Communications of the Korean Statistical Society
2011, Vol. 18, No. 1, 131–136

DOI: 10.5351/CKSS.2011.18.1.131

Accuracy of Multiple Outlier Tests in Nonlinear Regression

Myung-Wook Kahng[1,a]

[a]Department of Statistics, Sookmyung Women's University

Abstract

The original Bates-Watts framework applies only to the complete parameter vector. Thus, guidelines developed in that framework can be misleading when the adequacy of the linear approximation is very different for different subsets. The subset curvature measures appear to be reliable indicators of the adequacy of linear approximation for an arbitrary subset of parameters in nonlinear models. Given the specific mean shift outlier model, the standard approaches to obtaining test statistics for outliers are discussed. The accuracy of outlier tests is investigated using subset curvatures.

Keywords: Curvature measures, intrinsic curvature, outlier test, parameter-effects curvature, subset curvatures.

1. Introduction

A standard paradigm for the analysis of nonlinear models is to assume that results for the linear model hold at least approximately for large enough sample sizes. Confidence regions for parameters of a normal nonlinear regression model are commonly constructed by using linear regression methods, replacing the solution locus with the tangent plane at the maximum likelihood estimate. Such linear regions are generally easier to construct and comprehend than corresponding likelihood regions. Likelihood regions, on the other hand, are not influenced by parameter-effects nonlinearity and generally have true coverage closer to the nominal level than do linear regions. Under suitable regularity conditions and with a sufficiently large sample size, linear and likelihood regions will be in good agreement, but in any particular problem the strength of this agreement is uncertain.

Bates and Watts (1980) propose measures of intrinsic and parameter-effects curvature for assessing the adequacy of the linear approximation. These ideas are extended and refined by Cook and Goldberg (1986). They develop measures for assessing the agreement between linear and likelihood regions for an arbitrary subset of parameters from a nonlinear model. The subset curvatures developed in this paper appear to be reliable indicators of the adequacy of linear confidence regions for most nonlinear models. This ability to deal with subsets greatly extends the usefulness of the Bates-Watts methodology. Because the original Bates-Watts framework applies only to the complete parameter vector, guidelines developed in that framework can be misleading when the adequacy of the linear approximation is very different for different subsets. To ensure good agreement between the tangent plane and likelihood regions, the maximum curvature must be considerably smaller than the Bates-Watts guide. However, this criterion can be too stringent for certain parameter subsets if the whole-parameter curvatures is used. By contrast, the subset curvature describes the shape of the likelihood region in the parameter subspace of interest. Thus, the subset curvature is more directly relevant to the linearization adequacy question.

This research was supported by the Sookmyung Women's University Research Grants 2009.

[1] Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
E-mail: mwkahng@sookmyung.ac.kr

Section 2 considers the problem of testing multiple outliers in nonlinear regression. Given the specific mean shift outlier model, standard approaches to obtaining test statistics for outliers are discussed. The test based on linear approximation, namely the score test, is easy to calculate, but quite different from likelihood ratio tests. The accuracy of the linear approximation is investigated using subset curvature measures in Section 3. We provide an example in Section 4.

2. Outlier Tests in Nonlinear Regression

The standard nonlinear regression model can be written as

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \ldots, n$$

in which the i^{th} response y_i is related to the q-dimensional vector of known explanatory variables $\boldsymbol{x}_i$ through the known model function f, which depends on the p-dimensional unknown parameter $\boldsymbol{\theta} \in \Theta$, and ϵ_i is error. We assume that f is twice continuously differentiable in $\boldsymbol{\theta}$, and errors ϵ_i are *i.i.d.* normal random variables with mean 0 and variance σ^2. In matrix notation we may write,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{y}$ is an n-dimensional vector with elements $y_1, \ldots, y_n$, $\boldsymbol{X}$ is an $n \times q$ matrix with rows $\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T$, $\boldsymbol{\epsilon}$ is an n-dimensional vector with elements $\epsilon_1, \ldots, \epsilon_n$, and $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = (f(\boldsymbol{x}_1, \boldsymbol{\theta}), \ldots, f(\boldsymbol{x}_n, \boldsymbol{\theta}))^T = (f_1(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta}))^T = \boldsymbol{f}(\boldsymbol{\theta})$. Given the response vector $\boldsymbol{y}$, the least squares estimate of $\boldsymbol{\theta}$ is denoted $\hat{\boldsymbol{\theta}}$, and the predicted response vector is $\hat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{X}, \hat{\boldsymbol{\theta}}) = \boldsymbol{f}(\hat{\boldsymbol{\theta}})$. A tangent plane approximation to the expectation surface at $\hat{\boldsymbol{\theta}}$ is used to make inferences about $\boldsymbol{\theta}$ through the derived linear model $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{f}(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{V}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ where $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta}) = \partial \boldsymbol{f} / \partial \boldsymbol{\theta}^T$ is the $n \times p$ matrix and $\hat{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\theta}}) \partial \boldsymbol{f} / \partial \boldsymbol{\theta}^T$ evaluated at $\hat{\boldsymbol{\theta}}$.

