The Korean Communications
in Statistics Vol. 6, No. 3, 1999
pp. 939-946

Influence Analysis of the Likelihood Ratio Test in Multivariate Behrens-Fisher Problem

Kang-Mo Jung[1] and Myung Geun Kim[2]

Abstract

We propose methods for detecting influential observations that have a large influence on the likelihood ratio test statistic for the multivariate Behrens-Fisher problem. For this purpose we derive the influence curve and the derivative influence of the likelihood ratio test statistic. An illustrative example is given to show the effectiveness of the proposed methods on the identification of influential observations.

1. Introduction

The detection of outliers and influential observations has a long history. However, many diagnostic measures have been proposed for influence analyses in the context of estimation. A few works that treat detection of influential observations for test statistics in multivariate analysis are found. Among others, Kim (1995) investigated the influence of observations on the likelihood ratio test statistics in comparing covariance matrices based on influence curves. Influence analysis in testing problems is very important because in extreme situations, a single observation can dominate our conclusion about the hypotheses as can be seen in Section 5.

The likelihood ratio test (LRT) statistic in the multivariate Behrens-Fisher problem contains two different covariance matrices. It is well known that the covariance matrix is very sensitive to influential observations, and so is the LRT statistic. Case deletion diagnostics are widely used in many statistical analyses (Cook and Weisberg, 1982). However, case deletion diagnostics require amount of computation time. The influence curve was introduced by Hampel (1974) as a device to measure the effect of an infinitesimal contamination at an observation on a statistic. It has been used as a criterion for detecting outliers and influential observations. Another simple method for influence analysis is the derivative influence (De Gruttola et al., 1987), the differential change in an estimated parameter resulting from a slight perturbation in the weight assigned to a given observation.

In this work we derive the influence curve and the derivative influence of the LRT statistic in the multivariate Behrens-Fisher problem for the purpose of investigating the influence of observations. To get the influence curve we appropriately define statistical functionals for the

1) Assistant Professor, Department of Informatics and Statistics, Kunsan National University, Kunsan, 573-701, Korea.
2) Associate Professor, Department of Applied Statistics, Seowon University, Chongju, 361-742, Korea.

observations. To get the influence curve we appropriately define statistical functionals for the LRT statistic and parameters. We use the empirical influence curve as a sample version of the influence curve that provides useful information about the influence of observations on the LRT statistic. It is well known that the sample covariance matrix is more sensitive to influential observations than the sample mean vector. Thus, for getting the derivative influence the perturbation is chosen in which a weight is put on the covariance matrix for an observation.

Section 2 discusses the LRT statistic in the multivariate Behrens-Fisher problem. We derive the influence curve and the derivative influence on the LRT statistic in Sections 3 and 4, respectively. In Section 5, a numerical example is given and it will show that the case deletion diagnostic and proposed methods have the same results.

2. LRT Statistic in Multivariate Behrens-Fisher Problem

Suppose that two independent random samples $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n_1}$ and $\boldsymbol{x}_{n_1+1}, \cdots, \boldsymbol{x}_n$ $(n = n_1 + n_2)$ are drawn from p-variate normal distribution $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, respectively. The Behrens-Fisher problem is the test of the null hypothesis H_0 against the alternative hypothesis H_1 defined by

$$H_0: \ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 (= \boldsymbol{\mu}), \qquad H_1: \ \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{1}$$

Let F_i be the distribution function from the ith population for $i = 1, 2$, and let $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(F_i)$, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(F_i)$ be the statistical functionals for each population. The empirical distribution function based on n_i observations for each population is denoted by $\widehat{F}_i$. Then $\widehat{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_i(\widehat{F}_i)$ and $\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i(\widehat{F}_i)$ become the maximum likelihood estimators. We write $L_i(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ as the likelihood function based on n_i observations for each population. Under H_1, the maximized likelihood function becomes $L_1(\widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\Sigma}}_1)L_2(\widehat{\boldsymbol{\mu}}_2, \widehat{\boldsymbol{\Sigma}}_2)$, where $L_i(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i) = (2\pi)^{-n_i p/2} |\widehat{\boldsymbol{\Sigma}}_i|^{-n_i/2} e^{-n_i p/2}$. Therefore, the statistical functional related to this maximized likelihood is

$$LF_1(F_1, F_2) = (2\pi)^{-np/2} e^{-np/2} |\boldsymbol{\Sigma}_1|^{-n_1/2} |\boldsymbol{\Sigma}_2|^{-n_2/2}.$$

