Published in final edited form as:

Psychometrika. 2013 October ; 78(4): 624–647. doi:10.1007/s11336-013-9323-7.

A BAYESIAN MODELING APPROACH FOR GENERALIZED SEMIPARAMETRIC STRUCTURAL EQUATION MODELS

Xin-Yuan Song,

DEPARTMENT OF STATISTICS, THE CHINESE UNIVERSITY OF HONG KONG

Zhao-Hua Lu,

DEPARTMENT OF STATISTICS, THE CHINESE UNIVERSITY OF HONG KONG

Jing-Heng Cai, and

DEPARTMENT OF STATISTICS, SUN YAT-SEN UNIVERSITY

Edward Hak-Sing Ip

DEPARTMENT OF BIostatistical SCIENCES, DIVISION OF PUBLIC HEALTH SCIENCES,
WAKE FOREST UNIVERSITY HEALTH SCIENCES

Abstract

In behavioral, biomedical, and psychological studies, structural equation models (SEMs) have been widely used for assessing relationships between latent variables. Regression-type structural models based on parametric functions are often used for such purposes. In many applications, however, parametric SEMs are not adequate to capture subtle patterns in the functions over the entire range of the predictor variable. A different but equally important limitation of traditional parametric SEMs is that they are not designed to handle mixed data types—continuous, count, ordered, and unordered categorical. This paper develops a generalized semiparametric SEM that is able to handle mixed data types and to simultaneously model different functional relationships among latent variables. A structural equation of the proposed SEM is formulated using a series of unspecified smooth functions. The Bayesian P-splines approach and Markov chain Monte Carlo methods are developed to estimate the smooth functions and the unknown parameters. Moreover, we examine the relative benefits of semiparametric modeling over parametric modeling using a Bayesian model-comparison statistic, called the complete *deviance information criterion* (DIC). The performance of the developed methodology is evaluated using a simulation study. To illustrate the method, we used a data set derived from the National Longitudinal Survey of Youth.

Keywords

Bayesian P-splines; latent variables; MCMC methods; semiparametric models

1. Introduction

As an analytic tool deeply rooted in the psychological and social sciences, the structural equation model (SEM; Bollen, 1989; Lee, 2007) is one of the most widely used methods for assessing plausible causal assumptions through the modeling of relationships between latent variables. SEMs have found broad applications in various disciplines including psychology, social studies, economics, and medicine (see, for example, Bentler & Stein, 1992; Pugesek, Tomer, & von Eye, 2003; Sanchez, Budtz-Jorgensen, Ryan, & Hu, 2005, among others). Briefly, an SEM consists of two components—the measurement component is a factor-analysis model that groups correlated observed variables into the corresponding latent variables; the structural component is a set of regression equations that assess the effects of explanatory latent variables on outcome latent variables. It is possible for an SEM to have a set of outcome variables acting as regressors on another set of outcome variables.

A salient feature of the SEM is its ability to account for measurement error in the model. By constructing latent variables to represent single or multiple observed measures, the SEM explicitly models the error term, which often takes the form of a random Gaussian variable of mean zero and unknown variance. Traditional SEMs mostly focus on continuous response variables, so that such specification of measurement error tends to be adequate. However, as mixed data types—continuous, count, ordered, and unordered categorical—were increasingly used and collected in studies, the limitations of continuous-response SEM became apparent. A study in marketing of a specific product could involve the amount of money spent on the product (continuous data), the number of shopping trips per month (count data), response to customer satisfaction surveys recorded on the Likert scale (ordered categorical data), and shopping trip transportation modes (unordered categorical). It is therefore not surprising that the inclusion of mixed-mode data has consistently drawn a high level of interest in the development of SEMs (see, for example, Arminger & Küsters, 1988; Moustaki, 1996; Bartholomew & Knott, 1999; Shi & Lee, 2000; Jöreskog & Moustaki, 2001; Huber, Scaillet, & Victoria-Feser, 2009; Song, Lee, Ng, So, & Chan, 2007). Recently, Skrondal and Rabe-Hesketh (2004) introduced a general framework to accommodate latent variable models with data that follow the exponential family or other distributions (see also Moustaki & Victoria-Feser, 2006; Song & Lee, 2007; Cai, Song, Lam, & Ip, 2011, among others).

The aforementioned mixed-mode methods have been developed for parametric SEMs, in which the structural equations are specified either as linear or nonlinear, but all with known parametric functional forms. Although nonlinear SEMs represent an important extension of linear SEMs and have been found to be useful in practice (see Kenny & Judd, 1984; Schumacker & Marcoulides, 1998; Lee, 2007), parametric SEMs are still too restrictive for many applications. One limitation shared by both linear and nonlinear SEMs is that their functional forms for the relationships between variables have to be specified prior to data analysis. Generally, the functional form of a specific relationship is seldom known a priori. Furthermore, strict parametric functional forms are likely to miss subtle patterns. Customer satisfaction could rapidly increase with an initial increase in length of service hours but then tapers off with further increase in service. The effect of drinking on social interaction may increase with the first one or two drinks, plateau off, and then deteriorate, possibly

asymmetrically from the initial increasing phase after a certain amount of consumption. In many cases, it is challenging, if not impossible, to apply parametric modeling to each and every collected variable of interest for delineating the functional forms of their relationships with other variables.

With advances in statistical methods and computational technology, semi- and non-parametric modeling methods based on different smoothing techniques have become more accessible. These methods have the promise of breaking the straitjacket of requiring a priori specified parametric functions. They include, but are not limited to, smoothing splines (Green & Silverman, 1994), kernel methods with local polynomials (Fan & Gijbels, 1996), empirical likelihood (Owen, 2001), and penalized splines (Ruppert, Wand, & Carroll, 2003). In particular, the Markov chain Monte Carlo (MCMC) method has emerged as a powerful computational tool for estimating smooth functions, especially within the Bayesian paradigm (see DiMatteo, Genovese, & Kass, 2001; Panagiotelis & Smith, 2008). Our smoothing method of choice (see Section 2) is the Bayesian P-splines (Berry, Carroll, & Ruppert, 2002; Lang & Brezger, 2004; Fahrmeir & Raach, 2007), which is a Bayesian analogue of the penalized splines. The P-spline is highly flexible in fitting smooth functions by the introduction of random walk priors to the coefficients of a large number of B-spline basis functions. Song and Lu (2010) extended the Bayesian P-splines approach to a latent variable framework and developed an efficient method to model and estimate the functional relationships between latent variables. However, the method is restricted to only continuous data. Furthermore, the framework focused on estimation and did not discuss model selection and assessment.

In this paper, we extend the semiparametric SEM studied by Song and Lu (2010) to include discrete data. More importantly, the proposed integrated framework can simultaneously handle mixed data types, including continuous, ordered and unordered categorical, and count data. Yet another distinctive advantage of this modeling effort is the flexibility of fitting different functions, whose forms need not be specified a priori, to response data for individual predictors within the structural component of SEM. When used as an exploratory tool, the proposed semiparametric SEM allows users to visually examine and interpret the functional relationship between latent variables of interest. For example, based on a plot of the semiparametric function, a user can determine that the function is sufficiently smooth and linear to be captured by a simple linear equation. In addition to visual inspection, in this paper we propose strategies to (1) statistically assess whether the semiparametric approach is necessary and whether it offers meaningful benefits over its simpler parametric counterparts, and (2) compare the performance—in terms of a goodness-of-fit measure—between different semiparametric models as well as between these models and their parametric counterparts. To the best of our knowledge, no existing work has been devoted to addressing all of the aforementioned issues jointly.

Together with the modeling efforts described above, this paper also offers two technical contributions to the literature. First, to solve the problem of estimating an unknown function between latent variables, we develop an efficient MCMC estimation algorithm by introducing a “probit transformation” to the explanatory latent variables. While the solution suggested in Song and Lu (2010) does directly address the issue, the method requires both a

high level of effort in appropriately specifying the prior distribution and computationally intensive evaluations of the posterior distribution. These limitations were acknowledged and discussed by the authors. Compared with their method, the currently proposed MCMC algorithm is simpler in model-specifying effort and more efficient in the computation of the posterior. A second technical contribution of this paper is the inclusion of a model-comparison statistic—the complete *deviance information criterion* (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002)—which could be used for assessing the necessity of semiparametric models (over parametric ones). The original form of DIC is not appropriate for the proposed model framework because of the presence of the mixed data type and latent variables. We propose to use an adaptation—the complete DIC—which is based on the complete-data log-likelihood function, for the purpose of directly comparing between semiparametric and parametric SEMs.

This paper is organized as follows. Section 2 introduces a generalized semiparametric SEM, in which different link functions are used to model different types of data in the measurement equation, and an additive non-parametric structural equation is used for modeling the functional relations between latent variables. The Bayesian P-splines approach, on which the semiparametric SEM is based, is described in this section as well. Section 3 presents the Bayesian estimation procedure, as well as model-comparison strategies, both of which are based on an MCMC algorithm. In Section 4, as an illustration of the proposed approach, we provide an application to data collected from a highly visible study—the U.S. National Longitudinal Survey of Youth (NLSY). Section 5 reports the empirical performance of the proposed methodology assessed by a simulation study. Section 6 concludes the paper with a discussion.

2. A Generalized Semiparametric SEM

2.1. Model Description

We consider a $p \times 1$ random vector composed of ordered categorical, continuous, count, and unordered categorical variables. Without loss of generality, for $i = 1, \dots, n$, let $\mathbf{d}_i = (d_{i1}, \dots, d_{ip})' = (z_{i1}, \dots, z_{i,r_1}, y_{i,r_1+1}, \dots, y_{i,r_2}, v_{i,r_2+1}, \dots, v_{i,r_3}, u_{i,r_3+1}, \dots, u_{i,r_4})'$, where z_{ij} , y_{ij} , v_{ij} , and u_{ij} denote ordered categorical, continuous, count, and unordered categorical variables, respectively. To provide a unified framework for modeling different types of data, let

$\mathbf{d}_i^* = (z_{i,1}^*, \dots, z_{i,r_1}^*, y_{i,r_1+1}^*, \dots, y_{i,r_2}^*, v_{i,r_2+1}^*, \dots, v_{i,r_3}^*, \mathbf{w}_{i,r_3+1}', \dots, \mathbf{w}_{i,r_4}')'$ be the underlying vector, which is linked to the vector $\mathbf{d}_i$ as follows:

$$\begin{cases} z_{ij} = h_j(z_{ij}^*), & j=1, \dots, r_1, \\ y_{ij} = h_j(y_{ij}^*), & j=r_1+1, \dots, r_2, \\ \kappa_{ij} = h_j(v_{ij}^*), & j=r_2+1, \dots, r_3, \\ u_{ij} = h_j(\mathbf{w}_{ij}'), & j=r_3+1, \dots, r_4, \end{cases} \quad (1)$$

where $\kappa_{ij} = E(v_{ij})$, $h_j(\cdot)$'s are threshold, identity, exponential, and multinomial probit link functions, respectively, as defined below.

For ordered categorical variables z_{ij} that take integer values in $\{0, 1, \dots, H_j - 1\}$, a threshold link function $h_j(\cdot)$ is used as follows (Muthén, 1984; Lee, 2007):

$$z_{ij} = h_j(z_{ij}^*) = \sum_{l=0}^{H_j-1} l \cdot I(\alpha_{j,l} \leq z_{ij}^* < \alpha_{j,l+1}), \quad (2)$$

where $I(\cdot)$ is an indicator function, which takes the value 1 if $\alpha_{j,l} \leq z_{ij}^* < \alpha_{j,l+1}$ and 0 otherwise. Here, $\{-\infty = \alpha_{j,0} < \alpha_{j,1} < \dots < \alpha_{j,H_j-1} < \alpha_{j,H_j} = +\infty\}$ is a set of thresholds defining the H_j categories.