Suppose we suspect in advance that m cases indexed by an m-dimensional vector $\mathbf{I} = (i_1, \ldots, i_m)$ are outliers. A useful framework used to study outliers is the mean shift outlier model,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{D}\boldsymbol{\delta} + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\boldsymbol{\delta} = (\delta_{i_1}, \ldots, \delta_{i_m})^T$, $\boldsymbol{D} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m)$, and $\boldsymbol{d}_j$ is the i_j^{th} standard basis vector for $\boldsymbol{R}^n$. The testing of the hypothesis $\boldsymbol{\delta} = \mathbf{0}$ is equivalent to testing whether m cases in the set $\boldsymbol{I}$ are outliers. A single outlier case is discussed by Kahng and Kim (2009).

We denote the log-likelihood for model (2.2) by $L(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2)$ and obtain

$$L\left(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2\right) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\theta}, \boldsymbol{\delta}), \tag{2.3}$$

where $S(\boldsymbol{\theta}, \boldsymbol{\delta}) = (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) - \boldsymbol{D}\boldsymbol{\delta})^T (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) - \boldsymbol{D}\boldsymbol{\delta})$. Let $\hat{\boldsymbol{\theta}}_{(\mathbf{I})}$ be the corresponding estimate with cases indexed by I are excluded. Given σ^2, (2.3) is maximized with respect to $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\delta})$ when $S(\boldsymbol{\theta}, \boldsymbol{\delta})$ is minimized at the least squares estimates $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\theta}}_{(\mathbf{I})}, \hat{\boldsymbol{\delta}})$. Furthermore, $\partial L / \partial \sigma^2 = 0$ has solution $\sigma^2 = S(\boldsymbol{\theta}, \boldsymbol{\delta})/n$, which gives a maximum for given $\boldsymbol{\phi}$ as the second derivative is negative. This suggests that $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\theta}}_{(\mathbf{I})}, \hat{\boldsymbol{\delta}})$ and $\hat{\sigma}^2_{(\mathbf{I})} = S(\hat{\boldsymbol{\theta}}_{(\mathbf{I})}, \hat{\boldsymbol{\delta}})/n$ are the maximum likelihood estimates. Under the null hypothesis $\boldsymbol{\delta} = \mathbf{0}$, the maximum likelihood estimates are $\boldsymbol{\phi}_0 = (\hat{\boldsymbol{\theta}}, \mathbf{0})$ and $\hat{\sigma}^2 = S(\hat{\boldsymbol{\theta}}, \mathbf{0})/n$, which are the maximum likelihood estimates of model (2.1).

