The Korean Communications in Statistics **Vol. 14** No. 2, 2007, pp. 401–411

Some Results on the Log-linear Regression Diagnostics*

Miyoung Yang[1] Jimin Choi[2] and Choongrak Kim[3]

Abstract

In this paper we propose an influence measure for detecting potentially influential observations using the infinitesimal perturbation and the local influence in the log-linear regression model. Also, we propose a goodness-of-fit measure for variable selection. A real data set are used for illustration.

Keywords: Cook's distance; goodness-of-fit; influential perturbations; robust regression.

1. Introduction

In fitting a regression model, quantities like estimates of unknown parameters and overall goodness of fit can be substantially influenced by one or few observations. It is, therefore, important for an analyst to be able to identify such observations and assess their effects on various aspects of analysis. Generally, two approaches which attempt to quantify the effect of individual observations have been investigated: assessment by deletion (Cook, 1977; Andrews and Pregibon, 1978; Cook and Weisberg, 1982; Belsley *et al.*, 1980) and assessment by influential perturbations (Belsley *et al.*, 1980). In the assessment by deletion approach, one computes the change in some aspects of the fit incurred by deleting one or more data points. For single deletions (say the i^{th} observation), the effect is based

* This work was supported for two years by Pusan National University Research Grant

1) Member of Statistical Research Institute, Department of Statistics, Pusan National University, Jangjeon-dong, Geumjeong-gu,Busan 609-735, Korea.
E-mail : whymey@hanmail.net

2) Member of Statistical Research Institute, Pusan National University, Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea.
E-mail : stat.choi@gmail.com

3) Professor, Department of Statistics, Pusan National University, Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea.
Correspondence : crkim@pusan.ac.kr

on $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$, where $\hat{\boldsymbol{\beta}}_{(i)}$ is the estimate of $\boldsymbol{\beta}$ based on $n-1$ points without the i^{th} observation. The infinitesimal perturbation approach is based on the influence curve, a construction that relies on an appropriate functional of the true underlying distribution function. Much of the recent work is concerned with only the perturbation scheme in which the weights attached to individual or groups of cases are modified. For the most part, the case-weights are restricted to be either 0 or 1 so that a case is either deleted or retained at full weight. These ideas are adapted for use in logistic regression by Pregibon (1981). An important extension of these diagnostic approaches is to nonlinear regression model, where presumably the effects of outliers and leverage points could be worse. Also, we need the deletion of more than one observation due to the masking effect (see, Cook (1977) for details).

On the other hand, a robust regression method is also used to remove the effect of outlying observations by specifying appropriate influence function. This method reduces the effect of outlying observations by giving smaller weights than normal observations at the stage of fitting the model. See Huber (1981), Hampel *et al.* (1986) and Maronna *et al.* (2006) for details on the robust regression approach.

This paper proposes diagnostic measures which should accompany the "usual" output from a maximum likelihood fit of a log linear regression model, and suggests a robust goodness of fit measure to select good models. The proposed measures can be easily extended to other type of generalized linear models such as logistic regression model and constant coefficient of variation model by specifying appropriate link function. Since the analytic expression for a diagnostic in the log linear regression model is not available, one-step estimation is often used. Also, multiple cases deletion is considered in this paper. In Section 2, we introduce the model and the relevant notation. Section 3 is concerned with detecting influential observations and sets via Cook's distance (Cook, 1977) and the influence curve approach. A diagnostic measure which can simultaneously be used for goodness-of-fit and detection of influential observations is suggested in Section 4, and Section 5 deals with an example to illustrate influence diagnostics and goodness-of-fit measure. Finally, Section 6 gives remarks and future research.

2. Background and Notation

2.1. Log linear regression model

Consider a sample $\mathbf{y}' = (y_1, y_2, \ldots, y_n)$ of independent random variables such

that y_i has Poisson distribution $P(\mu_i)$ with μ_i unknown. If we let $\boldsymbol{\eta} = \log \boldsymbol{\mu}$, the log linear regression model utilizes the relationship

$$\eta_i = \log \mu_i = \mathbf{x}_i' \boldsymbol{\beta}, \quad i = 1, 2, \ldots, n,$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are p-vectors of explanatory variables and $\boldsymbol{\beta}$ is an unknown parameter vector. The log likelihood for $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ is

$$L(\boldsymbol{\eta}) = \sum_{i=1}^{n} \{y_i \boldsymbol{x}_i' \boldsymbol{\beta} - a_i(\boldsymbol{x}_i' \boldsymbol{\beta}) + b_i(y_i)\}, \tag{2.1}$$

where $a_i(\eta_i) = \mu_i$ and $b_i(y_i) = -\log(y_i!)$.

