Ann. Inst. Statist. Math.
Vol. 51, No. 2, 383–398 (1999)

COVARIATE TRANSFORMATION DIAGNOSTICS FOR GENERALIZED LINEAR MODELS

ANDY H. LEE* AND JOHN S. YICK**

Faculty of Science, Northern Territory University, Darwin, NT 0909, Australia

(Received April 30, 1996; revised October 20, 1997)

Abstract. Transformations of covariates are commonly applied in regression analysis. When a parametric transformation family is used, the maximum likelihood estimate of the transformation parameter is often sensitive to minor perturbations of the data. Diagnostics are derived to assess the influence of observations on the covariate transformation parameter in generalized linear models. Three numerical examples are presented to illustrate the usefulness of the proposed diagnostics.

Key words and phrases: Local influence, partial influence, profile likelihood displacement, transformation diagnostics.

1. Introduction

Transformations of variables have often been applied to data in statistical modelling. Parametric transformation families, such as the Box-Cox power transformation, is commonly used. Various diagnostics have been proposed to assess the sensitivity of the maximum likelihood estimate (MLE) of the transformation parameter; see e.g. Cook and Wang (1983), Atkinson (1986, 1988), Wang (1987), Tsai and Wu (1990). Most of these methods are concerned with transformation of the response or simultaneous transformation (transform-both-sides model). Transformation diagnostics for the covariates, however, have been studied to a lesser extent. Ezekiel and Fox (1959) introduced the partial residual plot. Box and Tidwell (1962) suggested constructed variables and added variable plots to assist the selection of suitable transformations for covariates. A review of such procedures can be found in Cook and Weisberg (1982) and Chatterjee and Hadi (1988).

Traditionally, transformation diagnostics are derived using the case deletion approach (Cook and Weisberg (1982), Wei and Shih (1994*a*)). Since Cook (1986) developed the local influence methodology as a general tool for assessing the effect

* Now at Department of Epidemiology and Biostatistics, Curtin University of Technology, GPO Box U 1987, Perth, WA 6845, Australia.

** Now at Statistics and Demography Unit, Northern Territory Department of Education, GPO Box 4821, Darwin, NT 0801, Australia.

of small departures from model assumptions, there is a large body of literature dealing with response transformation and simultaneous transformations based on this approach (Lawrance (1988), Hinkley and Wang (1988), Tsai and Wu (1992), Shih (1993), Wei and Shih (1994*b*)). In contrast, limited diagnostics are available for analyzing the transformation of covariates. Cook (1987) used a subset formula from local influence to derive diagnostics for partially nonlinear models, which include transformation of a single covariate as a special case. Wei and Hickernell (1996) considered further extensions to several covariates based on profile likelihood displacement and found that their diagnostics are related to those of Cook (1987). Nevertheless, all of the above methods are devoted exclusively to the linear regression setting.

The aim of this paper is to present influence diagnostics for assessing the effect of minor perturbations on the MLE of the covariate transformation parameters in generalized linear models. Two separate approaches based on analysis of the transformation parameter surface or profile likelihood displacement, and partial influence, are proposed in the next section. Specific perturbation schemes are outlined in Section 3 to examine the different aspects of influence. Three illustrative examples are provided in Section 4.

2. Profile likelihood displacement and local influence

We assume the responses $\boldsymbol{y} = \langle y_1, \ldots, y_n \rangle^T$ have a density or mass function of the form

$$f_y(y_i;\boldsymbol{\theta}) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\}$$

with $\theta_i = k(\eta_i)$, where η_i is the linear predictor and $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ are known functions. Without loss of generality the dispersion parameter ϕ is assumed known or may be replaced by an estimate $\hat{\phi}$ and write $\hat{a} = a(\hat{\phi})$, which gives an exponential-family density with natural parameter $\boldsymbol{\theta}$. The log-likelihood function is then

$$\hat{a}^{-1} \sum_{i=1}^{n} [y_i k(\eta_i) - b\{k(\eta_i)\}].$$

Goodness-of-fit of a generalized linear model may often be improved by transforming one or more covariates $\boldsymbol{z}$ of $X = \langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(p)}, \boldsymbol{z}_{(1)}, \ldots, \boldsymbol{z}_{(q)} \rangle$. Let the linear predictor of the transformation model be

$$\boldsymbol{\eta} = \boldsymbol{x}\boldsymbol{\delta} + G(\boldsymbol{z}, \boldsymbol{\lambda})\boldsymbol{\xi},$$

where the $n \times q$ matrix $G(\boldsymbol{z}, \boldsymbol{\lambda}) = \langle g_1(\boldsymbol{z}_{(1)}, \lambda_1), \ldots, g_q(\boldsymbol{z}_{(q)}, \lambda_q) \rangle$, and

$$g_j(\boldsymbol{z}_{(j)}, \lambda_j) = \langle g_j(z_{1j}, \lambda_j), \ldots, g_j(z_{nj}, \lambda_j) \rangle^T$$

represents a known, twice continuously differentiable transformation family indexed by λ_j $(j = 1, \ldots, q)$. Here the parameter vector $\boldsymbol{\lambda}$ is of special interest.

Let $\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda})$, $\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda})$ be functions that maximise $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})$ for fixed $\boldsymbol{\lambda}$ and denote the corresponding profile log-likelihood for $\boldsymbol{\lambda}$ by $L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}))$. To assess the global influence of individual cases on the MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$, one can adopt the case deletion approach of Cook and Weisberg (1982). The difference between $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}_{[i]}$, the MLE of $\boldsymbol{\lambda}$ without case i, can be measured through the profile likelihood displacement

$$LD_i = 2[L(\hat{\boldsymbol{\lambda}}) - L(\hat{\boldsymbol{\lambda}}_{[i]})] \tag{2.1}$$

where $L(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}))$. For the special case of linear regression, Wei and Hickernell (1996) obtained approximations to simplify computations of LD_i. A large value of LD_i indicates that $\hat{\boldsymbol{\lambda}}$ is likely to be dependent on case i.

2.1 *First and second order approach*

Wu and Luo (1993*a*, 1993*b*) studied the perturbation-formed MLE surface of a parameter of interest in regression. Unlike the likelihood displacement surface, such a MLE surface does not necessarily have zero first derivative at the null point of no perturbation, so that its slope as well as curvature can be used to examine local influence. They referred to the maximum slope direction of the fixed surface at the null point as the first order approach, whereas assessment of directions corresponding to large normal curvatures is called the second order approach.

For the MLE surface of the transformation parameter $\boldsymbol{\lambda}$, we now introduce small changes into our model through an $n \times 1$ vector $\boldsymbol{\omega} \in \Omega$, where Ω denotes the open set of relevant perturbations. Suppose there is a null point $\boldsymbol{\omega}_0$ in Ω representing no perturbation so that $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}_0} = \hat{\boldsymbol{\lambda}}$. In the manner of Wu and Luo (1993*a*), the MLE surface of $\boldsymbol{\lambda}$ is the geometric influence graph formed by $\boldsymbol{\alpha}(\boldsymbol{\omega}) = \langle \boldsymbol{\omega}^T, \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}} \rangle$, where $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ is the estimate of $\boldsymbol{\lambda}$ under perturbation $\boldsymbol{\omega}$.

