Path analysis, often applied to observational data to study causal structures, describes causal relationship between observed variables. The path analysis is of confirmatory nature and can make statistical tests for assumed causal relations based on comparison of the implied covariance matrix with a sample covariance one. Estimated path coefficients are useful in evaluating magnitude of causality. The traditional path analysis has potential difficulties: there exist equivalent models and the postulated model is assumed to represent “true” causal relations. One cannot determine the causal direction between two variables if only two variables are observed because they are equivalent to each other. The path coefficient estimates are biased if unobserved confounding variables exist, and one cannot detect their existence statistically in many cases. In this paper we develop a new model for causal inference using nonnormality of observed variables, and provide a partial solution to the above-mentioned problems of causal analysis via traditional path analysis using nonnormal data.

Keywords: Causal inference; higher-order moments and cumulants; independent component analysis; nonnormality; path analysis.

1 Introduction

Path analysis, originated by the biologist S. Wright in 1920’s, is a traditional tool for analysis of causal relations of observed variables in an empirical way. The path analysis is an extension of regression analysis in that many dependent and independent variables can be analyzed simultaneously. After Wright’s invention, the path analysis has been developed to create new concepts such as direct and indirect effects and reciprocal effects. In 1970’s, the path analysis was incorporated with factor analysis and allows latent variables in the model. The new framework is now called structural equation modeling or covariance structure analysis (e.g., Bollen, 1989) and is a powerful tool of causal analysis.

The structural equation modeling (SEM) is of confirmatory nature and researchers have to assume true causal relation (or link) before collecting or analyzing data (e.g., Goldberger, 1972). Of course, they can modify their model to better fit to the data collected, but such modifications should be minimized. An alternative to SEM is graphical modeling (e.g., Pearl, 2000; Lauritzen, 1996), which makes exploratory analysis for causality.

Consider how to determine causal direction empirically using SEM or path analysis. The simple regression models, Models 1 and 2 called here, are expressed as:

\[
\text{Model 1: } x_1 = a_{12}x_2 + e_1 \\
\text{Model 2: } x_2 = a_{21}x_1 + e_2,
\]

\[\text{(1.1)}\]

\[\text{(1.2)}\]
where the explanatory variable is assumed to be uncorrelated with the error \( e_1 \) or \( e_2 \). We cannot mention anything about which model is better from the two conventional regression analyses based on the two models above. Using the SEM terminology, the both models are saturated on the covariance matrix of \( [x_1, x_2] \).

There are many approaches to determining causal direction among which the following two are useful. One is to introduce instrumental variables \( w_1 \) and \( w_2 \) such that \( w_1 \) has no direct impact on \( x_2 \) and \( w_2 \) has no direct impact on \( x_1 \):

\[
\begin{align*}
\text{Model 1': } & \quad \begin{cases} x_1 = a_{12}x_2 + b_{11}w_1 + e_1 \\ x_2 = b_{12}w_2 + e_2 \end{cases} \quad (1.3) \\
\text{Model 2': } & \quad \begin{cases} x_1 = b_{21}w_1 + e_1 \\ x_2 = a_{21}x_1 + b_{22}w_2 + e_2 \end{cases} \quad (1.4)
\end{align*}
\]

Then, one can compare fits of these models to determine the causal direction. Correlation between instrumental variables is allowed.

The second approach uses longitudinal data in which one observes the variables \( x_1 \) and \( x_2 \) twice at \( t \) and \( t' \) \((t' > t)\), and constitute the following model:

\[
\begin{align*}
\{ x_1(t') &= a_{11}x_1(t) + a_{12}x_2(t) + e_1 \\
 x_2(t') &= a_{21}x_1(t) + a_{22}x_2(t) + e_2 \} \quad (1.5)
\end{align*}
\]

Statistical significance concerning the parameters \( a_{12} \) and \( a_{21} \) will provide important information on determination of the causal direction.

These approaches aim at removing the problem of the equivalence between Models 1 and 2 to obtain information about causal direction.

Some difficulties have been reported in these approaches. It is difficult to find suitable instrumental variables and also difficult to statistically examine whether \( w_1 \) and \( w_2 \) are actually instrumental variables. Testing instrumental variables and determining causal direction are confounding. In the second approach one has to collect data twice or more for the same subjects, which is expensive. The problems concerning the both approaches are as follows: The inference may be totally distorted if confounding variables between \( x_1 \) and \( x_2 \) exist and/or if the relation is nonlinear or involves interaction effects.

