Journal of the Korean Statistical Society (2004), **33**: 2, pp 233–244

LOCAL INFLUENCE ANALYSIS OF THE PROPORTIONAL COVARIANCE MATRICES MODEL

MYUNG GEUN KIM[1] AND KANG-MO JUNG[2]

ABSTRACT

The influence of observations is investigated in fitting proportional covariance matrices model. Local influence measures are obtained when all parameters or subsets of the parameters are of interest. We will also derive the local influence measure for investigating the influence of observations in testing the proportionality of covariance matrices. A numerical example is given for illustration.

AMS 2000 subject classifications. Primary 62H99; Secondary 62J20.
Keywords. Covariance matrix, local influence, proportionality, test.

1. INTRODUCTION

The proportionality between covariance matrices is a fundamental one in comparing covariance matrices (Manly and Rayner, 1987). In genetic studies we often meet the problem of testing the proportionality of covariance matrices. Knowledge of proportionality is important in selection work and in genetic studies on the inheritance of characters (Federer, 1951).

The inference problem in the proportional covariance matrices model has been treated by some authors, for example Guttman *et al.* (1985), Flury (1986) and Eriksen (1987). These authors considered the maximum likelihood estimation of the model parameters, leading to the likelihood ratio test of the proportionality among covariance matrices. The maximum likelihood estimators cannot be expressed in a closed form and are found by iteratively solving the likelihood equations.

Received October 2003; accepted February 2004.

[1]Department of Applied Statistics, Seowon University, Chongju 361-742, Korea (e-mail : mgkim@seowon.ac.kr)

[2]Department of Informatics & Statistics, Kunsan National University, Kunsan 573-701, Korea (e-mail : kmjung@kunsan.ac.kr)

A parametric method gives a good result when the model assumptions are satisfied. We do not know what will happen when distributional assumptions are violated. Even a single outlier may make the sample mean vector meaningless. Therefore the detection of outliers or influential observations is important and it has a long history. It is well known that the sample covariance matrix is sensitive to outlying observations. The maximum likelihood estimators of the model parameters are functions of the sample covariance matrices. Therefore it is necessary to perform the influence analysis in the proportional covariance matrices model. To this end we will adapt the local influence method that was introduced by Cook (1986) as a general method of investigating the influence of observations.

In this paper we derive the local influence measures using Cook's approach (1986) on the parameters of the proportional covariance matrices model and on the test statistic about the proportionality of this model. In Section 2 we review some estimation results for the proportional covariance matrices model. In Section 3 the local influence measures are derived when all parameters or subsets of the parameters are of interest. In Section 4 we derive the local influence measure for investigating the influence of observations in testing the proportionality of covariance matrices. In Section 5 a numerical example will be given for illustration.

2. Proportional Covariance Matrices Model

In this section we review some estimation results for the proportional covariance matrices model.

Let $\mathbf{x}_{kj}$ $(j = 1, \ldots, n_k)$ be a random sample from a p-variate normal distribution $N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ for each $k = 1, \ldots, K$. The K samples are assumed to be independent. The proportionality among the covariance matrices is also assumed to hold and it can be mathematically expressed as the following hypothesis

$$H_p \ : \ \boldsymbol{\Sigma}_k = c_k \boldsymbol{\Sigma}_1, \qquad k = 2, \ldots, K, \tag{2.1}$$

for some unknown positive constants $c_2, \ldots, c_K$. For convenience we introduce $c_1 = 1$. Let $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$ and $\mathbf{c} = (c_2, \ldots, c_K)^T$. We denote by $\mathbf{S}_k$ the sample covariance matrix with divisor n_k for the k^{th} group. Then $n_k \mathbf{S}_k$ are independently distributed according to Wishart distribution with expectation $\boldsymbol{\Sigma}_k$ and n_k degrees of freedom. The log likelihood function, ignoring unimportant terms, can be written as

$$L(\mathbf{c}, \boldsymbol{\Sigma}^{-1}) = \frac{1}{2}\left[n \log\left|\boldsymbol{\Sigma}^{-1}\right| - \sum_{k=1}^{K} n_k \left\{p \log(c_k) + \frac{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_k)}{c_k}\right\}\right], \qquad (2.2)$$

where $n = \sum_{k=1}^{K} n_k$. The maximum likelihood estimators of c_k and $\boldsymbol{\Sigma}$ are found by iteratively solving the following two equations

$$\widehat{c_k} = \frac{1}{p}\mathrm{tr}\left(\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_k\right),$$

$$\widehat{\boldsymbol{\Sigma}} = r_1\mathbf{S}_1 + \sum_{k=2}^{K} \frac{r_k}{\widehat{c_k}}\mathbf{S}_k,$$

where $r_k = n_k/n$. More details can be found in Manly and Rayner (1987).