2.2. Estimation

The maximum likelihood estimator (MLE) is obtained by maximizing (2.1) and is a solution to $(\partial/\partial\boldsymbol{\beta})L(\boldsymbol{\eta}) = 0$. In particular, $\hat{\boldsymbol{\beta}}$ satisfies the system of equations:

$$\sum_{i=1}^{n} x_{ij}(y_i - \dot{a}(\boldsymbol{x}_i' \hat{\boldsymbol{\beta}})) = 0, \quad j = 1, 2, \ldots, p,$$

where $\dot{a}(x)$ denotes the first derivative of $a(x)$. Writing $\boldsymbol{e} = \boldsymbol{y} - \dot{a}(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \mathbf{y} - \hat{\boldsymbol{\mu}}$, the matrix formulation of the likelihood equations is $\boldsymbol{X}'(\mathbf{y} - \hat{\mathbf{y}}) = 0$.

The MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is often found using Newton's method since the likelihood equations are nonlinear in $\boldsymbol{\beta}$. By the Newton-Raphson method,

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t + (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \hat{\mathbf{y}}), \quad t = 0, 1, \ldots,$$

where $\mathbf{V}$ be an $n \times n$ diagonal matrix with i^{th} diagonal $v_i = \hat{\mu}_i$. Let

$$\mathbf{z}^t = \mathbf{X}\boldsymbol{\beta}^t + \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) \tag{2.2}$$

be a psuedo vector evaluated at the t^{th} iteration, then

$$\boldsymbol{\beta}^{t+1} = (\mathbb{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{z}^t.$$

At convergence, we have $\mathbf{z} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbb{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$, and, therefore, we may write the MLE of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{z}$. This method is referred to as the iteratively reweighted least squares (IRLS) since the form of the estimator resembles the weighted least squares estimator in the classical linear models.

3. Regression Diagnostics

3.1. Deletion method

To see the influence of the i^{th} observation on $\hat{\boldsymbol{\beta}}$, it is useful to evaluate $\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}$. However, the analytic form for $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ is not available since iterative methods are required to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$. To overcome this difficulty, Pregibon (1981) suggested a one-step estimator $\hat{\boldsymbol{\beta}}^1_{(i)}$ of $\hat{\boldsymbol{\beta}}_{(i)}$ for the influence measure of Cook's distance (Cook, 1977) type. To be more specific, let $\hat{\mu}_i = \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})$ and let $\mathbf{r}$ be an n-vector with the i^{th} element $r_i = e_i/\sqrt{v_i}$ where $e_i = y_i - \hat{y}_i$. Pregibon (1981) showed that

$$\hat{\boldsymbol{\beta}}^1_{(i)} = \hat{\boldsymbol{\beta}} - \frac{(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}x_i e_i}{1 - h^*_{ii}}, \tag{3.1}$$

where h^*_{ii} is the i^{th} diagonal element of $\mathbf{H}^* = \mathbf{V}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{1/2}$. Pregibon (1981) discussed the accuracy of this one-step approximation and concluded that componentwise the approximation tends to underestimate the fully iterated value, but that this may be unimportant for identifying influential cases. Using (3.1), it can be easily shown that the one-step version of Cook's distance is

$$\begin{aligned} D_i &= (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(i)})'\mathbf{X}'\mathbf{V}\mathbf{X}((\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(i)})/p \\ &= r_i^2 h^*_{ii}/\{(1 - h^*_{ii})^2 p\} \\ &= \{(y_i - \hat{\mu}_i)/\sqrt{\hat{\mu}_i}\}^2 h^*_{ii}/\{(1 - h^*_{ii})^2 p\}. \end{aligned}$$

