In…nitesimal In‡uence Analysis

Eugene Demidenko
Dartmouth College
e-mail: eugene.demidenko@dartmouth.edu
Submitted to Statistical Science

November 19, 1998

1. Introduction

In‡uence analysis and in‡uential observations diagnostics is a necessary part of full statistical analysis. Statistical analyst must be sure that there are no in‡uential observations, irrelevant to the study which a¤ect the major statistical inference. Numerous number of articles and several books address the problem of in‡uential observations and outliers. Interested readers are referred to books by Belsley et al. (1980), Cook and Weisberg (1982), and Neter et al. (1992), and papers by Pregibon (1981), Chatterjee and Hadi (1986), and Cook (1986). Mainly, two types of in‡uence analyses for linear model are developed. In the …rst one the calculation of leverage and standardized residuals play central role. The second one is based on in‡uence of case deletion and is called 'case deletion diagnostics'.

In this book in‡uence is understood as the sensitivity of a statistic upon small perturbation in data or model, so that in‡uence analysis might be called sensitivity analysis. In fact, we make use of an old de…nition of in‡uence as formulated by Cook and Weisberg (1982, p.101): "The basic idea in in‡uence analysis is quite simple. We introduce small perturbation in the problem formulation, and monitor how the perturbation changes the outcome of the analysis". In order to distinguish this type of in‡uence analysis we call it in…nitesimal (in…nitely small). Two types of in…nitesimal in‡uence analysis can be considered: (i) in‡uence of small change in individual observation, (ii) in‡uence of model assumptions. The …rst analysis is referred as in…nitesimal data in‡uence and the second as in…nitesimal model in‡uence. Since we consider small changes, the use of derivative plays the central role in measuring the infinitesimal influence. The idea to measure the sensitivity of a statistic via infinitesimal approach based on derivative is not new. Probably, Hodges and Moore (1972) were the first who suggested to measure the influence of individual observation of explanatory variables on the Ordinary Least Squares (OLS) estimate via derivative (following our definition this is an example of infinitesimal data influence). Cook (1986) suggested to measure model sensitivity as the maximum curvature of the log-likelihood displacement, and called it 'local influence' (in our definition this is infinitesimal model influence). In the present paper we combine the two approaches under the umbrella of infinitesimal influence analysis and show how it works for many statistics and characteristics in variety of statistical settings. For instance, as it is shown in Section 4.2, the Cooks's and the Hodges and Moore approaches are equivalent for linear regression model.

The infinitesimal approach is especially useful for complex statistical model while standard influence measures are designated primarily for linear regression model. Infinitesimal influence analysis has several advantages: (i) it has an intuitively appealing interpretation – partial derivative indicates what is affected by what and with what magnitude, (ii) it is easy to compute, the analysis is based on the current estimate, (iii) it allows broad graphical support which is very useful for diagnostics, (iv) it is a general tool and can be applied to literally any statistic, as influence of data perturbation, or statistical model, as influence of model perturbation.

The goal of this paper is to introduce Infinitesimal influence analysis and illustrate it by several examples applied to linear, nonlinear and logistic regression models.

2. Influence analysis for linear regression

Influence analysis for linear regression model is concentrated around the following two major concepts: (i) leverage, as the diagonal element of the hat matrix, and standardized residuals, (ii) case deletion diagnostics and Cook's distance. The aim of this section is to provide a quick overview of these concepts.

The standard linear regression model is written in the form

$$y_i = \beta' x_i + \varepsilon_i; \qquad i = 1; \ldots; n \tag{2.1}$$

where x_i is the $m \times 1$ vector of explanatory or independent variables (covariates) and β is the $m-$parameter of interest. It is assumed that $\varepsilon_1; \ldots; \varepsilon_n$ are independent identically distributed (iid) with zero mean and constant variance σ^2. The Ordinary Least Squares (OLS) estimator of β is

$$\hat{\beta} = (X'X)^{-1}X'y$$

where the ith row of matrix X is x_i', and it is assumed X to have full rank. The goal of influence analysis is to detect influential observations. Such an observation might be an outlier, a wrongly recorded result of experiment, etc. In any event an influential observation deserves a close look in terms of its correctness and adequateness to the postulated model. Influence analysis provides some measures of detecting influential observations. It is worth to notice that standard regression characteristics like coefficient of determination or t-statistics are indicators of the overall regression quality and are not designated to detect individual observations or part of the data leading to a bad fitting. Standard influential analysis tries to detect influential cases which consist of observation on the dependent and explanatory variables, (y_i, x_i'). Later, we shall learn how to distinguish the influence with respect to dependent and explanatory variable by means of Infinitesimal influence analysis. It is a great deal of how to define influence: what is influenced by what, how to measure influence etc. Moreover, usually it is not difficult to find influential observations after a definition of influence is given.

The first step in finding an influential case is to check the OLS residual, $r_i = y_i - \hat{y}_i$ where $\hat{y}_i$ is the predicted value, $\hat{y}_i = \hat{\beta}'x_i$. However, it is not necessary true that a larger residual points to an influential case. Another important characteristic of influence analysis is leverage, the diagonal element of the hat matrix $H = X(X'X)^{-1}X'$. The diagonal element of this matrix is $p_i = x_i'(X'X)^{-1}x_i$. It is easy to prove that $0 \leq p_i < 1$ and $\sum_{i=1}^{n} p_i = m$. Cases with high leverage value are called influential; later we shall give an interpretation of leverage from infinitesimal point of view. A better indicator of influential case, in terms of their departure from the fitted model, is the Studentized residual, $r_i/(s\sqrt{1-p_i})$, where $s^2 = \sum r_i^2/(n-m)$ is the estimated variance.

Another measure of influence comes from the idea of case deletion. Indeed, a case might be called influential if the OLS estimate changes significantly after deletion this case from the data set and recomputing the estimate. Fortunately, it is not necessary to recalculate regression after case deletion. Muller (1974) proved that if $\hat{\beta}_{(i)}$ denotes the OLS estimate after the ith case deleted, then

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{r_i}{1-p_i}(X'X)^{-1}x_i. \qquad (2.2)$$

Proof. The following formula is used

$$(A - bb')^{-1} = A^{-1} + \frac{A^{-1}bb'A^{-1}}{1 - b'A^{-1}b}$$

where A is a positive definite matrix and b is a vector-column of the appropriate length. Using this formula one obtains

$$\begin{aligned}
&(X'_{(i)}X_{(i)})^{-1}x_i \\
&= (X'X)^{-1}x_i + (X'X)^{-1}\frac{x_i x_i'}{1 - x_i'(X'X)^{-1}x_i}(X'X)^{-1}x_i \\
&= \left(1 + \frac{x_i'(X'X)^{-1}x_i}{1 - x_i'(X'X)^{-1}x_i}\right)(X'X)^{-1}x_i = \frac{1}{1-p_i}(X'X)^{-1}x_i. \qquad (2.3)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat\beta_{(i)} &= (X'_{(i)}X_{(i)})^{-1}X'_{(i)}y_{(i)} \\
&= \left((X'X)^{-1} + (X'X)^{-1}\frac{x_i x_i'}{1 - x_i'(X'X)^{-1}x_i}(X'X)^{-1}\right)(X'y - x_i y_i) \\
&= \hat\beta + (X'X)^{-1}\frac{x_i x_i'}{1 - x_i'(X'X)^{-1}x_i}\hat\beta - (X'X)^{-1}x_i y_i \\
&\quad - (X'X)^{-1}\frac{x_i x_i'}{1 - x_i'(X'X)^{-1}x_i}(X'X)^{-1}x_i y_i = \hat\beta - \frac{r_i}{1-p_i}(X'X)^{-1}x_i,
\end{aligned}$$

which implies (2.2). ■

An alternative way to assess the influence of the ith case is to calculate the difference in predicted values. More precisely, if $\hat y_{(i)}$ is the $n \times 1$ vector of the predicted value after the ith case deleted we calculate

$$D_i = \frac{\| \hat y - \hat y_{(i)} \|^2}{ms^2},$$

which is called Cook's distance. Implying formula (2.2) it is possible to express D_i in terms of the original data:

$$D_i = \frac{r_i^2}{ms^2}\frac{p_i}{(1-p_i)^2}. \qquad (2.4)$$

There are some others measures of influence for linear regression, the interested reader is referred to the literature cited above.

3. The idea of infinitesimal influence

As was mentioned in the Introduction, we distinguish two types of influence: influence of an individual observation and model influence viewed as influence of underlying assumptions. This distinction is not rigid and sometimes we consider model perturbation to measure data influence, e.g. missclassification in binary data (sections 4.3 and 6.1).

3.1. Data influence

Let D be the data vector consisted of individual observations on the dependent and independent variables and $t = t(D)$ be any statistic or characteristic of interest. For instance, t might be an estimate itself, the vector of predicted values, test statistic, etc. Infinitesimal influence analysis suggests a measure of how slight perturbation in observation affects the statistic. Thus, if D_i is the ith element of the data, the change ΔD leads to the change in the statistic as $t(D + \Delta D e_i) - t(D)$ where e_i is the Kronecker vector, i.e., consisted of zero except the ith element which is one. Then, the infinitesimal change in the statistic, due to the data perturbation, can be defined as

$$\lim_{\Delta D \to 0} \frac{t(D + \Delta D e_i) - t(D)}{\Delta D}. \tag{3.1}$$

Observations with high values of limit (3.1) are called influential. Probably the reader had already realized that quantity (3.1) is just the partial derivative of t with respect to D_i; namely

$$\frac{\partial t(D)}{\partial D_i}. \tag{3.2}$$

The influence analysis based on the derivative (3.2) is referred as infinitesimal data influence analysis. The kth element of vector (3.2) indicates how small perturbation in the ith element of data affects the kth element of vector statistic. When statistic t does not admit a closed form solution, (3.2) is computed as the derivative of implicit function.

3.2. Model influence

The idea to nest the postulated model into a more general one and then apply infinitesimal influence analysis belongs to Cook (1986). However, our influence analysis seems to be more explicit and intuitively appealing.

Let $L(\theta)$ be the log-likelihood of the postulated model, subject to influence analysis. We nest our model into a more general model dependent on the additional parameter ω; we shall call it the 'parent-model'. The log-likelihood of the parent-model is denoted as $L(\theta \mid \omega)$. Mathematically, the nesting property, without loss of generality, can be written as $L(\theta \mid \omega = 0) = L(\theta)$; which means that the postulated model is just a specific case of the parent-model when $\omega = 0$. Let $t(\omega)$ be any statistic or characteristic of interest as a function of ω, e.g. the Maximum Likelihood Estimate (MLE) which maximizes the log-likelihood $L(\theta \mid \omega)$ assuming ω known. Then, the influence of t with respect to possible departure from the postulated model is measured as

$$\left.\frac{\partial t}{\partial \omega}\right|_{\omega=0}. \tag{3.3}$$

Influence analysis based on the derivative (3.3) is called infinitesimal model influence analysis. It is worth to mention that we do not need to reestimate the postulated model, all calculations of (3.3) are accomplished at the current estimate.

The difference with the Cook's local influence is that he took the likelihood displacement, difference of log-likelihood function, as the measure of the model departure, and the infinitesimal change has been measured as the maximum curvature of the likelihood displacement at $\omega = 0$. Our idea is more explicit because we measure influence of the model change on statistic directly as (3.3). Apparently, (3.3) has a clearer interpretation and easier to calculate. For instance, as it will be shown in section 4.2, the influence of individual observation of explanatory variable on the OLS coefficient in linear model based on the Cook's local influence approach is equivalent to the infinitesimal influence based on the partial derivative as obtained by Hodges and Moore (1972).

A few comments on how to calculate (3.2) or (3.3) when t does not admit a closed form solution. We illustrate how to calculate (3.3) when t is the MLE. Then, by the definition, t is the solution to the score equation $\partial L(\theta \mid \omega)/\partial\theta = 0$ and the needed derivative, as follows from the formula for the derivative of the implicit function, is given as

$$\left.\frac{\partial t}{\partial \omega}\right|_{\omega=0} = \left(- \left.\frac{\partial^2 L(t \mid \omega)}{\partial t^2}\right|_{\omega=0} \right)^{-1} \left(\left.\frac{\partial^2 L(t \mid \omega)}{\partial t \partial \omega}\right|_{\omega=0} \right). \tag{3.4}$$

It is interesting to note that the inversed matrix is an approximation to the asymptotic covariance of the MLE, so that (3.4) can be rewritten as

$$\left.\frac{\partial t}{\partial \omega}\right|_{\omega=0} = \mathrm{cov}(t)\left(\left.\frac{\partial^2 L(t \mid \omega)}{\partial t\,\partial \omega}\right|_{\omega=0}\right).$$

Below we show how the idea to use derivative as the measure of infinitesimal influence works for many statistics in many statistical models and settings. Further, the infinitesimal influence will be shortly called as I-influence.

