1 Introduction

Structural equation modeling originated (Jöreskog (1973), Bentler (1980), Bollen (1989)) as a method for modeling linear relations among observed and hypothesized latent variables. Despite limitation inherent in the linearity assumption of traditional structural equation modeling, it has indeed provided a revolutionary and popular framework for addressing research questions in the social, psychological and behavioral sciences where latent variable are quite common. In order to expand the flexibility and thus applicability of this already useful statistical modeling method, a natural extension is to include the possibility of modeling nonlinear relations among the latent variables in addition to linear relations.

There has been growing literature (some of which described later in this paper) developing different kinds nonlinear structural equation models and estimation methods for them. Generally, the estimation methods in this literature can be described as either making and relying on distributional assumptions for the underlying latent variables or instead leaving the distribution unspecified. Furthermore, the methods can be described as being either tailor-made to a specific sort of nonlinear structural model, i.e. polynomial or specifically low dimensional polynomial, or else being applicable to a more general nonlinear structural model.

In this paper we present a general nonlinear structural equation model and estimation methods for it. Section 2 presents the general nonlinear structural equation model as an extension of the linear structural equation model. Section 2 also describes special cases of
the nonlinear structural equation model including those that have been considered in the literature. Section 3 presents an estimation method for the general model which does not make strong distributional assumptions about the latent variables and can be implemented using a pseudo likelihood approach combined with the Monte Carlo Expectation Maximization algorithm (MCEM). Section 4 presents an example motivated by an investigation of cystic fibrosis patients where treatment adherence is examined in relation to social, familial and personal factors, and a nonlinear structural equation model is specified and the estimation method described herein is used. Section 5 provides some discussion.

2 General nonlinear structural equation model

2.1 Linear structural equation models

The traditional linear structural equation model is typically made up of two parts: the measurement model describing the relationships between the observed and latent variables and the structural model describing the relationships between the latent variables. Given a vector of $p$ observed variables $Z_i$ for the $i^{th}$ individual in a sample of size $n$ and a vector of $q$ latent variables $f_i$, the linear structural equation model system can be written:

\[ Z_i = \mu + \Lambda f_i + \epsilon_i \]  \hspace{1cm} (1)
\[ b_0 + B_0 f_i = \delta_{0i} \]  \hspace{1cm} (2)

where in the measurement model, the matrices $\mu$ ($p \times 1$) and $\Lambda$ ($p \times q$) contain fixed or unknown scalars describing the linear relation between the observations $Z_i$ and the common latent factors $f_i$, and $\epsilon_i$ represents the ($p \times 1$) vector of random measurement error independent of $f_i$ such that $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \Psi$ with fixed and unknown scalars in $\Psi$; and
in the structural model, the matrices \( b_0 \) \((d \times 1)\) and \( B_0 \) \((d \times q)\) contain fixed or unknown scalars defining \( d \) different additive linear simultaneous structural equations relating the factors to one another plus the \((d \times 1)\) vector of random equation error \( \delta_{0i} \), where \( E(\delta_{0i}) = 0 \) and \( Var(\delta_{0i}) = \Delta_0 \) with fixed and unknown scalars in \( \Delta_0 \).

The simultaneous linear structural model as written in (2) is very general. For many practical research questions which can be addressed by simultaneous structural models, it is useful to model specific variables in terms of the rest of the variables, i.e. it is useful to consider some of the latent variables as endogenous and others as exogenous, where endogenous variables are those that are functions of other endogenous and exogenous variables. Let \( f_i = (\eta_i', \xi_i')' \) where \( \eta_i \) are the \( d \) endogenous latent variables and \( \xi_i \) are the \( q - d \) exogenous latent variables. Then a commonly used form for the structural model (2) becomes:

\[
\eta_i = b + B\eta_i + \Gamma\xi_i + \delta_i
\]

where it is assumed the equation errors \( \delta_i \) have \( E(\delta_i) = 0, \) \( Var(\delta_i) = \Delta \) and are independent of the \( \xi_i \) as well as independent of \( \epsilon_i \) in (1), and the matrices \( b \) \((d \times 1)\), \( B \) \((d \times d)\), \( \Gamma \) \((d \times (q-d))\), and \( \Delta \) \((d \times d)\) are fixed or unknown scalars. The structural model (3) is said to be in implicit form, implicit because it has endogenous variables on both sides of the equations, i.e. it is not “solved” for the endogenous variables. It is assumed that the diagonal of \( B \) is zero so that no element of \( \eta_i \) is a function of itself. A sufficient condition for solving (3) is that \( (I - B) \) is invertible, then (3) can be solved for the endogenous variables and written as

\[
\eta_i = b^* + \Gamma^*\xi_i + \delta_i^*
\]

where \( b^* = (I - B)^{-1}b \), \( \Gamma^* = (I - B)^{-1}\Gamma \), and \( Var(\delta_i^*) = (I - B)^{-1}\Delta(I - B)^{-1}' \). The structural model (4) is said to be in reduced form as the \( \eta_i \) now appear only on the left hand side of the equations. It is important to note the assumption that the equation errors
\( \delta_i \) were additive and independent of the \( \xi_i \) in the implicit form (3) results in the equation errors \( \delta_i^* \) in the reduced form (4) also being additive and independent of the \( \xi_i \).

Given \( p, q \) and \( d \), additional restrictions must be placed on \( \mu, \Lambda, \Psi, b_0, B_0, \) and \( \Delta_0 \) in (1)-(2) in order to make all the unknown parameters identifiable. The assumption that (2) can be written in reduced form (4) is the typical restriction placed on the structural model. Additionally, a common restriction placed on the measurement model (1) is the errors-in-variables parameterization where \( q \) of the observed variables are each fixed to be equal to one of the \( q \) different latent variables plus measurement error. For a thorough discussion of identifiability in linear structural equation models see e.g. Bollen 1989. Finally, it should be noted that there is no inherent distributional assumptions needed for \( \epsilon_i, \delta_{0i}, \) nor \( f_i \) at this point of model specification although distributional assumptions may be added eventually to perform estimation.

### 2.2 Extension to nonlinear structural models

A natural way to examine many scientific theories empirically is by measuring some variables on a sample of a population then examining several possible relationships between the variables. The two parts of the structural equation model (1)-(2) match this idea where (1) is measuring the latent variables and (2) is relating the latent variables to one another. A straightforward extension then is to assume that the way the variables are measured in (1) is reasonable, but that there may be more complicated relationships between the latent variables of interest than just linear ones. Thus the general nonlinear structural equation model we introduce retains the linear measurement model but considers nonlinear structural
relations:

\[ Z_i = \mu + \Lambda f_i + \epsilon_i \]  \hspace{1cm} (5)

\[ H_0(f_i; \beta_0) = \delta_{0i} \]  \hspace{1cm} (6)

where the general simultaneous nonlinear structural model system is described by the \((d \times 1)\) vector function \(H_0\) which is a known function of \(f_i\) with unknown parameters \(\beta_0\). The parameters, \(\mu\), \(\Lambda\), and the specification of the errors \(\epsilon_i\) and \(\delta_{0i}\) to have zero expectation and variances \(\Psi\) and \(\Delta_0\) respectively are the same as for the linear structural equation model (1)-(2) above.