We extend the Cook's distance to the influence of set of observations in $I = \{i_1, \ldots, i_m\}$, then, following to the notation for influential observations $\mathbf{X}_{(I)}$ is an $(n - m) \times p$ matrix omitting the observations in I. Also, let $\mathbf{r}_I$ denote the m-subvector of $\mathbf{r}$ corresponding to cases in I and $\mathbf{H}^*_I$ is $m \times m$ submatrix of $\mathbf{H}^*$. Using this notation, Cook's distance for the observations in I becomes

$$\begin{aligned} D_I &= (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(I)})'\mathbf{X}'\mathbf{V}\mathbf{X}((\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^1_{(I)})/p \\ &= \mathbf{r}_I'(\mathbf{I} - \mathbf{H}^*_I)^{-1}\mathbf{H}^*_I(\mathbf{I} - \mathbf{H}^*_I)^{-1}\mathbf{r}_I/p. \end{aligned}$$

It is necessary to evaluate the influence of set of observations because of masking effect. However, the computation of D_I is quite expensive as m increases.

3.2. Infinitesimal perturbation method

The infinitesimal perturbation approach in the multiple linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ for the effect of the i^{th} observation is obtained by specifying $\epsilon_i \sim$

$N(0, \sigma^2/w_i)$, where $0 \leq w_i \leq 1$. According to this specification, the normal equations are modified as $\mathbf{X}'\mathbf{W}(\mathbf{y} - \hat{\mathbf{y}}) = 0$ with $\mathbf{W} = \text{diag}\{1, \ldots, 1, w_i, 1, \ldots, 1\}$, and one obtains $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_w$, where $\hat{\boldsymbol{\beta}}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$. Pregibon (1981) showed that

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_w = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i(1-w)e_i}{\{1-(1-w)h_{ii}\}}.$$

The effect of infinitesimal perturbations of the variance of the i^{th} data point is easily obtained by differentiation of $\hat{\boldsymbol{\beta}}_w$ leading to

$$\Delta\hat{\boldsymbol{\beta}}_w = \frac{\partial}{\partial w}\hat{\boldsymbol{\beta}}_w = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{e}_i}{\{1-(1-w)h_{ii}\}^2}$$

where h_{ii} is the i^{th} diagonal element of $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Note that $w_i \to 0$ implies $\hat{\boldsymbol{\beta}}_w \to \hat{\boldsymbol{\beta}}_{(i)}$, therefore, assessment by deletion is a special case of assessment by infinitesimal perturbations.

In the log linear regression model, consider a set of poisson data $\{y_1, \ldots, y_n\}$, the log likelihood of which may be written as the sum of n individual contributions, $l(\mathbf{x}_i'\boldsymbol{\beta}; y_i)$. The log likelihood function may be expressed as

$$L_w(\boldsymbol{\eta}) = \sum_{i=1}^{n} w_i l(\mathbf{x}_i'\boldsymbol{\beta}; y_i). \tag{3.2}$$

For the infinitesimal perturbation of the i^{th} observation, let

$$w_j = \begin{cases} w & \text{if } j = i \\ 1 & \text{otherwise,} \end{cases},$$

with $0 \leq w \leq 1$. Then, the MLE of $\boldsymbol{\beta}$ becomes a function of w, and this is equivalent to solving the equation: $\mathbf{X}'\mathbf{W}(\mathbf{y} - \hat{\mathbf{y}}) = 0$. By the Newton-Raphson iteration, we have

$$\boldsymbol{\beta}^{t+1}(w) = \boldsymbol{\beta}^t(w) + (\mathbf{X}'\mathbb{V}^{1/2}\mathbf{W}\mathbf{V}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{y} - \hat{\mathbf{y}}), \tag{3.3}$$

and, after one-step, equation (3.3) becomes

$$\boldsymbol{\beta}^1(w) = (\mathbf{X}'\mathbf{V}^{1/2}\mathbf{W}\mathbf{V}^{1/2}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{1/2}\mathbf{W}\mathbf{V}^{1/2}\mathbf{z}.$$

Pregibon (1981) showed that,

$$\boldsymbol{\beta}^1(w) = \hat{\boldsymbol{\beta}} - \frac{(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_i(1-w)e_i}{1-(1-w)h_{ii}^*}.$$

