IDENTIFICATION OF INFLUENTIAL OBSERVATIONS IN TESTING ZERO CORRELATION

YOUNG HEE PARK

1. Introduction

A test for zero correlation is equivalent to testing that a covariance matrix is diagonal. Generally, it is performed using the likelihood ratio test (LRT). However, The LRT is very sensitive to outliers and to a little contaminated multivariate normality. Hence it is required to detect the influential observations in testing zero correlation.

The local influence method has been proposed by Cook(1986) for assessing the local influence of minor perturbations of a statistical model using likelihood displacement. Wu and Luo (1993) extended it to the so-called second order local influence method that is based on evaluating the curvature for the perturbation-formed surface of a variable such as maximum likelihood estimator (MLE) of a parameter. The surface is called the MLE surface..

The second order local influence method is considered in this thesis and is applied to detecting observations which have large local influence on testing zero correlation. A perturbation scheme is considered in which the covariance matrix is not homogeneous over the sample. The perturbation vector and the perturbed likelihood ratio statistic form a surface in the Euclidean space. A measure of how the surface deviates from its tangent plane at the point representing no perturbation is obtained by studying the curvature of a certain curve on the surface. We consider the largest and the second largest curvatures and their associated direction vectors to investigate the influence of observations. Observations with highly significant direction cosine can be influential.

An illustrative example is given to compare the local influence method with the case-deletion method. This example shows that the local influence method clearly identifies individually and jointly influential observations.

Key words:Influential observations, Local influence, Correlation, Outliers, Case-deletion.

2. Local influence measure for testing zero correlation

Let $x_1, x_2, \ldots, x_n$ be a random sample from $N_p(\mu, \Sigma)$ and consider

$$x_j = \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix},$$

where x_j and μ are $p \times 1$ vectors.

For testing the hypothesis

$$H_0 : \Sigma_0 = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 & 0 \\ 0 & \sigma_2 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & \sigma_p \end{pmatrix}, \tag{1}$$

say, that the p variables $x_{j1}, x_{j2}, \ldots, x_{jp}$ are mutually independent, we will assess local influence. One of test procedures for the hypothesis is the likelihood ratio test. For more details, refer to Seber (1984).

We consider a simultaneously perturbed model specified by a perturbation vector $w = (w_1, \ldots, w_n)^T$ in which

$$x_r \sim N(\mu, \Sigma/w_r) \quad (1 \leq r \leq n) \tag{2}$$

where $w = 1_n + al$ ($w_i = 1 + al_i$, for $i = 1, 2, \ldots, n$), scalar a measures the magnitude of a perturbation in the direction vector l of unit length. The point called null point $w = 1_n$ corresponding to $a = 0$, represents no perturbation in any direction.

Let $S(w)$ be sample covariance matrix which is a maximum likelihood estimator for Σ under H_1 and perturbed model. And $\prod_{i=1}^{p} s_i$ is the determinant of a maximum likelihood estimator for $\Sigma_0(w)$ under H_0. Then the perturbation of the likelihood ratio criterion is given by

$$\eta(w) = |S(w)|/\Pi_{i=1}^{p} s_i \tag{3}$$

which is a function of $w = w(a)$. The likelihood ratio statistic for testing H_0 in (1) is an increasing function of $\eta(1_n)$ and therefore $\eta(w)$ is a sort of the perturbed likelihood ratio statistic.

We will express $\eta(w)$ in a convenient form. Let $A(w)$ be a lower triangular matrix with positive diagonal elements such that $S(w) = A(w)A^T(w)$. $A(w)$ is unique whenever $S(w)$ is positive definite with probability 1, that is $n > p$. Let $B^T(w)$ be $A^{-1}(w)$. Then $B(w)$ is an upper triangular matrix satisfying the identity $S^{-1}(w) = B(w)B^T(w)$. We denote by $b_{jk}(w)$ the (j, k) component of $B(w)$.