Motivated by the desire (as in the linear structural model) to model systems of structural models where certain sets of variables are written as functions of other variables plus error, we consider again endogenous and exogenous \(\eta_i\) and \(\xi_i\) and introduce the following class of nonlinear structural equation model

\[ \eta_i = H(\eta_i; \xi_i; \beta) + \delta_i \]  \hspace{1cm} (7)

where \(H\) is a \((d \times 1)\) vector function with unknown parameters \(\beta\), and \(\delta_i\) is random equation error independent of \(\xi_i\) and \(\epsilon_i\) with \(E(\delta_i) = 0\) and \(Var(\delta_i) = \Delta\) such that \(\Delta\) is a \((d \times d)\) matrix of fixed or unknown scalars. Note that since \(H\) is a function of both \(\eta_i\) and \(\xi_i\) we refer to this simultaneous nonlinear model (7) as being in implicit form. It is assumed that \(H\) is such that there are no elements of \(\eta_i\) which are functions of themselves.

In order that the parameters in (7) are identifiable, it is important that the model is written in an unambiguous way. One way of doing this is to focus on models that can be written in an explicit reduced form. In the linear structural model, choosing models that had reduced form meant restricting to the subset of models in (3) that had \((I-B)\) invertible. Here in the nonlinear case, the rules for knowing when (7) can be solved explicitly for \(\eta_i\) and
thus written in reduced form are not so simple. A substantial literature from econometrics investigates the solvability of systems of nonlinear simultaneous equations (where the $\eta_i$ and $\xi$ would be considered observed), see e.g. Benkard and Berry (2005) and no general set of rules is available. Nevertheless we continue the general development of the nonlinear structural model by assuming that the model of interest can be written in reduced form and then (in the next subsection) we describe useful subclasses of implicit form models which can be written in reduced form.

The general reduced form simultaneous nonlinear structural model of (7) (when a reduced form exists) can then be written

$$\eta_i = h(\xi_i, \delta_i; \beta^*),$$

where $h$ is a $(d \times 1)$ vector function with unknown parameters $\beta^*$ and $\eta_i$, $\xi_i$, and $\delta_i$ are as in (7). Note that in general, solving a nonlinear implicit form (7) results in the equation error term $\delta_i$ entering the reduced form function $h$ nonlinearly. Thus, additive equation error independent of $\xi_i$ in the implicit form of the nonlinear structural model does not necessary result in additive error in the reduced form.

### 2.3 Subclasses of nonlinear structural models

In the following we present several classes of simultaneous nonlinear structural models each of which can be written in reduced form and hence are subclasses of the general nonlinear structural model (8).

**Linear endogenous, nonlinear exogenous:** This name describes a class of nonlinear structural models that is linear in the endogenous variables but possibly nonlinear in the
where $g$ is a vector function of $\xi_i$ with unknown parameters $\gamma$ and the matrix $(I - B)$ is invertible. The equation errors $\delta_i$ are as before in (7). Note that because of the linearity in the endogenous variables, non-recursive models are also included here. It is straightforward to see how this model can be written in reduced form by multiplying both sides by the inverse of $(I - B)$. We note that the reduced form has separable (additive) equation error.

**Nonlinear recursive:** This name describes a class of nonlinear structural models that is possibly nonlinear in both the endogenous and exogenous variables but where the system of equations is recursive (i.e. one equation can be substituted into the next), that is,

\[
\eta_1 = g_1(\xi, \beta_1) + \delta_{1i} \\
\eta_2 = g_2(\eta_1, \xi, \beta_2) + \delta_{2i} \\
\vdots \\
\eta_d = g_d(\eta_1, \ldots, \eta_{(d-1)i}, \xi, \beta_d) + \delta_{di}
\]  

where $g_1 \ldots g_d$ are nonlinear functions of the corresponding latent variables and unknown parameters $\beta_1, \ldots \beta_d$ respectively and the vector of equation errors formed by taking $(\delta_{1i}, \delta_{2i}, \ldots \delta_{di})'$ is treated as $\delta_i$ in (7). As a result of the triangular recursive form, it is straightforward to see how the model can be written in reduced form by substitution, but we note that the equation error will not necessarily be separable in the reduced form.

**Linear endogenous, additive nonlinear exogenous:** This class restricts the model (9) so that $g(\xi_i, \gamma)$ is an additive function of possibly nonlinear terms involving only $\xi_i$. This model could also be described as linear in parameters, but nonlinear in the exogenous variables.
latent variables, that is,

$$\eta_i = B\eta_i + \Gamma g(\xi_i) + \delta_i \quad (14)$$

where $\Gamma$ is a $(d \times r)$ matrix of fixed or unknown scalars and $g(\xi_i) = (g_1(\xi_i), g_2(\xi_i), \ldots, g_r(\xi_i))'$ is a $(r \times 1)$ vector function of known functions of the exogenous variables.

This class of nonlinear structural models and particularly its subsets below for the polynomial and specifically the second order model is the one almost exclusively examined in the literature up to this point. Assuming normality for $\xi_i$, Arminger and Muthén (1998) and Zhu and Lee (1999) described the Bayesian method for (14) with a linear measurement model (5) while Lee and Zhu (2002) describe the full maximum likelihood method for it. The nonlinear structural model (14) has also been examined by Lee and Zhu (2000), Lee and Song (2003a, 2003b), Song and Lee (2002), Lee and Lu (2003), and Lee et.al (2003). An estimation method for (14) not relying on distribution assumptions for $\xi_i$ was developed by Bollen (1995, 1996) using a two-stage least squares. The method uses the instrumental variable technique where instruments are formed by taking functions of the observed indicators. One difficulty of the method comes from finding an appropriate instrument. Bollen (1995) and Bollen and Paxton (1998) show that the method works for the quadratic and interaction model but for general $g(\xi_i)$ it may be impossible to find appropriate instruments.

**General Polynomial:** Further restricting (14) so that $g(\xi_i)$ is taken to be all the pure powers and all the multi-way interactions of those powers of the elements in $\xi_i$ results in the polynomial structural equation model. An estimation method for the general order polynomial structural equation model was described by Wall and Amemiya (2000, 2003). The two stage method of moments estimator produces consistent estimators for the structural model parameters for virtually any distribution of the observed indicator variables where the
linear measurement model holds. The procedure uses factor score estimates and estimates of their measurement error in a form of nonlinear errors-in-variables regression and produces closed-form method of moments type estimators as well as asymptotically correct standard errors.