In this article, we explore possibility of using nonnormality to distinguish between Models 1 and 2 and to detect and adjust an unseen confounding variable that makes inevitable distortions of causal inference. We do not discuss influence of sample selection on causal inference here (see, e.g., Spirtes, Meek & Richardson, 1995).

## 2 Nonnormality and causal inference

### 2.1 Determination of causal direction

Let \( x_{1j} \) and \( x_{2j} \) \((j = 1, \ldots, n)\) be observations on random variables \( x_1 \) and \( x_2 \) with mean zero. Denote \( \bar{x}_i^2 = \frac{1}{n} \sum_{j=1}^{n} x_{ij}^2 \) \((i = 1, 2)\) and \( \bar{x}_1x_2 = \frac{1}{n} \sum_{j=1}^{n} x_{1j}x_{2j} \). We shall use similar notation in subsequent derivations without explicit definitions.
The second-order moment structure of Model 1 is obviously given as

\[
E \begin{bmatrix}
\bar{x}_1^2 \\
\bar{x}_1^2 \bar{x}_2 \\
\bar{x}_2^2
\end{bmatrix} = \begin{bmatrix}
a_{12}^2 E(x_2^2) + E(e_1^2) \\
a_{12} E(x_2^2) \\
E(x_2^2)
\end{bmatrix}
\] or equivalently \( E[m_2] = \sigma_2(\tau_2) \) \( (2.1) \)

The number of statistics (or moments) to be used and the number of parameters are both three and thus, the Models 1 and 2 are saturated and equivalent to each other. The both models receive a perfect fit to the sample covariance matrix.

Shimizu and Kano (2003b) assume that \([x_1, x_2]\) is nonnormally distributed and utilize higher-order moments to distinguish between Model 1 and Model 2. They further assume that explanatory and error variables are independently distributed.

First consider use of third-order moments. The expectations of the third-order moments can be expressed in a similar manner as

\[
E \begin{bmatrix}
\bar{x}_1^3 \\
\bar{x}_1^3 \bar{x}_2 \\
\bar{x}_1^2 \bar{x}_2^2 \\
\bar{x}_2^3
\end{bmatrix} = \begin{bmatrix}
a_{12}^3 E(x_2^3) + E(e_1^3) \\
a_{12}^2 E(x_2^3) \\
a_{12} E(x_2^3) \\
E(x_2^3)
\end{bmatrix}
\] or equivalently \( E[m_3] = \sigma_3(\tau_3) \) \( (2.2) \)

for Model 1. Note that \( \tau_3 \) contains some parameters in \( \tau_2 \).

In Model 1, we have three second-order moments and four third-order moments, whereas there are five parameters, namely, \( E(e_1^2), E(e_2^2), a_{12}, E(e_1^3) \) and \( E(x_2^3) \). Thus, if we define

\[
T = n \left( \begin{bmatrix}
m_2 \\
m_3
\end{bmatrix} - \begin{bmatrix}
\sigma_2(\hat{\tau}_2) \\
\sigma_3(\hat{\tau}_3)
\end{bmatrix} \right)^T \hat{M} \left( \begin{bmatrix}
m_2 \\
m_3
\end{bmatrix} - \begin{bmatrix}
\sigma_2(\hat{\tau}_2) \\
\sigma_3(\hat{\tau}_3)
\end{bmatrix} \right) \tag{2.3}
\]

with appropriate estimators \( \hat{\tau}_i \) and a correctly chosen weight matrix \( \hat{M} \), then \( T \) represents distance between data and the model employed and will be asymptotically distributed according to chi-square distribution with df=2 degrees of freedom.\(^2\) We can thus evaluate a fit of Model 1 using the statistic \( T \). The same argument holds for Model 2, and we can confirm that Models 1 and 2 are not equivalent to each other in general. Note that if one assumes that errors are normally distributed, the degrees of freedom of \( T \) becomes three.

