Use of non-normality in structural equation modeling: Application to direction of causation

Shohei Shimizu\textsuperscript{a,b,1}Yutaka Kano\textsuperscript{a}

\textsuperscript{a}Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, Japan
\textsuperscript{b}Helsinki Institute for Information Technology, Basic Research Unit, Department of Computer Science, University of Helsinki, Finland

Abstract

Structural equation modeling (SEM) typically utilizes first- and second-order moment structures. This limits its applicability since many unidentified models and many equivalent models that researchers would like to distinguish are created. In this paper, we relax this restriction and assume non-normal distributions on exogenous variables. We shall provide a solution to the problems of underidentifiability and equivalence of SEM models by making use of non-normality (higher-order moment structures). The non-normal SEM is applied to finding the possible direction of a path in simple regression models. The method of (generalized) least squares is employed to estimate model parameters. A test statistic for examining a fit of a model is proposed. A simulation result and a real data example are reported to study how the non-normal SEM approach works empirically.

Key words: Structural equation modeling, non-normality, equivalent model, unidentified model, higher-order moments, causal inference

1Corresponding author. Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, Toyonaka, 560-8531, Japan. Email: shimizu@sigmath.es.osaka-u.ac.jp. Phone and fax: +81-6-6850-6485.
2 We thank Aapo Hyvärinen, Yusuke Miyamoto, and the anonymous reviewers for their helpful comments. This research was partially supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture and Sports, and the Osaka University Kouenkai.

Preprint submitted to Elsevier Science 7 January 2006
Structural equation modeling (SEM) is an important branch in multivariate analysis that integrates factor analysis and path analysis (e.g., Bollen, 1989). SEM has been successfully applied to observational and experimental studies in social sciences. In typical situations, SEM uses first- and second-order moment structures that are derived from structural equations that prescribe causal connections between variables. Higher-order moment structures have almost never been utilized for model identification. This is partly due to the fact that SEM was originally developed based on the normal assumption. It should be noted, however, that observed variables in social sciences are almost never normally distributed (Micceri, 1989). Higher-order moment structures could be informative for SEM.

Higher-order moments have been used to help define the weight matrix in the method of generalized least squares to achieve asymptotic efficiency of estimators and chi-squaredness of test statistics. The approach is said to be asymptotic distribution-free (e.g., Browne, 1982, 1984), and many programs have incorporated the estimation method (e.g., Arbuckle and Wothke, 1999; Bentler, 1995; Jöreskog and Sörbom, 1993). A different usage of higher-order moments is found when examining nonlinear effects such as interaction and quadratic effects (e.g., Jöreskog and Yang, 1996; Kenny and Judd, 1984), where non-product exogenous variables are assumed to be normal and product indicators (e.g., $x_1 x_2$) are used to prescribe nonlinear terms of latent variables. Mixture modeling for latent variables (e.g., Muthén, 2002) uses higher-order moments implicitly.

There are some drawbacks to conventional SEM, namely unidentified models and equivalent models. SEM models are based on hypothetical causal relations developed by a practical researcher and are not always identified and estimable. For instance, one cannot distinguish between specific and error factors in a usual measurement model; any factor model with two indicators is not identified. Readers can find systematic discussions on identification in Bollen (1989). Even if models under consideration are identified, one may not be able to examine fits of the models to the data collected. Such a problem takes place if the model is saturated, i.e., has zero degree of freedom (see Section 2.4.2 for some details). A typical example is a regression model. Even if models considered are identified and not saturated, one may not be able to make a model comparison due to equivalence between the models to be compared. For a thorough treatment on equivalent models, see Bekker et al. (1994).

Bentler (1983) pointed out the possibilities that non-normality (higher-order moment structures) could resolve these problems if data do not follow normality. All studies in SEM using non-normality or higher-order moments since
Bentler (1983), however, have not focused on solving these problems. The work by Mooijaart (1985) is an exception which showed that use of third-order moment structures enables us to determine a factor rotation in exploratory factor analysis so that factor rotation by varimax for example is not necessary. A similar argument in the error-in-variables model can be found in Aigner et al. (1983); Fuller (1987); Pal (1980). Note that these authors did not mention anything about examination of model fit.

Our work applies to Bentler’s idea to simple regression models. We shall provide identification conditions for the regression models to be comparable in terms of model fit.

The paper is structured as follows. First, in Section 2, we propose a new framework of non-normal SEM, abbreviated nnSEM, and review the method of (generalized) least squares to estimate structural parameters using first- to fourth-order moment structures and a test statistic for goodness-of-fit. In Section 3, we discuss a specific nnSEM approach to find the possible direction of a path in a simple regression model, which cannot be fully examined within the traditional SEM framework. In Sections 4 and 5, we conduct a simulation study and provide a real data example to study how the nnSEM approach to the simple regression model works empirically. We end with some discussion in Section 6.