To find the direction of largest local change, we approximate the MLE surface by its tangent plane at $\boldsymbol{\omega}_0$, which is determined by $\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$. The direction of largest local change is just the direction of maximum slope on this tangent plane over Ω. The derivations below are similar to those of Shih (1993, p. 414) for the tranform-both-sides model. Write $L(\boldsymbol{\lambda} \mid \boldsymbol{\omega}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega}) \mid \boldsymbol{\omega})$ for the profile log-likelihood corresponding to the perturbed model, where $\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega})$ and $\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega})$ are functions that maximise $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi} \mid \boldsymbol{\omega})$ for fixed $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$. Then $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ satisfies the following equation:

$$\frac{\partial L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial \boldsymbol{\lambda}} = \mathbf{0}.$$

Differentiating with respect to ω_i yields

$$\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \left(\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} \right) + \frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} = \mathbf{0}.$$

Therefore,

$$\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} = \left[\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \right]^{-1} \left(\frac{-\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} \right). \tag{2.2}$$

Let $L_1^{(1)}$ and $L_2^{(1)}$ be the derivatives of $L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi} \mid \boldsymbol{\omega})$ with respect to $\boldsymbol{\lambda}$ and $(\boldsymbol{\delta},\boldsymbol{\xi})$ respectively, with superscript (t) denoting the t-th derivative of the function. It can be shown that

$$\frac{\partial^2 L(\boldsymbol{\lambda};\tilde{\boldsymbol{\delta}},\tilde{\boldsymbol{\xi}} \mid \boldsymbol{\omega})}{\partial\omega_i\partial\boldsymbol{\lambda}^T} = L_{1\omega i}^{(2)} + L_{12}^{(2)}\frac{\partial(\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega}),\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda} \mid \boldsymbol{\omega}))}{\partial\omega_i} = L_{1\omega i}^{(2)} - L_{12}^{(2)}[L_{22}^{(2)}]^{-1}L_{2\omega i}^{(2)}.$$

The partitions of $L^{(2)}$ are

$$\begin{aligned}
L_{11}^{(2)} &= -\sum_{i=1}^{n} S2_i G^{(1)}(z_i,\boldsymbol{\lambda})^T\boldsymbol{\xi}^T\boldsymbol{\xi}G^{(1)}(z_i,\boldsymbol{\lambda}) + S1_i\,\mathrm{diag}(\boldsymbol{\xi})G^{(2)}(z_i,\boldsymbol{\lambda})\\
L_{12}^{(2)} &= \sum_{i=1}^{n}\langle S2_i\,\mathrm{diag}(\boldsymbol{\xi})G^{(1)}(z_i,\boldsymbol{\lambda})^T\boldsymbol{x}_i,\\
&\qquad S2_i G(z_i,\boldsymbol{\lambda})^T G^{(1)}(z_i,\boldsymbol{\lambda})\,\mathrm{diag}(\boldsymbol{\xi}) + S1_i\,\mathrm{diag}(G^{(1)}(z_i,\boldsymbol{\lambda}))\rangle\\
L_{21}^{(2)} &= [L_{12}^{(2)}]^T\\
L_{22}^{(2)} &= \sum_{i=1}^{n} S2_i\langle\boldsymbol{x}_i, G(z_i,\boldsymbol{\lambda})\rangle^T\langle\boldsymbol{x}_i, G(z_i,\boldsymbol{\lambda})\rangle
\end{aligned}$$

where $\boldsymbol{x}_i = \langle x_{i1},\ldots,x_{ip}\rangle$ denotes the i-th row of $\boldsymbol{x}$,

$$G^{(1)}(z_i,\boldsymbol{\lambda}) = \left\langle \frac{g_1(z_{i1},\lambda_1)}{\partial\lambda_1},\ldots,\frac{g_q(z_{iq},\lambda_q)}{\partial\lambda_q}\right\rangle$$

is a $1\times q$ vector,

$$G^{(2)}(z_i,\boldsymbol{\lambda}) = \frac{\partial G(z_i,\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T}$$

is a $q\times q$ matrix,

$$\begin{aligned}
S1_i &= y_i k^{(1)}(\eta_i) - b^{(1)}(k(\eta_i))k^{(1)}(\eta_i),\\
S2_i &= y_i k^{(2)}(\eta_i) - \{b^{(2)}(k(\eta_i))[k^{(1)}(\eta_i)]^2 + b^{(1)}(k(\eta_i))k^{(2)}(\eta_i)\},
\end{aligned}$$

while $L_{1\omega i}^{(2)} = \frac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\omega_i\partial\boldsymbol{\lambda}^T}$ and $L_{2\omega i}^{(2)} = \frac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\omega_i\partial(\boldsymbol{\delta},\boldsymbol{\xi})^T}$ are entries on the corresponding columns of

$$\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega;\boldsymbol{\delta}_\omega,\boldsymbol{\xi}_\omega) = \frac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi} \mid \boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi})^T} \tag{2.3}$$

and further derived for various types of perturbations in Section 3. All of the above quantities are evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$.

To compute the maximum slope direction at the null point, $\boldsymbol{l}_{\text{slope}}^{\hat{\boldsymbol{\lambda}}}$, we evaluate $\frac{\partial\hat{\boldsymbol{\lambda}}_\omega}{\partial\boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$. For second order local influence, the direction $\boldsymbol{l}_{\max}^{\hat{\boldsymbol{\lambda}}}$ which corresponds to the maximum normal curvature of the MLE surface is the main diagnostic quantity to study the combined effect on $\hat{\boldsymbol{\lambda}}$. According to matrix theory, the local maximum curvatures and the corresponding directions are the solution of the generalized eigenvector equation

$$(A - eB)\boldsymbol{l} = \mathbf{0} \tag{2.4}$$

where the matrices

$$A = \frac{\partial^2 \hat{\boldsymbol{\lambda}}_{\omega}}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T} = \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\omega}}{\partial \boldsymbol{\omega}^T}\right]^T \left(\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} \mid \omega)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right) \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\omega}}{\partial \boldsymbol{\omega}^T}\right]$$

and

$$B = I + \frac{\partial \hat{\boldsymbol{\lambda}}_{\omega}}{\partial \boldsymbol{\omega}^T} \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\omega}}{\partial \boldsymbol{\omega}^T}\right]^T$$

are evaluated at $\boldsymbol{\omega}_0$. Alternatively, we can apply the subset formulation of Cook (1987) to the profile likelihood displacement

$$LD(\boldsymbol{\omega}) = 2[L(\hat{\boldsymbol{\lambda}}) - L(\hat{\boldsymbol{\lambda}}_{\omega})]. \tag{2.5}$$

An equivalent form to (2.4) is then

$$\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega}) \left\{ [L^{(2)}]^{-1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [L_{22}^{(2)}]^{-1} \end{pmatrix} \right\} \boldsymbol{\Delta}^T(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega}). \tag{2.6}$$

