Published in final edited form as:

Biometrics. 2012 June ; 68(2): 567–577. doi:10.1111/j.1541-0420.2012.01751.x.

Bayesian Lasso for Semiparametric Structural Equation Models

Ruixin Guo^{1,*}, Hongtu Zhu^{1,**}, Sy-Miin Chow^{2,***}, and Joseph G. Ibrahim^{1,****}

¹Department of Biostatistics, University of North Carolina at Chapel Hill, USA

²Department of Psychology, University of North Carolina at Chapel Hill, USA

Summary

There has been great interest in developing nonlinear structural equation models and associated statistical inference procedures, including estimation and model selection methods. In this paper a general semiparametric structural equation model (SSEM) is developed in which the structural equation is composed of nonparametric functions of exogenous latent variables and fixed covariates on a set of latent endogenous variables. A basis representation is used to approximate these nonparametric functions in the structural equation and the Bayesian Lasso method coupled with a Markov Chain Monte Carlo (MCMC) algorithm is used for simultaneous estimation and model selection. The proposed method is illustrated using a simulation study and data from the Affective Dynamics and Individual Differences (ADID) study. Results demonstrate that our method can accurately estimate the unknown parameters and correctly identify the true underlying model.

Keywords

Bayesian Lasso; Latent variable; Spline; Structural equation model

1. Introduction

Structural equation modeling (SEM) is a multivariate analysis tool that has been widely used in the behavioral, medical and social sciences where latent variables are frequently encountered (e.g., Jöreskog, 1973; Kenny and Judd, 1984; Bollen, 1989; Jöreskog and Sörbom, 1996; Schumacker and Marcoulides, 1998; Skrondal and Rabe-Hesketh, 2004; Lee, 2007). SEMs are designed to investigate the relationships among observed manifest variables and unobservable latent variables through a measurement equation and a structural equation. While the measurement equation relates the latent variables to the manifest (or response) variables, the structural equation expresses the relationships among the endogenous and exogenous latent variables. The latter thus provides the key foundation on which most research hypotheses are based and tested.

Traditional SEMs assume linear relations among latent variables (e.g., Jöreskog, 1973; Jöreskog and Sörbom, 1996). Recently, the nonlinear SEM (NSEM), originating from Kenny and Judd (1984), has generated considerable interest among SEM modelers, which

*rguo@bios.unc.edu
 **hzhu@bios.unc.edu
 ***symin@email.unc.edu
 ****ibrahim@bios.unc.edu

Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 3 & 4 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

allows nonlinear relations among latent variables such as the interactions and quadratic terms (see also Schumacker and Marcoulides, 1998, and the references therein). However, one fundamental issue bypassed by NSEMs lies in the fact that in real data analysis, the parametric form(s) of the nonlinearities are usually unknown. Misspecification of the corresponding parametric functions, in contrast, can lead to very misleading results, as we will show in our simulation study. To handle the unknown parametric forms at the latent level, several semiparametric approaches have been developed recently in the context of SEMs (e.g., Bauer, 2005; Fahrmeir and Raach, 2007; Song and Lu, 2010). However, the proposed modeling frameworks do not or only take into account limited kinds of interactions, which restricts their utility in practice.

Besides estimation, model selection is another issue we seek to address in this paper. Most contemporary work on SEMs focuses on model selection for parametric SEMs (e.g., Raftery, 1993; Lee and Tang, 2006; Lee and Song, 2008), as opposed to semiparametric modeling. Fahrmeir and Raach (2007) and Song and Lu (2010) did not consider the model selection issue in their semiparametric formulations. Classical model selection methods are based on pairwise hypothesis testing of the nested models under comparison, but these methods are usually difficult for comparing non-nested complex SEMs. In addition, popular criteria for non-nested models, like AIC and BIC as adopted by Bauer (2005), are computationally expensive and unstable when the search dimension is high – an inherent feature for most semiparametric SEMs. Recently, Bayesian model selection methods have been developed in the SEM context (e.g., Raftery, 1993; Lee and Song, 2003; Lee, 2007) due to the advantages of the Bayesian approach (see Lee, 2007), where the most commonly used measure is the Bayes factor (Kass and Raftery, 1995). Despite its popularity, computation of the Bayes factor is very intensive for SEMs because multiple intractable integrals are involved. Moreover, hypothesis testing–based methods such as the Bayes factor operate by performing pairwise comparisons of two competing models. If many models are under comparison, the associated computational costs can be very intensive.

The aim of this paper is to address these issues. The present article extends earlier work on semiparametric modeling by considering a broader framework that allows for linear and nonlinear relations in the structural equation with covariates. We utilize basis expansions to represent the nonparametric structural relations in our proposed semiparametric SEM (SSEM). With the basis representation, penalization methods such as the Bayesian Lasso (Park and Casella, 2008) can be readily used to achieve simultaneous parameter estimation and model selection. No additional measure of model selection is needed and the posterior estimates automatically provide guidance for identifying the true underlying model.

To motivate the proposed method, we consider the Affective Dynamics and Individual Differences (ADID) study (Emotions and Dynamic Systems Laboratory, 2010). The purpose of this study was to evaluate individual (e.g., age and personality) differences in the dynamics of emotions in laboratory and everyday life. We investigate possible effects of three latent exogenous variables, namely, perceived stress, emotional clarity and psychological resilience on two latent endogenous variables: positive and negative emotions. Generally, higher perceived stress is known to be associated with lower positive emotion and higher negative emotion (von Känel et al., 2005). There is, however, evidence that certain protective personality traits, such as psychological resilience (the ability to cope with negative experiences) and emotion clarity (the ability to differentiate discrete emotions with clarity and prevision), may help counteract the adverse effect of stress (Ong and Allaire, 2005). Although findings from the affect literature have suggested some general linkages among these exogenous variables on positive and negative emotions, the precise structural relationships among them and the extent to which such relations are nonlinear remain unclear.

The rest of this paper is organized as follows. In Section 2, we introduce our SSEM framework with its associated basis representation. In Section 3, we propose the Bayesian Lasso method to achieve simultaneous estimation and model selection. To illustrate our proposed method, a simulation study and the ADID study are presented in Section 4. Finally, Section 5 discusses some related issues and possible extensions for future work.