3. Local Influence

In this section we briefly review the local influence method introduced by Cook (1986) and then derive local influence measures for the proportional covariance matrices model when all parameters or subsets of the parameters are of interest.

Let $\mathbf{w} = (\mathbf{w}_1^T, \ldots, \mathbf{w}_K^T)^T$ be an n-dimensional vector of perturbations, where $\mathbf{w}_k = (w_{k1}, \ldots, w_{kn_k})^T$ is the n_k-dimensional vector. We consider the perturbed model in which the j^{th} observation in the k^{th} group $\mathbf{x}_{kj}$ is perturbed according to

$$\mathbf{x}_{kj} \sim N\left(\boldsymbol{\mu}_k, w_{kj}^{-1}\boldsymbol{\Sigma}_k\right) \qquad (3.1)$$

for $k = 1, \ldots, K$ and $j = 1, \ldots, n_k$. We write $\mathbf{1}_n$ as the n-dimensional vector with all elements equal to 1. When $\mathbf{w} = \mathbf{1}_n$, the perturbed model reduces to the unperturbed model.

Let $\boldsymbol{\theta}$ be the q-dimensional vector of parameters formed by stacking $c_2, \ldots, c_K$ and $\mathrm{vech}(\boldsymbol{\Sigma}^{-1})$, where $q = K-1+p(p+1)/2$ and $\mathrm{vech}(\mathbf{A})$ denotes the $p(p+1)/2$-dimensional vector formed by stacking the columns of the lower triangular portion of a $p \times p$ symmetric matrix $\mathbf{A}$. We denote the log-likelihoods for the unperturbed and perturbed models by $L(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta}|\mathbf{w})$, respectively. The likelihood displacement $LD(\mathbf{w})$ is defined by $LD(\mathbf{w}) = 2\{L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_w)\}$, where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_w$ are the maximum likelihood estimators of $\boldsymbol{\theta}$ under the unperturbed and perturbed models, respectively. Define the $q \times n$ matrix

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta} \mid \mathbf{w})}{\partial\boldsymbol{\theta}\partial\mathbf{w}^T}$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\mathbf{w} = \mathbf{1}_n$, and the $q \times q$ matrix

$$\ddot{\mathbf{G}} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Let $\ddot{\mathbf{F}}$ be the $n \times n$ matrix defined by

$$\ddot{\mathbf{F}} = \boldsymbol{\Delta}^T \ddot{\mathbf{G}}^{-1} \boldsymbol{\Delta}.$$

Note that $\partial \widehat{\boldsymbol{\theta}}_w / \partial \boldsymbol{w}^T|_{\mathbf{w}=\mathbf{1}_n} = -\ddot{\mathbf{G}}^{-1}\boldsymbol{\Delta}$ and $\ddot{\mathbf{F}} = L(\widehat{\boldsymbol{\theta}}_{\mathbf{w}})/\partial \mathbf{w} \partial \mathbf{w}^T|_{\mathbf{w}=\mathbf{1}_n}$.

Let $\boldsymbol{l}_{\max}$ be the eigenvector corresponding to the largest absolute eigenvalue of $-2\ddot{\mathbf{F}}$. And let $\mathbf{1}_{(i)}$ be the i-dimensional vector with its i^{th} element equal to 1 and the others being zero. The surface of interest is formed by the $(n+1)$-dimensional vector of the values $\mathbf{w}$ and $LD(\mathbf{w})$ as $\mathbf{w}$ varies over a certain space. Then the largest absolute eigenvalue is the maximum curvature of the curve which is the portion of the surface cut out by the plane spanned by the vectors $\mathbf{1}_{(n+1)}$ and $(\boldsymbol{l}_{\max}^T, 0)^T$. An index plot of the direction cosines of $\boldsymbol{l}_{\max}$ may be helpful for identifying influential observations, because this vector $\boldsymbol{l}_{\max}$ indicates how to perturb the model to obtain the greatest local change in the likelihood displacement (Cook, 1986, pp. 138–139). Observations that have large absolute direction cosines are potentially influential.

