Journal of the Korean
Statistical Society
Vol. 21, No. 1, 1992

DETECTING INFLUENTIAL OBSERVATIONS ON TRANSFORMATION PARAMETER IN BOX-COX MODEL[1]

Choongrak Kim[2] and Meeseon Jeong[3]

ABSTRACT

On Box-Cox transformation, one or few responses are influential on transformation parameter estimator. To detect influential observations, several diagnostics (Cook and Wang 1983, Hinkley and Wang 1988, Lawrance 1988, Tsai and Wu 1990) have been suggested. We compare these diagnostics and denote the necessity of multiple cases deletion which is important especially when the masking effect is present. Also, analytic expression of Tsai and Wu's diagnostic is given. We suggest a computationally feasible and useful algorithm based on the basic building blocks, and present descriptive examples using artificial data.

1. INTRODUCTION

When errors in regression models are not normally distributed, transformations of the response are considered. Let y denote an n-vector of observable responses, $\boldsymbol{\beta}$ be a p-vector of unknown parameters, and X be a $n \times p$ design matrix. Box and Cox (1964) suggested power transformation which assumes that the power transformed response

[1]This research was supported by the Ministry of Education (90-91)
[2]Department of Statistics, Pusan National University, Pusan, 609-735, Korea
[3]Department of Statistics, Pusan National University, Pusan, 609-735, Korea

$$\mathbf{y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

where

$$y_i^{(\lambda)} = \begin{cases} (y_i^{\lambda} - 1)/\lambda & , \lambda \neq 0 \\ \log y_i & , \lambda = 0 \end{cases} \tag{1.2}$$

is normally distributed. The maximum likelihood estimator(MLE) $\hat{\lambda}$ of transformation parameter λ is obtained by minimizing

$$\mathbf{Q}(\lambda) = \mathbf{Z}^{(\lambda)^t}(\mathbf{I} - \mathbf{H})\mathbf{Z}^{(\lambda)} \tag{1.3}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, $\mathbf{Z}^{(\lambda)} = \mathbf{y}^{(\lambda)}/J^{1/n}$ and J denote the Jacobian of the transformation from $\mathbf{y}$ to $\mathbf{y}^{(\lambda)}$. In fact, computation of $\hat{\lambda}$ requires iterative minimization of $\mathbf{Q}$.

As pointed out by Andrews (1971), $\hat{\lambda}$ is very sensitive to outlying responses and influential observations. To identify cases that influence $\hat{\lambda}$, Cook and Wang (1983) suggested one-step diagnostic using the mean-shift outlier model. Hinkley and Wang (1988) modified Cook and Wang's method, and provided an alternative power estimator. Tsai and Wu (1990), however, argued that both methods do not explain the deletion effects on J, therefore, accuracy can be rarely achieved. Instead, they considered diagnostic using the deletion model which takes into account the deletion effect on J, and claimed that diagnostic based on the deletion model could provide more accurate and reliable transformation power estimator. Further, they insisted that their estimator almost always converges to the true estimator. A quite different method using Cook's (1986) local influence approach was done by Lawrance (1988).

Ideally, Tsai and Wu's (TW) method should be better than Cook and Wang's (CW) or Hinkley and Wang's (HW). However, it is not necessarily always true since all three methods are based on $\hat{\lambda}$ which itself may not be true. Also, Tsai and Wu's suggestion have several defaults. First, their estimator should be calculated by the symbolic manipulation programs MACSYMA or SMP, i.e., analytic expression was not given. Secondly, the accuracy of the estimator is not guaranteed when k-step iteration ($k = 4$ in their paper) is not allowed. Therefore, computational feasibility could be a serious problem in this approach. Finally, they did not consider multiple cases deletion. When observations are jointly influential, case-deletion is not enough to detect jointly influential observations (influential set).

In this paper, we provide an analytic expression of Tsai and Wu's estimator and extend it to the multiple cases deletion. Also, we suggest some practical uses of the estimator providing basic building blocks of which the estimator consists. In Section 2, summaries of four estimators are given and an analytic form of Tsai and Wu's estimator is derived.

