Journal of the Faculty of Enviromental Science and Technology,
Okayama University
Vol. 7, No.1. pp. 23-29, March 2002

Studies on Covariance Selection Models :

Stepwise Procedure and Local Influence

JiMin SUNG* , WooSung YANG† and Yutaka TANAKA**

(Received November 30, 2001)

Analysis of covariance selection models is a useful multivariate method to analyze the covariance structure of a multivariate normal distribution. It is used to reveal cause-and-effect relationships. In the present paper we review the theory and study numerically how the stepwise procedure of covariance selection works in actual data analysis. Then we try to develop a method of influence analysis in covariance selection, and show a numerical example to illustrate the usefulness of the method of influence analysis.

keywords: *Covariance Selection, Local Influence, Influence Function, Stepwise Procedure*

1. INTRODUCTION

Graphical gaussian model or covariance selection model is originally proposed by Dempster (19 72) as a means of parameter reduction when the covariance structure of multivariate normal distribution is to be estimated. This method is characterized by specified variable pairs that have zero partial correlations, and the result of analysis is often expressed by a linear graph which consists of nodes and arcs. In this sense the analysis of covariance selection models is known as a member of the family of graphical modeling, which has been highlighted as a set of recently developed multivariate techniques to analyze cause-and-effect relationships in complex phenomena. This analysis seems interesting and effective in revealing causal relationships. It may be sensitive to outlying observations, however, like other multivariate methods such as principal component analysis and factor analysis. So it will be valuable to develop a method of influence analysis. In the present paper we first review the theory of covariance selection models and study numerically how a stepwise procedure of selecting partial correlations works in actual data analysis. Then we try to develop a method for assessing local influence in the analysis and show a numerical example to illustrate the usefulness of the method.

*Graduate School of Natural Science and Technology, Master Course, Okayama University
†Graduate School of Natural Science and Technology, Doctor Course, Okayama University
**Department of Environmental and Mathematical Sciences, Okayama University

2. ANALYSIS OF COVARIANCE SELECTION MODELS

Suppose we have observations $\{\underline{x}_i, i = 1, \cdots, n\}$, each of which follows independently a p-variate normal distribution $N(\underline{\mu}, \Sigma)$. A covariance selection model or graphical gaussian model is defined by specifying that some elements of inverse covariance matrix $\Phi = \Sigma^{-1} = (\phi_{ij})$ are zero, i.e.,

$$\phi_{ij} = 0 \ \text{ for } (i,j) \in I,$$

where I indicates a subset of index pairs of $\Omega = \{(i,j),\ i,j = 1, \cdots, p,\ i < j\}$. The remaining elements ϕ_{ij} for $(i,j) \in J$, are not specified, where J indicates the compliment of I in the whole set. It is known that $\phi_{ij} = 0$ is equivalent to the fact that the partial correlation between variables i and j is zero.

Parameters $\phi_{ij},\ (i,j) \in J$ can be estimated by maximizing the profile log-likelihood function

$$l(\Phi, S) = \frac{n}{2}\log|\Phi| - \frac{n}{2}tr(\Phi S),$$

where Φ contains unknown parameters ϕ_{ij}, $(i,j) \in J$ and zeros in the remaining elements. The maximum likelihood estimate $\hat{\underline{\phi}}_1$ satisfies

$$\frac{\partial l}{\partial \phi_{ij}} = 0, \ (i,j) \in J,$$

where

$$\begin{aligned} \frac{\partial l}{\partial \phi_{ij}} &= \frac{n}{2} tr(\Phi^{-1} E_{ij}) - \frac{n}{2} tr(E_{ij} S) \\ &= \begin{cases} n(\phi^{ij} - s_{ij}), & i \neq j, \\ \frac{n}{2}(\phi^{ii} - s_{ii}), & i = j, \end{cases} \end{aligned} \tag{1}$$

