector with an one in the the j th position and zeros in the other positions, for $j = 1, \dots, p$.

Non-spherical disturbance In this case, the relevant part of the log-likelihood function subject to the restriction $\mathbf{r} = \mathbf{R}\boldsymbol{\beta}$ is given by

$$\ell(\boldsymbol{\theta}_G) = -\sum_{g=1}^2 \frac{n_g}{2} \log(\sigma_g^2) - \sum_{g=1}^2 \frac{1}{2\sigma_g^2} (\mathbf{y}_g - (\mathbf{X}_g + \mathbf{W}_g\mathbf{A}_g)\boldsymbol{\beta})^\top (\mathbf{y}_g - (\mathbf{X}_g + \mathbf{W}_g\mathbf{A}_g)\boldsymbol{\beta}) - \boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}),$$

where \mathbf{W}_g is the $n_g \times p$ perturbed matrix, $\mathbf{A}_g = \text{diag}(a_{g1}, \dots, a_{gp})$ is the $p \times p$ diagonal matrix, and a_{gj} is the standard deviation of the j th column of \mathbf{X}_g , for $g = 1, 2$ and $j = 1, \dots, p$. We use matrix calculus to obtain

$$\Delta(\hat{\boldsymbol{\theta}}_G) = \begin{pmatrix} \frac{1}{\sigma_1^2} ((\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_G)^\top \otimes \mathbf{A}_1)\mathbf{K}_1 - \frac{1}{\sigma_1^2} \hat{\boldsymbol{\beta}}_G^\top \mathbf{A}_1 \otimes \mathbf{X}_1^\top & \frac{1}{\sigma_2^2} ((\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_G)^\top \otimes \mathbf{A}_2)\mathbf{K}_2 - \frac{1}{\sigma_2^2} \hat{\boldsymbol{\beta}}_G^\top \mathbf{A}_2 \otimes \mathbf{X}_2^\top \\ -\frac{1}{\sigma_1^4} \hat{\boldsymbol{\beta}}_G^\top \mathbf{A}_1 \otimes (\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_G)^\top & 0 \\ 0 & -\frac{1}{\sigma_2^4} \hat{\boldsymbol{\beta}}_G^\top \mathbf{A}_2 \otimes (\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_G)^\top \end{pmatrix},$$

where \mathbf{K}_g is the $n_g p \times n_g p$ commutation matrix such that $\text{vec} \mathbf{W}_g^\top = \mathbf{K}_g \text{vec} \mathbf{W}_g$ for the $n_g \times p$ matrix \mathbf{W} . In particular, if we perturb only \mathbf{x}_j to $\mathbf{x}_j + a_j \mathbf{w}$, we obtain

$$\Delta(\hat{\boldsymbol{\theta}}_G)_j = a_j \begin{pmatrix} \frac{1}{\sigma_1^2} \mathbf{s}_j (\mathbf{y}_1 - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_G)^\top - \frac{1}{\sigma_1^2} \hat{\boldsymbol{\beta}}_{Gj} \mathbf{X}_1^\top & \frac{1}{\sigma_2^2} \mathbf{s}_j (\mathbf{y}_2 - \mathbf{X}_2 \hat{\boldsymbol{\beta}}_G)^\top - \frac{1}{\sigma_2^2} \hat{\boldsymbol{\beta}}_{Gj} \mathbf{X}_2^\top \\ -\frac{1}{\sigma_1^2} \hat{\boldsymbol{\beta}}_{Gj} (\mathbf{y}_1 - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_G)^\top & 0 \\ 0 & -\frac{1}{\sigma_2^2} \hat{\boldsymbol{\beta}}_{Gj} (\mathbf{y}_2 - \mathbf{X}_2 \hat{\boldsymbol{\beta}}_G)^\top \end{pmatrix},$$

where \mathbf{s}_j is the $p \times 1$ vector with an one in the the j th position and zeros in the other positions, for $j = 1, \dots, p$.

3.3 Global influence

In addition to our local influence statistics F_{ii} and l_{\max} given in Subsection 3.1, using the case deletion method, we introduce three global influence statistics: LD, relative change (RC) and generalised Cook distance (GCD). These are defined by

- (a) $\text{LD}_i = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{(i)})]$,
- (b) $\text{RC}_i = \|\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}\| / \|\hat{\boldsymbol{\theta}}\|$,
- (c) $\text{GCD}_i = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top \mathbf{J}(\hat{\boldsymbol{\theta}})^{-1} (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}) / k$,

for measuring differences between fits with and without the i th observation, for $i = 1, \dots, n$. In the spherical disturbance case, to calculate these global influence measures, ℓ is given in (5), $\hat{\boldsymbol{\theta}}$ is obtained using all the observations, $\hat{\boldsymbol{\theta}}_{(i)}$ is computed using the data without the i th observation, $\mathbf{J}(\hat{\boldsymbol{\theta}})$ is given in (13) and $k = p + 1$ is the dimension of $\boldsymbol{\theta}$. In the non-spherical disturbance case, ℓ is given in (10), $\hat{\boldsymbol{\theta}}_G$ is obtained using all the observations, $\hat{\boldsymbol{\theta}}_{Gi}$ is calculated using the data without the i th observation, $\mathbf{J}(\hat{\boldsymbol{\theta}}_G)$ is in (14) and $k = p + 2$ is the dimension of $\boldsymbol{\theta}_G$.