4. Linear regression model

I-influence analysis for linear regression model (2.1) is quite developed although never has been considered on a systematic basis. The aim of this section is to revive relevant formulae and illustrate the approach graphically on several examples. Infinitesimal influence analyses for nonlinear regression and binary data are developed in sections 4 and 5.

In this section standard linear regression model in the form (2.1) is considered. To be specific, index i is referred to the ith individual. Since $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ and $\partial \hat{y}_i / \partial y_i = p_i$ we infer that, from the I-influence point of view, leverage measures the influence of observation y_i on the predicted value, $\hat{y}_i$. Although leverage is an important characteristic of influence analysis one should remember that it measures the influence of individual observation of the dependent variable on the predicted value for the same individual. However, we argue that the most important characteristic is the estimate itself rather than the predicted value, unless the only purpose of the model is to predict the ith individual. Therefore, the question we pose is: how the OLS estimate is affected by an individual observation of the dependent or explanatory variable? Thus, unlike traditional influence analysis we distinguish the influence of the dependent and independent variable. The influence with respect to the dependent variable is called $Y-$influence and with respect to the explanatory variable as $X-$influence. These types of influence reveal what is the cause of influence – the dependent or independent variable and which independent variable is most influential in the multivariate regression model setting. The difference between these two types of influences is shown in Figure 1. In $Y-$influence we change dependent variable and seek how regression line changes, particularly the slope does not change much and all changes are in the intercept term (the empty circled observation is changed to the filled one). On contrary, in $X-$influence we seek how a change in explanatory variable changes the regression line. In particular, if the outlier moves parallel to $x-$axe from empty circled to filled the slope changes visibly. Thus, we conclude that for this example Y-influence is less significant than X-influence.

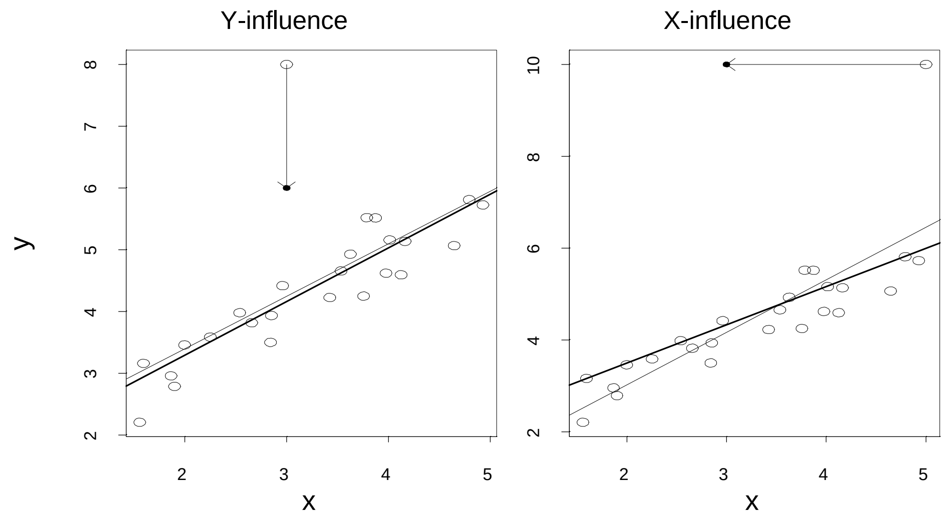

Figure 1. Y and X-influences. Solid – original regression line, bold - perturbated regression line after an individual observation changed. Left: Y-influence, the dependent variable changes. Right: X-influence, the independent variable changes. The same magnitude of perturbation in y and x leads to different changes in regression slope. Apparently, for this example the slope is more influential to explanatory variable.

4.1. Influence of dependent variable on the OLS estimate

The infinitesimal influence of the ith observation of the dependent variable on the OLS estimate is measured as the partial derivative,

$$\frac{\partial \hat{\mathbf{b}}}{\partial y_i} = (X'X)^{-1}x_i; \qquad i = 1; ...; n; \tag{4.1}$$

the m-dimensional vector. The kth element of vector (4.1) is interpreted as the rate of change in the estimate of the kth regression coefficient upon small change in y_i. High absolute values of the derivative are associated with influential observations of y: The vector $(X'X)^{-1}x_i$; as we shall learn later, is an important part of the influence analysis. It may be large when two conditions hold: (i) the length of x_i is relatively large, i.e. x_i lies outside the bulk of data $\{x_i\}$; (ii) x_i lies in the direction where the scatter plot is squeezed rather than stretched out. Thus, (4.1) reflects outstanding observations x_i and multicollinearity as well (Cook and Weisberg 1982, p. 13).

4.2. Influence of explanatory variable on the OLS estimate

We revive the idea of measuring the influence of individual observation of explanatory variable x_{ik} on the OLS estimate via derivative, as suggested by Hodges and Moore (1972). Thus, the question we pose is: how the OLS estimate is affected by small perturbation in the ith observation of the kth covariate, x_{ik}? To make the infinitesimal approach work we assume here that x_{ik} is continuous (in the next subsection an approach will be developed for binary explanatory variable). Following the line of the I-influence the sensitivity of $\hat{\beta}$ to small change in x_{ik} is measured via partial derivative

$$\frac{\partial \hat{\beta}}{\partial x_{ik}}; \qquad 1 \le i \le n;\ 1 \le k \le m. \tag{4.2}$$

Thus, the jth element of this vector is interpreted as the rate of departure from the jth OLS coefficient under small perturbation in the ith case of the kth covariate. To derive the expression for (4.2) we write

$$\begin{aligned}
&\frac{\partial}{\partial x_{ik}}\left(\sum x_i x_i'\right)^{-1}\left(\sum x_i y_i\right) \\
&= -(X'X)^{-1}\left(\frac{\partial}{\partial x_{ik}}\sum x_i x_i'\right)\hat{\beta} + (X'X)^{-1}\left(\frac{\partial}{\partial x_{ik}}\sum x_i y_i\right) \\
&= -(X'X)^{-1}(x_i e_k' + e_k x_i')\hat{\beta} + (X'X)^{-1} e_k y_i \\
&= (X'X)^{-1}\left[e_k(y_i - x_i'\hat{\beta}) - x_i(e_k'\hat{\beta})\right] = (X'X)^{-1}(e_k r_i - x_i\hat{\beta}_k);
\end{aligned}$$

where r_i is the ith OLS residual, $y_i - \hat{\beta}' x_i$; and $\hat{\beta}_k$ is the kth OLS coefficient, e_k is the $m \times 1$ Kronecker vector, i.e. consists of zeroes except the kth element which is one. Thus, the influence of observation x_{ik} on the OLS estimate can be measured as (Hodges and Moore, 1972)

$$\frac{\partial \hat{\beta}}{\partial x_{ik}} = (X'X)^{-1}(e_k r_i - x_i\hat{\beta}_k); \qquad 1 \le i \le n;\ 1 \le k \le m. \tag{4.3}$$

This influence measure is easy to interpret: one unit change of x_{ik} leads to $\hat{\beta}$ change specified by (4.3). Certainly, the influence analysis based on this formula does not make sense for certain explanatory variables, e.g. sex or intercept term. As we see, perturbation in the kth covariate implies changes in other coefficients unless covariates are orthogonal. The influence of x_{ik} on the estimate, as follows from (4.3), has two components. The first component, $r_i(X'X)^{-1}e_k$ is associated with large residual and the second component, $\hat{\beta}_k(X'X)^{-1}x_i$ is the coefficient times the vector of the $Y-$influence.

It is interesting to note that (4.3) is equivalent to what derived by Cook (1986) using his approach of local influence assuming the distribution of ε is normal. He suggested to measure the influence of x_{ik} on $\hat{\beta}_k$ as $r_i - \hat{\beta}_k q_i$ where q_i is the ith residual in regression of x_k on the rest of explanatory variables. Now we will show that $r_i - \hat{\beta}_k q_i$ is proportional to the kth component of vector (4.3), i.e. I-influence and local influence are equivalent. Without loss of generality we can assume that $k = 1$. Let X be partitioned as $[x_1, X_2]$ so that

$$(X'X)^{-1} = \begin{bmatrix} a & b' \\ b & A \end{bmatrix},$$

where a is a scalar, b is a $(m-1)\times 1$ vector and A is a $(m-1)\times(m-1)$ symmetric matrix. Then

$$a = \frac{1}{x_1'x_1} + \frac{1}{(x_1'x_1)^2}x_1'Mx_1, \qquad X_2b = -\frac{1}{x_1'x_1}Mx_1$$

where

$$M = H + \frac{1}{x_1'x_1 - x_1'Hx_1}Hx_1x_1'H,$$

and $H = X_2(X_2'X_2)^{-1}X_2'$ is the $n\times n$ projection matrix. After some algebra one obtains $a = (x_1'x_1 - x_1'Hx_1)^{-1}$. We find

$$\begin{aligned} x_1'Mx_1 &= x_1'\left(H + \frac{1}{x_1'x_1 - x_1'Hx_1}Hx_1x_1'H\right)x_1 \\ &= x_1'Hx_1 + \frac{(x_1'Hx_1)^2}{x_1'x_1 - x_1'Hx_1} = \frac{x_1'x_1 \cdot x_1'Hx_1}{x_1'x_1 - x_1'Hx_1} \end{aligned}$$

and

$$\begin{aligned} Mx_1 &= \left(H + \frac{1}{x_1'x_1 - x_1'Hx_1}Hx_1x_1'H\right)x_1 \\ &= Hx + \frac{x_1'Hx_1}{x_1'x_1 - x_1'Hx_1}Hx_1 = \left(\frac{x_1'x_1}{x_1'x_1 - x_1'Hx_1}\right)Hx_1. \end{aligned}$$

Then, the first column of matrix $X(X'X)^{-1}$ is

$$
\begin{aligned}
& a x_1 + X_2 b \\
= & \left(\frac{1}{x_1'x_1} I + \frac{1}{(x_1'x_1)^2} \cdot \frac{x_1'x_1 \cdot x_1'Hx_1}{x_1'x_1 - x_1'Hx_1} I - \frac{1}{x_1'x_1} \frac{x_1'x_1}{x_1'x_1 - x_1'Hx_1} H \right) x_1 \\
= & \left(\frac{1}{x_1'x_1} I + \frac{1}{x_1'x_1} \cdot \frac{x_1'Hx_1}{x_1'x_1 - x_1'Hx_1} I - \frac{1}{x_1'x_1 - x_1'Hx_1} H \right) x_1 \\
= & \left(\frac{1}{x_1'x_1 - x_1'Hx_1} I - \frac{1}{x_1'x_1 - x_1'Hx_1} H \right) x_1 = \frac{1}{x_1'x_1 - x_1'Hx_1} (I - H) x_1 \\
= & \; aq;
\end{aligned}
$$

where $q = (I - H)x_1$; the residual vector in regression x_1 on X_2. Then, returning to formula (4.3) in vector form

$$
\frac{\partial \widehat{\beta}_1}{\partial x_1} = (X'X)^{-1}_{11} r - \widehat{\beta}_1 \left(X(X'X)^{-1} \right)_{\cdot 1} = a(r - \widehat{\beta}_1 q);
$$

i.e. $\{r_i - \widehat{\beta}_1 q_i\}$ proportional to $\{\partial \widehat{\beta}_1 / \partial x_{i1}\}$ with the factor $a = (X'X)^{-1}_{11}$. ■

Formula (4.3) is even more general than the Cook's formula because it is distribution free and allows to assess the influence of the kth explanatory variable on the jth OLS-coefficient when $k \neq j$.

4.3. Influence of binary explanatory variable on the OLS estimate

In some cases the explanatory variable is not continuous, e.g. is binary and takes value either 0 or 1. We can still apply the I-influence employing the idea of missclassification. Following our definition of Section 3 this is an example of infinitesimal model influence analysis.

