The Statistician (1999)
48, *Part* 4, *pp.* 529–538

Leverage in inequality-constrained regression models

Gilberto A. Paula

Universidade de São Paulo, Brazil

[Received March 1998. Revised June 1999]

Summary. We derive leverage measures in inequality-constrained linear regression models. When the restricted and unrestricted least squares estimates agree, the usual leverage measure h_{ii} from the unrestricted linear regression is also extended to the restricted case. However, under violation, h_{ii} is decomposed into two new leverage measures which take account of the influence of the observations on their restricted fitted values and on the change in the difference between the restricted and unrestricted fitted values. The last measure may be helpful to assess observations which are discordant with the prior belief for the parameters. We discuss extensions to generalized linear models, and two illustrative examples are given.

Keywords: Generalized linear models; Hat matrix; Jacobian leverage matrix; Linear regression; Local influence; Restricted estimation

1. Introduction

Firstly, consider the traditional linear regression model $Y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \sigma\epsilon_i$, $i = 1, \ldots, n$, where $Y_1, \ldots, Y_n$ are independent random variables, ϵ_i has a standard normal distribution, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a $p \times 1$ vector of regression coefficients, $\sigma > 0$ and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ is a $p \times 1$ vector of regression variable values. Further, suppose that we have the prior belief $\mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{0}$, where $\mathbf{C}$ is a $k \times p$ known full row rank matrix. Let $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ denote the inequality-constrained least squares (ICLS) and the ordinary least squares (OLS) estimators respectively.

We may express the ICLS estimator (see, for instance, Paula (1993)) as

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{\mathrm{R}}, \tag{1}$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

is the OLS estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\Lambda}}_{\mathrm{R}} = -\{\mathbf{C}_{\mathrm{R}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}}\hat{\boldsymbol{\beta}}$$

is the estimated Lagrange multiplier vector, $\mathbf{X}$ is an $n \times p$ matrix with rows $\mathbf{x}_i^{\mathrm{T}}$, $i = 1, \ldots, n$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $\mathbf{C}_{\mathrm{R}}$ is a $q \times p$ matrix formed by q rows of $\mathbf{C}$, $q \leqslant k$. From Theil and Van de Panne (1960) the ICLS estimator (1) may be defined as the OLS estimator for the linear regression problem subject to the linear equality constraints $\mathbf{C}_{\mathrm{R}}\boldsymbol{\beta} = \mathbf{0}$. They proved that, if m rows of $\mathbf{C}\hat{\boldsymbol{\beta}}$ violate the parameter constraints, at least one of these corresponding violated inequality

Address for correspondence: Gilberto A. Paula, Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, 05315-970 São Paulo, SP, Brazil.
E-mail: giapaula@ime.usp.br

 0039–0526/99/48529

constraints is satisfied as an equality by the ICLS estimate. For example, if we have the violations $\mathbf{C}_{j_1}^{\mathrm{T}}\hat{\boldsymbol{\beta}} < 0$ and $\mathbf{C}_{j_2}^{\mathrm{T}}\hat{\boldsymbol{\beta}} < 0$, where $\mathbf{C}_{j_1}^{\mathrm{T}}$ and $\mathbf{C}_{j_2}^{\mathrm{T}}$ denote two rows of $\mathbf{C}$, we should have either $\mathbf{C}_{j_1}^{\mathrm{T}}\boldsymbol{\beta} = 0$, or $\mathbf{C}_{j_2}\boldsymbol{\beta} = 0$ or $\mathbf{C}_{j_1}^{\mathrm{T}}\boldsymbol{\beta} = 0$ and $\mathbf{C}_{j_2}^{\mathrm{T}}\boldsymbol{\beta} = 0$. Therefore, the matrix $\mathbf{C}_{\mathrm{R}}$ will be given by either $\mathbf{C}_{\mathrm{R}} = \mathbf{C}_{j_1}^{\mathrm{T}}$, or $\mathbf{C}_{\mathrm{R}} = \mathbf{C}_{j_2}^{\mathrm{T}}$ or $\mathbf{C}_{\mathrm{R}} = (\mathbf{C}_{j_1}, \mathbf{C}_{j_2})^{\mathrm{T}}$. We should take as $\mathbf{C}_{\mathrm{R}}$ the submatrix with rows of $\mathbf{C}$ corresponding to violated inequality conditions that leads to the best linear equality least squares solution. When there is no violation with the unrestricted estimate we have $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$.