Let $\boldsymbol{\mu} = \boldsymbol{\mu}(F_1, F_2)$ and $\boldsymbol{\Sigma}_{i0} = \boldsymbol{\Sigma}_{i0}(F_1, F_2)$ be statistical functionals satisfying the following equations

$$\boldsymbol{\mu} = (n_1 \boldsymbol{\Sigma}_{10}^{-1} + n_2 \boldsymbol{\Sigma}_{20}^{-1})^{-1} (n_1 \boldsymbol{\Sigma}_{10}^{-1} \boldsymbol{\mu}_1 + n_2 \boldsymbol{\Sigma}_{20}^{-1} \boldsymbol{\mu}_2), \tag{2}$$

$$\boldsymbol{\Sigma}_{i0} = \boldsymbol{\Sigma}_i + \boldsymbol{d}_i \boldsymbol{d}_i^T, \tag{3}$$

$\boldsymbol{d}_i = \boldsymbol{\mu}_i - \boldsymbol{\mu}$. Then $\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\widehat{F}_1, \widehat{F}_2)$ and $\widehat{\boldsymbol{\Sigma}}_{i0} = \boldsymbol{\Sigma}_{i0}(\widehat{F}_1, \widehat{F}_2)$ are the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{i0}$ under H_0 (Mardia et al., pp. 142-143, 1979). Under H_0 the maximized likelihood is $L_1(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}_{10})L_2(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}_{20})$. Then the statistical functional related to this maximized likelihood becomes

$$LF_0(F_1, F_2) = \prod_{i=1}^{2}(2\pi)^{-n_i p/2}|\boldsymbol{\Sigma}_{i0}|^{-n_i/2} \times \exp[-(n_i/2)\{\mathrm{tr}(\boldsymbol{\Sigma}_{i0}^{-1}\boldsymbol{\Sigma}_i) - \boldsymbol{d}_i^T\boldsymbol{\Sigma}_{i0}^{-1}\boldsymbol{d}_i\}.$$

Hence the statistical functional $\lambda = \lambda(F_1, F_2)$ related to the LRT statistic is easily computed as

$$\begin{aligned}\lambda &= -2\log LF_0(F_1, F_2)/LF_1(F_1, F_2) \\ &= \sum_{i=1}^{2} n_i \log(1+\boldsymbol{d}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}_i)\end{aligned} \tag{4}$$

using $\boldsymbol{\Sigma}_{i0}^{-1} = \boldsymbol{\Sigma}_i^{-1} - (1+\boldsymbol{d}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}_i)^{-1}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}_i\boldsymbol{d}_i^T\boldsymbol{\Sigma}_i^{-1}$ and $|\boldsymbol{\Sigma}_{i0}| = |\boldsymbol{\Sigma}_i|(1+\boldsymbol{d}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{d}_i)$.