If x_{ik} is binary and observed we can interpret it as an outcome of a classification procedure, without loss of generality we can assume $k = m$. The sensitivity of the OLS coefficient to binary variable is understood as the sensitivity to the probability of missclassification. To set up the model with missclassification we shall assume that the regression model is given as $E(y|z) = \alpha + \beta_m z$ where z is the true unobserved binary explanatory variable missclassified with the probability q; and $\alpha = \beta_1 x_1 + ... + \beta_{m-1} x_{m-1}$. However, we do not observe z but observe x_m such that $\Pr(z = 1|x_m = 0) = \Pr(z = 0|x_m = 1) = q$; the symmetric missclassification is assumed. Then, in terms of observed data the regression conditioned on $x_m = 1$

can be rewritten as

$$\begin{aligned} E(y|x_m &= 1) = E(y|z = 1; x_m = 0)\Pr(z = 1|x_m = 1) \\ +E(y|z &= 0; x_m = 0)\Pr(z = 0|x_m = 1). \end{aligned}$$

Also, we shall assume that the missclassification does not depend on the regression, or more precisely $\Pr(y|x_m; z) = \Pr(y|z)$. Hence, it is easy to see, $E(y|x_m) = \alpha + \beta_m(x_m + (1 - 2x_m)q)$. Therefore, assuming x_{im} binary, in notations of (4.3), the missclassification impact on the OLS estimate can be measured as

$$\left.\frac{\partial \hat{\beta}}{\partial q_i}\right|_{q_i=0} = (1 - 2x_{im})(X'X)^{-1}(e_k r_i - x_i \hat{\beta}_k). \tag{4.4}$$

As we see, this formula resembles (4.3), and they have the same absolute value.

4.4. Influence of explanatory variable on predicted value

As was mentioned above, diagonal element of the hat matrix measures the influence of the dependent variable on its predicted value. Not of less importance is how the predicted value is affected by the explanatory variable. Omitting fairly simple algebra using (4.3) one obtains

$$\frac{\partial \hat{y}_i}{\partial x_{ik}} = \hat{\beta}_k(1 - p_i) + x_i'(X'X)^{-1}e_k r_i \qquad 1 \le i \le n;\ 1 \le k \le m \tag{4.5}$$

where p_i is the leverage.

A few comments on how calculate derivatives when there are replicates, regression is curvilinear or explanatory variables are functionally related. In the case of replicates, i.e. when several observations of the dependent variable y_i are available for the same value of x_{ik}, formula (4.3) is written as

$$\frac{\partial \hat{\beta}}{\partial x_{ik}} = (X'X)^{-1}\Big(e_k \sum_{x=x_{ik}} r_i - x_i \hat{\beta}_k\Big).$$

In the case of curvilinear regression, i.e., when x_i enters regression model as $\tilde{x}_i = g(x_i)$ where $g(\cdot)$ is a known function, formula (4.3) is written as

$$\frac{\partial \hat{\beta}}{\partial x_{ik}} = (\tilde{X}'\tilde{X})^{-1}(e_k r_i g'(x_{ik}) - \tilde{x}_i \hat{\beta}_k), \tag{4.6}$$

where g' denotes the derivative of g, see section 4.8 where $g(s) = \log(s + 1)$. Also, it is easy to obtain the expression for the derivative in the case when some explanatory variables are functionally related, as in quadratic regression.

4.5. Case or group deletion

The theory of case deletion, and particularly infinitesimal deletion based on weights, has been developed by Belsley et al. (1980), Pregibon (1981), Cook and Weisberg (1982). We shortly review it for further extension to nonlinear regression and logistic regression.

We introduce weight w_i for case i and assume all other cases have weight 1. Then, the weighted normal equation for the least squares estimate is written as

$$\sum_{j\neq i}(y_j - \beta' x_j)x_j + w_i(y_i - \beta' x_i)x_i = 0. \tag{4.7}$$

Clearly, the solution to this equation is the weighted LS estimate, $\hat{\beta} = \hat{\beta}(w_i)$. In the infinitesimal deletion approach one assesses how decreasing of w_i from 1 to 0 affects regression coefficient estimate. The rate of this change is measured as the derivative of the estimate with respect to w_i. Two types of infinitesimal deletion may be distinguished according to at what point the derivative is computed: We call I-influence of deletion at inclusion when the according derivative is computed at $w_i = 1$. If the derivative is computed at $w_i = 0$ we call it I-influence of deletion at exclusion. The formula for the derivative at inclusion have been derived by Belsley et al. (1980):

$$\left.\frac{\partial\hat{\beta}}{\partial w_i}\right|_{w_i=1} = r_i(X'X)^{-1}x_i \tag{4.8}$$

where r_i is the OLS residual (it can be derived from (4.7) by differentiating with respect to w_i). Another option is to calculate the derivative at the point where the ith case is excluded, i.e. $w_i = 0$. We will prove that

$$\left.\frac{\partial\hat{\beta}}{\partial w_i}\right|_{w_i=0} = \frac{r_i}{(1-p_i)^2}(X'X)^{-1}x_i. \tag{4.9}$$

Proof. Using (2.3) one obtains

$$\begin{aligned}
\left.\frac{\partial\hat{\beta}}{\partial w_i}\right|_{w_i=0} &= \left.\frac{\partial}{\partial w_i}(X'X - w_i x_i x_i')^{-1}(X'y - w_i x_i y_i)\right|_{w_i=0} \\
&= (X_{(i)}'X_{(i)})^{-1}x_i x_i'(X_{(i)}'X_{(i)})^{-1}X_{(i)}'y_{(i)} - (X_{(i)}'X_{(i)})^{-1}x_i y_i \\
&= \frac{1}{(1-p_i)^2}(X'X)^{-1}x_i x_i'(X'X)^{-1}(X'y - x_i y_i) - \frac{1}{1-p_i}(X'X)^{-1}x_i y_i \\
&= \left(\frac{p_i y_i - \hat{y}_i}{(1-p_i)^2} + \frac{y_i}{1-p_i}\right)(X'X)^{-1}x_i = \frac{r_i}{(1-p_i)^2}(X'X)^{-1}x_i;
\end{aligned}$$

and formula (4.9) is proved. ■

As we see, all three measures (2.2) and (4.8), (4.9) look similar. Moreover, it is easy to show that approximately the latter is the average of the formers. Indeed, half of the sum (4.8) and (4.9), assuming $p_i^2 \simeq 0$, is

$$\begin{aligned} &\frac{1}{2} r_i \left(1 + \frac{1}{(1-p_i)^2}\right)(X'X)^{-1}x_i \\ = \; &\frac{1}{2} r_i \left(\frac{2}{1-p_i} + \frac{p_i^2}{(1-p_i)^2}\right)(X'X)^{-1}x_i \simeq \frac{r_i}{1-p_i}(X'X)^{-1}x_i. \end{aligned}$$

There is an important feature of measure (4.8): it enables to determine the influence of group deletion, which follows from the fact that the derivative of a sum is the sum of the derivatives. In fact, let a group of cases I is chosen, and we want to find how the deletion of this group of cases affects the OLS estimate. Introducing weights $w = w_i; i \in I$ we come to the estimating equation

$$\sum_{i \notin I} (y_i - \beta' x_i) x_i + w \sum_{i \in I} (y_i - \beta' x_i) x_i = 0.$$

Then, the influence of the entire group I deletion on the OLS estimate is measured as $(X'X)^{-1} \sum_{i \in I} r_i x_i$ which is the sum of individual deletion measures. In other words, by identifying a group of observations with large values of (4.8) we actually identify a group of influential cases deleted simultaneously.. It is worthwhile to notice that this additive property does not hold for usual case deletion diagnostics.

Derivatives (4.8) and (4.9) can be used to predict the effect of case deletion on OLS coefficients. Indeed, using the general approximation formula $f(x) - f(x_0) = f'(x_0)(x - x_0)$, we can write for deletion at inclusion

$$\hat{\beta} - \hat{\beta}_{(i)} = (1 - 0) \times \left.\frac{\partial \hat{\beta}(w_i)}{\partial w_i}\right|_{w_i = 1} = r_i (X'X)^{-1} x_i$$

and for deletion at exclusion

$$\hat{\beta}_{(i)} - \hat{\beta} = (0 - 1) \times \left.\frac{\partial \hat{\beta}(w_i)}{\partial w_i}\right|_{w_i = 0} = -\frac{r_i}{(1-p_i)^2}(X'X)^{-1} x_i.$$

These predictions are compared below, on a real life example. We benefit from analysis of case deletion based on I-influence especially in nonlinear models where no closed form solution exists for the estimate upon case deletion.

4.6. Influence on regression characteristics

Sometimes, besides regression coefficients themselves, we are interested in regression characteristics as a part of influence analysis. For instance, in investigating the effect of a new treatment the key characteristic would be the t-statistic of the OLS estimate at the treatment effect variable (1 =new treatment, 0 = old treatment), where y is the treatment outcome and the set of other covariates might include age, gender etc. No influence analysis is available for regression characteristics such as $t-$statistic or coefficient of determination in traditional regression diagnostics. As was mentioned above, one of the advantages of the I-influence is that it can be applied to any statistic, as a function of data. We will illustrate this feature by the I-influence analysis for the coefficient of determination and $t-$statistics. Following the line of previous discussion we distinguish two types of influences: the influence of the dependent ($Y-$influence) and independent ($X-$influence) variables.

4.6.1. Y-influence

Coefficient of determination. We start the analysis with the Residual Sum of Squares (RSS). The influence of the observation of the dependent variable y_i on RSS is measured as the partial derivative of RSS with respect to y_i. Then, if r_i denotes the ith OLS residual, the Y-influence of RSS is measured as

$$\frac{\partial RSS}{\partial y_i} = \frac{\partial}{\partial y_i}\sum_{j=1}(y_j - x_j'\hat{\beta})^2 = 2r_i - 2\sum_{j=1} r_j x_j'(X'X)^{-1}x_i = 2r_i, \tag{4.10}$$

since $\sum r_j x_j = 0$. As we see, the rate of change of RSS with respect to small change of the dependent variable y_i is proportional to the OLS residual. Omitting some algebra we obtain the formula for Y-influence of R^2:

$$\frac{\partial R^2}{\partial y_i} = \frac{2}{\sum (y_j - \overline{y})^2}\left((1 - R^2)(y_i - \overline{y}) - r_i\right). \tag{4.11}$$

As we see, the Y-influence for coefficient of determination consists of two parts: the first part is associated with the $y-$residual and the second part is associated with the OLS residual.

t-statistic. Let $D = \mathrm{diag}((X'X)^{-1})$ denote the $m \times m$ diagonal matrix of the inverse to $X'X$, then the vector of $t-$statistics can be written as $t = s^{-1}D^{-1/2}\hat{\beta}$

where $s^2 = RSS/(n - m)$. Omitting some algebra and using previously derived formulae (4.3) and (4.10), one obtains

$$\frac{\partial t}{\partial y_i} = \frac{1}{s} D^{-1/2}(X'X)^{-1}x_i - \frac{u_i}{RSS}. \tag{4.12}$$

Hence, plotting these derivatives against i one can identify influential observations of the dependent variable in terms of sensitivity of $t-$statistics to small changes in observation of the dependent variable (see the next section for an example).

4.6.2. $X-$influence

The influence of independent variable on RSS is trivial – it easy to show that the derivative of RSS and R^2 with respect to x_{ik} is proportional to the OLS residual. The $X-$influence of $t-$statistic is measured as the partial derivative,

$$\frac{\partial t}{\partial x_{ik}} = \frac{\hat{\beta}_k u_i}{RSS} t + \frac{1}{s} D^{-1/2}(X'X)^{-1}(e_k u_i - x_i \hat{\beta}_k) + \frac{1}{s} p$$

where p is the $m \times 1$ vector with the jth component

$$p_j = ((X'X)^{-1}_{jj})^{-3/2}(X'X)^{-1}_{jk} \sum_{l=1}^{m} (X'X)^{-1}_{jl} x_{il} \hat{\beta}_j$$

One can expect that this influence analysis will be especially useful for correlated explanatory variables: then it can identify influential observations of explanatory variable that make $t-$statistics low.

4.7. Example 1. Woman Body Fat

In order to illustrate the I-influence approach we consider an example of regression of 20 women Body Fat, y_i on Triceps skinfold thickness, x_{i1} and Thigh circumference, x_{i2} from Neter et al. (1990). The estimated multivariate regression is $\hat{y}_i = .2224x_1 + .6594x_2 - 19.174$. The standard technique in visualizing influential cases is to use proportional influence plot – it is a scatter plot where the ith observation point is represented by a circle of radius proportional to the leverage, p_i. Then, larger circles in the plot identify more influential cases. However, we can make such graphs more informative displaying the influence of both the dependent and independent variable where the latter is calculated by formula (4.5). Such a graph will be called cross influence plot. Since in our example there are two continuous variables it is relevant to study the influence of individual observation of Triceps and Thigh on the predicted value of Body Fat, see Figure 2. Each case is represented by a cross. The length of the vertical bar is proportional to the leverage, p_i and is equal to the diameter of the circle in the standard proportional influence graph. The length of the horizontal bar is proportional to (4.5). Two regression lines are displayed. The solid line for the left graph is the slice of the multivariate model by the plane $x_2 = \overline{x}_2$; i.e. $E(y|x_2 = \overline{x}_2) = .2224x_1 + .6594\overline{x}_2 - 19.174.$ The dotted line is the simple regression y on $x_1.$ The difference in slopes reflects the multicollinearity, that is the case in this example. In particular, if x_1 and x_2 are uncorrelated these lines must be parallel. Interesting that, as it is follows from the right graph, the coefficient at Thigh is not affected by the multicollinearity. As we see, the longest vertical bar corresponds to case #3. On the other hand, the horizontal bar for this case is relatively small, especially for Triceps-influence.