Then the test criterion, $\eta(w)$ can be expressed as

$$\eta(w) = \Pi_{k=1}^{p} b_{kk}^{-2}(w) \Pi_{k=1}^{p} s_k^{-1}. \tag{4}$$

The perturbed vector and the perturbed likelihood ratio criterion form a surface in the $(n + 1)$-dimensional Euclidean space. The signed normal curvature in direction l of the surface is obtained at the null point 1_n from the following equations (See Wu and Luo (1993), (2.2)):

$$C_l = \frac{l^T \ddot{\eta}^T l}{(1 + \dot{\eta}^T \dot{\eta})^{1/2} l^T (I + \dot{\eta}\dot{\eta}^T) l} \tag{5}$$

with

$$\frac{\partial \eta}{\partial a}|_{a=0} = \dot{\eta}^T l, \qquad \frac{\partial^2 \eta}{\partial a^2}|_{a=0} = l^T \ddot{\eta} l, \tag{6}$$

where $\dot{\eta}$ is $\partial\eta/\partial w$ and $\ddot{\eta}$ is $\partial^2\eta/\partial w \partial w^T$. The curvature C_l has the form $l^T F l / l^T D l$, where $F = \ddot{\eta}$ and $D = (I + \dot{\eta}\dot{\eta}^T)(1 + \dot{\eta}^T\dot{\eta})^{1/2}$. The local maximum curvature and its associated direction vector are obtained by solving the following generalized eigenvalue problem (Wilkinson, 1965)

$$|F - \lambda D| = 0. \tag{7}$$

This calculation is standard. The eigenvalue obtained from (7) corresponds to curvature and its associated eigenvector to direction vector. Observations corresponding to highly significant direction cosine of the absolutely largest curvature are possibly influential. A plot of two direction vectors, l_{max} and l_{2nd} corresponding to the first- and second-largest local maximum curvatures respectively is helpful in detecting the influential observations by locating the points which are separated from the rest of data. The results of the first order approach are the by-products of this second order method. The value of the maximum slope is $\dot{\eta}^T l$ which is in the direction with cosines $\dot{\eta}_i / \|\dot{\eta}\|$ for $i = 1, \ldots, n$. The plot of the directional cosines of maximum slope and observation numbers will be helpful to detect outliers.

3. Procedure of local influence method

To get D, we consider the following equation

$$\dot{\eta}^T l = \frac{\partial \eta(w)}{\partial a}|_{a=0} = \sum_{r=1}^{n} l_r \left[\sum_{i=1}^{q} \frac{\partial \eta}{\partial \theta_i}|_{\theta=\hat{\theta}} \frac{\partial \hat{\theta}_i(w)}{\partial w_r}|_{w=1_n} \right], \tag{8}$$

where $q(=2p)$ is the number of parameters, θ_i denotes the parameter in (4) and $\hat{\theta}_i(w)$ represents the maximum likelihood estimator of the θ_i under the perturbed model.

The partial derivatives of η with respect to the parameters are

$$\frac{\partial \eta}{\partial b_{kk}} = \frac{-2\eta}{b_{kk}}, \quad \frac{\partial \eta}{\partial \sigma_k} = -\frac{\eta}{\sigma_k}. \tag{9}$$

For the perturbed model, the maximum likelihood estimator of $\mu(w)$ is given by under the null and full models.

$$\hat{\mu}(w) = \bar{x}(w) = \sum_{j=1}^{n} w_j x_j / \sum_{j=1}^{n} w_j \tag{10}$$

and that of σ_i under H_0 is $s_i(w)$ which satisfies

$$s_i(w) = \frac{\sum_{r=1}^{n}\{w_r(x_{ri} - \bar{x}_i(w))^2\}}{n}. \tag{11}$$

Under no constraint, the maximum likelihood estimator of Σ is $S(w)$ and we have

$$B(w)^T S(w) B(w) = I_p, \tag{12}$$

where $S(w) = (1/n)\left[XH(w)X^T\right]$, $X = (x_1, x_2, \ldots, x_n)$, and $H(w) = \text{diag}(w_1, \ldots, w_n) - ww^T/1_n^T w$.