3. Influence Curve

Let $T = T(F_1, F_2)$ be a parameter which is expressed as a functional of the distribution functions F_1, F_2. Assume that F_1 is perturbed as $\widetilde{F}_1 = (1-\varepsilon)F_1 + \varepsilon\delta_{\boldsymbol{x}}$, where $\varepsilon > 0$and $\delta_{\boldsymbol{x}}$ is the distribution function having unit mass at $\boldsymbol{x}$. Then the influence function $I(\boldsymbol{x}, T)$ for the statistical functional $T = T(F_1, F_2)$ at $\boldsymbol{x}$ is defined (Hampel, 1974) by

$$I(\boldsymbol{x}, T) = \lim_{\varepsilon \to 0}\frac{\widetilde{T} - T}{\varepsilon},$$

where $\widetilde{T} = T(\widetilde{F}_1, F_2)$. We expand $\widetilde{T} = T(\varepsilon)$ as a convergent power series of ε as follows

$$T(\varepsilon) = T(\widetilde{F}_1, F_2) = T + T^{(1)}\varepsilon + O(\varepsilon^2). \tag{5}$$

The influence curve $I(\boldsymbol{x}, T)$ is obtained as the first order differential coefficient of $T(\varepsilon)$ at $\varepsilon = 0$, that is, the coefficient $T^{(1)}$ of the first order term of ε in the power series expansion (5).

The definition $\boldsymbol{\mu}_1(\varepsilon) = \int \boldsymbol{x}\, d\widetilde{F}_1$ yields $\boldsymbol{\mu}_1(\varepsilon) = \boldsymbol{\mu}_1 + \boldsymbol{z}\varepsilon$, $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{\mu}_1$. Also we have

$$\boldsymbol{\Sigma}_1(\varepsilon) = \boldsymbol{\Sigma}_1 + (\boldsymbol{z}\boldsymbol{z}^T - \boldsymbol{\Sigma}_1)\varepsilon + O(\varepsilon^2),$$

whereas $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ are not perturbed since only F_1 is perturbed.

If a perturbed matrix $\boldsymbol{A}(\varepsilon)$ for a matrix $\boldsymbol{A}$ is assumed to be expanded as $\boldsymbol{A}(\varepsilon) = \boldsymbol{A} + \boldsymbol{A}^{(1)}\varepsilon + O(\varepsilon^2)$, where $\boldsymbol{A}^{(1)} = d\boldsymbol{A}(\varepsilon)/d\varepsilon|_{\varepsilon=0}$, then the identity $\boldsymbol{I} = \boldsymbol{A}(\varepsilon)\boldsymbol{A}(\varepsilon)^{-1}$ gives

$$A(\varepsilon)^{-1} = A^{-1} + A^{-1(1)}\varepsilon + O(\varepsilon^2), \tag{6}$$

where $A^{-1(1)} = -A^{-1} A^{(1)} A^{-1}$.

By using (6), we get

$$\Sigma_1(\varepsilon)^{-1} = \Sigma_1^{-1} + (\Sigma_1^{-1} - \Sigma_1^{-1} z z^T \Sigma_1^{-1})\varepsilon + O(\varepsilon^2). \tag{7}$$

The perturbed functionals for Σ_{i0} can be written as

$$\begin{aligned} \Sigma_{10}(\varepsilon) &= \Sigma_1(\varepsilon) + d_1(\varepsilon) d_1(\varepsilon)^T = \Sigma_{10} + \Sigma_{10}^{(1)}\varepsilon + O(\varepsilon^2), \\ \Sigma_{20}(\varepsilon) &= \Sigma_2(\varepsilon) + d_2(\varepsilon) d_2(\varepsilon)^T = \Sigma_{20} + \Sigma_{20}^{(1)}\varepsilon + O(\varepsilon^2), \end{aligned}$$

where $d_1(\varepsilon) = \mu_1(\varepsilon) - \mu(\varepsilon)$, $d_2(\varepsilon) = \mu_2 - \mu(\varepsilon)$, and for $i = 1,2$

$$\Sigma_{i0}^{(1)} = \Sigma_i^{(1)} + d_i d_i^{(1)T} + d_i^{(1)} d_i^T.$$

Equation (6) yields $\Sigma_{i0}^{-1(1)} = -\Sigma_{i0}^{-1} \Sigma_{i0}^{(1)} \Sigma_{i0}^{-1}$.