**Quadratic and Interactions** Finally, if the general polynomial model is restricted to simply the second order model, we have the quadratic and/or interaction structural equation model. In particular,

\[
\eta_i = \gamma_0 + \gamma_1 \xi_i + \gamma_2 \xi_i^2 + \delta_i \quad \text{or} \\
\eta_i = \gamma_0 + \gamma_1 \xi_{1i} + \gamma_2 \xi_{2i} + \gamma_3 \xi_{1i} \xi_{2i} + \delta_i
\]

each taken as the structural model underlying its own linear measurement model are the models presented and estimated by the pioneering paper of Kenny and Judd (1984). In fact, these came to be known by some literature as the “Kenny and Judd model” and attracted much methodological discussions and alterations by a number of papers, including Hayduk (1987), Ping (1996), Jaccard and Wan (1995), Joreskog and Yang (1996, 1997), Schumacker and Marcoulides (1998), Li et al. (1998) and within growth curve modeling Li et al. (2000) and Wen et al. (2002). The method of estimation proposed by Kenny and Judd (1984) involved taking products of the observed indicators \(Z_i\) and treating these products as themselves indicators of the quadratic or interaction terms. This results in many (tedious) constraints on the model covariance matrix but nevertheless is possible to implement in existing *linear* structural equation modeling software programs (e.g. LISREL). The Kenny and Judd (1984) method relied on the normality assumption for \(\xi_i\) and was shown to produce inconsistent estimators when the observed indicators are not normally distributed (Wall and Amemiya, 2001). Building on the products of indicators method, Wall and Amemiya (2001)
developed an estimation method practical for the quadratic and interaction model that produces consistent estimators without assuming any distributional form for the underlying factors or errors. Comparisons via simulation study between several different approaches for the interaction model were examined in Marsh et al. (2004).

3 Pseudo likelihood estimation for the general nonlinear structural equation model

3.1 Motivation and setup

Estimation for the parameters in the general nonlinear structural model comprised of the linear measurement model (5) and the reduced form nonlinear structural model (8) will be the aim of this section. The error terms $e_i$ and $\delta_i$ will be assumed to be normally distributed which can often be considered reasonable in most applications. The distribution of the latent variables $\eta_i$ and $\xi_i$ on the other hand will not be specified as normal. Because the structural model can be written in reduced form where $\eta_i$ is a direct function of $\xi_i$ and $\delta_i$, only the distribution of $\xi$ remains unspecified. The aim is to develop an estimator of $\beta^*$ and $\Delta$ that work well for weakly specified distributions of $\xi_i$. The method presented here follows closely the work of Amemiya and Zhao (2001, 2002)

For individual $i$, the joint distribution of the observed data and the latent variables can be written under the nonlinear structural equation model (5)-(8)

$$P(Z_i; f_i; \theta) = P(Z_i|f_i; \theta_m)P(f_i; \theta_s)$$

$$= P(Z_i|\eta_i, \xi_i; \theta_m)P(\eta_i, \xi_i; \theta_s)$$

$$= P(Z_i|\eta_i, \xi_i; \theta_m)P(\eta_i|\xi_i; \theta_1)P(\xi_i; \theta_\xi) \quad (17)$$
where \( \theta_m \) represents the measurement model parameters, \( \theta_m = \{ \mu, \Lambda, \Psi \} \), and \( \theta_s \) represents the structural model parameters which are made up of the parameters in the nonlinear structural function (8), i.e. \( \theta_1 = \{ \beta^*, \Delta \} \) and the parameters \( \theta_\xi \) describing the distribution of \( \xi \). Note that the parameters in the three parts are all distinct.

Given a known distribution for \( P(\xi_i; \theta_\xi) \) with nice form, estimation for all the parameters given data could proceed via maximum likelihood, or with the addition of prior information proceed within a fully Bayesian setting. Treating the latent variables as missing data, the expectation maximization algorithm can be used for full maximum likelihood estimation although difficulty arises in the integration of the E-step since no closed form is available. Taking the distribution of \( P(\xi_i; \theta_\xi) \) to be normally distributed, Amemiya and Zhao (2001) performed the full maximum likelihood for the general nonlinear model using the Monte Carlo EM algorithm.

Very commonly it is assumed that the distribution of the exogenous variables, \( P(\xi_i; \theta_\xi) \) are normally distributed only for computational convenience. This restrictive assumption will be weakened in the current method by taking the hypothetically assumed distribution for \( \xi_i \) to be a multivariate normal mixture, i.e.

\[
\xi_i \sim \sum_{j=1}^{J} \pi_j N(\mu_j, \Sigma_\xi) \quad (18)
\]

where \( \mu_j \) is the \((q - d) \times 1\) mean vectors for the \( j^{th} \) component of the mixture and the covariance matrix \( \Sigma_\xi \) is assumed to be the same for all components. The normal mixture is considered as it can approximate a large class of distributions reasonably well and it is practical. That is, essential aspects of the estimation method, latent variable distribution deconvolution and Monte Carlo simulation from an estimated density, can be carried out readily using the normal mixture form.
While it would be possible to consider a full likelihood or fully Bayesian approach incorporating the finite mixture distribution for $P(\xi_i; \theta_\xi)$, the current paper presents an estimate for $\theta_1$ utilizing the pseudo maximum likelihood estimation procedure proposed by Gong and Samaniego (1981) and Parke (1986). In this approach, instead of maximizing the likelihood with respect to $\theta_1$, $\theta_m$, and $\theta_\xi$, some consistent estimators of the nuisance parameters $\theta_m$, and $\theta_\xi$ are substituted into the likelihood, and the resulting function is maximized only with respect to $\theta_1$. The pseudo-likelihood approach is computationally simpler than full likelihood while not losing the ability to consider flexible distributions for $\xi_i$.

3.2 Estimating the nuisance parameters

Estimation of the structural model parameter $\theta_1$ is of primary interest, thus $\theta_m$ and $\theta_\xi$ are considered nuisance parameters and will be estimated separately. The goal is to use estimators for these nuisance parameters which are consistent under weak distributional assumptions for the latent variables. The approach presented here is similar to that presented by Amemiya and Zhao (2002).

We start with describing our estimator $\hat{\theta}_m$ of $\theta_m$. It has been shown that the maximum normal likelihood estimators of the factor loadings and error variances in the linear factor analysis are consistent and have nice properties for nearly any unspecified distribution of the factor vector. See, e.g., Amemiya, Fuller, and Pantula (1987), Anderson and Amemiya (1988), and Browne and Shapiro (1988). Hence, we apply the maximum likelihood estimation to the linear measurement models (5) treating $f_i$ as unrestricted normal, and obtain $\hat{\theta}_m$.

