Maximum likelihood estimation for nonlinear structural equation models with normal latent variables

Yan D. Zhao\textsuperscript{a,}\textsuperscript{*} and Dewi Rahardja\textsuperscript{b}
\textsuperscript{a}Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN, USA
\textsuperscript{b}Department of Statistical Science, Baylor University, Waco, TX, USA

Abstract. Structural equation analysis has been widely used in behavioral and social sciences. In practice, for continuous observed variables, linear structural equation models have been used nearly exclusively. The use of models that are nonlinear in latent variables has been limited to simple situations with a linear measurement model and with a single polynomial structural relationship, usually containing only one quadratic or cross-product term. This paper introduces a general structural equation model with a nonlinear measurement model and a simultaneous system of nonlinear and non-polynomial structural relationships. For such a model, a parameterization useful for identification and interpretation is presented. For model fitting and parameter inferences, the maximum likelihood approach is considered. A method for obtaining parameter estimates and their asymptotic covariance matrix estimate is developed based on a new version of the Monte Carlo EM algorithm. The performance of the algorithm is examined using simulation studies.

Keywords: Errors-in-variables, latent variable analysis, Monte Carlo EM algorithm, seemingly unrelated regression

1. Introduction

Structural equation analysis is a broadly used statistical method for assessing the quantitative relationships among underlying latent variables. In many applications, some concepts are not directly observable, but measurements or indicators related to the underlying variables can be obtained. A structural equation system generally consists of two sub-models, a measurement model relating the observable indicators to latent variables, and a structural (path) model describing simultaneous relationships among the latent variables. A system with a linear measurement model and a linear structural model has been extensively applied in practice. For general reference, see, e.g., Bollen [3], Bentler [2], and Jöreskog and Sörbom [6].

For continuous observed variables, the literature on nonlinear structural equation system has been limited to situations with simple models. For a system with a linear measurement model and a simple nonlinear structural model with an interaction or square term, Kenny and Judd [8] proposed a model fitting method. This method was further discussed by Jaccard and Wan [5] and Jöreskog and Yang [7]. Wall and Amemiya [14] gave a modified and generalized version of the method by Kenny and Judd [8], covering non-normal underlying variables. For the same system, Bollen [4] proposed the use of instrumental variables, and Arminger and Muthen [1] applied a Bayesian approach. Despite the value of these methods for simple models, extending any of these to more general situations is impractical, if not impossible. Wall and Amemiya [13] considered a system with a linear measurement model and a single polynomial structural model. They introduced a new estimator that is consistent for any polynomial structural model without specifying the distributional form of the latent variables. They also developed accurate asymptotic
inference procedures. Extending their estimation method to a system of polynomial reduced-form structural models is straightforward. However, their method and idea are not directly applicable to non-polynomial models.

In this paper, we introduce a very general nonlinear structural equation model. The structural part of the model is a simultaneous system of multiple nonlinear equations. The measurement part is also a set of nonlinear relationships, none of which needs to be polynomial. The standard approach for fitting a linear structural equation model is that of maximum likelihood. Unfortunately, for a model nonlinear in underlying variables, the likelihood cannot be given explicitly. We overcome the difficulty in the likelihood approach by developing a new version of the Monte Carlo EM algorithm. Previously, various EM algorithms have been proposed for simple models involving latent variables. See, e.g., Meng and Schilling [12] and, Klein et al. [9]. Our approach differs from the existing ones and utilizes a particular structure of our general model. The new algorithm developed here is also practical by capitalizing on a simple routine for weighted seemingly unrelated regression.

The remainder of this paper is organized as follows. In Section 2, we define a general structural equation system with a nonlinear measurement model and a simultaneous nonlinear structural model. Our Monte Carlo EM algorithm is described in Section 3. Section 4 is devoted to simulation studies. Some discussion can be found in Section 5.

2. Nonlinear structural equation system

A general nonlinear structural equation system can be expressed in the following way. Let \( i = 1, 2, \ldots, n \) denote individuals in a sample of size \( n \), and let a \( p \times 1 \) vector \( Z_i \) include observed variables from the \( i^{th} \) individual. We consider a model appropriate for situations where all observed measurements are continuous or scale-type measurements taking more than a few distinct values. Suppose that \( Z_i \) is an observed indicator of a \( k \times 1 \) underlying latent vector \( z_i \). A structural equation system consists of a measurement model relating the observation \( Z_i \) to the latent variable \( z_i \), and a structural or path model representing relationships among the elements of \( z_i \). We assume that both measurement and structural models can be nonlinear in the latent variable \( z_i \).