Note that $w = 0$ corresponds to deleting the i^{th} observation, and $w = 1$ corresponds to the usual maximum likelihood fit. The influence of the i^{th} observation can be evaluated at various values of w. Based on the motivation by the local influence (Cook, 1986) and the replacement measure (Kim, 1996), we can evaluate the influence of the i^{th} observation by the derivative of $\hat{\boldsymbol{\beta}}^1(w)$ with respect to w at $w = 1$, *i.e.*,

$$\Delta\hat{\boldsymbol{\beta}}^1(w) = \frac{\partial}{\partial w}\hat{\boldsymbol{\beta}}^1(w) = \frac{(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_i e_i}{\{1-(1-w)h_{ii}^*\}^2},$$

and let $\Delta\hat{\boldsymbol{\beta}}^1(1)$ be the value of $\Delta\hat{\boldsymbol{\beta}}^1(w)$ evaluated at $w = 1$. Therefore,

$$\Delta\hat{\boldsymbol{\beta}}^1(1) = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_i\mathbf{e}_i. \tag{3.4}$$

Using (3.4), we can define an influence measure for the i^{th} observation, the one-step version of influence curve, as

$$\begin{aligned} R_i &= (\Delta\hat{\boldsymbol{\beta}}^1(1))'\mathbf{X}'\mathbf{V}\mathbf{X}(\Delta\hat{\boldsymbol{\beta}}^1(1))/p \\ &= {r_i}^2 h_{ii}^*/p \\ &= \{(y_i-\hat{\mu}_i)/\sqrt{\hat{\mu}_i}\}^2 h_{ii}^*/p. \end{aligned} \tag{3.5}$$

Also, for the influence of set of observations in $I = \{i_1, \ldots, i_m\}$, we have

$$R_I = \mathbf{r}_I'\mathbf{H}_I^*\mathbf{r}_I/p.$$

The proposed influence measure R_i in (3.5) can be interpreted as the generalized linear model diagnostic version of the local influence and the replacement measure in the classical linear model diagnostics (see, Kim (1996) for detailed discussion).

4. Goodness of Fit Measures

4.1. Deviance and Pearson's Chi-square

In the generalized linear models, two important measures for the goodness of fit are the deviance and the Pearson χ^2 statistics. The deviance is twice the difference between the maximum log likelihood achievable and that achieved by the model under investigation. If we denote by $\hat{\theta} = \theta(\hat{\boldsymbol{\mu}})$ and $\tilde{\theta} = \theta(\mathbf{y})$ by estimates of the canonical parameters under the two models, the deviance is

$$D(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\sum\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\}/a_i(\phi).$$

The Pearson χ^2 statistic is defined as

$$\chi^2 = \sum(y_i - \hat{\mu})^2/V(\hat{\mu}_i),$$

where $V(\hat{\mu})$ is the estimated variance function for the distribution concerned.

Both the deviance and Pearson χ^2 have exact χ^2 distributions for normal-theory models, and asymptotic results are available for the other distributions. However, asymptotic results may not be specially relevant to statistics calculated from limited amounts of data, and for these either D or χ^2 may prove superior in its distributional properties. The deviance has a general advantage as a measure of discrepancy in that it is additive for nested sets of models if maximum likelihood estimates are used, whereas χ^2 in general is not. However, χ^2 may sometimes be preferred because of easier interpretation. For the asymptotic behavior of D or χ^2, see Pierce and Schafer (1986) and McCullagh and Nelder (1989).

4.2. A robust goodness of fit measure

In Section 3, we suggested an influence measure R_i for the diagnostics on $\hat{\boldsymbol{\beta}}$. Let $\Lambda = \Sigma R_i$, then Λ is the total influence of observations on $\hat{\boldsymbol{\beta}}$. In this sense, Λ can be regarded as a goodness of fit measure, and thus it can be used a criterion for variable selection. If we choose variables such that Λ is small enough, then the model based on variables chosen by Λ would be robust. Note that

$$\Lambda = \sum r_i^2 h_{ii}^* / p$$

and

$$\chi^2 = \sum r_i^2 \ ,$$

therefore Λ is a weighted version of the Pearson χ^2. Since the Pearson χ^2 is usually compared with χ^2 distribution and $h_{ii}^* \approx p/n$, we better use

$$C_\Lambda = n\Lambda = \frac{n}{p} \sum r_i^2 h_{ii}^*$$

than Λ to compare C_Λ with χ^2 distribution (see, Pierce and Schafer (1986) for details). Note that C_Λ can be used both for the evaluation of each observation and for the overall goodness of fit since it is a weighted average of the influence of each observation.