2 Non-normal structural equation modeling

2.1 Model

Let $x$ be a $p$-dimensional observed vector, $f$ a $q$-dimensional latent vector, and $e$ an error vector of appropriate dimension. Let $\xi$ and $\eta$ be vectors consisting of exogenous and endogenous (construct) variables in $x$ and $f$, respectively. Let $\nu$ be an intercept vector of $\eta$. The well-known structural equation model (Bentler, 1995; Bentler and Weeks, 1980) is written as

$$\eta = \nu + B\eta + \Gamma\xi + e,$$

where $B$ and $\Gamma$ are matrices of path coefficients, the expectation of $e$ is zero, and $\xi$ and $e$ are independent. Here we assume that the components of $\xi$ and $e$ have unknown non-normal distributions and that their moments up to fourth-order exist. One can assume normality for errors if it is relevant. (Existence of the eighth-order moments of observed variables will be assumed when asymptotic distributions of estimators are derived below.)
It follows that, after obtaining the reduced form of (1), \( x \) can be expressed as

\[
x = G \begin{bmatrix} \eta \\ \xi \end{bmatrix} = G \begin{bmatrix} I - B & -\Gamma \\ O & I \end{bmatrix}^{-1} \left( \begin{bmatrix} \nu \\ E[\xi] \end{bmatrix} + \begin{bmatrix} e \\ \xi - E[\xi] \end{bmatrix} \right),
\]

where \( G \) is a known selection matrix that selects observed variables from \( \eta \) and \( \xi \), and \( I - B \) is assumed to be nonsingular.

### 2.2 Identification

It is difficult to develop identification conditions even in traditional SEM models (Bollen, 1989). Therefore, it will be more difficult to study identification conditions for the general \( \text{nnSEM} \) model in (1). Instead, some conditions for specific models have been provided.

Mooijaart (1985) presented an identification condition for the factor analysis model with non-normal independent common factors:

\[
x = Af + e,
\]

where \( f \) denotes a \( q \)-dimensional independent common factor vector and \( e \) denotes a \( p \)-dimensional independent unique factor vector. He showed that the model above is identified and has no rotation indeterminacy if the skewnesses of common (independent) factors are distinct from each other.

Comon (1994) studied the model (2) theoretically for the case where \( e = 0 \) and provided conditions for the model to be identified in the context of independent component analysis (Hyvärinen et al., 2001). The conditions include: i) all components of \( f \), with a possible exception of one component, must be non-normal and not be degenerate; ii) the number \( p \) of observed variables must be at least as large as the number \( q \) of independent factors, that is, \( p \geq q \); iii) the matrix \( A \) must be of full column rank.

What kind of non-normality is required in i) depends on methods of estimation. For instance, nonzero skewnesses will be required if third-order moments are utilized in estimation.

Let us denote the first- to fourth-order moment structures of the \( \text{nnSEM} \) model (1) as follows:
\[ \sigma_1(\tau) = E[x] \]
\[ \sigma_i(\tau) = H_i E\left[ (x - E[x]) \otimes \cdots \otimes (x - E[x]) \right] \quad i = 2, 3, 4, \]

where the symbol \( \otimes \) denotes the Kronecker product and \( H_i \) is a selection matrix of order \( \left( \binom{p+i-1}{i} \right) \times p^i \) that selects non-duplicated elements (equivalently, removing redundancy). The structural parameter vector \( \tau \) consists of free parameters of \( \nu, E[\xi], B, \Gamma, \) and higher-order moments of \( e \) and \( \xi \) (see e.g., Bentler, 1983, for more details). The moment structures depend only on the model parameters. For example, the first-order moment structure is:

\[ \sigma_1(\tau) = G \left[ \begin{array}{cc} I - B & -\Gamma \\ O & I \end{array} \right]^{-1} \left[ \begin{array}{c} \nu \\ E(\xi) \end{array} \right], \]

which is a function of the model parameters in \( \tau \) only.

When using the moment structures \( \sigma_i(\tau) \) in estimation of \( \tau \), we must assume that the identifiability of the model (1) based on the moment structures holds, that is,

\[ \text{if } \sigma_i(\tau_1) = \sigma_i(\tau_2) \quad \text{for} \quad i = 1, 2, 3, 4, \quad \text{then} \quad \tau_1 = \tau_2. \]

We shall discuss the identification problem for a specific model in Section 3.

2.3 Estimation

Bentler (1983) and Mooijaart (1985) have considered the generalized least squares approach (GLS) in estimation that utilizes higher-order moments (see also Hansen, 1982). We also employ the GLS approach to estimate structural parameters.