4 Numerical illustration

In this section, we illustrate and validate the diagnostic methodology proposed in this work with three empirical examples of real-world data. The first two examples are for the spherical disturbance case and the third one is for the non-spherical disturbance case.

4.1 Example 1

We analyze the data for a response with seven covariables observed in 40 metropolitan areas used in a study by Ramanathan (1993, Table 10.1) and Cysneiros & Paula (2005). The objective of this study is to regress the number (in thousands) of subscribers with cable TV (Y) against the number (in thousands) of homes in the area (X_1), the per capita income for each TV market with cable (X_2), the installation fee (X_3), the monthly service charge (X_4), the number of TV signals carried by each cable system (X_5) and the number of TV signals received with good quality without cable (X_6). As Y is a count, we take its square root to try to stabilize the variance. Thus, the model is

$$\sqrt{Y}_i = \beta_0 + \sum_{j=1}^6 x_{ij} \beta_j + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ are mutually independent errors, for $i = 1, \dots, 40$. It is reasonable to expect the effect of each coefficient to be unidirectional, such as in Cysneiros & Paula (2005), so that the opposite direction is theoretically impossible. We may focus on assessing whether the number of subscribers changes as the monthly service charge changes, that is, to assess whether $\beta_4 = 0$ or not, which can be treated as an exact linear restriction. Of course, in the same way for the remaining covariables, we may be interested in assessing other equality restrictions. Therefore, we use $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$, where

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and $\beta = (\beta_0, \beta_1, \dots, \beta_6)^\top$, such as in Cysneiros & Paula (2005). This corresponds to the exact linear restrictions $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. For these data, Cysneiros & Paula (2005) found case #14 to be most influential using the standardised residuals, and both cases #1 and #14 to be the most influential using the total local influence based on a quadratic penalty function. Employing our formulas provided in Section 3, we obtained the RML estimates and the plots of the diagonal elements of \mathbf{F} and \mathbf{l}_{\max} . The plot of standardised residuals in Figure 1 may indicate two extremal observations. However, the potentially influential observations in Figure 2 include cases #21 and #16 identified by $|\mathbf{F}_{ii}|$, plus those found by Cysneiros & Paula (2005) and our \mathbf{l}_{\max} . In addition, the three global influence statistics in Figure 3 indicate that only cases #14 and #1 are influential.

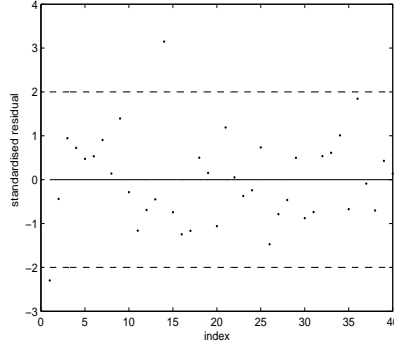


Fig. 1 Plot of standardised residual versus index for data of Example 1.

4.2 Example 2

We use the data considered in Paula & Cysneiros (2010, Application 1). The response (Y) is the weight to height ratio (scaled by a factor of 100) versus age (X) for 72 children from birth to 71.5 months. A restricted normal model is proposed with response required to be cubic for $x \leq t_0$ and linear for $x > t_0$ ($t_0 = 16$ months), that is,

$$Y_i = \beta_0 + \sum_{j=1}^3 x_i^j \beta_j + \beta_4 (x_i - t_0)_+^3 + \varepsilon_i, \quad i = 1, \dots, 72,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ are mutually independent errors, and $(x_i - t_0)_+ = 0$ if $x \leq t_0$, or $(x_i - t_0)_+ = x_i - t_0$ if $x \geq t_0$ with the restrictions $\mathbf{R}\beta = \mathbf{0}$ to guarantee a linear tendency after t_0 . Thus,

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & 1 & 0 & -3t_0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

and $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^\top$.

Using a quadratic penalty function under a model perturbation scheme, Paula & Cysneiros (2010, Application 1) applied the total local influence method to β and then to σ^2 . They found two young children identified as cases #2 and #8 have a large total local influence on β , and cases #8, #21 and #25 have a large influence on σ^2 . Using our formulas provided in Section 3, we obtain the RML estimates and the plots of the diagonal elements of \mathbf{F} and \mathbf{l}_{\max} . The plot of standardised residuals in Figure 4 indicates four extremal observations. However, the potentially influential cases in Figure 5 include #1, #72 and #9, and even #2 and #8, identified by \mathbf{F}_{ii} and \mathbf{l}_{\max} , in addition to those found by Paula & Cysneiros (2010, Application 1), with case #1 or #72 being the most influential ones in one or two perturbation schemes. In Figure 6, the LD statistics may indicate that cases #8, #2, #1 and #21 are influential. The other two global influence statistics indicate that only cases #2 and #1 are influential.