3.1. All parameters are of interest

Let $\mathbf{D}_p$ and vec$(\mathbf{A})$ be the duplication matrix and the p^2-dimensional vector formed by stacking the columns of $\mathbf{A}$, respectively such that $\mathbf{D}_p \text{vech}(\mathbf{A}) = \text{vec}(\mathbf{A})$ (Magnus and Neudecker, 1988, p. 49). Since $\text{tr}(\mathbf{S}_k \boldsymbol{\Sigma}^{-1}) = \text{vec}^T(\mathbf{S}_k)\mathbf{D}_p \text{vech}(\boldsymbol{\Sigma}^{-1})$ in (2.2), we obtain the following partial derivatives evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ to get $\ddot{\mathbf{G}}$

$$\begin{aligned} \frac{\partial^2 L}{\partial c_k^2} &= -\frac{n_k p}{2\widehat{c}_k^2}, \\ \frac{\partial^2 L}{\partial c_k \partial \text{vech}^T(\boldsymbol{\Sigma}^{-1})} &= \frac{n_k}{2\widehat{c}_k^2} \text{vec}^T(\mathbf{S}_k)\mathbf{D}_p, \\ \frac{\partial^2 L}{\partial \text{vech}(\boldsymbol{\Sigma}^{-1}) \partial \text{vech}^T(\boldsymbol{\Sigma}^{-1})} &= -\frac{n}{2}\mathbf{D}_p^T(\widehat{\boldsymbol{\Sigma}} \otimes \widehat{\boldsymbol{\Sigma}})\mathbf{D}_p, \end{aligned}$$

where $\otimes$ denotes the Kronecker product of matrices. The last equality is easily obtained since

$$\frac{\partial}{\partial \text{vech}(\boldsymbol{\Sigma}^{-1})} \log |\boldsymbol{\Sigma}^{-1}| = \mathbf{D}_p^T \text{vec}(\boldsymbol{\Sigma}), \quad \frac{\partial}{\partial \text{vech}^T(\boldsymbol{\Sigma}^{-1})} \text{vec}(\boldsymbol{\Sigma}) = -(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}_p.$$

Thus we have

$$\ddot{\mathbf{G}} = -\frac{1}{2}\begin{pmatrix} p\ \mathrm{diag}(n_2/\widehat{c}_2^2, \ldots, n_K/\widehat{c}_K^2) & \mathbf{G}_{21}^T \\ \mathbf{G}_{21} & n\mathbf{D}_p^T(\widehat{\boldsymbol{\Sigma}} \otimes \widehat{\boldsymbol{\Sigma}})\mathbf{D}_p \end{pmatrix}, \tag{3.2}$$

where $\mathbf{G}_{21} = [-(n_2/\widehat{c}_2^2)\mathbf{D}_p^T\mathrm{vec}(\mathbf{S}_2), \ldots, -(n_K/\widehat{c}_K^2)\mathbf{D}_p^T\mathrm{vec}(\mathbf{S}_K)]$ has the size $p(p+1)/2 \times (K-1)$.

Under the perturbation (3.1), we have the maximum likelihood estimators of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ as

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_k(\mathbf{w}) &= \frac{\sum_{j=1}^{n_k} w_{kj}\mathbf{x}_{kj}}{\sum_{j=1}^{n_k} w_{kj}}, \\ \mathbf{S}_k(\mathbf{w}) &= \frac{1}{n_k}\sum_{j=1}^{n_k} w_{kj}\big(\mathbf{x}_{kj} - \widehat{\boldsymbol{\mu}}_k(\mathbf{w})\big)\big(\mathbf{x}_{kj} - \widehat{\boldsymbol{\mu}}_k(\mathbf{w})\big)^T, \end{aligned}$$

respectively, and they are independent. The $n_k\mathbf{S}_k(\mathbf{w})$ $(k = 1, \ldots, K)$ are independently distributed as the Wishart distribution with expectation $\boldsymbol{\Sigma}_k$ and n_k degrees of freedom under the perturbed model. For the proportional covariance matrices model (2.1), the log-likelihood function for the perturbed model, ignoring unimportant terms, can be written as

$$L\big(\mathbf{c}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{w}\big) = \frac{1}{2}\left[n\log\left|\boldsymbol{\Sigma}^{-1}\right| - \sum_{k=1}^{K} n_k\left\{p\log(c_k) + \frac{\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\mathbf{S}_k(\mathbf{w})\big)}{c_k}\right\}\right].$$