Next consider the fourth-order moments. In a similar manner, we have\(^3\)

\[
E \begin{bmatrix}
\bar{x}_1^4 \\
\bar{x}_1^4 \bar{x}_2 \\
\bar{x}_1^3 \bar{x}_2^2 \\
\bar{x}_1^2 \bar{x}_2^3 \\
\bar{x}_2^4
\end{bmatrix} = \begin{bmatrix}
b_{12}^4 E(x_2^4) + 6b_{12}^3 E(x_2^2) E(e_1^2) + E(e_1^4) \\
b_{12}^3 E(x_2^4) + 3b_{12}^2 E(x_2^2) E(e_1^2) \\
b_{12}^2 E(x_2^4) + E(x_2^2) E(e_1^2) \\
b_{12} E(x_2^4) \\
E(x_2^4)
\end{bmatrix}
\] or equivalently \( E[m_4] = \sigma_4(\tau_4) \) \( (2.4) \)

Use of \([m_2, m_4] \) or \([m_2, m_3, m_4] \) enables us to distinguish between Models 1 and 2, and we can use the test statistic \( T \) as defined similarly to compare between the two models.

\(^2\)See Section 2 for some details.

\(^3\)Please note that in the proceedings, there were typos in (2.4), (2.9), (2.12); the typos have been corrected in this web version.

3
2.2 Detection of confounding variables

It is known that the regression-based causal analysis for observational data may be totally distorted if there are unseen (or unobserved) confounding variables between $x_1$ and $x_2$. Let $z$ be a confounding variable, and let us assume that

$$x_2 = a_{21}x_1 + a_{23}z + e_2$$
$$x_1 = a_{13}z + e_1.\tag{2.5}$$

We then have that

$$\text{Cov}(x_1, x_2) = a_{21}V(x_1) + a_{23}a_{13}V(z).\tag{2.7}$$

There could be nonzero covariance between $x_1$ and $x_2$ even if $a_{21} = 0$, and one could make an interpretation that causality from $x_2$ to $x_1$ or its opposite exists; on the other hand, there could be zero covariance between $x_1$ and $x_2$ even if $a_{21}$ is substantial. If researchers can notice and observe the confounding variable $z$, they could adjust or model the effect of $z$. If not, the problems mentioned above will take place in the conventional regression analysis.

Now we shall show that using nonnormality one can identify an unseen confounding variable for some cases. Recall that an explanatory and error variables are assumed to be independent in Models 1 and 2. We can see that the explanatory and error variables are not independent if Model 1 or 2, a simple regression model, is employed for the case where a confounding variable exists. In fact, if we make a simple regression analysis ignoring the existing confounding variable $z$, the regression model can be rewritten as

$$x_2 = \beta x_1 + e'_2$$
$$e'_2 = (a_{21} - \beta)x_1 + a_{23}z + e_2\tag{2.9}$$

with $\beta = \text{Cov}(x_1, x_2)/V(x_1)$. Here $x_1$ and $e'_2$ are uncorrelated but we cannot expect them to be independent of each other. Thus, if dependency between $x_1$ and $e'_2$ is introduced with the confounding variable $z$, one can detect it by observing that the both Models 1 and 2 receive a poor fit.

Assuming independence among $z$, $e_1$ and $e_2$, we can derive the moment structure as

$$E \begin{bmatrix}
  x_1^2 \\
  x_1 x_2 \\
  x_2^2
\end{bmatrix} = \begin{bmatrix}
  a_{13}^2(V(z) + E(e'_2)) \\
  a_{13}(a_{21}a_{13} + a_{23})V(z) + a_{21}E(e'_2) \\
  a_{23}^2 + E(e'_2)
\end{bmatrix}, \tag{2.10}$$

$$E \begin{bmatrix}
  x_1^2 \\
  x_1 x_2 \\
  x_2^2
\end{bmatrix} = \begin{bmatrix}
  a_{13}^2E(z^3) + E(e'_2) \\
  a_{13}^2(a_{21}a_{13} + a_{23})E(z^3) + a_{21}E(e'_2) \\
  a_{13}(a_{21}a_{13} + a_{23})^2E(z^3) + a_{21}^2E(e'_2)
\end{bmatrix}, \tag{2.11}$$

$$E \begin{bmatrix}
  x_1^2 \\
  x_1 x_2 \\
  x_2^2
\end{bmatrix} = \begin{bmatrix}
  a_{13}^4E(z^4) + 6a_{13}^2V(z)(e'_2^2) + E(e'_2) \\
  a_{13}^3c_{23}E(z^4) + 3a_{13}(a_{21}a_{13} + c_{23})V(z)(e'_2^2) + a_{21}E(e'_2) \\
  a_{13}c_{23}E(z^4) + 3a_{21}c_{23}(a_{13}a_{21} + c_{23})V(z)(e'_2^2) + 3a_{13}c_{23}V(z)(e'_2^2) + E(e'_2)E(e'_2) + a_{21}E(e'_2)
\end{bmatrix}, \tag{2.12}$$

