Sensitivity analysis in covariance selection models

JiMin SUNG* , Yutaka TANAKA**

1. Introduction

Graphical gaussian model or covariance selection model is originally proposed by Dempster (1972) 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 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. 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 try to develop a method for assessing local influence in covariance selection models.

2. Analysis of covariance selection models

Suppose we have observations \( \{x_i, i = 1, \ldots, n\} \), each of which follows independently a \( p \)-variate normal distribution \( N(\mu, \Sigma) \). 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.,

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

where \( I \) indicates a subset of index pairs of \( \Omega = \{(i, j), i, j = 1, \ldots, p, i < j\} \). The remaining elements \( \phi_{ij} \) 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

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

where \( \Phi \) contains unknown parameters \( \phi_{ij}, (i, j) \in J \) and zeros in the remaining elements. The maximum likelihood estimate \( \hat{\phi}_{ij} \) satisfies

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

where

\[
\frac{\partial l}{\partial \phi_{ij}} = \frac{n}{2} \text{tr}(\Phi^{-1} E_{ij}) - \frac{n}{2} \text{tr}(E_{ij} S)
\]

\[
= \begin{cases} n(\phi_{ij} - s_{ij}), & i \neq j, \\ \frac{n}{2}(\phi_{ii} - s_{ii}), & i = j, \end{cases}
\]

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 \( \Phi^{-1} \). An iterative algorithm is given by Wermuth and Scheidt (1977) for computing \( \hat{\phi}_{ij} \) in a typical covariance selection model, where several elements of \( \Phi \) 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

---

*Graduate School of Natural Science and Technology, Doctor Course, Okayama University
**Department of Environmental and Mathematical Sciences, Okayama University
likelihood estimation a consistent estimate $V_{11}$ for the asymptotic covariance matrix $\text{acov}(\hat{\phi}_{1})$ can be obtained by inverting the Hessian matrix of $-l$ with respect to $\phi_{1}$, i.e.,

$$V_{11} = \left[-\nabla^{2}l_{11}\right]_{\phi_{1},\phi_{2}=0}^{-1} = \left[-\frac{\partial^{2}l}{\partial \phi_{1} \partial \phi_{1}^{T}}\right]_{\phi_{1},\phi_{2}=0}^{-1}$$

where the elements of the Hessian matrix are given by

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

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.

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

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

where $G$ is compared with the upper $\alpha$ 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 $\phi_{ij}$ 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 $\alpha$ 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**: $z_{i} \sim N(\mu, w_{i}^{-1}\Sigma)$.

**Type2**: $z_{i} \sim N(\mu, [nw_{i}/ \sum_{j} w_{j}]^{-1}\Sigma)$.

Introducing perturbations from $w_{0} = (1, \ldots, 1)^{T}$ (unperturbed) to $w = (w_{1}, \ldots, w_{n})^{T}$ (perturbed), in particular, to a certain direction $d$ as $w = w_{0} + td$, $||d|| = 1$, we search for influential directions in the sense that the likelihood displacement defined by

$$D(w) = 2[l((\hat{\phi}_{1}|w_{0}) - l(\hat{\phi}_{1}|w_{0})],$$

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

$$C_{d} = 2 \left\{ \partial_{\phi_{i}} \phi_{i}^{T} \right\} V_{11}^{-1} \left\{ \partial_{\phi_{i}} \phi_{i}^{T} \right\} d,$$

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

$$\left\{ 2 \left\{ \partial_{\phi_{i}} \phi_{i}^{T} \right\} V_{11}^{-1} \left\{ \partial_{\phi_{i}} \phi_{i}^{T} \right\} - \lambda I \right\} d = 0. \quad (4)$$

The partial derivative $\partial \phi_{i} / \partial w^{T}$ can be derived by expanding $\partial l(\hat{\phi}_{1}, w|w)/\partial \phi_{i}$ in $w$ around $w_{0}$ assuming that the log-likelihood function $l(\hat{\phi}_{1}|w)$ is twice continuously differentiable in $(\hat{\phi}_{1}^{T}, w^{T})^{T}$.