The matrix $\boldsymbol{\Delta}$ in $\ddot{\mathbf{F}}$ is obtained by using the following partial derivatives evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\mathbf{w} = \mathbf{1}_n$. The identity

$$\frac{\partial}{\partial w_{ij}}\mathbf{S}_k(\mathbf{w}) = \frac{1}{n_k}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T I(i = k) \quad \text{for } k = 1, \ldots, K, \tag{3.3}$$

where $I(\cdot)$ is the indicator function, yields

$$\frac{\partial}{\partial w_{ij}}L\big(\mathbf{c}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{w}\big) = -\frac{1}{2c_k}\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T\big)I(i = k).$$

Thus we have

$$\frac{\partial^2}{\partial c_k \partial w_{ij}}L\big(\mathbf{c}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{w}\big) = \frac{1}{2\widehat{c}_k^2}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)I(i = k) \tag{3.4}$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\mathbf{w} = \mathbf{1}_n$. Since

$$\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T\big) = \mathrm{vech}^T\big(\boldsymbol{\Sigma}^{-1}\big)\mathbf{D}_p^T\mathrm{vec}\big((\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T\big).$$

we have

$$\frac{\partial^2}{\partial \mathrm{vech}(\boldsymbol{\Sigma}^{-1})\partial w_{kj}} L(\mathbf{c}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{w}) = -\frac{1}{2\widehat{c}_k}\mathbf{D}_p^T \mathrm{vec}\big((\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T\big). \quad (3.5)$$

The $q \times n$ matrix $\boldsymbol{\Delta}$ is obtained in a partitioned form as $\boldsymbol{\Delta}^T = (\boldsymbol{\Delta}_1^T, \boldsymbol{\Delta}_2^T)$, where the components of $\boldsymbol{\Delta}_1^T$ and $\boldsymbol{\Delta}_2^T$ are given by (3.4) and (3.5), respectively. Therefore, we can get $\ddot{\mathbf{F}} = \boldsymbol{\Delta}^T \ddot{\mathbf{G}}^{-1} \boldsymbol{\Delta}$ using (3.2), (3.4) and (3.5).

3.2. Subsets of parameters are of interest

Suppose that $\boldsymbol{\theta}$ is partitioned into $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$, where only the subset $\boldsymbol{\theta}_1$ is of interest. In this case Cook (1986) obtained the normal curvature as the eigenvalue of $-2\ddot{\mathbf{H}}$ with

$$\ddot{\mathbf{H}} = \boldsymbol{\Delta}^T (\ddot{\mathbf{G}}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}, \quad (3.6)$$

where

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{G}_{\boldsymbol{\theta}_2}^{-1} \end{pmatrix}$$

and $\mathbf{G}_{\boldsymbol{\theta}_2} = \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^T$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$.

When $c_2, \ldots, c_K$ are of interest, we have $\mathbf{G}_{\boldsymbol{\theta}_2}^{-1} = -(2/n)\{\mathbf{D}_p^T (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \mathbf{D}_p\}^{-1}$. From (3.6) we get

$$\ddot{\mathbf{H}} = \ddot{\mathbf{F}} - \boldsymbol{\Delta}_2^T \mathbf{G}_{\boldsymbol{\theta}_2}^{-1} \boldsymbol{\Delta}_2.$$

Next consider the case in which only $\boldsymbol{\Sigma}$ is of interest. We rearrange $\boldsymbol{\Delta}$ and $\ddot{\mathbf{G}}$ so that the terms related to $\boldsymbol{\Sigma}$ appear in the left upper corner. Similarly to the previous case we obtain

$$\mathbf{G}_{\boldsymbol{\theta}_2}^{-1} = -\frac{2}{p} \mathrm{diag}\left(\frac{\widehat{c}_2^2}{n_2}, \ldots, \frac{\widehat{c}_K^2}{n_K}\right)$$

and

$$\ddot{\mathbf{H}} = \ddot{\mathbf{F}} - \boldsymbol{\Delta}_1^T \mathbf{G}_{\boldsymbol{\theta}_2}^{-1} \boldsymbol{\Delta}_1.$$

4. INFLUENCE ON A TEST OF FITTING THE PROPORTIONAL COVARIANCE MODEL

In this section we derive the local influence measure for investigating the influence of observations in testing the proportionality of covariance matrices.