Multiple cases deletion and basic building blocks are considered based on artificial data in Section 3. Section 4 gives our concluding comments.

2. DIAGNOSTICS IN TRANSFORMATION PARAMETER

2.1 Four Diagnostics

To detect influential observations, we generally compare $\hat{\lambda}$ with $\hat{\lambda}_{(I)}$, where $\hat{\lambda}_{(I)}$ is the maximum likelihood estimator of λ based on $n-m$ observations after deleting m observations in $I = \{i_1, i_2, \ldots, i_m\}$. However it is computationally difficult because the Jacobian J differs for each value of I. To overcome this difficulty, Cook and Wang (1983) used the mean-shift outlier model,

$$\mathbf{y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}_I\gamma + \boldsymbol{\varepsilon} \tag{2.1.1}$$

where $\mathbf{E}_I$ is a $n \times m$ matrix containing 1 in the i_jth position as indexed by I and 0's elsewhere and γ is unknown constructed variable. Under the outlier model, the MLE of λ is obtained by minimizing

$$\mathbf{Q}^C(\lambda) = \mathbf{Z}^{(\lambda)^t}(\mathbf{I} - \mathbf{H}_E)\mathbf{Z}^{(\lambda)} \tag{2.1.2}$$

where $\mathbf{H}_E = \mathbf{X}_E({\mathbf{X}_E}^t\mathbf{X}_E)^{-1}{\mathbf{X}_E}^t$ and $\mathbf{X}_E = (\mathbf{X}, \mathbf{E}_I)$. To obtain one-step estimator of the MLE, $\mathbf{Z}^{(\lambda)}$ in $\mathbf{Q}^C(\lambda)$ is replaced by its first order Taylor expansion about $\hat{\lambda}$,

$$\mathbf{Z}^{(\lambda)} \simeq \mathbf{Z}^{(\hat{\lambda})} + (\lambda - \hat{\lambda})\mathbf{W}^{(\hat{\lambda})}$$

where $\mathbf{W}^{(\lambda)} = \partial\mathbf{Z}^{(\lambda)}/\partial\lambda$. The corresponding one-step estimator $\hat{\lambda}_{(I)}^{CW}$ can be expressed as

$$\hat{\lambda}_{(I)}^{CW} = \hat{\lambda} + \frac{\mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I}}{\mathbf{r}_W{}^t\mathbf{r}_W - \mathbf{r}_{W,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I}} \tag{2.1.3}$$

where $\mathbf{H}_I$ is the $m \times m$ submatrix of $\mathbf{H}$ indexed by I, $\mathbf{r}_{Z,I}$ and $\mathbf{r}_{W,I}$ are the indicated $m \times 1$ subvectors of $\mathbf{r}_Z = (\mathbf{I} - \mathbf{H})\mathbf{Z}^{(\hat{\lambda})}$ and $\mathbf{r}_W = (\mathbf{I} - \mathbf{H})\mathbf{W}^{(\hat{\lambda})}$, respectively.

Hinkley and Wang (1988) also used the mean-shift outlier model and considered quadratic approximation providing

$$\mathbf{Z}^{(\lambda)} \simeq \mathbf{Z}^{(\hat{\lambda})} + (\lambda - \hat{\lambda})\mathbf{W}^{(\hat{\lambda})} + (\lambda - \hat{\lambda})^2\mathbf{U}^{(\hat{\lambda})}/2$$

where $\mathbf{U}^{(\lambda)} = \partial^2\mathbf{Z}^{(\lambda)}/\partial\lambda^2$. Then, they obtained the modified one-step estimator

$$\hat{\lambda}_{(I)}^{HW} = \hat{\lambda} + \frac{\mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I}}{\mathbf{r}_W{}^t\mathbf{r}_W+\mathbf{r}_Z{}^t\mathbf{r}_U-\mathbf{r}_{W,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I}-\mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{U,I}} \tag{2.1.4}$$

where $\mathbf{r}_{U,I}$ is the indicated $m \times 1$ subvector of $\mathbf{r}_U = (\mathbf{I}-\mathbf{H})\mathbf{U}^{(\hat{\lambda})}$. Since $\hat{\lambda}_{(I)}^{CW}$ and $\hat{\lambda}_{(I)}^{HW}$ do not explain the Jacobian effect for each value I, they may not be accurate when the observations in I are outliers or influential cases.