For continuous variables y_{ij} , $h_j(\cdot)$ is the identity link; that is,

$$y_{ij} = h_j(y_{ij}^*) = y_{ij}^*. \quad (3)$$

For count variables v_{ij} , we assume $p(v_{ij}) \stackrel{D}{=} \text{Pois}(\kappa_{ij})$, where “ $p(\cdot) \stackrel{D}{=}$ ” denotes “the distribution of $p(\cdot)$ that follows,” and $\text{Pois}(\cdot)$ indicates the Poisson distribution. Under the generalized linear model framework (McCullagh & Nelder, 1989), the log link $h_j^{-1}(\cdot)$ or the exponential link $h_j(\cdot)$ is taken as follows:

$$\log(\kappa_{ij}) = h_j^{-1}(E(v_{ij})) = v_{ij}^* \quad \text{or} \quad \kappa_{ij} = E(v_{ij}) = h_j(v_{ij}^*) = \exp(v_{ij}^*). \quad (4)$$

For unordered categorical variables u_{ij} with L_j categories, we assume that u_{ij} take values from $\{0, 1, \dots, L_j - 1\}$. For simplicity, we assume that $L_j = L$ for all j . This assumption can be relaxed without difficulty. Following Imai and van Dyk (2005) and Song et al. (2007), u_{ij} is modeled in terms of an underlying vector $\mathbf{w}_{ij} = (w_{ij,1}, \dots, w_{ij,L-1})'$ with a multinomial probit link function $h_j(\cdot)$ defined as follows:

$$u_{ij} = h_j(\mathbf{w}_{ij}) = \begin{cases} 0, & \text{if } \max(\mathbf{w}_{ij}) \leq 0, \\ l, & \text{if } \max(\mathbf{w}_{ij}) = w_{ij,l} > 0. \end{cases} \quad (5)$$

For example, for an unordered categorical variable u_{ij} with three categories, we have $\mathbf{w}_{ij} = (w_{ij,1}, w_{ij,2})'$, $u_{ij} = 0$ if both $w_{ij,1}$ and $w_{ij,2}$ are less than zero; $u_{ij} = 1$ if $w_{ij,1} > w_{ij,2}$ and $w_{ij,1} > 0$; and $u_{ij} = 2$ if $w_{ij,2} > w_{ij,1}$ and $w_{ij,2} > 0$.

The measurement equations for $\mathbf{d}_i$ are defined through Equations (1) and (6)–(9) below. For z_{ij} , y_{ij} , and v_{ij} , the measurement equations are given by

$$z_{ij}^* = \mu_j + \mathbf{A}_j' \boldsymbol{\omega}_i + \varepsilon_{ij}, \quad (6)$$

$$y_{ij}^* = \mu_j + \mathbf{A}_j' \boldsymbol{\omega}_i + \varepsilon_{ij}, \quad (7)$$

$$v_{ij}^* = \mu_j + \mathbf{A}_j' \boldsymbol{\omega}_i, \quad (8)$$

where μ_j is an intercept, $\mathbf{A}_j (q \times 1)$ is a vector of factor loadings, $\boldsymbol{\omega}_i (q \times 1)$ is a vector of latent variables, and ε_{ij} is a random error following $N[0, \psi_j]$ that is independent of $\boldsymbol{\omega}_i$. For u_{ij} , the measurement equation is given by

$$\mathbf{w}_{ij} = \boldsymbol{\mu}_j + \mathbf{1}_{L-1} \mathbf{A}_j' \boldsymbol{\omega}_i + \varepsilon_{ij}, \quad (9)$$

where $\boldsymbol{\mu}_j = (\mu_{j1}, \dots, \mu_{j,L-1})'$ is the intercept vector, $\mathbf{1}_{L-1}$ is an $(L-1) \times 1$ vector with the 1 at all elements, $\mathbf{A}_j$ is the factor-loading vector defined previously, and ε_{ij} is the random-error vector, which is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_{wj}]$ and is independent of $\boldsymbol{\omega}_i$. The probability mass function of u_{ij} is

$$p(u_{ij} = l | \boldsymbol{\omega}_i, \boldsymbol{\mu}_j, \mathbf{A}_j, \boldsymbol{\Psi}_{wj}) = \int_{\mathbf{R}_l} \phi_{L-1}(\mathbf{w}_{ij}; \boldsymbol{\mu}_j + \mathbf{1}_{L-1} \mathbf{A}_j' \boldsymbol{\omega}_i, \boldsymbol{\Psi}_{wj}) d\mathbf{w}_{ij}, \quad (10)$$

where $\phi_{L-1}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density function of the $(L-1)$ -variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and

$$\mathbf{R}_l = \begin{cases} \{\mathbf{w}_{ij} : \max(\mathbf{w}_{ij}) < 0\}, & l=0, \\ \{\mathbf{w}_{ij} : \max(\mathbf{w}_{ij}) = w_{ij,l} > 0\}, & l=1, 2, \dots, L-1. \end{cases}$$

Hence, $p(\mathbf{w}_{ij} | u_{ij} = l, \boldsymbol{\omega}_i, \boldsymbol{\mu}_j, \mathbf{A}_j, \boldsymbol{\Psi}_{wj})$ is proportional to the truncated multivariate normal

distribution with the density function $\phi_{L-1}(\mathbf{w}_{ij}; \boldsymbol{\mu}_j + \mathbf{1}_{L-1} \mathbf{A}_j' \boldsymbol{\omega}_i, \boldsymbol{\Psi}_{wj}) I(\mathbf{w}_{ij} \in \mathbf{R}_l)$, where $I(\mathbf{w}_{ij} \in \mathbf{R}_l)$ is an indicator function which takes value 1 if $\mathbf{w}_{ij} \in \mathbf{R}_l$ and 0 otherwise.

For the structural component of the proposed semiparametric SEM, we consider a partition of $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i', \boldsymbol{\xi}_i')'$, where $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{iq_1})'$ is a $q_1 \times 1$ outcome latent vector, and $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{iq_2})'$ is a $q_2 \times 1$ explanatory latent vector assumed to follow $N[\mathbf{0}, \boldsymbol{\Phi}]$, where $\boldsymbol{\Phi}$ is a covariance matrix. In order to examine the functional effects of explanatory latent variables in $\boldsymbol{\xi}_i$ and covariates in $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{im})'$ on outcome latent variables in $\boldsymbol{\eta}_i$, we propose the following non-parametric structural equation. For an arbitrary element η_{ij} in $\boldsymbol{\eta}_i$,

$$\eta_{ij} = g_{j1}(c_{i1}) + \cdots + g_{jm}(c_{im}) + f_{j1}(\xi_{i1}) + \cdots + f_{j,q_2}(\xi_{i,q_2}) + \delta_{ij}, \quad j=1, \dots, q_1, \quad (11)$$

where $g_{j1}(\cdot), \dots, g_{jm}(\cdot), f_{j1}(\cdot), \dots, f_{j,q_2}(\cdot)$ are unspecified smooth functions with continuous second-order derivatives, and δ_{ij} is a random residual distributed as $N[0, \psi_{\delta j}]$ and independent of ξ_i . For notational simplicity, we suppress the subscript j in (11) by assuming $q_1 = 1$ throughout the following subsections and in Section 3.

2.2. The Bayesian P-Splines

In the analysis of the generalized semiparametric SEM defined in the previous subsection, one important issue is the modeling of the unknown smooth functions in (11). Based on the robustness of P-splines in estimating unknown smooth functions (Lang & Brezger, 2004; Fahrmeir & Raach, 2007), we consider a Bayesian P-splines approach that can be implemented using MCMC methods. Briefly, P-splines are penalized splines using B-splines as basis functions. Following Song and Lu (2010), the smooth functions $g(\cdot)$'s and $f(\cdot)$'s in (11) can be modeled by a sum of B-splines (De Boor, 1978) based on a large number of knots in the domain of c 's or ξ 's. To understand this concept, consider the following simple example. If we suppose $\omega_i = (\eta_i, \xi_i)'$, the structural equation is defined as $\eta_i = f(\xi_i) + \delta_i$, which is a special case of (11). Utilizing the B-splines, $f(\xi_i)$ is modeled as follows:

$$f(\xi_i) = \sum_{k=1}^K \gamma_k B_k(\xi_i), \quad (12)$$

where K is the number of splines determined by the number of knots, γ_k 's are the unknown parameters, and the functions $B_k(\cdot)$ are B-splines with appropriate order. A natural choice of $B_k(\cdot)$ is the cubic B-spline, of which K ranges from 10 to 30, a range that should allow sufficient flexible functions. Note that the traditional B-spline basis functions $B_k(\cdot)$ are defined in a predetermined finite interval. However, the observations of the latent variables generated in the MCMC iterations may lie outside the fixed interval. Moreover, the unknown scales of latent variables make the determination of the ranges of latent variables and positions of the knots in advance impossible. Song and Lu (2010) proposed a solution of using natural cubic splines in place of the B-splines functions. The authors pointed out that with the use of the natural cubic splines, the right-hand side of (12) would always be well defined even if the observations of ξ_i in certain MCMC iterations exceeded the boundaries of the predetermined interval. Furthermore, they introduced an additional scale parameter for constructing flexible knots for latent variables with unknown scales. Their method, however, requires both careful specification of the prior distribution of the scale parameter and non-trivial posterior simulation from the associated complex conditional distribution (see Song & Lu, 2010). In the current paper, we consider a simpler method to solve the aforementioned difficulties through "probit transformation" applied to explanatory latent variables. Specifically, instead of using $B_k(\xi_i)$ in a B-splines approximation directly, we model $f(\xi_i)$ as follows:

$$f(\xi_i) = \sum_{k=1}^K \gamma_k B_k(\Phi^*(\xi_i)), \quad (13)$$

where $\Phi^*(\cdot)$ is the cumulative distribution of $N[0, 1]$. Equation (13) transforms the original scale of ξ_i to the scale of the cumulative probability $P(\xi \leq \xi_i)$. This transformation is applied to the positions of knots in the domain of ξ_i $(-\infty, \infty)$, so that we can restrict these positions in the finite interval $[0, 1]$ and determine them beforehand when using P-splines to approximate the non-parametric function $f(\xi_i)$. Given that the probit transformation function is monotonically increasing, the composite function of ξ_i in the right-hand side of (13) will not change the relation and interpretation between $f(\xi_i)$ and ξ_i . Once $f(\xi_i)$ is estimated, the functional effect of ξ_i on the outcome latent variable η_i can be interpreted according to the shape or change pattern of $f(\xi_i)$. The modeling method in (13) has the following advantages. (1) It forces $\Phi^*(\xi_i)$ to fall into a fixed interval $[0, 1]$, regardless of original ranges of ξ_i . (2) It is easy to determine the positions of knots in the domain of ξ_i a priori. For example, the knots can be chosen as equally spaced grids. (3) As mentioned above, Song and Lu (2010) introduced an additional scale parameter to overcome the difficulty of constructing B-spline basis on $(-\infty, \infty)$. The modeling (13) renders this additional scale parameter unnecessary. The proposed procedure therefore saves substantial efforts for prior specification and posterior simulation. In other words, the proposed computational algorithm for estimating the unknown function $f(\xi_i)$ is simpler, more efficient, and easier to use.

The problem of over-fitting may occur if too large a number of knots is used. Eilers and Marx (1996) proposed the P-spline by penalizing the coefficients of adjacent B-splines to prevent over-fitting of the B-splines approximation. More specifically, the penalized coefficient method minimizes

$$\sum_{i=1}^n \left(\eta_i - \sum_{k=1}^K \gamma_k B_k(\Phi^*(\xi_i)) \right)^2 + \beta \sum_{k=t+1}^K (\Delta^t \gamma_k)^2, \quad (14)$$

where β is a smoothness tuning parameter for controlling the amount of penalty, and $\Delta^t \gamma_k$ is defined in a recursive manner as follows: $\Delta^1 \gamma_k = \gamma_k - \gamma_{k-1}$, $\Delta^2 \gamma_k = \Delta^1 \gamma_k - \Delta^1 \gamma_{k-1} = \gamma_k - 2\gamma_{k-1} + \gamma_{k-2}$, ..., $\Delta^t \gamma_k = \Delta^{t-1} \gamma_k - \Delta^{t-1} \gamma_{k-1}$. Using matrix notation, (14) can be rewritten as

$$\sum_{i=1}^n \left(\eta_i - \sum_{k=1}^K \gamma_k B_k(\Phi^*(\xi_i)) \right)^2 + \beta \boldsymbol{\gamma}' \mathbf{M}_\gamma \boldsymbol{\gamma}, \quad (15)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$, and $\mathbf{M}_\gamma$ is the penalty matrix derived from the specified difference penalty.

Within a Bayesian framework (Lang & Brezger, 2004), the coefficient γ_k is regarded as random, and the difference penalty in (15) is replaced by its stochastic analogue, which is given by $\gamma_k = \gamma_{k-1} + e_k$, where e_k 's are independently distributed as $N[0, \tau_\gamma]$. Within this modeling framework, the amount of smoothness is controlled by the additional variance parameter τ_γ , which corresponds to the inverse of the smoothing parameter in (15). Thus, τ_γ can be treated as a new smoothing parameter. This idea can be easily extended to the more general situation defined by (11). Accordingly, the structural equation (11) is reformulated as follows:

$$\eta_i = \sum_{l=1}^m \sum_{k=1}^{K_{bl}} b_{lk} B_{lk}^c(c_{il}) + \sum_{l=1}^{q_2} \sum_{k=1}^{K_l} \gamma_{lk} B_{lk}(\Phi^*(\xi_{il})) + \delta_i. \quad (16)$$

2.3. Model Identification

The model defined in the previous sections is not identified without imposing identifiability constraints on model parameters. We discuss several model indeterminacies and provide appropriate identifiability constraints as follows.