4.3 Example 3

The objective of this example is to apply our results with restrictions for the non-spherical case. We use the data analysed in Examples 9.8 and 9.9 in Wooldridge (2013) to study if the R&D intensity increases with firm size, while an

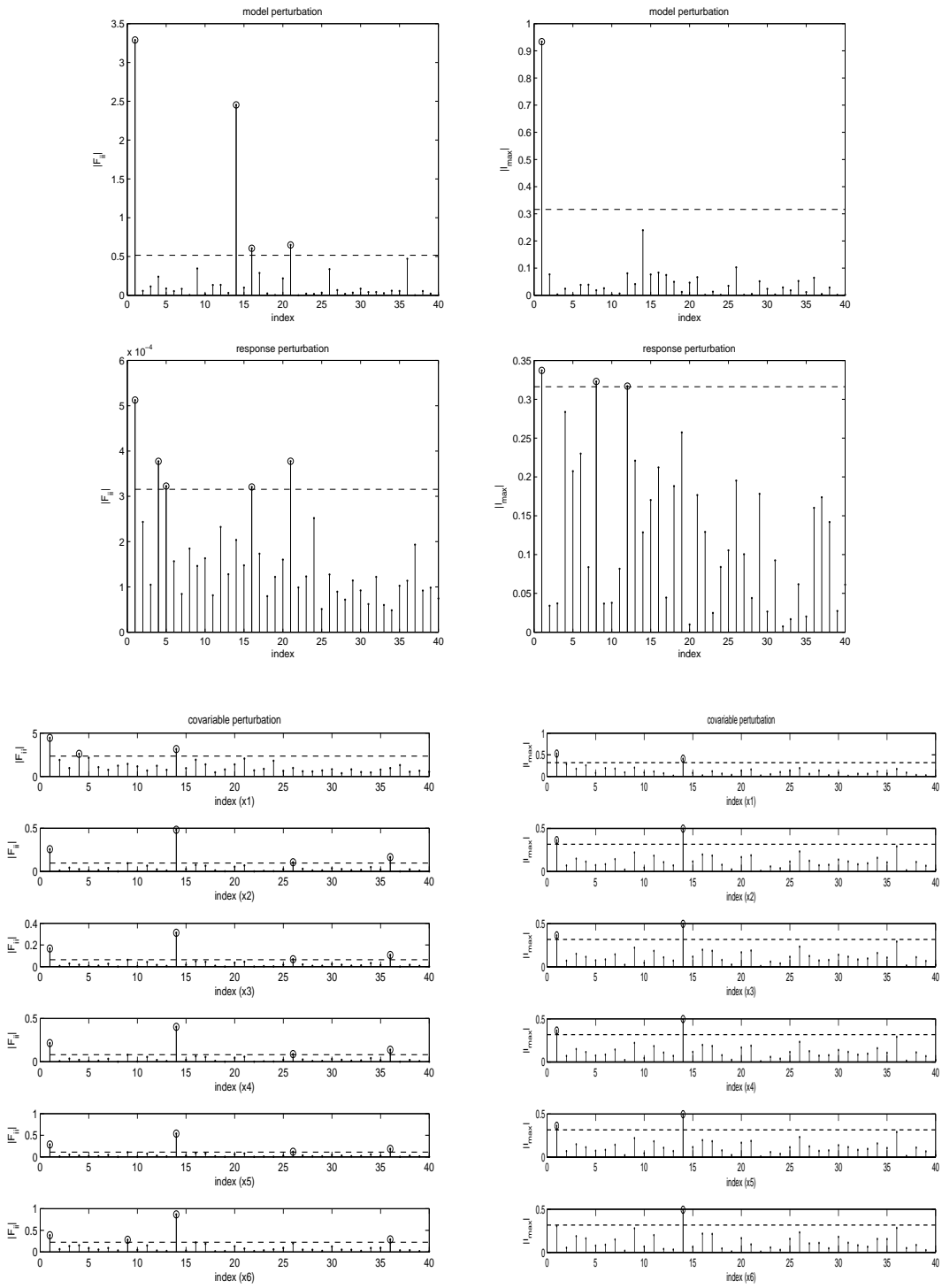


Fig. 2 Plots of diagonal element of F versus index (left) and of element of l_{\max} versus index (right) for data of Example 1.

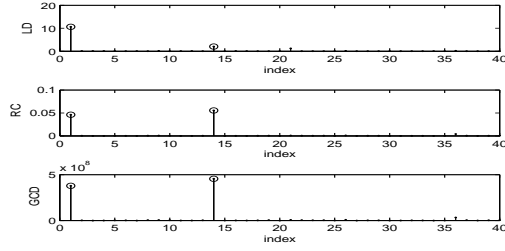


Fig. 3 Plots of global influence statistics versus index for data of Example 1.