A general nonlinear measurement model with additive measurement error can be expressed in the form

\[
Z_i = H_0(z_i) + u_i, 
\]

However, this model is not identifiable in the sense that the factor vector \( z_i \) can be transformed without altering the meaning of the model. Although the general identification issue for models nonlinear in latent variables is not trivial, the so-called errors-in-variables parameterization provides a relatively simple way to represent an identifiable nonlinear model. That is, we consider a parametric nonlinear measurement model that can be written in the form

\[
Z_i = \left( H(z_i; \lambda) \right)_{z_i} + u_i, \tag{1}
\]

where \( \lambda \) is an unknown parameter vector, the form of the function \( H \) is known, and \( u_i \) is a zero-mean measurement error vector independent of \( z_i \). In this form, the latent vector \( z_i \) is identified and interpreted as the true underlying values of the last \( k \) components of \( Z_i \). In considering possible nonlinear structural relationships among the elements of \( z_i \), this parameterization with interpretable latent variables is helpful. For all the parameters in model (1) to be estimable, the dimension of the latent factor \( k \) needs to be sufficiently small as compared to the number of observed indicators \( p \), and the distribution of \( u_i \) needs to be sufficiently simple (e.g., with independent components, or with a limited number of nonzero covariances). This requirement has to be assessed for each given case. The identification of model (1) can depend on the assumed latent structural model for \( z_i \), if it is a non-trivial model with real restrictions on \( z_i \). Since we may be interested in considering several different structural models, a safe approach might be to check the identification of model (1) without restriction given by a structural model for \( z_i \).

For a structural model specifying nonlinear relationships among the elements of \( z_i \), an implicit simultaneous model can be written as

\[
G_0(z_i) = e_i, \tag{2}
\]

where \( e_i \) is a \( q \times 1 \) equation error with zero mean, and \( q \) is the number of (non-redundant) relationships. In general, the covariance matrix of \( e_i \) is unrestricted, contributing to the simultaneity of model (2). However, in some
applications, a diagonal covariance matrix is assumed. Once again, this general model form is ambiguous, because, e.g., the whole equation can be multiplied by a matrix without changing the essential meaning of the relationships. In the linear simultaneous system, a model is uniquely specified by setting some coefficients to be one, and assuming the existence of a nonsingular coefficient matrix, i.e., the existence of an explicitly solved reduced form. Using the same idea, we assume that model (2) can be solved explicitly for \( q \) elements of \( z_i \) in terms of the other \( k - q \) elements. To express such a reduced form representation of the structural model, write

\[
\mathbf{z}_i = \begin{pmatrix} y_i \\ x_i \end{pmatrix}, \quad \mathbf{w}_i = \begin{pmatrix} x_i \\ e_i \end{pmatrix},
\]

where \( y_i = 1 \times q, x_i = (k - q) \times 1, \) and \( e_i \) is \( k \times 1 \). We note that \( x_i, y_i, \) and \( e_i \) are unobservable. We assume that model (2) can be expressed in an identifiable and explicit form

\[
y_i = G(w_i; \beta),
\]

where \( \beta \) is a vector of relationship parameters. Note that, when a model based on subject-matter meaning has the form Eq. (2) with a zero mean error \( e_i \), then the reduced form Eq. (4) can be nonlinear in \( e_i \).

Some examples of nonlinear structural models (all with \( q = 2 \) for illustration) are as follows:

1. A quadratic mediator model in form Eq. (2) is

\[
z_{1i} = \beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i}^2 + e_{1i}, \quad z_{2i} = \beta_4 + \beta_5 z_{3i} + e_{2i}.
\]

This system can be written in form Eq. (4) by substituting the second equation into the first, and setting \( x_i = z_{3i} \).

2. A simultaneous quadratic model (2) might take the form

\[
z_{1i} = \beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i} + \beta_4 z_{4i}^2 + e_{1i}, \quad z_{2i} = \beta_5 + \beta_6 z_{1i} + \beta_7 z_{4i} + \beta_8 z_{3i}^2 + e_{2i},
\]

which can be written in the reduced form Eq. (4) by the usual linear transformation.

3. For two endogenous variables \( z_{1i} \) and \( z_{2i} \) with possibly bounded ranges and an unbounded exogenous variable \( z_{3i} \), a reasonable model can be

\[
z_{\ell i} = \beta_1 + \frac{\beta_2}{1 + e^{-\beta_3 z_{3i}}} + e_{\ell i}, \quad \ell = 1, 2,
\]

which is already in the reduced form Eq. (4).