The first model indeterminacy is due to the existence of ordered and unordered categorical variables. For ordered categorical variables, the indeterminacy is caused by the fact that the scale of the underlying continuous variables z_{ij}^* in (2) is unknown. Hence, the μ_j and ψ_j in (6) and the threshold parameters $\alpha_{j,l}$ in (2) cannot be simultaneously estimable. One

common method is to fix $\alpha_{j,1} = \Phi^{*-1}(f_{j,1}^*)$ and $\alpha_{j,H_j-1} = \Phi^{*-1}(f_{j,H_j-1}^*)$, where $\Phi^*(\cdot)$ is the standard normal distribution function, $f_{j,1}^*$ and f_{j,H_j-1}^* are the frequency of the first category, and the cumulative frequencies of categories $z_j < H_j - 1$, respectively (see Shi & Lee, 2000; Lee, 2007). For unordered categorical variables, multiplying each $w_{ji,l}$ by an arbitrary positive constant does not change the values of u_{ij} in (5). This indeterminacy can be solved by fixing $\Psi_{wj} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix of appropriate order (see Dunson, 2000; Song et al., 2007).

The second model indeterminacy is associated with the measurement equations defined by (6)–(9). For instance, for an arbitrary nonsingular matrix $\mathbf{C}$, measurement equation (7) can be rewritten as $y_{ij}^* = \mu_j + \mathbf{A}'_j \boldsymbol{\omega}_i + \varepsilon_{ij} = \mu_j + \mathbf{A}'_j \mathbf{C} \mathbf{C}^{-1} \boldsymbol{\omega}_i + \varepsilon_{ij} = \mu_j + \mathbf{A}'_j^* \boldsymbol{\omega}_i^* + \varepsilon_{ij}$, where $\mathbf{A}_j^* = \mathbf{A}_j \mathbf{C}$, and $\boldsymbol{\omega}_i^* = \mathbf{C}^{-1} \boldsymbol{\omega}_i$. To tackle this non-identifiability problem, we follow a common method in SEMs (Bollen, 1989; Lee, 2007) to fix appropriate elements in the loading matrix $\mathbf{A}' = [\mathbf{A}_1 \cdots \mathbf{A}_p]$ such that the only nonsingular matrix $\mathbf{C}$ that satisfies the imposed conditions is the identity matrix. A simple and common method is to use a $\mathbf{A}$ with the non-overlapping structure. For example, in the study of the NLSY data set in Section 4, we use the non-overlapping $\mathbf{A}$ in (23) to identify the measurement equations and to provide a clear interpretation of each latent variable. One of the elements at each column of $\mathbf{A}$ is fixed at 1 in order to introduce a scale to the corresponding latent variable.

The third model indeterminacy is associated with the structural equation defined by (11), in which each unknown function is not identified up to a constant. For instance, in the study of

the NLSY data set in Section 4, the outcome latent variable η in (24) remains unchanged if we replace $f_1(\xi_1)$ and $f_2(\xi_2)$ by $f_1^*(\xi_1)=f_1(\xi_1)+c$ and $f_2^*(\xi_2)=f_2(\xi_2)-c$, where c is an arbitrary constant. Following Panagiotelis and Smith (2008) and Song and Lu (2010), we impose the following constraint on $f_k(\cdot)$ at each MCMC iteration: for $l = 1, \dots, q_2$,

$\sum_{i=1}^n f_l(\xi_{il})=0$, that is, $\mathbf{1}_n' \mathbf{F}_l=0$, where $\mathbf{F}_l = (f_l(\xi_{1l}), \dots, f_l(\xi_{nl}))'$, and $\mathbf{1}_n$ is an $n \times 1$ vector with all elements being 1. This constraint is equivalent to $\mathbf{1}_n' \mathbf{B}_l \gamma_l=0$, where $\mathbf{B}_l = [B_{lk}(\Phi^*(\xi_{il}))]_{n \times K_l}$ (see Song & Lu, 2010). For a general structural equation (11), additional constraints $\mathbf{1}_n' \mathbf{B}_l^c \mathbf{b}_l=0$ with $\mathbf{B}_l^c = [B_{lk}^c(c_{il})]_{n \times K_{bl}}$ should be imposed on $g(\cdot)$, $l = 1, \dots, m$. After restricting the mean of each function in (11) to be zero, the additive structural equation is fully identified.

Given that general necessary and sufficient conditions to guarantee the identifiability of an SEM are difficult to find, we are mainly concerned with the sufficient identification conditions in this article. Recently, San Martín, Jara, Rolin, and Mouchart (2011) provided a rigorous discussion of the identification of IRT-type models. They pointed out that the identification of the measurement equation and the structural equation does not imply the identification of the whole model. Such a discussion can be applied to SEMs. In addition, Jara, Quintana, and San Martín (2008) pointed out that the identification of the parameters of the structural equation is a necessary condition for the identification of the parameters of the whole model provided that the prior distributions of the parameters in the structural equation and in the measurement equations are independent. This result is in general true. The theories and methodologies developed by Jara et al. (2008) and San Martín et al. (2011) may also be used to help provide a rigorous proof for the identification of the proposed model. Further research is required for this investigation.

3. Bayesian Analysis with the MCMC Methods

3.1. Prior Distributions

Let $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_{r_3}, \mu_{r_3+1}, \dots, \mu_{r_4}\}$, $\boldsymbol{\Lambda} = \{\Lambda_1, \dots, \Lambda_p\}$, $\boldsymbol{\Psi} = \{\psi_1, \dots, \psi_{r_2}\}$, and $\boldsymbol{\theta}_s$ denote the collection of unknown structural parameters in the model; i.e., $\boldsymbol{\theta}_s = \{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \psi_\delta, \boldsymbol{\Phi}\}$. Furthermore, let $\mathbf{b} = \{b_1, \dots, b_m\}$, $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_{q_2}\}$, $\boldsymbol{\tau}_b = \{\tau_{b1}, \dots, \tau_{bm}\}$, $\boldsymbol{\tau}_\gamma = \{\tau_{\gamma 1}, \dots, \tau_{\gamma q_2}\}$, and $\boldsymbol{\theta}_f$ denote the collection of parameters that are associated with the P-splines approximation; i.e., $\boldsymbol{\theta}_f = \{\mathbf{b}, \boldsymbol{\gamma}, \boldsymbol{\tau}_b, \boldsymbol{\tau}_\gamma\}$. Finally, let $\mathbf{a} = \{a_1, \dots, a_{r_1}\}$ be the unknown thresholds. In a full Bayesian analysis, one important issue is the specification of appropriate prior distributions for the unknown parameters.

We first consider the prior distributions for $\mathbf{b}$ and $\boldsymbol{\gamma}$. The stochastic analogues to the difference penalty are equivalent to assigning the Gaussian prior distributions to the coefficients of B-splines. Moreover, the identifiability constraints, $\mathbf{1}_n' \mathbf{B}_l^c \mathbf{b}_l=0$ and $\mathbf{1}_n' \mathbf{B}_l \gamma_l=0$, should be, respectively, imposed on $\mathbf{b}_l$ and γ_l in order to identify the unknown functions $g(\cdot)$'s and $f(\cdot)$'s. With these constraints, the truncated Gaussian priors for $\mathbf{b}_l$ and γ_l are specified as

$$p(\mathbf{b}_l|\tau_{bl}) = \left(\frac{1}{2\pi \tau_{bl}} \right)^{(K_{bl}^*/2)} \exp \left\{ -\frac{1}{2\tau_{bl}} \mathbf{b}_l' \mathbf{M}_{bl} \mathbf{b}_l \right\} I(\mathbf{1}_n' \mathbf{B}_l' \mathbf{b}_l = 0), \quad l=1, \dots, m, \quad (17)$$

$$p(\boldsymbol{\gamma}_l|\tau_{\gamma l}) = \left(\frac{1}{2\pi \tau_{\gamma l}} \right)^{(K_l^*/2)} \exp \left\{ -\frac{1}{2\tau_{\gamma l}} \boldsymbol{\gamma}_l' \mathbf{M}_{\gamma l} \boldsymbol{\gamma}_l \right\} I(\mathbf{1}_n' \mathbf{B}_l' \boldsymbol{\gamma}_l = 0), \quad l=1, \dots, q_2, \quad (18)$$

where $B_{lk}^{(c)}(\cdot)$, $B_{lk}(\cdot)$, $\mathbf{b}_l$, $\boldsymbol{\gamma}_l$, $\mathbf{M}_{bl}$, $\mathbf{M}_{\gamma l}$, K_{bl} , K_l , τ_{bl} and $\tau_{\gamma l}$ are similarly defined as those in the previous section, $K_{bl}^* = \text{rank}(\mathbf{M}_{bl})$, and $K_l^* = \text{rank}(\mathbf{M}_{\gamma l})$.

Second, for the smoothing parameters τ_{bl} and $\tau_{\gamma l}$ we follow the existing literature (Lang & Brezger, 2004; Song & Lu, 2010) and assign highly dispersed (but proper) inverse gamma priors given by

$$\begin{aligned} p(\tau_{bl}^{-1}) &\stackrel{D}{=} \text{Gamma}[\alpha_{b0}, \beta_{b0}], \quad \text{for } l=1, \dots, m, \\ p(\tau_{\gamma l}^{-1}) &\stackrel{D}{=} \text{Gamma}[\alpha_{\gamma 0}, \beta_{\gamma 0}], \quad \text{for } l=1, \dots, q_2, \end{aligned} \quad (19)$$

where α_{b0} , β_{b0} , $\alpha_{\gamma 0}$, and $\beta_{\gamma 0}$ are hyperparameters with preassigned values. To obtain highly dispersed priors, common choices for these hyperparameters are $\alpha_{b0} = \alpha_{\gamma 0} = 1$; β_{b0} and $\beta_{\gamma 0}$ are small. Here, we set $\alpha_{b0} = \alpha_{\gamma 0} = 1$ and $\beta_{b0} = \beta_{\gamma 0} = 0.005$ in the subsequent numerical studies.

Third, for the structural parameters that are involved in the measurement equations (6)–(9) and the structural equation (11), we consider the following conjugate priors:

$$\begin{aligned} p(\mu_j) &\stackrel{D}{=} N[\mu_{j0}, \sigma_{j0}^2], \quad \text{for } j=1, \dots, r_3, \\ p(\boldsymbol{\mu}_j) &\stackrel{D}{=} N[\boldsymbol{\mu}_{j0}, \mathbf{H}_{\boldsymbol{\mu}_{j0}}], \quad \text{for } j=r_3+1, \dots, r_4, \\ p(\boldsymbol{\Lambda}_j|\psi_j) &\stackrel{D}{=} N[\boldsymbol{\Lambda}_{j0}, \psi_j \mathbf{H}_{\boldsymbol{\Lambda}_{j0}}], \quad p(\psi_j^{-1}) \stackrel{D}{=} \text{Gamma}[\alpha_{j0}, \beta_{j0}], \quad \text{for } j=1, \dots, r_2, \\ p(\boldsymbol{\Lambda}_j) &\stackrel{D}{=} N[\boldsymbol{\Lambda}_{j0}, \mathbf{H}_{\boldsymbol{\Lambda}_{j0}}], \quad \text{for } j=r_2+1, \dots, r_4, \\ p(\psi_\delta^{-1}) &\stackrel{D}{=} \text{Gamma}[\alpha_{\delta 0}, \beta_{\delta 0}] \quad \text{and} \quad p(\boldsymbol{\Phi}^{-1}) \stackrel{D}{=} \text{Wishart}[\mathbf{R}_0, \rho_0], \end{aligned} \quad (20)$$

where μ_{j0} , σ_{j0}^2 , $\boldsymbol{\mu}_{j0}$, $\mathbf{H}_{\boldsymbol{\mu}_{j0}}$, $\boldsymbol{\Lambda}_{j0}$, $\mathbf{H}_{\boldsymbol{\Lambda}_{j0}}$, α_{j0} , β_{j0} , $\alpha_{\delta 0}$, $\beta_{\delta 0}$, $\mathbf{R}_0$, and ρ_0 are hyperparameters with preassigned values.