Now, we consider procedures for testing $H_0 : \boldsymbol{\delta} = \mathbf{0}$ against $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$. The asymptotically equivalent test methods are discussed by Kahng (1995), namely the likelihood ratio(LR) test introduced by Neyman and Pearson (1928) and the score(S) test due originally to Rao (1947) and developed further by Silvey (1959). These test statistics are defined as LR $= n[\log S(\hat{\boldsymbol{\theta}}, \mathbf{0}) - \log S(\hat{\boldsymbol{\theta}}_{(I)}, \hat{\boldsymbol{\delta}})]$ and $S = \boldsymbol{e}_{\mathbf{I}}^T(\boldsymbol{I} - \hat{\boldsymbol{H}}_{\mathbf{I}})\boldsymbol{e}_{\mathbf{I}}/\hat{\sigma}^2$, where $\tilde{\boldsymbol{H}}_{\mathbf{I}}$ is $m \times m$ minor of $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{V}}(\tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}})^{-1}\tilde{\boldsymbol{V}}^T$ with rows and columns indexed by $\mathbf{I}$, $\tilde{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\theta}}_{(\mathbf{I})})$ is $\partial \boldsymbol{f}/\partial \boldsymbol{\theta}^T$ evaluated at $\hat{\boldsymbol{\theta}}_{(\mathbf{I})}$, $\hat{\boldsymbol{H}}_{\mathbf{I}}$ is the $m \times m$ minor of $\hat{\boldsymbol{H}} = \hat{\boldsymbol{V}}(\hat{\boldsymbol{V}}^T\hat{\boldsymbol{V}})^{-1}\hat{\boldsymbol{V}}^T$ with rows and columns indexed by $\mathbf{I}$ and $\boldsymbol{e}_{\mathbf{I}}$ is $m \times 1$ vector indexed by $\mathbf{I}$.

The two statistics differ in computational features. Unlike the likelihood ratio test, the score test requires only quantities calculated under the null hypothesis. Nevertheless, all statistics are invariant under the reparametrization of $\boldsymbol{\theta}$, and have the same asymptotic distribution under the null hypothesis $H_0 : \boldsymbol{\delta} = \mathbf{0}$.

3. Accuracy of Linear Approximation

Let us consider the mean shift outlier model (2.2). Since parameter $\boldsymbol{\delta}$ is the parameter subset of this model, the subset curvatures for $\boldsymbol{\delta}$ can be found by the methods described in Cook and Goldberg (1986).

Let $\boldsymbol{h}(\boldsymbol{\delta}) = \boldsymbol{f}(\boldsymbol{X}, \tilde{\boldsymbol{\theta}}(\boldsymbol{\delta})) + \boldsymbol{D}\boldsymbol{\delta}$, where $\tilde{\boldsymbol{\theta}}(\boldsymbol{\delta})$ denotes the p-dimensional vector-valued function that maximizes $L(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2)$ over $\boldsymbol{\theta}$ given $\boldsymbol{\delta}$. In order to obtain simple reduced forms for the curvatures, Cook and Goldberg (1986) assume that the intrinsic curvature of $\boldsymbol{h}$ at $\hat{\boldsymbol{\delta}}$ is negligible. This assumption seems to be a reasonable approximation for many nonlinear models, as in the examples given in numerous literatures such as Bates and Watts (1980). In case the intrinsic curvature is large, it can be substantially reduced by reparametrization. The intrinsic curvature measures the property of the whole expectation surface. If the intrinsic curvature is zero, then the expectation and tangent plane are identical and $\boldsymbol{h}(\boldsymbol{\delta})$ is a curve in this plane. The intrinsic curvature should be calculated to check this assumption, however, experiment has shown that they are typically small.

To obtain precise expressions, we need the following definitions. Define $n \times p \times p$ array $\tilde{\boldsymbol{W}} = \boldsymbol{W}(\hat{\boldsymbol{\theta}}_{(\mathbf{I})})$ which is $\partial^2 \boldsymbol{f}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ evaluated at $\hat{\boldsymbol{\theta}}_{(\mathbf{I})}$. We consider the QR decomposition of $n \times (p+m)$ matrix $\boldsymbol{V}_{\boldsymbol{\phi}}$, namely

$$\boldsymbol{V}_{\boldsymbol{\phi}} = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\phi}^T}\right|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} = \left(\tilde{\boldsymbol{V}}\boldsymbol{D}\right) = \boldsymbol{Q}\boldsymbol{R}$$

where $\boldsymbol{Q}$ is an $n \times (p+m)$ matrix with orthogonal columns and $\boldsymbol{R}$ is a $(p+m) \times (p+m)$ upper triangular matrix. Now partition $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \mathbf{0} & \boldsymbol{R}_{22} \end{bmatrix} \quad \text{and} \quad \boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{R}_{11}^{-1} & -\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\boldsymbol{R}_{11}^{-1} \\ \mathbf{0} & \boldsymbol{R}_{22}^{-1} \end{bmatrix},$$