To express the overall system in a single model, the explicit structural model (4) can be substituted into the errors-in-variables measurement model (1) to obtain a reduced form for observations

\[
\mathbf{Z}_i = \mathbf{K}(w_i; \alpha) + \mathbf{u}_i, \quad i = 1, 2, \ldots, n,
\]

where

\[
\mathbf{K}(w_i; \alpha) = \begin{pmatrix} \mathbf{H}(G(w_i; \beta), x_i; \lambda) \\ G(w_i; \beta) \\ x_i \end{pmatrix},
\]

\[
\alpha = \begin{pmatrix} \beta \\ \lambda \end{pmatrix}.
\]

We consider estimation of the parameters in the nonlinear structural equation system Eq. (5). We assume that the underlying latent variables \( x_i, u_i, \) and \( e_i \) are independent normally distributed random vectors

\[
x_i \sim N(\mu_x, \Sigma_x), \quad u_i \sim N(0, \Sigma_u), \quad e_i \sim N(0, \Sigma_e).
\]
The normality of the measurement error \( \mathbf{u}_i \) and the equation error \( \mathbf{e}_i \) can be considered reasonable in most applications. The distribution of the latent factor \( \mathbf{x}_i \), characterizing the target population may not be automatically assumed normal. But, the errors-in-variables parameterization of the measurement model (1) can be useful in assessing the normality of \( \mathbf{x}_i \), or in suggesting possible transformation, because some observed measurements are linear functions of \( \mathbf{x}_i \) and \( \mathbf{u}_i \). Once the measurement model (1) can be expressed using a transformed \( \mathbf{x}_i \), with normality, the structural model (4) needs to be re-expressed in terms of the new \( \mathbf{x}_i \), but without changing the general nonlinear form Eq. (4).

Under the errors-in-variables parameterization, the latent vector \( \mathbf{x}_i \) is treated as unrestricted. Thus, the mean \( \mu_{\mathbf{x}_i} \) is any \((k - q) \times 1\) vector, and the covariance matrix \( \Sigma_{\mathbf{x}_i} \) is any unrestricted \((k - q) \times (k - q)\) positive definite matrix. Let \( \mathbf{\sigma}_{\mathbf{x}} \) be a vector containing distinct elements of \( \Sigma_{\mathbf{x}_i} \). The measurement error covariance matrix \( \Sigma_{\mathbf{u}_i} \) is often assumed to be diagonal, although some covariances may be assumed to be present. We use a general notation \( \mathbf{\sigma}_{\mathbf{u}} \) to denote a vector of unknown elements of \( \Sigma_{\mathbf{u}_i} \). In most applications, no information is available for the equation error \( \mathbf{e}_i \), and \( \Sigma_{\mathbf{e}} \) is considered an unrestricted covariance matrix. But, in some applications, researchers may be willing to assume some zero covariances. To cover such situations, we denote a vector of unknown elements of \( \Sigma_{\mathbf{e}} \) by \( \mathbf{\sigma}_{\mathbf{e}} \). Define a vector containing all parameters as

\[
\mathbf{\theta} = \begin{pmatrix} \mathbf{\theta}_r \\ \mathbf{\theta}_w \end{pmatrix},
\]

where

\[
\mathbf{\theta}_r = \begin{pmatrix} \mathbf{\alpha} \\ \mathbf{\sigma}_{\mathbf{u}} \end{pmatrix}, \quad \mathbf{\theta}_w = \begin{pmatrix} \mu_{\mathbf{x}} \\ \mathbf{\sigma}_{\mathbf{x}} \\ \mathbf{\sigma}_{\mathbf{e}} \end{pmatrix}.
\]

Note that \( \mathbf{\theta}_r \) contains the parameters of the observation reduced model (5) including \( \mathbf{\alpha} \) defined in Eq. (7), and that \( \mathbf{\theta}_w \) includes the distributional parameters for \( \mathbf{w}_i \) defined in Eq. (3).