To obtain $\hat{\lambda}_{(I)}$, Tsai and Wu (1990) used the case-deletion model

$$\mathbf{y}_{(I)}^{(\lambda)} = \mathbf{X}_{(I)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(I)}.$$

The MLE $\hat{\lambda}_{(I)}$ of $\lambda_{(I)}$ from case-deletion model minimizes

$$\mathbf{Q}_{(I)}(\lambda) = \left(\prod_{i\in I} y_i\right)^{2(\lambda-1)/(n-m)} \left\{\tilde{\mathbf{Z}}^{(\lambda)^t}(\mathbf{I}-\mathbf{H}_E)\tilde{\mathbf{Z}}^{(\lambda)}\right\} \tag{2.1.5}$$

where $\tilde{\mathbf{Z}}^{(\lambda)} = \mathbf{y}^{(\lambda)}/J^{1/(n-m)}$. Then the one-step estimator of $\hat{\lambda}_{(I)}$ is given by

$$\hat{\lambda}_{(I)}^{TW} = \hat{\lambda} - \frac{\dot{\mathbf{Q}}_{(I)}(\hat{\lambda})}{\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda})} \tag{2.1.6}$$

where $\dot{\mathbf{Q}}_{(I)}(\lambda) = \partial\mathbf{Q}_{(I)}(\lambda)/\partial\lambda$ and $\ddot{\mathbf{Q}}_{(I)}(\lambda) = \partial^2\mathbf{Q}_{(I)}(\lambda)/\partial\lambda^2$, by using the first order Taylor expansion of $\dot{\mathbf{Q}}_{(I)}(\lambda)$ about $\hat{\lambda}$,

$$\dot{\mathbf{Q}}_{(I)}(\lambda) \simeq \dot{\mathbf{Q}}_{(I)}(\hat{\lambda}) + (\lambda-\hat{\lambda})\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda}).$$

On the other hand, Lawrance (1988) obtained diagnostic using local influence approach of Cook (1986) instead of deleting cases. The diagnostic for the influence of the ith case is

$$l_{(i)} = \frac{\mathbf{r}_{Z,i}\mathbf{r}_{W,i}}{\sqrt{\sum_{i=1}^{n}(\mathbf{r}_{Z,i}\mathbf{r}_{W,i})^2}}, \quad i = 1, 2, \ldots, n \tag{2.1.7}$$

which is the case direction that has the greatest effect when the constant model variances are perturbed.

2.2 Analytic Expression of $\hat{\lambda}_{(I)}^{TW}$

Tsai and Wu insisted that their estimator provides a more accurate approximation than $\hat{\lambda}_{(I)}^{CW}$ and $\hat{\lambda}_{(I)}^{HW}$, through results for artificial data, and that the calculation of $\hat{\lambda}_{(I)}^{TW}$ is not difficult because $\dot{\mathbf{Q}}_{(I)}(\lambda)$ and $\ddot{\mathbf{Q}}_{(I)}(\lambda)$ can be computed analytically with the symbolic manipulation program MACSYMA or SMP. Similar to $\hat{\lambda}_{(I)}^{CW}$ and $\hat{\lambda}_{(I)}^{HW}$, however, it would be very useful to express $\hat{\lambda}_{(I)}^{TW}$ as a function of basic building blocks. Using the following notations

$$\begin{aligned}
S_{ZZ,I} &= \mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{Z,I} \\
S_{WW,I} &= \mathbf{r}_{W,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I} \\
S_{ZW,I} &= \mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{W,I} \\
S_{ZU,I} &= \mathbf{r}_{Z,I}{}^t(\mathbf{I}-\mathbf{H}_I)^{-1}\mathbf{r}_{U,I} \\
G_I &= \log\left(\textstyle\prod_{i\in I} y_i/\dot{y}^m\right)/(n-m)
\end{aligned}$$

where $\dot{y}$ is the geometric mean of y_i's, we obtained the analytic expression of Tsai and Wu's diagnostic as follows;

$$\hat{\lambda}_{(I)}^{TW} = \hat{\lambda} - \left[2G_I + \frac{\mathbf{r}_W{}^t\mathbf{r}_W + \mathbf{r}_Z{}^t\mathbf{r}_U - S_{WW,I} - S_{ZU,I} - 2G_I S_{ZW,I}}{G_I\{\mathbf{r}_Z{}^t\mathbf{r}_Z - S_{ZZ,I}\} - S_{ZW,I}}\right]^{-1} \tag{2.2.1}$$