Now, given estimates for $\theta_m$, we focus on estimating the parameters $\theta_\xi$ describing the distribution of $\xi_i$. To obtain an estimate of $\theta_\xi$ in the latent variable normal mixture distribution, we use a method referred to as a measurement error deconvolution. This method starts with
obtaining the so-called factor score estimator $\hat{\xi}_i$ of each $\xi_i$ based on the measurement model and its estimated parameters. The factor score estimator for $f_i$ is

$$\hat{f}_i = \left[ \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} \right]^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} [Z_i - \hat{\mu}]$$  \hspace{1cm} (19)

and $\hat{\xi}_i$ is taken as the corresponding subset of elements from $\hat{f}_i$. To use these factor score estimates for making inference about the distribution of $\xi_i$, it is necessary to have an estimate of the measurement error that exists in $\hat{f}_i$ as a measure of $f_i$. Ignoring the errors of $O_p(n^{-1/2})$ in estimation of $\theta_m$ and recalling the $\epsilon_i$ are normally distributed and independent of $f_i$, we have

$$\hat{f}_i = f_i + r_i,$$  \hspace{1cm} (20)

where $r_i \sim N(0, \Sigma_r)$, and $\Sigma_{r\xi}$ denotes the elements of $\Sigma_r$ corresponding to $Var(\hat{\xi}_i - \xi_i)$, and a consistent estimator $\hat{\Sigma}_r$ of $\Sigma_r$ is

$$\hat{\Sigma}_r = [\hat{\Lambda} \hat{\Psi}^{-1} \hat{\Lambda}']^{-1}.$$  \hspace{1cm} (21)

An alternate form of (19) and (21) are given in Wall and Amemiya (2000, 2003) for the case when $\hat{\Psi}$ is singular. Denote the elements of $\hat{\Sigma}_r$ corresponding to $\Sigma_{r\xi}$ as $\hat{\Sigma}_{r\xi}$. It follows from (20) and (18) that

$$\hat{\xi}_i \sim \sum_{j=1}^{J} \pi_j N(\mu_j, \Sigma_{\xi})$$  \hspace{1cm} (22)

where

$$\Sigma_{\xi} = \Sigma_{\xi} + \Sigma_{r\xi}$$  \hspace{1cm} (23)

and so by fitting a normal mixture to $\hat{\xi}$ via maximum likelihood following, for example, a standard EM algorithm for normal mixtures from McLachlan and Peel (2000) to obtain
\{ \hat{\pi}_j, j = 1 \ldots J \}, \{ \hat{\mu}_j, j = 1 \ldots J \}, \text{ and } \hat{\Sigma}_\xi, \text{ we can then obtain an estimate of } \Sigma_\xi \text{ by subtraction (deconvolution), i.e. } \hat{\Sigma}_\xi = \hat{\Sigma}_\xi - \hat{\Sigma}_r\xi. \text{ Note that in practice this sort of difference estimator for a covariance matrix can lead to non-negative definite estimate. When this occurs some modification similar to that given in Amemiya (1985) can be used.}

### 3.3 MCEM for the pseudo likelihood

Given consistent estimators \( \hat{\theta}_m \) and \( \hat{\theta}_\xi \) for the nuisance parameters, the pseudo maximum likelihood estimator (PMLE) for \( \theta_1 \) is obtained by maximizing the likelihood evaluated at \( \hat{\theta}_m \) and \( \hat{\theta}_\xi \) with respect to \( \theta_1 \). Since the likelihood function does not have an explicit expression, we consider performing the maximization using a Monte Carlo EM (MCEM) algorithm. The complete data pseudo likelihood is

\[
L_c = \prod_{i=1}^n P(\mathbf{Z}_i, \mathbf{f}_i; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi)
\]

Then the E-step obtains the expectation of the complete data pseudo likelihood given the observations and the current parameter estimates \( \theta_1^{(t)} \)

\[
E(\log L_c | \mathbf{Z}_1 \ldots \mathbf{Z}_n, \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi)
\]

\[
= \sum_{i=1}^n \int \log P(\mathbf{Z}_i, \mathbf{f}_i; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi) \frac{P(\mathbf{Z}_i | \mathbf{f}_i; \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi)}{\int P(\mathbf{Z}_i | \mathbf{f}_i; \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi) d \mathbf{f}_i} d \mathbf{f}_i
\]

\[
= \sum_{i=1}^n E \left( \log P(\mathbf{Z}_i, \mathbf{f}_i; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi) \frac{P(\mathbf{Z}_i | \mathbf{f}_i; \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi)}{\int P(\mathbf{Z}_i | \mathbf{f}_i; \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi) d \mathbf{f}_i} \right)
\]

\[
\equiv g_{\theta_1^{(t)}}(\theta_1; \mathbf{Z}, \hat{\theta}_m, \hat{\theta}_\xi)
\]

where the expectation is taken with respect to the random latent variables \( \mathbf{f}_i \). The Monte Carlo method can then be used to approximate this expectation. Given the current \( \theta_1^{(t)} \) along with \( \hat{\theta}_\xi \), a Monte Carlo sample \( (\mathbf{f}_i^1, \ldots, \mathbf{f}_i^M) \) is generated. The \( m^{th} \) sample for individual \( i \) is
generating as follows:

\[ \xi_i^m \sim P(\xi_i; \hat{\theta}_\xi) \]  

\[ \delta_i^m \sim N(0, \Delta^{(t)}) \]  

then take

\[ \eta_i^m = h(\xi_i^m, \delta_i^m; \beta^{(t)}) \]  

so \( f_i^m = (\eta_i^m, \xi_i^m)' \). The expectation can then be approximated as

\[ g_{\theta_1(i)}(\theta_1; Z, \hat{\theta}_m, \hat{\theta}_\xi) \approx \sum_{m=1}^{M} \frac{1}{p} \sum_{m=1}^{M} \log P(Z_i; f_i^m; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi)W_i^m \equiv g_{\theta_1(i)}^{MC}(\theta_1; Z, \hat{\theta}_m, \hat{\theta}_\xi) \]

where \( W_i^m = \frac{p(Z_i; f_i^m; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi)}{p(Z_i; f_i^m; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi)p(f_i^m; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi)} \) is a weight which can be calculated straightforwardly. The value for \( P(Z_i; f_i^m; \theta_1, \hat{\theta}_m, \hat{\theta}_\xi) \) can be calculated directly from the multivariate normal distribution. Note that when calculating

\[ P(f_i^m; \theta_1^{(t)}, \hat{\theta}_m, \hat{\theta}_\xi) = P(\eta_i^m|\xi_i^m; \beta^{(t)}, \Delta^{(t)})P(\xi_i^m; \hat{\theta}_m), \]

if the normally distributed equation error \( \delta_i \) is not additive in the reduced form for \( h \) in (8), then the probability \( P(\eta_i^m|\xi_i^m; \beta^{(t)}, \Delta^{(t)}) \) will not follow a multivariate normal distribution. In cases where the structural model has the nonlinear recursive structure described in section 2.3 with independent equation errors, then it will be possible to calculate \( P(\eta_i^m|\xi_i^m; \beta^{(t)}, \Delta^{(t)}) \) by appropriate recursive conditioning on the sequential endogenous variables using a product of conditional normal distributions. Generally, the probability can be calculated by taking the multivariate normal probability for \( \delta_i^m \) appropriately scaled by the Jacobian of \( \delta = h^{-1}(\eta) \) taken with respect to \( \eta \). That is

\[ P(\eta_i^m|\xi_i^m; \beta^{(t)}, \Delta^{(t)}) = P(\delta_i^m; \Delta^{(t)}) \left| \frac{\partial \delta}{\partial \eta} \right|_{\eta_i^m}. \]