### 3. Monte Carlo EM algorithm

The likelihood function based on observations \( \mathbf{Z}_i, i = 1, 2, \ldots, n \), satisfying Eqs (5) and (8) does not have an explicit form in general. From the appearance of the observation reduced model (5), it is natural to consider some form of the EM algorithm treating \( \mathbf{w}_i \) in Eq. (3) as missing. Our approach takes advantage of the fact that the conditional distribution of the observation \( \mathbf{Z}_i \) given \( \mathbf{w}_i \) is particularly simple. Using the \( f \) notation for density functions (and not using the lower-case notation for the observed values), we write the complete-data likelihood as

\[
L_c(\mathbf{\theta}) = \prod_{i=1}^{n} f(\mathbf{Z}_i|\mathbf{w}_i; \mathbf{\theta}_r) f(\mathbf{w}_i; \mathbf{\theta}_w).
\]

Under Eq. (8), both density functions appearing in Eq. (10) are normal. Note that the conditional density of \( \mathbf{Z}_i \) given \( \mathbf{w}_i \) depends only on \( \mathbf{\theta}_r \), and that the marginal density of \( \mathbf{w}_i \) depends only on \( \mathbf{\theta}_w \). In describing our approach for developing an EM type algorithm, we denote the value of \( \mathbf{\theta} \) at the \( j^{th} \) iteration by

\[
\mathbf{\theta}^{(j)} = \begin{pmatrix} \mathbf{\theta}_r^{(j)} \\ \mathbf{\theta}_w^{(j)} \end{pmatrix}.
\]

The E-step in the \((j + 1)^{st}\) iteration is designed to obtain the conditional expectation of \( \log (L_c(\mathbf{\theta})) \) given \( \mathbf{Z}_i, i = 1, 2, \ldots, n \), under the distribution specified by \( \mathbf{\theta}^{(j)} \).

\[
E^{(j)}(\mathbf{\theta}) = E \{ \log (L_c(\mathbf{\theta})) | \mathbf{Z}_1, \ldots, \mathbf{Z}_n \}
\]

\[
= \sum_{i=1}^{n} \left\{ \int \log (f(\mathbf{Z}_i|\mathbf{w}_i; \mathbf{\theta}_r)) + \log (f(\mathbf{w}_i; \mathbf{\theta}_w)) \right\} f(\mathbf{w}_i|\mathbf{Z}_i; \mathbf{\theta}^{(j)}) d\mathbf{w}_i,
\]
Using a simple representation of the joint density in Eq. (10), we express the conditional density $f(w_i|Z_i; \theta^{(j)})$ in a particular ratio form

$$f(w_i|Z_i; \theta^{(j)}) = \frac{f(Z_i|w_i; \theta^{(j)}) f(w_i; \theta^{(j)})}{\int f(Z_i|w; \theta^{(j)}) f(w; \theta^{(j)}) dw}.$$ \hspace{1cm} (13)

When Eq. (13) is substituted into Eq. (12), the $i^{th}$ term in the summation involves two integrals. We propose to compute these integrals by Monte Carlo simulations generating pseudo random variables from the $f(w; \theta^{(j)})$ distribution. We choose to use this approach, because $f(Z_i|w; \theta^{(j)})$ is simple to evaluate, and because random numbers from $f(w; \theta^{(j)})$ can be readily generated. For each individual $i$, we generate a pseudo random sample

$$\{\tilde{w}_{im} : m = 1, 2, \ldots, M^{(j)} \}$$ \hspace{1cm} (14)

of size $M^{(j)}$, where

$$\tilde{w}_{im} = \left( \tilde{x}_{im}, \tilde{e}_{im} \right),$$

$$\tilde{x}_{im} \sim N(\mu_X^{(j)}, \Sigma_X^{(j)}),$$

$$\tilde{e}_{im} \sim N(0, \Sigma_{e}^{(j)}).$$

To decrease the simulation error in the eventual estimates, we generate $\{\tilde{w}_{im} : m = 1, 2, \ldots, M^{(j)} \}$ separately and independently for different $i$'s and for different iterations. But, we use the same set of size $M^{(j)}$ for the two integral computations involved for each $i$. This is because the denominator integral plays the role of a normalizing constant for weights adding up to one for each $i$ as given in Eq. (18) below. Then, our simulation-based E-step gives

$$\hat{E}^{(j)}(\theta) = \hat{E}_r^{(j)}(\theta_r) + \hat{E}_w^{(j)}(\theta_w),$$ \hspace{1cm} (15)

where

$$\hat{E}_r^{(j)}(\theta_r) = \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} \log (f(Z_i|\tilde{w}_{im}; \theta_r)) \ a_{im},$$ \hspace{1cm} (16)

$$\hat{E}_w^{(j)}(\theta_w) = \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} \log (f(\tilde{w}_{im}; \theta_w)) \ a_{im},$$ \hspace{1cm} (17)

$$a_{im} = \frac{f(Z_i|\tilde{w}_{im}; \theta^{(j)}_r)}{\sum_{\ell=1}^{M^{(j)}} f(Z_i|\tilde{w}_{im}; \theta^{(j)}_w)}.$$ \hspace{1cm} (18)