Let \( x_1, \ldots, x_N \) be a random sample from a \( \text{nmSEM} \) model in (1), and define the sample counterparts to the moments in (3) as

\[ m_1 = \frac{1}{N} \sum_{j=1}^{N} x_j \]
\[ m_i = \frac{1}{N} H_i \sum_{j=1}^{N} (x_j - m_1)^{i \times} \cdots \otimes (x_j - m_1) \quad (i = 2, 3, 4). \]
We then have that $E[m_i] = \sigma_i(\tau_0) + o(1)$ ($i = 1, \ldots, 4$), where $\tau_0$ is the true parameter vector.

Let us denote

$$m = \begin{bmatrix} m_1^T, m_2^T, m_3^T, m_4^T \end{bmatrix}^T$$

(6)

$$\sigma(\tau) = \begin{bmatrix} \sigma_1(\tau)^T, \sigma_2(\tau)^T, \sigma_3(\tau)^T, \sigma_4(\tau)^T \end{bmatrix}^T.$$  

(7)

Then, the GLS estimator of $\tau$ is obtained as

$$\hat{\tau} = \arg\min_{\tau} \{m - \sigma(\tau)\}^T \hat{U}^{-1} \{m - \sigma(\tau)\}.$$  

(8)

Here $\hat{U}$ is a weight matrix in GLS estimation that converges in probability to a certain positive definite matrix $U$.

Assume that the eighth-order moments of $x_i$ are finite, and then we can define

$$V = \lim_{N \to \infty} N \cdot \text{Var}(m).$$  

(9)

Denote $J = \partial\sigma(\tau)/\partial\tau^T$ and assume that $J$ is of full column rank. Under some regularity conditions including the identifiability condition in (5) and the rank condition of $J$, the GLS estimator $\hat{\tau}$ obtained by (8) is consistent and asymptotically normal with

$$\frac{1}{N} (J^T U^{-1} J)^{-1} J^T U^{-1} V U^{-1} J (J^T U^{-1} J)^{-1}$$

as the asymptotic variance-covariance matrix (see e.g., Browne, 1982, 1984; Ferguson, 1958).

Let $\hat{V}$ be a usual (nonparametric) estimation for $V$. One often takes the estimator $\hat{V}$ as $\hat{U}$ in estimation to achieve asymptotic efficiency among the class of all estimators that are functions of $m$ (Browne and Armiger, 1995). In this paper we adopt $\hat{V}$ as $\hat{U}$ to compute the estimates.
2.4 Testing a hypothesis

2.4.1 Some test statistics to evaluate a model fit

Testing model assumptions is an important process of SEM and mnSEM as well. Here, the null and alternative hypotheses $H_0$ and $H_1$ are as follows:

$$H_0 : \mathbb{E}(m) = \sigma(\tau) \quad \text{versus} \quad H_1 : \mathbb{E}(m) \neq \sigma(\tau).$$

(10)

Assume again that the $V$ in (9) is finite and positive definite and $J$ is of full column rank. Define

$$F(\hat{\tau}) = \{m - \sigma(\hat{\tau})\}^T \hat{V}^{-1} \{m - \sigma(\hat{\tau})\}. \quad (11)$$

Then a test statistic $T_1 = N \times F(\hat{\tau})$ could be used to test the null hypothesis $H_0$, that is, to examine a fit of the model considered to data. Under $H_0$, the statistic $T_1$ asymptotically approximates to a chi-square variate with degrees $u - v$ of freedom where $u$ is the number of distinct moments employed and $v$ is the number of parameters estimated (e.g., Browne, 1982, 1984). Acceptance of $H_0$ implies that the model assumptions fit data. Rejection of $H_0$ suggests that at least one model assumption is in error so that $H_1$ holds (e.g., Bollen, 1989).

However, it is often pointed out that this type of test statistics requires large sample sizes for $T_1$ to behave like a chi-square variate (e.g., Hu et al., 1992). Therefore, we would apply Yuan-Bentler’s proposal (Yuan and Bentler, 1997) to $T_1$ to improve its chi-square approximation and employ the following test statistic $T_2$:

$$T_2 = \frac{T_1}{1 + F(\hat{\tau})}. \quad (12)$$

We shall make a simulation experiment to study finite sample behavior of $T_2$ in Section 4.

2.4.2 Saturated and equivalent models

A simple and typical procedure to compare two models is as follows: First we test the models and see if each model is accepted or rejected. If one model is accepted and the other is rejected, we adopt the accepted model. However, such a procedure does not work at all if the models are saturated or equivalent. Saturated models are those that always fit any data perfectly since the number
of parameters is equal to the number of distinct moments employed. One cannot evaluate saturated models using the test statistics above. Equivalent models are those that provide the same moment structures. Therefore, one cannot say anything about which model is better than the other since both models provide the same value of the test statistics. See Bekker et al. (1994) and Bollen (1989) for details.