2. Semiparametric Structural Equation Models

2.1 Model Formulation

We observe a $p \times 1$ vector of response (or manifest) variables $\mathbf{y}_i$ and an $s \times 1$ vector of fixed covariates $\mathbf{x}_i$ for $i = 1, \dots, n$, where n is the number of subjects. For each subject, it is common to introduce a $q \times 1$ vector of latent variables, denoted by $\boldsymbol{\omega}_i$ for $i = 1, \dots, n$. The measurement equation is commonly defined as a factor analysis model given by

$$\mathbf{y}_i = \mathbf{A}\mathbf{c}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \quad \text{for } i = 1, 2, \dots, n, \quad (1)$$

where $\mathbf{c}_i$ is an $r \times 1$ vector of known functions of $\mathbf{x}_i$, $\boldsymbol{\varepsilon}_i$ is a $p \times 1$ vector of measurement errors, $\mathbf{A}$ and $\boldsymbol{\Lambda}$ are $p \times r$ and $p \times q$ unknown parameter matrices. Assume that $\boldsymbol{\omega}_i$ and $\boldsymbol{\varepsilon}_i$ are independent, $\boldsymbol{\varepsilon}_i$ are independent copies of a $N(\mathbf{0}, \boldsymbol{\Psi}_\varepsilon)$ distribution with $\boldsymbol{\Psi}_\varepsilon = \text{diag}(\psi_{\varepsilon 1}, \dots, \psi_{\varepsilon p})$, and $\boldsymbol{\omega}_i$ ($i = 1, \dots, n$) are independently distributed. Due to the unobservable $\boldsymbol{\omega}_i$, the model is not identifiable unless certain constraints are imposed. The most common strategy is to have some elements of $\boldsymbol{\Lambda}$ and/or $\mathbf{A}_m$ to be fixed (Bollen, 1989; Lee, 2007).

The relations among the latent variables are modeled by the structural equation. Let

$\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$, where $\boldsymbol{\eta}_i$ is a $q_1 \times 1$ vector of endogenous latent variables and $\boldsymbol{\xi}_i$ is a $q_2 \times 1$ vector of exogenous latent variables with $q_2 = q - q_1$. Instead of using a parametric structural equation, we propose a general nonparametric structural equation as follows:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i) + \boldsymbol{\zeta}_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where $\boldsymbol{\zeta}_i$ is a $q_1 \times 1$ vector of residuals which follows a $N(\mathbf{0}, \boldsymbol{\Psi}_\zeta)$ distribution with covariance matrix $\boldsymbol{\Psi}_\zeta = \text{diag}(\psi_{\zeta 1}, \dots, \psi_{\zeta q_1})$, and $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i) = \{F_1(\mathbf{x}_i, \boldsymbol{\xi}_i), \dots, F_{q_1}(\mathbf{x}_i, \boldsymbol{\xi}_i)\}^T$ is a $q_1 \times 1$ vector of unknown functions of $\{\mathbf{x}_i, \boldsymbol{\xi}_i\}$. It is assumed that $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ are independently distributed, and $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{\Phi})$. For simplicity, we assume that $\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$ is nonsingular and $|\boldsymbol{\Pi}_0|$ is independent of the elements of $\boldsymbol{\Pi}$. We refer to (1) and (2) as our Semiparametric Structural Equations Model (SSEM). Generally, $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i)$ could take any arbitrary form of the covariates and exogenous latent variables, including standard linear or nonlinear functions of $\mathbf{x}_i$ and/or $\boldsymbol{\xi}_i$ as special cases (Jöreskog, 1973; Bollen, 1989; Jöreskog and Sörbom, 1996; Skrondal and Rabe-Hesketh, 2004; Lee, 2007). Consider two latent variables for example, as in the case of Jöreskog (1973), $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i) = \boldsymbol{\beta}_1 \boldsymbol{\xi}_{i1} + \boldsymbol{\beta}_2 \boldsymbol{\xi}_{i2}$; in Kenny and Judd (1984), $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i) = \boldsymbol{\beta}_1 \boldsymbol{\xi}_{i1} + \boldsymbol{\beta}_2 \boldsymbol{\xi}_{i2} + \boldsymbol{\beta}_{12} \boldsymbol{\xi}_{i1} \boldsymbol{\xi}_{i2}$; for NSEMs in Lee (2007), $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i) = \boldsymbol{\Pi} \boldsymbol{\eta}_i + \mathbf{B} \mathbf{x}_i + \mathbf{H}(\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i2})$, where $\mathbf{H}(\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i2})$ is a vector-valued function containing non-zero differentiable functions; etc. Particularly, our SSEM also includes some semiparametric latent variable models, such as the one proposed by Song and Lu (2010) when $\boldsymbol{\Pi} = \mathbf{0}$ and each component of $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i)$ equals $g_{h1}(x_{i1}) + \dots + g_{hs}(x_{is}) + f_{h1}(\boldsymbol{\xi}_{i1}) + \dots + f_{hq_2}(\boldsymbol{\xi}_{iq_2}) + \zeta_{hi}$, $h = 1, \dots, q_1$, where $g_{hj}(x_{ij})$ and $f_{hj}(\boldsymbol{\xi}_{ij})$ are all one-dimensional functions. Although their model allows an unspecified structure of each exogenous latent variable $\boldsymbol{\xi}_{ij}$ and covariate x_{ij} on $\boldsymbol{\eta}_{hi}$, they did not consider the possible interactions among these covariates and/or the exogenous latent variables, which restricts the utility of this model in many applications. In parametric SEMs, it is very common to compare models with or without terms like $\boldsymbol{\xi}_{i1} \boldsymbol{\xi}_{i2}$. Notice that (2) in our SSEM allows interactions among the covariates and/or exogenous variables.

2.2 Modeling Nonparametric Functions

We approximate the nonparametric functions in (2) by using linear basis expansions. As an illustration, we consider a simple structural equation with one endogenous and two exogenous latent variables $\xi_i = (\xi_{i1}, \xi_{i2})^T$ and no covariates, given by

$$\eta_i = F(\xi_{i1}, \xi_{i2}) + \zeta_i = f_1(\xi_{i1}) + f_2(\xi_{i2}) + f_{12}(\xi_{i1}, \xi_{i2}) + \zeta_i. \quad (3)$$