where $\boldsymbol{R}_{11}$ is $p \times p$, $\boldsymbol{R}_{12}$ is $p \times m$, and $\boldsymbol{R}_{22}$ is $m \times m$. Consider the transformation $\check{\boldsymbol{W}}_{\boldsymbol{\phi}} = \boldsymbol{R}^{-T}\tilde{\boldsymbol{W}}_{\boldsymbol{\phi}}\boldsymbol{R}^{-1}$, where the i^{th} face of $\tilde{\boldsymbol{W}}_{\boldsymbol{\phi}}$ is

$$\left(\tilde{\boldsymbol{W}}_{\boldsymbol{\phi}}\right)_i = \left.\frac{\partial^2 h_i}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^T}\right|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} = \begin{bmatrix} \tilde{\boldsymbol{W}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$\tilde{\boldsymbol{W}}_i$ is the i^{th} face of $\tilde{\boldsymbol{W}}$ and h_i is the i^{th} element of $\boldsymbol{h}(\boldsymbol{\delta})$. Then the i^{th} face of $\check{\boldsymbol{W}}_{\boldsymbol{\phi}}$ can be expressed as

$$\left(\check{\boldsymbol{W}}_{\boldsymbol{\phi}}\right)_i = \begin{bmatrix} \boldsymbol{R}_{11}^{-T}\tilde{\boldsymbol{W}}_i\boldsymbol{R}_{11}^{-1} & \boldsymbol{R}_{11}^{-T}\tilde{\boldsymbol{W}}_i\left(\boldsymbol{R}^{-1}\right)_{12} \\ \left(\boldsymbol{R}^{-1}\right)_{12}^T\tilde{\boldsymbol{W}}_i\boldsymbol{R}_{11}^{-1} & \left(\boldsymbol{R}^{-1}\right)_{12}^T\tilde{\boldsymbol{W}}_i\left(\boldsymbol{R}^{-1}\right)_{12} \end{bmatrix} = \begin{bmatrix} \left(\check{\boldsymbol{W}}_{\boldsymbol{\phi}11}\right)_i & \left(\check{\boldsymbol{W}}_{\boldsymbol{\phi}12}\right)_i \\ \left(\check{\boldsymbol{W}}_{\boldsymbol{\phi}21}\right)_i & \left(\check{\boldsymbol{W}}_{\boldsymbol{\phi}22}\right)_i \end{bmatrix}$$

Table 1: Tetracycline data

i	x_i	y_i
1	1	0.7
2	2	1.2
3	3	1.4
4	4	1.4
5	6	1.1
6	8	0.8
7	10	0.6
8	12	0.5
9	16	0.3

where $(\check{\boldsymbol{W}}_{\boldsymbol{\phi} jk})_i$ $(\check{\boldsymbol{W}}_{\boldsymbol{\phi} jk})_i$ is the i^{th} face of $\check{\boldsymbol{W}}_{\boldsymbol{\phi} jk}$ for $j, k = 1, 2$.

Under the above settings and by applying the result of Kahng (1995), we may define the maximum parameter-effects and intrinsic curvatures of $\boldsymbol{h}$ at $\hat{\boldsymbol{\delta}}$ as

$$\Gamma_s^{\tau}(\boldsymbol{\delta}) = \sqrt{m}\, s_{(\mathbf{I})} \max_{\|\boldsymbol{b}\|=1} \left\| \boldsymbol{b}^T \boldsymbol{A}_{22} \boldsymbol{b} \right\|$$

and

$$\Gamma_s^{\eta}(\boldsymbol{\delta}) = 2\sqrt{m}\, s_{(\mathbf{I})} \max_{\|\boldsymbol{b}\|=1} \left\| \left[\boldsymbol{b}^T\right] [\boldsymbol{A}_{12}]\, \boldsymbol{b} \right\|$$