3 Finding the direction of a path between two observed variables

3.1 Purpose and conventional SEM approach

Consider the two simple regression models, namely, Models 1 and 2:

Model 1: \[ x_1 = b_{12} x_2 + e_1 \]
Model 2: \[ x_2 = b_{21} x_1 + e_2. \]

There is a situation in which researchers want to have empirical evidence as to which model is more appropriate or is fitted better to a data set collected. Although it is difficult to determine which variable is causal, i.e., which causal direction is true, such an analysis would provide an important insight for future research.

The traditional approach uses first- and second-order moment structures only and assumes that the explanatory variable is uncorrelated with the error variable. The two models above are then saturated in terms of first- and second-moment structures, and it is impossible to obtain any information on relative model fit of the two regression models.

There are some approaches that compare Models 1 and 2 within the conventional SEM framework, among which the following two are often applied. 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 \( x_1 \). Then, one can compare fits of these models to find the direction of a path. 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 \( t' \) (\( t < t' \)). Statistical significance tests concerning the path coefficient of \( x_1(t) \) to \( x_2(t') \) and that of \( x_2(t) \) to \( x_1(t') \) will provide information on the comparison. These approaches aim at removing the problem of saturation and equivalence between Models 1 and 2. See Bollen (1989) for these approaches.

Here we suggest an alternative modeling approach that applies nnSEM to this problem. In this modeling, one needs no instrumental variables and no
longitudinal data. The model analyzes only \( x_1 \) and \( x_2 \) at one occasion.

### 3.2 Non-normal SEM approach

Let \( x_{1k} \) and \( x_{2k} \) \((k = 1, \ldots, N)\) be observations on random variables \( x_1 \) and \( x_2 \) with mean zero. Denote \( m_{ij} = \frac{1}{N} \sum_{k=1}^{N} x_{1k}^i x_{2k}^j \). The assumption of zero means is made in order to avoid unnecessary complications. The following derivations will hold by ignoring smaller order terms if the assumption is not made and sample means are used as their estimators. In other words, all the results below hold exactly but the derivations are more complicated.

The simple regression models in nnSEM, Models 1' and 2' are expressed in the form:

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

where the explanatory variable is assumed to be independent of the error in each model. Some or all of the exogenous variables are assumed to be non-normal. We shall give a precise condition on the non-normality in the propositions later.

The second-order moment structure of Model 1' is obviously given as

\[
E \begin{bmatrix} m_{20} \\ m_{11} \\ m_{02} \end{bmatrix} = \begin{bmatrix} b_{12}^2 & 1 & E(x_2^2) \\ b_{12} & 0 & E(e_1^2) \\ 1 & 0 & \end{bmatrix} \quad \text{or equivalently} \quad E[m_2] = \sigma_2(\tau_2).
\]

Here \( \tau_2 = [E(x_2^2), E(e_1^2), b_{12}]^T \). It is known that the number of sample moments and the number of parameters are both three, and thus the Models 1' and 2' are saturated and equivalent to each other as far as second-order moments alone are concerned. Both models receive a perfect fit to the sample covariance matrix.

The expectations of the third- and fourth-order moments can be expressed in a similar manner as
\[
E \begin{bmatrix}
m_{30} \\
m_{21} \\
m_{12} \\
m_{03}
\end{bmatrix} = \begin{bmatrix}
b_{12}^3 & 1 \\
b_{12}^2 & b_{01} \\
b_{12} & 0 \\
1 & 0
\end{bmatrix} \begin{bmatrix}
E(x_2^3) \\
E(e_1^3)
\end{bmatrix}
\]

or equivalently \( E[m_3] = \sigma_3(\tau_3) \)

and

\[
E \begin{bmatrix}
m_{40} \\
m_{31} \\
m_{22} \\
m_{13} \\
m_{04}
\end{bmatrix} = \begin{bmatrix}
b_{12}^4 & 6b_{12}^2 & 1 \\
b_{12}^3 & 3b_{12} & 0 \\
b_{12}^2 & 1 & 0 \\
b_{12} & 0 & 0 \\
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
E(x_2^4) \\
E(x_2^3)E(e_1^2) \\
E(x_2^2)E(e_1^3) \\
E(e_1^4)
\end{bmatrix}
\]

or equivalently \( E[m_4] = \sigma_4(\tau_4) \)

for Model 1’, where \( \tau_3 = [E(x_2^3), E(e_1^3), b_{12}]^T \) and \( \tau_4 = [E(x_2^4), E(e_1^4), E(x_2^3), E(e_1^3), b_{12}]^T \).

Note that there is some duplication among \( \tau_i \).

In Model 1’, we have twelve sample moments, whereas there are seven parameters, namely, \( E(x_2^3), E(e_1^3), E(x_2^2), E(e_1^2), E(x_2^4), E(e_1^4) \) and \( b_{12} \). We could thus evaluate a fit of Model 1’ using the statistic \( T_2 \) in (12). If we further assume the error variable \( e_1 \) is normal, the number of parameters reduces to five, so that the estimation problem will become easier.