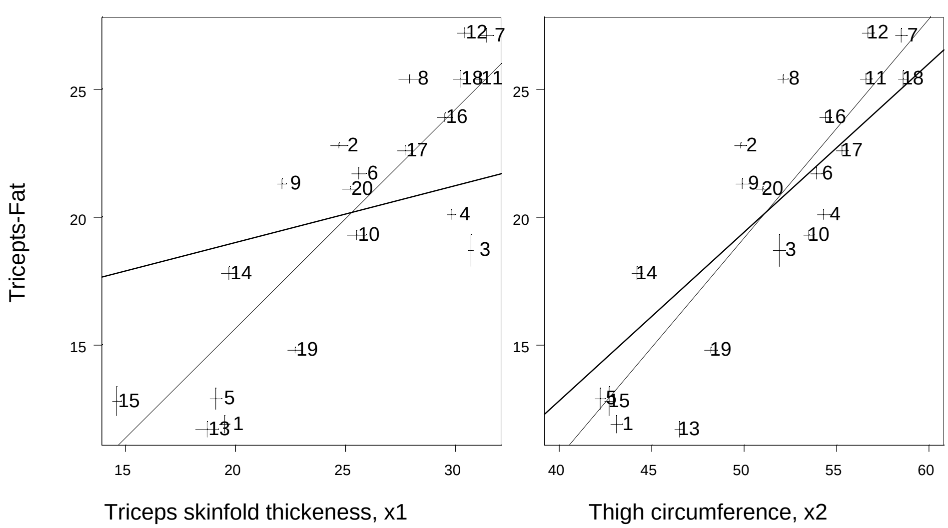

Figure 2. Cross Influence plot for the Woman Body Fat example. Solid line – simple regression, bold line – slice of the multivariate OLS-regression plane at the mean of the other variable. The length of the vertical bar of the cross is proportional to the influence of the according Body Fat observation on the predicted value, leverage. The length of the horizontal bar is proportional to the influence of the according independent variable on the predicted value of Fat in the multivariate regression.

It means that the predicted value for case #3 is sensitive to perturbation of the dependent variable and not so much sensitive to perturbation in the explanatory variables.

Star in‡uence plot shows the in‡uence of individual observations on the OLS coe¢cients, Figure 3. In order to compare magnitudes of the in‡uence, the percent change in beta-coe¢cients is displayed. Coe¢cient $\hat{\beta}_1 = .2224$ is more sensitive to perturbation and its rate lies in the range from -20% to 35%.

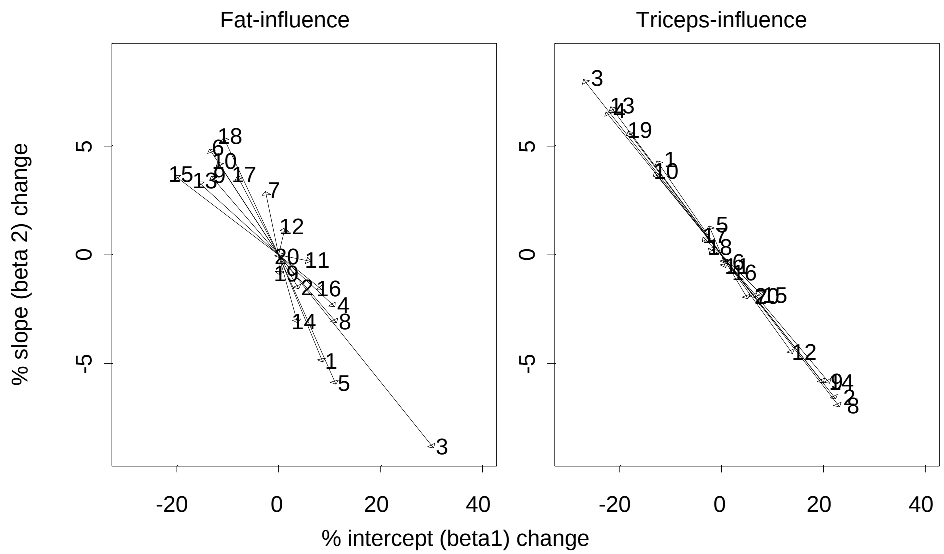

Figure 3. Star In‡uence Plot for Woman Body Fat example. For the Fat-in‡uence formula (4.1) is used, for the Triceps-in‡uence formula (4.3) is used. The number at the end of the arrow displays case #. As follows from the Fat-in‡uence analysis an increase of the third observation of Fat by one leads approximately to 35% increase of the intercept and 8% decrease of the second slope. The arrows in the right graph lies on the line because of multicollinearity.

Small changes in observation of Fat for case #3 a¤ect both coe¢cients: $\hat{\beta}_2$ decreases and $\hat{\beta}_1$ increases, see the left graph. Speci…cally, the intercept increases approximately by 35% and the …rst slope ($\hat{\beta}_2$) decreases by 8% if the observation of Fat in the third case increases by 1. Vice versa, small positive change of Triceps in case #3 increases $\hat{\beta}_2 = .6594$ and decreases $\hat{\beta}_1$; the right graph. In‡uence arrows lie on the line because of multicollinearity.. Star In‡uence plot is useful when simultaneous changes of OLS coe¢cients are under investigation.

The bar in‡uence plot is shown in Figure 4 and displays relative changes in the beta-coe¢cients in a di¤erent manner.

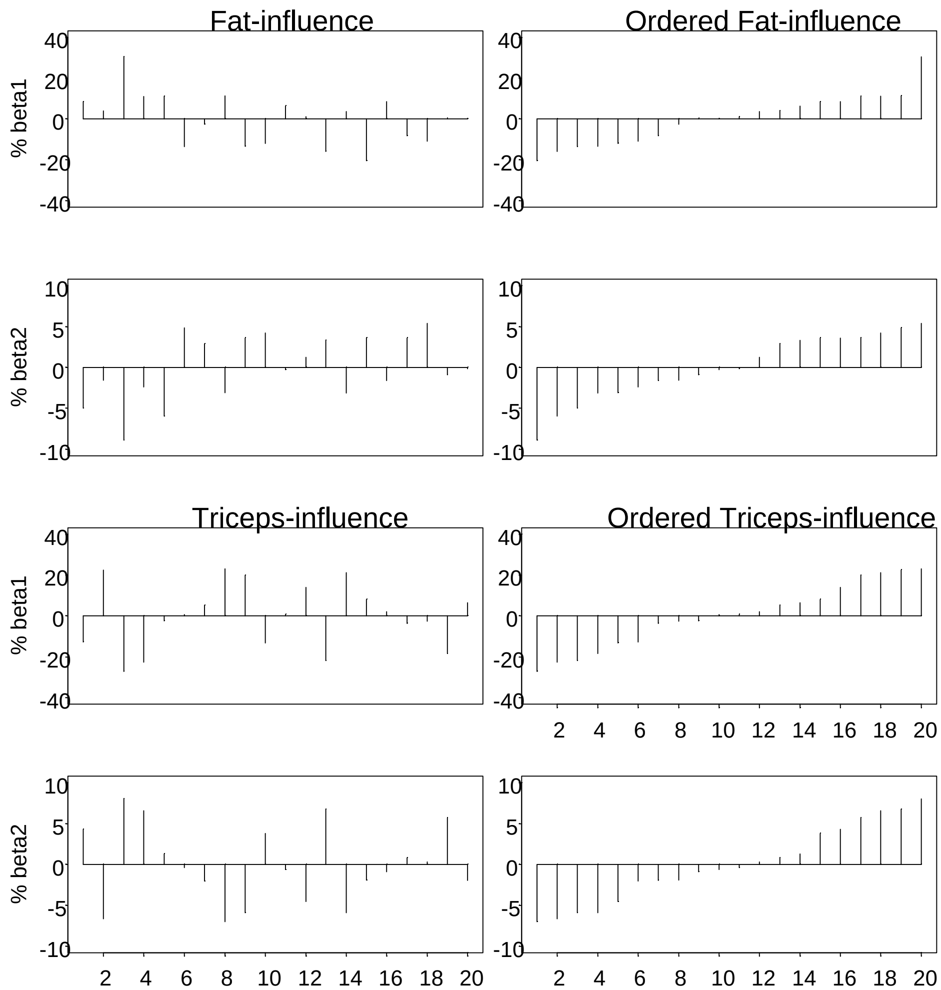

Figure 4. Bar In‡uence plot for the Woman Body Fat example. In the graphs at right the in‡uence is ordered so that in‡uential observations are located at left and right. The in‡uential case number can be found on the left graph projecting the in‡uence bar from right to left. This type of graph is convenient for group in‡uence detection. For instance, we can identify four observations of Triceps which are in‡uential on $\bar{\mathbf{b}}_1$ – they are located at the right side.

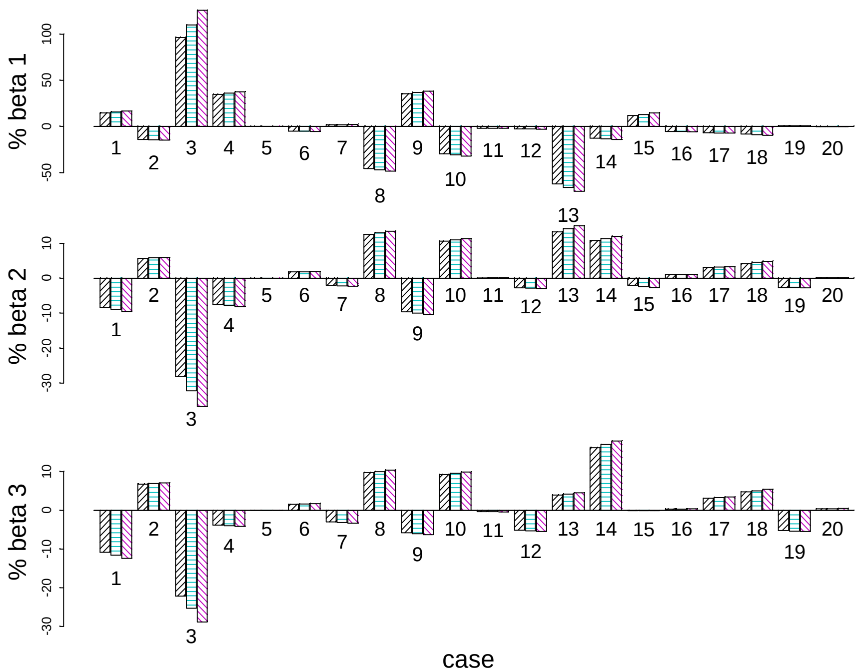

Figure 5. Case deletion in‡uence plot for the Woman Body Fat Example as % of change in the OLS-estimate. For each case the left bar corresponds to in…nitesimal deletion at inclusion (4.8), the middle bar corresponds to full case deletion (2.2), and the third bar corresponds to deletion at exclusion (4.9). The length of the in‡uence bar is proportional to the change in the beta-coe¢cient after the ith case is deleted.

This graph should be read from right to left. At the right graph one can identify maximum in‡uence located at the end-points. Projecting the bars on the left graph it is easy to localize the in‡uential cases. Bar in‡uence plots are useful for detecting groups of in‡uential observations and comparison of in‡uence for di¤erent regression coe¢cients. As we see, $\hat{b}_1$ is slightly more sensitive to perturbation both in dependent and independent variables, it ranges from -20 to 30% and for $\hat{b}_2$ from -10 to 7%.

The in‡uence of case deletion is illustrated in Figure 5 where relative changes in beta-coe¢cients are shown. Three bars according to formulae (4.8), (2.2), and (4.9) for three regression coe¢cients are displayed. As was noticed in subsection

3.3, the in‡uence of deletion approximately is equal the half sum of the two others. The most in‡uential case is #3. Looking back in Figure 4 we see that this is due to observations on Body Fat and Triceps. Bar in‡uence in combination with deletion in‡uence plots can help in identifying what individual observation of what variable is in‡uential. Notice, that the standard deletion diagnostics do not allow to identify which variable makes the case in‡uential.