The detailed proof is given in the Appendix. For Tsai and Wu's artificial data, the results using equation (2.2.1) equal up to third decimal of theirs.

2.3 Summaries of Four Diagnostics

For easier comparisions of the diagnostics mentioned above, note that

$$\hat{\lambda}_{(I)}^{CW} = \hat{\lambda} + \frac{S_{ZW,I}}{\mathbf{r}_W{}^t\mathbf{r}_W - S_{WW,I}}$$

$$\hat{\lambda}_{(I)}^{HW} = \hat{\lambda} + \frac{S_{ZW,I}}{\mathbf{r}_W{}^t\mathbf{r}_W + \mathbf{r}_Z{}^t\mathbf{r}_U - S_{WW,I} - S_{ZU,I}}$$

$$\hat{\lambda}_{(I)}^{TW} = \hat{\lambda} - \left[2G_I + \frac{\mathbf{r}_W{}^t\mathbf{r}_W + \mathbf{r}_Z{}^t\mathbf{r}_U - S_{WW,I} - S_{ZU,I} - 2G_I S_{ZW,I}}{G_I\{\mathbf{r}_Z{}^t\mathbf{r}_Z - S_{ZZ,I}\} - S_{ZW,I}}\right]^{-1}$$

$$l_{(i)} = \frac{\mathbf{r}_{Z,i}\mathbf{r}_{W,i}}{\sqrt{\sum_{i=1}^n (\mathbf{r}_{Z,i}\mathbf{r}_{W,i})^2}}$$

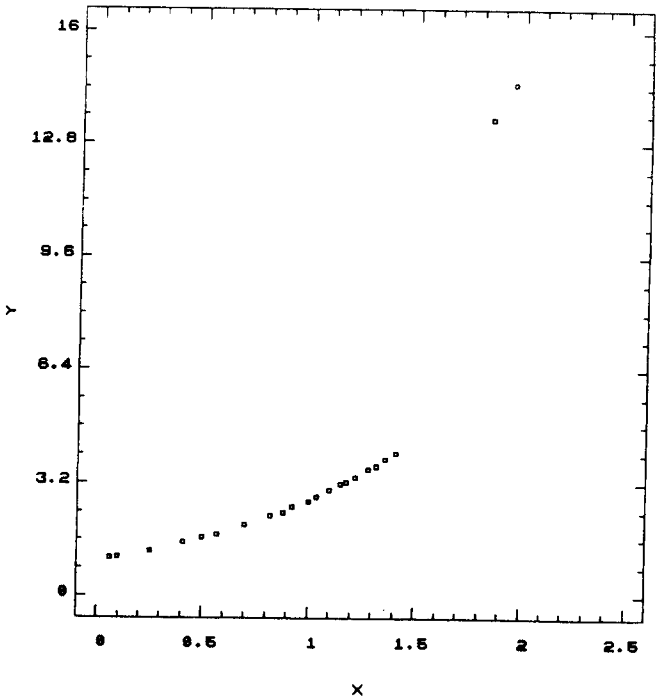
From the final form of each diagnostics, we note that $\hat{\lambda}_{(I)}^{HW}$ can be obtained by adding the information of $\mathbf{r}_Z{}^t\mathbf{r}_U - S_{ZU,I}$ to $\hat{\lambda}_{(I)}^{CW}$ and that $\hat{\lambda}_{(I)}^{TW}$ is the same as $\hat{\lambda}_{(I)}^{HW}$ if G_I is small enough. Also, $l_{(i)}$ is equal to $S_{ZW,i}$, numerator term in $\hat{\lambda}_{(I)}^{CW}$, except the leverage effect $\mathbf{H}_I$.