In Section 2 two leverage matrices are obtained, under violation, for the linear inequality regression model. Then, some leverage measures are derived and a sensitivity study under small changes in the response values is performed for a particular example. Extensions to generalized linear models are discussed in Section 3 and the last section gives some concluding remarks.

2. Leverage

Suppose that the observed vector $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ is perturbed by adding b to the lth element such that $\mathbf{y}_{l,b} = \mathbf{y} + b\mathbf{f}_l$, where $\mathbf{f}_l$ is the lth standard basis vector in $\mathbb{R}^n$. The ICLS and OLS estimators for the perturbed data are denoted respectively by $\tilde{\boldsymbol{\beta}}_l(b)$ and $\hat{\boldsymbol{\beta}}_l(b)$. Similarly, the predicted response vectors for the perturbed data become $\tilde{\mathbf{y}}_{l,b} = \mathbf{X}\tilde{\boldsymbol{\beta}}_l(b)$ and $\hat{\mathbf{y}}_{l,b} = \mathbf{X}\hat{\boldsymbol{\beta}}_l(b)$. Emerson *et al.* (1984) proposed a vector of *generalized leverages* due to perturbation of the lth observation in unrestricted regression. St Laurent and Cook (1992) extended the methodology to unrestricted non-linear regression. The derivation of a *Jacobian leverage matrix* and the use of its principal diagonal elements as leverage measures have proved to be more informative than the usual leverages in non-linear regression.

In inequality-constrained regression, however, the definition of leverage may be thought of in a different way rather than that used in unrestricted regression. When there is no violation, $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$, the usual leverage measures h_{ii}, $i = 1, \ldots, n$, obtained from the principal diagonal of the hat matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}},$$

may be used to assess the influence of the observations on their own fitted values. However, under violation, h_{ii} does not correspond to a leverage measure for the restricted predicted values, which suggests the calculation of a new leverage matrix. Then, following Emerson *et al.* (1984), we propose the vector of leverages

$$\mathbf{F}(b;\ l) = \frac{1}{b}(\tilde{\mathbf{y}}_{l,\mathrm{b}} - \tilde{\mathbf{y}}),$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}_{l,b}$ are the predicted values for the unperturbed and perturbed data respectively. Note that $\mathbf{C}_{\mathrm{R}}$ is a matrix that depends on the violations and therefore on the values of $\mathbf{Y}$. Then, if $\mathbf{C}_{\mathrm{R}}$ does not change under small values for b we obtain (see the proof in Appendix A) the vector of Jacobian leverages

$$\mathbf{J}(l) = \lim_{b\to 0}\{\mathbf{F}(b;\ l)\} = \mathbf{M}\mathbf{f}_l, \tag{2}$$

where $\mathbf{M} = \mathbf{H} - \mathbf{G}$ is the new leverage matrix,

$$\mathbf{G} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$$

is the projection matrix onto $C(\mathbf{Z})$, the subspace spanned by the columns of the $n \times q$ matrix

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}.$$

The matrix $\mathbf{G}$ projects $\mathbf{y}$ into $C(\mathbf{Z})$, by setting $\mathbf{G}\mathbf{y} = \hat{\mathbf{y}} - \tilde{\mathbf{y}}$. Thus, $C(\mathbf{Z})$ may be interpreted as the subspace of the differences between the solutions (unrestricted and restricted) for **Y**, and **G** may be obtained in an alternative way (see Appendix A) by considering the vector of leverages

$$\mathbf{F}^*(b;\ l) = \frac{1}{b}\{\hat{\mathbf{y}}_{l,b} - \tilde{\mathbf{y}}_{l,b} - (\hat{\mathbf{y}} - \tilde{\mathbf{y}})\},$$

for which we obtain

$$\mathbf{J}^*(l) = \lim_{b\to 0}\{\mathbf{F}^*(b;\ l)\} = \mathbf{G}\mathbf{f}_l. \tag{3}$$

Therefore, **G** is interpreted as the leverage matrix which takes account of the influence of perturbing the observations on the change in the difference between the unrestricted and restricted predicted values. Note that $\mathrm{var}\{(\hat{Y}_i - \tilde{Y}_i)|\mathbf{C}_\mathrm{R}\} = g_{ii}\sigma^2$, where g_{ii} is the ith diagonal element of the matrix **G**. Hence, under violation, the larger is g_{ii}, the larger tends to be the difference between the predicted values. Observations with large value for the difference $\hat{Y} - \tilde{Y}$ may be discordant with the prior belief for the parameters.