We note that any two-dimensional function $F(\xi_{i1}, \xi_{i2})$ can be decomposed into f_1 , f_2 , and f_{12} as in (3), where f_1 and f_2 could be constant but f_{12} must be a function of both ξ_{i1} and ξ_{i2} if an interaction between the two exists. For $j = 1, 2$, we model

$$f_j(\xi_{ij}) = \sum_{m_j=1}^{M_j} \beta_{jm_j} h_{jm_j}(\xi_{ij}) \quad (4)$$

as a linear basis expansion in ξ_{ij} , where $\{h_{jm_j}(\cdot), m_j = 1, \dots, M_j\}$ are basis functions for ξ_{ij} , such as piecewise polynomials and natural cubic splines, among many others (Hastie et al., 2009). For f_{12} , we model it by using the tensor product basis as follows:

$$f_{12}(\xi_{i1}, \xi_{i2}) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \beta_{m_1 m_2}^{(12)} h_{1m_1}(\xi_{i1}) h_{2m_2}(\xi_{i2}), \quad (5)$$

where $\{h_{1m_j}(\cdot), m_j = 1, \dots, M_1\}$ and $\{h_{2m_j}(\cdot), m_j = 1, \dots, M_2\}$ are basis functions for ξ_{i1} and ξ_{i2} , respectively. Specifically, if $h_{jm}(\xi_{ij}) = \xi_{ij}$ for $j = 1, 2$, (3) recovers the common nonlinear structural equation: $\eta_i = \beta_1 \xi_{i1} + \beta_2 \xi_{i2} + \beta_{12} \xi_{i1} \xi_{i2} + \zeta_i$ (Kenny and Judd, 1984).

It is worth noting that some spline bases may contain constant basis functions, as in the case of natural cubic splines and regression splines. This leads to redundant basis functions when both (4) and (5) are used, and consequently, potential identifiability problems. Without loss of generality, suppose that $h_{j1}(\xi_{ij}) = 1$. To solve the redundant basis functions problem, we combine all the constant bases together as a single intercept term β_0 , and use

$$f_j(\xi_{ij}) = \sum_{m_j=2}^{M_j} \beta_{jm_j} h_{jm_j}(\xi_{ij}) \text{ for each one-dimensional function and}$$

$$f_{12}(\xi_{i1}, \xi_{i2}) = \sum_{m_1=2}^{M_1} \sum_{m_2=2}^{M_2} \beta_{m_1 m_2}^{(12)} h_{1m_1}(\xi_{i1}) h_{2m_2}(\xi_{i2}) \text{ for the two-dimensional function } f_{12},$$

where the summation is taken from $m_j = 2$ to exclude the constant basis. The spline model for higher dimensional functions can be modified accordingly. It follows from (4) and (5) that the structural equation can be represented by a linear model of basis functions given by $\eta_i = \beta^T H(\xi_i) + \zeta_i$, where $H(\xi_i) = \{h_{1m_1}(\xi_{i1}), \dots, h_{1M_1}(\xi_{i1}), h_{2m_2}(\xi_{i2}), \dots, h_{2M_2}(\xi_{i2}), h_{1m_1}(\xi_{i1})h_{2m_2}(\xi_{i2}), \dots, h_{1M_1}(\xi_{i1})h_{2M_2}(\xi_{i2})\}^T$ and β is the coefficient parameter associated with $H(\xi_i)$.

The linear basis representation can be easily extended to the general $F(x_i, \xi_i)$ with covariates and any number of latent variables following the same techniques above, where the interactions between x_i and ξ_i are also allowed and can be modeled in a similar vein. The basis representation (5) can be generalized to higher dimensional functions if interactions among more than two variables are of interest. In practice, we usually only interested in the interactions between every two are variables. Let $H(x_i, \xi_i)$ be an $N_H \times 1$ vector of basis functions formed similarly by (4) and/or (5) for each component of $\{x_i, \xi_i\}$. By using the linear basis expansion, the general structural equation model in (2) can be represented as:

$$\eta_i = \Pi \eta_i + \mathbf{B} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\xi}_i) + \zeta_i, \quad (6)$$

where $\mathbf{B} \mathbf{H}(\mathbf{x}_i, \boldsymbol{\xi}_i)$ is the basis expansion of $\mathbf{F}(\mathbf{x}_i, \boldsymbol{\xi}_i)$ and $\mathbf{B}_{q_1 \times N_H}$ is the unknown parameter matrix. Based on (6), we adopt the idea of the Bayesian Lasso proposed by Park and Casella (2008) and use it in conjunction with MCMC methods to estimate the parameters in our SSEM. As detailed below, this approach provides a much more computationally efficient and flexible methodology for simultaneous parameter estimation and model selection.

3. Bayesian Lasso in the Semiparametric SEM

3.1 Bayesian Lasso (BLasso)

Consider the linear model: $\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, where $\mathbf{y}$ is the response vector and $\mathbf{X}$ is a standardized design matrix. The main interest is usually to estimate $\boldsymbol{\beta}$ to identify important covariates. Tibshirani (1996) proposed the least absolute shrinkage and selection operator (Lasso) by introducing an L_1 penalty to the residual sum of squares for estimating

$\boldsymbol{\beta}$, i.e., $\arg \min_{\boldsymbol{\beta}} \left\{ (\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X} \boldsymbol{\beta})^T (\mathbf{y} - \mu \mathbf{1}_n - \mathbf{X} \boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j| \right\}$. By the nature of the L_1 -norm, Lasso shrinks some of the β_j 's corresponding to the unimportant covariates exactly to 0, which automate the simultaneous model selection and estimation. Park and Casella (2008) proposed a Bayesian framework of the Lasso (BLasso) by imposing a conditional Laplace

prior: $\pi(\boldsymbol{\beta} | \sigma^2) = \prod_{j=1}^p \lambda e^{-\lambda |\beta_j| / \sqrt{\sigma^2}} / (2 \sqrt{\sigma^2})$, where the form $\lambda e^{-\lambda |z| / 2}$ can be represented as a scale mixture of normals with an exponential mixing density:

$\lambda e^{-\lambda |z| / 2} = \int_0^\infty \{e^{-z^2 / (2s)} / (\sqrt{2\pi s})\} (a^2 e^{-a^2 s / 2} / 2) ds$, $a > 0$. This motivates a hierarchical representation:

$$\begin{aligned} \mathbf{y} | \mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim N_n(\mu \mathbf{1}_n + \mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \\ \boldsymbol{\beta} | \sigma^2, \tau_1^2, \tau_2^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2 \mathbf{D}_\tau), \quad \mathbf{D}_\tau = \text{diag}(\tau_1^2, \dots, \tau_p^2) \\ \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim \pi(\sigma^2) d\sigma^2 \prod_{j=1}^p \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} d\tau_j^2, \quad \sigma^2 > 0, \end{aligned} \quad (7)$$

which greatly simplifies the computation, since all of the full conditional distributions have known closed forms and can be sampled by the Gibbs sampler (Geman and Geman, 1984).