The M-step in the $(j+1)^{st}$ iteration finds $\theta^{(j+1)}$ that maximizes $\hat{E}^{(j)}(\theta)$ in Eq. (15). Our way of formulating the E-step also simplifies the M-step, because the parameters $\theta_r$ and $\theta_w$ in Eq. (15) are separated in two terms, and because each of $\hat{E}_r^{(j)}(\theta_r)$ and $\hat{E}_w^{(j)}(\theta_w)$ is a weighted sum of the log-transformed normal densities with weights $a_{im}$. In particular, $\hat{E}_w^{(j)}(\theta_w)$ is maximized by taking

$$\mu_X^{(j+1)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} \tilde{x}_{im} \ a_{im},$$

$$\Sigma_X^{(j+1)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} \left( \tilde{x}_{im} - \mu_X^{(j+1)} \right) \left( \tilde{x}_{im} - \mu_X^{(j+1)} \right)^t \ a_{im},$$ \hspace{1cm} (19)

$$\Sigma_{e}^{(j+1)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} \tilde{e}_{im} \tilde{e}_{im}^t \ a_{im}.$$
where we have used $\sum_{m=1}^{M^{(j)}} a_{im} = 1$ for every $i$. If some covariance elements of $\Sigma_{e}$ are known to be zero (e.g., diagonal), then the corresponding elements of $\Sigma_{e}^{(j+1)}$ are set to be zero. We combine the values in Eq. (19) to form $\theta_{(j+1)}^{(j+1)}$. As for the maximization of $\tilde{E}_{r}^{(j)}(\theta)$, no explicit formula exists in general. We provide an iterative procedure to obtain $\theta_{(j+1)}^{(j+1)}$ as described in the Appendix. Finally, the $(j + 1)^{st}$ EM stage estimate of $\theta$ is

$$
\theta_{(j+1)}^{(j+1)} = \begin{pmatrix} \theta_{(j+1)}^{(j+1)} \\ \theta_{(j+1)}^{(j+1)} \end{pmatrix}.
$$

To estimate the asymptotic covariance matrix of the maximum likelihood estimator $\hat{\theta}$, the final estimate from the MCEM algorithm, the observed or expected information matrix needs to be obtained. In a typical EM setting, the method proposed by Louis [10] is used. However, due to the large number of parameters in our model, this can be algebraically tedious or even intractable. Instead, we propose the use of an estimated information matrix given in McLachlan and Krishnan [11, pp. 120–122]. Let the score vector for the observed data log-likelihood function be

$$
S(Z; \theta) = \sum_{i=1}^{n} s(Z_i; \theta),
$$

where

$$
s(Z_i; \theta) = \partial \log f(Z_i; \theta) / \partial \theta.
$$

Then, the expected information matrix can be estimated by its empirical form

$$
I_{e}(\theta; Z) = \sum_{i=1}^{n} s(Z_i; \theta)s'(Z_i; \theta) - \frac{1}{n}S(Z; \theta)S'(Z; \theta).
$$

Evaluated at $\theta = \hat{\theta}$, $I_{e}(\theta; Z)$ reduces to

$$
\hat{I}_{e}(\theta; Z) = \sum_{i=1}^{n} \hat{s}(Z_i; \theta)s'(Z_i; \theta).
$$

Each component of $s(Z_i; \hat{\theta})$ is derived and provided in the Appendix. It can be shown that $s(Z_i; \theta) = E(\partial l_{e_{i}}(\theta) / \partial \theta | Z; \theta)$, where $l_{e_{i}}(\theta) = \log f(Z_i, w_i; \theta)$. Written in this form, an estimate $\hat{s}(Z_i; \theta)$ of $s(Z_i; \theta)$ can be obtained using our Monte Carlo integration method. Then, we estimate the asymptotic covariance matrix of $\hat{\theta}$ by the inverse of $\hat{I}_{e}(\theta, Z)$, where

$$
\hat{I}_{e}(\theta; Z) = \sum_{i=1}^{n} \hat{s}(Z_i; \theta)s'(Z_i; \theta).
$$