In addition, we have $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y} = (\mathbf{H} - \mathbf{G})\mathbf{y}$. Consequently, we can study the influence of the ith observation on its own restricted fitted value by the relationship

$$\tilde{y}_i = m_{ii}y_i + \sum_{j=1, j\neq i}^{n} m_{ij}y_j, \tag{4}$$

where $m_{ij} = h_{ij} - g_{ij}$. Since **M** is an idempotent matrix it follows that $0 \leqslant m_{ii} \leqslant 1$, which implies $h_{ii} - g_{ii} \geqslant 0$. Thus, from equations (4), observations with large values for m_{ii} (h_{ii} large and g_{ii} small) are potentially influential on their restricted fitted values. Similarly to Hoaglin and Welsch (1978), a criterion here would be to select for further analysis those observations with m_{ii} greater than twice the mean leverage, given by

$$\overline{m} = \sum_{i=1}^{n} m_{ii}/n = \mathrm{tr}(\mathbf{M})/n = (p - q)/n.$$

Therefore, the results above suggest the following leverage measures:

(a) h-leverage, for assessing the influence of the observations on their fitted values, when $\mathbf{C}\hat{\boldsymbol{\beta}} \geqslant \mathbf{0}$,
(b) m-leverage, for assessing the influence of the observations on their restricted fitted values and
(c) g-leverage, for assessing the influence of the observations on the difference between the restricted and unrestricted fitted values, when $\hat{\boldsymbol{\beta}}$ is not in accordance with $\mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{0}$.

2.1. Simple linear regression

Suppose now that $E(Y_i) = \beta_1 + \beta_2 x_i$ with the constraint $\beta_2 \geqslant 0$. If we have the violation $\hat{\beta}_2 < 0$, then $\tilde{\beta}_1 = \overline{y}$ and $\tilde{\beta}_2 = 0$. We may show in this case that

$$m_{ii} = 1/n$$

and

$$g_{ii} = (x_i - \overline{x})^2 / \sum_{j=1}^{n} (x_j - \overline{x})^2,$$

for $i = 1, \ldots, n$. Therefore, under violation, no observation is influential on its restricted fitted value. However, those observations that are remote in the subspace $C(\mathbf{X})$ are influential on the change in the difference between the predicted values.

2.2. Example

To present an illustration, consider the data set given in Table 1, which describes the level of application of three different materials used in various concentrations in an attempt to increase the chance of survival (percentage) of a certain type of animal semen (Walpole and Myers (1972), p. 311). Armstrong and Frome (1976) suggested the use of a linear regression $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \sigma\epsilon_i$, where $\epsilon_i \sim N(0, 1)$, $i = 1, \ldots, 13$, with the constraints $\beta_j \geqslant 0$, $j = 1, 2, 3$. They imposed these constraints to remove those variables which would decrease the semen's chance of survival.

Table 2 presents the OLS and ICLS estimates and the corresponding standard deviations. Because there is no closed form for var($\tilde{\boldsymbol{\beta}}$), we consider the standard deviation of $\tilde{\boldsymbol{\beta}}$ from

$$\mathrm{var}(\tilde{\boldsymbol{\beta}}|\mathbf{C}_{\mathrm{R}}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\{\mathbf{I} - \mathbf{C}_{\mathrm{R}}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{C}_{\mathrm{R}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\},$$

with σ^2 being replaced by its restricted estimate

$$\tilde{\sigma}^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2/(n-p).$$

Note that we have two violations, namely $\hat{\beta}_2 < 0$ and $\hat{\beta}_3 < 0$. From Armstrong and Frome (1976) it

Table 1. Results for 13 experiments involving the concentration of three different materials x_i used to increase the survival of animal semen†

Observation	y	x_1	x_2	x_3
1	25.5	1.74	5.30	10.80
2	31.2	6.32	5.42	9.40
3	25.9	6.22	8.41	7.20
4	38.4	10.52	4.63	8.50
5	18.4	1.19	11.60	9.40
6	26.7	1.22	5.85	9.90
7	26.4	4.10	6.62	8.00
8	25.9	6.32	8.72	9.10
9	32.0	4.08	4.42	8.70
10	25.2	4.15	7.60	9.20
11	39.7	10.15	4.83	9.40
12	35.7	1.72	3.12	7.60
13	26.5	1.70	5.30	8.20

†Source: Armstrong and Frome (1976).