where $s_{(\mathbf{I})}^2 = S(\hat{\boldsymbol{\theta}}_{(\mathbf{I})}, \hat{\boldsymbol{\delta}})/(n - p - m)$, $\check{\boldsymbol{W}}_{\boldsymbol{\phi}_{12}}$ and $\check{\boldsymbol{W}}_{\boldsymbol{\phi}_{22}}$ are the subarrays of $\check{\boldsymbol{W}}_{\boldsymbol{\phi}}$, $\boldsymbol{A}$ is the $(p+m) \times (p+m) \times (p+m)$ parameter-effects curvature array $\boldsymbol{A} = [\boldsymbol{Q}^T][\check{\boldsymbol{W}}_{\boldsymbol{\phi}}]$ (Bates and Watts, 1981), $\boldsymbol{A}_{12}$ and $\boldsymbol{A}_{22}$ are the subarrays of $\boldsymbol{A}$ with i^{th} faces $\boldsymbol{A}_{i12}, \boldsymbol{A}_{i22}$, for $i = p+1, \ldots, p+m$.

Combining these two subset curvatures, the total subset curvature $\Gamma_s(\boldsymbol{\delta})$ of $\boldsymbol{h}$ at $\hat{\boldsymbol{\delta}}$ is $\Gamma_s(\boldsymbol{\delta}) = (\Gamma_s^{\tau}(\boldsymbol{\delta}) + \Gamma_s^{\eta}(\boldsymbol{\delta}))^{1/2}$. If $\Gamma_s(\boldsymbol{\delta})$ (or both $\Gamma_s^{\tau}(\boldsymbol{\delta})$ and $\Gamma_s^{\eta}(\boldsymbol{\delta})$) is sufficiently small, the likelihood and linear confidence regions for δ will be similar, otherwise we can expect these confidence regions to be dissimilar. Following Ratkowsky (1983, p.18) and Cook and Goldberg (1986), $1/\{2\sqrt{F_{\alpha}(m, n-p-m)}\}$ may be used as a rough guide for judging the size of these curvatures. This method can be used to judge the adequacy of the test procedures which are based on the linear approximation. When $\Gamma_s^{\tau}(\boldsymbol{\delta})$ and $\Gamma_s^{\eta}(\boldsymbol{\delta})$ are greater than the guide, the linearization based test, namely the score test, is quite different from the likelihood ratio test.

All quantities in formulas $\Gamma_s^{\tau}(\boldsymbol{\delta})$, $\Gamma_s^{\eta}(\boldsymbol{\delta})$ and $\Gamma_s(\boldsymbol{\delta})$ can be found or estimated without knowing the response, $\boldsymbol{y}_{\mathbf{I}}$, of the case that is suspected to be an outlier. This implies that the curvature measures, $\Gamma_s^{\tau}(\boldsymbol{\delta})$, $\Gamma_s^{\eta}(\boldsymbol{\delta})$ and $\Gamma_s(\boldsymbol{\delta})$, do not depend on how large $\boldsymbol{\delta}$ is or how severe the outliers are. We may have larger curvature measures for $\boldsymbol{\delta}$ even if the cases that are being tested have small test statistics.

There is one problem that has not yet been resolved. If the curvature measures $\Gamma_s^{\tau}(\boldsymbol{\delta})$, $\Gamma_s^{\eta}(\boldsymbol{\delta})$ and $\Gamma_s(\boldsymbol{\delta})$ can be found from the fit of a full data set, we can assess the accuracy of the linear approximation based on the outlier test for each subset of size m prior to fitting all deletion models. We can then use tests based on the linear approximation, such as the score test, for the subsets in which the linear approximation is valid. This substantially reduces the computational cost for detecting outliers. Thus, it is desirable to express curvature measures as a function of a full set of data, as is usual in this kind of an investigation. However, it is not easy to obtain this expression in nonlinear regression because $\tilde{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\theta}}_{(\mathbf{I})})$ and $\tilde{\boldsymbol{W}} = \boldsymbol{W}(\hat{\boldsymbol{\theta}}_{(\mathbf{I})})$ change when cases are deleted.