A test of the hypothesis (2.1) is usually performed by using the likelihood ratio statistic given by

$$T = \sum_{k=1}^{K} n_k \Big\{ p \log(\widehat{c}_k) - \log \big|\widehat{\boldsymbol{\Sigma}}^{-1}\big| - \log \big|\mathbf{S}_k\big| \Big\},$$

where $\widehat{c}_1 = 1$ for convenience. Then T is approximately distributed as a chi-squared distribution with $(K-1)(p^2+p-2)/2$ degrees of freedom. The null hypothesis H_p in (2.1) would be rejected for a significantly large value of T.

To investigate the influence of observations on the likelihood ratio statistic T, we consider the perturbed statistic $T(\mathbf{w})$ under the perturbation scheme (3.1). The maximum likelihood estimators of c_k and $\boldsymbol{\Sigma}$ under the perturbed model are found by iteratively solving the following equations

$$\widehat{c}_k(\mathbf{w}) = \frac{1}{p} \operatorname{tr}\big(\mathbf{S}_k(\mathbf{w}) \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\big), \tag{4.1}$$

$$\widehat{\boldsymbol{\Sigma}}(\mathbf{w}) = r_1 \mathbf{S}_1(\mathbf{w}) + \sum_{k=2}^{K} \frac{r_k}{\widehat{c}_k(\mathbf{w})} \mathbf{S}_k(\mathbf{w}). \tag{4.2}$$

Then the perturbed statistic becomes

$$T(\mathbf{w}) = \sum_{k=1}^{K} n_k \Big\{ p \log\big(\widehat{c}_k(\mathbf{w})\big) - \log \big|\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\big| - \log \big|\mathbf{S}_k(\mathbf{w})\big| \Big\},$$

where $\widehat{c}_1(\mathbf{w}) = 1$ for convenience. We have a surface $(\mathbf{w}^T, T(\mathbf{w}))$ from which an influence measure can be obtained, and a large absolute value of $\partial T(\mathbf{w})/\partial w_{kj}$ evaluated at $\mathbf{w} = \mathbf{1}_n$ implies that the corresponding observation vector $\mathbf{x}_{kj}$ is possibly influential (Lawrance, 1988). The first order derivative of $T(\mathbf{w})$ with respect to $\mathbf{w}$ evaluated at $\mathbf{w} = \mathbf{1}_n$ is given by

$$\begin{aligned} \dot{T} = \frac{\partial T(\mathbf{w})}{\partial \mathbf{w}} = & -\sum_{k=1}^{K} n_k \frac{1}{|\mathbf{S}_k|} \left(\frac{\partial}{\partial \mathbf{w}} |\mathbf{S}_k(\mathbf{w})| \right) - n \frac{1}{\big|\widehat{\boldsymbol{\Sigma}}^{-1}\big|} \left(\frac{\partial}{\partial \mathbf{w}} \big|\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\big| \right) \\ & + \sum_{k=2}^{K} n_k \frac{p}{\widehat{c}_k} \left(\frac{\partial}{\partial \mathbf{w}} \widehat{c}_k(\mathbf{w}) \right). \end{aligned} \tag{4.3}$$

Now we compute equation (4.3). First, we derive the first term of (4.3). Since

$$\frac{\partial}{\partial w_{ij}} |\mathbf{S}_k(\mathbf{w})| = |\mathbf{S}_k(\mathbf{w})| \operatorname{tr}\left(\mathbf{S}_k(\mathbf{w})^{-1} \left\{ \frac{\partial}{\partial w_{ij}} \mathbf{S}_k(\mathbf{w}) \right\} \right).$$

(3.3) yields

$$\frac{\partial}{\partial w_{ij}}\big|\mathbf{S}_k(\mathbf{w})\big| = \frac{1}{n_k}\big|\mathbf{S}_k\big|(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k)^T\mathbf{S}_k^{-1}(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k)I(i=k).$$