3. MULTIPLE CASES DELETION

3.1 Masking Effect

When a data set contains more than one influential observation, they may not be detected by a single case deletion. This phenomenon could be serious especially when they are masked. As an illustration, we create an artificial data by a method similar to that of Cook and Wang (1983). The second and third columns of Table 1 contains this data. (The remaining columns will be discussed later). These data were constructed by arbitrarily choosing 22 X values and generating the first 20 responses through the model $\log y_i = x_i + \varepsilon_i, i = 1, 2, \ldots, 20$, where the ε_i's are pseudo-random normal variables with mean 0 and variance $\sigma^2 = 0.01$. The last two responses y_{21}, y_{22} were selected so that y_{21} and y_{22} are influential set, and the log transformation is appropriate for $y_i, i = 1, 2, \ldots, 20$. Plot of these data is given in Figure 1.
For simplicity, let $d_{(I)} = \hat{\lambda} - \hat{\lambda}_{(I)}, d_{(I)}^{CW} = \hat{\lambda} - \hat{\lambda}_{(I)}^{CW}, d_{(I)}^{HW} = \hat{\lambda} - \hat{\lambda}_{(I)}^{HW}$ and $d_{(I)}^{TW} = \hat{\lambda} - \hat{\lambda}_{(I)}^{TW}$. First, we delete one observation and obtain four diagnostics ($d_{(i)}^{CW}, d_{(i)}^{HW}, d_{(i)}^{TW}$ and $l_{(i)}$) which are listed in Table 1. As shown in this table, none of them suggest that both y_{21} and y_{22} are outliers. Note that $\hat{\lambda}_{(21,22)} = 0.0172$ implying log transformation. Also, $\hat{\lambda} = -0.5434$. Therefore, true difference $d_{(21,22)}$ is -0.5606. Corresponding values for three diagnostics are $d_{(I)}^{CW} = -.48$, $d_{(I)}^{HW} = -.37$, and $d_{(I)}^{TW} = -.45$. These values are not close enough to the true difference to be used safely. Also, no one is consistently better than the others. Hence, accuracy cannot be guaranteed in these diagnostics. Tsai and Wu (1990) argued that this accuracy can be achieved by k-step estimation (4-step iteration in their paper), however, this iteration scheme is almost computationally infeasible. This infeasibility is certainly getting worse as n and m increases. Therefore, it is necessary to have some computationally feasible algorithm.

Figure 1. Artificial Data



3.2 Outliers Detecting Algorithm

As mentioned in the previous section, none of suggested diagnostics provides an accurate estimate of the transformation parameter. Therefore, we first decide candidates of potentially influential sets, and, for I's chosen as candidates, compute $\lambda_{(I)}$ by the k-step iteration using equation(2.2.1). To decide those candidates, we need some measure to obtain them. This measure could, of course, be one of the diagnostics mentioned above, however, computational burden is too much unless n is small. For example, $S_{ZW,I}$ and $S_{WW,I}$, common terms in $\hat{\lambda}^{CW}_{(I)}$, $\hat{\lambda}^{HW}_{(I)}$, and $\hat{\lambda}^{TW}_{(I)}$, require the computation of $(\mathbf{I}-\mathbf{H}_I)^{-1}$ for each I. As alternatives to $S_{ZW,I}$ and $S_{WW,I}$, we suggest using

$$t_{ZW,I} = \mathbf{r}_{Z,I}{}^t\,\mathbf{H}_I\,\mathbf{r}_{W,I}$$

$$t_{WW,I} = \mathbf{r}_{W,I}{}^t\,\mathbf{H}_I\,\mathbf{r}_{W,I}$$

based on the idea that $\mathbf{H}_I \propto (\mathbf{I}-\mathbf{H}_I)^{-1}$ in some sense. The use of $\mathbf{H}_I$ instead of $(\mathbf{I}-\mathbf{H}_I)^{-1}$ in the classical linear model diagnostics was discussed by Draper and John (1981), Cook and Weisberg (1980), and Kim (1989). The use of $t_{ZW,I}$ and $t_{WW,I}$ has a great computational advantage over the use of $S_{ZW,I}$ and $S_{WW,I}$ because $t_{ZW,I}$ and $t_{WW,I}$ can be expressed as functions of basic building blocks resulting from the estimation process of $\hat{\lambda}$. In fact,