### 4. Simulation study

This section reports two simulation studies using two different structural models: exponential and simultaneous quadratic model. The properties of the maximum likelihood estimation and the performance of our MCEM algorithm are examined. For both studies, 1,000 samples of sample size $n = 200$ were generated. For both studies, we used a linear measurement model with the separate indicator structure and a diagonal measurement error covariance matrix. The first simulation study uses the exponential model

$$
y_{i} = \beta_{0} \exp(-\beta_{1} x_{i}) + \epsilon_{i},
$$

with four observed variables $Z_{1i}$, $Z_{2i}$, $Z_{3i}$, and $Z_{4i}$ satisfying the measurement model

$$
\begin{pmatrix} Z_{1i} \\ Z_{2i} \\ Z_{3i} \\ Z_{4i} \end{pmatrix} = \begin{pmatrix} \lambda_{01} \\ \lambda_{02} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{i} \\ y_{i} \end{pmatrix} + \begin{pmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \\ u_{4i} \end{pmatrix},
$$

(23)
Table 1

<table>
<thead>
<tr>
<th></th>
<th>bias</th>
<th>rmse</th>
<th>cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>0.256</td>
<td>1.960</td>
<td>0.951</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.003</td>
<td>0.035</td>
<td>0.956</td>
</tr>
</tbody>
</table>

Table 1: The empirical bias, root mean square error, and the empirical coverage probability of nominal 95% CI for the exponential structural model (1,000 samples).

where \(u_1, \ldots, u_4\) are known to be independent. The true parameter values were set to be \(\beta_0 = 20\), \(\beta_1 = 0.5\), \(\lambda_{01} = 2\), \(\lambda_{11} = 0.8\), \(\lambda_{02} = 3\), and \(\lambda_{12} = 0.6\). Note that the model (22) with a non-polynomial relation cannot be handled by any existing procedure.

Since the equation error and the measurement error \(u_4\) are once again confounded, We write

\[ Z_{4i} = \beta_0 \exp(-\beta_1 x_i) + e_i + u_{4i} = \beta_0 \exp(-\beta_1 x_i) + v_i, \]

and estimate \(\text{Var}\{v_i\}\). The latent variable \(x_i\) is normally distributed with mean 3 and variance 1. The measurement errors \(u_{1i}, u_{2i}, u_{3i}\), and \(v_i\) were generated from zero mean normal distribution with the variance selected so that the reliability for all of \(Z_{1i}, Z_{2i}, Z_{3i}\), and \(Z_{4i}\) is 0.75.

Figure 1 presents the box-plots of the 1,000 estimates of \(\beta_0\) and \(\beta_1\), showing small bias and nearly symmetric empirical distribution. Table 1 gives the empirical bias and root mean square error, along with empirical coverage probabilities for the nominal 95% confidence intervals. We see that the confidence intervals using the standard error estimates in Eq. (21) are nearly accurate.

In the second simulation study, we consider a simultaneous quadratic structural model

\[
\begin{align*}
y_{1i} &= \beta_1 + \beta_2 y_{2i} + \beta_3 x_{1i} + \beta_4 x_{1i}^2 + e_{1i}, \\
y_{2i} &= \beta_5 + \beta_6 y_{1i} + \beta_7 x_{2i} + \beta_8 x_{2i}^2 + e_{2i},
\end{align*}
\]

where the true values for \(\beta_1, \beta_2, \ldots, \beta_8\) were taken to be 3, 2, 5, \(-3\), 5, 3, 4, and \(-6\). The measurement model has the following form
Table 2
The empirical bias and the square root of the mean squared error for the simultaneous quadratic structural model (1,000 samples)

<table>
<thead>
<tr>
<th></th>
<th>bias</th>
<th>rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_4$</td>
<td>-0.312</td>
<td>0.913</td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>-1.372</td>
<td>3.889</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.006</td>
<td>0.150</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.271</td>
<td>1.124</td>
</tr>
</tbody>
</table>

$\begin{pmatrix} Z_{1i} \\ Z_{2i} \\ Z_{3i} \\ Z_{4i} \\ Z_{5i} \\ Z_{7i} \\ Z_{8i} \end{pmatrix} = \begin{pmatrix} \lambda_{01} \\ \lambda_{02} \\ \lambda_{03} \\ \lambda_{04} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_{11} \\ \lambda_{12} \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ y_{1i} \\ y_{2i} \end{pmatrix} + \begin{pmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \\ u_{4i} \\ u_{5i} \\ u_{7i} \\ u_{8i} \end{pmatrix},$

with true values

$\begin{pmatrix} \lambda_{01} \\ \lambda_{11} \\ \lambda_{02} \\ \lambda_{12} \\ \lambda_{03} \\ \lambda_{13} \\ \lambda_{04} \\ \lambda_{14} \end{pmatrix} = \begin{pmatrix} 3 & 0.9 & 4 & 0.6 \\ 3 & 0.9 & 4 & 0.6 \end{pmatrix}.$