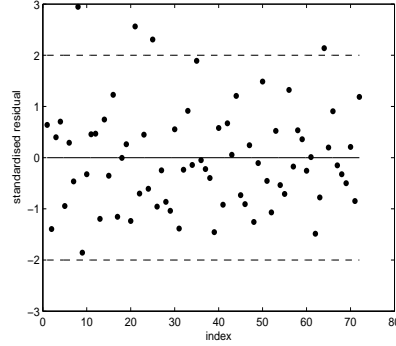


Fig. 4 Plot of standardised residual versus index for data of Example 2.

equality restriction on the coefficient may be involved. Following Wooldridge (2013), we suppose that R&D expenditures as a percentage of sales are related to sales and profits as a percentage of sales by

$$Y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i, \quad i = 1, \dots, 32,$$

where the responses Y_i are for the R&D intensity, the covariables x_{i1} and x_{i2} are the firm sales (in millions) and the profit margins, respectively. For a simple illustration, we impose $\mathbf{R} = (0, 1, 0)$ and $\mathbf{r} = 0.00005$, which agrees with the data analysis in Wooldridge (2013). After ordering the data by the sizes of x_{i1} , that is, the sales, and splitting the data into two groups of 16 observations, we assume $\varepsilon_i \sim N(0, \sigma_1^2)$, for $i = 1, \dots, 16$, and $\varepsilon_i \sim N(0, \sigma_2^2)$, for $i = 17, \dots, 32$, are mutually independent errors. We fit the model and conduct the Goldfeld-Quandt test for heteroscedasticity with p-value = 0.012 < 0.05 = α , which supports the appropriateness of a non-spherical disturbance model with $\sigma_1^2 \neq \sigma_2^2$.

We present a plot of the standardised residuals in Figure 7. Note that this residual for case #1 is greater than 2. In Figure 8, the local influence statistics find case #1 to be most influential and cases #22, #30, #32 and #10 are possibly influential. In Figure 9, the global influence statistics suggest cases #1 and #30 are influential, with cases #32 and #10 to be noted. Using a dummy variable approach, Wooldridge (2013) found cases #10 and #1 to be individually influential. They are the largest firm (case #10) and the firm with the highest value of R&D intensity (case #1). We find that case #1 is more influential than case #10, and identify additional observations including cases #22 and #30 as potentially influential using our local influence method with F_{ii} and even l_{\max} . These cases are not identified by the global influence or dummy variable approaches.