$$t_{ZW,I} = \sum_{i\in I} \mathbf{r}_{Z,i}\mathbf{r}_{W,i}h_{ii} + \sum_{(i,j)\in I,\, i\neq j} h_{ij}(\mathbf{r}_{W,i}\mathbf{r}_{Z,j} + \mathbf{r}_{W,j}\mathbf{r}_{Z,i})$$

$$t_{WW,I} = \sum_{i\in I} \mathbf{r}_{W,i}{}^2 h_{ii} + 2\sum_{(i,j)\in I,\, i\neq j} \mathbf{r}_{W,i}\mathbf{r}_{W,j}h_{ij}$$

Now, we present an algorithm based on 2 measures $t_{ZW,I}$, $t_{WW,I}$ at each m. When $m=1$, compute $\hat{\lambda}_{(i)}$ for i's with large values of $t_{ZW,i}$ and $t_{WW,i}$. Observations with large $d_{(i)} = \hat{\lambda} - \hat{\lambda}_{(i)}$ are influential. When $m=2$, compute $\hat{\lambda}_{(I)}$ for I's with large values of $t_{ZW,I}$ and $t_{WW,I}$. Sets with large $d_{(I)} = \hat{\lambda} - \hat{\lambda}_{(I)}$ can be interpreted as follows. If I contains an observation, say i, which is already detected as influential, and $d_{(I)} \approx d_{(i)}$, then I cannot be claimed as influential since $d_{(I)}$ is large due to swamping by i. On the other hand, if I does not contain any influential observations and $d_{(I)}$ is large, this set is influential. We can proceed this step for $m = 3, 4, \cdots$.

To illustrate the algorithm easily, $\hat{\lambda}$ was -.54 suggesting inverse square root transformation, and candidates could be 1, 2, 21, 22 as shown in Table 1. But $d_{(i)}$'s for these observations are small, i.e., no observation is influential when $m=1$. For $m=2$, candidates could be (1,2) and (21,22) (see Table 2). For these sets, $d_{(1,2)} = .10$ and $d_{(21,22)} = -.56$, therefore, only set (21,22) is influential. When $m=3$, all candidates contain (21,22) as subset, and the most influential set is (9,21,22) with $d_{(9,21,22)} = -.57$ which is almost equal to $d_{(21,22)}$. Therefore, we can conclude that (21,22) is the most and the only influential set with $\hat{\lambda}_{(21,22)} = -.54 - (-.56) = .02$ suggesting log transformations.

4. CONCLUDING REMARKS

Masking effect is more serious in Box-Cox transformation model than in the classical linear model. Therefore, it is necessary to obtain the effect of multiple cases on the transformation parameter. Previous suggestions by many authors are computationally infeasible, and accuracy is not guaranteed. This article develops a useful algorithm using computable measures. Also, we derived analytic expression for the Tsai and Wu's (1990) diagnostic which can be used to estimate the correct value of $\lambda_{(I)}$ by the k-step iteration. Of course, that formula is easily implemented in FORTRAN, and it is much easier to use than the symbolic programs such as MACSYMA or SMP. As pointed out by referees, the accuracy of our algolithm cannot be guaranteed unless we use k-step iteration using the analytic expression of Tsai and Wu's diagnostic in Section 2.2. Therefore, accurate estimate of $\lambda_{(I)}$ should be calculated by the k-step iteration for those sets detected as potentially influential via our algorithm.