where $c_{23}$ denotes $a_{21}a_{13} + a_{23}$. Assume that $V(z) = 1$ without any loss of generality. We can estimate and evaluate the model that uses second-, third- and fourth-order
moments because there are 11 parameters, namely $a_{21}, a_{13}, a_{23}, E(z^3), E(z^4), E(e_1^2), E(e_2^2), E(e_3^1), E(e_4^1)$ and $E(e_4^2)$ for 12 statistics. We cannot estimate the model that uses second- and third-order moments, because there are 8 parameters for 7 statistics. We can estimate the model that uses second- and fourth-order moments but cannot evaluate it because there are 8 parameters for 8 statistics.

If $e_1$ and $e_2$ are normally distributed, one can estimate and evaluate the model that uses only the second- and third-order moments or uses only the second- and fourth-order moments.

As a result, one can suspect that there is a confounding variable between $x_1$ and $x_2$ if the both Models 1 and 2 receive a poor fit and the model with an unseen confounding variable is fitted statistically. Furthermore, we can determine the causal direction whether $x_1 \rightarrow x_2$ or $x_2 \rightarrow x_1$ is appropriate even in the case where a confounding variable exits.

It should be noted that the independence assumption between explanatory and error variables is crucial in our setting.

We cannot identify more than one unseen confounding variables between $x_1$ and $x_2$ in the current setting because the parameters to be estimated enormously increases in number as the confounding variables increase. This problem will be noticed when the model with an unseen confounding variable $z_1$ is not fitted well and one needs to introduce another confounding variable $z_2$. There may be cases where one can identify what the confounding variable $z_1$ is and it is observable. In the case, one can study effects of another unseen confounding variable, constituting a new model with observed $x_1$, $x_2$ and $z_1$ and an unseen $z_2$ as follows:

\begin{align*}
  x_2 &= a_{21}x_1 + a_{23}z_1 + a_{24}z_2 + e_2 \\
  x_1 &= a_{13}z_1 + a_{14}z_2 + e_1,
\end{align*}

where $z_1$ and $z_2$ may be correlated. One can identify as many confounding variables as one needs if the iterative process is repeated.

### 2.3 Estimation and examination of fit

Bentler (1983) and Mooijaart (1985) have considered the generalized least squares approach (GLS) that uses higher-order moments. We also employ the GLS approach to estimate structural parameters.

Let $m_1$, $m_2$ and $m_h$ be the vectorized first-, second- and $h^{th}$-order moments after removing the redundant elements, where $h$ means here 3 or 4 or both. Let $\boldsymbol{\tau}$ be a vector that contains the model parameters involving all moments employed and let $\boldsymbol{\sigma}_1(\boldsymbol{\tau}) = E(m_1)$, $\boldsymbol{\sigma}_2(\boldsymbol{\tau}) = E(m_2)$ and $\boldsymbol{\sigma}_h(\boldsymbol{\tau}) = E(m_h)$. The GLS estimator of $\boldsymbol{\tau}$ is then obtained as

\begin{equation}
\hat{\boldsymbol{\tau}} = \arg \min_{\boldsymbol{\tau}} \left( \begin{bmatrix} m_1 \\ m_2 \\ m_h \end{bmatrix} - \begin{bmatrix} \boldsymbol{\sigma}_1(\boldsymbol{\tau}) \\ \boldsymbol{\sigma}_2(\boldsymbol{\tau}) \\ \boldsymbol{\sigma}_h(\boldsymbol{\tau}) \end{bmatrix} \right)^T \hat{U}^{-1} \left( \begin{bmatrix} m_1 \\ m_2 \\ m_h \end{bmatrix} - \begin{bmatrix} \boldsymbol{\sigma}_1(\boldsymbol{\tau}) \\ \boldsymbol{\sigma}_2(\boldsymbol{\tau}) \\ \boldsymbol{\sigma}_h(\boldsymbol{\tau}) \end{bmatrix} \right). \tag{2.15}
\end{equation}

When variables to be analyzed have been centered in advance, the terms involving $m_1$ in (2.15) are removed.