$\underline{\phi}_1$ indicating a vector of ϕ_{ij} for $(i,j) \in J$. Here E_{ij} is a $p \times p$ matrix which has 1's as the (i,j)-th and (j,i)-th elements and 0's as the other elements, and $\phi^{ij}(=\sigma_{ij})$is the (i,j)-th element of Φ^{-1}. An iterative algorithm is given by Wermuth and Scheidt(1977) for computing $\hat{\underline{\phi}}_1$ in a typical covariance selection model, where several elements of Φ are forced to zero. They use 'INVEST operator', which gives the closed form of the maximum likelihood estimate for the inverse covariance matrix with one zero element. For several zero elements the INVEST-operator is applied to each of the prespecified or selected variable pairs in turn repeatedly. The cycling ends when all these elements are close enough to zero. Based on the standard theory of maximum likelihood estimation a consistent estimate V_{11} for the asymptotic covariance matrix $acov(\hat{\underline{\phi}}_1)$ can be obtained by inverting the Hessian matrix of $-l$ with respect to $\underline{\phi}_1$, i.e.,

$$V_{11} = \left[-\ddot{l}_{11}\right]^{-1} = \left[-\frac{\partial^2 l}{\partial \underline{\phi}_1 \partial \underline{\phi}_1^T}\right]^{-1}_{\phi_1, \phi_2 = 0}$$

where the elements of the Hessian matrix are given by

$$\begin{aligned} \frac{\partial^2 l}{\partial \phi_{ij} \partial \phi_{kl}} &= \frac{n}{2} tr\left(-\Phi^{-1} \frac{\partial \Phi}{\partial \phi_{kl}} \Phi^{-1} \frac{\partial \Phi}{\partial \phi_{ij}}\right) \\ &= -\frac{n}{2} tr(\Phi^{-1} E_{kl} \Phi^{-1} E_{ij}), \\ &\quad \text{for } (i,j),(k,l) \in J. \end{aligned} \tag{2}$$

The above discussions can be applied to the case where there is no constraint on partial correlations, except that ordinary procedure of maximum likelihood estimation is used instead of the iterative application of the INVEST-operator.

Now let $\underline{\phi}_1$ be partitioned to $\underline{\phi}_1 = (\underline{\phi}_2^T, \underline{\phi}_3^T)^T$ and suppose we are interested in $\underline{\phi}_2$. Then it is well known that the maximum likelihood estimate $\hat{\underline{\phi}}_2$ is given from partitioned vector $\hat{\underline{\phi}}_1 = (\hat{\underline{\phi}}_2^T, \hat{\underline{\phi}}_3^T)^T$ and a consistent estimate V_{22} for the asymptotic covariance matrix $acov(\hat{\underline{\phi}}_2)$ is given by the corresponding part of V_{11}, or more precisely,

$$V_{22} = \left[-\ddot{l}_{22\cdot 3}\right]^{-1} = -\left[\ddot{l}_{22} - \ddot{l}_{23}\ddot{l}_{33}^{-1}\ddot{l}_{32}\right]^{-1}.$$

The goodness of fit of the assumed model is measured with the so-called deviance

$$G = -n \log|S| - n \log|\hat{\Phi}|, \tag{3}$$

where G is compared with the upper α point of a chi-squared distribution with f degrees of freedom, f indicating the number of partial correlations forced to zero, and the significance of each element ϕ_{ij} assuming model M can be tested by comparing the difference of the deviances of model M and model $M \cap (H : \phi_{ij} = 0)$ with the upper α point of a chi-squared distribution with one degree of freedom.

3. COOK'S LOCAL INFLUENCE

In this section we develop a method of influence analysis in covariance selection models based on the idea of Cook(1986). As perturbation schemes we consider the following two types of case-weight perturbations.

Type1 : $\underline{x}_i \sim N(\underline{\mu}, w_i^{-1}\Sigma)$.

Type2 : $\underline{x}_i \sim N(\underline{\mu}, [nw_i/\sum_j w_j]^{-1}\Sigma)$.