The partial derivative $\partial\hat{\mu}/\partial w_r$ obtained by differentiating (10) with respect to w_r, evaluated at $w = 1_n$ is

$$\frac{\partial \hat{\mu}}{\partial w_r}|_{w=1_n} = \frac{1}{n}(x_r - \bar{x}) \quad (1 \le r \le n). \tag{13}$$

If we differentiate both sides of (11) and (12) with respect to w_r and use (13) then

$$\frac{\partial s_k(w)}{\partial w_r}|_{w=1_n} = \frac{(x_{rk} - \bar{x}_k)^2}{n} \tag{14}$$

for $1 \le k \le p$ and $\partial B(w)/\partial w_r|_{w=1_n}$ satisfies the following identities:

$$\begin{aligned} \left(\frac{\partial B(w)}{\partial w_r}|_{w=1_n}\right)^T A \;&+\; A^T\left(\frac{\partial B(w)}{\partial w_r}|_{w=1_n}\right) \\ &= -\frac{1}{n} B^T (x_r - \bar{x})(x_r - \bar{x})^T B \end{aligned} \tag{15}$$

for $1 \leq r \leq n$. Then the $(\partial b_{kk}(w)/\partial w_r)|_{w=1_n}$ is given by $c_{k,k}/(2v_{k,k})$, where each $v_{k,k}$ and $c_{k,k}$ is the (k,k) component of A and the right-hand side of (15) $(k = 1, \ldots, p)$.

To obtain F, we consider $n \times n$ matrix $\ddot{\eta}$ of which the (r,s) component is as follows:

$$\frac{\partial^2 \eta(\theta)}{\partial w_r \partial w_s} = \sum_{i=1}^{q} \left[\sum_{j=1}^{q} \left\{ \frac{\partial^2 \eta(\theta)}{\partial \theta_j \partial \theta_i}\Big|_{\theta=\hat{\theta}} \cdot \frac{\partial \hat{\theta}_j(w)}{\partial w_r} \right\} \frac{\partial \hat{\theta}_i(w)}{\partial w_s} \right] + \sum_{i=1}^{q} \frac{\partial \eta(\theta)}{\partial \theta_i}\Big|_{\theta=\hat{\theta}} \cdot \frac{\partial^2 \hat{\theta}_i(w)}{\partial w_r \partial w_s}. \tag{16}$$

It is necessary to obtain $\partial^2\eta/\partial\theta_j\partial\theta_i|_{\theta=\hat{\theta}}$ and $\partial^2\hat{\theta}_i(w)/\partial w_r\partial w_s$ to obtain the $\partial^2\eta(\theta)/\partial w_r\partial w_s$ in (16).

The $\partial^2\eta/\partial\theta_j\theta_i|_{\theta=\hat{\theta}}$ is obtained easily from (9). The partial derivative of B_k $(k = 1, 2)$ with respect to w_r and w_s is obtained by twice differentiating both sides of (12).

Hence we get

$$\begin{aligned}
&\left(\frac{\partial^2 B(w)}{\partial w_r \partial w_s}\Big|_{w=1_n}\right)^T A + A^T \left(\frac{\partial^2 B(w)}{\partial w_r \partial w_s}\Big|_{w=1_n}\right) \\
&= -\frac{1}{n}\left[\left\{\left(\frac{\partial B(w)}{\partial w_s}\Big|_{w=1_n}\right)^T + B^T\right\}(x_r - \bar{x})(x_r - \bar{x})^T \left\{B + \left(\frac{\partial B(w)}{\partial w_s}\Big|_{w=1_n}\right)\right\}\right] \\
&\quad - \left(\frac{\partial B(w)}{\partial w_s}\Big|_{w=1_n}\right)^T S \left(\frac{\partial B(w)}{\partial w_r}\Big|_{w=1_n}\right) - \left(\frac{\partial B(w)}{\partial w_r}\Big|_{w=1_n}\right)^T S \left(\frac{\partial B(w)}{\partial w_s}\Big|_{w=1_n}\right) \\
&\quad - \frac{1}{n}\left[\left\{\left(\frac{\partial B(w)}{\partial w_r}\Big|_{w=1_n}\right)^T + B^T\right\}(x_s - \bar{x})(x_s - \bar{x})^T \left\{B + \left(\frac{\partial B(w)}{\partial w_r}\Big|_{w=1_n}\right)\right\}\right] \\
&\quad + \frac{1}{n^2} B^T \left\{(x_r - \bar{x})(x_s - \bar{x})^T + (x_s - \bar{x})(x_r - \bar{x})^T\right\} B.
\end{aligned} \tag{17}$$