If λ is a scalar parameter (transformation of a single covariate, i.e. $q = 1$), then $\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} | \omega)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}$ becomes a scalar quantity so that $\boldsymbol{l}_{\max}^{\hat{\lambda}}$ is proportional to $\boldsymbol{l}_{\text{slope}}^{\hat{\lambda}} = (\partial \hat{\lambda}_{\omega} / \partial \boldsymbol{\omega}^T)^T$. Upon simplifying,

$$\frac{\partial \hat{\lambda}_{\omega}}{\partial \boldsymbol{\omega}^T} \propto L_{1\omega}^{(2)} - L_{12}^{(2)} [L_{22}^{(2)}]^{-1} L_{2\omega}^{(2)}. \tag{2.7}$$

2.2 *Partial influence approach*

Let the linear predictor before transformation of covariate $\boldsymbol{z}$ be given by

$$\boldsymbol{\eta}_0 = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{z}\gamma. \tag{2.8}$$

The model after transformation is

$$\boldsymbol{\eta} = \boldsymbol{x}\boldsymbol{\delta} + G(\boldsymbol{z}, \boldsymbol{\lambda})\boldsymbol{\xi}, \tag{2.9}$$

so that $G(\boldsymbol{z}, \boldsymbol{\lambda}_0) = \boldsymbol{z}$ represents no transformation. A test of the hypothesis H_0: $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ can be based on $D_0 - D$, the reduction in deviance from model (2.8) to model (2.9). It is important to judge whether any particular observation has an undue impact on this test. Denote $D_{0[i]}$ and $D_{[i]}$ for the deviances of (2.8) and (2.9) respectively after deleting case i. In the manner of Lee (1988), a partial influence measure for the impact of case i on the transformation can be formulated as

$$d_i = (D_0 - D) - (D_{0[i]} - D_{[i]}), \tag{2.10}$$

which represents the change in deviance due to the transformation of $\boldsymbol{z}$ when the i-th observation is excluded.

Consider the log-likelihood $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})$ of the transformation model (2.9). The full MLE of $\boldsymbol{\lambda}$, $\boldsymbol{\delta}$, $\boldsymbol{\xi}$ are denoted by $\hat{\boldsymbol{\lambda}}$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\xi}}$ respectively. Similarly, let $L_0(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be the log-likelihood of model (2.8), with MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. Under minor perturbations, the respective log-likelihood becomes $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi} \mid \boldsymbol{\omega})$ and $L_0(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{\omega})$, with associated MLEs $(\hat{\boldsymbol{\lambda}}_\omega, \hat{\boldsymbol{\delta}}_\omega, \hat{\boldsymbol{\xi}}_\omega)$ and $(\hat{\boldsymbol{\beta}}_\omega, \hat{\boldsymbol{\gamma}}_\omega)$. Suppose that $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi} \mid \boldsymbol{\omega}_0) = L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})$ and $L_0(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{\omega}_0) = L_0(\boldsymbol{\beta}, \boldsymbol{\gamma})$. The partial influence on the transformation due to perturbation $\boldsymbol{\omega}$ can be assessed by

$$d(\boldsymbol{\omega}) = 2\{[L_0(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) - L(\hat{\boldsymbol{\lambda}}; \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\xi}})] - [L_0(\hat{\boldsymbol{\beta}}_\omega, \hat{\boldsymbol{\gamma}}_\omega) - L(\hat{\boldsymbol{\lambda}}_\omega; \hat{\boldsymbol{\delta}}_\omega, \hat{\boldsymbol{\xi}}_\omega)]\}. \tag{2.11}$$

Analogous to (2.10) in case-deletion, the log-likelihood displacement $d(\boldsymbol{\omega})$ measures the local effect on the transformation parameter with respect to the contours of the unperturbed deviance reduction.

Let $F = \frac{1}{2} d(\boldsymbol{\omega})$. We obtain the normal curvature at $F(\boldsymbol{\omega}_0)$ along the direction $\boldsymbol{l}$ as

$$C(\boldsymbol{l}) = 2|\boldsymbol{l}^T F^{(2)} \boldsymbol{l}|,$$
$$F^{(2)} = \frac{\partial^2 L(\hat{\boldsymbol{\lambda}}_\omega; \hat{\boldsymbol{\delta}}_\omega, \hat{\boldsymbol{\xi}}_\omega)}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T} - \frac{\partial^2 L_0(\hat{\boldsymbol{\beta}}_\omega, \hat{\boldsymbol{\gamma}}_\omega)}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}.$$

Applying the chain rule of differentiation yields

$$\begin{aligned} F^{(2)} = {} & \boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega)[L^{(2)}(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})]^{-1} \boldsymbol{\Delta}^T(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) \\ & - \nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega)[L_0^{(2)}(\boldsymbol{\beta}, \boldsymbol{\gamma})]^{-1} \nabla^T(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega) \end{aligned} \tag{2.12}$$

where $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega)$ is defined in (2.3),

$$\nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega) = \frac{\partial^2 L_0(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial (\boldsymbol{\beta}, \boldsymbol{\gamma})^T},$$

evaluated at $\boldsymbol{\omega}_0$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$. Expressions for $L_0^{(2)}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and $\nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega)$ are derived under the original generalized linear model (2.8), which can be found in Thomas and Cook (1989). Let $\boldsymbol{l}^d_{\max}$ be the direction cosines of maximum normal curvature, which is the perturbation direction that produces the greatest local change in $\boldsymbol{\lambda}$ as measured by (2.11). Standard matrix theory shows that $\boldsymbol{l}^d_{\max}$ is just the eigenvector associated with the largest eigenvalue of

$$\boldsymbol{\Delta}[L^{(2)}]^{-1}\boldsymbol{\Delta}^T - \nabla[L_0^{(2)}]^{-1}\nabla^T.$$

The most influential elements of the data on the transformation may be identified by their large components of $\boldsymbol{l}^d_{\max}$. We also recommend to plot $\hat{\boldsymbol{\lambda}}_\omega$ against the perturbation scale for each local direction $\boldsymbol{l}$ of interest. The characteristics of such curves should be informative for further investigation on the relationship between local and global influences.

3. Perturbation schemes

A number of perturbation schemes have been suggested to examine the different aspects of influence (Cook (1987), Thomas and Cook (1989)). In the following, we consider relevant perturbation schemes and derive the corresponding $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega})$ quantity.

3.1 *Perturbation of case weights*

We define a vector of weights $\boldsymbol{\omega} = \langle \omega_1, \ldots, \omega_n \rangle^T$, $\omega_i \geq 0$, to perturb the contribution of each case to the log-likelihood. The point representing no perturbation is $\boldsymbol{\omega}_0 = \langle 1, \ldots, 1 \rangle^T$. We obtain $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega}) = \langle \Delta_i^w \rangle$ evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$, where

$$\Delta_i^w = S1_i \langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}), G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) \operatorname{diag}(\boldsymbol{\xi}) \rangle.$$

The case weight perturbation scheme actually generalizes case deletion, where ω_i is limited to the values 0 or 1. Furthermore, if the deletion of the i-th case is of interest (as revealed by $\boldsymbol{l}_{\max}$), it may be considered as the perturbation located in direction $\boldsymbol{l}_{[i]}$ from the null point, where $\boldsymbol{l}_{[i]}$ is the direction cosines with i-th component -1 but zeros elsewhere (Wu and Luo (1993*a*, 1993*b*)). A plot of $\hat{\boldsymbol{\lambda}}_{\omega}$ in the direction $\boldsymbol{l}_{[i]}$ can then monitor the global effects of downweighting the i-th case.