Table 2: Subset curvatures for $\boldsymbol{\delta}$

I	$\Gamma_s^{\tau}(\boldsymbol{\delta})$	$\Gamma_s^{\eta}(\boldsymbol{\delta})$	$\Gamma_s(\boldsymbol{\delta})$
1, 2	269.7246	0.7103	**269.7255**
1, 3	37.8376	1.2550	**37.8584**
2, 3	15.3341	1.7805	**15.4371**
1, 5	1.6124	0.6677	**1.7452**
1, 9	0.8548	0.5117	**0.9963**
1, 4	0.8356	0.5286	**0.9888**
1, 8	0.8158	0.5354	**0.9758**
1, 7	0.7931	0.5212	**0.9490**
1, 6	0.7758	0.4977	**0.9217**
2, 5	0.4613	0.4383	**0.6363**
2, 4	0.2213	0.3343	**0.4009**
2, 8	0.1988	0.3123	**0.3702**
2, 7	0.1889	0.3176	**0.3695**
2, 9	0.2288	0.2773	**0.3595**
2, 6	0.1927	0.2747	**0.3356**
5, 9	0.0840	0.1839	**0.2022**
5, 6	0.0873	0.1645	**0.1862**
8, 9	0.0365	0.1409	0.1456
3, 9	0.0374	0.1359	0.1410
6, 9	0.0369	0.1323	0.1374
4, 5	0.0880	0.0714	0.1133
3, 6	0.0199	0.0948	0.0969
3, 5	0.0340	0.0890	0.0953
7, 9	0.0110	0.0943	0.0949
4, 9	0.0262	0.0755	0.0799
5, 8	0.0258	0.0697	0.0743
5, 7	0.0261	0.0680	0.0728
6, 7	0.0110	0.0663	0.0672
6, 8	0.0059	0.0613	0.0616
3, 8	0.0119	0.0557	0.0570
4, 6	0.0218	0.0469	0.0517
4, 7	0.0037	0.0507	0.0508
3, 4	0.0137	0.0466	0.0486
3, 7	0.0025	0.0470	0.0471
7, 8	0.0202	0.0408	0.0455
4, 8	0.0125	0.0202	0.0238

4. Example

The data on the metabolism of tetracycline were presented in Bates and Watts (1988) and are reproduced in Table 1. In this experiment, a tetracycline compound was administered orally to a subject and the concentration(y_i) of tetracycline hydrochloride in the serum in micrograms per milliliter was measured over a period(x_i) of 16 hours. The proposed model is the following:

$$f(x, \boldsymbol{\theta}, \delta) = \theta_3 \left[\exp\left(-\theta_1 \left(x - \theta_4\right)\right) - \exp\left(-\theta_2 \left(x - \theta_4\right)\right)\right].$$

We consider the multiple outlier problem, supposing that we suspect two outliers. We calculated the subset curvatures $\Gamma_s^{\tau}(\boldsymbol{\delta})$, $\Gamma_s^{\eta}(\boldsymbol{\delta})$ and $\Gamma_s(\boldsymbol{\delta})$ for each of the 36 pairs of 9 locations. They are listed in descending order of $\Gamma_s(\boldsymbol{\delta})$ in Table 2. The corresponding guide is $1/2\sqrt{F_{.05}(2,3)} = .1618$. For the 17 pairs of subsets with $\Gamma_s(\boldsymbol{\delta})$ values greater than the guide, the score test based on linear approximation is inadequate for testing outliers. This indicates that estimations, inferences, and diagnostics based on the linear approximation about $\boldsymbol{\delta}$ may give misleading results.

References

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion), *Journal of the Royal Statistical Society Series, B,* **42**, 1–25.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications,* John Wiley and Sons, New York.

Cook, R. D. and Goldberg, M. L. (1986). Curvatures for parameter subsets in nonlinear regression, *The Annals of Statistics*, **14**, 1399–1418.

Cook, R. D. and Weisberg, S. (1990). Confidence curves in nonlinear regression, *Journal of the American Statistical Association,* **85**, 544–551.

Kahng, M. (1995). Testing outliers in nonlinear regression, *Journal of the Korean Statistical Society*, **24**, 419–437.

Kahng, M. and Kim, B. (2009). Assessing the accuracy of outlier tests in nonlinear regression, *Communications of the Korean Statistical Society,* **16**, 163–168.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference, *Biometrika,* **20A**, 175–240 and 263–294.

Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society,* **44**, 50–57.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach,* Marcel Dekker, New York.

Silvey, S. D. (1959). The Lagrangian multiplier test, *Annals of Mathematical Statistics,* **30**, 389–407.

Received November 2010; Accepted December 2010