The element of the first term in (4.3) corresponding to w_{kj} becomes

$$-(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k)^T\mathbf{S}_k^{-1}(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k). \tag{4.4}$$

Second, similarly to the first order derivative of $|\mathbf{S}_k(\mathbf{w})|$, the identity

$$\frac{\partial}{\partial w_{kj}}\left|\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\right| = \left|\widehat{\boldsymbol{\Sigma}}^{-1}\right| \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\left\{\frac{\partial}{\partial w_{kj}}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\right\}\right)$$

yields the element of the second term in (4.3) corresponding to w_{kj} as

$$-n\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1}\right), \tag{4.5}$$

where $\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1} = \partial\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})/\partial w_{kj}$ evaluated at $\mathbf{w}=\mathbf{1}_n$.

Finally, we derive the element of the last term in (4.3) corresponding to w_{kj}. For convenience let $\dot{\mathbf{S}}_{i,kj}$ be $\partial\mathbf{S}_i(\mathbf{w})/\partial w_{kj}$ evaluated at $\mathbf{w}=\mathbf{1}_n$ given in (3.3). Since (4.1) gives

$$\frac{\partial}{\partial w_{kj}}\widehat{c}_i(\mathbf{w}) = \frac{1}{p}\left\{\operatorname{tr}\left(\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1}\mathbf{S}_i\right) + \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}^{-1}\dot{\mathbf{S}}_{i,kj}\right)\right\},$$

we have

$$\begin{aligned}\sum_{i=2}^{K} n_i\frac{p}{\widehat{c}_i}\left\{\frac{\partial\widehat{c}_i(\mathbf{w})}{\partial w_{kj}}\right\} &= n\operatorname{tr}\left(\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1}(\widehat{\boldsymbol{\Sigma}}-r_1\mathbf{S}_1)\right) + \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{i=1}^{K}\frac{n_i}{\widehat{c}_i}\dot{\mathbf{S}}_{i,kj} - n_1\dot{\mathbf{S}}_{1,kj}\right)\right)\\ &= n\operatorname{tr}\left(\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1}\widehat{\boldsymbol{\Sigma}}\right) + (\mathbf{x}_{kj}-\bar{\mathbf{x}}_k)^T\widehat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k),\end{aligned} \tag{4.6}$$

where $\widehat{\boldsymbol{\Sigma}}_k = \widehat{c}_k\widehat{\boldsymbol{\Sigma}}$. The last equality in (4.6) is due to the fact that

$$0 = \operatorname{tr}\left(\dot{\widehat{\boldsymbol{\Sigma}}}_{kj}^{-1}\mathbf{S}_1\right) + \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}^{-1}\dot{\mathbf{S}}_{1,kj}\right)$$

obtained by differentiating $p = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{w})\mathbf{S}_1(\mathbf{w}))$ with respect to w_{kj}.

Therefore, equations (4.4) to (4.6) yield the first order derivative of $T(\mathbf{w})$ with respect to w_{kj} evaluated at $\mathbf{w}=\mathbf{1}_n$ as

$$\dot{T}_{kj} = (\mathbf{x}_{kj}-\bar{\mathbf{x}}_k)^T\left(\widehat{\boldsymbol{\Sigma}}_k^{-1} - \mathbf{S}_k^{-1}\right)(\mathbf{x}_{kj}-\bar{\mathbf{x}}_k). \tag{4.7}$$

A large absolute value of $\dot{T}_{kj}$ indicates that observation $\mathbf{x}_{kj}$ is influential in testing H_p in (2.1).

5. NUMERICAL EXAMPLE

For illustrating the local influence method, we consider the painted turtle data (Johnson and Wichern, 1992, p. 364, Table 6.7) in which 24 measurements on three variables (the shell length, width and height) are obtained for each of female and male turtles. The observations are labelled as 1 to 24 for female (the first group) and 25 to 48 for male (the second group). In what follows our analysis will be confined to the first two variables (the shell length and width).