Then the $(\partial^2 b_{kk}(w)/\partial w_r \partial w_s)|_{w=1_n}$ is given by $c_{k,k}/(2v_{k,k})$, where each $v_{k,k}$ and $c_{k,k}$ is the (k,k) component of A and the right-hand side of (17).

From this second order approach, we can obtain the first order approach based on the slope. The first order approach gives the direction which maximizes the slope $\dot{\eta}$. The corresponding directional cosines are $\dot{\eta}/\|\dot{\eta}\|$. The first order and the second order methods are complementary measure which reflect a significant influence.

Consider transformation

$$x_r^* = T x_r + b \quad (r = 1, 2, \ldots, n),$$

where T is a $p \times p$ nonsingular lower triangular matrix and b is $p \times 1$ vector. Let $\eta(w, x^*)$ be the perturbed likelihood ratio criterion based on the transformed data x^*. The likelihood function based on x^* under the null model (or full model) is equivalent to the likelihood function based on x under the null model (or full model) multiplied by $|T|^{-1}$. The $\eta(w, x^*)$ reduces to the $\eta(w, x)$. Hence F and D in (7) are invariant under the transformations and so are the local maximum curvatures and their associated direction cosines.

4. Example

The milk transportation-cost data is considered to illustrate the local influence method in testing zero correlation. The data is taken from Johnson and Wichern (1992, p. 276) and is composed of two groups. Each one is measured on three variables. We use the data of Group I for the gasoline trucks, composed of 36 observations. The three variables are fuel, repair and capital per-mile basis. The Group I was previously analyzed by Bacon-Shone and Fung (1987), and Caroni and Prescott (1992). The result is that observations 9 and 21 are possible outliers.

Figure 1 is the index-directional cosines of l_{slope}. It provides observations 21,20,2,23 are influential. The result using the local influence method is as follows. The absolute value of the largest curvature is 9.133×10^{-2} and the second largest one is 5.123×10^{-2}. Thus it seems that the information from the first direction vector l_{max} is a little larger than that of the second direction vector l_{2nd}.

Figure 2 shows the plot of the direction vectors l_{max} and l_{2nd} in which n pairs of components of l_{max} and l_{2nd} in the same position are plotted. From Figure 2, we see that observation 9 is highly significant along l_{max} while observations 36, 21 and 25 are a little significant, and observations 21, 9, 23 and 36 are highly significant along l_{2nd}. They are separated from the rest of data.