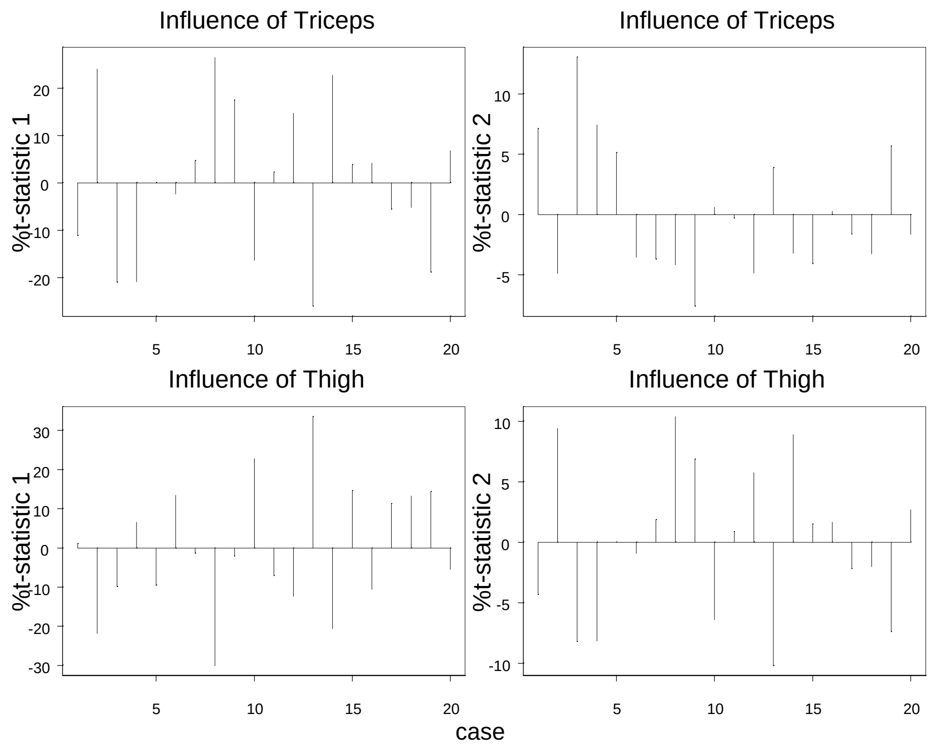

Figure 6. I-in‡uence of t-statistics. The right bottom graph reveals that an increase of second observation of Thigh by one approximately increase t-statistic for the second slope by 10%.

4.8. Example 2. Gypsy moth studies

In this section we illustrate how I -in‡uence analysis can be applied to curvilinear regression. We use burlap data on gypsy moth study (Bounaccorsi 1994) to illustrate the I-in‡uence in curvilinear regression. The objective, as formulated by Bounaccorsi, was "... to see how well counts of gypsy moth egg mass found under burlap bands on trees can be used to predict the egg mass density for a large area." The data consist of 51 measurements of egg mass, denoted as megg, in the area and egg mass under burlap, denoted as mburlap. Due to the objective of prediction we look for relationship of megg on mburlap, the left plot in Figure 7.

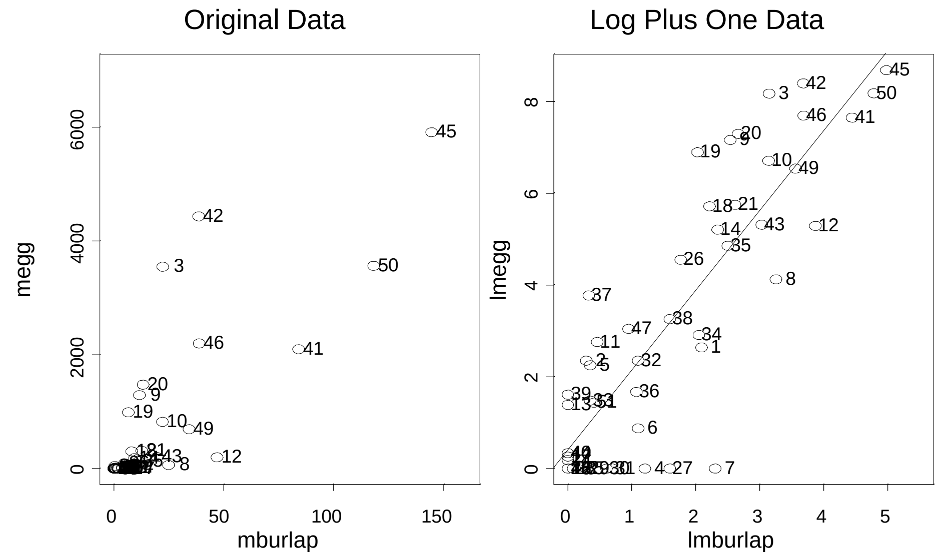

Figure 7. Gypsy moth study, burlap data. The left graph: scatter plot of original data. The right graph: scatter plot of log plus 1 data with the regression line. The transformation makes the scatter plot more informative.

As wee see, there is a cluster of observations in the neighborhood of zero because some observations are zero or close to zero. Thus, unlike Bounaccorsi who studied the relationship of mburlap on megg, we suggest: (i) to use the reversed relationship, i.e. megg on mburlap, (ii) to take logarithms of variables plus 1. There is an interesting property of this transformation: for large mburlap and megg we obtain a power relationship and for small it is linear. In fact, if $\log(y+1) = a + b\log(x+1)$ then $y = c(x+1)^b - 1$ where $c = \exp(a)$; and we have $\lim y/x^b \simeq c$ for large x: If $x \simeq 0$ we can approximate $(x+1)^b \simeq 1 + bx$ and then $y \simeq (c-1) + (cb)x$: Therefore, the regression model we suggest is curvilinear,

$$\text{lmegg}_i = a + b \cdot \text{lmburlap}_i + \varepsilon_i$$

where $\text{lmegg}_i = \log(\text{megg}_i + 1)$ and $\text{lmburlap}_i = \log(\text{mburlap}_i + 1)$: The transformation makes the scatter plot more informative, compare two graphs in Figure

7.

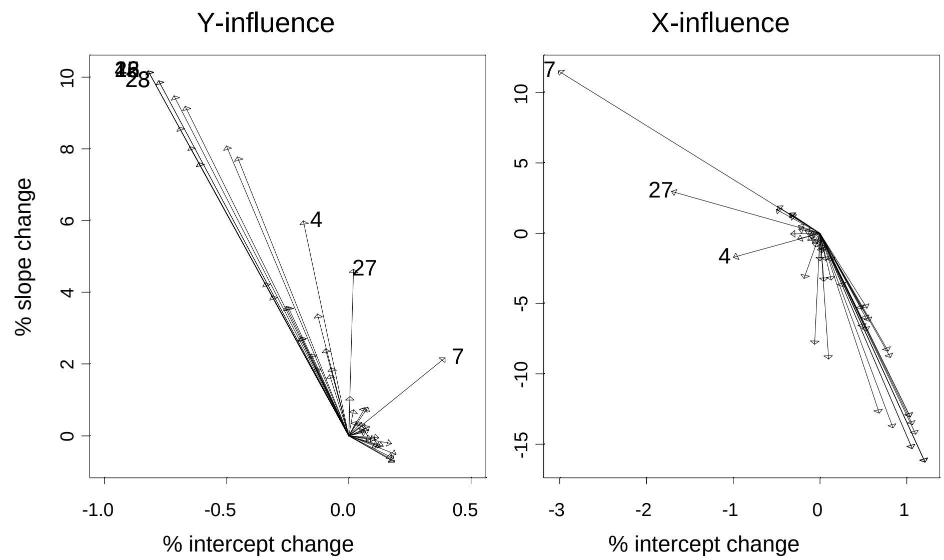

Figure 8. Star In‡uence Plot, burlap data. Case #7 is in‡uential – an increase of mburlap by one would increase the slope by 10%.

We show only Star In‡uence Plot for this example – it shows how small perturbation in megg and mburlap a¤ects the OLS estimate for the intercept and slope, Figure 8. The derivatives are calculated similar to formula (4.6) where $g(s) = \log(s + 1)$: The left plot shows the impact of small changes in megg, and the right one shows the impact of mburlap. The star in‡uence plot helps us to view directions of data in‡uence in the parameter space. Clearly, case #7 is in‡uential. However, the 7th observation of megg is not in‡uential on the slope, and has a positive outstanding in‡uence on the intercept. The 7th observation of mburlap is in‡uential on both slope and intercept. Therefore, one can infer that the 7th case is in‡uential because of 'wrong' measurement of mburlap, not megg. We notice that we could detect the impact of mburlap using standard technique based on leverage or case deletion diagnostics because the in‡uence would be associated with the case as whole.

5. Nonlinear regression model

I-in‡uence analysis is especially useful in complex statistical models, like nonlinear regression where case deletion diagnostics lead to time expensive regression recalculations. The nonlinear regression model is written as $y_i = f_i(\beta; x_i) + \varepsilon_i$ where x_i is the vector of explanatory variables subject to influence analysis. The Least Squares (LS) estimate, $\widehat{\beta}$ satisfies the vector normal equation

$$\sum_{i=1} (y_i - f_i(\beta; x_i)) \frac{\partial f_i(\beta; x_i)}{\partial \beta} = 0. \tag{5.1}$$

Apparently, $\widehat{\beta}$ can be viewed as a function of y_i and x_i. How $\widehat{\beta}$ is affected by individual observation of the dependent or explanatory variable? Following the approach of I-influence analysis this influence is measured as $\partial\widehat{\beta}/\partial y_i$ and $\partial\widehat{\beta}/\partial x_{ik}$. There is no closed form solution to the LS estimate in nonlinear regression, so that we need to find the above derivatives treating $\widehat{\beta}$ as an implicit function of y_i and x_i.

5.1. Influence of dependent variable on the LS-estimate

The according derivative can be found either by direct differentiation of (5.1) or applying formula for the derivative of implicit function, as in (3.4); either way leads to

$$\frac{\partial\widehat{\beta}}{\partial y_i} = H^{-1} \frac{\partial f_i}{\partial \beta} \tag{5.2}$$

where

$$H = \sum_{j=1} \left[\left(\frac{\partial f_j}{\partial \beta} \right) \left(\frac{\partial f_j}{\partial \beta} \right)' - (y_j - f_j) \frac{\partial^2 f_j}{\partial \beta^2} \right]$$

is the one-half of the Hessian of the sum of squares; all derivatives are calculated at $\beta = \widehat{\beta}$. Following common suggestion matrix H may be approximated by $G'G$ where G is the $n \times m$ matrix of first derivatives of $\{f_i\}$, Bates and Watts (1987).

5.2. Influence of explanatory variable on the LS-estimate

Again, to find $\partial\widehat{\beta}/\partial x_{ik}$ we can either differentiate the normal equation (5.1) or apply the formula for the derivative of implicit function,

$$\frac{\partial\widehat{\beta}}{\partial x_{ik}} = H^{-1} \left[r_i \frac{\partial^2 f_i}{\partial \beta \partial x_{ik}} - \frac{\partial f_i}{\partial x_{ik}} \frac{\partial f_i}{\partial \beta} \right]. \tag{5.3}$$

It is not difficult to obtain formula for influence in case when explanatory variable is binary, employing the idea of missclasifications, as was done for linear model in section 4.3.

5.3. Influence of dependent variable on the predicted value

Following the I-influence approach, we measure the influence of individual observation y_i on $f(\hat{\beta}) = (f_1(\beta; x_1), \ldots, f_n(\beta; x_n))'$ as $\partial f(\hat{\beta})/\partial y_i$. The idea to use this derivative to conduct influence analysis in nonlinear regression was suggested by Emerson et al. (1984) and later generalized by Laurent and Cook (1992, 1993). In order to find $\partial f(\hat{\beta})/\partial y_i$ we use formula (5.2) and then apply the chain rule:

$$\frac{\partial f(\hat{\beta})}{\partial y_i} = \frac{\partial f(\hat{\beta})}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial y_i} = G \frac{\partial \hat{\beta}}{\partial y_i} = G H^{-1} \frac{\partial f_i}{\partial \beta}. \tag{5.4}$$

Laurent and Cook called the matrix consisted of vectors (5.4) Jacobian leverage. For linear regression model the ith component of vector (5.4) is usual leverage, p_i.