One advantage of BLasso is that it provides a posterior sample which can be used to summarize the entire distribution of $\boldsymbol{\beta}$, whereas Lasso only gives a point estimate of $\boldsymbol{\beta}$. In particular, the posterior mode of $\boldsymbol{\beta}$ is equivalent to the Lasso estimator. Although BLasso does not shrink the posterior mean or median exactly to zero, it does shrink them close to 0 much faster than ridge regression and yields better prediction accuracy. Another advantage is that, with the hierarchical representation, BLasso is likely to be adopted in a more complex modeling framework and it provides an efficient procedure for simultaneous parameter estimation and model selection. More discussions about (B)Lasso can be found in Appendix A of the supplementary document.

3.2 Prior Specification

We discuss the BLasso prior specifications for the structural equation (6) first. For simplicity, let $\boldsymbol{\Lambda}_\omega = (\Pi, \mathbf{B})$ denote the unknown matrices in the basis representation (6). We specify BLasso priors for each row of $\boldsymbol{\Lambda}_\omega$, i.e., $\boldsymbol{\Lambda}_{\omega h}$ for $h = 1, \dots, q_1$. That is,

$$\Lambda_{\omega h} | \psi_{\zeta h}, \tau_{\Lambda_{\omega h}} \sim N(\mathbf{0}, \psi_{\zeta h} \mathbf{D}_{\Lambda_{\omega h}}), \quad \psi_{\zeta h}^{-1} \sim \text{Gamma}(\alpha_{0\zeta h}, \beta_{0\zeta h}),$$

$$\pi(\tau_{\Lambda_{\omega h}}) \propto \prod_{j=1}^{q_1} \frac{\lambda_{\Pi_h}^2}{2} e^{-\lambda_{\Pi_h}^2 \tau_{\Pi_h j}^2 / 2} \prod_{j=1}^{N_H} \frac{\lambda_{B_h}^2}{2} e^{-\lambda_{B_h}^2 \tau_{B_h j}^2 / 2}, \quad (8)$$

where $\mathbf{D}_{\Lambda_{\omega h}} = \text{diag}(\tau_{\Lambda_{\omega h}})$ with $\tau_{\Lambda_{\omega h}} = (\tau_{\Pi_h}, \tau_{B_h})^T$, $\tau_{\Pi_h} = (\tau_{\Pi_h 1}^2, \dots, \tau_{\Pi_h q_1}^2)^T$ and $\tau_{B_h} = (\tau_{B_h 1}^2, \dots, \tau_{B_h N_H}^2)^T$, λ_{Π_h} and λ_{B_h} are regularization parameters from the Laplace priors for Π_h and B_h respectively, and $\alpha_{0\zeta h}$ and $\beta_{0\zeta h}$ are pre-specified hyperparameters. By using (8), Λ_{ω} can be estimated efficiently and at the same time the estimation provides guidance for identifying the true structural relations. The regularization parameters are also assigned gamma priors (Park and Casella, 2008). For Φ , we use the inverse-Wishart prior, i.e., $\Phi \sim IW(\mathbf{R}_0, \rho_0)$ with pre-specified hyperparameters $\mathbf{R}_0$ and ρ_0 . For the case of constant basis functions, a flat prior can be assigned for the nuisance parameter β_0 .

For the measurement equation, let $\Lambda_y = (\mathbf{A}, \Lambda)$. In practice some elements of Λ and/or $\mathbf{A}$ are fixed for model identifiability purposes. To carry out the prior specification, we follow Lee and Zhu (2000). We create an index matrix $\mathbf{M} = (m_{kj})_{p \times (r+q)}$ with elements $m_{kj} = 1$ if the kj -th element of Λ_y , $\lambda_{y kj}$, is unknown and 0 otherwise. Let $r_{yk} = \sum_{j=1}^{r+q} m_{kj}$ for $k = 1, \dots, p$. If $r_{yk} > 0$, let Λ_{yk}^* be the $r_{yk} \times 1$ vector of unknown parameters in the k -th row of Λ_y and assign a conjugate prior for $\{\Lambda_{yk}^*, \psi_{\varepsilon k}\}$, i.e., $\Lambda_{yk}^* | \psi_{\varepsilon k} \sim N_{r_{yk}}(\mu_{0yk}^*, \psi_{\varepsilon k} \mathbf{H}_{0yk}^*)$, $\psi_{\varepsilon k}^{-1} \sim \text{Gamma}(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$, where μ_{0yk}^* , $\mathbf{H}_{0yk}^*$, $\alpha_{0\epsilon k}$, and $\beta_{0\epsilon k}$ are pre-specified hyperparameters. See Appendix B in the supplementary document for guidance in selecting the hyperparameters.

3.3 Related Issues

Unobservable latent variables cause several difficulties when applying the BLasso to our SSEM. One difficulty is related to the use of the spline basis, where the scales of the knots are unknown beforehand because of the unknown range of the latent variables. There are two ways to solve this problem. A simple way is to determine the range based on empirical results, or generate a separate MCMC chain using a simple structural equation or use a software procedure like SAS PROC TCALIS to run a simple SEM in order to get rough estimates of the latent variables. Another way to solve this problem is to introduce a free scale parameter s_j for each ξ_{ij} (Song and Lu, 2010). Specifically, we choose K_j percentiles from a $N(0, 1)$, say $(\kappa_1, \dots, \kappa_{K_j})$, for the spline basis, and then use $(s_j \kappa_1, \dots, s_j \kappa_{K_j})$ as the knots. To determine s_j , a prior can be assigned and an additional MCMC step will be added to update s_j . Another difficulty caused by latent variables concerns the use of the BLasso. The magnitude of the BLasso estimates can automatically guide the model selection since the covariates are standardized to the same scale. However, the mean and standard deviation of the basis functions involving ξ in $\mathbf{H}(\mathbf{x}, \xi)$ cannot be pre-determined because ξ is unobservable. We solve this problem by generating a separate MCMC chain using either a simple structural equation or a non-standardized $\mathbf{H}(\mathbf{x}, \xi)$ to obtain samples of ξ so as to calculate the scale of the basis functions for standardization purposes.

Note that formulations such as the decomposed functions f_1 , f_2 , and f_{12} in (3) are identified only up to a constant. In Song and Lu (2010), they imposed constraints on $\mathbf{f}_j = \{f_j(\xi_{1j}), \dots, f_j(\xi_{nj})\}^T$ and $\mathbf{g}_k = \{g_k(x_{1k}), \dots, g_k(x_{nk})\}^T$ for all j, k in their model by setting $\mathbf{1}^T \mathbf{f}_j = 0$ and $\mathbf{1}^T \mathbf{g}_k = 0$ to ensure unique identification of each f_j and g_k . However, this is unnecessary in our model representation (6). By standardizing $\{\eta, \mathbf{H}(\mathbf{x}, \xi)\}$ for the use of the BLasso, the constraints are automatically satisfied. For the case with constant basis functions, all

constant terms are combined into an intercept β_0 , and identifying the component functions up to a constant is enough in practice.