Let $\Gamma(\varepsilon) = (\sum_{i=1}^{2} n_i \Sigma_{i0}(\varepsilon)^{-1})^{-1}$ and $\eta(\varepsilon) = \sum_{i=1}^{2} n_i \Sigma_{i0}(\varepsilon)^{-1} \mu_i(\varepsilon)$. Then

$$\begin{aligned} \Gamma^{(1)} &= \sum_{i=1}^{2} n_i \Gamma \Sigma_{i0}^{-1} \Sigma_{i0}^{(1)} \Sigma_{i0}^{-1} \Gamma, \\ \eta^{(1)} &= \sum_{i=1}^{2} n_i (\Sigma_{i0}^{-1} \mu_i^{(1)} - \Sigma_{i0}^{-1} \Sigma_{i0}^{(1)} \Sigma_{i0}^{-1} \mu_i). \end{aligned}$$

Since $\mu(\varepsilon) = \Gamma(\varepsilon)\eta(\varepsilon)$, $\mu^{(1)}$ satisfies the linear equation

$$\mu^{(1)} = \Delta \mu^{(1)} + \alpha, \tag{8}$$

where

$$\Delta = \sum_{i=1}^{2} n_i \Gamma \Sigma_{i0}^{-1}(d_i^T \Sigma_{i0}^{-1} d_i \cdot I_p + d_i d_i^T \Sigma_{i0}^{-1}), \tag{9}$$

$$\alpha = n_1 \Gamma\{\Sigma_{10}^{-1} z - \Sigma_{10}^{-1}(z z^T - \Sigma_1 + d_1 z^T + z d_1^T) \Sigma_{10}^{-1} d_1\}, \tag{10}$$

where I_p is the identity matrix of order p. By solving (8) we get $\mu^{(1)}$, $d_1^{(1)} = z - \mu^{(1)}$ and $d_2^{(1)} = -\mu^{(1)}$. Also we can obtain $\Sigma_1^{-1(1)}$ from (7), and $\Sigma_2^{-1(1)} = 0$. When F_2 is perturbed, we will get necessary results by interchanging the subscripts 1 and 2.

Then the first order differential coefficient of the perturbed LRT functional $\lambda(\varepsilon)$ becomes

$$\lambda^{(1)} = \sum_{i=1}^{2} n_i (1+u_i)^{-1}(d_i^T \Sigma_i^{-1(1)} d_i + 2 d_i^{(1)T} \Sigma_i^{-1} d_i), \tag{11}$$

where $u_i = d_i^T \Sigma_i^{-1} d_i$.

Among three popular sample versions of the theoretical influence curve, for example, the empirical influence curve (EIC), the deleted empirical influence curve, and the sample influence curve (Cook and Weisberg, 1982), we will use EIC for simplicity. It is well known that all three sample versions yield similar results. The EIC is obtained by substituting the empirical distribution functions $\widehat{F}_1, \widehat{F}_2$ and the observation x_j for F_1, F_2 and x in the definition, respectively.

4. Derivative Influence

Let $\hat{T}$ be an estimator of T. The derivative influence is derived by considering a perturbation scheme represented by w in which the distribution for only one observation is perturbed. When we denote by $\hat{T}(w)$ the estimator of T under this perturbation scheme, we assume that $\hat{T}=\hat{T}(w_0)$ for some w_0 . If only the jth observation is perturbed, the derivative influence is defined by

$$D_j(\hat{T})=\left.\frac{d\hat{T}(w)}{dw}\right|_{w=w_0} \qquad (12)$$

(De Gruttola et al., 1987).