The M-step is to maximize \( g_{\theta_1(i)}^{MC}(\theta_1; Z, \hat{\theta}_m, \hat{\theta}_\xi) \) w.r.t \( \theta_1 \). From (17), we note that \( \theta_1 \) only appears in the term in \( P(\eta_i|\xi_i; \theta_1) \), hence to maximize \( g_{\theta_1(i)}^{MC}(\theta_1; Z, \hat{\theta}_m, \hat{\theta}_\xi) \) we only need to
maximize

\[ G_{\theta_1}(\theta_1; Z, \hat{\theta}_m, \hat{\theta}_\xi) = \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} [\log P(\eta_i^m|\xi_i^m; \theta_1)] W_i^m \]

Given the general nonlinear structural model \( \eta_i = h(\xi_i, \delta_i; \beta^*) \) and given the equation errors \( \delta_i \) are multivariate normal with \( \text{Var}\delta_i = \Delta \), the maximization to obtain \( \theta_1^{(t+1)} = (\beta^{*(t+1)}, \Delta^{(t+1)}) \) can be accomplished by multivariate nonlinear weighted least squares. Note that for simpler forms of the nonlinear structural model, less computationally involved methods of maximization may be possible to implement. For example, multivariate linear regression can be used when the nonlinear structural model has the linear in endogenous, additive nonlinear in exogenous form as in (14).

The MCEM algorithm will iterate between the E-step and the M-step until the parameters converge according to some criteria. It has been pointed out that it is inefficient to choose a large Monte Carlo sample size \( M \) when theta is far from the ML estimate (Wei and Tanner, 1990; Booth and Hobert, 1999) and that it is preferable to start with a small \( M \) and increase it for each iteration by some fixed number. The convergence of the EM algorithm can be monitored by plotting theta versus the iteration number.

### 3.4 Standard errors estimation

The computation of the estimated covariance matrix for the PMLE was discussed in Parke (1986). Let \( \theta_2 = (\theta_m, \theta_\xi) \) represent all the nuisance parameters. Let \( (\theta_1^0, \theta_2^0) \) be the true values for \( (\theta_1, \theta_2) \), and let the information matrix for \( (\theta_1, \theta_2) \) at \( (\theta_1^0, \theta_2^0) \) for the full likelihood be denoted by

\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix},
\]

(32)
partitioned corresponding to \((\theta_1, \theta_2)\). Parke (1986) showed that if
\[
\sqrt{n}(\hat{\theta}_2 - \theta_2^0) \xrightarrow{L} N(0, \Upsilon), \quad \text{as} \quad n \to \infty,
\]
then the PMLE \(\hat{\theta}_1\) satisfies
\[
\sqrt{n}(\hat{\theta}_1 - \theta_1^0) \xrightarrow{L} N(0, \Xi), \tag{33}
\]
where \(n\) is the sample size, and
\[
\Xi = \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} \Upsilon \Sigma_{21} \Sigma_{11}^{-1}.
\]

While an estimator for the part of \(\Upsilon\) corresponding to \(\theta_m\) may be readily available from standard software packages that fit confirmatory factor analysis models, the part of \(\Upsilon\) corresponding to \(\hat{\theta}_i\) and the covariance between the two are not readily available from canned software. Thus an estimator \(\hat{\Upsilon}\) of \(\Upsilon\) can be obtained using a nonparametric bootstrap covariance matrix to estimate \(\Upsilon\). To estimate \(\Sigma_{11}\) and \(\Sigma_{12}\) in (32), we use an approximation to the expected information matrix, as described in McLachlan and Krishnan (1997, pp. 120–122). The observed data log-likelihood is
\[
\sum_{i=1}^{n} \log P(Z_i; \theta) = \sum_{i=1}^{n} \log \int P(Z_i; f_i; \theta) P(f_i; \theta) \, df_i,
\]
and the corresponding individual score vector is
\[
s(Z_i; \theta) = \partial \log P(Z_i; \theta)/\partial \theta.
\]
We propose to use an estimator of the form
\[
\sum_{i=1}^{n} s(Z_i; \hat{\theta}) s'(Z_i; \hat{\theta}),
\]
using our estimator \(\hat{\theta}\), and to extract \(\hat{\Sigma}_{11}\) and \(\hat{\Sigma}_{12}\) parts. It can be shown that \(s(Z_i; \theta) = E\{\partial l_{ci}(\theta)/\partial \theta | Z_i; \theta\}\), where \(l_{ci}(\theta) = \log P(Z_i; f_i; \theta)\). Thus \(s(Z_i; \hat{\theta})\) can be computed using
Monte Carlo method, with \( \{(\eta^m_i, \xi^m_i) : m = 1, 2, \ldots, M\} \) and \( \{W^m_i : m = 1, 2, \ldots, M\} \) obtained in the last step of the MCEM algorithm. Then, \( \hat{\Sigma}_{11} \) and \( \hat{\Sigma}_{12} \) can be obtained in
\[
\sum_{i=1}^{n} \hat{s}(Z_i; \hat{\theta})\hat{s}'(Z_i; \hat{\theta}).
\]
Combining \( \hat{\Sigma}_{11}, \hat{\Sigma}_{12}, \) and \( \hat{\Upsilon} \), we obtain our estimate of the asymptotic covariance matrix of the PMLE \( \hat{\theta}_1 \) as
\[
n^{-1} \left[ \hat{\Sigma}_{11}^{-1} + \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Upsilon} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \right].
\]

4 Example

4.1 The Data

The data used in this section is not real life data (due to data privacy issues) but is instead computer generated data mimicking and motivated by a real life study. The data is motivated by a study of children with cistic fibrosis (CF) which was interested in examining the influences that stressors in the child’s life, self esteem, and feelings of dejection have on the child’s adherence to the treatment regimes. The data have been generated directly from the nonlinear structural equation model described below and are used only to provide an example of the kind of nonlinear structural equations that might be considered and how to apply the pseudo likelihood approach for inference. The results are not intended to represent or even reflect the results in the motivating study.