Let $n$ be a sample size and define $V$ as

\begin{equation}
V = \lim_{n \to \infty} n \cdot V[m_1^T, m_2^T, m_h^T]^T. \tag{2.16}
\end{equation}
A typical choice of $\hat{U}$ will be a consistent estimator of $V$. The resultant GLS estimator $\hat{\tau}$ determined by (2.15) is then consistent and achieves the smallest asymptotic variance-covariance matrix among the class of estimators that are functions of the sample moments employed. The asymptotic variance-covariance matrix of $\hat{\tau}$ is expressible in the form:

$$\frac{1}{n}(J^TV^{-1}J)^{-1},$$

(2.17)

where

$$J = \frac{\partial[\sigma_1(\tau)^T, \sigma_2(\tau)^T, \sigma_h(\tau)^T]^T}{\partial \tau^T}$$

(2.18)

(see, e.g., Ferguson, 1958). In the case, however, one requires to compute the $2^h$th-order moments to implement this estimation, and it is known that one needs to have extremely large samples to stably estimate higher-order moments and that inverting such a huge matrix $\hat{V}$ may cause difficulties in solving the equations (see e.g., Hu, Bentler & Kano, 1992). Thus, rather than using $\hat{V}$ it could be more effective to choose, as $\hat{U}$, a simpler weight matrix such as $\text{diag}(\hat{V})$ and the identity matrix particularly for moderate sample sizes. The asymptotic variance-covariance matrix of the estimator can then be estimated in the form:

$$\frac{1}{n}(J^TU^{-1}J)^{-1} J^TU^{-1}VV^{-1}J (J^TU^{-1}J)^{-1},$$

(2.19)

where $U$ is the convergence limit of $\hat{U}$.

In this paper we adopt the diagonal matrix $\text{diag}(\hat{V})$ as $\hat{U}$ to compute the estimates and the estimates are used to evaluate a model fit.

One can use the following test statistic

$$T = n \left( \begin{bmatrix} m_1 \\ m_2 \\ m_h \end{bmatrix} - \begin{bmatrix} \sigma_1(\hat{\tau}) \\ \sigma_2(\hat{\tau}) \\ \sigma_h(\hat{\tau}) \end{bmatrix} \right)^T \hat{M} \left( \begin{bmatrix} m_1 \\ m_2 \\ m_h \end{bmatrix} - \begin{bmatrix} \sigma_1(\hat{\tau}) \\ \sigma_2(\hat{\tau}) \\ \sigma_h(\hat{\tau}) \end{bmatrix} \right)$$

(2.20)

to examine the model assumption, with

$$M = V^{-1} - V^{-1}J(J^TV^{-1}J)^{-1}J^TV^{-1}.$$  

(2.21)

The statistic $T$ approximates to a chi-square variate with degrees $\text{tr}[VM]$ of freedom where $n$ is large enough (see e.g. Shimizu & Kano, 2003a). The required assumption for this is that $\hat{\tau}$ is a $\sqrt{n}$-consistent estimator. No asymptotic normality is needed. See Browne (1984) for details.

### 3 Real data example

Questionnaire data about popularity evaluation of male and female comedians, actors, actresses, singers, and pop stars in Japan were analyzed as an example to illustrate the effectiveness of our method described here. The survey was conducted at several universities in 1998 (cf., Kano, 1998). The sample size is 614. We took the diagonal matrix $\text{diag}(\hat{V})$ as $\hat{U}$ in (2.15) and estimated model parameters and
evaluated a model fit. The means of observed variables are subtracted to be centered and all estimates shown here are standardized where all variables have unit variance.

First we examined Model 1 and Model 2 that use second-, third- and fourth-order moments to determine the direction of a path between $x_1$ and $x_2$. The labels of the observed variables $x_1, x_2$ are shown in Table 1. These variables were measured by a seven-point Likert scale. We show model fit chi-squares $T$ in (2.20), associated degrees of freedom and estimates of path coefficients in Figure 1. The both models receive a poor fit and were statistically rejected. As a result, $x_2$ may not be a cause of $x_1$; $x_1$ may not be a cause of $x_2$; or confounding variables exist between $x_1$ and $x_2$.