$$\frac{\partial \phi_{i}}{\partial w^{T}} = -\left[ \frac{\partial^{2}l}{\partial \phi_{i} \partial \phi_{i}^{T}} \right]_{\phi_{1},\phi_{2}=0}^{-1} \left[ \frac{\partial^{2}l}{\partial \phi_{i} \partial \phi_{i}^{T}} \right] \frac{\partial \phi_{i}}{\partial w^{T}}$$

(5)
where $s = vech(S)$, both of $\frac{\partial \hat{\phi}}{\partial w}$ and $\frac{\partial s}{\partial w}$ being evaluated at $w_0$. Differentiate both sides of (1) in $s_{kl}$,

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

(6)

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

(7)

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

$$
\frac{\partial s}{\partial w} = n^{-1} \{(z_0 - \bar{z})(z_0 - \bar{z})^T - S\}
$$

(8)

for the type 2 case-weight perturbation. Note that for the type 2 perturbation $\frac{\partial s}{\partial w_0}$ 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 $\frac{\partial s}{\partial w_0}$ and the empirical influence function of $\hat{\phi}$. In this sense we call partial derivatives of parameters with respect to $w_0$ by the name of influence functions in a broad sense.

4. Influence analysis based in influence functions

So far we have obtained the influence function $\frac{\partial \hat{\phi}}{\partial w}$. It can be easily verified that the approximate relation

$$
\hat{\phi}_{i(A)} \approx \hat{\phi}_{i} - c \sum \frac{\partial \hat{\phi}}{\partial w_0}
$$

holds, where $A$ indicates a subset of observations, $\hat{\phi}_{i(A)}$ 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.** We compute the influence functions $\frac{\partial \hat{\phi}}{\partial w_0}$, $\alpha = 1, \cdots, n$, using eq.(5) with eqs.(2), (6), (7) and (8).

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

$$
D_\alpha = \left[\frac{\partial \hat{\phi}}{\partial w_0}\right]^T V_1^{-1} \left[\frac{\partial \hat{\phi}}{\partial w_0}\right]
$$

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

**Step 3.** For multiple-case diagnostics apply PCA with matrix $V_1^{-1}$ to the data set of $\{\frac{\partial \hat{\phi}}{\partial w_0}\}$, 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 matrix $V^{-1}$ is to take into account the covariances among variables. The PCA with matrix $V_1^{-1}$ of the influence functions $\{\frac{\partial \hat{\phi}}{\partial w_0}\}$ is formulated by

$$
\left\{ \frac{1}{n} \begin{bmatrix} \frac{\partial \hat{\phi}}{\partial w} \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{\phi}}{\partial w} \end{bmatrix}^T \right\} \alpha = 0.
$$

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

$$
\frac{\partial G}{\partial w} = -\text{tr}(S^{-1} \frac{\partial s}{\partial w}) = \text{tr}(\hat{\phi} \frac{\partial \hat{\phi}}{\partial w}).
$$
5. Numerical example

We use "fertility of the Swiss soil and social economics index"(1888) data which consists of six variables and 47 observations. We calculate partial correlations and search for the pair of variables which gives the smallest absolute value. We select the model with zero partial correlations at variable pairs (5,6),(3,6),(2,6) and (1,3) through stepwise procedure. The estimated partial correlations based on this final model are given in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>-0.3569</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>0.0000</td>
<td>-0.2319</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>-0.6491</td>
<td>-0.4863</td>
<td>0.4509</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>0.4701</td>
<td>0.3289</td>
<td>-0.5042</td>
<td>0.6133</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>0.3928</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1660</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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.1 shows the index plot of Cook's D and the scatter plot of PCA.

![Fig.1 Index plot of Cook's D (left) and scatter plot PC2 vs PC1 for $\hat{\beta}_i$ (right)](image)

In this figure observation #45 is found influential in single-case diagnostics. But it does not form an 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.

References