Table 2. Unconstrained and constrained estimates for the animal semen example

β_0	β_1	β_2	β_3	σ^2	R^2
Unconstrained estimates					
39.157	1.016	−1.862	−0.343	4.297†	0.912
(5.887)	(0.191)	(0.267)	(0.617)		
Constrained estimates					
23.380	1.234	0.000	0.000	27.869†	0.427
(2.622)	(0.476)	(0.000)	(0.000)		

†9 degrees of freedom.

follows that the ICLS estimates agree with the least squares estimates subject to the linear equality constraints $\mathbf{C}_R\boldsymbol{\beta} = \mathbf{0}$, where

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

and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$. Fig. 1 displays the index plot for g_{ii} and m_{ii}. We see from Fig. 1(a) four outstanding values for observations 1, 3, 5 and 12, which may be influential on the change in the difference between the predicted values. Looking at the data set in Table 1, we notice some disagreements between these observations and the prior assumption of non-negative coefficients β_2 and β_3. For instance, observation 1 has a small value for Y and a large value x_3, observation 5 has a small value for Y and a large value x_2 and observation 12 has a large value for Y and small values x_2 and x_3. From Fig. 1(b) we notice observations 4 and 11 with high m-leverage. These observations have a disproportionate influence on their restricted fitted values.

The local influence of the four outstanding observations on the change in the difference between the predicted values (Fig. 1(a)) is confirmed in Fig. 2, which plots the difference $d_i = \hat{y}_{i,b} - \tilde{y}_{i,b} - (\hat{y} - \tilde{y})$ against b changing in the interval $[-2, 2]$, for $i = 1, \ldots, 13$. Clearly, the variations corresponding to these four outstanding observations are much more accentuated than the variations in the remaining observations.

3. Generalized linear models

Suppose now that $Y_1, \ldots, Y_n$ are independent random variables and belong to the exponential family of distributions such that $E(Y_i) = \mu_i$ and $\text{var}(Y_i) = V(\mu_i)\phi^{-1}$, where $V(\mu_i)$ is the variance function of μ_i and ϕ^{-1} is the dispersion parameter. Further, assume that $\mu_i = g^{-1}(\eta_i)$, where $\eta_i = \mathbf{x}_i^T\boldsymbol{\beta}$, $g(\cdot)$ is the link function and the parameters are constrained to $\mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{0}$. To motivate the construction of the leverage matrices for generalized linear models, consider the iterative process