5. Example

Fisher (1949) published data set consisting of 16 measurements of tuberculin response to four treatments which we call W, X, Y, Z. These were applied in a

Latin square design so that the effects were not confounded with type of cow and site. The data were as follows with the corresponding treatment in Table 5.1 (see, Baker and Nelder (1978) for more details).

Table 5.1: Fisher's data on tuberculin response

site	cow class			
	1	2	3	4
1	454 (W)	249 (X)	349 (Y)	249 (Z)
2	408 (X)	322 (W)	312 (Z)	347 (Y)
3	523 (Y)	268 (Z)	411 (W)	285 (X)
4	364 (Z)	283 (Y)	266 (X)	290 (W)

Fisher believed that the variances of the observations were proportional to their expectation and that the systematic part of the model was linear on the log-scale, indicating the use of the log link with Poisson errors. The treatments were considered as arising from a 2×2 factorial arrangement indexed by two factors A, B in the way described in Table 5.2.

Table 5.2: The treatments indexed by two factors A and B

		A	
		1	2
B	1	W	X
	2	Y	Z

We fit the data to the log linear model using the GLIM. Criteria for selecting best model are three goodness of fit measures: deviance D, Pearson χ^2, and the proposed measure C_Λ defined in Section 4, and the results are summarized in Table 5.3. As is clear from the table, the best parsimonious model is

$$\text{site} + \text{cow} + \text{B}$$

and three criteria show very similar results.

For the influence of observations in the selected model, we evaluate D_I and R_I. Regard the data set in Table 5.1 as 4×4 matrix and apply *vec* operation to define the k^{th} observation. For example, 249 is the 5^{th} observation. Table 5.4 shows five largest values for $m = 1, 2, 3, 4$ defined in Section 3.1, and observations 4 and 5 seem to be potentially influential, however, they are not so apparent to be regarded as influential. Other observations detected when $m = 3, 4$ are due to the swamping phenomenon, but are not influential. To see the effect of the

Table 5.3: Goodness of fit measures D, χ^2, and C_Λ

model	d.f.	deviance(D)	Pearson (χ^2)	C_Λ	$\chi^2_{df}(.95)$
null(1)	15	265.30	278.73	278.73	25.00
A	14	265.28	278.69	278.69	23.68
B	14	203.09	210.85	210.84	23.68
A+B	13	203.07	210.79	212.72	22.36
site	12	232.38	238.09	228.60	21.03
cow	12	91.76	92.43	98.51	21.03
site+A	11	232.37	238.14	238.66	19.68
cow+A	11	91.74	92.39	98.16	19.68
site+B	11	170.17	177.60	174.73	19.68
cow+B	11	29.55	29.47	30.78	19.68
site+A+B	10	170.15	177.45	175.18	18.31
cow+A+B	10	29.53	29.48	30.68	18.31
site+cow	9	58.84	58.78	55.27	16.92
site+cow+A	8	58.82	58.74	56.01	15.51
site+cow+B	8	1.41	1.42	1.35	15.51
site+cow+A+B	7	1.40	1.41	1.30	14.07

Table 5.4: Five largest values of D_I and R_I for $m = 1, 2, 3, 4$

m	set	D_I	set	R_I
1	4	0.05142	4	0.02078
	5	0.02540	5	0.01646
	1	0.01779	11	0.01005
	14	0.01719	14	0.00950
	11	0.01491	12	0.00626
2	4,5	0.09720	4,5	0.04248
	4,9	0.09588	11,14	0.03464
	11,14	0.09196	4,11	0.03103
	5,16	0.07955	5,16	0.03095
	1,4	0.07940	1,4	0.02840
3	1,4,9	0.22912	4,5,16	0.05617
	2,4,5	0.21168	4,5,11	0.04538
	1,3,6	0.18618	4,5,9	0.04491
	1,6,9	0.15808	5,11,14	0.04456
	4,5,16	0.14920	4,5,8	0.04289
4	1,3,6,9	0.55832	4,5,9,16	0.05931
	1,4,6,9	0.44201	4,5,6,16	0.05720
	2,4,5,16	0.42023	4,5,11,14	0.05712
	1,2,3,6	0.40121	4,5,15,16	0.05539
	1,4,9,13	0.33831	4,5,8,16	0.05533