The condition on the numbers of sample moments and parameters is mathematically ambiguous to guarantee identifiability of Model 1’, distinguishability between Models 1’ and 2’ and chi-squaredness of \( T_2 \). (Here, we say that the Models 1’ and 2’ are distinguishable if the third- or fourth-moment structures of the models are different.) We shall give a precise condition in the following propositions.

We can easily verify the identifiability of Model 1’ in nnSEM using the moments up to fourth-order. In fact, \( \sigma_i(\tau_i) = \sigma_i(\tilde{\tau}_i) \) implies \( \tau_i = \tilde{\tau}_i \) \((i = 2, 3, 4)\). Notice that identifiability holds even if some or all of higher-order cumulants of \( x_2 \) and \( e_1 \) are zero.

We state a precise condition for the distinguishability between Models 1’ and 2’ in Proposition 1 below. Assume that Model 1’: \( x_1 = b_{12} x_2 + e_1 \) holds true, and then the condition to be given in the following is on the parameter \( b_{12} \) and the moments of \( x_2 \) and \( e_1 \). Roughly speaking, the higher-order moment structures of the two regression models are different if the following three conditions are met: i) the exogenous variable \( x_2 \) or \( e_1 \) is non-normal and has different values of the higher-order moments from their counterparts of normal variables; ii) the \( x_1 \) and \( x_2 \) are correlated; iii) the error variable \( e_1 \) has a positive variance,
that is, \( x_1 \) and \( x_2 \) do not have correlation 1 nor -1.

**Proposition 1** Let \( x_1 \) and \( x_2 \) be random variables with positive variances and finite fourth-order moments. Denote by \( r \) the correlation coefficient between \( x_1 \) and \( x_2 \), and denote by \( \sigma_i^{(1)}(\tau_i^{(1)}) \) and \( \sigma_i^{(2)}(\tau_i^{(2)}) \), respectively, the \( i \)-th order moment structures of Models 1' and 2' defined above. Let \( \text{cum}_4(z) = E(z^4) - 3E(z^2)^2 \) for a random variable \( z \) with \( E(z) = 0 \). Assume Model 1' to be true. For simplicity, assume \( E(x_i) = 0 \) and \( b = \sigma_i^{(1)}(\tau_i^{(1)}) \) not hold, then

\[
\text{Proposition 1:}
\]

\[
\begin{align*}
\text{If i) } & 0 < |r| < 1 \text{ and ii) either } E(x_2^4), E(e_1^3), \text{cum}_4(x_2) \text{ or } \text{cum}_4(e_1) \text{ is not zero, then Models 1' and 2' are distinguishable from each other using the} \\
& \text{the assumptions above, that is, it holds that } \sigma_3^{(1)}(\tau_3^{(1)}) \neq \sigma_3^{(2)}(\tau_3^{(2)}) \text{ or } \\
& \sigma_4^{(1)}(\tau_4^{(1)}) \neq \sigma_4^{(2)}(\tau_4^{(2)}). \\
\text{Proof:} & \text{ From the second-order structure, we first obtain that } b_{12} = \text{Cov}(x_1, x_2)/V(x_2) \\
& \text{and } b_{21} = \text{Cov}(x_1, x_2)/V(x_1), \text{ and hence } r^2 = b_{12}b_{21}. \text{ The condition on } r \text{ requires} \\
& 0 < b_{12}b_{21} < 1. \quad (13) \\
\text{The assumption } & \sigma_3^{(1)}(\tau_3^{(1)}) = \sigma_3^{(2)}(\tau_3^{(2)}) \text{ means that} \\
& \begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & b_{21} \\
0 & 0 & b_{21} \\
1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
E(x_2^3) \\
E(e_1^3) \\
E(x_2^3) \\
E(e_1^3)
\end{bmatrix}
= \\
\begin{bmatrix}
1 & b_{21} \\
0 & b_{21} \\
0 & b_{21} \\
0 & b_{21}
\end{bmatrix}
\begin{bmatrix}
E(x_1^3) \\
E(e_1^3)
\end{bmatrix}
,
\end{align*}
\]

from which it follows that \( b_{12}(1 - b_{12}b_{21})E(x_2^3) = 0 \). Using (13), we have \( E(x_2^3) = 0 \) and hence \( E(e_1^3) = 0 \). As a result, if \( E(x_2^3) = E(e_1^3) = 0 \) does not hold, then \( \sigma_3^{(1)}(\tau_3^{(1)}) \neq \sigma_3^{(2)}(\tau_3^{(2)}) \).