Under H_1, assume that only observation $\boldsymbol{x}_j$ $(1\le j\le n_1)$ is drawn from $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1/w)$. This perturbation is based on the fact that the sample covariance is more sensitive than the sample mean vector. Let $\boldsymbol{\mu}_i(\hat{F}_i)=\overline{\boldsymbol{x}_i}$ and $\boldsymbol{\Sigma}_i(\hat{F}_i)=S_i$. Under the perturbation scheme, the maximum likelihood estimators become

$$\begin{aligned}\widehat{\boldsymbol{\mu}_1}(w) &= (n_1+w-1)^{-1}\{n_1\overline{\boldsymbol{x}_1}+(w-1)\ \boldsymbol{x}_j\},\\ \widehat{\boldsymbol{\Sigma}_1}(w) &= S_1+(w-1)(n_1+w-1)^{-1}(\boldsymbol{x}_j-\overline{\boldsymbol{x}_1})(\boldsymbol{x}_j-\overline{\boldsymbol{x}_1})^T,\end{aligned}$$

whereas $\widehat{\boldsymbol{\mu}_2}=\overline{\boldsymbol{x}_2}$, $\widehat{\boldsymbol{\Sigma}_2}=S_2$. Note that $\widehat{\boldsymbol{\mu}_1}(1)=\widehat{\boldsymbol{\mu}_1}=\overline{\boldsymbol{x}_1}$ and $\widehat{\boldsymbol{\Sigma}_1}(1)=\widehat{\boldsymbol{\Sigma}_1}=S_1$. Then we get the derivatives of $\widehat{\boldsymbol{\mu}_1}(w)$, $\widehat{\boldsymbol{\Sigma}_1}(w)$ at $w=1$ as

$$D_j(\widehat{\boldsymbol{\mu}_1})=\left.\frac{d\,\widehat{\boldsymbol{\mu}_1}(w)}{dw}\right|_{w=1}=\frac{1}{n_1}(\boldsymbol{x}_j-\overline{\boldsymbol{x}_1}),\qquad D_j(\widehat{\boldsymbol{\Sigma}_1})=\frac{1}{n_1}(\boldsymbol{x}_j-\overline{\boldsymbol{x}_1})(\boldsymbol{x}_j-\overline{\boldsymbol{x}_1})^T,$$

and also $D_j(\widehat{\boldsymbol{\Sigma}_1}^{-1})=-\widehat{\boldsymbol{\Sigma}_1}^{-1}D_j(\widehat{\boldsymbol{\Sigma}_1})\widehat{\boldsymbol{\Sigma}_1}^{-1}$.

Similarly, under H_0 we consider the perturbation scheme in which $\boldsymbol{x}_j$ $(1\le j\le n_1)$ is drawn from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{10}/w)$. The likelihood function gives the maximum likelihood estimators

$$\begin{aligned}\hat{\boldsymbol{\mu}}(w) &= \hat{\Gamma}(w)\hat{\boldsymbol{\eta}}(w),\\ \hat{\boldsymbol{\Sigma}}_{10}(w) &= S_1+(\overline{\boldsymbol{x}_1}-\hat{\boldsymbol{\mu}}(w))(\overline{\boldsymbol{x}_1}-\hat{\boldsymbol{\mu}}(w))^T+\frac{(w-1)}{n_1}(\boldsymbol{x}_j-\hat{\boldsymbol{\mu}}(w))(\boldsymbol{x}_j-\hat{\boldsymbol{\mu}}(w))^T,\\ \hat{\boldsymbol{\Sigma}}_{20}(w) &= S_2+(\overline{\boldsymbol{x}_2}-\hat{\boldsymbol{\mu}}(w))(\overline{\boldsymbol{x}_2}-\hat{\boldsymbol{\mu}}(w))^T,\end{aligned}$$

where $\hat{\Gamma}(w)=\{(n_1+w-1)\widehat{\boldsymbol{\Sigma}_{10}}(w)^{-1}+n_2\widehat{\boldsymbol{\Sigma}_{20}}(w)^{-1}\}^{-1}$ and $\hat{\boldsymbol{\eta}}(w)=n_1\widehat{\boldsymbol{\Sigma}_{10}}(w)^{-1}\overline{\boldsymbol{x}_1}+(w-1)\widehat{\boldsymbol{\Sigma}_{10}}(w)^{-1}\boldsymbol{x}_j+n_2\widehat{\boldsymbol{\Sigma}_{20}}(w)^{-1}\overline{\boldsymbol{x}_2}$. As in Section 3, the chain rule of the differentiation yields the linear equation