Finally, for the thresholds that define ordered categorical variables, we propose a non-informative prior distribution to reflect the fact that very little information is available for these parameters. Given that $\alpha_{j,1}$ and α_{j,H_j-1} are fixed beforehand for identification, we consider the prior distribution for $\alpha_{j,2} < \dots < \alpha_{j,H_j-2}$ as follows:

$$p(\boldsymbol{\alpha}_j) = p(\alpha_{j,2}, \dots, \alpha_{j,H_j-2}) \propto \text{constant}, \quad j=1, \dots, r_1.$$

This prior distribution is improper; following Cowles (1996), however, the joint conditional distribution of the thresholds and underlying continuous variables is proper, and they are sampled simultaneously in the MCMC procedure (see the Appendix).

3.2. Posterior Inference

Let $\mathbf{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}$ with $\mathbf{d}_i = \{\mathbf{z}_i, \mathbf{y}_i, \mathbf{v}_i, \mathbf{u}_i\}$, where $\mathbf{z}_i, \mathbf{y}_i, \mathbf{v}_i$ and $\mathbf{u}_i$ are the vectors of ordered categorical, continuous, count, and unordered categorical variables, respectively. Let $\boldsymbol{\theta}_* = \{\boldsymbol{\theta}_s, \boldsymbol{\theta}_f, \mathbf{a}\}$. The Bayesian estimate of $\boldsymbol{\theta}_*$ can then be obtained through the sample mean of a sufficiently large number of observations drawn from the posterior distribution $p(\boldsymbol{\theta}_*|\mathbf{D})$. However, this posterior distribution is intractable due to the existence of latent variables and discrete data. To solve the problem, consider the following procedure. First, let

$\mathbf{z}_i^* = (z_{i1}^*, \dots, z_{i,r_1}^*)'$, $\mathbf{w}_i = \{\mathbf{w}_{i,r_3}, \dots, \mathbf{w}_{i,r_4}\}$, $\mathbf{Z}^* = \{\mathbf{z}_1^*, \dots, \mathbf{z}_n^*\}$, $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$, and $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_n\}$. Augment the observed data $\mathbf{D}$ with the latent quantities $\mathbf{F} = \{\mathbf{Z}^*, \mathbf{W}, \boldsymbol{\Omega}\}$ in the posterior analysis. Draw a sequence of random observations from the joint posterior distribution $p(\boldsymbol{\theta}_*, \mathbf{F}|\mathbf{D})$ using MCMC methods, such as the Gibbs sampler (Geman & Geman, 1984) and the Metropolis-Hastings (MH) algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970). More specifically, $(\mathbf{a}, \mathbf{Z}^*)$, $\mathbf{W}$, $\boldsymbol{\Omega}$, $\boldsymbol{\theta}_s$, and $\boldsymbol{\theta}_f$ are generated from the corresponding full conditional distributions iteratively. These full conditional distributions are derived in the Appendix. Note that sampling from some of these full conditional distributions is not straightforward. For example, the full conditional distribution of $\mathbf{w}_{ij}$ is a multivariate truncated normal distribution, and the parameter spaces of $\mathbf{b}_l$ and $\boldsymbol{\gamma}_l$ are constrained and degenerate. Hence, additional efforts are required to simulate observations from such non-standard distributions (see the Appendix).

3.3. Model Comparison Using DIC

A specific question regarding model selection is whether or not an advanced (more complex) model is actually “better” than an elementary (simple) one. The issue is particularly relevant here: if it can be shown that a parametric SEM using simple linear equations can provide better fit to the observed data, then the case for using a semiparametric model would be difficult to establish. One common Bayesian model comparison statistic designed for this purpose is the Bayes factor (Kass & Raftery, 1995). Song et al. (2007) demonstrated that the path-sampling procedure proposed by Gelman and Meng (1998) is useful for computing the Bayes factor. However, defining a linked model between the proposed semiparametric SEM and conventional parametric SEMs within the path-sampling procedure is nontrivial. Moreover, as pointed out by Song et al. (2007), the path-sampling procedure for calculating the Bayes factor is highly computationally intensive. Apparently, the computational cost for the current complex model is even higher. In this paper, we propose to use an index based on the DIC (Spiegelhalter et al., 2002) for conducting model comparison. The index is designed to seek an appropriate model by finding a balance between goodness-of-fit to the data and model complexity. The model with a smallest DIC-based index is selected in a model-comparison procedure. Celeux, Forbes, Robert, and Titterton (2006) explored several

versions of DIC and pointed out that the complete DIC based on the complete-data log-likelihood function is the best among all of the versions when latent variables or missing data are included. The complete DIC is defined as follows:

$$\text{DIC} = -4E_{\theta_*, \mathbf{F}} \{ \log p(\mathbf{D}, \mathbf{F} | \theta_*) | \mathbf{D} \} + 2E_{\mathbf{F}} \{ \log p(\mathbf{D}, \mathbf{F} | E_{\theta_*} [\theta_* | \mathbf{D}, \mathbf{F}]) | \mathbf{D} \}, \quad (21)$$

where $\log p(\mathbf{D}, \mathbf{F} | \theta_*)$ is the complete-data log-likelihood function. For the proposed generalized semiparametric SEMs, $\log p(\mathbf{D}, \mathbf{F} | \theta_*)$ can be written as follows:

$$\log p(\mathbf{D}, \mathbf{F} | \theta_*) = \sum_{i=1}^n \log p(\mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i, \boldsymbol{\omega}_i | \theta_*),$$

where

$$\begin{aligned} & \log p(\mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i, \boldsymbol{\omega}_i | \theta_*) \\ &= \log p(\mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i | \boldsymbol{\omega}_i, \theta_*) + \log p(\eta_i | \boldsymbol{\xi}_i, \theta_*) + \log p(\boldsymbol{\xi}_i | \theta_*) \\ &= -\frac{1}{2} \left[\sum_{j=1}^{r_2} \log(\psi_j) + \sum_{j=1}^{r_1} \left(z_{ij}^* - \mu_j - \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right)^2 / \psi_j + \sum_{j=r_1+1}^{r_2} \left(y_{ij} - \mu_j - \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right)^2 / \psi_j \right] \\ & \quad + \sum_{j=r_2+1}^{r_3} \left[v_{ij} \left(\mu_j + \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right) - \exp \left(\mu_j + \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right) - \log(v_{ij}!) \right] \\ & \quad - \frac{1}{2} \sum_{j=r_3+1}^{r_4} \left(\mathbf{w}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}_{L-1} \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right)' \left(\mathbf{w}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}_{L-1} \boldsymbol{\Lambda}'_j \boldsymbol{\omega}_i \right) \\ & \quad - \frac{1}{2\psi_\delta} \left[\eta_i - \sum_{l=1}^m \sum_{k=1}^{K_{bl}} b_{lk} B_{lk}^c(c_{il}) - \sum_{l=1}^{q_2} \sum_{k=1}^{K_l} \gamma_{lk} B_{lk}(\Phi^*(\xi_i)) \right]^2 - \frac{1}{2} \log(\psi_\delta) \\ & \quad - \frac{1}{2} \left(\log |\boldsymbol{\Phi}| + \boldsymbol{\xi}_i' \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i \right) + \text{constant}, \quad \text{for } \mathbf{z}_i^* \in \mathbf{R}_i^*, \end{aligned} \quad (22)$$

and $\mathbf{R}_i^* = \prod_{j=1}^{r_1} [\alpha_{j, z_{ij}}, \alpha_{j, z_{ij}+1}]$. The first expectation in (21) can be approximated by the following:

$$E_{\theta_*, \mathbf{F}} \{ \log p(\mathbf{D}, \mathbf{F} | \theta_*) | \mathbf{D} \} = \int \log p(\mathbf{D}, \mathbf{F} | \theta_*) p(\mathbf{F}, \theta_* | \mathbf{D}) d\mathbf{F} d\theta_* \approx \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{D}, \mathbf{F}^{(t)} | \theta_*^{(t)}),$$

where $\{(\mathbf{F}^{(t)}, \theta_*^{(t)}): t=1, \dots, T\}$ are the MCMC samples drawn from the joint posterior distribution $p(\mathbf{F}, \theta_* | \mathbf{D})$. Since the smoothing parameters $\boldsymbol{\tau}_b$ and $\boldsymbol{\tau}_\gamma$ are only involved in the prior distributions of $\mathbf{b}_l$ and $\boldsymbol{\gamma}_l$ for controlling the smoothness of P-spline approximations, they are not included in the complete-data log-likelihood function (22) and the computation of DIC. Furthermore, if we let $\{\theta_*^{(m,t)}: m=1, \dots, M\}$ denote the observations generated from $p(\theta_* | \mathbf{D}, \mathbf{F}^{(t)})$ via the MCMC methods described above, then we have the following approximation:

$$E_{\theta_*} [\theta_* | \mathbf{D}, \mathbf{F}^{(t)}] \approx \bar{\theta}_*^{(t)} = \frac{1}{M} \sum_{m=1}^M \theta_*^{(m,t)}.$$

Consequently, the second expectation in (21) can be approximated by

$$E_{\mathbf{F}} \{ \log p(\mathbf{D}, \mathbf{F} | E_{\theta_*}[\theta_* | \mathbf{D}, \mathbf{F}]) | \mathbf{D} \} \approx \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{D}, \mathbf{F}^{(t)} | \bar{\theta}_*^{(t)}).$$

Note that the complete DIC is computed based on MCMC samples, which are by-products in the estimation procedure. As a result, the computational burden for obtaining DIC values is light. Spiegelhalter, Thomas, Best, and Lunn (2003) pointed out that if the difference in DIC is small, say less than 5, and the models make very different inferences, then just reporting the model with the lowest DIC is not enough. In such case, other Bayesian model selection statistics such as the Bayes factor (Kass & Raftery, 1995) and L_V -measure (Ibrahim, Chen, & Sinha, 2001), and goodness-of-fit assessment techniques such as the posterior predictive (PP) p -value (Meng, 1994) and residual plots may be used to help achieve a decisive conclusion.

4. The Study of the NLSY Data Set

In this section, we illustrate our proposed methodology using NLSY data (Center for Human Resource Research, 2004). The four-decade-long NLSY is one of the most comprehensive longitudinal studies of youths conducted in North America. Although the NLSY survey originally focused on labor force behavior, the content of the survey has evolved and become considerably broader in subsequent years. The NLSY is a nationally representative sample of youths who were 14–21 years old in 1979 and 29–36 years in 1994. The data set derived for this analysis included 1,660 observations with a few missing entries. For simplicity, the missing data in this analysis were imputed under a missing-at-random (MAR) assumption.

In the current study, we focus on the effects of a child's "behavior problem," "home environment," "mother's alcoholic problem," and "friendship" on his/her "academic achievement." Given that one can never directly measure these variables of primary interest, they were modeled as latent constructs that are measured by two or more observed variables. Following the existing literature, "academic achievement" was measured by the Peabody Individual Achievement Tests (PIAT) with continuous scales in the following domains: math, reading comprehension, and reading recognition (Dunn & Markwardt, 1970). "Behavior problem" was measured by the Behavior Problem Index (BPI) with an ordinal scale in the following domains: anti-social, anxious, dependent, headstrong, and hyperactive behaviors (Zill, 1985). In the BPI, higher scores represent greater levels of behavior problems. "Home environment" was measured by two continuous variables, namely cognitive stimulation and emotional support, and one unordered categorical variable coded with {0, 1, 2} representing a child who "lives with a biological father," "lives with a step father," or "neither." The probabilities of the unordered categorical variable taking 0, 1, and

2 is related to the latent score of “home environment” through (5), (9), and (10). The result obtained later shows that the estimated factor loading of “home environment” corresponding to this unordered categorical variable is negative, indicating that a larger probability of $u_{ij} = 0$ (lives with a biological father) is associated with a higher score of “home environment” (better home environment). The instruments for measuring these constructs come from a short form of the Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell & Bradley, 1984). “Mother’s alcoholic problem” was measured by two dichotomous variables, namely “whether mother is a binge drinker” and “whether mother is a heavy drinker.” Finally, “friendship” was measured by two count variables, namely the number of boyfriends and the number of girlfriends. In a preliminary analysis with a linear SEM, we find that children’s “behavior problem,” “home environment,” “mother’s alcoholic problem,” and “friendship” with others have significant effects on their “academic achievement.” Substantive studies have demonstrated that the conventional linear or quadratic structural equation may not be adequate to capture the true relationships between latent variables (see, for example, Fahrmeir & Raach, 2007; Song & Lu, 2010, among others). For this purpose, we propose a generalized semiparametric SEM, in which the observed variables are mixed continuous, count, and ordered and unordered categorical, and the structural equation is formulated using a series of unknown smooth functions of explanatory latent variables. The path diagram, which shows the relationships between the observed variables and their corresponding latent variables, as well as the interrelationships between latent variables, is presented in Figure 1. Operationally, the main goal of this analysis is to investigate how the explanatory latent variables “behavior problem (ξ_1),” “home environment (ξ_2),” “mother’s alcoholic problem (ξ_3),” and “friendship (ξ_4)” influence the outcome latent variable “academic achievement (η).” Given that the continuous variables have very different scales, they were all standardized prior to the analysis.