3.2 *Perturbation of individual covariates $\boldsymbol{x}$*

We modify an individual covariate, say the j-th column $\boldsymbol{x}_{(j)}$ of $\boldsymbol{x}$, to $\boldsymbol{x}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{x}_{(j)} + t\boldsymbol{\omega}$, as long as the covariate is not an indicator variable. Here, t is the scaling factor used to convert the generic perturbation $\boldsymbol{\omega}$ to the appropriate size and units, and $\boldsymbol{\omega}_0 = \langle 0, \ldots, 0 \rangle^T$ represents no perturbation. It can be verified that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega}) = \langle \Delta_i^x \rangle$ where

$$\Delta_i^x = S2_i \delta_j \langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}) + S1_i \boldsymbol{u}_j, G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) \operatorname{diag}(\boldsymbol{\xi}) \rangle,$$

δ_j being the regression coefficient associated with $\boldsymbol{x}_{(j)}$ and $\boldsymbol{u}_j$ denotes a $1 \times q$ row vector with j-th component 1 but zeros elsewhere.

3.3 *Perturbation of individual transformed covariates $\boldsymbol{z}$*

We perturb the j-th transformed covariate $\boldsymbol{z}_{(j)}$ to $\boldsymbol{z}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{z}_{(j)} + t\boldsymbol{\omega}$. Again, t is the appropriate scaling factor and $\boldsymbol{\omega}_0 = \mathbf{0}$ indicates no perturbation. We find that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\omega}; \boldsymbol{\delta}_{\omega}, \boldsymbol{\xi}_{\omega}) = \langle \Delta_i^z \rangle$, where

$$\begin{aligned}\Delta_i^z = \langle & S2_i \xi_j \boldsymbol{x}_i, S2_i \xi_j G(\boldsymbol{z}_i, \boldsymbol{\lambda}) + S1_i G^{(1)}(z_{ij}, \boldsymbol{\lambda}) \boldsymbol{u}_j, \\ & S2_i \xi_j G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) \operatorname{diag}(\boldsymbol{\xi}) + S1_i \xi_j G^{(2)}(z_{ij}, \boldsymbol{\lambda}) \boldsymbol{u}_j \rangle\end{aligned}$$

with ξ_j being the regression coefficient associated with $\boldsymbol{z}_{(j)}$, and $G^{(1)}(z_{ij}, \boldsymbol{\lambda})$ is the j-th entry of $G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda})$, $G^{(2)}(z_{ij}, \boldsymbol{\lambda})$ is the (j, j)-th entry of $G^{(2)}(\boldsymbol{z}_i, \boldsymbol{\lambda})$.

3.4 *Perturbation of responses*

We then consider altering the responses by taking $\boldsymbol{y}(\boldsymbol{\omega}) = \boldsymbol{y} + t\boldsymbol{\omega}$ where $t = \text{diag}\{[\hat{a}b^{(2)}(k(\eta_i))]^{1/2}\}$. As with other additive perturbations, $\boldsymbol{\omega}_0 = \mathbf{0}$ gives the unperturbed state. It follows that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) = \langle \Delta_i^y \rangle$, where

$$\Delta_i^y = k^{(1)}(\eta_i)\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}), G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda})\, \text{diag}(\boldsymbol{\xi})\rangle.$$

Note that this perturbation scheme may not be meaningful for discrete response, such as those in binary logistic regression.

4. Examples

4.1 *Snow geese data*

Consider the snow geese data as reported by Weisberg (1985) and further analyzed in Wei and Hickernell (1996). The data set consists of observations on the response y = true flock size as obtained by count from aerial photographs and covariate x = visually estimated flock size for a sample of $n = 45$ flocks of snow geese.

The original fitted regression model is

$$\hat{y}_i = \underset{(8.61)}{26.65} + \underset{(0.08)}{0.883}x_i$$

with standard errors of the coefficients enclosed in parentheses. In view of the heteroscedasticity evident in the data, Wei and Hickernell (1996) proposed the following covariate transformation model

$$y_i = \delta + \left(\frac{x_i^\lambda - 1}{\lambda}\right)\xi + \varepsilon_i.$$

Parameter estimates for δ and ξ are -35.759 (10.83) and 8.604 (0.63) respectively, and $\hat{\lambda} = 0.538$.

Figure 1 shows the (re-scaled) partial influence measure d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$, based on case deletions. Case 29 is the most influential observation, which is consistent with the index plot of profile likelihood displacement LD_i (Wei and Hickernell (Fig. 2)). It affects the estimate of λ significantly, $\hat{\lambda}_{[29]} = 1.38$. Indeed, case 29 is a leverage point recording the highest observer count of 500 birds.

We next examine the local effect of each case on $\hat{\lambda}$. Based on the partial influence approach, the direction cosines $\boldsymbol{l}_{\text{max}}^d$ from perturbing case weight, response, and transformed covariate are plotted against case index in Fig. 2. It is evident that the greatest local change in $\hat{\lambda}$ depends essentially on case 29. This result is consistent with the $\boldsymbol{l}_{\text{slope}}^{\hat{\lambda}}$ vectors displayed in Fig. 3 under case weight and response perturbation schemes. However, no cosine in $\boldsymbol{l}_{\text{slope}}^{\hat{\lambda}}$ appears to be outlying with respect to perturbations of the transformed covariate. Since λ is a scalar parameter, $\boldsymbol{l}_{\text{max}}^{\hat{\lambda}} \propto \boldsymbol{l}_{\text{slope}}^{\hat{\lambda}}$. It is interesting to note that Wei and Hickernell had

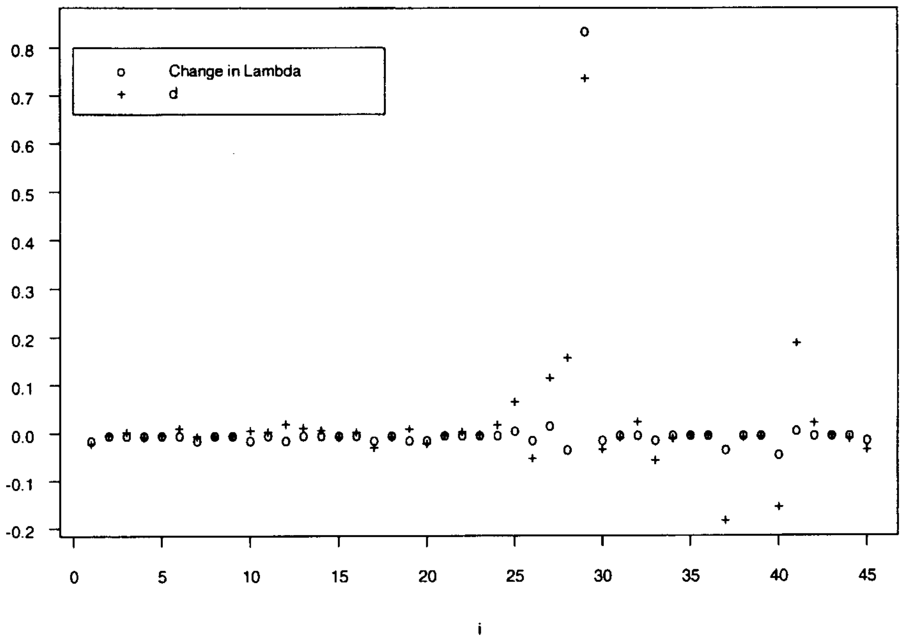

Fig. 1. Case deletion diagnostics (rescaled) d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for snow geese data.