$$D_j(\boldsymbol{\mu})=\hat{\Delta}D_j(\boldsymbol{\mu})+\widehat{\boldsymbol{\beta}_j}, \qquad (13)$$

where $\widehat{\beta}_j = \{1-(x_j-\widehat{\mu})^T \widehat{\Sigma}_{10}^{-1}(\overline{x}_1-\widehat{\mu})\}\widehat{\Gamma}\,\widehat{\Sigma}_{10}^{-1}(x_j-\widehat{\mu})$ and $\widehat{\Delta}$ is the same with Δ in (9) when we used the empirical distribution functions $\widehat{F}_1$ and $\widehat{F}_2$. Similarly to (8) we can get the derivative influence $D_j(\mu)$ at x_j by solving (13).

Some algebra yields the perturbed LRT statistic as

$$\widehat{\lambda}(w) = n_1 \log\{1+n_1^{-1}(n_1+w-1)\widehat{u}_1(w)\} + n_2 \log\{1+\widehat{u}_2(w)\},$$

where $\widehat{u}_i(w) = (\widehat{\mu}_i(w)-\widehat{\mu}(w))^T \widehat{\Sigma}_i(w)^{-1}(\widehat{\mu}_i(w)-\widehat{\mu}(w))$. Therefore, we get the derivative influence of $\widehat{\lambda}$ as following

$$D_j(\widehat{\lambda}) = \frac{n_1}{1+\widehat{u}_1}\{\frac{\widehat{u}_1}{n_1} + 2\,\widehat{d}_1^T\,\widehat{\Sigma}_1^{-1} D_j(\widehat{d}_1) + \widehat{d}_1^T D_j(\widehat{\Sigma}_1^{-1})\,\widehat{d}_1\} + \frac{n_2}{1+\widehat{u}_2}\,2\,\widehat{d}_2^T\, S_2^{-1} D_j(\widehat{d}_2), \tag{14}$$

where $D_j(\widehat{d}_1) = D_j(\widehat{\mu}_1) - D_j(\widehat{\mu})$ and $D_j(\widehat{d}_2) = -D_j(\widehat{\mu})$.

5. Numerical Example

The empirical influence curve and the derivative influence described in Sections 3 and 4 are applied to the hook-billed kites data (Johnson and Wichern, 1992, Tables 5.7 and 6.7). The data set has measurements on two variables (tail length and wing length in millimeters) for 45 male and 45 female kites, respectively. The observations are labelled as 1 to 45 for male kites and 46 to 90 for female kites.

The maximum likelihood estimators based on the full data set are $\overline{x}_1 = (189.3, 280.9)^T$, $\overline{x}_2 = (193.6, 279.8)^T$, $\widehat{\mu} = (193.1, 280.2)^T$, $S_1 = \begin{pmatrix} 288.2 & 77.7 \\ 77.7 & 165.5 \end{pmatrix}$, $S_2 = \begin{pmatrix} 118.0 & 119.6 \\ 119.6 & 203.9 \end{pmatrix}$, $\widehat{\Sigma}_{10} = \begin{pmatrix} 302.7 & 75.1 \\ 75.1 & 165.9 \end{pmatrix}$, $\widehat{\Sigma}_{20} = \begin{pmatrix} 118.3 & 119.4 \\ 119.4 & 204.1 \end{pmatrix}$. The above results are obtained using S-PLUS by solving the simultaneous equations (2) and (3) iteratively with the termination rule that the maximum difference between two values of $\widehat{\mu}$ in the previous step and in the current step is less than 0.001. The LRT statistic based on the full data set is $\widehat{\lambda}$ = 3.62, and therefore we conclude that the null hypothesis is not rejected by comparing $\chi^2_2(0.05) = 5.99$, where $\chi^2_k(\alpha)$ is the upper αth percentile of the χ^2 distribution with k degrees of freedom.