The maximum likelihood estimates of c_2 and $\widehat{\mathbf{\Sigma}}$ are given by

$$\widehat{c}_2 = 0.407 \quad \text{and} \quad \widehat{\mathbf{\Sigma}} = \begin{pmatrix} 379.83 & 223.08 \\ 223.08 & 141.24 \end{pmatrix},$$

respectively. The likelihood ratio statistic T for testing the proportionality given in Section 4 is 1.32 and the corresponding p-value is 0.52. Thus we can assume at reasonable significance levels that the proportionality between two covariance matrices exists.

First, we investigate the influence of observations for the proportional covariances model with the index plot of $\boldsymbol{l}_{\max}$ given in Figure 5.1 when all parameters are of interest. The largest eigenvalue of $\ddot{\mathbf{F}}$ is 4.37. From Figure 5.1 we can see that observation 48 is distinctly separated from the other observations. Observations 23, 24, 25 and 47 are located at the outer side of the main body. Hence observation 48 has the largest influence in fitting the proportional covariance matrices model and the influence of the other four observations is not severe compared with that of observation 48. This result can be confirmed by using the likelihood distance in the next paragraph.

For confirmation of the results in Figure 5.1, we compute the likelihood distance LD_r defined by

$$LD_r = 2\Big\{L\Big(\widehat{\mathbf{c}}, \widehat{\mathbf{\Sigma}}^{-1}\Big) - L\Big(\widehat{\mathbf{c}}_{(r)}, \widehat{\mathbf{\Sigma}}_{(r)}^{-1}\Big)\Big\},$$

where $L(\widehat{\mathbf{c}}, \widehat{\mathbf{\Sigma}}^{-1})$ in (2.2) is the log likelihood based on the full data and $\widehat{\mathbf{c}}_{(r)}, \widehat{\mathbf{\Sigma}}_{(r)}^{-1}$ are the maximum likelihood estimates based on the reduced data without the r^{th} observation (Cook and Weisberg, 1982). The results are summarized in the left half of Table 5.1. Numbers in Table 5.1 are arranged in decreasing order of LD_r. We observe that observation 48 has the largest influence and the influence of the other four observations is not large compared with that of observation 48. Thus the result of the likelihood distance is similar to that of the local influence

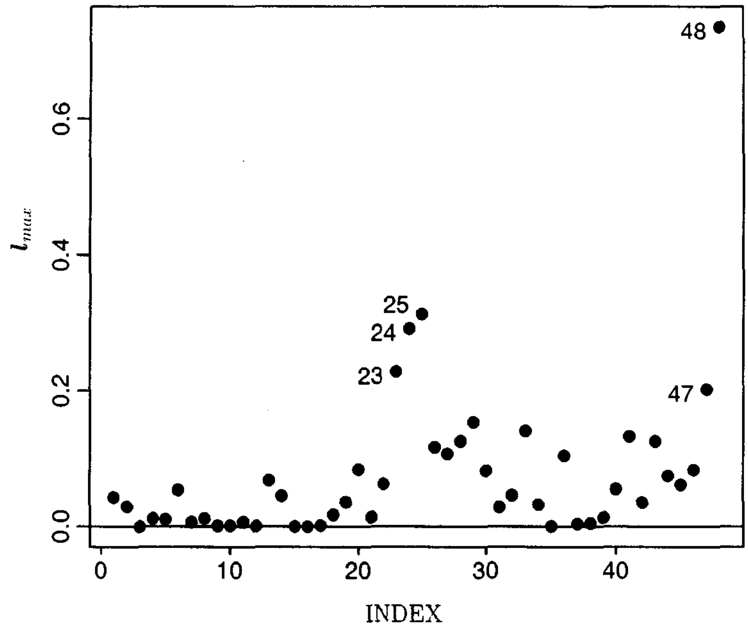

FIGURE 5.1 *The index plot of $l_{\max}$ when all parameters are of interest*

method. However, computation of LD_r is a tedious work due to iterations for getting $\widehat{\mathbf{c}}_{(r)}$ and $\widehat{\Sigma}_{(r)}^{-1}$ while it is simple to perform the local influence method.