The latent variables $(x_{1i}, x_{2i})$ were generated from

$N \left( \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \right)$

distribution. The first 6 measurement errors $u_{1i}, \ldots, u_{6i}$ were generated as zero mean normal random variables with variances chosen so that the reliability for each of $Z_{1i}, \ldots, Z_{6i}$ was 0.75. The variables $Z_{7i}$ and $Z_{8i}$ were generated from the reduced form

$\begin{pmatrix} Z_{7i} \\ Z_{8i} \end{pmatrix} = A \left( \beta_1 + \beta_3 x_{1i} + \beta_4 x_{1i}^2 \right) + v_i,$

where

$A = \begin{pmatrix} 1 & -\beta_2 \\ -\beta_0 & 1 \end{pmatrix}^{-1}, \quad v_i = A \begin{pmatrix} u_{7i} + e_{1i} \\ u_{8i} + e_{2i} \end{pmatrix}.$

The $v_i$ were zero mean normal vectors so that the reliability for each of $Z_{7i}$ and $Z_{8i}$ was 0.75, and the correlation between the two components of $v_i$ was 0.4.

For the estimated structural parameters $\beta_4, \beta_8, \beta_2$ and $\beta_6$, Table 2 presents the empirical bias and the root mean square error, and Fig. 2 gives the box-plots of the 1,000 estimates. Again we see small bias and nearly symmetric empirical distribution of the estimates.

5. Discussion

In this paper we introduced a general structural equation model with a nonlinear measurement model and a simultaneous system of nonlinear and non-polynomial structural relationships. For such a model, a parameterization useful for identification and interpretation is presented. For model fitting and parameter inferences, the maximum likelihood approach is considered. A method for obtaining parameter estimates and their asymptotic covariance matrix estimate is developed based on a new version of the Monte Carlo EM algorithm. The performance of the algorithm is examined using simulation studies. Our simulation shows that the performance of our algorithm is very good.
Fig. 2. Box-plots for the simultaneous quadratic structural model (1,000 samples).

The MCEM algorithm converges fairly fast. For the exponential model with sample size \( n = 200 \) in the Section 4, on the average each sample required approximately 20 minutes on a DEC Alpha workstation with 256 MB RAM and 333 MHZ CPU. Our simulation program was written in C Language with some of the functions called in from NAG. The C programs are available upon request.

For future work, the normal assumption on the latent vector may be relaxed and a distribution-free method is desired. In addition, there may be data where the latent vector is categorical which takes a few distinct values. Much work is in need for such data.

Appendix

Maximization of \( \hat{E}_r^{(j)}(\theta_r) \)

The conditional reduced form Eq. (5) is in a nonlinear multivariate regression, and the likelihood in Eq. (16) is a weighted average of normal log-likelihood functions. Taking advantage of these particular forms involved in \( \hat{E}_r^{(j)}(\theta_r) \), we propose the use of an iteratively re-weighted generalized least squares (i.e., iteratively weighted nonlinear seemingly unrelated regression) with \( n M^{(j)} \) observations and \( p \) equations. Given \( \Sigma_{\mathbf{u}}^{(j)} \) and \( a_{im} \) in Eq. (18), we first obtain the value \( \hat{\alpha} \) of \( \alpha \) that minimizes

\[
\sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} a_{im} \left[ Z_i - K(\tilde{w}_{im}; \alpha) \right]' \left( \Sigma_{\mathbf{u}}^{(j)} \right)^{-1} \left[ Z_i - K(\tilde{w}_{im}; \alpha) \right].
\]

This nonlinear regression minimization can be carried out using, e.g., the modified Gauss-Newton algorithm. Specifically, Let \( \hat{\alpha}^{(t)}(0) \) be the \( t^{th} \) step estimate of \( \hat{\alpha} \) in the modified Gauss-Newton algorithm with \( \hat{\alpha}^{(0)} = \alpha^{(j)} \) as the initial value. Then, the Gauss-Newton correction term for the \( (t + 1)^{st} \) step is

\[
D^{(t)} = \left[ \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} a_{im} F_{im}' \left( \Sigma_{\mathbf{u}}^{(j)} \right)^{-1} F_{im} \right]^{-1} \left[ \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} a_{im} F_{im}' \left( \Sigma_{\mathbf{u}}^{(j)} \right)^{-1} \left[ Z_i - K(\tilde{w}_{im}; \hat{\alpha}^{(t)}) \right] \right],
\]

where
\[
F_{im} = \frac{\partial K(\hat{\omega}_{im} ; \hat{\alpha}(t))}{\partial \alpha} = \frac{\partial K(\hat{\omega}_{im} ; \alpha)}{\partial \alpha} |_{\alpha = \alpha(t)}.
\]