The assumption \( \sigma_4^{(1)}(\tau_4^{(1)}) = \sigma_4^{(2)}(\tau_4^{(2)}) \) along with \( \sigma_2^{(1)}(\tau_2^{(1)}) = \sigma_2^{(2)}(\tau_2^{(2)}) \) implies the following relationship between the fourth-order cumulant structures:

\[
\begin{align*}
\text{The assumption } & \sigma_4^{(1)}(\tau_4^{(1)}) = \sigma_4^{(2)}(\tau_4^{(2)}) \text{ also with } \sigma_2^{(1)}(\tau_2^{(1)}) = \sigma_2^{(2)}(\tau_2^{(2)}) \text{ implies the following relationship between the fourth-order cumulant structures:} \\
& \begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & b_{21} \\
0 & 0 & b_{21} \\
1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\text{cum}_4(x_2) \\
\text{cum}_4(e_1) \\
\text{cum}_4(x_1) \\
\text{cum}_4(e_2)
\end{bmatrix}
= \\
\begin{bmatrix}
1 & b_{21} \\
0 & b_{21} \\
0 & b_{21} \\
0 & b_{21}
\end{bmatrix}
\begin{bmatrix}
\text{cum}_4(x_1) \\
\text{cum}_4(e_2)
\end{bmatrix}
,
\end{align*}
\]
from which it follows that \( b_{12}(1 - b_{12}b_{21})\text{cum}_4(x_2) = 0 \). Using (13), we have \( \text{cum}_4(x_2) = 0 \) and hence \( \text{cum}_4(e_1) = 0 \). As a result, if \( \text{cum}_4(x_2) = \text{cum}_4(e_1) = 0 \) does not hold, then \( \sigma_4^{(1)}(\tau_4^{(1)}) \neq \sigma_4^{(2)}(\tau_4^{(2)}) \). Q.E.D.

Next we shall state a precise condition for the chi-squaredness of the test statistic \( T_2 \).

**Proposition 2** Assume the eighth-order moments of \((x_1, x_2)\) to be finite and assume \( V \) in (9) to be nonsingular. Suppose that Model 1’ is true. For simplicity, assume \( E(x_2) \) to be zero. Then the statistic \( T_2 \) in (12) for testing a model fit of Model 1’ asymptotically follows the chi-squared distribution with degrees of freedom \( u - v \).

The proof is omitted here (Readers interested in details can refer to Yuan and Bentler, 1997, for example). Note that the nonsingularity of \( V \) is guaranteed if \( x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_1^2x_2^2, x_1x_2^3, x_2^3, x_1x_2^4, x_1x_2^5, x_1^2x_2^3, x_1^3x_2^2, x_1^4x_2, x_1^5x_2, x_1^6x_2, x_1^7x_2, x_1^8x_2 \), and \( x_2^4 \) are linearly independent with probability one.

The conditions in Propositions 1 and 2 have to be met to compare between Models 1’ and 2’, whether the comparison is made inferentially or descriptively.

The two regression models are testable using either third- or fourth-order moments. It is not necessary to model both third- and fourth-order moments. However, there are two benefits of modeling both of the higher-order moments: i) more degrees of freedom can be obtained; ii) symmetrically-distributed variables can be analyzed. Therefore, we would model both third- and fourth-order moments in this paper (Modeling both third- and fourth-order moments might require more sample sizes than does modeling third- or fourth-order moments only. It should be figured out how many sample sizes are required for the two approaches. However, this would exceed the scope of the paper).

### 4 Simulation experiment

We conducted a simulation experiment to study empirical performance of our method to find the possible direction of a path between two observed variables developed in Section 3.

Here we employed GLS estimation with \( \hat{U} = \hat{V} \) and the test statistic \( T_2 \) in Sections 2.3 and 2.4 utilizing second-, third- and fourth-order moment structures of observed variables.

The simulations consisted of 10,000 trials. In each trial, we generated two-dimensional independent random variates as exogenous variables \([x_2, e_1]^T\) of
We first generated \( x_2 \) and \( \tilde{e}_1 \) independently, each with unit variance. We took 0.8, 0.5 or 0.2 as values of the regression coefficient \( b \). The error term \( e_1 \) is then obtained as \( \sqrt{1 - b^2} \tilde{e}_1 \), so that \( V(bx_2 + e_1) = 1 \).

In the strong non-normality condition, we employed the gamma-distribution \( \Gamma(2.5, 1/\sqrt{2.5}) \) that yields variance 1, skewness 1.27 and kurtosis 2.40 as the distribution of \( x_2 \) and \( \tilde{e}_1 \). In the weak non-normality condition, we employed the gamma-distribution \( \Gamma(7.5, 1/\sqrt{7.5}) \) that yields variance 1, skewness 0.73 and kurtosis 0.80.