Ordered categorical data are frequently encountered in subject-matter areas such as public health and social science research. If ordered categorical manifest variables are present, we may follow the approach adopted by Lee and Zhu (2002) to deal with such data and all the prior specifications above remain the same. See the supplemental document for more details.

3.4 MCMC Algorithm

To fit our model, we use MCMC methods to sample from the joint posterior distribution of the parameters. The parameters of interest here are $\theta = (\theta_1^T, \theta_2^T)^T$, where $\theta_1 = \{\Lambda_y, \Psi_\varepsilon\}$ denotes the unknown parameters in the measurement equation and $\theta_2 = \{\Lambda_\omega, \Psi_\zeta, \Phi\}$ contains the unknown parameters in the structural equation. For the case with constant basis functions, there is also a nuisance parameter β_0 . Let $Y = \{y_1, \dots, y_n\}$, $X = \{x_1, \dots, x_n\}$, $C = \{c_1, \dots, c_n\}$, and $\Omega = \{\omega_1, \dots, \omega_n\}$. Since Ω is an unobserved quantity, the joint posterior distribution of interest is $p(\Omega, \theta | Y, X, C)$, which can be sampled using the Gibbs sampler along with the Metropolis Hastings (MH) algorithm. Under the prior specifications of Section 3.2, it is straightforward to derive the full conditional distributions. The main steps of the MCMC algorithm are summarized below. The detailed MCMC implementation with all the full conditional distributions and the guidance for hyperparameter specifications are given in Appendix B of the supplementary document. For the notations below, “G” denotes the gamma distribution, “IG” is the inverse-gamma distribution, “IN” denotes the inverse-Gaussian distribution. The explicit expressions for the quantities $\mu_{\Lambda_{yk}}, H_{\Lambda_{yk}}$ etc.... are given in Appendix B of the supplementary document.

A. generate Ω from $p(\Omega | Y, X, C, \theta)$ by the MH algorithm;

B. generate θ from $p(\theta | \Omega, Y, X, C)$ using the Gibbs sampler:

(B1) sample $\theta_1 = (\Lambda_y, \Psi_\varepsilon)$,

- sample $[\Lambda_{yk}^* | \cdot] \sim N_{r_{yk}}(\mu_{yk}, \psi_{\varepsilon k} H_{yk})$, and $[\psi_{\varepsilon k} | \cdot] \sim \text{IG}(\alpha_{\varepsilon k}, \beta_{\varepsilon k})$, for $k = 1, \dots, p$;

(B2) sample $\theta_2 = (\Lambda_\omega, \Psi_\zeta, \Phi)$ and $\beta_0 \sim N(\mu_{\beta_0}, \Sigma_{\beta_0})$ if it exists,

- sample $[\Lambda_{\omega h} | \cdot] \sim N(\mu_{\Lambda_{\omega h}}, \psi_{\zeta h} \Sigma_{\Lambda_{\omega h}})$, $[\psi_{\zeta h} | \cdot] \sim \text{IG}(\alpha_{\zeta h}, \beta_{\zeta h})$, $[\Phi | \cdot] \sim \text{IW}(\mathbf{R}, n + \rho_0)$
- sample $[\tau_{\Pi_h}^{-2} | \cdot] \sim \text{IN}(\mu_{\Pi_h}, \lambda_{\Pi_h})$, and $[\tau_{B_h}^{-2} | \cdot] \sim \text{IN}(\mu_{B_h}, \lambda_{B_h})$,
- sample $[\lambda_{\Pi_h}^2 | \cdot] \sim G(r_{\Pi_h}, \delta_{\Pi_h})$, and $[\lambda_{B_h}^2 | \cdot] \sim G(r_{B_h}, \delta_{B_h})$;

C. if ordered categorical manifest variables are present or unknown scale s_j is used to determine the range of ξ , additional MCMC steps using the MH algorithm are needed.

4. Examples

We conducted simulation studies and a real data analysis to examine the performance of our proposed Bayesian lasso in SSEM for simultaneous estimation and model selection.

4.1 Simulation Study

The data was simulated based on (1) and (2) with $n = 500$, $p = 9$, $q_1 = 1$, $q_2 = 2$. We used $c_i = (1, \dots, 1)^T$, $A = \text{diag}(0.0^*, 0.0^*, 0.0^*, \mu_4, \dots, \mu_9)$, and a non-overlapping structure for

$$\Lambda^T = \begin{bmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{bmatrix},$$

where the values with asterisks were fixed for identifiability purposes (see Lee, 2007). The true values used for the simulation are: $\mu_4 = \dots = \mu_9 = \lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = \lambda_{83} = \lambda_{93} = \psi_{e1} = \dots = \psi_{e9} = \psi_{\zeta} = \phi_{12} = b_1 = b_2 = 0.36$, and $\{\phi_{11}, \phi_{12}, \phi_{22}\} = \{1.0, 0.25, 1.0\}$ for Φ . For a clear illustration, we simulated a linear structure for the covariates and focused on the structural relationships among the latent variables: $\eta_i = b_1 x_{i1} + b_2 x_{i2} + f(\xi_{i1}, \xi_{i2}) + \zeta_i$ where $f(\xi_{i1}, \xi_{i2}) = f_1(\xi_{i1}) + f_2(\xi_{i2})$ with $f_1(\xi_1) = \sin(\xi_1) - \xi_1 + 0.5$ and $f_2(\xi_2) = \exp(\xi_2)/2.5 - 3.0$, and $x_1 \sim 2\text{-binomial}(1, 0.5)$, $x_2 \sim N(-0.5, 1.0)$. In practice, the true relations between the latent variables are unknown and the number of latent variables and the full model to be used mainly depend on the research questions of interest. Suppose that we are also interested in $x_3 \sim \text{binomial}(1, 0.4)$ and the interaction between ξ_1 and ξ_2 . Then the full model we use to explore the unknown structural relations at the latent level is $\eta_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + f_1(\xi_{i1}) + f_2(\xi_{i2}) + f_{12}(\xi_{i1}, \xi_{i2}) + \zeta_i$. We expect that the BLasso would help identify the true model by shrinking b_3 and the spline coefficients for $f_{12}(\xi_{i1}, \xi_{i2})$ towards zero.