Suppose we have data collected from a self-report questionnaire asking adolescents who have cistic fibrosis about the strains and stresses they encounter and feel, their self esteem, their feelings of dejection, and their frequency of skipping (non-adhering) to their treatments. The model considered of interest is shown in Figure 1 where
• *parental/youth strain* is measured by 3 items (Z1-Z3), e.g. You get into hassles/fights with your parents. Denote this latent factor as $\xi_1$.

• *peer/youth strain* is measured by 3 items (Z4-Z6), e.g. None of your friends seem to understand what having CF is like. Denote this latent factor as $\xi_2$.

• *personal worries and strains* is measured by 5 items (Z7 - Z11), e.g. You worry about the future or You stay at home when you really don’t want to. Denote this latent factor as $\xi_3$.

• *self-esteem* is measured by 6 items (Z12-Z17), e.g. I feel that I have a number of good qualities. Denote this latent factor as $\eta_1$.

• *feelings of dejection* is measured by 2 items (Z18-Z19), e.g. You get so sick of all you have to do to take care of yourself that you just want to give up. Denote this latent factor as $\eta_2$.

• *non-adherence* is a score (Z20) created as a frequency of not adhering to a number of items including e.g. You skip doing chest physical therapy treatments. This observed variable will be treated as an observed latent variable, denoted $\eta_3$.

From Figure 1 we see that there are three exogenous and three endogenous variables of interest. The explicit relationships specified among these variables are described below. Denote $\mathbf{Z} = (Z_1, \ldots Z_{20})$ as in Figure 1. Then the nonlinear structural equations model considered is
Figure 1: Structural equation model - Unobserved variables represented by ovals or circles (circles for the errors), rectangles represent observed variables. Nonlinear relationships are not explicitly represented in the figure.

\[
\begin{align*}
Z &= \mu + \Lambda (\xi_1, \xi_2, \xi_3, \eta_1, \eta_2, \eta_3)' + \epsilon \\
\eta_1 &= \beta_{10} + \beta_{11} \xi_1 + \beta_{12} \xi_2 + \beta_{13} \xi_3 + \delta_1 \\
\eta_2 &= \beta_{20} + \beta_{21} \exp (\beta_{22} \eta_1 + \beta_{23} \xi_1 + \beta_{24} \xi_2 + \beta_{25} \xi_3) + \delta_2 \\
\eta_3 &= \beta_{30} + \beta_{31} \eta_1 + \beta_{32} \eta_2 + \beta_{33} \eta_1 \eta_2 + \delta_3 \\
\mu' &= 
\begin{pmatrix}
0 & \mu_1 & \mu_2 & 0 & \mu_3 & \mu_4 & 0 & \mu_5 & \mu_6 & \mu_7 & \mu_8 & 0 & \mu_9 & \mu_{10} & \mu_{11} & \mu_{12} & \mu_{13} & 0 & \mu_{14} & 0
\end{pmatrix} \\
\Lambda' &= 
\begin{pmatrix}
1 & \lambda_{11} & \lambda_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \lambda_{21} & \lambda_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{31} & \lambda_{32} & \lambda_{33} & \lambda_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{41} & \lambda_{42} & \lambda_{43} & \lambda_{44} & \lambda_{45} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{51} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
and the last element of $\epsilon$ is set equal to zero since the endogenous variable $\eta_3 = \text{nonadherence}$ is treated as directly observed by $Z_{20}$. The data shown in Figure 2 and that used for the rest of this example were generated from the model above as follows: $(\xi_1, \xi_2, \xi_3)' = sqrt(exp(x_1, x_2, x_3))$, where $(x_1, x_2, x_3)'$ is multivariate normal with all means zero, variances equal to 1 and $Cov(x_1, x_2) = .5$, $Cov(x_1, x_3) = .6$, $Cov(x_2, x_3) = .4$; $\epsilon_1 \ldots \epsilon_{19}, \delta_1, \delta_2, \delta_3$ are independent and normally distributed each with mean zero (recall $\epsilon_{20} = 0$); variances of $\epsilon$’s are equal to 0.25, variance of $\delta_1$ is 1, and variances of $\delta_2$ and $\delta_3$ are 0.25; all unknown values in $\mu$ are set at zero and all unknown elements of $\Lambda$ are set at one for data generation; and $\beta_{10} = 6$, $\beta_{11} = -0.5$, $\beta_{12} = -0.5$, $\beta_{13} = -0.5$; $\beta_{20} = 0$, $\beta_{21} = 1$, $\beta_{22} = -0.05$, $\beta_{23} = 0.25$, $\beta_{24} = 0.25$, $\beta_{25} = 0.25$; $\beta_{30} = 6$, $\beta_{31} = -0.7$, $\beta_{32} = 1$, $\beta_{33} = -0.125$.

Note that the exogenous variables $(\xi_1, \xi_2, \xi_3)$ are not normally distributed due to the transformation taken. One dataset with 1000 independently sampled vectors $Z$ was generated under these specifications.

### 4.2 Description of the nonlinearities

The particular nonlinearity considered between the strains, self-esteem, feelings of dejection and the unhealthy behavior of non-adhering to treatment extends the types of models usually considered for relating stress, self-esteem, and unhealthy behaviors. Many studies have considered linear relationships between stress, self-esteem and different unhealthy behaviors, e.g. related to suicide, Wilburn and Smith (2005); related to smoking, Byrne and Mazanov (2001); while some others have considered interaction effects between stress and self-esteem, e.g. Roberts and Kassel (1997), and Abel (1996). In contrast to the studies given as examples here where the measurement of the latent variables was considered to be done exactly using an observed scale treated with no measurement error, the full nonlinear structural equation
model not only considers more general nonlinear relationships but also takes into account the measurement error inherent in the latent variables through the measurement model.

A description of the theoretical reasons for considering the nonlinear structural models (36) and (37), for \( \eta_2 \), i.e. feelings of dejection and \( \eta_3 \) non-adherence, is best given by examining the behavior of the functions seen for the generated data found in Figure 2. Dejection (Parrot, 2001) is a state of sadness in particular describing a feeling of being defeated. As self-esteem decreases and as overall stress increases, feelings of dejection increase. But the nonlinearities suggest that dramatically increased feelings of dejection come at a sort of tipping point or breaking point. That is, after a certain level of stress or a certain lack of self esteem, the feelings of dejection are much higher. This can be modelled by the exponential model. The interaction term in the model for non-adherence is motivated by the fact that higher self-esteem is expected to weaken the effect that feelings of dejection have on non-adherence and likewise high levels of feelings of dejection would be expected to weaken the protective effect that self-esteem has on nonadherence. This is seen for example in the bottom left figure in Figure 2 where the for low levels of dejection (represented with circles) the increase of non-adherence as self-esteem decreases is much milder than when dejection is high (represented with plus signs).

### 4.3 Estimation

The pseudo maximum likelihood method described in section 3 will be used to fit model (34)-(37) to the generated data.