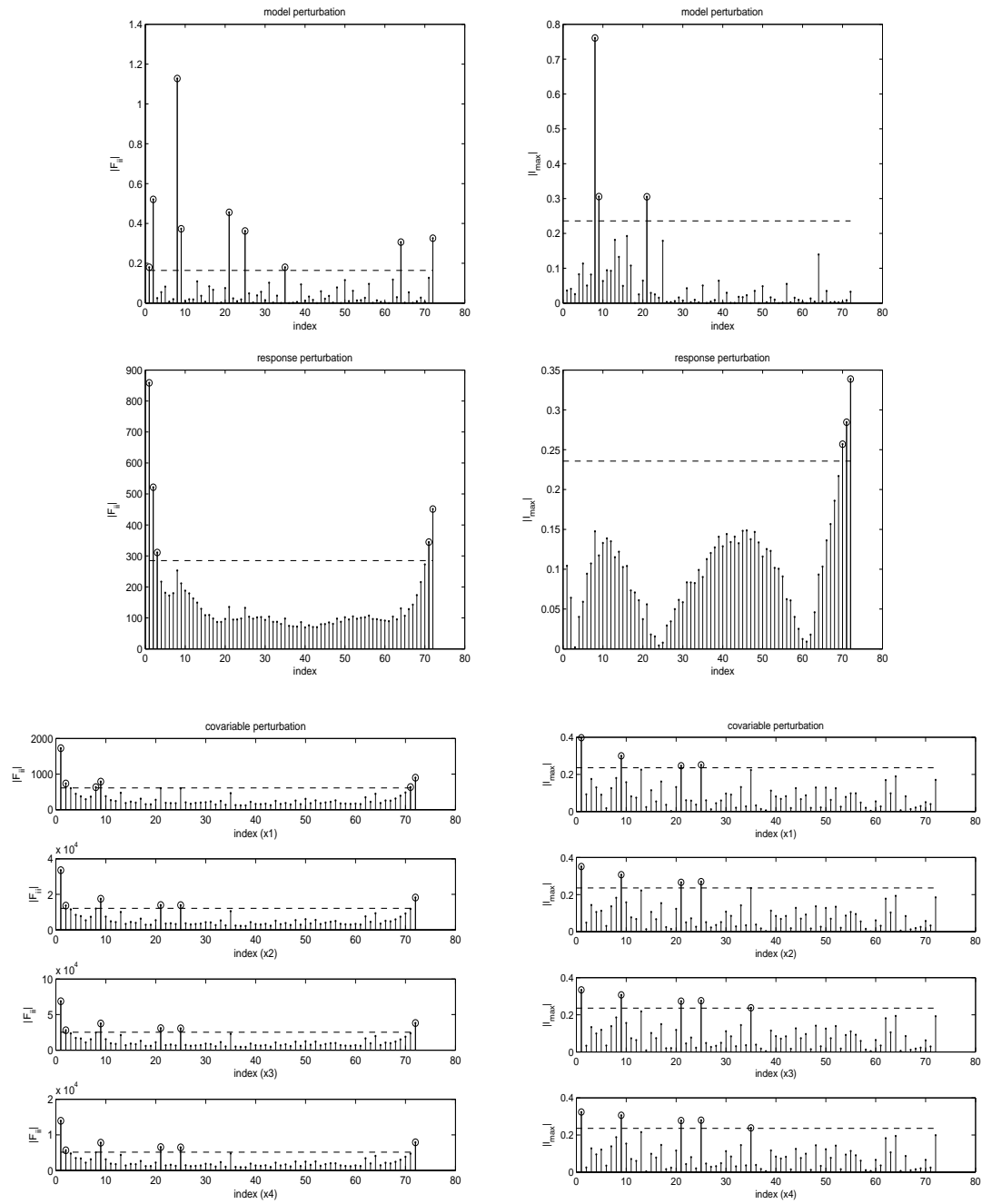


Fig. 5 Plots of diagonal element of F versus index (left) and of element of l_{\max} versus index (right) for data of Example 2.

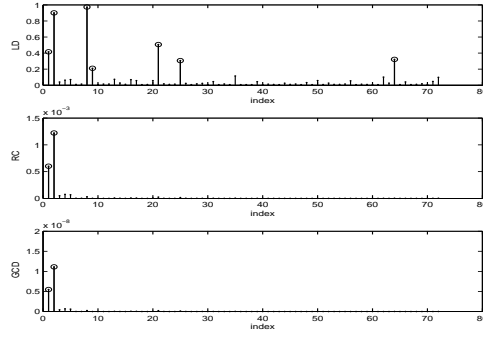


Fig. 6 Plots of global influence statistics versus index for data of Example 2.

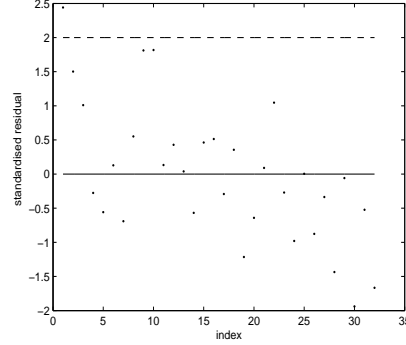


Fig. 7 Plot of standardised residual versus index for data of Example 3.

5 Concluding remarks

We have established results for influence diagnostics in the possibly heteroskedastic regression linear model with exact linear restrictions. We have used the restricted maximum likelihood estimators with Lagrange multipliers for the linear penalty function to find the diagnostic matrix. On one hand, the empirical examples have indicated that our results can be used to make findings similar to those on the same datasets provided in Cysneiros & Paula (2005) and Paula & Cysneiros (2010). On the other hand, we have seen that our results are different from those given in Cysneiros & Paula (2005) and Paula & Cysneiros (2010) in a number of ways. Our methodology may be used in the non-spherical disturbance case and further extended models. For diagnostics, we have used not just the total local influence statistics which are based on the diagonal elements of the diagnostic matrix, but also the direction eigenvector associated with its largest eigenvalue. We have compared the local influence statistics with three global influence statistics by examining the possible influential observations identified by these sets of influence statistics. Our local influence statistics have identified more possible influential observations than the local influence statistics in Cysneiros & Paula (2005) and Paula & Cysneiros (2010) and our global influence statistics. These results are directly related to the restricted least squares estimators, which have been widely used in econometrics and statistics. They also complement those results for the linear model established by Liu & Neudecker (2007) using a sensitivity analysis approach. Our results can be used and implemented in a reasonably easy way by computer packages. Our Matlab codes are available on request.

Appendix A: differentials for the Hessian matrix

We use matrix calculus as studied in Magnus & Neudecker (1999) to establish our results in both disturbance cases. In the spherical disturbance case, we present the differentials for the Hessian matrix in Appendix A and for the Δ matrices in Appendix B. In the non-spherical disturbance case, we obtain the differentials and matrices in a similar manner, so that they are omitted here.

First, we take the differential of ℓ given in (5) with respect to β and σ^2 and obtain

$$d_{\beta}\ell = \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^{\top} \mathbf{X}d\beta - \lambda^{\top} \mathbf{R}d\beta, \quad (19)$$

$$d_{\sigma^2}\ell = -\frac{n}{2\sigma^2}d\sigma^2 + \frac{(\mathbf{y} - \mathbf{X}\beta)^{\top}(\mathbf{y} - \mathbf{X}\beta)}{2\sigma^4}d\sigma^2. \quad (20)$$