Based on the apparent meaning of the latent constructs and the identifiability constraints, we consider the following non-overlapping structure of the loading matrix for the corresponding latent variables:

$$A' = [A_1 \quad \cdots \quad A_{13}]$$

$$= \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{52} & \lambda_{62} & \lambda_{72} & \lambda_{82} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{10,3} & \lambda_{11,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{13,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{15,5} \end{bmatrix}. \quad (23)$$

To examine the functional effects of the explanatory latent variables ξ_1 to ξ_4 on the outcome latent variable η , the following non-parametric structural equation is proposed:

$$\eta = f_1(\xi_1) + f_2(\xi_2) + f_3(\xi_3) + f_4(\xi_4) + \delta. \quad (24)$$

A total of 25 equidistant knots were used to construct cubic P-splines for ξ s. The second-order random-walk penalty was used for the Bayesian P-splines in estimating the unknown smooth functions. With very little information about the unknown parameters in this study, we used a relatively vague prior input in (20) as follows: μ_{j0} and the elements in μ_{j0} and Λ_{j0} were zeros, and $\mathbf{H}_{\mu j0}$, $\mathbf{H}_{j0}$, and $\mathbf{R}_0$ were identity matrices with appropriate dimensions, $\alpha_{j0} = \alpha_{80} = 9$, $\beta_{j0} = \beta_{80} = 4$, and $\rho = 6$. Our preliminary tests suggested that the MCMC algorithm converged within 10,000 iterations. Based on this information, we collected 10,000 observations subsequent to discarding the first 10,000 burn-in iterations for conducting the proposed Bayesian inference. Table 1 and Figure 1 present the Bayesian estimates (Est) of the unknown structural parameters and their corresponding standard error estimates (SE). Table 1 shows that the intercept parameters μ_j , $j = 1, \dots, 10$ are not significantly different from zero. This is not surprising because the observed continuous variables y_{ij} , $j = 1, 2, 3, 9, 10$, were standardized prior to the analysis, and for the ordered categorical variables z_{ij} , $j = 4, \dots, 8$, the identifiability constraints of fixing $\alpha_{j,1}$ and α_{j,H_j-1} at the empirical standard normal quantiles, $\Phi^*(f_{j,1}^*)$ and $\Phi^*(f_{j,H_j-1}^*)$, imply that the means of the underlying continuous variables z_{ij}^* , $j = 4, \dots, 8$ are zero. Figure 2 depicts the Bayesian estimates of the unknown functions, together with the 5 % and 95 % pointwise quantiles. From this figure, we observed that—perhaps with the exception of the function between behavioral problems and academic achievement—the fitted curves were neither linear nor quadratic. This suggests that semiparametric models could be a better choice than their parametric counterparts. We further compared the performances of conventional parametric SEMs and the proposed semiparametric SEM by considering the model comparison statistics of the following competing models, which were specified by the same measurement equations (see Figure 1) but by different structural equations:

M_S : Semiparametric SEM: the structural equation is defined as (24);

M_L : Linear SEM: $\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \delta$; and

M_Q : Quadratic SEM: $\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \gamma_5\xi_1^2 + \gamma_6\xi_2^2 + \gamma_7\xi_3^2 + \gamma_8\xi_4^2 + \delta$.

The DIC values corresponding to M_S , M_L , and M_Q are 93,542, 93,613, and 94,133, respectively, suggesting that M_S is the best model and also affirming the necessity of its usage.

The above results are instructive in that they demonstrate the power of semiparametric SEM in revealing subtle patterns in the functional relationship between latent variables. They are also illustrative because they show how the semiparametric SEM can be used both as an exploratory tool for investigating functional forms and as a confirmatory tool for selecting models through statistically based criteria.

The interpretations of the obtained functional effects are described as follows. (1) “Behavior problem” has an essentially negative effect on “academic achievement.” This negative effect is mild when a child’s behavior problem is not severe, but becomes significant when his/her behavior problem gets worse. (2) “Home environment” has a positive effect on “academic achievement.” In general, a child with a better home environment is more likely to perform better in his/her academic achievement. This positive effect, however, appears to be neither

linear nor quadratic. The functional form exhibits a logistic function-like pattern over the range of the home environment score: the rate of change of the home effect is slow, fast, and then back to slow, respectively, at the low, medium, and high levels of the home environment score. Such a pattern reveals that at least a moderately good home environment is important for academic achievement. (3) “Mother’s alcoholic problem” appears to have negative impact on a child’s academic achievement. This effect is especially pronounced when the consumption level increases beyond moderation. We are, however, puzzled by the upward bounce toward the end of the alcoholism scale. One plausible explanation is that the mother was no longer taking care of the home environment for raising the child. Further investigation will be required to resolve this dilemma. (4) “Friendship” has a positive effect on “academic achievement.” The effect is not unexpected; what is surprising is that the effect appears to be nonlinear—it increases steeply at low level of the friendship scale, attains maximum at around the medium score of friendship, and then remains more or less stable after that.

To assess the sensitivity of the Bayesian results to the prior input, the above analysis was repeated with some ad hoc perturbations to the current prior input. For instance, μ_{j0} and the elements in $\boldsymbol{\mu}_{j0}$ and $\boldsymbol{\Lambda}_{j0}$ were set to 1, and $\mathbf{H}_{\mu j0}$, $\mathbf{H}_{j0}$, and $\mathbf{R}_0$ were set to values 10 times the identity matrices, and we set $\alpha_{j0} = \alpha_{g0} = 5$, $\beta_{j0} = \beta_{g0} = 6$, and $\rho = 14$. Compared with the reported results, the sensitivity analysis showed that both complete DIC values and estimates of the unknown parameters and smooth functions were similar. Results for the sensitivity analysis are not included here because of space limitation. In summary, our Bayesian results are robust to prior specification and input.

Using a single PC with an Intel Pentium Dual CPU E2180@2.00 GHz and 2.00 GB RAM, the computing time for obtaining the Bayesian estimates of parameters and the DIC value under M_S was about 40 minutes. The computer program is written in R with compiled C codes. The codes are freely available to the general public through the following link: <http://www.sta.cuhk.edu.hk/xy-song/psy/>.

5. Simulation Study

The main objective of this section is to report the empirical performance of the proposed semiparametric SEM and the Bayesian P-splines approach. Here, we present two experiments: the first experiment generated simulated data from a semiparametric model using a fixed set of parameter values and then assessed parameter recovery; the second experiment generated data from both semiparametric and parametric models that were structurally similar to the NLSY model and then compared the performance between the two types of model.

5.1. Experiment 1

Let $\mathbf{d}_i = (z_{i1}, z_{i2}, z_{i3}, y_{i4}, y_{i5}, y_{i6}, v_{i7}, v_{i8}, v_{i9}, u_{i10}, u_{i11}, u_{i12})'$ be a random vector defined by the measurement equations (6)–(9), where z_{ij} ’s are ordered categorical variables with five categories, and u_{ij} ’s are unordered categorical variables with three categories. The factor loadings are defined as follows:

$$= \begin{bmatrix} & \mathbf{A}' = [& \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{12} &] \\ 1 & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{83} & \lambda_{93} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{11,4} & \lambda_{12,4} & \end{bmatrix},$$

where the 1's and 0's are fixed parameters that identify the model. Let

$\omega_i = (\eta'_i, \xi'_i)' = (\eta_{i1}, \eta_{i2}, \xi_{i1}, \xi_{i2})'$, where ξ_i is drawn from $M[0, \Phi]$, and let c_i be a fixed covariate coming from $M[0, 1]$. For the structural component, we consider the following non-parametric equations:

$$\begin{aligned} \eta_1 &= g_1(c) + f_{11}(\xi_1) + f_{12}(\xi_2) + \delta_1, \\ \eta_2 &= g_2(c) + f_{21}(\xi_1) + f_{22}(\xi_2) + \delta_2, \end{aligned} \quad (25)$$

where $g_1(c) = (c/2)^3$, $f_{11}(\xi_1) = \sin(1.5\xi_1) - \xi_1$, $f_{12}(\xi_2) = 1.65 - \exp(\xi_2)$, $g_2(c) = c/4$, $f_{21}(\xi_1) = \cos(2\xi_1) - \xi_1$, $f_{22}(\xi_2) = -0.5 + \exp(2\xi_2)/[1 + \exp(2\xi_2)]$. The true values of the unknown parameters are: $\mu_1 = \cdots = \mu_9 = \mu_{10,1} = \cdots = \mu_{12,2} = 0.5$, $\lambda_{21} = \lambda_{31} = \cdots = \lambda_{93} = 0.8$, $\psi_1 = \cdots = \psi_6 = 0.3$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.3$, and $\{\phi_{11}, \phi_{12}, \phi_{22}\} = \{1.0, 0.8, 1.0\}$. Finally, for $j = 1, 2, 3$, $\alpha_j = \{-2.0, -0.6, 0.6, 2.0\}$, where the -2.0 and 2.0 are fixed to identify the ordered categorical variables. On the basis of the above settings, 100 data sets with a sample size $n = 800$ were generated for the analysis.

In this simulation study, the prior input in (20) was assigned as follows: μ_0 and the elements in μ_0 and Λ_0 were zeros; and $\mathbf{H}_{\mu 0}$, $\mathbf{H}_0$, and $\mathbf{R}_0$ were identity matrices with appropriate dimensions, $\alpha_0 = \alpha_{\delta 0} = 9$, $\beta_0 = \beta_{\delta 0} = 4$, and $\rho = 4$. A total of 25 equidistant knots were used to construct cubic P-splines in the domains of c_i , ξ_{i1} , and ξ_{i2} . The second-order random-walk penalty was used for the Bayesian P-splines in estimating the unknown smooth functions. Based on some pilot runs, we found that the algorithm converged within 4,000 iterations. We therefore took a burn-in phase of 4,000 iterations and used an additional 10,000 observations to obtain the Bayesian estimates of unknown parameters and their standard error estimates. The simulation results were based on 100 replications of each simulation condition. The averages for the bias (BIAS) of the Bayesian estimates and the root mean square errors (RMS) between the Bayesian estimates and the true population values of the structural parameters are reported in Table 2 (Columns M_S). The b 's and γ 's in our approach were estimated by fitting true parametric functions to $\hat{g}(\cdot)$'s and $\hat{h}(\cdot)$'s. To save space, the intercept terms in μ are not presented.

We observe that the values of BIAS and RMS are reasonably small, indicating that the proposed methods are able to recover the unknown parameters. The BIAS and RMS values of some parameters associated with the unordered categorical variables, such as $\lambda_{11,4}$, $\lambda_{12,4}$ and ϕ_{22} , are relatively large compared with corresponding values from the other data types. This is consistent with the fact that unordered categorical variables generally provide less information than other data types such as continuous data. Figure 3 shows the averages of

the pointwise posterior means of non-parametric functions, together with the 5 % and 95 % pointwise quantiles. These results suggest that the estimated curves satisfactorily capture the true underlying functional relationships between latent and observed variables.

Using the same computer setting as in the real analysis, the computation time for one replication is about 30 minutes.

5.2. Experiment 2

The second experiment generated data from the following three model settings. The measurement equations were identical to those defined in Experiment 1, and the structural equations were defined, respectively, as follows:

M_S : Semiparametric structural equation (25);

M_L : Linear structural equation: $\eta_j = b_j c + \gamma_{j1} \xi_1 + \gamma_{j2} \xi_2 + \delta_j$;

M_Q : Quadratic structural equation: $\eta_j = b_j c + \gamma_{j1} \xi_1 + \gamma_{j2} \xi_2 + \gamma_{j3} \xi_1^2 + \gamma_{j4} \xi_2^2 + \delta_j$, $j = 1, 2$;

M_O : Oracle structural equation: $\eta_j = b_j g_j(c) + \gamma_{j1} f_{j1}(\xi_1) + \gamma_{j2} f_{j2}(\xi_2) + \delta_j$, $j = 1, 2$, where the exact forms of $g_j(\cdot)$, $f_{j1}(\cdot)$, and $f_{j2}(\cdot)$ are given.