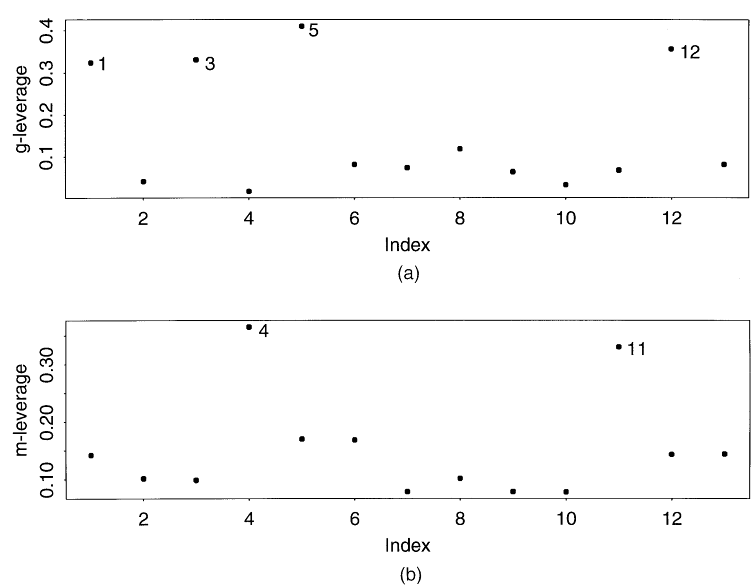

Fig. 1. Index plots for (a) g_{ii} and (b) m_{ii}

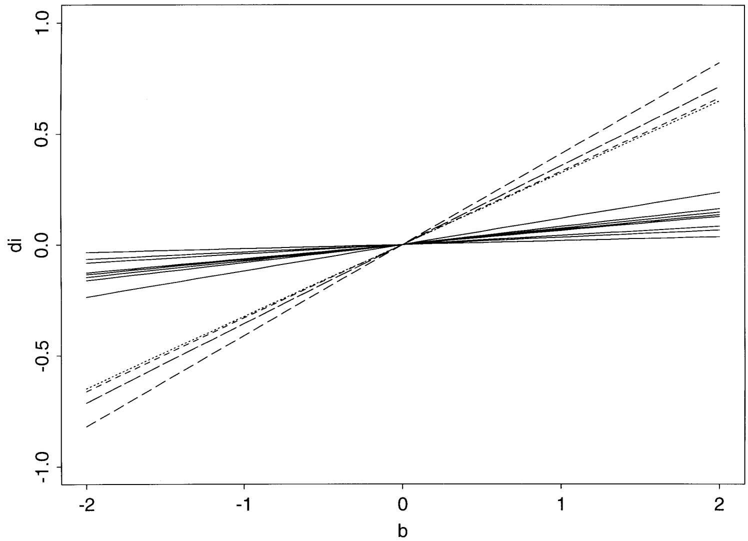

Fig. 2. Variation in the difference between predicted values: ······, observation 1; -----, observation 3; — — —, observation 5; ——, observation 12; ——, other observations

for obtaining the unrestricted maximum likelihood estimate of $\boldsymbol{\beta}$ (see, for instance, McCullagh and Nelder (1989), p. 43), which is given at convergence by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}\hat{\mathbf{z}}, \tag{5}$$

where $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $\mathbf{V} = \mathrm{diag}(V_1, \ldots, V_n)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ and $V_i = V(\mu_i)$. Equation (5) is the least squares solution for the linear regression model

$$\hat{\mathbf{Z}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{v}, \qquad \boldsymbol{v} \sim N_n(\mathbf{0}, \hat{\mathbf{V}}^{-1}), \tag{6}$$

where $\hat{\mathbf{Z}}$ denotes a random variable with observed value $\hat{\mathbf{z}}$. The leverage matrix for model (6) is given by

$$\hat{\mathbf{H}} = \hat{\mathbf{V}}^{1/2}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{1/2},$$

which may be interpreted as the projection matrix onto the subspace spanned by the columns of $\hat{\mathbf{V}}^{1/2}\mathbf{X}$. Indeed, $\hat{\mathbf{H}}$ is the projection matrix onto the tangent plane to the surface $\mathcal{S} = \{\boldsymbol{\mu}(\boldsymbol{\beta}); \boldsymbol{\beta} \in \mathbb{R}^p\}$ at $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$. The principal diagonal elements of $\hat{\mathbf{H}}$ have been suggested as local leverage measures in unrestricted generalized linear models (see, for instance, Pregibon (1981) and McCullagh and Nelder (1989), p. 397). However, because

$$\hat{h}_{ii} = \hat{V}_i\mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{x}_i,$$

observations remote in $C(\mathbf{X})$, which have

$$b_{ii} = \mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i$$

large, do not necessarily have a large value for $\hat{h}_{ii}$.

Hosmer and Lemeshow (1989), p. 153, presented some numerical studies for $\hat{h}_{ii}$ in logistic regression. They showed that the leverage increases as the estimated probability moves from 0.5 to 0.9 or from 0.5 to 0.1. However, $\hat{h}_{ii}$ decreases rapidly as the estimated probability approaches 0 or 1. Therefore, $\hat{h}_{ii}$ has an interpretation similar to that of h_{ii} for an estimated probability in the interval [0.1, 0.9]. It may be useful to display the plot of $\hat{h}_{ii}$ against the restricted fitted values.