Table 1. One case deletion in artificial data

case	x_i	y_i	$S_{WW,i}$	$S_{ZW,i}$	$d_{(i)}$	$d_{(i)}^{CW}$	$d_{(i)}^{HW}$	$d_{(i)}^{TW}$	$l_{(i)}$
1	.06	1.07	1.90	-.33	.03	.04	.03	.03	-.46
2	.10	1.11	1.43	-.23	.02	.02	.02	.02	-.33
3	.25	1.28	.36	-.03	.00	.00	.00	.00	-.05
4	.41	1.53	.01	.01	-.00	-.00	-.00	-.00	.02
5	.50	1.67	.01	-.02	.00	.00	.00	.00	-.03
6	.57	1.76	.04	-.03	.00	.00	.00	.00	-.04
7	.70	2.04	.14	-.08	.01	.01	.01	.01	-.13
8	.82	2.27	.21	-.07	.01	.01	.01	.01	-.11
9	.88	2.37	.23	-.04	.00	.00	.00	.00	-.07
10	.92	2.52	.25	-.07	.01	.01	.01	.01	-.11
11	1.00	2.66	.25	-.01	.00	.00	.00	.00	-.02
12	1.04	2.81	.25	-.02	.00	.00	.00	.00	-.04
13	1.10	3.00	.24	-.01	.00	.00	.00	.00	-.02
14	1.15	3.18	.22	.00	.00	.00	.00	.00	-.01
15	1.18	3.22	.22	.03	-.00	-.00	-.00	-.00	.05
16	1.22	3.37	.20	.04	-.00	-.00	-.00	-.00	.06
17	1.28	3.60	.16	.05	-.00	-.00	-.00	-.00	.08
18	1.32	3.66	.15	.08	-.01	-.01	-.01	-.01	.14
19	1.36	3.87	.12	.08	-.01	-.01	-.01	-.01	.12
20	1.41	4.04	.09	.09	-.01	-.01	-.01	-.01	.14
21	1.85	13.50	3.07	.53	-.06	-.07	-.06	-.06	.75
22	1.95	14.50	3.81	.05	-.00	-.01	-.01	-.00	.07

Table 2. Four largest (in absolute values) $t_{ZW,I}, t_{WW,I}$, and $d_{(I)}$ when deleting two cases, i.e, $m = 2$

set	$t_{ZW,I}$	set	$t_{WW,I}$	set	$d_{(I)}$
(21,22)	.20	(21,22)	2.29	(21,22)	-.56
(1,2)	-.17	(1,2)	1.02	(1,2)	.10
(1,3)	-.08	(6,22)	.69	(18,21)	-.06
(10,21)	-.08	(5,22)	.69	(19,21)	-.06

APPENDIX : DERIVATION OF EQUATION (11)

$\mathbf{Q}_{(I)}(\lambda)$ in (2.1.5) can be written as

$$\begin{aligned}\mathbf{Q}_{(I)}(\lambda) &= \left(\prod_{i\in I} y_i\right)^{2(\lambda-1)/(n-m)} \times \\ &\left\{\tilde{\mathbf{Z}}^{(\lambda)t}(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}^{(\lambda)} - (\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}^{(\lambda)})^t(\mathbf{I}-\mathbf{H}_I)^{-1}(\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}^{(\lambda)})\right\}.\end{aligned}$$

Note that $\tilde{\mathbf{Z}}^{(\lambda)} = c(\lambda)^{1/2}\mathbf{Z}^{(\lambda)}$ where $c(\lambda) = J^{-2m/(n(n-m))}$.

Let $k(\lambda) = (\prod_{i\in I} y_i)^{2(\lambda-1)/(n-m)}$ and $f(\lambda) = (\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}^{(\lambda)})^t(\mathbf{I}-\mathbf{H}_I)^{-1}(\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\tilde{\mathbf{Z}}^{(\lambda)})$ by supressing the index I.