Natural cubic splines were used to approximate the unknown functions. Cubic splines are linear beyond the boundaries so that we do not need to be concerned if an MCMC sample of ξ_{ij} exceeds the boundaries. For each ξ_j , $K = 5$ knots were used (κ_k , $k = 1, \dots, K$). A simple version of natural cubic splines derived from a truncated power series basis was used for illustration in this study (Hastie et al., 2009): $h_{j1}(\xi_{ij}) = 1$, $h_{j2}(\xi_{ij}) = \xi_{ij}$, and $h_{j,k+2}(\xi_{ij}) = d_k(\xi_{ij}) - d_{K-1}(\xi_{ij})$ for $k = 1, \dots, K-2$, where $d_k(\xi_{ij}) = \{(\xi_{ij} - \kappa_k)_+^3 - (\xi_{ij} - \kappa_k)_+^3\} / (\kappa_K - \kappa_k)$.

The hyperparameters used here are $\alpha_{0ek} = \alpha_{0\zeta k} = 10$, $\beta_{0ek} = \beta_{0\zeta k} = 4$, $\rho_0 = 8$, $R_0 = 5\Phi_0$ where Φ_0 is the true value of Φ , μ_{0yk} is the true value, and $H_{0yk} = 0.25I$. The hyperparameters for the regularization parameters were taken to be $r_{0B} = 1.0$, $\delta_{0B} = 0.1$. Based on our sensitivity analysis (see Appendix D of the supplementary document), the posterior distributions are not very sensitive to the choice of the hyperparameters.

We generated MCMC chains of size 10,000 with the first half used as burn-in samples, which appears to be good enough for convergence based on the Estimated Potential Scale Reduction (EPSR; Gelman, 1996) values. As suggested by Gelman (1996), the MCMC procedure is said to have converged if all the EPSR values of the unknown parameters are less than 1.2. Figure 1 plots all the EPSR values, where the convergence is achieved very fast after the first few hundred iterations. The convergence can be further observed from three additional diagnostics (see Appendix C in the supplementary document). The simulation study was repeated 100 times (100 simulated datasets) and the average posterior estimates based on the 100 replicates are presented. Before we start, a separate chain of size 5,000 was generated to obtain estimates of ξ so as to calculate the scale of $H(x, \xi)$ for standardization purposes.

For performance comparison purposes, we also analyzed the same simulated data by applying the BLasso (without using splines) to two alternative parametric NSEM models, NSEM-I and NSEM-II. For NSEM-I, we specify: $\eta_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \zeta_i$, where the posterior estimates of b_3 and γ_3 are expected to be zero since x_3 and the interaction term are not in the true model. The structural equation considered in NSEM-II is: $\eta_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \gamma_1 f_1(\xi_{i1}) + \gamma_2 f_2(\xi_{i2}) + \gamma_3 f_{12}(\xi_{i1}, \xi_{i2}) + \zeta_i$, where f_1 and f_2 are the true nonlinear functions used to simulate the data and $f_{12}(\xi_{i1}, \xi_{i2}) = \exp(\xi_{i1}) / \{1 + \exp(\xi_{i2})\}$ is an interaction term that is not present in the true model but is of interest to test.

Thus, the true values of the γ_j 's are $\gamma_1 = \gamma_2 = 1$ and $\gamma_3 = 0$. The interaction term is included here to illustrate the ability of our method to detect spurious interaction effects. The main average results of NSEM-I, NSEM-II, and the non-spline part of SSEM based on 100 simulation replicates are presented in Table 1, where only important and representative results are reported to save space. Table 2 gives the posterior estimates for the spline coefficients in the SSEM. Briefly, the results demonstrate that the BLasso for our SSEM performs very well in both estimation and model selection. The results of NSEM-II show that with the correct parametric form, the BLasso can accurately estimate the unknown parameters and identify the true model. In contrast, when the associated parametric functions are misspecified as in NSEM-I, very poor estimates and misleading model selection conclusions were obtained.

We begin with the results of NSEM-II. Table 1 shows that the unknown parameters are accurately estimated with small biases and standard deviations. Moreover, the BLasso identifies the true model by shrinking b_3 and γ_3 towards zero. Thus, if the correct parametric structural relation can be recognized, then the BLasso provides an efficient way for accurate estimation and model selection. However, in practice, the actual parametric forms are usually unknown and misspecification can yield very misleading results as shown from the results for NSEM-I. Although the parametric form of NSEM-I is very commonly used in the NSEM literature, it led to poor results here in that none of the b_j s were estimated correctly and the large magnitude of γ_3 misleadingly implies a strong interaction effect. The estimates associated with the measurement equation are relatively accurate, but with larger biases and standard deviations than SSEM and NSEM-II.

For our SSEM, non-spline parameter estimation results in Table 1 are as good as those obtained from NSEM-II wherein the true parametric functions are known, which demonstrates its superior performance. Estimates of all the spline parameters, including the intercept β_0 , parameters for f_1 , f_2 and interaction f_{12} , are given in Table 2, which are based on the standardized basis so their relative magnitudes reflect the significance of the corresponding terms in the model. As seen, the spline coefficients for the interaction term are all relatively small and close to zero, which suggests that no strong interaction effect exists and that the main effects need to be further evaluated. For the basis functions of f_1 and f_2 , there are large spline coefficients (marked by a *), and thus strong main effects of ξ_1 and ξ_2 are implied. These can be validated by inspecting Figure 2, which plots the true two-dimensional surface $f(\xi_1, \xi_2)$ and the one estimated by our SSEM. It can be seen that the estimated surface identifies the true shape. The shape of the curves of ξ_1 given different values of ξ_2 does not change with ξ_2 and vice versa, which implies that no interaction exists between the two. The function plots in Figure 2 further illustrate the spline estimation accuracy, where the true parametric structural relations were correctly identified by our SSEM. Since the relations are identified up to a constant as discussed in Section 3.3, there are slight shifts in the plot.