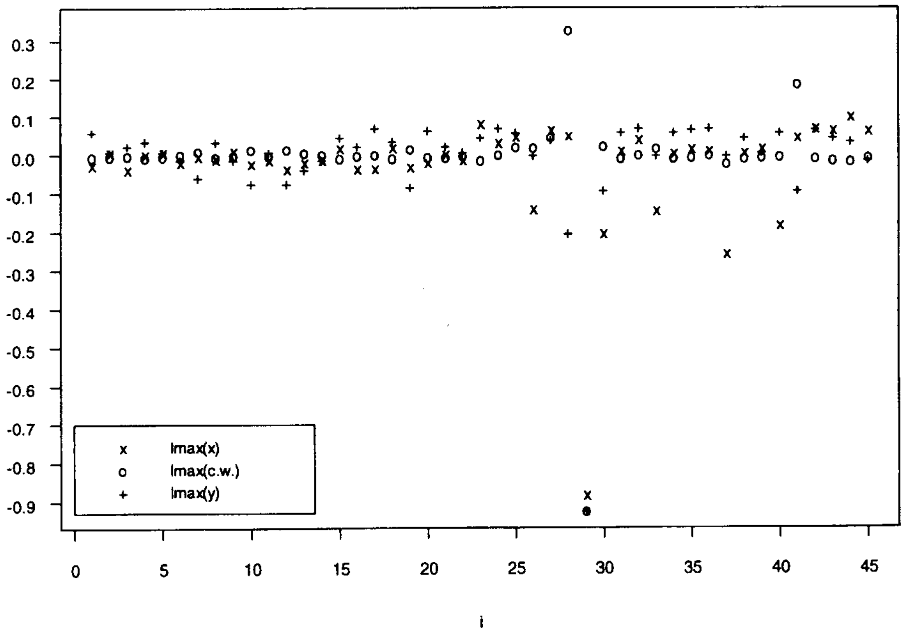

Fig. 2. Direction cosines $l^d_{\max}$ from local perturbations for snow geese data.

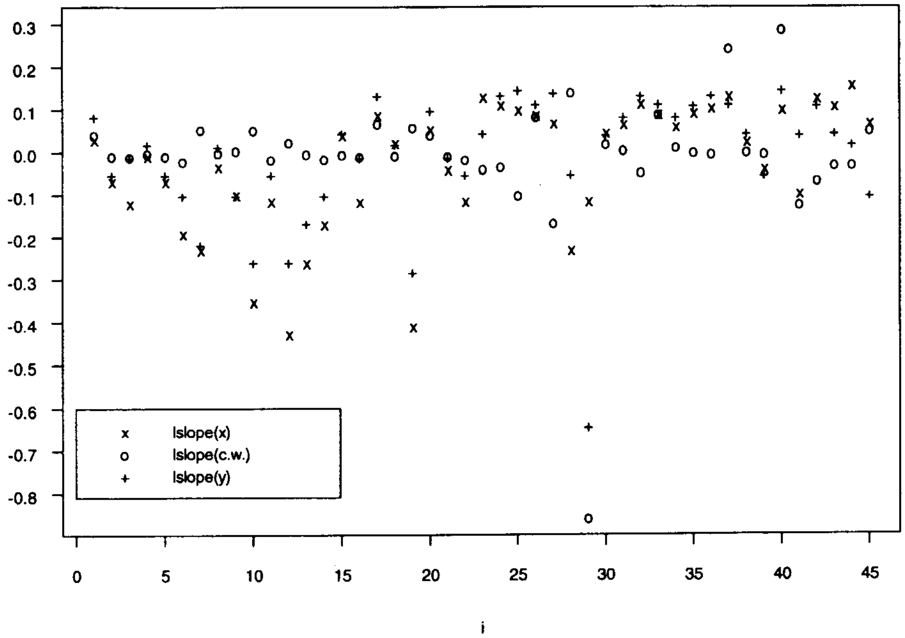

Fig. 3. Direction cosines $l^{\hat{\lambda}}_{\text{slope}}$ from local perturbations for snow geese data.

to resort to an alternative perturbation scheme (proportional instead of additive) before case 29 becomes discordant.

To confirm the indications of the proposed local influence diagnostics, we plot $\hat{\lambda}_\omega$ against the perturbation scale for each maximizing local direction $l^d_{\max}$ in Fig. 4. The effects of downweighting case 29 (in direction $l_{[29]}$) are almost the same as those of simultaneously perturbing all case weights, except when the perturbation scale approaches 1, where $\hat{\lambda}$ increases rapidly as the contribution of case 29 is downweighted to zero. It is worth noting that while the curve associated with response perturbations has the greatest slope at the null state of no perturbation, the analysis is quite insensitive to minor modifications in the covariate.

4.2 *ESR data*

We next illustrate the proposed diagnostics with data from Collett ((1991), p. 8) relating the chronic disease state y_i (0 = healthy; 1 = unhealthy) of 32 individuals, judged from the erythrocyte sedimentation rate (ESR) reading, to the plasma fibrinogen level x_i (in gm/ℓ). The fitted logistic regression model is

$$\begin{aligned}\text{logit}(\hat{\mu}_i) &= -6.845 + 1.827x_i \\ &\quad\;(2.764)\quad(0.899)\end{aligned}$$

with deviance 24.84 on 30 d.f. A constructed variable plot suggests cases 15 and 23 are outliers and that a non-linear transformation of x_i is required; see Collett ((1991), p. 167). Collett then proceeded to include a quadratic term in the model.

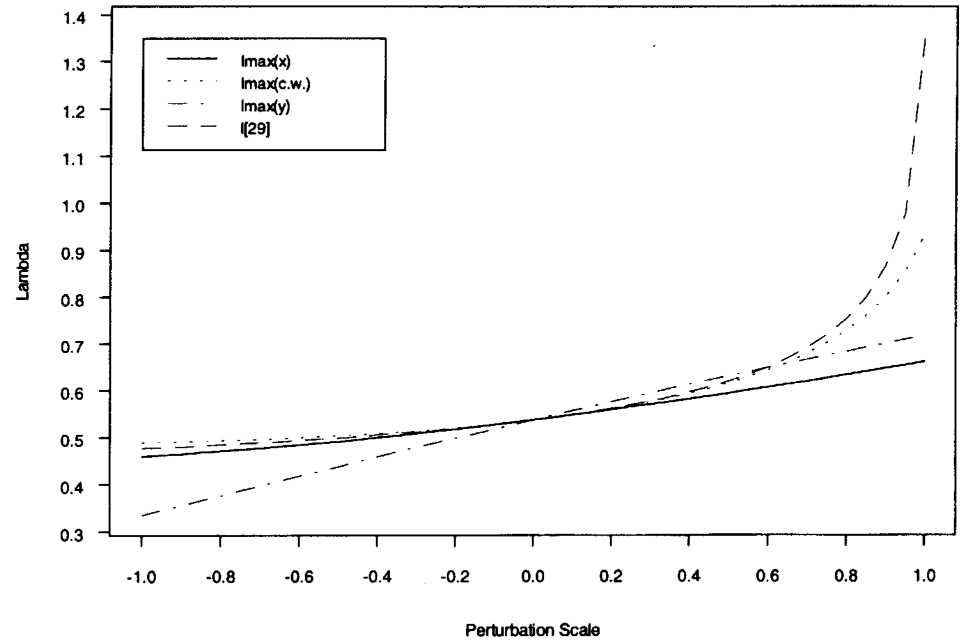

Fig. 4. $\hat{\lambda}$ in directions of local influence for snow geese data.