Table 5.5: Comparison of estimates based on the full data set and the data set after deleting the 4^{th} observation, respectively

	full data set			case 4 deleted		
parameter	estimate	S.E.	p value	estimate	S.E.	p value
intercept	5.4753	0.0404	<.0001	5.4747	0.0404	<.0001
site 1	0.0480	0.0403	0.2340	0.0337	0.0443	0.4468
site 2	0.1438	0.0394	0.0003	0.1428	0.0394	0.0003
site 3	0.1816	0.0391	<.0001	0.1797	0.0392	<.0001
cow 1	0.3967	0.0379	<.0001	0.3974	0.0379	<.0001
cow 2	-0.0427	0.0418	0.3062	-0.0570	0.0457	0.2125
cow 3	0.1289	0.0401	0.0013	0.1295	0.0401	0.0012
B 1	0.2095	0.0277	<.0001	0.2171	0.0295	<.0001

influential observation (here we take the 4^{th} observation) on estimates, we list both estimates based on the full data set and the data set after deleting the 4^{th} observation, respectively, in Table 5.5. We see that some estimates (especially "site1") change significantly even though we delete just one observation.

6. Remarks and Future Research

In this paper, we mentioned two issues in the log linear model; regression diagnostics and goodness of fit measure. In the log linear regression diagnostics, influence measures based on the influence curve derived from the infinitesimal perturbation approach, and the replacement method. These measures can be easily extended to the subset deletion for detecting the masking effect. We compared these measures with the one-step version of the Cook's distance. As goodness of fit measures in the generalized linear models, the deviance and the Pearson χ^2 are often used. We suggest a robust goodness of fit measure.

For future research, two issues should be studied. First, some reference values for the influence measure R_I is worth pursuing. It can be a function of the number of observations, the number of independent variables, and the number of cases deleted. Possible approaches are a Monte Carlo study in the linear regression diagnostics by Kim and Storer (1996) and a bootstrap approach in the nonlinear or nonparametric model by Kim *et al.* (2001). Second, we need more justifications for the robust goodness of fit measure although it has a meaning of minimizing the sum of case influences. For example, sensitivity analysis to see the robustness of C_Λ should be done.

References

Andrews, D. F. and Pregibon, D. (1978). Finding outliers that matter. *Journal of the Royal Statistics Society*, Ser. B, **40**, 85–93.

Baker, R. J. and Nelder, J. A. (1978). *The GLIM System. Release 3, Generalized Linear Interactive Modelling.* Numerical Algorithms Group, Oxford.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980) *Regression Diagnostics: Identifying Influential Data and Source of Collinearity.* John Wiley & Sons, New York.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **48**, 133–169.

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

Fisher, R. A. (1949). A biological assay of tuberculins. *Biometrics*, **5**, 300–316.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P.J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York.

Huber, P. J. (1981). *Robust Statistics.* John Wiley & Sons, New York.

Kim, C. (1996). Local influence and replacement measure. *Communications in Statistics - Theory and Methods*, **25**, 49–61.

Kim, C. and Storer, B. E. (1996). Reference values for Cook's distance. *Communications in Statistics - Simulation and Computation*, **25**, 691–708.

Kim, C., Lee, Y. and Park, B-U. (2001). Cook's distance in local polynomial regression. *Statistics & Probability Letters*, **54**, 33–40.

Maronna, R. A., Martin, D. R. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods.* John Wiley & Sons, New York.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd ed., Chapman & Hall/CRC, New York.

Pierce, D. A. and Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, **81**, 977–986.

Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, **9**, 705–724.

[Received April 2007, Accepted July 2007]