5.4. Influence of case deletion

The influence of case deletion on characteristics of nonlinear regression has been studied by Cook and Weisberg (1982), Ross (1987). Strictly speaking, case deletion leads to regression re-estimation.. To avoid this, one can use a one-step approximation, as was suggested by Pregibon (1981) and Preisser and Qaqish (1996) for generalized linear model. Analogously to linear model two kinds of I-influence of case deletion can be considered: the influence at inclusion and exclusion. In the first type of influence the derivative is computed at $w_i = 1$ and in the second at $w_i = 0$. Omitting fairly simple algebra, the influence at inclusion and exclusion is measured as

$$\left.\frac{\partial \hat{\beta}}{\partial w_i}\right|_{w_i=1} = r_i H^{-1} \frac{\partial f_i}{\partial \beta}; \qquad \left.\frac{\partial \hat{\beta}}{\partial w_i}\right|_{w_i=0} \simeq \frac{r_i}{(1-p_i)^2} (G'G)^{-1} \frac{\partial f_i}{\partial \beta}; \tag{5.5}$$

where r_i is the ith LS residual and

$$p_i = \left(\frac{\partial f_i}{\partial \beta}\right)' (G'G)^{-1} \left(\frac{\partial f_i}{\partial \beta}\right)$$

is the analog of leverage. The second derivative in (5.5) is called 'exclusion 1'. We can use another definition of leverage based on (5.4), that gives

$$\left.\frac{\partial \hat{\beta}}{\partial w_i}\right|_{w_i=0} \simeq \frac{r_i}{(1-p_i^*)^2} H^{-1} \frac{\partial f_i}{\partial \beta} \tag{5.6}$$

where

$$p_i^* = \left(\frac{\partial f_i}{\partial \beta}\right)' H^{-1} \left(\frac{\partial f_i}{\partial \beta}\right)$$

which is called ‘exclusion 2’. Notice, that all derivatives are computed at the LS estimate so that we do not need to reestimate the regression. The three measures are compared in the following example.

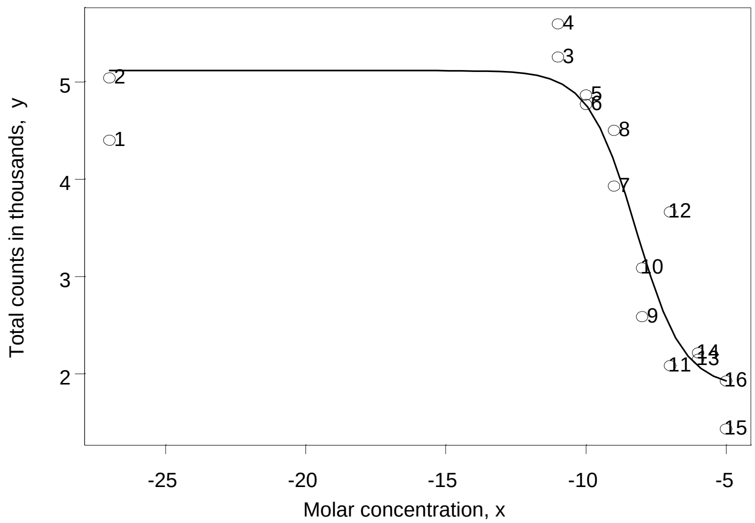

Figure 6. Observation points and the fitted curve for the nonlinear regression model (5.7). For each molar concentration there is a pair of total counts.

5.5. Example 3. Logistic growth curve model

We take an example considered by Laurent and Cook (1993), borrowed from the book by Bates & Watts (1988). The dependent variable y; radioactivity counts in rat heart tissue, is related to molar concentration of nifedipene (NIF), $x = \log_{10}$(concentration NIF) via the logistic growth curve model:

$$f(\theta; x_i) = \theta_1 + \frac{\theta_2}{1 + e^{\theta_4 (x_i - \theta_3)}}; \quad i = 1; \ldots; 16: \tag{5.7}$$

An interesting feature of this data is that for the first two cases ($i = 1; 2$) the concentration is zero, i.e., formally $x = -\infty$ and $f(\theta; x_1) = f(\theta; x_2) = \theta_1 + \theta_2$:

Apparently, a real concentration could be positive, so that one might admit that the measurement tool was not precise enough to measure tiny concentration. Therefore, it is worthwhile to assess the influence of these points. Also, since the sample size is fairly small, one can expect that each case is influential at some degree. The observation points with the fitted curve are shown in Figure 9

In order to display points we set $x_1 = x_2 = -27$, as Laurent and Cook did. The LS estimates with $t-$statistics are $\hat{\theta}_1 = 1923.52\,(5.2)$, $\hat{\theta}_2 = 3194.92\,(6.7)$, $\hat{\theta}_3 = -8.3214\,(21)$, $\hat{\theta}_4 = 535.6\,(2.6)$. We start our influence analysis with assessing how small perturbation in individual observation of the dependent or independent variable affects the LS estimates: formulae (5.2) and (5.3), Figure 10 (influence plots for the second and the third theta-parameters are not shown).

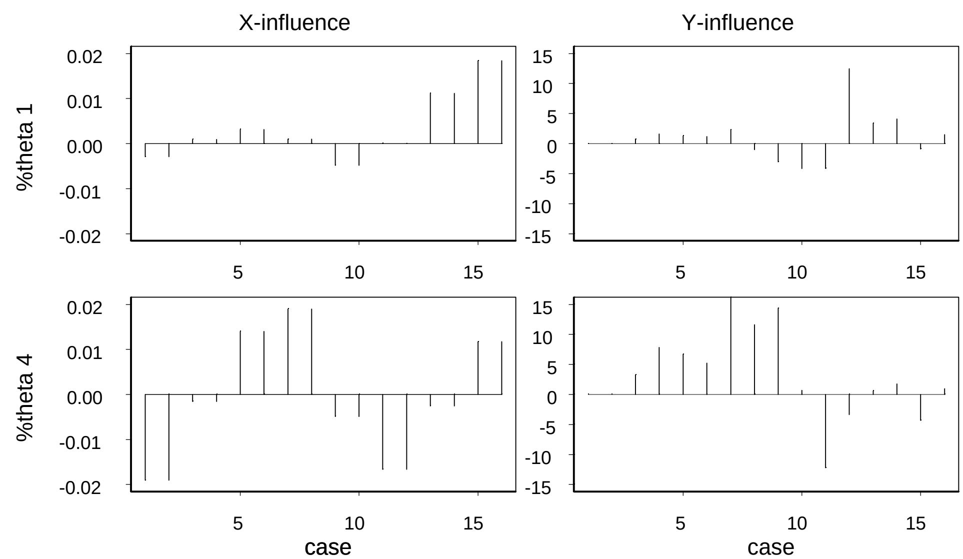

Figure 10. Bar Influence Plot for logistic growth curve model (5.7). The 12th observation on total counts affects $\hat{\theta}_1$ (the increase of y_{12} by one thousand increases $\hat{\theta}_1$ approximately by 12%). $\hat{\theta}_4$ is affected by the 7th observation of counts.

The influence analysis with respect to x could answer the question how well the experiment was designed and what should be done to improve it. First of all, we observe that $\hat{\theta}_1$ is more sensitive than $\hat{\theta}_4$ probably because $\hat{\theta}_1$ is less significant, the fact is likely to be general. Secondly, parameters are much more sensitive to $y-$observations than to $x-$observations. This fact could not be revealed using standard case deletion diagnostics because the in‡uence of the dependent and independent variables are not separated. As we also see, the two left-end points have little e¤ect on parameters. From the other hand, the right-end points have maximum in‡uence on the …rst and fourth parameters. A close look in Figure 10 clari…es the reason for that: parameter μ_4 corresponds to the rate of y change with respect to x and two right x-observations bring substantial information for estimation parameter μ_4. Hence, in order to get more precise estimates of the rate-parameter the experimentalist has to add design points with $x > ¡5$: Now, let us consider the in‡uence of case deletion on parameters estimation, Figure 11 (…rst and fourth parameters).

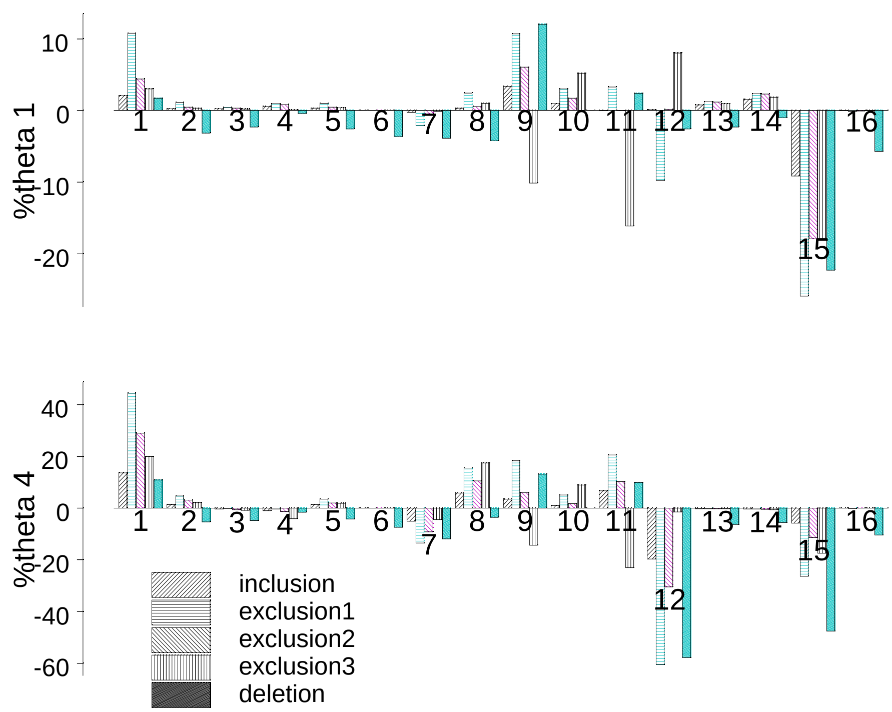

Figure 11. Case deletion diagnostics for nonlinear regression (5.7).

Five measures of case deletion as relative changes to the LS estimates are computed. The …rst four measures are calculated by formulas (5.5-??), the …fth one, 'deletion' corresponds to precise case deletion and regression recalculation. As we see, the in‡uence of case deletion in a certain way accumulates the in‡uences driven by the dependent and independent variable considered above. Case #15 is influential for all parameters, this part of influence inference coincides with the conclusion made by Laurent and Cook (1993) based on Jacobian leverage (5.4). However, in contrast to their analysis case #16 is not influential, that can be verified looking back in Figure 10. For this example all five measures behave quite differently for some cases, however generally they are correlated.

6. Logistic regression

In this section we deal with binary dependent variable. Thus, let y_i code the occurrence of a certain event: $y_i = 1$ means the event took place, $y_i = 0$ did not. For instance, $y_i = 1$ may mean individual i has disease, and $y_i = 0$ means individual i is disease free. As before, $m \times 1$ vector x_i denotes the correspondent vector of explanatory variables (covariates). In logistic regression the occurrence of the event given x_i is modelled via probability defined as

$$\Pr(y_i = 1) = \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}}, \qquad i = 1, \ldots, n \tag{6.1}$$

where the $m \times 1$ vector β is the parameter of interest. Commonly, logistic regression is estimated by maximum likelihood method. The log-likelihood function for data (y_i, x_i) has the form $\beta' \sum_{y_i=1} x_i - \sum_{i=1}^{n} \ln(1 + e^{\beta' x_i})$, and the MLE is determined by the score equation

$$\sum_{i=1} \frac{1}{1 + e^{\beta' x_i}} x_i - \sum_{y_i=0} x_i = 0. \tag{6.2}$$

The aim of this section is to provide measures of influence of the dependent variable and covariates on the MLE, $\hat{\beta}$ as the solution to (6.2). Following our approach we measure this influence as the derivative of $\hat{\beta}$ with respect to individual observation on either dependent or independent variable.

6.1. Influence of dependent variable on the MLE

How sensitive is the MLE to changes in the ith observation of the dependent variable, y_i? Since y_i takes values 0 or 1, a straightforward solution would be recalculating the logistic regression replacing $y_i = 1$ by $y_i = 0$, and vise versa. However, it is possible to avoid massive recalculations employing the idea of missclassification, as was done before for explanatory variable in linear model (Section

3). Thus, let us assume the observed event is symmetrically missclassified with certain probability q_i ($q_j = 1; j \neq i$): Then, the probability of the event $y_i = 1$ under missclassification is

$$\Pr(y_i = 1) = (1 - q_i)\frac{e^{-\beta' x_i}}{1 + e^{-\beta' x_i}} + q_i \frac{1}{1 + e^{-\beta' x_i}} = \frac{(1 - q_i)e^{-\beta' x_i} + q_i}{1 + e^{-\beta' x_i}}. \tag{6.3}$$

We notice that in special case when missclassification is absent ($q_i = 0$); one obtains the former probability (6.1). Vise versa, if $q_i = 1$ one comes to a reverse coding. Thus, given q_i the true model is (6.3) with the according log-likelihood. The MLE $\hat{\beta}$, as the solution to the log-likelihood with probability (6.3), is a function of q_i. Therefore, the derivative $\partial\hat{\beta}/\partial q_i$ at $q_i = 0$ can be interpreted as a measure of influence of observation y_i on the MLE. After some algebra we obtain the formula for the derivative

$$\left.\frac{\partial \hat{\beta}}{\partial q_i}\right|_{q_i = 0} = \frac{e^{\mathbf{b}' x_i} - 1}{e^{\mathbf{b}' x_i} + 1} H^{-1} x_i \tag{6.4}$$

where

$$H = \sum_{j=1} \frac{e^{\mathbf{b}' x_j}}{(1 + e^{\mathbf{b}' x_j})^2} x_j x_j'$$

is the Hessian of the negative log-likelihood. It is interesting to notice that the influence is zero if $\exp(\hat{\beta}' x_j) = 1$, i.e. when the probability of event is $1/2$.