Next, we introduced an unseen confounding variable $z$ to explain the relationship between $x_1$ and $x_2$. Figure 2 shows the three models considered here and estimation

---

Table 1: Variable labels

| $x_1$: How much do you like (Comedian) A |
| $x_2$: How much do you like (Comedian) B |

---

4 Comedians A and B are Tokoro Joji and Hisamoto Masami.
results. It is seen that introduction of the confounding variable greatly reduces model fit chi-squares and that Model 3 is now statistically acceptable. We thus conclude that there are a confounding variable between \( x_1 \) and \( x_2 \) and a negative causal impact from \( x_1 \) to \( x_2 \). Chi-squares for fit of the models considered are summarized in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>( T ) (df)</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>13.55 (5)</td>
<td>0.02</td>
</tr>
<tr>
<td>Model 2</td>
<td>14.20 (5)</td>
<td>0.01</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.21 (1)</td>
<td>0.63</td>
</tr>
<tr>
<td>Model 4</td>
<td>5.72 (1)</td>
<td>0.02</td>
</tr>
<tr>
<td>Model 5</td>
<td>10.56 (2)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

We could interpret the confounding variable \( z \) as tendency of liking comedians. It is interesting to note that \( x_1 \) negatively influences upon \( x_2 \) after adjusting the confounding variable whereas the simple correlation between \( x_1 \) and \( x_2 \) are positive. We could not tell that the relation is causal based only on the analysis, but we could say that \( x_1 \) and \( x_2 \) are negatively correlated for the population with a given value of \( z \).

The Model 5 in Figure 2 is a one-factor model with two variables. It is known that the model is not identifiable and not estimable in a traditional factor analysis model, where only second-order moments are used (see e.g., Kano, 1997).

4 Relation with independent component analysis

The important assumptions of our approach described in the previous sections are nonnormality and independence (between explanatory and error variables). These are basic assumptions of independent component analysis (ICA) recently discussed extensively.

ICA is one of multivariate analysis techniques which aims at separating or recovering linearly-mixed unseen multidimensional independent signals from the mixed observable variables. See e.g., Hyvärinen, Karhunen and Oja (2001) for a thorough description of ICA. Let \( x \) and \( s \) be \( p \)- and \( q \)-vectors of unseen (or blind) signals and observed variables that are linear mixtures of \( s \). The ICA model is then written as

\[
x = As, \tag{4.1}
\]

where \( A \) is called a mixing matrix. The main process of ICA is to estimate the mixing matrix.

Comon (1994) studied this model theoretically for the typical case where \( q \leq p \) and provided conditions for the model to be estimable. The conditions include that \( s \) are mutually independent and contains at most one normal component. Mooijaart (1985) discussed a similar problem in the context of nonnormal factor analysis.

\(^5\)Model 3 and Model 4 are not estimable, either.
Comon’s conditions are weaker and more elegant than Mooijaart’s. The estimation is implemented by finding the demixing matrix $W$ such that the components of $\hat{s} = Wx$ have maximal nonnormality. Nonnormality is often measured by cumulants.

We write Model 1 as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a_{21} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ e_2 \end{bmatrix},$$

(4.2)

where $x_1$ and $e_2$ are nonnormal and independent. Thus, Model 1 is nothing but an ICA model.

Noisy ICA is a model of ICA with independent error terms as defined

$$x = As + e,$$

(4.3)

where $e$ is typically assumed to follow according to a normal distribution. Shimizu and Kano (2003c) pointed out that the normality assumption is often violated and that the violation leads to a serious bias in estimation of $A$. They then proposed a new model with independent and nonnormal errors. The model is an overcomplete ICA model and is known to be difficult to estimate. Shimizu and Kano (2003c) used the fact that the mixing matrix has a certain structure to develop a feasible estimation method.

The model with a confounding factor in (2.5) and (2.6) is written as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{13} & 1 & 0 \\ a_{21}a_{13} + a_{23} & a_{21} & 1 \end{bmatrix} \begin{bmatrix} z \\ e_1 \\ e_2 \end{bmatrix}.$$ 

(4.4)

Here $z$, $e_1$ and $e_2$ are mutually independent. The model is an overcomplete ICA model with a structured mixing matrix. The method by Shimizu and Kano (2003c) will also be applied to estimate the model in (4.4).

A bi-directed causal model$^6$ has been discussed by Asher (1976), Bollen (1989) and Spirtes et al. (1995) among others. The model for two observed variables $x_1$ and $x_2$ are written as

$$\begin{cases} x_1 = a_{12}x_2 + e_1 \\ x_2 = a_{21}x_1 + e_2 \end{cases}.$$ 

(4.5)

The model can be expressed as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.$$ 

(4.6)

This model is a model of ICA if the independence assumption between $e_1$ and $e_2$ is appropriate.

The connection between ICA and path analysis in nonnormal SEM may play an important role in this area.

References

$^6$It is a nonrecursive model and is also called a reciprocal model.