First the measurement model is fit using SAS Proc Calis and the \( \hat{\mu}, \hat{\Lambda}, \) and \( \hat{\Psi} \) are obtained. Then using equation (19) and (21), the \( \hat{\xi} \) are obtained and also \( \hat{\Sigma}_r\xi \). Then using the \textit{EMclust} function from the \textit{mclust} library (Fraley and Raftery (2002)) in the R
statistical package, several finite mixture models were considered for fitting the model (18) to \( \hat{\xi} \). Based on the BIC criterion, a mixture model with 5 components and spherical covariance matrix \( \hat{\Sigma}_\xi \) fit best. Thus the estimates for \( \{\hat{\pi}_j, j = 1 \ldots 5\} \), \( \{\hat{\mu}_j, j = 1 \ldots 5\} \), and \( \hat{\Sigma}_\xi \), are obtained and \( \hat{\Sigma}_\xi \) can be obtained by subtracting \( \hat{\Sigma}_r \). Appendix A presents all the estimated nuisance parameters. Figures 3 and 4 compare the marginal and bivariate distributions of the true underlying non normally distributed exogenous latent variables with the the estimated distributions fitted with the mixture model. Marginally (Figure 3) we see that the mixture model with 5 components appears to adequately capture the positive skew. Bivariately, the major fanning feature between the true latent variables is captured by the mixture model via the two distant components in opposite directions from the rest. It is possible that a mixture model allowing different volumes (i.e. different \( \Sigma_\xi \)) for each of the 5 components would capture the distribution more closely, although the much simpler model fit here with common \( \Sigma_\xi \) appears reasonable.

Given the estimates for the nuisance parameters (shown in Appendix A), we can proceed with the MCEM for doing maximum pseudo likelihood. Taking advantage of the recursive nature of the nonlinear structural model considered and the equation errors being independent we have

\[
P(\eta^m_1|\xi^m_i; \theta_1) = P(\eta^m_2|\eta^m_1, \xi^m_i; \beta_2) P(\eta^m_3|\eta^m_1, \xi^m_i; \beta_3)
\]

where each of the three components has a univariate normal distribution and is a function of separate parameters. Hence calculation of the weights in the E-step is straightforward as a product of normal densities and maximization in the M-step can be performed separately for each of the three parts using least squares. For the \( P(\eta^m_1|\xi^m_i; \beta_1) \) and \( P(\eta^m_3|\eta^m_2, \eta^m_1; \beta_3) \),
ordinary linear least squares is used since the equations are linear in the parameters \( \beta_1 \) and \( \beta_3 \). For \( P(\eta_t^m|\eta_1^m, \xi_i^m; \beta_2) \) a nonlinear least squares is necessary but is easily implemented for example using the \textit{nlm} maximization function in R. A program for the MCEM pseudo likelihood approach for this model has been implementing in R and is available from the authors.

The standard errors as described in section 3.4 are also computed. The bootstrap method was used to obtain \( \hat{\Upsilon} \). This involved obtaining 5000 bootstrap samples for which \( \hat{\theta}_m^{(B)} \) \( B = 1 \ldots 5000 \) were computed for the measurement model from SAS PROC CALIS, and then using the same bootstrap samples, \( \hat{\theta}_\xi^{(B)} \) was computed using the \textit{Mclust} function in R. Denoting \( \hat{\theta}_2^{(B)} = (\hat{\theta}_m^{(B)', \hat{\theta}_\xi^{(B)'})' \) and taking \( \text{mean}(\hat{\theta}_2^{(B)}) = \frac{1}{5000} \sum_{B=1}^{5000} \hat{\theta}_2^{(B)} \), then the estimator \( \hat{\Upsilon} \) is computed as \( \sum_{B=1}^{5000} (\hat{\theta}_2^{(B)} - \text{mean}(\hat{\theta}_2^{(B)}))(\hat{\theta}_2^{(B)} - \text{mean}(\hat{\theta}_2^{(B)}))' \). For obtaining standard errors of the structural model parameters \( \hat{\theta}_1 \), the estimator \( \hat{\Upsilon} \) is combined with results from the MCEM. In particular, the estimators \( \hat{s}(Z_i; \hat{\theta}) \) are obtained from the last step of the MCEM and the estimated asymptotic covariance is calculated for \( \hat{\theta}_1 \) using (34). The standard errors are taken as the square root of the diagonal and 95% confidence intervals can be formed based on an assumption of asymptotic normality. Results are shown in Table 1. Not surprisingly since this is simulated data with a sample of size 1000 the resulting estimates are close to the true values.

5 Discussion

In this paper, the nonlinear structural equation model was described generally and an estimation procedure was presented using a flexible mixture model for the underlying exogenous factors and a pseudo likelihood approach for estimating the parameters of interest. Particular
Table 1: Estimates based on pseudo maximum likelihood for $\theta_1$

<table>
<thead>
<tr>
<th>parameter</th>
<th>truth</th>
<th>estimate</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{10}$</td>
<td>6</td>
<td>6.005</td>
<td>0.0575</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.5</td>
<td>-0.452</td>
<td>0.0488</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-0.5</td>
<td>-0.533</td>
<td>0.0431</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>-0.5</td>
<td>-0.504</td>
<td>0.0472</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>1</td>
<td>0.982</td>
<td>0.0434</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>0</td>
<td>0.016</td>
<td>0.144</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>1</td>
<td>0.940</td>
<td>0.109</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-0.05</td>
<td>-0.037</td>
<td>0.0097</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>0.25</td>
<td>0.263</td>
<td>0.0160</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>0.25</td>
<td>0.235</td>
<td>0.0176</td>
</tr>
<tr>
<td>$\beta_{25}$</td>
<td>0.25</td>
<td>0.262</td>
<td>0.0163</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.25</td>
<td>0.252</td>
<td>0.0104</td>
</tr>
<tr>
<td>$\beta_{30}$</td>
<td>6</td>
<td>6.084</td>
<td>0.131</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>-0.7</td>
<td>-0.720</td>
<td>0.0280</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>1</td>
<td>0.963</td>
<td>0.030</td>
</tr>
<tr>
<td>$\beta_{33}$</td>
<td>-0.125</td>
<td>-0.117</td>
<td>0.00968</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.25</td>
<td>0.238</td>
<td>0.0105</td>
</tr>
</tbody>
</table>

Attention was placed on distinguishing between techniques that make strong distributional assumptions for the underlying factors (i.e. normally distributed) verses those techniques not relying on (or more robust) to these assumptions. Sound statistical methods should generally aim to be practicably feasible and applicable to many problems while minimizing the number of uncheckable assumptions. That is the aim of the model and method presented in this paper.