Extensions of **M** and **G** to restricted generalized linear models may be performed in a similar way to that given in Section 2, by assuming the constraints $\mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{0}$ in model (6). Then, we find the local leverage matrix $\hat{\mathbf{M}} = \hat{\mathbf{H}} - \hat{\mathbf{G}}$, where

$$\hat{\mathbf{G}} = \hat{\mathbf{Z}}(\hat{\mathbf{Z}}^{\mathrm{T}}\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^{\mathrm{T}}$$

and

$$\hat{\mathbf{Z}} = \hat{\mathbf{V}}^{1/2}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}.$$

The principal diagonal elements of $\hat{\mathbf{M}}$ and $\hat{\mathbf{G}}$ may be used for revealing observations which are influential in the same sense as m_{ii} and g_{ii}.

3.1. Example

As an illustration, consider the data set described in Table 3 (McDonald and Diamond (1983), Table 1) on the distribution of pregnancy and natural abortion according to the degree of consanguinity between the parents in three districts of Shizuoka City, Japan. Let π_{ij} denote the probability of natural abortion for the ith degree of consanguinity (1, no relation; 2, second cousins; 3, $1\text{–}\frac{1}{2}$ cousins; 4, first cousins) and jth district (1, rural; 2, intermediate; 3, urban). They argued that the probability of natural abortion should increase with the degree of consanguinity and suggested fitting the linear logistic model

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha + \delta_i + \gamma_j, \qquad i = 1, \ldots, 4 \text{ and } j = 1, 2, 3, \tag{7}$$

to the data, where $\delta_1 = \gamma_1 = 0$ with the parameters subject to the constraints $0 \leqslant \delta_2 \leqslant \delta_3 \leqslant \delta_4$.

Table 4 presents the unrestricted and restricted maximum likelihood estimates and the approximate standard deviations. Since we have the violation $\hat{\delta}_3 > \hat{\delta}_4$, the restricted maximum likelihood estimates may be obtained by fitting model (7) constrained to $\mathbf{C}_{\mathrm{R}}\boldsymbol{\beta} = \mathbf{0}$ (see, for instance, McDonald and Diamond (1983)), where $\boldsymbol{\beta} = (\alpha, \delta_2, \delta_3, \delta_4, \gamma_2, \gamma_3)^{\mathrm{T}}$ and $\mathbf{C}_{\mathrm{R}} =$ (0, 0, −1, 1, 0, 0). Algorithms to fit linear inequality generalized linear models may also be found in McDonald and Diamond (1990) and Fahrmeir and Klinger (1994). An approximate standard deviation for $\tilde{\boldsymbol{\beta}}$ may be obtained from Nyquist (1991):

Table 3. Outcomes of 6358 pregnancies in three districts of Shizuoka City, Japan, according to the degree of consanguinity of the parents

Case	*Residence*	*Consanguinity*	*Pregnancies*	*Abortions*	*% abortions*
1		No relation	958	27	2.82
2	Rural	2nd cousins	160	1	0.62
3	District	$1\text{–}\frac{1}{2}$ cousins	65	3	4.61
4		1st cousins	293	12	4.09
5		No relation	2670	67	2.51
6	Intermediate	2nd cousins	338	11	3.25
7	District	$1\text{–}\frac{1}{2}$ cousins	237	11	4.64
8		1st cousins	654	23	3.52
9		No relation	543	7	1.29
10	Urban	2nd cousins	70	4	5.71
11	District	$1\text{–}\frac{1}{2}$ cousins	110	3	2.73
12		1st cousins	260	7	2.69

Table 4. Unconstrained and constrained estimates for the pregnancy and natural abortion example

α	δ_2	δ_3	δ_4	γ_2	γ_3	*Deviance*
Unconstrained estimates						
−3.647	0.152	0.598	0.402	−0.010	−0.387	9.041†
(0.169)	(0.273)	(0.269)	(0.187)	(0.182)	(0.271)	
Constrained estimates						
−3.651	0.153	0.454	0.454	−0.004	−0.378	9.473†
(0.169)	(0.273)	(0.168)	(0.168)	(0.182)	(0.271)	

†6 degrees of freedom.