Introducing perturbations from $\underline{w}_0 = (1, \cdots, 1)^T$ (unperturbed) to $\underline{w} = (w_1, \cdots, w_n)^T$ (perturbed), in particular, to a certain direction $\underline{d}$ as $\underline{w} = \underline{w}_0 + t\underline{d}$, $||\underline{d}|| = 1$, we search for influential directions in the sense that the likelihood displacement defined by

$$D(\underline{w}) = 2[l(\hat{\phi_1}|\underline{w}_0) - l(\hat{\phi_{1w}}|\underline{w}_0)],$$

changes most as t varies slightly from zero, where $\hat{\underline{\phi}}_{1w}$ indicates the maximum likelihood estimate after the perturbation. The normal curvature along $\underline{d}$ of the influence graph $(\underline{w}, D(\underline{w}))$ is given by

$$\begin{aligned} C_d &= 2\left|\underline{d}^T \left[\frac{\partial \hat{\underline{\phi}}_1}{\partial \underline{w}^T}\right]^T \left[-\ddot{l}_{11}\right] \left[\frac{\partial \hat{\underline{\phi}}_1}{\partial \underline{w}_1^T}\right] \underline{d}\right| \\ &= 2\left|\underline{d}^T \left[\frac{\partial \hat{\underline{\phi}}_1}{\partial \underline{w}^T}\right]^T V_{11}^{-1} \left[\frac{\partial \hat{\underline{\phi}}_1}{\partial \underline{w}^T}\right] \underline{d}\right|, \end{aligned}$$

and the influential directions are obtained as the eigenvectors associated with dominant eigenvalues of the eigenproblem

$$\left\{2\left[\frac{\partial\hat{\underline{\phi}}_1}{\partial\underline{w}^T}\right]^T V_{11}^{-1}\left[\frac{\partial\hat{\underline{\phi}}_1}{\partial\underline{w}^T}\right]-\lambda I\right\}\underline{d}=\underline{0}. \qquad (4)$$

Suppose that the parameter vector $\underline{\phi}_1$is partitioned as $\underline{\phi}_1=(\underline{\phi}_2^T,\underline{\phi}_3^T)^T$ and also that we are interested in only $\underline{\phi}_2$ in $\underline{\phi}_1$. Then the normal curvature of the profile likelihood displacement

$$D(\underline{w})=2[l(\hat{\phi_2},\hat{\phi_3}|\underline{w}_0)-l(\hat{\phi_{2w}},\phi_3(\hat{\phi_{2w}})|\underline{w}_0)]$$

is given by

$$\begin{aligned} C_d(\underline{w}) &= 2\left|\underline{d}^T\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]^T\left[-\ddot{l}_{22\cdot 3}\right]\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]\underline{d}\right| \\ &= 2\left|\underline{d}^T\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]^T V_{22}^{-1}\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]\underline{d}\right|, \end{aligned}$$

where $\phi_3(\phi_{2w})$ indicates the function which maximize l for fixed ϕ_{2w}. Thus the influential directions are obtained as the eigenvectors associated with dominant eigenvalues of the eigenproblem

$$\left\{2\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]^T V_{22}\left[\frac{\partial\hat{\underline{\phi}}_2}{\partial\underline{w}^T}\right]-\lambda I\right\}\underline{d}=\underline{0}. \qquad (5)$$

The partial derivative $\partial\hat{\phi}_1/\partial\underline{w}^T$ can be derived by expanding $\partial l(\hat{\phi}_{1w},s_w|\underline{w})/\partial\phi_1$ in $\underline{w}$ around $\underline{w}_0$ assuming that the log-likelihood function $l(\underline{\phi}_1|\underline{w})$ is twice continuously differentiable in $(\underline{\phi}_1^T,\underline{w}^T)^T$.