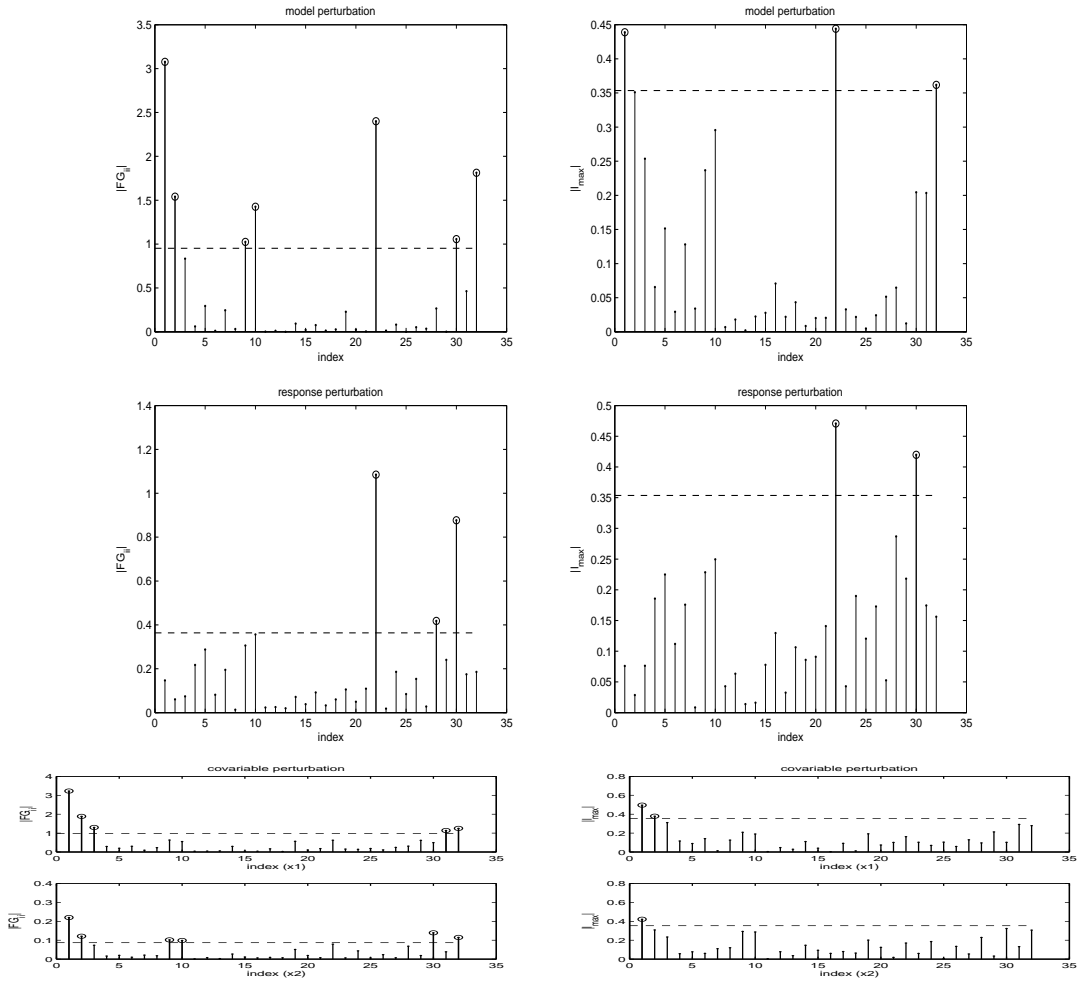


Fig. 8 Plots of diagonal element of F and l_{\max} versus index for data of Example 3.

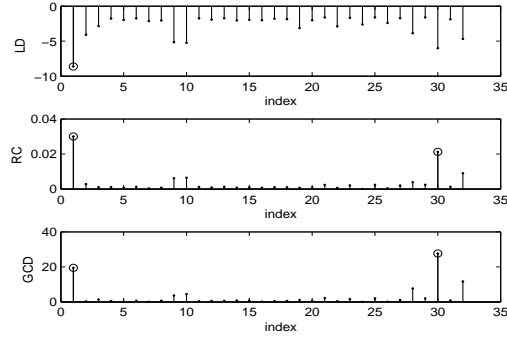


Fig. 9 Plots of global influence statistics versus index for data of Example 3.

Then, we take the differentials of the elements of the score vector given in (19) and (20) with respect to β and σ^2 as

$$d_{\beta}^2 \ell = -\frac{1}{\sigma^2} d\beta X^{\top} X d\beta, \quad (21)$$

$$d_{\sigma^2}^2 \ell = \frac{n}{2\sigma^4} d\sigma^2 d\sigma^2 - \frac{1}{\sigma^6} d\sigma^2 (\mathbf{y} - \mathbf{X}\beta)^{\top} (\mathbf{y} - \mathbf{X}\beta) d\sigma^2, \quad (22)$$

$$d_{\beta\sigma^2}^2 \ell = -\frac{1}{\sigma^4} d\beta^{\top} X^{\top} (\mathbf{y} - \mathbf{X}\beta) d\sigma^2. \quad (23)$$

We establish the Hessian matrix $\mathbf{H}(\theta)$ from the differentials given in (21), (22) and (23).

Appendix B: differentials for the perturbation schemes

We present the differentials for the perturbation schemes in the spherical disturbance case defined in Subsection 3.2 considering the log-likelihood functions ℓ_{w_1} , ℓ_{w_2} and ℓ_{w_3} established in (16), (17) and (18), respectively. From the differentials, we get the Δ matrices.