Alternatively, we consider fitting the Box-Cox transformation model:

$$\text{logit}(\hat{\mu}_i) = \delta_0 + x_i\delta_1 + \left(\frac{x_i^\lambda - 1}{\lambda}\right)\xi.$$

The resulting reduction in deviance, 8.06, is significant at the 5% level. The MLE for λ is 6.81, while parameter estimates for δ_0, δ_1 and ξ are 20.731 (10.094), -10.425 (5.095), 0.023 (0.013), respectively. Case deletion diagnostics (re-scaled) LD_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ displayed in Fig. 5 identify case 15 only, but large values of d_i are found for case 15 and to a certain extent, case 23.

The direction cosines $l^{\hat{\lambda}}_{\text{slope}}$ and $l^{d}_{\max}$ from minor perturbations are plotted against case index in Fig. 6. Results based on the first/second order approach are consistent with those of the partial influence approach. Upon perturbing the transformed covariate, case 15, followed by case 23, have components that are separated from those of the other individuals. An inspection of the data reveals that these two observations correspond to unhealthy patients with unusually low fibrinogen counts, despite the plasma protein concentration tends to increase under inflammatory disease conditions. Under perturbation of case weight, cases 5 and 14 emerge as influential on the transformation. We note that case 5 recorded the highest fibrinogen level compared to other healthy individuals in the sample. Meanwhile, case 14 has near average fibrinogen level among the unhealthy group, yet its standardized deviance residual is the second largest (after case 15) on fitting a quadratic logistic regression (Collett (1991), p. 168).

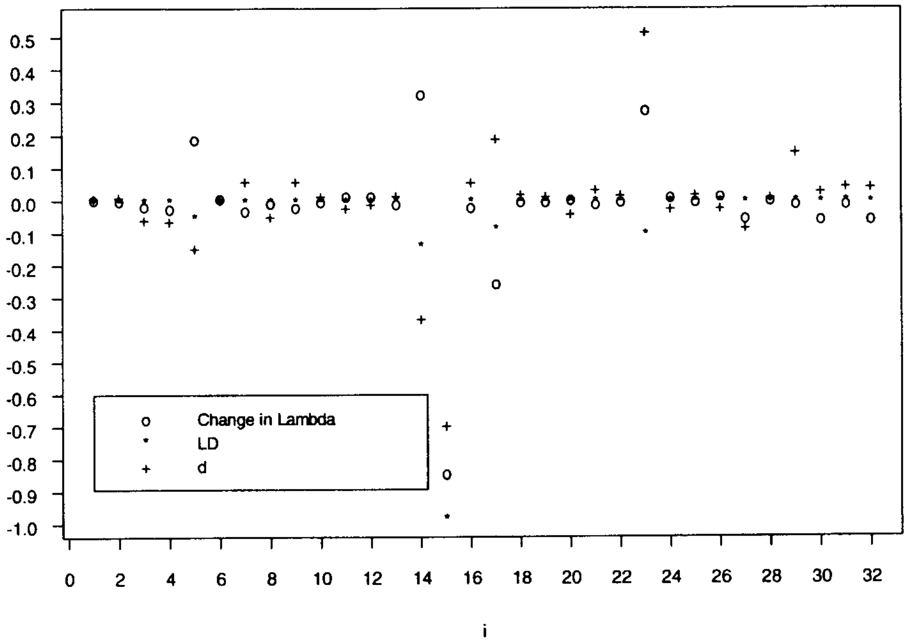

Fig. 5. Case deletion diagnostics (rescaled) LD_i, d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for ESR data.

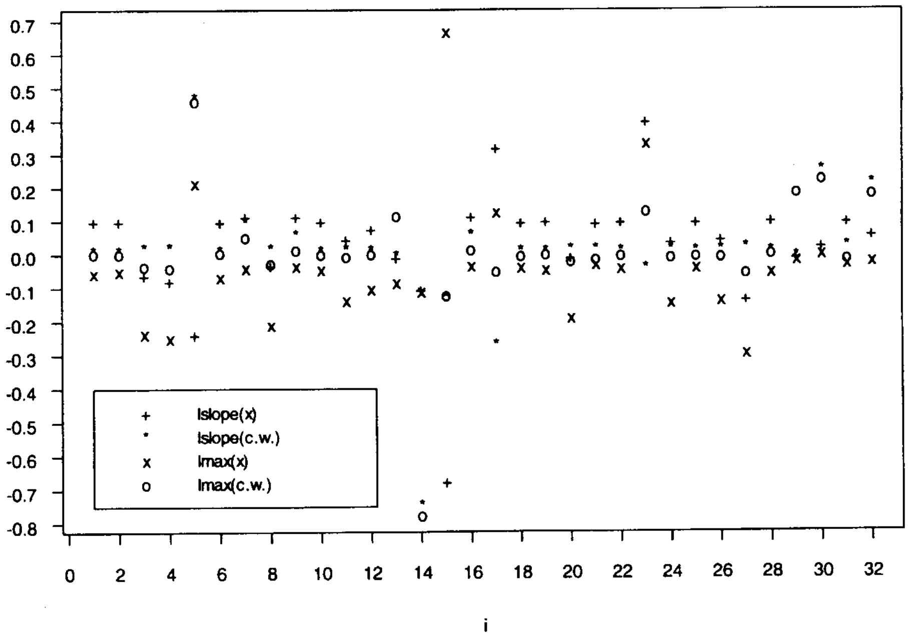

Fig. 6. Direction cosines $l^{\hat{\lambda}}_{\mathrm{slope}}$ and $l^{d}_{\max}$ from local perturbations for ESR data.