$$\frac{\partial\hat{\underline{\phi}}_1}{\partial\underline{w}^T}=-\left[\frac{\partial^2 l}{\partial\underline{\phi}_1\partial\underline{\phi}_1^T}\right]^{-1}_{\hat{\phi}_1}\left[\frac{\partial^2 l}{\partial\underline{\phi}_1\partial\underline{s}^T}\right]\frac{\partial\underline{s}}{\partial\underline{w}^T} \qquad (6)$$

where $\underline{s}=vech(S)$, both of $\partial\hat{\underline{\phi}}_1/\partial\underline{w}^T$ and $\partial\underline{s}/\partial\underline{w}^T$ being evaluated at $\underline{w}_0$. Differentiate both sides of (1) in s_{kl},

$$\frac{\partial^2 l}{\partial\phi_{ij}\partial s_{ij}} = -n,\ i\neq j,\ (i,j)\in J, \qquad (7)$$

$$\frac{\partial^2 l}{\partial\phi_{ij}\partial s_{kl}} = 0,\ (i,j)\neq(k,l), \qquad (8)$$

and, as shown by Tanaka and Zhang(1999), the elements of partial derivative $\partial\underline{s}/\partial\underline{w}^T$ are given as the elements of

$$\frac{\partial S}{\partial w_\alpha}=n^{-1}(\underline{x}_\alpha-\bar{\underline{x}})(\underline{x}_\alpha-\bar{\underline{x}})^T \qquad (9)$$

for the type 1 case-weight perturbation, and

$$\frac{\partial S}{\partial w_\alpha}=n^{-1}\left\{(\underline{x}_\alpha-\bar{\underline{x}})(\underline{x}_\alpha-\bar{\underline{x}})^T-S\right\} \qquad (10)$$

for the type 2 case-weight perturbation. Note that for the type 2 perturbation $\partial S/\partial w_\alpha$ is just n^{-1} times the ordinary empirical influence function of the sample covariance matrix, and it can also be verified that the similar relation holds between $\partial\hat{\underline{\phi}}_1/\partial w_\alpha$ and the empirical influence function of $\hat{\underline{\phi}}_1$. In this sense we call partial derivatives of parameters with respect to w_α by the name of influence functions in a broad sense.

4. INFLUENCE ANALYSIS BASED ON INFLUENCE FUNCTIONS

So far we have obtained the influence function $\partial\hat{\underline{\phi}}_1/\partial w_\alpha$. It can be easily verified that the approximate relation

$$\hat{\underline{\phi}}_{1(A)}\cong\hat{\underline{\phi}}_1-c\sum_{\underline{x}_\alpha\in A}\frac{\partial\hat{\underline{\phi}}_1}{\partial w_\alpha}$$

holds, where A indicates a subset of observations, $\hat{\underline{\phi}}_1(A)$ the estimate based on the sample without the observations belonging to A, and c a constant. Making use of this additivity relation a general procedure has been proposed for influence analysis based on influence functions as below (see, Tanaka, 1994; Tanaka and Zhang, 1999).

Step 1. Compute the influence functions $\partial\hat{\underline{\phi}}_1/\partial w_\alpha$, $\alpha=1,\cdots,n$, using eq.(6) with eqs.(2), (7), (8) and (9) or (10).

Step 2. For single-case diagnostics compute Cook's D defined by

$$D_\alpha = \left[\partial\underline{\hat{\phi}}_1/\partial w_\alpha\right]^T V_{11}^{-1}\left[\partial\underline{\hat{\phi}}_1/\partial w_\alpha\right]$$

or

$$D_\alpha = \left[\partial\underline{\hat{\phi}}_2/\partial w_\alpha\right]^T V_{22}^{-1}\left[\partial\underline{\hat{\phi}}_2/\partial w_\alpha\right]$$

for each observation after computing the second partial derivatives $\partial^2 l/\partial\phi_{ij}\partial\phi_{kl}$ using eq.(2). Regard the observations with large values of D as singly influential observations.