Then the exogenous variables were centered to have zero mean and mixed by the following regression model:

\[
x_1 = bx_2 + e_1.
\]

Next we analyzed the data with Model 1\(^\prime\) (\( x_1 \leftarrow x_2 \)) and Model 2\(^\prime\) (\( x_1 \rightarrow x_2 \)). Here these models assume that the explanatory variables are independent of the errors and that all the exogenous variables have nonzero skewnesses and kurtoses.

Table 1 shows the numbers of rejected null hypotheses at the significance level 0.05 in the first and second rows, and the numbers of cases, in the third row, where the chi-square value of Model 1\(^\prime\) (chi1) was smaller than that of Model 2\(^\prime\) (chi2), for each of the six conditions on the level of \( b \) and non-normality.

First we shall examine the empirical significance levels (number of rejections) from the first row [Model 1\(^\prime\) (\( x_1 \leftarrow x_2 \))]. Overall, we would say that the null models were rejected more often than the nominal number 500 for almost all the six conditions and levels of sample sizes. Of course, the levels become closer to the nominal one as the sample size \( N \) gets larger. It may be interesting to see that the empirical levels for the weak non-normality condition are better than for the strong non-normality condition.

Second, we shall examine the statistical power of the test from the second row [Model 2\(^\prime\) (\( x_1 \rightarrow x_2 \))]. The power of 0.8 (8,000 rejections) was achieved when \( N = 200, 300, 200, 300, 500 \) and 1,000, respectively, for the six conditions. Clearly power depends on the magnitude of the regression coefficient \( b \) and the degree of non-normality. Greater power of the test is observed for more severely non-normal populations.

Third, in the third rows where we made comparisons of values of chi-squares for the two models. One can compare these values to find descriptively which model is preferable. A better model receives a smaller value of chi-square, and
Table 1
Numbers of rejected null hypotheses and numbers of cases with \( \chi_1 < \chi_2 \) (10,000 replications).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b = 0.8 ), strong non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>454</td>
<td>661</td>
<td>744</td>
<td>735</td>
<td>770</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>1,085</td>
<td>7,029</td>
<td>9,999</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>7,888</td>
<td>9,630</td>
<td>9,984</td>
<td>9,999</td>
<td>10,000</td>
</tr>
<tr>
<td>( b = 0.8 ), weak non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>273</td>
<td>494</td>
<td>577</td>
<td>633</td>
<td>655</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>450</td>
<td>2,100</td>
<td>7,543</td>
<td>9,718</td>
<td>10,000</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>6,234</td>
<td>8,099</td>
<td>9,609</td>
<td>9,928</td>
<td>9,998</td>
</tr>
<tr>
<td>( b = 0.5 ), strong non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>460</td>
<td>687</td>
<td>720</td>
<td>762</td>
<td>737</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>836</td>
<td>6,199</td>
<td>9,982</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>7,587</td>
<td>9,448</td>
<td>9,976</td>
<td>9,997</td>
<td>10,000</td>
</tr>
<tr>
<td>( b = 0.5 ), weak non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>234</td>
<td>490</td>
<td>569</td>
<td>665</td>
<td>632</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>378</td>
<td>1,816</td>
<td>6,436</td>
<td>9,324</td>
<td>9,993</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>6,241</td>
<td>7,881</td>
<td>9,448</td>
<td>9,851</td>
<td>9,993</td>
</tr>
<tr>
<td>( b = 0.2 ), strong non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>425</td>
<td>702</td>
<td>755</td>
<td>750</td>
<td>728</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>537</td>
<td>1,440</td>
<td>4,605</td>
<td>7,585</td>
<td>9,750</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>6012</td>
<td>7,709</td>
<td>9,255</td>
<td>9,779</td>
<td>9,977</td>
</tr>
<tr>
<td>( b = 0.2 ), weak non-normality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1' (( x_1 \leftarrow x_2 ))</td>
<td>263</td>
<td>528</td>
<td>653</td>
<td>639</td>
<td>613</td>
</tr>
<tr>
<td>Model 2' (( x_1 \rightarrow x_2 ))</td>
<td>263</td>
<td>695</td>
<td>1,448</td>
<td>2,328</td>
<td>4,530</td>
</tr>
<tr>
<td>( \chi_1 &lt; \chi_2 )</td>
<td>5,411</td>
<td>6,192</td>
<td>7,568</td>
<td>8,395</td>
<td>9,272</td>
</tr>
</tbody>
</table>

hence a larger number in the row is preferable. The comparison more correctly chose Model 1' as \( N \) gets larger, and the correct choice with probability 0.8 was achieved when \( N = 100, 100, 100, 200, 200 \) and 300. The necessary sample sizes for the descriptive comparison were not unusually large.
The nnSEM approach worked fairly well other than the condition with the smallest path-coefficient and weak non-normality, although more simulation studies are clearly needed to study to what extent the results above can be generalized.