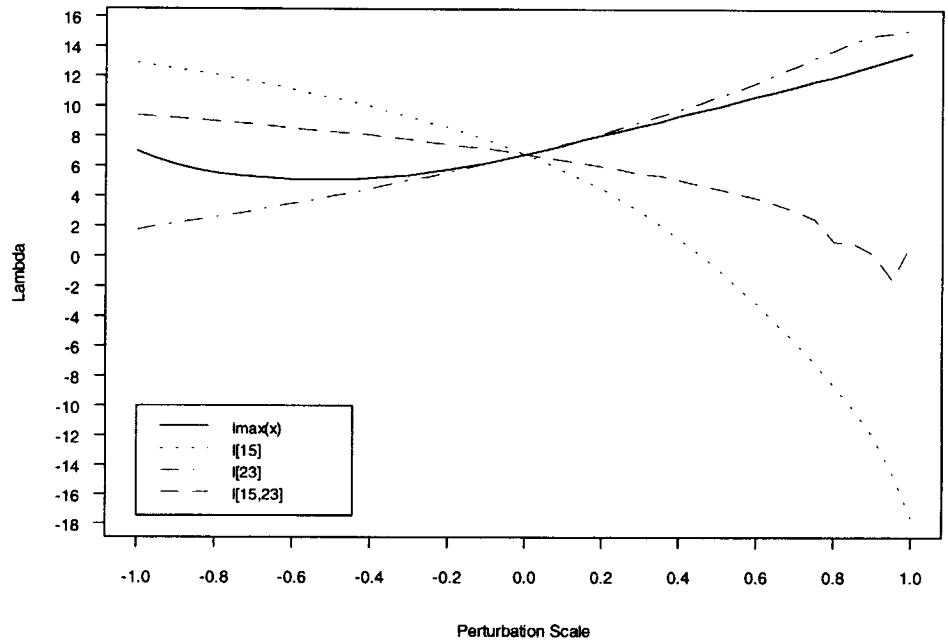

Fig. 7. $\hat{\lambda}$ in local influence directions associated with transformed covariate perturbations for ESR data.

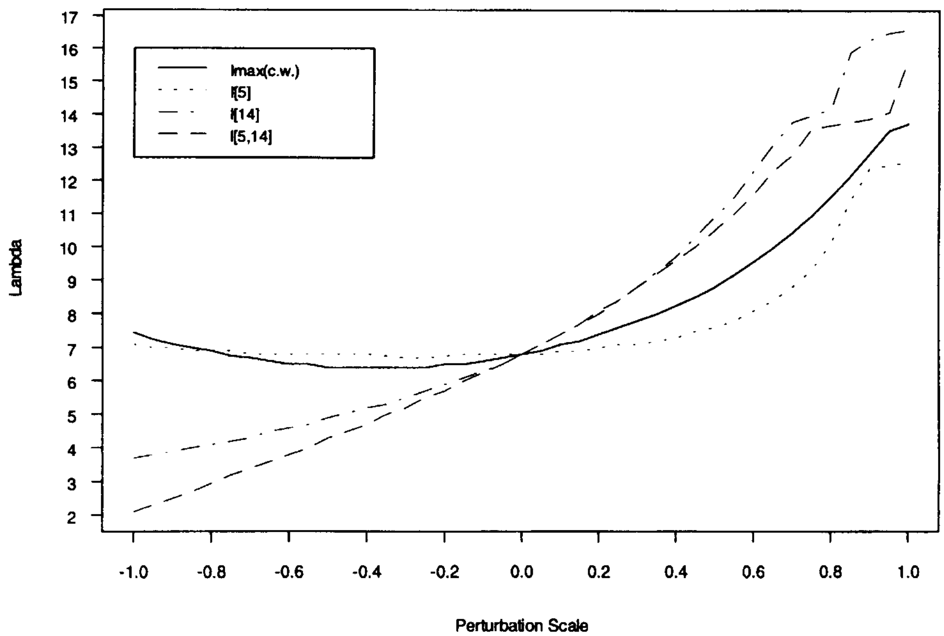

Fig. 8. $\hat{\lambda}$ in local influence directions associated with case weight perturbations for ESR data.

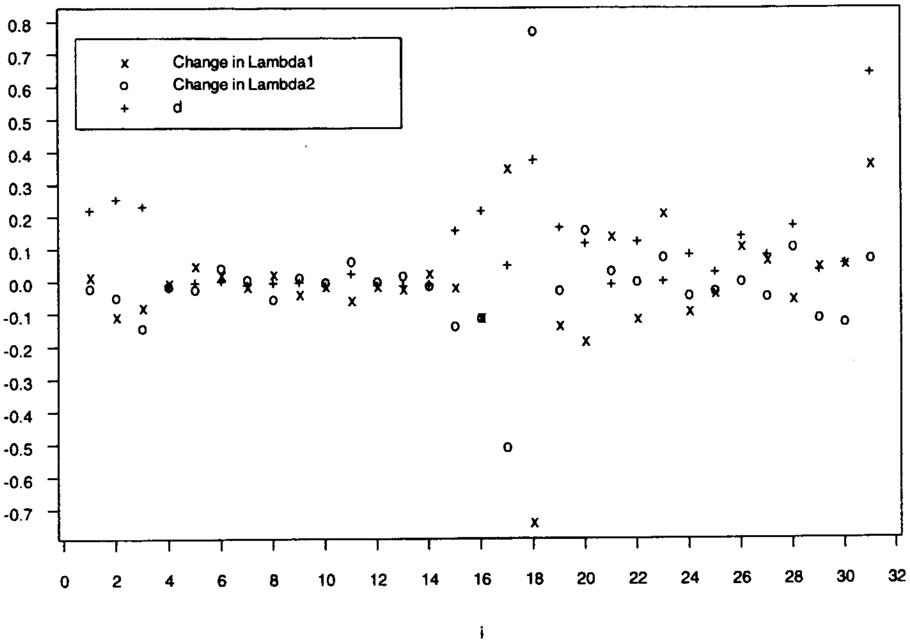

Fig. 9. Case deletion diagnostics (rescaled) d_i, $\hat{\lambda}_{[i]1} - \hat{\lambda}_1$ and $\hat{\lambda}_{[i]2} - \hat{\lambda}_2$ for tree data.

To further assess the extent of the perturbation effects, we plot the actual $\hat{\lambda}$ in Figs. 7 and 8 for selected local directions $\boldsymbol{l}$ of interest, including those related to the deletion of cases. The curve $\boldsymbol{l}_{[15]}$ associated with the downweighting of case 15 alone has the greatest slope at the state of no perturbation, in addition to producing the maximum global change in $\hat{\lambda}$. We also found that as the contributions of cases 15 and 23 are being reduced to zero (direction $\boldsymbol{l}_{[15,23]}$), $\hat{\lambda}$ approaches 1, representing no transformation. Therefore, once these two observations are removed, there is no evidence for covariate transformation. The net overall change due to covariate perturbations is not dramatic since the impact exerted by case 15 apparently has been compensated by the other cases. Besides, perturbations in this direction give similar effects as those of simultaneously modifying all case weights. Figure 8 also confirms that sensitivity of the transformation parameter depends considerably on the weights attached to cases 5 and 14.