Step 3. For multiple-case diagnostics apply PCA with metric V_{11}^{-1} or V_{22}^{-1} to the data set of $\{\partial\underline{\hat{\phi}}_1/\partial w_\alpha\}$ or $\{\partial\underline{\hat{\phi}}_2/\partial w_\alpha\}$, and draw scatter plot of the PC scores. Then search for observations which are located far from and on similar directions from the origin, and regard them as candidates for influential subsets of observations. The reason why we introduce metric V^{-1} is to take into account the covariances among variables. The PCA with metric V_{11}^{-1} or V_{22}^{-1} of the influence functions $\{\partial\underline{\hat{\phi}}_1/\partial w_\alpha\}$ or $\{\partial\underline{\hat{\phi}}_2/\partial w_\alpha\}$ are formulated by

$$\left\{\frac{1}{n}\left[\frac{\partial\underline{\hat{\phi}}_1}{\partial\underline{w}^T}\right]\left[\frac{\partial\underline{\hat{\phi}}_1^T}{\partial\underline{w}}\right] - \lambda' V_{11}\right\}\underline{a} = \underline{0}, \quad (11)$$

or a similar eigenproblem for $\underline{\hat{\phi}}_2$.

As discussed by Tanaka and Zhang(1999) there is a close relationship between two eigenproblems (4) and (11). It can be confirmed that these two eigenproblems are equivalent and there exist relations

$$\lambda = 2n\lambda', \quad \sqrt{n\lambda'}\underline{d} = \frac{\partial\underline{\hat{\phi}}_1^T}{\partial\underline{w}}\underline{a},$$

between the eigenvalues and eigenvectors of (4) and (11). Therefore we can obtain the influential directions in the sense of Cook's local influence in the above general procedure described in terms of influence functions.

In step 2 we can also consider the influence on the goodness-of-fit of the assumed model with the influence function for the deviance given by

$$\frac{\partial G}{\partial w_\alpha} = -ntr(S^{-1}\frac{\partial S}{\partial w_\alpha}) - ntr(\hat{\Phi}\frac{\partial\hat{\Phi}}{\partial w_\alpha}). \quad (12)$$

5. NUMERICAL EXAMPLE

We use "fertility of the Swiss soil and social economics index" data to illustrate the method of covariance selection and its influence analysis proposed in this paper. This data set consists of six variables and 47 observations, where the six variables are **a**: fertility of the soil , **b**: agricultural work person ratio, **c**: ratio of the person who takes the record of the top with test of the troop, **d**: ratio of the person who has the educational background above elementary school, **e**: ratio of the Catholic believer, and **f**: infant death rate within one year. These variables are measured at 47 autonomous states in Switzerland in 1888.

5.1 MODEL FITTING

Covariance selection models are fitted successively in the following manner. First, we calculate partial correlations and search for the pair of variables which gives the smallest absolute value. In **Table 1**, the smallest absolute partial correlation is 0.0007 for the pair of variables (5,6). Then, the model with a restriction $\phi_{56} = 0$ (model M_1) is fitted. The deviance difference is 0.00002 between the assumed model M_1 and the full model (the model without restriction, denoted by model M_0).

Table 1 Matrix of partial correlations (no restriction)

1.0000					
-0.3571	1.0000				
-0.1568	-0.2819	1.0000			
-0.5965	-0.4921	0.3724	1.0000		
0.4188	0.3245	-0.4488	0.5887	1.0000	
0.4032	-0.0651	0.0599	0.1116	0.0007	1.0000

Table 2 Matrix of partial correlations (with restriction $\phi_{56} = 0$)

1.0000					
-0.3571	1.0000				
-0.1568	-0.2819	1.0000			
-0.5965	-0.4921	0.3724	1.0000		
0.4188	0.3245	-0.4488	0.5887	1.0000	
0.4032	-0.0651	0.0599	0.1116	0.0000	1.0000

The estimated partial correlations under model M_1 is given in **Table 2**. Next we search for the pair of variables which gives the smallest partial correlation among the partial correlations not restricted to zero, and we can find that the pair of variables (3,6) has the smallest absolute value. Thus, the model with restrictions $\phi_{56} = \phi_{36} = 0$ (model M_2)

is fitted. The deviance difference between models M_1 and M_2 is 0.20978 ($p = 0.6469$).