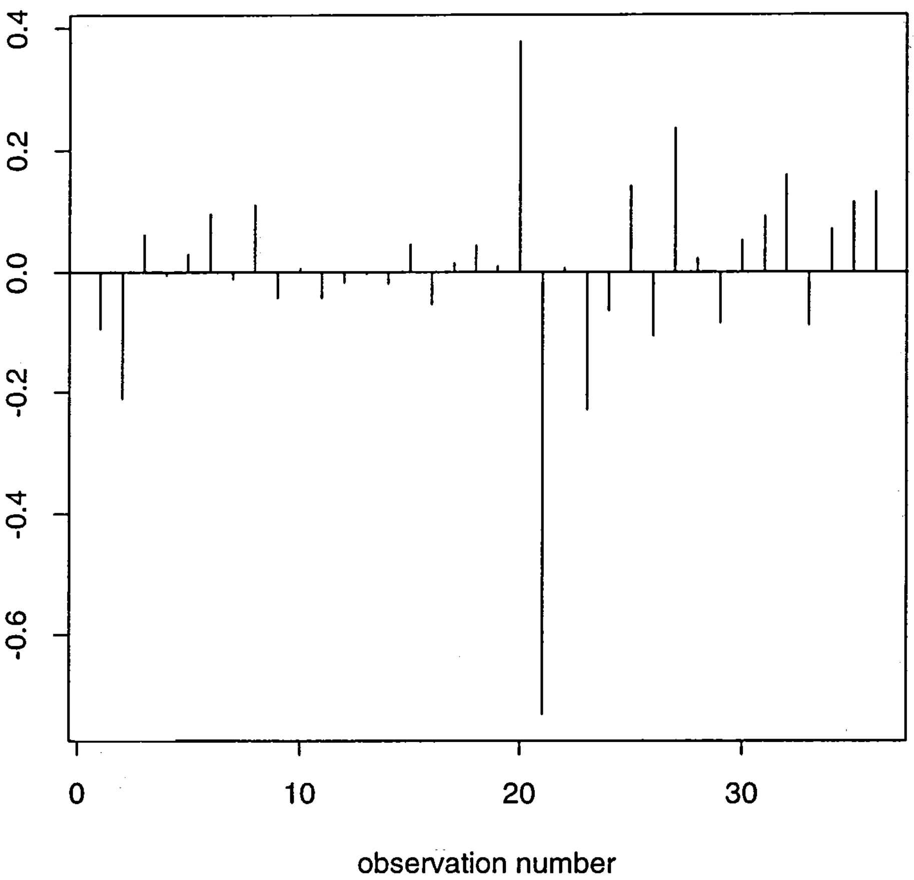

Figure 1. Index plot of the directional cosines corresponding to maximum slope at the null point.