Table 3 presents summaries of the DIC values corresponding to the above three model settings. We observe that the means of the DIC values follow the order: $M_O < M_S < M_Q < M_L$, and that the DIC value of M_S is always smaller than those of M_L and M_Q in each replication, suggesting that the semiparametric SEM performs better than the linear and quadratic SEMs. In addition, when the semiparametric model is compared with the benchmark “oracle” model M_O , in which the exact forms of functions $g_j(\cdot)$, $f_{j1}(\cdot)$, and $f_{j2}(\cdot)$ are given and known, we can see that the empirical performance of the semiparametric model is comparable to that of the oracle model. Table 2 presents estimates of the structural parameters of M_L , M_Q , and M_O . From the table, several observations can be made. (1) The results obtained from our proposed model are very similar to those obtained from M_O . (2) The parameter estimates that are associated with the structural equations in M_L and M_Q are biased (see the boldfaced parameter estimates in Table 2). In particular, the estimates of ψ_δ obtained from M_L and M_Q tend to contain substantial biases, indicating that the linear and quadratic SEMs are not sufficiently flexible to capture unknown functional relationships between latent and observed variables. (3) Given that model misspecifications in M_L and M_Q are only present in the structural component, the parameter estimates associated with the measurement equations in M_L and M_Q are now similar to those obtained in M_S and M_O .

6. Discussion

In this paper, we propose a generalized semiparametric SEM for handling mixed data types. The generalized semiparametric SEM can also simultaneously model the functional relationships between latent and observed variables. By introducing a probit transformation to explanatory variables, we developed an efficient MCMC algorithm to estimate the underlying functions, as well as the structural and threshold parameters. In addition, we adapt the DIC index for model comparison. The adapted version of the complete DIC can be

used to support decision making in choosing between semiparametric SEMs and conventional parametric SEMs.

We foresee the proposed semiparametric SEM to be used in both exploratory SEM and confirmatory SEM analyses (we acknowledge that in practice sometimes the distinction between the two types of analysis could be difficult to draw). As evidenced in the NLSY analysis, the functional form of relationship between latent variables can be readily explored, either visually or through a more formal procedure of statistical testing. For example, Figure 2 suggests that the relationship between academic achievement and home environment is unlikely to be linear, implying that a semiparametric model is more appropriate. Thus, as an exploratory tool, the semiparametric SEM can be used to confirm or disconfirm hypothesized forms of relationships. Perhaps more importantly, unexpected patterns in the graphs may lead to new insight and findings about subtle relationship between variables that cannot be made available using traditional approaches. Combined with the ability of the proposed model for handling mixed data type model, the visualization tool could greatly expand the power of SEM in (1) making it more flexible in accommodating broad data input type, and (2) allowing prompt assessment of the form of association between latent constructs of interest. When exploratory analysis is not the ultimate goal of a study, a researcher could “fix” the model after the exploratory phase—e.g., determine whether to include links between certain variables in the path diagram, and whether to model links between latent variables as semiparametric or parametric—and then confirmatory analysis can proceed, using the proposed MCMC-based tool for inference.

In this paper, we proposed the complete DIC as a model comparison statistic. As pointed out by Celeux et al. (2006), the complete DIC is not invariant with respect to the augmented data and it does not necessarily lead to the most suitable model. We did not encounter such problems in the use of the complete DIC in our simulations. One should not routinely apply it to substantive research without a preliminary simulation study. In complicated circumstances, other Bayesian model comparison statistics such as the Bayes factor (Kass & Raftery, 1995) and L_V -measure (Ibrahim et al., 2001), and Bayesian goodness-of-fit assessment techniques such as the posterior predictive (PP) p -value (Meng, 1994) and residual plots can also be used to help choosing a suitable model.

There are several limitations in the current research. First, while we have discussed model indeterminacies and provided appropriate identifiability constraints, we did not provide a rigorous proof of identification in the proposed SEM. Based on the methodologies developed by Jara et al. (2008) and San Martín et al. (2011), a further research on this important topic is of interest in the future. Second, the proposed model assumes the normality of latent variables and random errors. This assumption may be violated in practice. As pointed out by Song, Pan, Kwok, Vandenput, Ohlsson, and Leung (2010), for SEMs with continuous latent variables, the Bayesian estimation is robust to misspecification of the distribution of explanatory latent variables but sensitive to that of random errors (or response variables). Semiparametric modeling approaches such as Dirichlet process with the stick-breaking prior (Yang & Dunson, 2010; Yang, Dunson, & Baird, 2010; Song et al., 2010) and non-parametric transformation (Song & Lu, 2012) can be used to handle non-normal random errors or response variables. For SEMs with discrete latent variables, latent

class modeling technique (Nylund & Muthén, 2007; Pan, Song, & Ip, 2013) can be employed to handle categorical latent variables. Third, Fox and Glas (2003) considered a structural model where explanatory latent variables, defined at different hierarchical levels, were related to an outcome latent variable. An extension of the proposed model to allow latent variables defined at different hierarchical levels might be interesting for further research. Fourth, interactions between explanatory latent variables are not considered in the current study. Within the semiparametric SEM framework, an interaction term involves two-dimensional non-parametric functions of latent variables. Substantial effort will be required to solve the technical challenges in the representation, computation, and interpretation of such models. Finally, the current method only applies to homogeneous and cross-sectional data. Cai, Song, and Hser (2010) proposed parametric mixture SEMs to deal with heterogeneous data, and Lu and Song (2012) further considered a mixture varying coefficient model to analyze longitudinal heterogeneous data. Their methods can be extended to the current model framework. Hence, a dynamic mixture semiparametric SEM for analyzing multi-dimensional longitudinal heterogeneous data would be a valuable addition to the SEM toolbox. However, such longitudinal or mixture semiparametric SEMs could also engender similar technical challenges as in the analysis of interaction and will require further research initiatives.

Acknowledgments

This paper is supported by the grants GRF 446609 and GRF 404711 from the Research Grant Council of the Hong Kong Special Administration Region; NSFC 11101443 from the National Natural Science Foundation of China, and the Fundamental Research Funds for the Central Universities. The last author is supported by the NIH grant R01AG031827A and 1U01HL101066. This study is also supported by the high-performance grid computing platform of Sun Yatsen University.

References

- Arminger, G.; Küsters, U. Latent trait models with indicators of mixed measurement level. In: Langeheine, R.; Rost, J., editors. Latent trait and latent class models. New York: Plenum; 1988. p. 51-73.
- Bartholomew, DJ.; Knott, M. Kendall's library of statistics: Vol. 7. Latent variable models and factor analysis. 2. London: Arnold Publishers; 1999.
- Bentler PM, Stein JA. Structural equation models in medical research. Statistical Methods in Medical Research. 1992; 1:159–181. [PubMed: 1341656]
- Berry SM, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association. 2002; 97:160–169.
- Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.
- Cai JH, Song XY, Hser YI. A Bayesian analysis of mixture structural equation models with non-ignorable missing responses and covariates. Statistics in Medicine. 2010; 29:1861–1874. [PubMed: 20680980]
- Cai JH, Song XY, Lam KH, Ip HS. A mixture of generalized latent variable models for mixed mode and heterogeneous data. Computational Statistics & Data Analysis. 2011; 55:2889–2907.
- Caldwell, BM.; Bradley, RH. Home observation for measurement of the environment. Little Rock: University of Arkansas at Little Rock, Center for Child Development and Education; 1984.
- Celeux G, Forbes F, Robert CP, Titterton DM. Deviance information criteria for missing data models. Bayesian Analysis. 2006; 1:651–674.
- Center for Human Resource Research. The national longitudinal surveys NLSY user's guide, 1979–2004. Columbus: U.S. Department of Labor, Bureau of Labor Statistics; 2004.

- Cowles MK. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*. 1996; 6:101–111.
- De Boor, C. A practical guide to splines. New York: Springer; 1978.
- DiMatteo I, Genovese CR, Kass RE. Bayesian curve fitting with free-knot splines. *Biometrika*. 2001; 88:1055–1071.
- Dunn, LM.; Markwardt, F. Peabody individual achievement test manual. Circle Pines: American Guidance Services; 1970.
- Dunson DB. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society. Series B*. 2000; 62:355–366.
- Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 11:89–121.
- Fahrmeir L, Raach A. A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*. 2007; 72:327–346.
- Fan, J.; Gijbels, I. Local polynomial modelling and its applications. London: Chapman & Hall; 1996.
- Fox JP, Glas CAW. Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*. 2003; 68:169–191.
- Gelman A, Meng XL. Simulating normalizing constant: from importance sampling to bridge sampling to path sampling. *Statistical Science*. 1998; 13:163–185.
- Geman S, Geman D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984; 6:721–741. [PubMed: 22499653]
- Green, PJ.; Silverman, BW. Nonparametric regression and generalized linear models: a roughness penalty approach. London: Chapman & Hall; 1994.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their application. *Biometrika*. 1970; 57:97–109.
- Huber P, Scaillet O, Victoria-Feser MP. Assessing multivariate predictors of financial market movements: a latent factor framework for ordinal data. *Annals of Applied Statistics*. 2009; 3:249–271.
- Ibrahim JG, Chen MH, Sinha D. Criterion-based methods for Bayesian model assessment. *Statistica Sinica*. 2001; 11:419–443.
- Imai K, van Dyk DA. Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*. 2005; 124:311–334.
- Jara A, Quintana F, San Martín E. Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Computational Statistics & Data Analysis*. 2008; 52:5033–5045.
- Jöreskog KG, Moustaki I. Factor analysis for ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research*. 2001; 36:347–387. [PubMed: 26751181]
- Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association*. 1995; 90:773–795.
- Kenny DA, Judd CM. Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*. 1984; 96:201–210.
- Lang S, Brezger A. Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 2004; 13:183–212.
- Lee, SY. Structural equation modeling: a Bayesian approach. Chichester: Wiley; 2007.
- Lu ZH, Song XY. Finite mixture varying coefficient models for analyzing longitudinal heterogeneous data. *Statistics in Medicine*. 2012; 31:544–560. [PubMed: 22161474]
- McCullagh, P.; Nelder, JA. Generalized linear models. New York: Chapman & Hall; 1989.
- Meng XL. Posterior predictive *p*-values. *The Annals of Statistics*. 1994; 22:1142–1160.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machine. *Journal of Chemical Physics*. 1953; 21:1087–1091.
- Moustaki I. A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical & Statistical Psychology*. 1996; 49:313–334.

- Moustaki I, Victoria-Feser MP. Bounded-influence robust estimation in generalized linear latent variable models. *Journal of the American Statistical Association*. 2006; 101:644–653.
- Muthén B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984; 49:115–132.
- Nylund KL, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling A Multidisciplinary Journal*. 2007; 14:535–569.
- Owen, AB. Empirical likelihood. Boca Raton: Chapman & Hall; 2001.
- Pan JH, Song XY, Ip HS. A Bayesian analysis of generalized latent curve mixture models. *Statistics and Its Interface*. 2013 in press.
- Panagiotelis A, Smith M. Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *Journal of Econometrics*. 2008; 143:291–316.
- Pugesek, BH.; Tomer, A.; von Eye, A. Structural equation modeling applications in ecological and evolutionary biology. New York: Cambridge University Press; 2003.
- Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric regression. Cambridge: Cambridge University Press; 2003.
- San Martín E, Jara A, Rolin JM, Mouchart M. On the Bayesian nonparametric generalization of IRT-type models. *Psychometrika*. 2011; 76:385–409.
- Sanchez BN, Budtz-Jorgensen E, Ryan LM, Hu H. Structural equation models: a review with applications to environmental epidemiology. *Journal of the American Statistical Association*. 2005; 100:1443–1455.
- Schumacker, RE.; Marcoulides, GA., editors. Interaction and nonlinear effects in structural equation models. Mahwah: Lawrence Erlbaum Associates, Publishers; 1998.
- Shi JQ, Lee SY. Latent variable model with mixed continuous and polytomous data. *Journal of the Royal Statistical Society. Series B*. 2000; 62:77–87.
- Skrondal, A.; Rabe-Hesketh, S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton: Chapman & Hall/CRC; 2004.
- Song XY, Lee SY. Bayesian analysis of latent variable models with nonignorable missing outcomes from exponential family. *Statistics in Medicine*. 2007; 26:681–693. [PubMed: 16538704]
- Song XY, Lee SY, Ng MCY, So WY, Chan JCN. Bayesian analysis of structural equation models with multinomial variables and an application to type 2 diabetic nephropathy. *Statistics in Medicine*. 2007; 26:2348–2369. [PubMed: 17016863]
- Song XY, Lu ZH. Semiparametric latent variable models with Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 2010; 19:590–608.
- Song XY, Lu ZH. Semiparametric transformation models with Bayesian P-splines. *Statistics and Computing*. 2012; 22:1085–1098.
- Song XY, Pan JH, Kwok T, Vandenberg L, Ohlsson C, Leung PC. A semiparametric Bayesian approach for structural equation models. *Biometrical Journal*. 2010; 52:314–332. [PubMed: 20533410]
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society. Series B*. 2002; 64:583–639.
- Spiegelhalter, DJ.; Thomas, A.; Best, NG.; Lunn, D. WinBUGS user manual. Version 1.4. Cambridge: MRC Biostatistics Unit; 2003.
- Yang MG, Dunson DB. Bayesian semiparametric structural equation models with latent variables. *Psychometrika*. 2010; 75:675–693.
- Yang MG, Dunson DB, Baird D. Semiparametric Bayes hierarchical models with mean and variance constraints. *Computational Statistics & Data Analysis*. 2010; 54:2172–2186. [PubMed: 24363478]
- Zill, N. Behavior problem scales developed from the 1981 child health supplement to the national health interview survey. Washington: Child Trends, Inc; 1985.