The \((t+1)^{st}\) step estimate is \(\hat{\alpha}_{(t+1)} = \hat{\alpha}_t + \tau(t)D(t)\), where the step length \(\tau(t)\) may be taken to be some sequential powers of 0.5 to guarantee the decrease of Eq. (25) at each iteration. Let \(\hat{\alpha}\) be the final estimate after this modified Gauss-Newton algorithm converges. Then, a new estimate of \(\Sigma_u\) is

\[
\hat{\Sigma}_u = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M^{(j)}} a_{im} [Z_i - K(\hat{\omega}_{im} ; \hat{\alpha})] [Z_i - K(\hat{\omega}_{im} ; \hat{\alpha})]^t.
\]

The processes (25) and (27) may need to be iterated only a few times (or even once), because these iterative steps are nested in the overall EM algorithm, and because some relationship between the iteratively re-weighted generalized least squares and the weighted normal likelihood form in Eq. (16). If some covariances in \(\Sigma_u\) are predetermined to be zero (e.g., diagonal \(\Sigma_u\)), then the corresponding covariance components in Eq. (27) are set to be zero. The new estimate \(\hat{\theta}^{(j+1)}\) consists of \(\hat{\alpha}\) from Eq. (25) and \(\hat{\Sigma}_u\) in Eq. (27) after a few iterations of the re-weighted least squares.

**Components of \(s(Z_i; \hat{\theta})\)**

With \(\{\hat{\omega}_{im}\}\) and \(\{a_{im}\}\) obtained in the last step of the MCEM algorithm, the explicit formulas for our general model are

\[
\hat{s}(Z_i, \hat{\mu}_X) = \sum_{m=1}^{M^{(j)}} a_{im} \hat{\Sigma}_u^{-1} \hat{x}_{im} - \hat{\Sigma}_x^{-1} \hat{\mu}_X,
\]

\[
\hat{s}(Z_i, \hat{\sigma}_x) = \frac{1}{2} \sum_{m=1}^{M^{(j)}} a_{im} (\hat{x}_{im} - \hat{\mu}_X)^t \hat{\Sigma}_x^{-1} \hat{\Sigma}_x^{-1} (\hat{x}_{im} - \hat{\mu}_X) - \frac{1}{2} tr(\hat{\Sigma}_x^{-1} \frac{\partial \hat{\Sigma}_x}{\partial \sigma_{xk}}),
\]

\[
\hat{s}(Z_i, \hat{\sigma}_e) = \frac{1}{2} \sum_{m=1}^{M^{(j)}} a_{im} e_{im}^t \hat{\Sigma}_e^{-1} e_{im} - \frac{1}{2} tr(\hat{\Sigma}_e^{-1} \frac{\partial \hat{\Sigma}_e}{\partial \sigma_{kl}}),
\]

\[
\hat{s}(Z_i, \hat{\alpha}) = \sum_{m=1}^{M^{(j)}} a_{im} \frac{\partial K(\hat{\omega}_{im} ; \hat{\alpha})}{\partial \alpha} \hat{\Sigma}_u^{-1} (Z_i - K(\hat{\omega}_{im} ; \hat{\alpha})),
\]

\[
\hat{s}(Z_i, \hat{\Sigma}_u) = \frac{1}{2} \sum_{m=1}^{M^{(j)}} a_{im} (Z_i - K(\hat{\omega}_{im} ; \hat{\alpha}))^t \hat{\Sigma}_u^{-1} (Z_i - K(\hat{\omega}_{im} ; \hat{\alpha}))
\]

\[
- \frac{1}{2} tr(\hat{\Sigma}_u^{-1} \frac{\partial \hat{\Sigma}_u}{\partial \sigma_{us}}),
\]

where \(\sigma_{xk}\) is the \(k^{th}\) element of \(\sigma_{X}\), \(\sigma_{ei}\) is the \(l^{th}\) element of \(\sigma_{E}\), and \(\sigma_{us}\) is the \(s^{th}\) element of \(\sigma_{U}\).

**References**