Model perturbation Taking the differential of ℓ_{w_1} with respect to β and σ^2 , we obtain

$$\begin{aligned} d_{\beta} \ell_{w_1} &= \frac{1}{\sigma^2} d\beta^{\top} \mathbf{X}^{\top} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta) - d\beta^{\top} \mathbf{R}^{\top} \boldsymbol{\lambda}, \\ d_{\sigma^2} \ell_{w_1} &= -\frac{n}{2\sigma^2} d\sigma^2 + \frac{(\mathbf{y} - \mathbf{X}\beta)^{\top} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)}{2\sigma^4} d\sigma^2. \end{aligned}$$

Taking the differential of $d\ell_{w_1}$ with respect to w , we obtain

$$\begin{aligned} d_{\beta w}^2 \ell_{w_1} &= \frac{1}{\sigma^2} d\beta^{\top} \mathbf{X}^{\top} d\mathbf{W} (\mathbf{y} - \mathbf{X}\beta) \\ &= \frac{1}{\sigma^2} d\beta^{\top} ((\mathbf{y} - \mathbf{X}\beta)^{\top} \otimes \mathbf{X}^{\top}) S dw, \\ d_{\sigma^2 w}^2 \ell_{w_1} &= \frac{1}{2\sigma^4} d\sigma^2 (\mathbf{y} - \mathbf{X}\beta)^{\top} d\mathbf{W} (\mathbf{y} - \mathbf{X}\beta) \\ &= \frac{1}{2\sigma^4} d\sigma^2 ((\mathbf{y} - \mathbf{X}\beta)^{\top} \otimes (\mathbf{y} - \mathbf{X}\beta)^{\top}) S dw. \end{aligned}$$

Response perturbation Taking the differential of ℓ_{w_2} with respect to β and σ^2 , we obtain

$$\begin{aligned} d_{\beta} \ell_{w_2} &= \frac{1}{\sigma^2} d\beta^{\top} \mathbf{X}^{\top} (\mathbf{y} + w - \mathbf{X}\beta) - d\beta^{\top} \mathbf{R}^{\top} \boldsymbol{\lambda}, \\ d_{\sigma^2} \ell_{w_2} &= -\frac{n}{2\sigma^2} d\sigma^2 + \frac{(\mathbf{y} + w - \mathbf{X}\beta)^{\top} (\mathbf{y} + w - \mathbf{X}\beta)}{2\sigma^4} d\sigma^2. \end{aligned}$$

Taking the differential of $d\ell_{w_2}$ with respect to w we obtain

$$\begin{aligned} d_{\beta w}^2 \ell_{w_2} &= \frac{1}{\sigma^2} d\beta^{\top} \mathbf{X}^{\top} dw, \\ d_{\sigma^2 w}^2 \ell_{w_2} &= \frac{1}{\sigma^4} d\sigma^2 (\mathbf{y} + w - \mathbf{X}\beta)^{\top} dw. \end{aligned}$$

Covariable perturbation Taking the differential of ℓ_{w_3} with respect to β and σ^2 , we obtain

$$\begin{aligned} d_{\beta} \ell_{w_3} &= \frac{1}{\sigma^2} d\beta^{\top} (\mathbf{X} + \mathbf{W}\mathbf{A})^{\top} (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\beta) - d\beta^{\top} \mathbf{R}^{\top} \boldsymbol{\lambda}, \\ d_{\sigma^2} \ell_{w_3} &= -\frac{n}{2\sigma^2} d\sigma^2 + \frac{1}{2\sigma^4} d\sigma^2 (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\beta)^{\top} (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\beta). \end{aligned}$$

Taking the differential of $d\ell_{w_3}$ with respect to w , we obtain

$$\begin{aligned} d_{\beta w}^2 \ell_{w_3} &= \frac{1}{\sigma^2} d\beta^{\top} \mathbf{A} d\mathbf{W}^{\top} (\mathbf{y} - (\mathbf{X} + \mathbf{W}\mathbf{A})\beta) - \frac{1}{\sigma^2} d\beta^{\top} (\mathbf{X} + \mathbf{W}\mathbf{A})^{\top} d\mathbf{W} \mathbf{A} \beta, \\ d_{\sigma^2 w}^2 \ell_{w_3} &= -\frac{1}{\sigma^4} d\sigma^2 (\mathbf{y} - \mathbf{X}\beta - \mathbf{W}\mathbf{A}\beta)^{\top} d\mathbf{W} \mathbf{A} \beta. \end{aligned}$$

Acknowledgments

We would like to thank Professor Andrea Cerioli, Editor-in-Chief, and the referees for their constructive comments which have led to an improved version of the manuscript. We would also like to thank Professor Francisco J.A. Cysneiros at Universidade Federal de Pernambuco (Brazil) for providing us with the data and for his valuable assistance. The research for the present paper has been initiated and conducted since the authors met in pairs at the Australian National University (Australia), Universidad de Valparaiso (Chile) and Southwestern University of Finance and Economics (China). The authors are very grateful to the institutions and the colleagues for their kind support and hospitality. An early version of the paper was presented at the ASC-IMS Conference in Sydney in July 2014. The research of Dr. Victor Leiva has been supported by FONDECYT 1120879 (Chile).