Appendix: The Full Conditional Distributions

The full conditional distributions involved in the Gibbs sampler are given as follows.

1. The full conditional distribution of

$$\Omega: p(\Omega | \mathbf{D}, \mathbf{Z}^*, \mathbf{W}, \boldsymbol{\theta}_*) = \prod_{i=1}^n p(\omega_i | \mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i, \boldsymbol{\theta}_*), \text{ and}$$

$$\begin{aligned} p(\omega_i | \mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i, \boldsymbol{\theta}_*) &\propto p(\mathbf{d}_i, \mathbf{z}_i^*, \mathbf{w}_i | \omega_i, \boldsymbol{\theta}_*) p(\eta_i | \xi_i, \boldsymbol{\theta}_*) p(\xi_i | \boldsymbol{\theta}_*) \\ &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^{r_1} (z_{ij}^* - \mu_j - \mathbf{A}'_j \omega_i)^2 / \psi_j + \sum_{j=r_1+1}^{r_2} (y_{ij} - \mu_j - \mathbf{A}'_j \omega_i)^2 / \psi_j \right] \right. \\ &\quad + \sum_{j=r_2+1}^{r_3} \left[v_{ij} (\mu_j + \mathbf{A}'_j \omega_i) - \exp(\mu_j + \mathbf{A}'_j \omega_i) \right] \\ &\quad - \frac{1}{2} \sum_{j=r_3+1}^{r_4} (\mathbf{w}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}_{L-1} \mathbf{A}'_j \omega_i)' (\mathbf{w}_{ij} - \boldsymbol{\mu}_j - \mathbf{1}_{L-1} \mathbf{A}'_j \omega_i) \\ &\quad \left. - \frac{1}{2\psi_\delta} \left[\eta_i - \sum_{l=1}^m \sum_{k=1}^{K_{bl}} b_{lk} B_{lk}^c(c_{il}) - \sum_{l=1}^{q_2} \sum_{k=1}^{K_l} \gamma_{lk} B_{lk}(\Phi^*(\xi_i)) \right]^2 - \frac{1}{2} \xi_i' \boldsymbol{\Phi}^{-1} \xi_i \right\}. \end{aligned}$$

(A.1)

2. The full conditional distribution of

$$\mathbf{W}: p(\mathbf{W} | \mathbf{D}, \mathbf{Z}^*, \Omega, \boldsymbol{\theta}_*) = \prod_{i=1}^n \prod_{j=r_3+1}^{r_4} p(\mathbf{w}_{ij} | u_{ij}, \omega_i, \boldsymbol{\theta}_*), \text{ and}$$

$$p(\mathbf{w}_{ij} | u_{ij}=l, \omega_i, \boldsymbol{\theta}_*) \stackrel{D}{=} N \left[\boldsymbol{\mu}_j + \mathbf{1}_{L-1} \mathbf{A}'_j \omega_i, \mathbf{I}_{L-1} \right] \mathbf{I}(\mathbf{w}_{ij} \in \mathbf{R}_l), \quad (\text{A.2})$$

where $\mathbf{I}_{L-1}$ is an identity matrix with order $L-1$, and

$$\mathbf{R}_l = \begin{cases} \{\mathbf{w}_{ij} : \max(\mathbf{w}_{ij}) < 0\}, & \text{if } l=0, \\ \{\mathbf{w}_{ij} : \max(\mathbf{w}_{ij}) = w_{ij,l} > 0\}, & \text{if } l=1, 2, \dots, L-1. \end{cases}$$

The distribution of $p(\mathbf{w}_{ij} | u_{ij}, \omega_i, \boldsymbol{\theta}_*)$ is a truncated multivariate normal distribution. Following Song et al. (2007), we use the following partitioning of variables to simulate observations $\{w_{ij,1}, \dots, w_{ij,L-1}\}$ via Gibbs sampler. Let $\mathbf{w}_{ij,-l}$ be $\mathbf{w}_{ij}$ with $w_{ij,l}$ deleted, the distribution of $w_{ij,l}$ given $u_{ij}, \mathbf{w}_{ij,-l}, \omega_i$ and $\boldsymbol{\theta}_*$ is a univariate truncated normal distribution as follows:

$$p(w_{ij,l} | u_{ij}, \mathbf{w}_{ij,-l}, \omega_i, \boldsymbol{\theta}_*) \stackrel{D}{=} \begin{cases} N[\mu_{jl} + \mathbf{A}'_j \omega_i, 1] \mathbf{I}(w_{ij,l} \geq \max\{\mathbf{w}_{ij,-l}, 0\}), & \text{if } u_{ij}=l, \\ N[\mu_{jl} + \mathbf{A}'_j \omega_i, 1] \mathbf{I}(w_{ij,l} < \max\{\mathbf{w}_{ij,-l}, 0\}), & \text{if } u_{ij} \neq l. \end{cases}$$

(A.3)

3. The full distribution of $\mathbf{b}_l$ is

$$p(\mathbf{b}_l|\cdot) \stackrel{D}{=} N\left[\mathbf{b}_l^*, \sum_{bl}^*\right], \quad l=1, \dots, m, \quad (\text{A.4})$$

with the constraint $\mathbf{1}_n' \mathbf{B}_l^c \mathbf{b}_l = 0$, where

$$\sum_{bl}^* = (\mathbf{B}_l^{c'} \mathbf{B}_l^c / \psi_\delta + \mathbf{M}_{bl} / \tau_{bl})^{-1}, \mathbf{b}_l^* = \sum_{bl}^* \mathbf{B}_l^{c'} \boldsymbol{\eta}_c^* / \psi_\delta, \text{ and}$$

$$\boldsymbol{\eta}_c^* = (\eta_{c1}^*, \dots, \eta_{cn}^*)' \text{ with}$$

$$\eta_{ci}^* = \eta_i - \sum_{j \neq l} \sum_{k=1}^{K_{bl}} b_{jk} B_{jk}^c(c_{ij}) - \sum_{j=1}^{q_2} \sum_{k=1}^{K_l} \gamma_{jk} B_{jk}(\Phi^*(\xi_{ij})).$$

To sample an observation $\mathbf{b}_l$ from its full conditional distribution with the constraint, we can sample an observation $\mathbf{b}_l^{(\text{new})}$ from $N[\mathbf{b}_l^*, \sum_{bl}^*]$, then transform $\mathbf{b}_l^{(\text{new})}$ to $\mathbf{b}_l$ as follows:

$$\mathbf{b}_l = \mathbf{b}_l^{(\text{new})} - \sum_{bl}^* \mathbf{B}_l^{c'} \mathbf{1}_n \left(\mathbf{1}_n' \mathbf{B}_l^c \sum_{bl}^* \mathbf{B}_l^{c'} \mathbf{1}_n \right)^{-1} \mathbf{1}_n' \mathbf{B}_l^c \mathbf{b}_l^{(\text{new})}.$$

4. The full distribution of $\boldsymbol{\gamma}_l$ is

$$p(\boldsymbol{\gamma}_l|\cdot) \stackrel{D}{=} N\left[\boldsymbol{\gamma}_l^*, \sum_{\gamma l}^*\right], \quad l=1, \dots, q_2, \quad (\text{A.5})$$

with the constraint $\mathbf{1}_n' \mathbf{B}_l \boldsymbol{\gamma}_l = 0$, where

$$\sum_{\gamma l}^* = (\mathbf{B}_l' \mathbf{B}_l / \psi_\delta + \mathbf{M}_{\gamma l} / \tau_{\gamma l})^{-1}, \boldsymbol{\gamma}_l^* = \sum_{\gamma l}^* \mathbf{B}_l' \boldsymbol{\eta}^* / \psi_\delta, \text{ and } \boldsymbol{\eta}^* = (\eta_1^*, \dots, \eta_m^*)'$$

with

$$\eta_i^* = \eta_i - \sum_{j=1}^m \sum_{k=1}^{K_{bl}} b_{jk} B_{jk}^c(c_{ij}) - \sum_{j \neq l} \sum_{k=1}^{K_l} \gamma_{jk} B_{jk}(\Phi^*(\xi_{ij})).$$

Similar to $\mathbf{b}_l$, we can sample an observation $\boldsymbol{\gamma}_l^{(\text{new})}$ from $N[\boldsymbol{\gamma}_l^*, \sum_{\gamma l}^*]$, then transform $\boldsymbol{\gamma}_l^{(\text{new})}$ to $\boldsymbol{\gamma}_l$ as follows:

$$\boldsymbol{\gamma}_l = \boldsymbol{\gamma}_l^{(\text{new})} - \sum_{\gamma l}^* \mathbf{B}_l' \mathbf{1}_n \left(\mathbf{1}_n' \mathbf{B}_l \sum_{\gamma l}^* \mathbf{B}_l' \mathbf{1}_n \right)^{-1} \mathbf{1}_n' \mathbf{B}_l \boldsymbol{\gamma}_l^{(\text{new})}.$$

5. The full conditional distributions of $\boldsymbol{\tau}_b$ and $\boldsymbol{\tau}_\gamma$ are as follows:

$$\begin{aligned} p(\tau_{bl}^{-1}|\cdot) &\stackrel{D}{=} \text{Gamma} \left[\alpha_{b0} + K_{bl}^*/2, \beta_{b0} + \mathbf{b}_l' \mathbf{M}_{bl} \mathbf{b}_l / 2 \right], \quad l=1, \dots, m, \\ p(\tau_{\gamma l}^{-1}|\cdot) &\stackrel{D}{=} \text{Gamma} \left[\alpha_{\gamma 0} + K_l^*/2, \beta_{\gamma 0} + \boldsymbol{\gamma}_l' \mathbf{M}_{\gamma l} \boldsymbol{\gamma}_l / 2 \right], \quad l=1, \dots, q_2. \end{aligned} \quad (\text{A.6})$$

6)

6. The full conditional distribution of $\boldsymbol{\mu}$ is given as follows:

$$p(\mu_j|\cdot) \stackrel{D}{=} N \left[\mu_j^*, \sigma_j^* \right], \quad j=1, \dots, r_1, \quad p(\mu_j|\cdot) \stackrel{D}{=} N \left[\mu_j^{**}, \sigma_j^* \right], \quad j=r_1+1, \dots, r_2,$$

(A.7)

$$p(\mu_j|\cdot) \propto \exp \left\{ \sum_{i=1}^n \left[v_{ij} \left(\mu_j + \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i \right) - \exp \left(\mu_j + \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i \right) \right] - \frac{(\mu_j - \mu_{j0})^2}{2\sigma_{j0}} \right\}, \quad j=r_2+1, \dots, r_3,$$

(A.8)

$$p(\boldsymbol{\mu}_j|\cdot) \stackrel{D}{=} N[\tilde{\boldsymbol{\mu}}_j, \sum_{\mu_j}], \quad j=r_3+1, \dots, r_4, \quad (\text{A.9})$$

where $\sigma_j^* = (\sigma_{j0}^{-1} + n\psi_j^{-1})^{-1}$, $\mu_j^* = \sigma_j^* (n\bar{z}_j^* / \psi_j + \mu_{j0} / \sigma_{j0})$, with

$\bar{z}_j^* = n^{-1} \sum_{i=1}^n (z_{ij}^* - \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i)$, $\mu_j^{**} = \sigma_j^* (n\bar{y}_j / \psi_j + \mu_{j0} / \sigma_{j0})$ with

$\bar{y}_j = n^{-1} \sum_{i=1}^n (y_{ij} - \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i)$, $\sum_{\mu_j} = (n\mathbf{I}_{L-1} + \mathbf{H}_{\mu j 0}^{-1})^{-1}$, $\tilde{\boldsymbol{\mu}}_j = \sum_{\mu_j} (n\bar{\mathbf{w}}_j + \mathbf{H}_{\mu j 0}^{-1} \boldsymbol{\mu}_{j0})$

with $\bar{\mathbf{w}}_j = n^{-1} \sum_{i=1}^n (\mathbf{w}_{ij} - \mathbf{1}_{L-1} \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i)$.