$$\text{var}(\tilde{\boldsymbol{\beta}}|\mathbf{C}_{\text{R}}) \cong (\mathbf{X}^{\text{T}}\tilde{\mathbf{V}}\mathbf{X})^{-1}[\mathbf{I} - \mathbf{C}_{\text{R}}^{\text{T}}\{\mathbf{C}_{\text{R}}(\mathbf{X}^{\text{T}}\tilde{\mathbf{V}}\mathbf{X})^{-1}\mathbf{C}_{\text{R}}^{\text{T}}\}^{-1}\mathbf{C}_{\text{R}}(\mathbf{X}^{\text{T}}\tilde{\mathbf{V}}\mathbf{X})^{-1}].$$

Figs 3(a) and 3(b) display respectively the plots of $\hat{g}_{ii}$ and $\hat{m}_{ii}$ against the restricted fitted values. We see in Fig. 3(a) outstanding influence for case 7. Its observed proportion of natural abortion is larger than the proportion for case 8, contradicting the prior assumption $\delta_3 \leqslant \delta_4$. In Fig. 3(b), case 5 appears with a large value for $\hat{m}_{ii}$, greater than $2(p-q)/n$. Owing to the large number of pregnancies case 5 appears remote in the subspace spanned by the columns of $\hat{\mathbf{V}}^{1/2}\mathbf{X}$.

4. Concluding remarks

The identification of observations which are discordant with the prior assumption for the parameters in inequality-constrained linear regression models may be very useful for finding informative points in the data set. The well-known **H** hat matrix from the linear regression model may be not appropriate when the constraints are violated by the unrestricted estimates. In these cases, **H** is decomposed into $\mathbf{H} = \mathbf{M} + \mathbf{G}$, and the information about leverage should be obtained

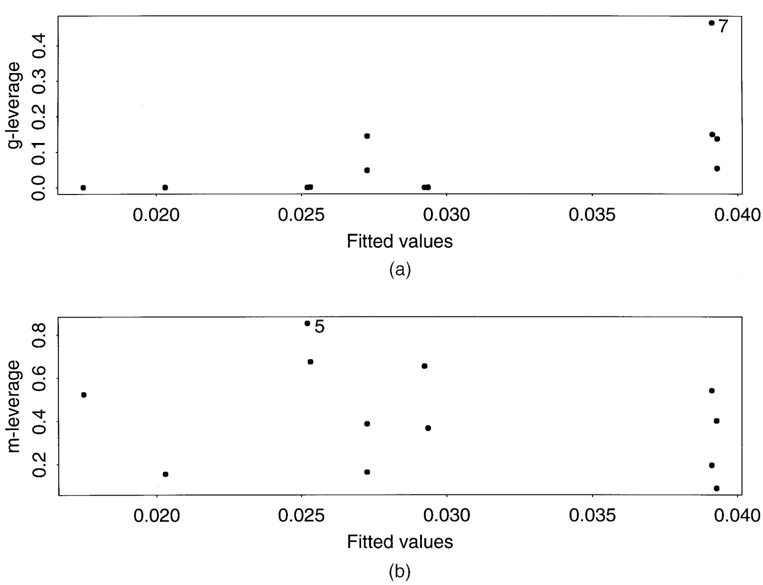

Fig. 3. Plots of (a) $\hat{g}_{ii}$ and (b) $\hat{m}_{ii}$ against the fitted values

from the principal diagonal elements of **M** and **G**, which may reveal respectively observations with high influence on the restricted predicted values and violations. It is interesting that observations with high g-leverage should not have a great effect on their restricted fitted values. High m-leverage is associated with large values for h_{ii} and small values for g_{ii}. The extension of the procedures to generalized linear models is straightforward but care should be taken in the interpretation of the diagnostic graphs, since the measures $\hat{m}_{ii}$ and $\hat{g}_{ii}$ depend on the fitted values.

Acknowledgements

This work was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico and Fundação de Amparo e a Pesquisa do Estado de São Paulo, Brazil. The author is grateful to two referees whose valuable comments and suggestions have improved the exposition of this paper.