Table 3 Stepwise procedure

Pairs of variables set to zero	Deviance	Deviance difference	p-value
$1step : (5,6)$	0.0000	0.0000	0.996
$2step : 1step + (3,6)$	0.2098	0.2097	0.646
$3step : 2step + (2,6)$	0.8789	0.6691	0.413
$4step : 3step + (1,3)$	1.8795	1.0005	0.317
$5step : 4step + (4,6)$	5.1846	3.3050	0.069
$6step : 5step + (2,3)$	8.2596	3.0749	0.079

In the similar manner we apply stepwise (backward elimination) procedure. The results are summarized in **Table 3**. As the partial correlation between variables 4 and 6 is almost significant in step 5, we stop the successive process, and select the model with zero partial correlations at variable pairs (5,6),(3,6),(2,6) and (1,3) as the final model. The hypothesis $H_0 : \phi_{56} = \phi_{36} = \phi_{26} = \phi_{13} = 0$ can be tested with log likelihood ratio(LR) or the deviance difference between the final model and the full model. That is, LR=1.8795 (DF=4), hence, the hypothesis is accepted (p=0.785) and therefore those partial correlations can be regarded as zero. The estimated partial correlations based on this final model are given in **Table 4**.

Table 4 Estimated partial correlations (Final model)

	a	b	c	d	e	f
a	1.0000					
b	-0.3569	1.0000				
c	0.0000	-0.2319	1.0000			
d	-0.6491	-0.4863	0.4509	1.0000		
e	0.4701	0.3289	-0.5042	0.6133	1.0000	
f	0.3928	0.0000	0.0000	0.1660	0.0000	1.0000

The obtained final model can be expressed in a linear graph in **Fig.1**. From this figure we can see that Catholic and Infant are conditionally independent given Fertility and Education and that Infant and (Agriculture, Examination) are conditionally independent given Fertility, Education and Catholic.

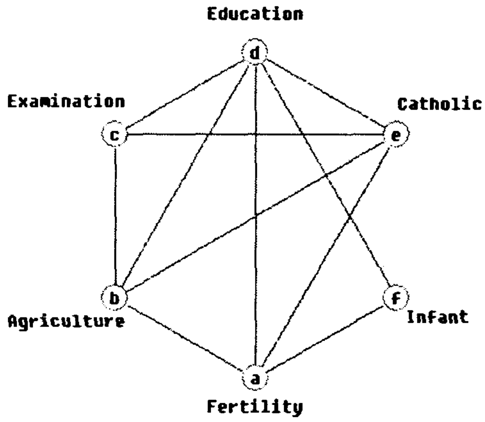

Fig.1 Linear graph of the final model (abde//adf//bcde)

5.2 INFLUENCE ANALYSIS

5.2.1 INFLUENCE ON THE INVERSE COVARIANCE MATRIX

Now let us investigate the influence of observations on the estimated elements of the inverse covariance matrix for the assumed model. The objective is to study the stability or sensitivity of results of analysis under the assumed model.

For single-case diagnostics the influence functions are computed and Cook's D is evaluated. **Fig.2** shows the index plot of Cook's D. It is noticed that observation #45 is much more influential than the others. So, we regard observation #45 as a singly influential observation.