Next we investigate the influence of observations on the test statistic about the hypothesis H_p in (2.1). The index plot of $\dot{T}$ given in (4.7) is presented in Figure 5.2. From this plot we may conclude that observations 13, 18, 24, 25 and 48 are candidates for influential observations in testing H_p. Similarly to the case of model assessment, we conduct the deletion influences by computing

TABLE 5.1 *Likelihood distance and deletion influence on the likelihood ratio test statistic in proportional covariances model*

r	LD_r	r	$T_{(r)}$	p-value
48	1.96	48	2.97	0.23
24	0.84	18	2.23	0.33
23	0.42	25	2.18	0.34
47	0.33	13	2.04	0.36
25	0.20	24	0.82	0.66

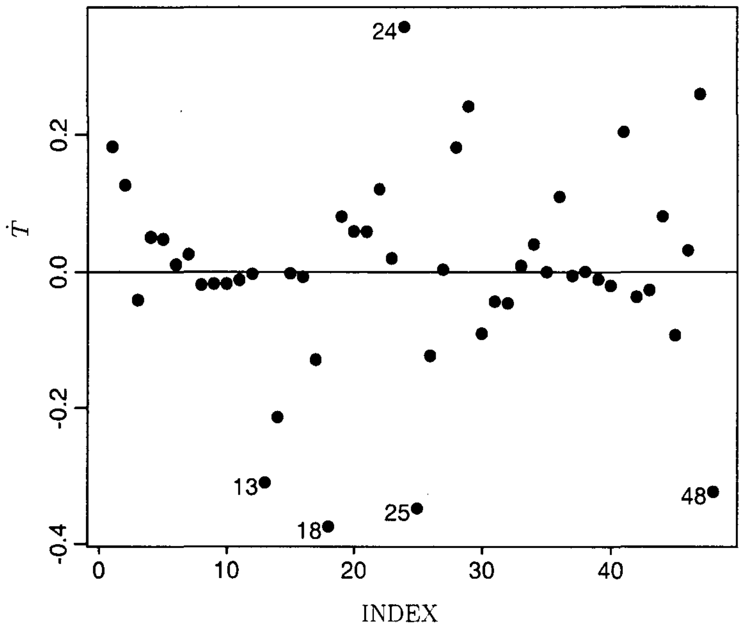

FIGURE 5.2 *Local influence measure $\dot{T}$ for the test statistic*

$T_{(r)}$, where $T_{(r)}$ denotes the test statistic after deletion of the corresponding observation. The case-deletion is a simple method for influence diagnostics. The deletion result is summarized in the right half of Table 5.1 and numbers are arranged in decreasing order of $|T - T_{(r)}|$. The p-value in Table 5.1 represents the p-value of the likelihood ratio statistic T based on the remaining data without the corresponding observation. Recall that the p-value for the full data set is 0.52. Table 5.1 indicates that the omission of observation 48 makes the greatest change in the p-value. Deletion of observation 24 enhances our belief about the proportionality while deletion of each of the other four observations reduces it. From Figure 5.2 we note that observation 24 is located in the upper part while the other four observations are located in the lower part. The results Figures 5.1 and 5.2 are different and they reflect the local influences of the parameters and the test statistic in the proportional covariance matrices model, respectively.

This example shows that the local influence method is useful for getting information about influential observations.

REFERENCES

COOK, R. D. (1986). "Assessment of local influence (with discussion)", *Journal of the Royal Statistical Society*, **B48**, 133–169.

COOK, R. D. AND WEISBERG, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, London.

ERIKSEN, P. S. (1987). "Proportionality of covariance matrices", *The Annals of Statistics*, **15**, 732–748.

FEDERER, W. T. (1951). "Testing proportionality of covariance matrices", *The Annals of Mathematical Statistics*, **22**, 102–106.

FLURY, B. K. (1986). "Proportionality of k covariance matrices", *Statistics & Probability Letters*, **4**, 29–33.

GUTTMAN, I., KIM, D. Y. AND OLKIN, I. (1985). "Statistical inference for constants of proportionality between covariance matrices", In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.), 257–280, North-Holland, New York.

JOHNSON, R. A. AND WICHERN, D. W. (1992). *Applied Multivariate Statistical Analysis*, 3rd ed., Prentice-Hall, Englewood Cliffs.

LAWRANCE, A. J. (1988). "Regression transformation diagnostics using local influence", *Journal of the American Statistical Association*, **83**, 1067–1072.

MAGNUS, J. R. AND NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, Chichester.

MANLY, B. F. J. AND RAYNER, J. C. W. (1987). "The comparison of sample covariance matrices using likelihood ratio tests", *Biometrika*, **74**, 841–847.