Appendix A: Proofs of equations (2) and (3)

To prove equation (2) consider that

$$\tilde{\mathbf{y}}_{l,b} - \tilde{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}_l(b) - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{\mathrm{R}_b} - \mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{\mathrm{R}}),$$

where $\mathbf{C}_{\mathrm{R}_b}$ denotes the matrix $\mathbf{C}_R$ under the perturbation on the lth observation,

$$\hat{\boldsymbol{\beta}}_l(b) = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{y} + b\mathbf{f}_l)$$

and

$$\hat{\boldsymbol{\Lambda}}_{\mathrm{R}_b} = -\{\mathbf{C}_{\mathrm{R}_b}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}_b}\hat{\boldsymbol{\beta}}_l(b).$$

Therefore,

$$\lim_{b\to 0}\left[\frac{1}{b}\{\mathbf{X}\hat{\boldsymbol{\beta}}_l(b) - \mathbf{X}\hat{\boldsymbol{\beta}}\}\right] = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}_l,$$

and by assuming that $\mathbf{C}_{\mathrm{R}}$ is constant in a neighbourhood of $\mathbf{y}$ we have

$$\begin{aligned}
&\lim_{b\to 0}\left[\frac{1}{b}\{\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{\mathrm{R}_b} - \mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{\mathrm{R}})\}\right]\\
&\quad = \lim_{b\to 0}\left(\frac{1}{b}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}[\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\{\mathbf{C}_{\mathrm{R}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}} - \mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\{\mathbf{C}_{\mathrm{R}_b}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}_b}](\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right)\\
&\qquad - \lim_{b\to 0}\left[\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\{\mathbf{C}_{\mathrm{R}_b}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}_b}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}_b}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}_l\right]\\
&\quad = \mathbf{0} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\{\mathbf{C}_{\mathrm{R}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}\}^{-1}\mathbf{C}_{\mathrm{R}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}_l.
\end{aligned}$$

By making $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$ and $\mathbf{G} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, where $\mathbf{Z} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}_{\mathrm{R}}^{\mathrm{T}}$, we find that

$$\lim_{b\to 0}\left\{\frac{1}{b}(\tilde{\mathbf{y}}_{l,b} - \tilde{\mathbf{y}})\right\} = (\mathbf{H} - \mathbf{G})\mathbf{f}_l. \tag{8}$$

To prove equation (3) consider that

$$\lim_{b\to 0}\left\{\frac{1}{b}(\hat{\mathbf{y}}_{l,b} - \hat{\mathbf{y}})\right\} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{f}_l.$$

Then, using equation (8) we obtain

$$\lim_{b\to 0}\left[\frac{1}{b}\{\hat{\mathbf{y}}_{l,b} - \hat{\mathbf{y}} - (\tilde{\mathbf{y}}_{l,b} - \tilde{\mathbf{y}})\}\right] = \mathbf{H}\mathbf{f}_l - (\mathbf{H} - \mathbf{G})\mathbf{f}_l = \mathbf{G}\mathbf{f}_l.$$

References

Armstrong, R. D. and Frome, E. L. (1976) A branch-and-bound solution of a restricted least squares problem. *Technometrics*, **18**, 447–450.

Emerson, J. D., Hoaglin, D. C. and Kempthorne, P. J. (1984) Leverage in least squares additive-plus-multiplicative fits for two-way tables. *J. Am. Statist. Ass.*, **79**, 329–335.

Fahrmeir, L. and Klinger, J. (1994) Estimating and testing generalized linear models under inequality restrictions. *Statist. Pap.*, **35**, 211–229.

Hoaglin, D. C. and Welsch, R. E. (1978) The hat matrix in regression and ANOVA. *Am. Statistn*, **32**, 17–22.

Hosmer, D. W. and Lemeshow, S. (1989) *Applied Logistic Regression*. New York: Wiley.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

McDonald, J. M. and Diamond, I. (1983) Fitting generalized linear models with linear inequality constraints. *GLIM Newslett.*, **6**, 29–36.

———(1990) On the fitting of generalized linear models with nonnegative parameter constraints. *Biometrics*, **46**, 201–206.

Nyquist, H. (1991) Restricted estimation of generalized linear models. *Appl. Statist.*, **40**, 133–141.

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

Pregibon, D. (1981) Logistic regression diagnostics. *Ann. Statist.*, **9**, 705–724.

St Laurent, R. T. and Cook, R. D. (1992) Leverage and superleverage in nonlinear regression. *J. Am. Statist. Ass.*, **87**, 985–990.

Theil, H. and Van de Panne, C. (1960) Quadratic programming as an extension of classical quadratic maximization. *Mangmnt Sci.*, **7**, 1–20.

Walpole, R. E. and Myers, R. H. (1972) *Probability and Statistics for Engineers and Scientists*. New York: Macmillan.