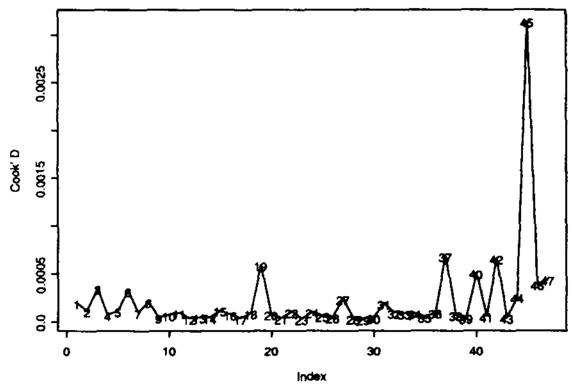

Fig.2 Index plot of Cook's D for $\hat{\underline{\phi}}_1$

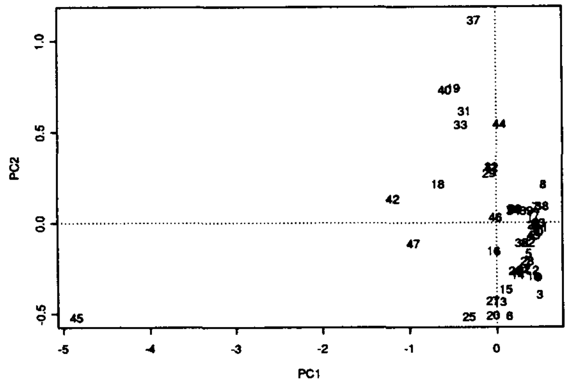

Fig.3 Scatter plot PC2 vs PC1 ($\hat{\underline{\phi}}_1$)

For multiple-case diagnostics we solved eigenproblem (11). The eigenvalues obtained are 31.5178 > 6.0771 > 4.2079 > 3.6671 > $\cdots$ in order of their magnitudes. **Fig.3** gives the scatter plot of PC2 versus PC1. In this figure we can see that the influential observation #45 found in single-case diagnostics does not form influential subset with other observations, because it is located far from the origin but no other observation is located on the similar direction with it.

Table 5 Estimated partial correlations without observation #45 (Final model)

1.0000					
-0.3545	1.0000				
0.0000	-0.2283	1.0000			
-0.5930	-0.4759	0.4189	1.0000		
0.4696	0.3226	-0.5021	0.5332	1.0000	
0.3944	0.0000	0.0000	0.1740	0.0000	1.0000

Table 6 Difference of partial correlations estimated with and without observation #45

0.0000					
-0.0025	0.0000				
0.0000	-0.0036	0.0000			
-0.0561	-0.0104	0.0320	0.0000		
0.0005	0.0063	-0.0021	0.0801	0.0000	
-0.0016	0.0000	0.0000	-0.0080	0.0000	0.0000

Table 5 shows the partial correlations estimated assuming the final model based on the sample without the #45 observation. **Table 6** gives the differences of the partial correlations estimated with and without the #45 observation. The differences are not so extreme, but rather large for partial correlations of (4,5),(1,4) and (3,4). Partial correlations become large in variable pairs (4,5) and (3,4), while it become smaller in (1,4), by omitting observation #45.

5.2.2 INFLUENCE ON COVARIANCE SELECTION

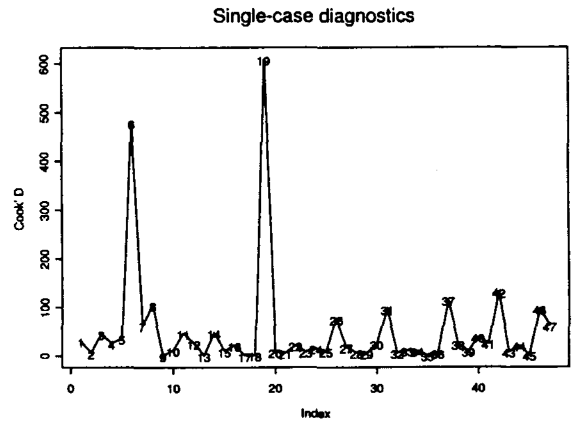

Fig.4 Index plot of Cook's D for $\hat{\underline{\phi}}_2$

Here we consider the stability or sensitivity of the process of selecting variable pairs. We study this problem in two different ways.

First we assume the full model and study influence on the estimated elements of the inverse covariance matrix which are forced to zero in our final model.