The performance of the proposed approach in recovering the true structural relations among the latent variables is further measured by the root mean square error of the function:

$$RMSE(\hat{f}) = \sqrt{\sum_{i=1}^N \{\hat{f}(\xi_{i1}, \xi_{i2}) - f(\xi_{i1}, \xi_{i2})\}^2 / N}$$
, where $f(\xi_{i1}, \xi_{i2}) = f_1(\xi_{i1}) + f_2(\xi_{i2})$ is the true function used to simulate the data. In NSEM-I, $\hat{f}(\xi_{i1}, \xi_{i2}) = \hat{\gamma}_1 \xi_{i1} + \hat{\gamma}_2 \xi_{i2} + \hat{\gamma}_3 \xi_{i1} \xi_{i2}$, while $\hat{f}(\xi_{i1}, \xi_{i2}) = \hat{\gamma}_1 f_1(\xi_{i1}) + \hat{\gamma}_2 f_2(\xi_{i2}) + \hat{\gamma}_3 f_{12}(\xi_{i1}, \xi_{i2})$ for NSEM-II. The corresponding RMSEs are reported in Table 1. With the known parametric specification, NSEM-II leads to the smallest RMSE, whereas the misspecification of NSEM-I yields a large RMSE. In SSEM, $f(\xi_1, \xi_2)$ is estimated using spline methods and the resultant RMSE is reasonably small, indicating a satisfactory performance even though the parametric form is unknown.

No true interaction is included above to illustrate the model selection ability of our method. We also simulated data with an interaction term using the structural equation of NSEM-II. Due to space constraints, the results of this part of the simulation are included in the supplementary document. In this case, our method still leads to accurate estimation. In particular, the estimates of the spline parameters imply a strong interaction effect, namely, the corresponding estimates are large in magnitude. The computation time of the proposed method is comparable to other commonly used Bayesian methods discussed in Lee (2007).

4.2 Affective Dynamics and Individual Differences (ADID) Study

To illustrate the empirical utility of the proposed method, we use a subset (271 participants with baseline trait measures) of the data from the ADID Study for our analysis. The participants enrolled in a laboratory study during which a series of behavioral, psychophysiological and trait measures were collected, followed by an experience sampling study. After removing data from participants who did not complete a sufficient portion of the study, a total of 247 participants were retained in the final analysis. The two endogenous variables of interest here are the participants' general levels of positive emotion (PE) and negative emotion (NE). The three latent exogenous variables used here include the individuals' scores on the emotion clarity (EC), perceived stress (PSS) and the participants' psychological resilience (PR). Therefore, $\eta_i = (PE_i, NE_i)^T$ and $\xi_i = (PSS_i, EC_i, PR_i)^T$ with $q_1 = 2$ and $q_2 = 3$. More information about this data set is given in Appendix F in the supplementary document. Our main objective is to study the relationships among the latent variables, specifically, whether the effects of PSS on PE and NE are moderated by positive traits such as PR and EC. In addition, three demographic variables, namely, Age (continuous), Income (categorical), and Education (categorical), are used as covariates in the structural equation.

Due to the relatively large number of ordinal manifest items involved and the lack of variability in some participants' item responses, we created three composite indicators (so-called "parcels") for each latent variable (see Appendix F in the supplementary document). Thus, we have in total $p = 15$ manifest variables, three parcels for each of the five latent variables. As each latent variable is related to its corresponding observable three parcels, we consider the commonly used non-overlapping structure for Λ in the measurement equation:

$$\Lambda^T = \begin{bmatrix} 1^* & \lambda_{21} & \lambda_{31} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1^* & \lambda_{52} & \lambda_{62} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{83} & \lambda_{93} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{11,4} & \lambda_{12,4} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{14,5} & \lambda_{15,5} \end{bmatrix} \quad (9).$$

We take $\mathbf{A} = \text{diag}(0^*, \dots, 0^*, \mu_7, \dots, \mu_{15})$ and $\mathbf{c}_i = (1, \dots, 1)^T$. To study the research questions of interest, we consider the following structural equation model:

$$\eta_{hi} = b_{h1}x_1 + b_{h2}x_2 + b_{h3}x_3 + f_{h1}(\xi_{i1}) + f_{h2}(\xi_{i2}) + f_{h3}(\xi_{i3}) + f_{h,12}(\xi_{i1}, \xi_{i2}) + f_{h,13}(\xi_{i1}, \xi_{i3}) + \zeta_{hi},$$

where $h = 1, 2$ and the effects of $\mathbf{x}_j$ are constrained to be linear since this part of the model is not our main interest and preliminary graphical explorations of the data indicated no obvious nonlinear patterns between $\boldsymbol{\eta}$ and $\mathbf{x}$.

To analyze the data using the proposed method, we used natural cubic splines for $f_{hj}(\xi_{ij})$ with $K_j = 5$ knots and the same choices of the hyperparameters as in the simulation study, except that $\boldsymbol{\mu}_{0,yk} = \mathbf{0}$ and $\boldsymbol{\Phi}_0 = \mathbf{I}$. MCMC chains of size 5,000, with different starting values,

were generated, which is sufficient for convergence as shown from the EPSR plot and other diagnostics (see Appendix C.2 for details).

The main results of the representative parameters are given in Table 3. The structural equation related coefficient estimates are for the standardized bases, and thus the relative magnitude of each of the coefficient estimates reflects the importance of the corresponding term. As seen from the table, all the estimates of the b_{ij} 's are small and close to zero, which implies weak effects of the covariates in predicting PE or NE. The lack of age differences in PE and NE, in contrast to some findings in the aging literature, may relate to the fact that the manifest items used to identify PE and NE here included both low- and high-intensity items, which differed from other studies that utilized primarily high-intensity affect items.

In Table 3, $\beta_{h,ij}$ indicates the coefficient corresponding to the j -th basis of $f_h(\xi_i)$ for the h -th endogenous variable and $\beta_{h,ij}^{(kl)}$ represents the coefficient corresponding to the product of the i -th basis of $f_{hk}(\xi_k)$ and the j -th basis of $f_{hl}(\xi_l)$ for $f_{h,k}(\xi_k, \xi_l)$. Specifically, the $\beta^{(12)}$ weights are for PSS*EC while the $\beta^{(13)}$ weights are associated with PSS*PR. Table 3 only reports the spline coefficients for PE to save space. The results suggest a strong interaction effect of PSS*PR on PE (some spline coefficients for PSS*PR, e.g., $\beta_{1,22}^{(13)}, \beta_{1,32}^{(13)}, \beta_{1,33}^{(13)}$, and $\beta_{1,34}^{(13)}$ are large in magnitude). The main effects of PSS and PR are thus significant, as confirmed by the spline estimates for PSS and PR. In contrast, all of the spline coefficients for PSS*EC are small and close to zero, which implies a weak interaction effect PSS*EC on PE. We thus proceed to examining the main effect of EC instead. The small corresponding coefficients for EC ($\beta_{1,22}, \dots, \beta_{1,25}$) indicate a weak main effect of EC on PE. Therefore, the underlying structure identified by the BLasso for PE has the form: $\eta_i = F_1(PSS, PR) + \zeta_{1i}$ with $F_1(PSS, PR) = f_{11}(\xi_{11}) + f_{13}(\xi_{13}) + f_{1,13}(\xi_{11}, \xi_{13})$. The surface of $F_1(PSS, PR)$ is plotted in the left panel of Figure 3, where high PSS lowers PE among individuals with low PR. However, for those with moderate to high PR, PSS has little or even a positive effect on PE. For NE, all the spline coefficients are relatively small and close to zero for both PSS*EC and PSS*PR, which implies weak interaction effects on NE. The estimates for the one-dimensional functions of ξ do not suggest the existence of other main effects, except for PSS. As shown in the right panel of Figure 3, high PSS is associated with high NE.