References

- Atkinson AC, Riani M (2000) *Robust Diagnostic Regression Analysis*. Springer:Berlin
- Atkinson AC, Riani M, Cerioli A (2004) *Exploring Multivariate Data with the Forward Search*. Springer:Berlin
- Barros M, Galea M, González M, Leiva V (2010) Influence diagnostics in the tobit censored response model. *Stat Method Appl* 19:379-397
- Billor N, Loynes RM (1993) Local influence: a new approach. *Comm Stat Theor M* 22:1595-1611
- Billor N, Loynes RM (1999) An application of the local influence approach to ridge regression. *J Appl Stat* 2:177-183
- Chatterjee S, Hadi AS (1988) *Sensitivity Analysis in Linear Regression*. Wiley:New York
- Cook D (1986) Assessment of local influence. *J R Stat Soc B* 48:133-169
- Chipman JS, Rao MM (1964) The treatment of linear restrictions in regression analysis. *Econometrica* 32:198-209
- Cook D, Weisberg S (1982) *Residuals and Influence in Regression*. Chapman & Hall:New York
- Cysneiros FJA, Paula GA (2005) Restricted methods in symmetrical linear regression models. *Comput Stat Data Anal* 49:689-708
- de Castro M, Galea M, Bolfarine H (2007) Local influence assessment in heteroscedastic measurement error models. *Comput Stat Data Anal* 52:1132-1142
- Diaz-García JA, Galea M, Leiva V (2003) Influence diagnostics for multivariate elliptical regression linear models. *Comm Stat Theor M* 32:625-641
- Efron B, Hinkley D (1978) Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* 65:457-487
- Galea M, de Castro M (2012) Influence assessment in an heteroscedastic errors-in-variables model. *Comm Stat Theor M* 41:1350-1363
- Galea M, Diaz-Garcia JA, Vilca F (2008) Influence diagnostics in the capital asset pricing model under elliptical distributions. *J Appl Stat* 35:179-192
- Galea M, Paula GA, Bolfarine H (1997) Local influence for elliptical linear models. *J R Stat Soc D* 46:71-79
- Greene WH (2007) *Econometric Analysis*. Prentice Hall:New York
- Gross J (2003) *Linear Regression*. Springer:Berlin
- Hocking R (2003) *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. Wiley:New York
- Judge G, Hill RC, Griffiths WE, Lutkepohl, Lee TC (1988) *Introduction to the Theory and Practice of Econometrics*. Wiley:New York
- Kleiber C, Zeileis A (2008) *Applied Econometrics with R*. Springer:Berlin
- Leiva V, Barros M, Paula GA, Galea M (2007) Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. *Comput Stat Data Anal* 51:5694-5707
- Leiva V, Rojas E, Galea M, Sanhueza A (2014) Diagnostics in Birnbaum-Saunders accelerated life models with an application to fatigue data. *Appl Stoch Mod Bus Ind* 30:115-131
- Liu S (2000) On local influence for elliptical linear models. *Stat Pap* 41:211-224
- Liu S (2002) Local influence in multivariate elliptical linear regression models. *Lin Alg Appl* 354:211-224
- Liu S (2004) On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. *J Appl Prob* 41A:393-405
- Liu S, Ahmed SE, Ma LY (2009) Influence diagnostics in the linear regression model with linear stochastic restrictions. *Pak J Stat* 25:647-662
- Liu S, Ma T, Polasek W (2014) Spatial system estimators for panel models: a sensitivity and simulation study. *Math Comp Simul* 101:78-102
- Liu S, Neudecker H (2007) Local sensitivity of the restricted least squares estimator in the linear model. *Stat Pap* 48:525-525
- Magnus JR, Neudecker H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley:Chichester
- Neudecker, H., Liu, S., Polasek, W. (1995) The Hadamard product and some of its applications in statistics. *Statistics* 26:365-373
- Paula GA (1993) Assessing local influence in restricted regression models. *Comput Stat Data Anal* 16:63-79
- Paula GA, Cysneiros FJA (2010) Local influence under parameter constraints. *Comm Stat Theor M* 39:1212-1228
- Paula GA, Leiva V, Barros M, Liu S (2012) Robust statistical modeling using the Birnbaum-Saunders- t distribution applied to insurance. *Appl Stoch Mod Bus Ind* 28:16-34
- Poon WY, Poon YS (1999) Conformal normal curvature and assessment of local influence. *J R Stat Soc B* 61:51-61
- Rao CR, Toutenburg H, Shalabh, Heumann C (2008) *Linear Models and Generalizations*. Springer:Berlin
- Ramanathan R (1993) *Statistical Methods in Econometrics*. Wiley:New York
- Shi L, Chen G (2008) Local influence in multilevel models. *Can J Stat* 36:259-275
- Shi L, Huang M (2011) Stepwise local influence analysis. *Comput Stat Data Anal* 55:973-982
- Trenkler G (1987) Mean square error matrix comparisons among restricted least squares estimators. *Sankhyā A* 49:96-104
- Wooldridge JM (2013) *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning:Mason-OH