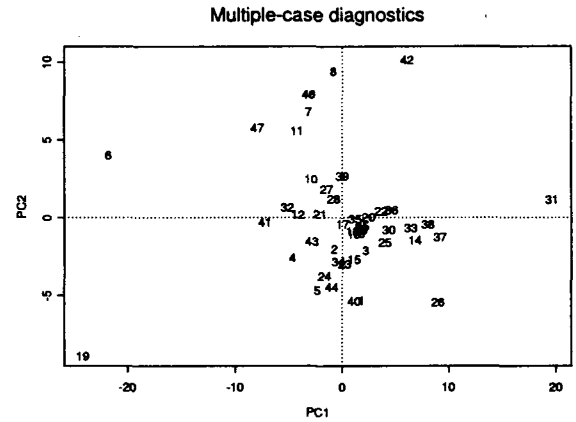

Fig.5 Scatter plot PC2 vs PC1 ($\hat{\underline{\phi}}_2$)

Fig.4 shows the index plot of Cook's D and **Fig.5** gives the scatter plot of PC2 versus PC1 obtained by PCA with matrix V^{-1}, where the eigenvalues are 2157.0547 > 674.7341 > 347.7068 > 174.3873 in order of their magnitudes. These results suggest that observations #6 and #19 are potentially influential to the results of covariance selection.

Another method to study the influence on the process of covariance selection is to evaluate the influence on the element to be forced to zero in each step of the successive procedure. Though we do not explain it in detail, we can find #6 and #19 as influential to the covariance selection procedure.

Table 7 Stepwise procedure (without #6 and #19)

Pairs of variables set to 0	Deviance	Deviance difference	p-value
$1step : (2,6)$	0.0015	0.0015	0.9689
$2step : 1step + (5,6)$	0.0615	0.0599	0.8066
$3step : 2step + (4,6)$	1.2124	1.1509	0.2834
$4step : 3step + (1,3)$	4.0858	2.8734	0.0901
$5step : 4step + (2,3)$	6.4615	2.3757	0.1232

Table 7 shows the result of the covariance selection procedure based on the sample without observations #6 and #19. Here we stop at step 4, because the p-value is less than 0.1. The final model in this case is the model with $\phi_{26} = \phi_{56} = \phi_{46} = 0$. The first two ϕ_{26} and ϕ_{56} are common. But, ϕ_{36} and ϕ_{13} are regarded zero in the former analysis, which ϕ_{46} is regarded zero in the analysis without the observation (#6 and #19). It is a suprise that only two observations among 47 cause such large effects.

6. CONCLUDING REMARKS

In the present paper we have first studied numerically how the stepwise procedure of covariance selection works. In a numerical example the result can be expressed in a simple linear graph, which is convenient for explanation. Then we have tried to develop a procedure of influence analysis, which treats the influence on the estimated parameters in the obtained model and the influence on the process of covariance selection. In our numerical example we could detect observations which are very influential to the estimated parameters or to the covariance selection process. It is interesting to know that the detected observations are different for the above two aspects of influence.

REFERENCES

Dempster, A. P. (1972). Covariance selection, Biometrics, **28**, 157-175.

Cook, R. D. (1986). Assessment of local influence. J. R. Statist. Soc., **B48**, 133-169

Wermuth, N. and Scheidt, E. (1977). Fitting a covariance selection model to a matrix, Algorithm AS **105**., Appl.Statist., **26**, 88-92

Tanaka, Y. (1994). Recent advance in sensitivity analysis in multivariate statistical method, Jour.Japan Soc.Comp.Statist., **7**, 1-25.

Tanaka, Y. and Zhang, F. H.(1999). R-mode and Q-mode influence analysis in statistical modelling: Relationship between influence function approach and local influence approach. Computational Statistics and Data Analysis, **32**, 197-218.

Edwards, D. (1995). *Introduction to Graphical Modelling, Springer-Verlag.*

Whittaker, J. L. (1990). *Graphical Models in Applied Multivariate Statistics, Wiley.*