7. The full conditional distributions of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are given as follows:

$$p(\boldsymbol{\Lambda}_j|\cdot) \stackrel{D}{=} N \left[\boldsymbol{\Lambda}_j^*, \psi_j \mathbf{H}_j^* \right], \quad p(\psi_j^{-1}|\cdot) \stackrel{D}{=} \text{Gamma} \left[n/2 + \alpha_{j0}, \beta_j^* \right], \quad j=1, \dots, r_1,$$

(A.10)

$$p(\mathbf{A}_j|\cdot) \stackrel{D}{=} N[\mathbf{A}_j^{**}, \psi_j \mathbf{H}_j^*], \quad p(\psi_j^{-1}|\cdot) \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{j0}, \beta_j^{**}], \quad j=r_1+1, \dots, r_2,$$

(A.11)

$$p(\mathbf{A}_j|\cdot) \propto \exp \left\{ \sum_{i=1}^n \left[v_{ij} (\mu_j + \mathbf{A}_j' \boldsymbol{\omega}_i) - \exp(\mu_j + \mathbf{A}_j' \boldsymbol{\omega}_i) \right] - \frac{1}{2} (\mathbf{A}_j - \mathbf{A}_{j0})' \mathbf{H}_{j0}^{-1} (\mathbf{A}_j - \mathbf{A}_{j0}) \right\}, \quad j=r_2+1, \dots, r_3,$$

(A.12)

$$p(\mathbf{A}_j|\cdot) \stackrel{D}{=} N[\mathbf{A}_j^{***}, \mathbf{H}_j^{**}], \quad j=r_3+1, \dots, r_4, \quad (\text{A.13})$$

$$\begin{aligned} \mathbf{H}_j^* &= (\mathbf{H}_{j0}^{-1} + \boldsymbol{\Omega} \boldsymbol{\Omega}')^{-1}, \mathbf{A}_j^* \\ &= \mathbf{H}_j^* [\mathbf{H}_{j0}^{-1} \mathbf{A}_{j0} \\ &\quad + \boldsymbol{\Omega} \tilde{\mathbf{Z}}_j^*], \mathbf{A}_j^{**} \\ &= \mathbf{H}_j^* [\mathbf{H}_{j0}^{-1} \mathbf{A}_{j0} \\ &\quad + \boldsymbol{\Omega} \tilde{\mathbf{Y}}_j], \beta_j^* = \beta_{j0} + 2^{-1} [\tilde{\mathbf{Z}}_j^{*'} \tilde{\mathbf{Z}}_j^* \\ &\quad - \mathbf{A}_j^{*'} \mathbf{H}_j^{*-1} \mathbf{A}_j^* \\ &\quad + \mathbf{A}_{j0}' \mathbf{H}_{j0}^{-1} \mathbf{A}_{j0}], \beta_j^{**} \\ &= \beta_{j0} + 2^{-1} [\tilde{\mathbf{Y}}_j' \tilde{\mathbf{Y}}_j \\ &\quad - \mathbf{A}_j^{**'} \mathbf{H}_j^{*-1} \mathbf{A}_j^{**} \\ &\quad + \mathbf{A}_{j0}' \mathbf{H}_{j0}^{-1} \mathbf{A}_{j0}] \end{aligned}$$

where $\tilde{\mathbf{Z}}_j^* = (z_{1j}^* - \mu_j, \dots, z_{nj}^* - \mu_j)'$

and $\tilde{\mathbf{Y}}_j = (y_{1j} - \mu_j, \dots, y_{nj} - \mu_j)'$. Finally, $\mathbf{H}_j^{**} = (\mathbf{H}_{j0}^{-1} + (L-1)\boldsymbol{\Omega} \boldsymbol{\Omega}')^{-1}$ and

$\mathbf{A}_j^{***} = \mathbf{H}_j^{**} [\mathbf{H}_{j0}^{-1} \mathbf{A}_{j0} + \boldsymbol{\Omega} \tilde{\mathbf{W}}_j]$ with

$$\tilde{\mathbf{W}}_j = (\mathbf{1}_{L-1}' (\mathbf{w}_{1j} - \boldsymbol{\mu}_j), \dots, \mathbf{1}_{L-1}' (\mathbf{w}_{nj} - \boldsymbol{\mu}_j))'.$$

8.

The full conditional distributions of $\boldsymbol{\psi}_\delta$ and $\boldsymbol{\Phi}$ are as follows:

$$p(\psi_\delta^{-1}|\cdot) \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{\delta 0}, \beta_\delta^*], \quad p(\boldsymbol{\Phi}^{-1}|\cdot) \stackrel{D}{=} \text{Wishart}[\mathbf{R}^*, n/2 + \rho_0],$$

(A.14)

where

$$\beta_{\delta}^* = \beta_{\delta 0} + 2^{-1} \left[\eta_i - \sum_{l=1}^m \sum_{k=1}^{K_{bl}} b_{lk} B_{lk}^c(c_{il}) - \sum_{l=1}^{q_2} \sum_{k=1}^{K_l} \gamma_{lk} B_{lk}(\Phi^*(\xi_i)) \right]^2,$$

and $\mathbf{R}^* = \mathbf{R}_0 + \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\xi}_i'$.

9. The full conditional distributions of $\boldsymbol{\alpha}$ and $\mathbf{Z}^*$ are given as follows. For $j =$

$1, \dots, r_1$, let $\mathbf{Z}_j^* = (z_{1j}^*, \dots, z_{n_j}^*)'$, $\mathbf{Z}_j = (z_{1j}, \dots, z_{n_j})'$ we have

$$p(\boldsymbol{\alpha}, \mathbf{Z}^* | \mathbf{D}, \mathbf{W}, \boldsymbol{\Omega}, \boldsymbol{\theta}_*) = \prod_{j=1}^{r_1} p(\boldsymbol{\alpha}_j, \mathbf{Z}_j^* | \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*), \text{ and}$$

$$p(\boldsymbol{\alpha}_j, \mathbf{Z}_j^* | \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*) = p(\boldsymbol{\alpha}_j | \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*) p(\mathbf{Z}_j^* | \boldsymbol{\alpha}_j, \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*), \text{ with}$$

$$p(\boldsymbol{\alpha}_j | \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*) \propto \prod_{i=1}^n \left\{ \Phi^* \left[\psi_j^{-1/2} \left(\alpha_{j, z_{ij}+1} - \mu_j - \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i \right) \right] - \Phi^* \left[\psi_j^{-1/2} \left(\alpha_{j, z_{ij}} - \mu_j - \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i \right) \right] \right\}.$$

(A.15)

Finally, $p(\mathbf{Z}_j^* | \boldsymbol{\alpha}_j, \mathbf{Z}_j, \boldsymbol{\Omega}, \boldsymbol{\theta}_*) = \prod_{i=1}^n p(z_{ij}^* | \boldsymbol{\alpha}_j, z_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\theta}_*)$, and

$$p(z_{ij}^* | \boldsymbol{\alpha}_j, z_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\theta}_*) \stackrel{D}{=} N \left[\mu_j + \boldsymbol{\Lambda}_j' \boldsymbol{\omega}_i, \psi_j \right] I_{[\alpha_{j, z_{ij}}, \alpha_{j, z_{ij}+1})}(z_{ij}^*). \quad (\text{A.16})$$

Figure 1.

The path diagram, together with the estimated factor loadings and their standard error estimates (*in parentheses*), of the proposed SEM in the analysis of the NLSY data set.

Figure 2.

Estimates of the unknown smooth functions in the analysis of the NLSY data set. The *solid curves* represent the pointwise mean curves, and the *dot-dashed curves* represent the 5 % and 95 % pointwise quantiles.

Figure 3.

Estimates of the unknown smooth functions in the simulation study. The *solid curves* represent the true curves, the *dashed* and *dot-dashed curves*, respectively, represent the estimated posterior means and the 5 % and 95 % pointwise quantiles on the basis of 100 replications.

Table 1

Bayesian estimates of the structural parameters in the study of the NLSY data set.

Par.	Est	SE	Par.	Est	SE
μ_1	-0.003	0.025	$\lambda_{13,4}$	1.367	0.369
μ_2	-0.010	0.024	$\lambda_{15,5}$	1.024	0.013
μ_3	-0.007	0.025	ψ_1	0.457	0.020
μ_4	0.102	0.027	ψ_2	0.317	0.018
μ_5	-0.002	0.029	ψ_3	0.272	0.017
μ_6	0.003	0.030	ψ_4	0.388	0.031
μ_7	0.004	0.030	ψ_5	0.449	0.025
μ_8	0.000	0.029	ψ_6	0.646	0.042
μ_9	-0.006	0.024	ψ_7	0.373	0.023
μ_{10}	-0.008	0.025	ψ_8	0.419	0.024
$\mu_{11,1}$	-0.599	0.040	ψ_9	0.676	0.038
$\mu_{11,2}$	-0.304	0.038	ψ_{10}	0.618	0.043
μ_{12}	-1.325	0.154	ϕ_{11}	0.485	0.045
μ_{13}	-3.244	0.494	ϕ_{21}	-0.135	0.018
μ_{14}	1.874	0.025	ϕ_{22}	0.322	0.040
μ_{15}	1.919	0.026	ϕ_{31}	0.169	0.056
λ_{21}	1.118	0.033	ϕ_{32}	-0.288	0.059
λ_{31}	1.150	0.033	ϕ_{33}	3.909	0.961
λ_{32}	1.061	0.058	ϕ_{41}	0.001	0.019
λ_{62}	0.831	0.054	ϕ_{42}	-0.048	0.019
λ_{72}	1.147	0.060	ϕ_{43}	0.087	0.071
λ_{82}	1.103	0.057	ϕ_{44}	0.900	0.038
$\lambda_{10,3}$	1.096	0.109	ψ_5	0.421	0.031
$\lambda_{11,3}$	-0.995	0.115			

Table 2

Bayesian estimates of the structural parameters in the simulation study.

Par.	M_S		M_L		M_O		M_O	
	BIAS	RMS	BIAS	RMS	BIAS	RMS	BIAS	RMS
λ_{21}	0.002	0.033	0.035	0.042	0.029	0.039	0.009	0.030
λ_{31}	0.007	0.036	0.038	0.045	0.034	0.044	0.016	0.035
λ_{52}	-0.007	0.025	-0.002	0.023	-0.003	0.024	-0.006	0.025
λ_{62}	-0.009	0.023	-0.004	0.022	-0.005	0.022	-0.008	0.022
λ_{83}	0.009	0.030	0.004	0.028	-0.005	0.029	-0.002	0.027
λ_{93}	0.011	0.033	0.006	0.029	-0.003	0.030	-0.001	0.029
$\lambda_{11,4}$	0.157	0.216	0.045	0.140	0.056	0.139	0.124	0.174
$\lambda_{12,4}$	0.142	0.202	0.034	0.140	0.039	0.145	0.108	0.164
ψ_1	0.033	0.051	0.068	0.082	0.071	0.081	0.027	0.047
ψ_2	0.031	0.047	0.067	0.078	0.061	0.072	0.027	0.044
ψ_3	0.023	0.036	0.056	0.063	0.080	0.085	0.019	0.033
ψ_4	0.001	0.021	0.006	0.024	0.017	0.028	0.002	0.022
ψ_5	0.007	0.021	0.006	0.021	0.027	0.033	0.007	0.021
ψ_6	0.005	0.019	0.005	0.020	0.019	0.027	0.004	0.018
ϕ_1	-0.044	0.085	-0.181	0.192	-0.017	0.079	0.003	0.059
ϕ_{12}	-0.106	0.141	-0.030	0.113	0.000	0.113	-0.093	0.111
ϕ_{22}	-0.203	0.295	0.181	0.350	0.026	0.268	-0.206	0.234
b_{11}	0.014	0.106	-0.223	0.242	-0.240	0.256	-0.003	0.091
b_{21}	0.014	0.124	-0.017	0.147	-0.002	0.126	-0.034	0.114
γ_{11}	-0.180	0.234	-0.351	0.413	-0.644	0.669	-0.081	0.164
γ_{12}	-0.110	0.249	-0.539	0.549	-0.315	0.363	0.150	0.210
γ_{21}	0.058	0.101	0.130	0.223	0.123	0.184	-0.017	0.043
γ_{22}	0.091	0.205	0.442	0.647	0.399	0.543	-0.041	0.159
ψ_{81}	0.037	0.059	0.547	0.562	0.308	0.334	0.070	0.082
ψ_{92}	0.003	0.032	0.300	0.308	0.174	0.188	0.028	0.039

Note: The true function of $\mathcal{G}(\cdot)$ is linear (parametric), so the estimates of μ_1 are similar under all model settings.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Summaries of the DIC values in the simulation study.

	Mean	SD
M_S	35,610	377
M_L	37,281	329
M_Q	36,604	348
M_O	34,995	304