Then,

$$\mathbf{Q}_{(I)}(\lambda) = k(\lambda)\left\{c(\lambda)\,\mathbf{Z}^{(\lambda)^t}(\mathbf{I}-\mathbf{H})\mathbf{Z}^{(\lambda)} - f(\lambda)\right\}$$

and, can easily show that

$$\dot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = \mathbf{r}_Z{}^t\mathbf{r}_Z\,[k(\lambda)c(\lambda)]'_{\lambda=\hat{\lambda}} - [k(\lambda)f(\lambda)]'_{\lambda=\hat{\lambda}}$$

and

$$\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = \mathbf{r}_Z{}^t\mathbf{r}_Z\,[k(\lambda)c(\lambda)]''_{\lambda=\hat{\lambda}} - [k(\lambda)f(\lambda)]''_{\lambda=\hat{\lambda}} + 2k(\hat{\lambda})c(\hat{\lambda})\left\{\mathbf{r}_Z{}^t\mathbf{r}_U + \mathbf{r}_W{}^t\mathbf{r}_W\right\}.$$

Since $f(\lambda) = c(\lambda)p(\lambda)$ where

$$p(\lambda) = (\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\mathbf{Z}^{(\lambda)})^t(\mathbf{I}-\mathbf{H}_I)^{-1}(\mathbf{E}_I^t(\mathbf{I}-\mathbf{H})\mathbf{Z}^{(\lambda)}),$$

we have $k(\lambda)f(\lambda) = k(\lambda)c(\lambda)p(\lambda)$. Now, let $q(\lambda) = c(\lambda)k(\lambda)$, then,

$$\dot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = q'(\hat{\lambda})(\mathbf{r}_Z{}^t\mathbf{r}_Z - p(\hat{\lambda})) - p'(\hat{\lambda})q(\hat{\lambda})$$

and

$$\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = (\mathbf{r}_Z{}^t\mathbf{r}_Z - p(\hat{\lambda}))q''(\hat{\lambda}) - 2p'(\hat{\lambda})q'(\hat{\lambda}) + 2q(\hat{\lambda})(\mathbf{r}_W{}^t\mathbf{r}_W + \mathbf{r}_Z{}^t\mathbf{r}_U - p''(\hat{\lambda})/2).$$

To get explicit form of $\dot{\mathbf{Q}}$ and $\ddot{\mathbf{Q}}$, note that

$$q(\lambda) = \left(\prod_{i\in I} y_i/\dot{y}^m\right)^{2(\lambda-1)/(n-m)},$$

$$q'(\hat{\lambda}) = q(\hat{\lambda}) \log (\prod_{i \in I} y_i/\dot{y}^m) \frac{2}{n-m} = 2q(\hat{\lambda})G$$

$$q''(\hat{\lambda}) = 4q(\hat{\lambda})G^2$$

where $G = G_I$ is defined in Section 2.2. Note that

$$p(\hat{\lambda}) = S_{ZZ,I},$$

$$p'(\hat{\lambda}) = 2S_{ZW,I}$$

and

$$p''(\hat{\lambda}) = 2\{S_{WW,I} + S_{ZU,I}\}.$$

Therefore,

$$\dot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = q(\hat{\lambda}) \left\{ 2G(\mathbf{r}_Z{}^t\mathbf{r}_Z - p(\hat{\lambda})) - p'(\hat{\lambda}) \right\}$$

and

$$\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda}) = 2G\dot{\mathbf{Q}}_{(I)}(\hat{\lambda}) + 2\,q(\hat{\lambda}) \left\{ \mathbf{r}_W{}^t\mathbf{r}_W + \mathbf{r}_Z{}^t\mathbf{r}_U - \frac{p''(\hat{\lambda})}{2} - Gp'(\hat{\lambda}) \right\}.$$

Hence, we have the ratio $\dot{\mathbf{Q}}_{(I)}(\hat{\lambda})/\ddot{\mathbf{Q}}_{(I)}(\hat{\lambda})$ equal to the second factor in the right-hand member of the equation (2.1.1).

REFERENCES

(1) Andrews, D. F. (1971). A note on the selection of data transformations. *Biometrika*, 58, 249-254.

(2) Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society*, Ser. B, 26, 211-252.

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

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

(5) Cook, R. D. and Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22, 495-508.

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

(7) Draper, N. R. and John, J. A. (1981). Influential observations and outlier in regression. *Technometrics*, 23, 21-26.

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

(9) Kim, C. (1989). A study on influential sets in regression. Unpublished ph.D. Thesis. University of Wisconsin-Madison.

(10) Lawrance, A. J. (1988). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association*, 83, 1067-1072.

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