6.2. Influence of covariate on the MLE

The influence of the individual observation x_{ik} on the MLE is measured as $\partial\hat{\beta}/\partial x_{ik}$. We find this derivative differentiating (6.2) with respect to x_{ik};

$$\frac{\partial \hat{\beta}}{\partial x_{ik}} = H^{-1}\left(e_k r_i - \frac{e^{\mathbf{b}' x_i}}{(1 + e^{\mathbf{b}' x_i})^2} x_i \hat{\beta}_k \right) \tag{6.5}$$

where $r_i = y_i - e^{\mathbf{b}' x_i}/(1 + e^{\mathbf{b}' x_i})$ is the ith residual of logistic regression. As the reader can see, formula (6.5) resembles its linear analog (4.3).

6.3. Influence on the predicted probability

In some instances we may interested in prediction of probabilities (e.g. Johnson 1985). Then, the characteristic of interest is $\hat{p}_i = e^{\mathbf{b}' x_i}/(1 + e^{\mathbf{b}' x_i})$ which may be analyzed with respect to influence of missclassification of the binary variable, individual observation of covariate, or case deletion. We find the derivative of $\hat{p}_i$ based on the derivative of the MLE applying the chain rule. We start with calculating the influence with respect to x_{ik} based on formula (6.5). Hence applying the chain rule we obtain

$$\frac{\partial \hat{p}_i}{\partial x_{ik}} = \left(\frac{\partial \hat{p}_i}{\partial \hat{\beta}}\right)' \left(\frac{\partial \hat{\beta}}{\partial x_{ik}}\right) = \frac{e^{\mathbf{b}' x_i}}{(1+e^{\mathbf{b}' x_i})^2} x_i' H^{-1} \left(r_i e_k - \frac{e^{\mathbf{b}' x_i}}{(1+e^{\mathbf{b}' x_i})^2} x_i \hat{\beta}_k \right).$$

Now we find how $\hat{p}_i$ is sensitive to missclassification q_i. Using (6.4) one obtains

$$\frac{\partial \hat{p}_i}{\partial q_i} = \left(\frac{\partial \hat{p}_i}{\partial \hat{\beta}}\right)' \left(\frac{\partial \hat{\beta}}{\partial q_i}\right) = \frac{e^{\mathbf{b}' x_i}(e^{\mathbf{b}' x_i} - 1)}{(1+e^{\mathbf{b}' x_i})^3} x_i' H^{-1} x_i. \tag{6.6}$$

We recall that in linear model the influence of the dependent variable on the predicted value is measured via the diagonal element of the hat matrix, leverage. Therefore, (6.6) can be considered as a generalization of the leverage for logistic regression. It is interesting to notice that the factor $x_i' H^{-1} x_i$ looks similar to linear model; however, the scalar factor in (6.6) is specific to logistic regression and reflects the binary nature of the dependent variable.

6.4. Influence of case deletion on the MLE

The theory of I-influence for case deletion in logistic regression was developed by Pregibon (1981). We can apply the technique developed for linear and nonlinear regression models to logistic regression. Thus, let i be fixed and w_i be the weight of the ith case. The MLE is a function of w_i, and we aim to find $\partial \hat{\beta} / \partial w_i$. The score equation for $y_i = 1$ under the assumption that the ith case has weight w_i has the form

$$\sum_{j \neq i} \frac{1}{1+e^{-\beta' x_i}} x_i + \frac{w_i}{1+e^{-\beta' x_i}} x_i - \sum_{y_i = 0} x_i = 0.$$

It is easy to write a similar score equation for $y_i = 0$. Differentiation with respect to w_i leads to two formulae corresponding to deletion at inclusion and exclusion,

$$\left.\frac{\partial \hat{\beta}}{\partial w_i}\right|_{w_i = 1} = r_i H^{-1} x_i; \qquad \left.\frac{\partial \hat{\beta}}{\partial w_i}\right|_{w_i = 0} \simeq \frac{r_i}{(1-p_i)^2} H^{-1} x_i;$$

where r_i is the ith residual defined above. It is easy to calculate the derivative for the predicted probability based on the chain rule.

6.5. Example 4. Finney data

We consider an example of logistic regression from Pregibon (1981) which is based on Finney data. The dependent variable indicates the occurrence (1) or nonoccurence (0) of vaso-constriction in the skin of the digits, x_1 and x_2 are the Volume and Rate of air inspired on a transient vaso-constriction, number of cases $n = 40$. The logistic regression in logarithms has the form $\text{logit}(y) = \beta_1 \log(\text{Volume}) + \beta_2 \log(\text{Rate}) + \beta_3$ and estimated by ML gives $\hat{\beta}_1 = 5.18$ and $\hat{\beta}_2 = 4.56$. The analysis of case deletion and infinitesimal deletion was conducted by Pregibon. Here we concentrate on I-influence of missclassification and individual observations of the independent variables, Volume and Rate, based on formulae (6.4) and (6.5), Figure 12.

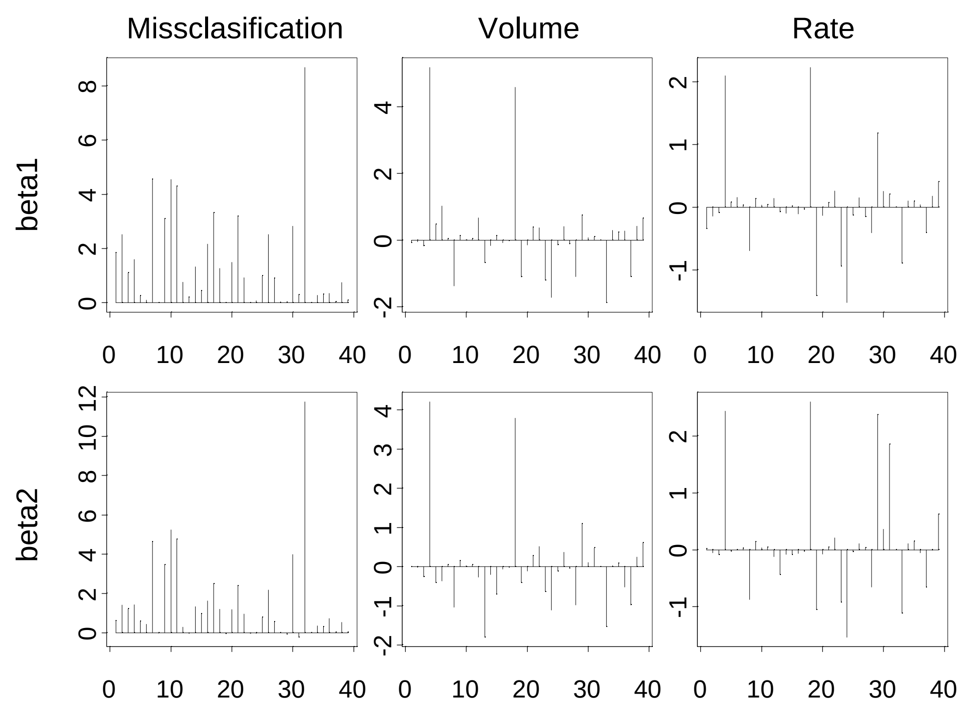

Figure 12. Bar influence plot for logistic regression, Finney data. Case #32 is influential – missclassification is suspected.

We notice that these formulae should be modified because variables are in logarithms. As we see from the left plots, $y_{32} = 0$ has an outstanding effect on the coe¢cients. Interesting, that there is almost no e¤ect with respect to Volume and Rate for this case. Indeed, if $y_{32} = 0$ is replaced by $y_{32} = 1$ coe¢cients change dramatically: from 5:18 to 3:24 and from 4:56 to 0:99 respectively. Also, we can see that the 32th observation lies outside of the bulk of the data from the 3D plot in Figure 13. Apparently, this case deserves a close look in terms of correctness of measurement and recording. Interestingly, standard case deletion diagnostics accomplished by other authors do not reveal this fact.

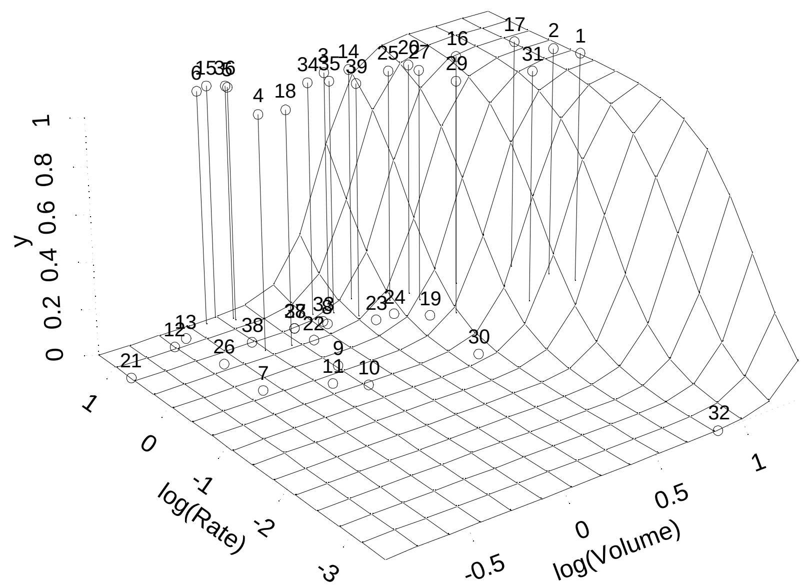

Figure 13. 3D plot of logistic regression model, Finney data. Actual observations are shown by circles. Case #32 is apart from the bulk of the data.

7. In‡uence of correlation structure

One of the assumptions of ordinary regression is that residuals are uncorrelated. How strong is this assumption in a concrete regression, what is the in‡uence of this assumption on the OLS estimate, i.e. what are consequences of possibly missing correlation structure? In particular, this question might be reasonable when regression analysis is applied to time series data.

We write the standard linear model as $y_t = \beta' x_t + \varepsilon_t$ were t denotes time index. Let us assume the correlation structure is induced by the first-order autoregression: $\varepsilon_t = \rho\varepsilon_t + \eta_t$ where η_t are i.i.d. The (i,j)th element of the correlation matrix $\Omega = \Omega(\rho)$ for $\varepsilon_1, \ldots, \varepsilon_n$ is equal $\rho^{|i-j|}$ and the weighted least squares estimate has the form $\hat{\beta}(\rho) = (X'\Omega^{-1}(\rho)X)^{-1}X'\Omega^{-1}(\rho)y$. Apparently, small departure of ρ from zero changes the estimate with the rate

$$\left.\frac{d\hat{\beta}(\rho)}{d\rho}\right|_{\rho=0} = -(X'X)^{-1}X'u \qquad (7.1)$$

where $u' = (r_2, r_1 + r_3, \ldots, r_{n-1})$ and r_t is the tth OLS residual. Similar formula can be obtained for nonlinear regression. One can assess the influence of possibly missing correlation structure by examining the components of vector (7.1).

8. Influence of measurement error in binary model

Often explanatory variables contain measurement errors. There is a well established theory of errors-in-variables (Fuller 1987; Carroll et al. 1996). Generally, if there are measurement errors in explanatory variables, parameters of the model are not identifiable unless additional information is provided. In particular, the MLE of parameters exist if variance of measurement error, σ^2 is known. In practice, a calibration or validation study must be undertaken to obtain an estimate of the variance. However, since such studies usually are expensive it might be very useful to assess the effect of measurement error on parameters estimates prior to variance estimation. It is well known that in simple linear regression with errors-in-variables the OLS-slope is attenuated. In multivariate linear regression, due to correlation among covariates, the effect of measurement error may be not so straightforward. Things are more complicated in nonlinear models, e.g. generalized linear models or nonlinear regressions. Several papers address the problem of influence analysis in regression models with measurement error in explanatory variables: Wellman and Gunst (1991) use influence function and one-step approximation to assess the influence in the case deletion diagnostics.. This idea is further generalized in two papers by Zhao and Lee (1994, 1995). While in those papers the variance of measurement error, σ^2 is known and positive, we are interested in how the MLE is sensitive to small departure of σ^2 from zero, i.e. what is the influence of possible measurement error. Similar approach has been taken by Chesher (1991), but he investigated the effect of small measurement error on the distributions. On contrary, our primary interest is the estimate itself.