The index plots of the empirical influence curve and the derivative influence for the LRT statistic from (11) and (14) are shown in Figures 1 and 2, respectively. We can observe that observation 31 has high influence.

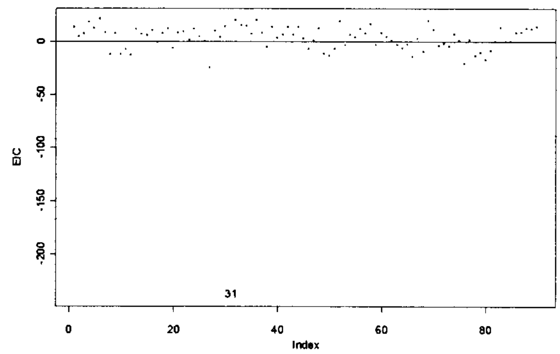

Figure 1. Empirical influence curve.

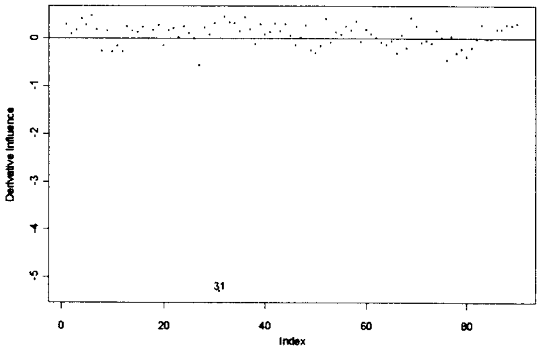

Figure 2. Derivative influence

The case deletion results are presented in Figure 3. The vertical axis denotes $\hat{\lambda} - \hat{\lambda}_{(j)}$, where $\hat{\lambda}_{(j)}$ is the LRT statistic without the jth observation. If this measure is large, then the corresponding observation has large influence on $\hat{\lambda}$. The case deletion diagnostic method yields the same results as the empirical influence curve and the derivative influence. Furthermore, $\hat{\lambda}_{(31)} = 22.47$ gives that the null hypothesis is rejected. It implies that opposite conclusions are made by removing only observation 31 or not. We conclude that observation 31 is a large influential observation on the LRT statistic.

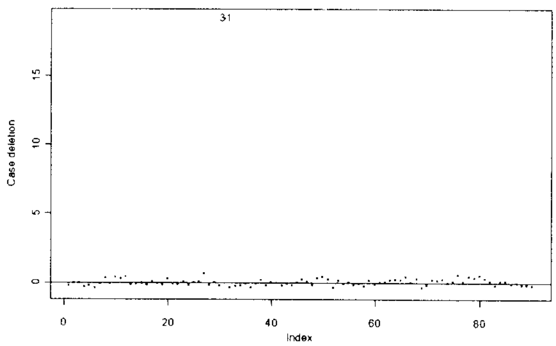

Figure 3. Case deletion diagnostic

The case deletion method could be a giant time-consuming job, particularly in the Behrens-Fisher problem, because we should solve equations (2) and (3) iteratively. Therefore, methods of the influence curve and the derivative influence are efficient in detecting influential observations.

References

[1] Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression*, Chapman and Hall, New York.

[2] De Gruttola, V., Ware, J. H. and Louis, T. A. (1987). Influence analysis of generalized least squares estimators, *Journal of the American Statistical Association*, Vol. 82, 911-917.

[3] Hampel, F. R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, Vol. 69, 383-393.

[4] Johnson, R. A. and Wichern, D. W. (1992). *Applied multivariate statistical analysis, 3rd Ed.*, Prentice Hall, New Jersey.

[5] Kim, M. G. (1995). The influence in comparing covariance matrices, *Communications in Statistics: Theory and Methods,* Vol, 24, 1431-1441.

[6] Mardia, K. V., Kent. J. T. and Bibby, J. M. (1979). *Multivariate analysis*, Academic Press, New York.