4.3 *Tree data*

To provide a numerical illustration of the diagnostics when $\boldsymbol{\lambda}$ is a vector quantity of interest, we consider the tree data from Ryan *et al.* ((1976), p. 278). The data consist of measurements on tree volume y (in ft^3), tree height x_2 (in ft), and tree diameter x_1 (in *inches*) at $4.5ft$ above ground level for a sample of $n = 31$ black cherry trees. The following covariate transformation model,

$$y_i = \delta + \left(\frac{x_{i1}^{\lambda_1} - 1}{\lambda_1}\right)\xi_1 + \left(\frac{x_{i2}^{\lambda_2} - 1}{\lambda_2}\right)\xi_2 + \varepsilon_i,$$

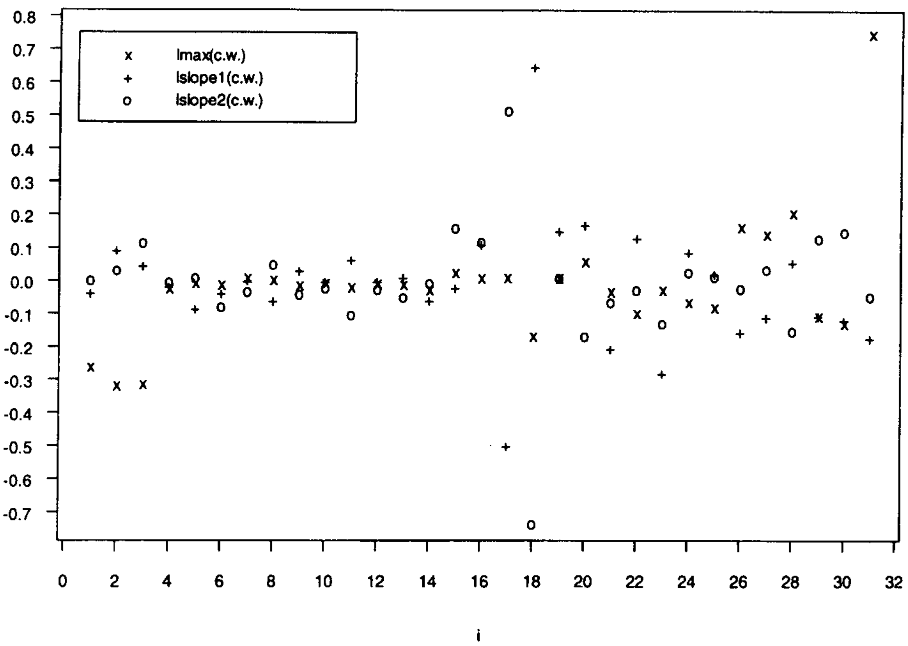

Fig. 10. Direction cosines $l_{\text{slope}}^{\hat\lambda_1}$, $l_{\text{slope}}^{\hat\lambda_2}$ and $l_{\max}^{d}$ from case weight perturbations for tree data.

suggested by Wei and Hickernell (1996), will be adopted in our analysis.

The deviance of the fitted model is 188.247 with $\hat{\boldsymbol\lambda} = (2.583, 1.738)^T$. Figure 9 gives (re-scaled) $\hat\lambda_{[i]1} - \hat\lambda_1$, $\hat\lambda_{[i]2} - \hat\lambda_2$, and d_i. In addition to having the most extreme LD_i values (see Wei and Hickernell (Fig. 8)), cases 17 and 18 induce substantial changes in the transformation parameter estimates upon their deletion, $\hat{\boldsymbol\lambda}_{[17]} = (2.71, -1.137)^T$, $\hat{\boldsymbol\lambda}_{[18]} = (2.325, 6.25)^T$. Meanwhile, the partial influence measure shows that the evidence for covariate transformation depends mainly on case 31.

The local influence diagnostics under case weight perturbations are plotted in Fig. 10. Cases 17 and 18 are clearly influential according to the first order diagnostics $l_{\text{slope}}^{\hat\lambda_1}$ and $l_{\text{slope}}^{\hat\lambda_2}$, whereas case 31 is also influential due to its large component of $l_{\max}^{d}$. These direction cosines provide different diagnostic information to that of the $l_{\max}^{\hat{\boldsymbol\lambda}}$ vector (Wei and Hickernell (Fig. 9)). On the other hand, case 31 emerges as the only influential observation under transformed covariate perturbations. It should be remarked that while case 31 corresponds to the largest tree in the sample, cases 17 and 18 are medium sized trees but with distinctive height of 85 ft and 86 ft respectively.

Acknowledgements

We are grateful to an associate editor and referees for suggestions that improved the presentation of this work. Thanks are also due to Fred Hickernell and Shih Jian-Qing for helpful discussions.

References

Atkinson, A. C. (1986). Diagnostic tests for transformations, *Technometrics*, **28**, 29–38.

Atkinson, A. C. (1988). Transformation unmasked, *Technometrics*, **30**, 311–318.

Box, G. E. P. and Tidwell, P. W. (1962). Transformation of the independent variables, *Technometrics*, **4**, 531–550.

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

Collett, D. (1991). *Modelling Binary Data*, Chapman and Hall, New York.

Cook, R. D. (1986). Assessment of local influence (with discussion), *J. Roy. Statist. Soc. Ser. B*, **48**, 133–169.

Cook, R. D. (1987). Influence assessment, *J. Appl. Statist.*, **14**, 117–131.

Cook, R. D. and Wang, P. C. (1983). Transformations and influential cases in regression, *Technometrics*, **25**, 337–343.

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

Ezekiel, M. and Fox, F. A. (1959). *Methods of Correlation and Regression Analysis*, Wiley, New York.

Hinkley, D. V. and Wang, S. (1988). More about transformations and influential cases in regression, *Technometrics*, **30**, 435–440.

Lawrance, A. J. (1988). Regression transformation diagnostics using local influence, *J. Amer. Statist. Assoc.*, **83**, 1067–1072.

Lee, A. H. (1988). Assessing partial influence in generalized linear models, *Biometrics*, **44**, 71–77.

Ryan, T. A., Joiner, B. L. and Ryan, B. F. (1976). *Minitab Student Handbook*, Duxbury Press, Boston.

Shih, J. Q. (1993). Regression transformation diagnostics in transform-both-sides model, *Statist. Probab. Lett.*, **16**, 411–420.

Thomas, W. and Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models, *Biometrica*, **76**, 741–749.

Tsai, C.-L. and Wu, X. (1990). Diagnostics in transformation and weighted regression, *Technometrics*, **32**, 315–322.

Tsai, C.-L. and Wu, X. (1992). Transformation-model diagnostics, *Technometrics*, **34**, 197–202.

Wang, P. C. (1987). Simultaneous transformations and influence in regression, *Comm. Statist. Theory Methods*, **16**, 3417–3425.

Wei, B. C. and Hickernell, F. J. (1996). Regression transformation diagnostics for explanatory variables, *Statist. Sinica*, **6**, 433–454.

Wei, B. C. and Shih, J. Q. (1994*a*). On statistical models for regression diagnostics, *Ann. Inst. Statist. Math.*, **46**, 267–278.

Wei, B. C. and Shih, J. Q. (1994*b*). Local influence analysis for regression transformation models, *Acta Math. Appl. Sinica*, **17**, 132–143.

Weisberg, S. (1985). *Applied Linear Regression*, 2nd ed., Wiley, New York.

Wu, X. and Luo, Z. (1993*a*). Second-order approach to local influence, *J. Roy. Statist. Soc. Ser. B*, **55**, 929–936.

Wu, X. and Luo, Z. (1993*b*). Residual sum of squares and multiple potential, diagnostics by a second order local approach, *Statist. Probab. Lett.*, **16**, 289–296.