A related modeling area to the one presented in this paper is nonlinear measurement error models, e.g. Fuller (1987), and Carrol Ruppert and Stefanski (1995). When the unknown measurements are considered random, it is then sometimes called nonlinear structural models, e.g. Patefield, M. (2002). Generally a nonlinear measurement error model considers a nonlinear relationship between latent variables (variables that cannot be observed directly) similar to those in the current paper, but a confirmatory factor analysis type model is not typically used as the measurement model. Instead, only one variable is typically used to
measure each latent variable and there is some assumptions made or external information used about the magnitude of the measurement error. Estimation methods including non-parametrics are well developed in this related field and may provide insight into methods useful for nonlinear structural equation modeling.

The nonlinear structural equation model presented in this paper could be made more general in two natural ways. One would be to allow for a measurement model with categorical observed variables. That is, a deviation from the linear measurement model. One of the possible difficulty with this is related to the desire to keep the underlying exogenous factor distribution flexible via the finite mixture model. It is not clear that the good results expected for estimating a flexible distribution for $\xi$ will work well when the measurement model is not linear. The other extension is to include observed covariates into the structural model. In fact, this can already be considered within the model presented, although it may not be immediately obvious. An observed covariate can be included directly into the structural model by taking it to be a perfect measure of its own exogenous latent variable with error equal to zero (similar to what was done for Z20 in the example). Then the observed variable can be considered as a special element in $\xi$ that is always equal to itself and never generated within the MCEM.

While general nonlinear relationships between latent variables may be natural to consider, there appears to be a lack of current theories motivating their use in the behavioral sciences (based on the authors’ own experience and reading of literature). This may be do in part to the small signal to noise ratio often expected in data collected in the social sciences which does not lend itself to thinking of any more than just linear relationships, but it may also be due to a lack (until more recently) of models and methods for fitting nonlinear relationships between latent variables. The hope is that if there are nonlinear theories out there just
waiting for a method, that the already existing methods (including that in this paper) will be discovered and become implemented.

References


Lee SY, Song XY. (2003a) Maximum likelihood estimation and model comparison of nonlinear structural equation models with continuous and polytomous variables *COMPUT STAT DATA AN*, 44, 125-142.


Song XY, Lee SY. (2002) A Bayesian approach for multigroup nonlinear factor analysis”

*Struct Equ Modeling, 9*(4), 523-553.


*JASA, 95*, 929-940.


Appendix A

> Lambda

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>1.0000 0.0000 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2,]</td>
<td>0.9431 0.0000 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3,]</td>
<td>0.9901 0.0000 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[4,]</td>
<td>0.0000 1.0000 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5,]</td>
<td>0.0000 0.9654 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[6,]</td>
<td>0.0000 0.9691 0.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[7,]</td>
<td>0.0000 0.0000 1.0000 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[8,]</td>
<td>0.0000 0.0000 1.0058 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[9,]</td>
<td>0.0000 0.0000 0.9731 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[10,]</td>
<td>0.0000 0.0000 1.0439 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[11,]</td>
<td>0.0000 0.0000 0.9804 0.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[12,]</td>
<td>0.0000 0.0000 0.0000 1.0000 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[13,]</td>
<td>0.0000 0.0000 0.0000 0.9815 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[14,]</td>
<td>0.0000 0.0000 0.0000 1.0145 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[15,]</td>
<td>0.0000 0.0000 0.0000 1.0165 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[16,]</td>
<td>0.0000 0.0000 0.0000 1.0040 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[17,]</td>
<td>0.0000 0.0000 0.0000 0.9898 0.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[18,]</td>
<td>0.0000 0.0000 0.0000 0.0000 1.0000 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[19,]</td>
<td>0.0000 0.0000 0.0000 0.0000 0.9631 0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[20,]</td>
<td>0.0000 0.0000 0.0000 0.0000 0.0000 1.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
> mu

\[
\begin{array}{ccccccccc}
[1] & 0.00000 & 0.08254 & 0.03744 & 0.00000 & 0.07588 & 0.07115 \\
[7] & 0.00000 & 0.04293 & 0.06978 & -0.01494 & 0.02725 & 0.00000 & 0.07329 & -0.07574 \\
[15] & -0.07303 & -0.00933 & 0.04260 & 0.00000 & 0.05085 & 0.00000
\end{array}
\]

> diag(Psi)

\[
\begin{array}{ccccccccc}
[1] & 0.22879 & 0.25550 & 0.26387 & 0.22457 & 0.24857 & 0.25313 & 0.24940 \\
[8] & 0.24991 & 0.26225 & 0.23446 & 0.26237 & 0.24384 & 0.23845 & 0.25896 & 0.25449 \\
[16] & 0.24356 & 0.25510 & 0.23300 & 0.28825 & 0.00000
\end{array}
\]

> output$mu, mu_1...mu_5 for mixture model using 5 components

\[
\begin{array}{ccccccccc}
1 & 2 & 3 & 4 & 5 \\
[1,] 1.450040 & 2.158594 & 3.445283 & 0.7947789 & 1.891052 \\
[2,] 1.321064 & 1.770537 & 1.559521 & 0.8865848 & 2.919759 \\
[3,] 1.431786 & 3.415500 & 2.113866 & 0.8422715 & 1.368160
\end{array}
\]

> output$pro, pi_1...pi_5 for mixture model

\[
\begin{array}{ccccccccc}
[1] & 0.27121709 & 0.01595621 & 0.02719309 & 0.64279488 & 0.04283873
\end{array}
\]

> output$sigma, Sigma_ksihat for mixture model for ksihat

\[
\begin{array}{ccccccc}
[1,] & 0.2304556 & 0.0000000 & 0.0000000 \\
[2,] & 0.0000000 & 0.2304556 & 0.0000000 \\
[3,] & 0.0000000 & 0.0000000 & 0.2304556
\end{array}
\]

34
> Sigrksihat, from equation (21)

\[
\begin{bmatrix}
[1,] 0.08645236 & 0.00000000 & 0.00000000 \\
[2,] 0.00000000 & 0.08394503 & 0.00000000 \\
[3,] 0.00000000 & 0.00000000 & 0.05005088 \\
\end{bmatrix}
\]

> Sigma_ksi = Sigma_ksihat - Sigrksihat

\[
\begin{bmatrix}
[1,] 0.1440033 & 0.0000000 & 0.0000000 \\
[2,] 0.0000000 & 0.1465106 & 0.0000000 \\
[3,] 0.0000000 & 0.0000000 & 0.1804047 \\
\end{bmatrix}
\]
Figure 2: Relationships between the true generated latent variables. In upper right figure, "combinestress" is the true sum of the three strain variables. To aid in viewing the interaction effect, plus signs and circles in bottom two figures represent those observations with high or low (respectively) values of the other nonplotted variable predicting nonadherence, i.e. dejection on the left and self-esteem on the right.
Figure 3: Marginal distribution of true underlying exogenous latent variables and respective estimated distributions for them using a normal mixture with 5 components

Figure 4: Bivariate distribution of true underlying exogenous latent variables and respective estimated distributions for them using a normal mixture with 5 components