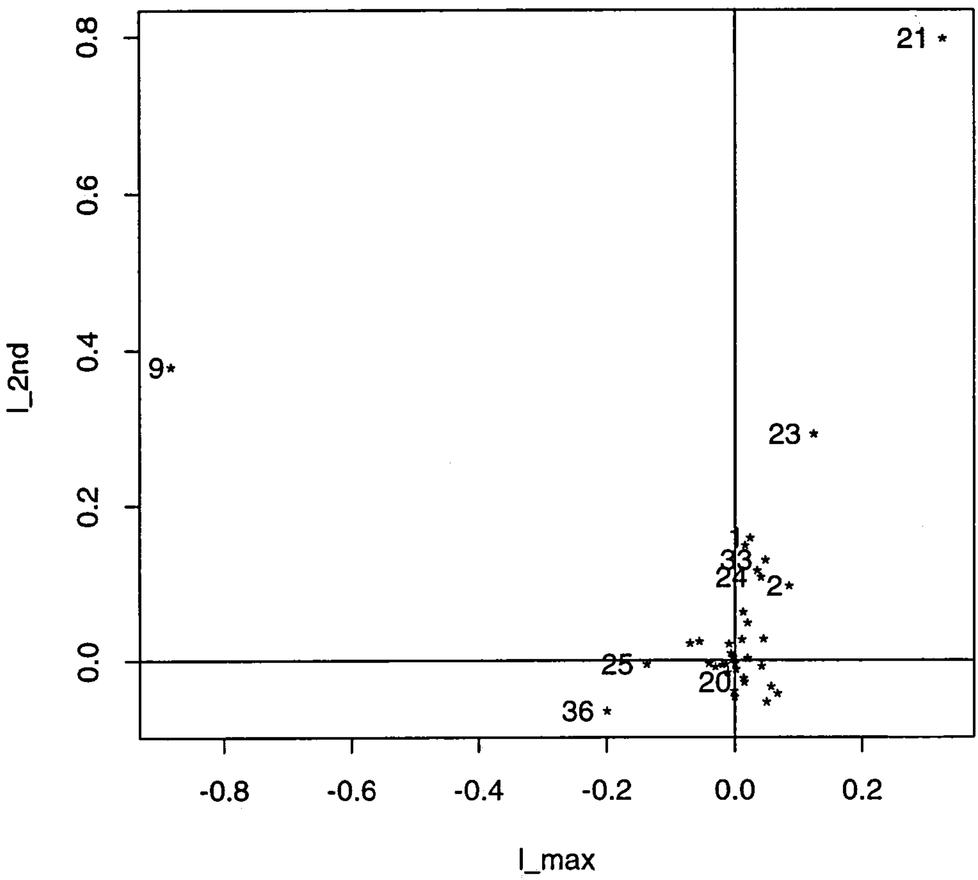

Figure 2. The plot of the first direction vector versus the second direction vector.

Table 1. The influential observations identified by the case-deletion method.

$n(I)$	$\lvert(\eta - \eta_{(I)})/\eta\rvert$	I
	0.2098	21
1	0.1096	20
	0.0999	9
	0.2585	21 23
2	0.2514	2 21
	0.2383	1 21
	0.3781	9 21 23
3	0.3467	9 20 36
	0.3182	9 25 36
	0.4724	2 9 21 23
4	0.4563	1 9 21 23
	0.4534	9 21 23 33
	0.5483	1 2 9 21 23
5	0.5403	2 9 21 23 33
	0.5398	1 9 21 23 33

The case-deletion method which has been generally used for detecting outliers is also applied to this data and the result is included in Table 1. Let $n(I)$ be the number of observations in the index set I of observations deleted and let $\eta_{(I)}$ be the likelihood ratio statistic obtained after deleting observations in I. We examine $|(\eta - \eta_{(I)})/\eta|$, the absolute value of relative difference between η and $\eta_{(I)}$ based on η. In Table 1, $n(I)$, $|(\eta - \eta_{(I)})/\eta|$ and observation numbers in I which have the largest, the second largest and the third largest $|(\eta - \eta_{(I)})/\eta|$ are included for each $n(I) = 1, 2, \ldots, 5$. In Table 1, observation 1 is not individually influential because the value of $|(\eta - \eta_{(I)})/\eta|$ for $I = \{1\}$ is smaller than the tenth largest for $n(I) = 1$. The influence of observation 20 is highly significant only for $n(I) = 1, 3$.

As shown in this example, observations which are highly significant along l_{max} and l_{2nd} are also highly influential for the overall case-deletion except observation 20. But the observation 20 is identified by the first order method.

The local influence result given by the plot of direction vectors corresponding to the largest and the second largest curvatures is compared with result given by case-deletion method which consumes great computation time so that we can know that both results are very similar. Hence the result allows us to save much of the computational effort. Moreover, we can know simply the jointly influential observations by the plot.

REFERENCES

Bacon-Shone, J. and Fung, W. K. (1987). A new graphical method for detecting single and multiple outliers in univariate and multivariate data, *Applied Statistics*, **36**, 153-162.

Caroni, C. and Prescott, P. (1992). Sequential application of Wilk's multivariate outlier test, *Applied Statistics*, **41**, 355-364.

Cook, R. D. (1986). Assessment of local influence, *J. R. Statist. Soc. B*, **48**, No. 2, 133-169.

Johnson, R. A. and Wichern, D. W. (1992). *Applied multivariate statistical analysis*, 3rd ed., Prentice-Hall.

Seber, G. A. F. (1984). *Multivariate observations*, John wiley & Sons.

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*, Oxford University Press.

Wu, X. and Luo, Z. (1993) Second-order approach to local influence, *J. R. Statist. Soc. B*, **55**, No. 4, 929-936.

DEPARTMENT OF MATHEMATICS EDUCATION, CHONGJU NATIONAL UNIVERSITY OF EDUCATION, CHONGJU 361-150, KOREA.