We finally note that the real data example in the next section is a case similar to the case of \( b = 0.5 \) and the strong non-normality condition in the simulation study.

5 Example with real data

Questionnaire data about criminal psychology were analyzed as an example to illustrate the effectiveness of our method. The survey was conducted with students at Osaka University, Japan (Murakami, 2000). The sample size was 222. Observed variables were standardized to have zero means and unit variances.

We employed nnSEM to explore the possible direction of a path between \( x_1 \) and \( x_2 \). The labels of the variables \( x_1 \) and \( x_2 \) are shown in Table 2. The \( x_1 \) was preceding in time to \( x_2 \). Therefore, the possible direction of a path from background knowledge was \( x_1 \rightarrow x_2 \). However, this was of course unknown to our method. The aim in this real example was to know if our method was really able to find the correct direction of the path without any background knowledge.

The skewnesses and kurtoses of the variables were 1.39, 2.79 for \( x_1 \) and 0.95 and 1.18 for \( x_2 \), and a Kolmogorov-Smirnov test showed that all the variables could not be assumed to come from a normal distribution (significance level 1%). Thus, statistical methods based on the non-normal assumption including our method can be applied to this kind of non-normal data.

The results are shown in Table 3. The Model 1' (\( x_1 \leftarrow x_2 \)) was rejected, whereas the Model 2' with the opposite direction (\( x_1 \rightarrow x_2 \)) was not rejected (\( p \) value = 0.67). The results look reasonable to the substantive argument that \( x_1 \) preceded \( x_2 \) in time.

Table 2

<table>
<thead>
<tr>
<th>Variable labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 ): Sum of scores of items that ask subjective evaluation on frequency of your criminal behavior (fraud) when you were a high school student</td>
</tr>
<tr>
<td>( x_2 ): Sum of scores of items that ask subjective evaluation on frequency of your criminal behavior (fraud) last year</td>
</tr>
</tbody>
</table>
Table 3
Estimated path coefficients, standard errors and model fit indices

<table>
<thead>
<tr>
<th></th>
<th>$b_{12}$ or $b_{21}$</th>
<th>$T_2$ (df)</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model $1'$: $x_1 \leftarrow x_2$</td>
<td>0.59 (0.40)</td>
<td>14.64 (5)</td>
<td>0.01</td>
</tr>
<tr>
<td>Model $2'$: $x_1 \rightarrow x_2$</td>
<td>0.65 (0.12)</td>
<td>3.21 (5)</td>
<td>0.67</td>
</tr>
</tbody>
</table>

6 Discussion

In this paper, we have used higher-order moment structures to extend traditional SEM. We took statistical causal inference in Section 3 as an example of the nnSEM discussed here. We used a simple bivariate model since it is simple, comprehensive and illustrates a problem that many researchers encounter commonly. We conducted a small simulation experiment.

There are many SEM models that are not identified but are identified in the nnSEM framework. There are also many saturated SEM models that are not saturated and whose fit can be evaluated statistically in the nnSEM framework. A simplest one is the one-factor analysis model with two indicators, which is essentially equivalent to the error-in-variable model as noted in Introduction. Literature has only discussed the models with a non-normal factor and normal errors. The general framework of the nnSEM also can estimate these models. Further nnSEM can examine a model fit and estimate models with non-normal errors. It is well-known that a multiple-indicator model with two correlated factors $f_1$ and $f_2$ is equivalent to the model with a simple regression between $f_1$ and $f_2$. Use of nnSEM can distinguish among the three models, namely (i) correlated $f_1$ and $f_2$ (confirmatory factor analysis model), (ii) $f_1 = \gamma_2 f_2 + d_1$, (iii) $f_2 = \gamma_1 f_1 + d_2$. Many complicated equivalent models have been presented in the literature, see e.g., Hershberger (1994). The nnSEM has a possibility to resolve the equivalence problem for those models. We would continue to study when nnSEM works.

It is often claimed that causal inference is not a statistical issue but a substantive one. We realize the limitation of causal analysis with nnSEM as well as with traditional SEM (e.g., Freedman, 1987; Rogosa, 1987; Bullock et al., 1994; Holland, 1986; Mulaik and James, 1995; Rosenbaum, 2002). One critique is that unobserved confounding variables can distort statistical causal inference and that it will be impossible to eliminate their effects in observational studies. However, there are cases where only observational data are available and one cannot conduct any experimental studies. For these cases, SEM and nnSEM discussed here will be powerful tools for causal inference, although one has to pay very careful attention to confounding variables. One caution should be made here. Practical users must not assert that causation is established based solely on the results derived from our procedure but should make sub-
stantive arguments as well. Statistical methods need any help of substantive arguments, and substantive arguments need any help of statistical methods, particularly in observational studies.

References


Micceri, T., 1989. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin 105 (1), 156–166.