Thus, the aim of this section is to show how to assess the effect of measurement error in explanatory variable on the MLE in the neighborhood of $\sigma^2 = 0$; via the spirit of the I-influence approach: This quick and easy to accomplish influence analysis does not require knowledge of σ^2 and may give rise to collect additional data or conduct validation study to estimate σ^2 if the influence of the measurement error on parameters of interest is substantial. This idea, in fact, is not new, e.g. Stefanski (1985) and Stefanski and Carroll (1985) studied statistical implications of measurement error via Taylor series expansion.

The plan is: (i) set up the model with known error variance σ^2, (ii) calculate $d\hat{\beta}(\sigma^2 = 0)/d\sigma^2$ where $\hat{\beta}$ is the MLE. Hence, we obtain a linear approximation to the MLE in the neighborhood of small variances: $\hat{\beta} \simeq \hat{\beta}_0 + \sigma^2 d\hat{\beta}(\sigma^2 = 0)/d\sigma^2$ where $\hat{\beta}_0$ is the MLE with no measurement error: This method provides: (a) information what coefficients are most sensitive to possible measurement error, (b) direction in changes of the OLS-estimates, (c) preliminary coefficients estimates given values of the error variance.

Let the set of covariates consist of fixed vector u_i measured without error and unobserved univariate covariate x_i measured with error. To simplify, we shall assume Berkson measurement error model (Berkson 1950, Fuller 1987, Carroll et.al. 1996), i.e. $x_i = z_i + \sigma\delta_i$ where z_i is the design variable and δ_i is the standardized measurement error, i.e. $E(\delta_i) = 0$ and $var(\delta_i) = 1$; where σ^2 is the variance of measurement error. We do not specify the distribution of measurement error. The true regression model is expressed in terms of x_i and in order to construct the log-likelihood for available observations $(y_i; z_i)$ we have to integrate out measurement error to obtain the 'observed' model, Carroll et.al. (1996).

The binary model is defined via conditional probability $\Pr(y_i = 1 \mid x_i) = H(\gamma' u_i + \tau x_i)$; where H is the inversed link function, and $\beta = (\gamma'; \tau)'$ is the common vector of parameters. If $H = \exp/(1+\exp)$ we come to logistic regression model considered in the previous section, if $H = \Phi$; normal distribution function, we come to probit regression. The 'observed' model can be written as

$$\Pr(y_i = 1 \mid z_i) = E_\delta H(\gamma' u_i + \tau z_i + \sigma\tau\delta)): \qquad (8.1)$$

Omitting tedious derivation, the influence of the measurement error on the MLE is measured as the expected derivative,

$$\left.\frac{d\hat{\beta}}{d\sigma^2}\right|_{\sigma^2=0} = -\frac{1}{2}\tau^2 \left(\sum_{i=1} \frac{H_i'^2}{H_i(1-H_i)} \begin{bmatrix} u_i \\ z_i \end{bmatrix} \begin{bmatrix} u_i \\ z_i \end{bmatrix}'\right)^{-1} \left(\sum_{i=1} \frac{H_i'' H_i'}{H_i(1-H_i)} \begin{bmatrix} u_i \\ z_i \end{bmatrix}\right); \tag{8.2}$$

where H_i and its derivatives H_i and H_i'' are computed at $\hat{\beta}'u_i + \hat{\tau}x_i$ where $\hat{\beta}$ and $\hat{\tau}$ are the MLEs for standard binary model computed without measurement error. Similar result for logistic regression, formulated in terms of asymptotic behaivior, can be found in Stefanski and Carroll (1985). According I-influence approach this derivative indicates how measurement error affects the MLE. We notice that the derivative is zero if $H'' = 0$ – it complies a well-known fact that the MLE does not change in linear regression model with Berkson measurement error, Berkson (1950).

We apply formula (8.2) to logistic regression model with Finney data, analyzed in section 5.4. How sensitive are the MLE to possible measurement error of Volume and Rate? Is the measurement error (m.e.) influential for the beta coefficients? Obviously, the rate of the MLE change, as a function of the variance of the m.e., computed at $\sigma^2 = 0$ may be a reasonable approximation at least for small m.e. We assume multiplicative Berkson measurement error in explanatory variables that implies the error is additive for logarithms. Then, we choose an interval of reasonable change of σ as $(0; .1)$. Apparently, $\sigma = .1$ can be interpreted as 10% measurement error SD. Figure 12 displays change of the MLE for β_1 and β_2 as functions of σ based on linear approximation $\hat{\beta} \simeq \hat{\beta}_0 + \sigma^2 d\hat{\beta}(\sigma^2 = 0)/d\sigma^2$ where the derivative is computed by formula (8.2) with the inversed link function $H = \exp/(1+\exp)$. Two m. e. scenarios are considered: m.e. in Volume and Rate. The influence of measurement error on the MLE is moderate. Interestingly, the MLEs are more sensitive to measurement error in Volume. For instance, 10% m.e. SD leads to increase of MLE by 8%. Positiveness of the derivative (8.2) complies with the well known fact that the MLE computed without m.e. attenuates the true estimate – this explains why our curves increase with σ.

In order to illustrate this measure we consider the probit model with normally distributed measurement error. This choice of the link function provides that the 'observed' model (8.1) remains probit, Carroll et al. (1984), Tosteson et al. (1989). More precisely, for probit model with the normally distributed measurement error

$$\Pr(y_i = 1 \mid z_i) = \Phi\left(\frac{\gamma' u_i + \tau z_i}{\sqrt{1 + \sigma^2\tau^2}}\right).$$

This implies that the exact MLE, as a function of σ^2 can be expressed via naive

($\sigma^2 = 0$) probit estimate $\hat\alpha_0$ and $\hat\beta_0$ as

$$\hat\alpha_{ML} = \frac{\hat\alpha_0}{\sqrt{1-\sigma^2\hat\beta_0^2}}; \qquad \hat\beta_{ML} = \frac{\hat\beta_0}{\sqrt{1-\sigma^2\hat\beta_0^2}}. \tag{8.3}$$

Clearly, the sensitivity of the MLE to small measurement error can be measured as

$$\left.\frac{d\hat\beta_{ML}}{d\sigma^2}\right|_{\sigma^2=0} = \frac{1}{2}\hat\beta_0^3 \tag{8.4}$$

The linear approximation $\hat\beta_{ML} \simeq \hat\beta_0 + .5\sigma^2\hat\beta_0^3$ works well in the neighborhood of small measurement error. We refer the reader to the graphs on pp. 91 and 93 in the book by Carroll et al. (1995), where the MLEs in the logistic regression model are displayed as functions of the variance of measurement error for Framingham Heart Study data, as a result of very expensive and time consuming simulations. Those functions look quite linear in the neighborhood of zero, and therefore linear approximation based on (8.2) must be pretty accurate. Again, formula (8.2) does not require any recalculations or knowledge the variance of measurement error.

9. Conclusions

Several approaches to influence analyses and related concepts have been developed in the literature: leverage and generalized leverage for nonlinear regression, case deletion diagnostics, influence function and local influence. Each of them deals with a specific feature of influence and, in fact, uses its own definition of influence. For instance, leverage measures how predicted values are affected by individual observation of the dependent variable, case deletion diagnostics seek the effect of case deletion on the estimate or the squared distance form the OLS estimate (Cook's distance), the measure of influence in local influence is the maximum curvature of the likelihood displacement. The strength of infinitesimal influence analysis is that it is distribution free. It suggests another measure of influence, as the derivative of statistic of interest with respect to observation either of the dependent or independent variable. As we assert, this quite a straightforward approach has certain advantages over other more complicated measures. In particular, the following four major features make the I-influence approach different from standard influence analyses such as case deletion diagnostics and local influence:

- The influence of a statistic with respect to a pertubated observation or model assumption is measured as the partial derivative. Thus, the influence measure has a clear interpretation as the rate of statistic change upon small perturbation of the observation (data influence) or the model (model influence).

- Two types of influences are distinguished in the data influence analysis: $Y-$influence and $X-$influence. In the $Y-$influence we seek how observation of the dependent variable affects statistic, in the $X-$influence we seek the influence of individual observation of explanatory variable. Therefore, I-influence is more detailed and may reveal why certain case is influential: because of high influence of observation of the dependent variable or independent variable.

- I-influence analysis is applicable to any statistic or characteristic of interest such as estimate, coefficient of determination, t-statistics, test statistic, etc.

- I-influence analysis does not require likelihood setting unlike local influence approach. For example, I-influence can be applied to $M-$estimators (Huber 1981), quasi-likelihood approach (McCullagh and Nelder 1989), generalized estimating equation approach (Liang and Zeger 1986) where the likelihood is not specified.

A practical strength of I-influence analysis is that it is easy to interpret. For example, in a new treatment study we may say that the observation of blood pressure, as one of the explanatory variable, for the 7th patient is influential on the $t-$statistic of the new treatment effect because 1 unit on the blood pressure scale changes the $t-$statistic by 30% whereas the same change for other patients leads only not more than to 10%. Maybe just this observation spoil the effect of the whole study? It can happen that case deletion diagnostics may not reveal this fact because it seeks the influence of the case as whole, without splitting the influence with respect to dependent and explanatory variable.

ACKNOWLEDGMENTS

This work was supported by grants CA 52192 and CA 61109 from the National Cancer Institute.

References

[1] Belsley, D.A., Kuh, E., Welsh, R.E. (1980). Regression Diagnostics: Identifying In‡uential Data and Sources of Collinearity. New York: Wiley.

[2] Buonaccorsi, J. Measurement error models for gypsy moth studies. in Eds. Lange, B., Billard, L., Conquest, L., Ryan, L., Brillinger, D., Greenhouse, J. (1994). Case Studies in Biometry. New York: Wiley.

[3] Carroll, R.J., Spiegelman, C.H., Lan, K.K., Bailey, K.T. & Abott, R.D. (1984). On errors-in-variables problem for binary regression models. Biometrika 71, 19-25.

[4] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). Measurement Error in Nonlinear Models. New York: Chapman and Hall.

[5] Chatterjee, S. and Hadi, A.S. (1986). In‡uential observations, high leverage points, and outliers in linear regression, Statistical Science 1, 379-416.

[6] Chatterjee, S. and Price, B. (1991). Regression Analysis by Example. Second Edition. New York: Wiley.

[7] Cook, R.D. and Weisberg, S. (1982). Residuals and In‡uence in Regression. New York: Chapman and Hall.

[8] Cook, R.D. (1986). Assessment of local in‡uence (with discussion). Journal of the Royal Statistical Society, ser.B. 2, 139-169.

[9] Fuller, W.A. (1987). Measurement Error Methods. New York: Wiley.

[10] Emerson, J.D., Hoaglin, D.C., and Kempthorne, P.J. (1984). Leverage in least squares additive-plus-multiplicative …ts for two-way tables. Journal of American Statistical Association 79, 329-335.

[11] Hodges, S.D. and Moore, P.G. (1972). Data uncertainties and least squares regression. Applied Statistics 21, 185-195.

[12] Johnson, W. (1985). In‡uence measures for logistic regression: Another point view. Biometrika 72, 59-65.

[13] Laurent, R.S.ST. and Cook, R.D. (1992). Leverage and superleverage in nonlinear regression. Journal of American Statistical Association 87, 985-990.

[14] Laurent, R.S.ST. and Cook, R.D. (1993). Leverage, local in‡uence and curvature in nonlinear regression. Biometrika 80, 99-106.

[15] Miller, R.G. (1974). An unbalanced jackknife. Annals of Statistics 2, 880-891.

[16] Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. (1990). Applied Linear Statistical Models. Chicago: IRWIN.

[17] Pregibon, D. (1981). Logistic regression diagnostic. Annals of Statistics 4, 705-724.

[18] Preisser, J.S. and Qaqish, B.F. (1996). Deletion diagnostics for generalized estimating equations, Biometrika 83, 551-562.

[19] Ross, W.H. (1987). The geometry of case deletion and the assessment of in‡uence in nonlinear regression. The Canadian Journal of Statistics 15, 91-103.

[20] Tosteson, T., Stefanski, L.A. & Scha¤er, D.W. (1989). A measurement error model for binary and ordinal regression. Statistics in Medicine 8, 1139-47.