SAS PROC TCALIS, as a commonly used and widely accepted procedure, was also used to analyze the data to evaluate the potential input of our analysis. The modeling framework of PROC TCALIS only handles linear structural equations without covariates and interactions, i.e., $\eta_{hi} = \gamma_{h1}\xi_{1i} + \gamma_{h2}\xi_{2i} + \gamma_{h3}\xi_{3i} + \zeta_{hi}$. Results from PROC TCALIS, also given in Table 3, provide a basis for comparing the results from our SSEM, which can accommodate both linear and nonlinear relations of latent exogenous variables and covariates. The SAS results concurred largely with the results from the linear portion of our SSEM. Significance of $\gamma = (\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{22}, \gamma_{23})$ can be tested based on the SAS output of their corresponding t -values $(-4.0349, -2.1137, 5.5670, 6.3625, 1.5956, -1.2263)$, which indicates that PSS is significant in predicting both PE and NE in terms of a large absolute value of the t -values, PR is weakly related to PE and NE, and PR is significantly related to PE but not NE. Similar relationships were also found by SSEM, but our SSEM can be used to extract additional nonlinear structural relations that cannot be evaluated using PROC TCALIS.

In summary, using our newly proposed SSEM, we were able to further clarify the role of perceived stress on PE and NE and how such relationships are moderated in part by PR in the case of PE but not NE. That is, whereas individuals with higher PSS inevitably experience higher NE, PR serves as a protective factor in helping individuals maintain a high level of PE despite the heightened perceived stress. This helps provide a basis for

formulating other theories concerning the interrelationships among emotions and personality traits.

5. Discussion

In this paper, we present a general semiparametric structural equation model and propose basis expansions as a nonparametric representation of the relations at the structural level. The main advantage of our model is that it allows for any linear or nonlinear relationships between the endogenous variables and other exogenous variables/covariates. Moreover, by using the proposed basis representation, the BLasso method can be applied to our SSEM to automate model selection during estimation. The performance of our proposed method was evaluated in both simulation studies and a real data analysis. As we have seen, our BLasso in SSEM provides an accurate and efficient way for simultaneous estimation and model selection. Commonly encountered methodological difficulties were also considered and the corresponding solutions were discussed in Section 3.3.

Other pertinent issues are discussed here, with suggestions for future research. Firstly, the proposed method was illustrated by using a simple version of natural cubic splines in our examples. More sophisticated spline models may be used to improve the performance of the proposed procedures. Secondly, we only used a small number of knots in our examples. Although more knots may provide more accurate function estimation results, the number of basis functions and the associated unknown parameters also correspondingly increase, which may lead to numerical difficulties if the sample size is relatively small. Thirdly, we use the BLasso as an efficient way for simultaneous estimation and model selection. There are many variants of the (B)Lasso, we refer the reader to Appendix A of the supplementary document for more details. One possible alternative here is to use the Bayesian Group Lasso (Kyung et al., 2010) to penalize the spline parameters for each unknown function as a group.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the editor, associate editor and two referees for helpful comments, which have led to an improvement of this article. This research was partially supported by NSF grant ..., BCS-0826844, ...

...
...

References

- Bauer DJ. A Semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*. 2005; 12:513–535.
- Bollen, KA. *Structural Equations with Latent Variables*. New York: Wiley; 1989.
- Emotions and Dynamic Systems Laboratory. Unpublished manual. University of North Carolina; Chapel Hill: 2010. The Affective Dynamics and Individual Differences (ADID) study: Developing non-stationary and network-based methods for modeling the perception and physiology of emotions.
- Fahrmeir L, Raach A. A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*. 2007; 72:327–346.
- Gelman, A. Inference and monitoring convergence. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov Chain Monte Carlo in Practice*. Florida: Chapman and Hall/CRC; 1996. p. 131-143.

- Geman S, Geman D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984; 6:721–741. [PubMed: 22499653]
- Hastie, T.; Tibshirani, R.; Friedman, J. *Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2. New York: Springer; 2009.
- Jöreskog, KG. A general method for estimating a linear structural equation system. In: Goldberger, AS.; Duncan, OD., editors. *Structural Equation Models in the Social Sciences*. New York: Seminar Press; 1973. p. 85-112.
- Jöreskog, KG.; Sörbom, D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Hove and London: Scientific Software International; 1996.
- 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.
- Kyung M, Gill J, Ghosh M, Casella G. Penalized regression, standard errors, and Bayesian Lassos. *Bayesian Analysis*. 2010; 5:369–412.
- Lee, S-Y. *Structural Equation Modeling: A Bayesian Approach*. Chichester, England: Wiley; 2007.
- Lee, S-Y.; Song, X-Y. Bayesian model comparison of structural equation models. In: Dunson, D., editor. *Random Effect and Latent Variable Model Selection*. New York: Springer; 2008. p. 121-150.
- Lee SY, Tang NS. Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statistica Sinica*. 2006; 16:1117–1141.
- Lee SY, Zhu HT. Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*. 2000; 53:209–232. [PubMed: 11109705]
- Lee SY, Zhu HT. Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*. 2002; 67:189–210.
- Ong AD, Allaire JC. Cardiovascular intraindividual variability in later life: The influence of social connectedness and positive emotions. *Psychology and Aging*. 2005; 20:476–485. [PubMed: 16248706]
- Park T, Casella G. The Bayesian Lasso. *Journal of the American Statistical Association*. 2008; 103:681–686.
- Raftery, AE. Bayesian model selection in structural equation models. In: Bollen, KA.; Long, JS., editors. *Testing Structural Equation Models*. Beverly Hills, CA: Sage; 1993. p. 163-180.
- Schumacker, RE.; Marcoulides, GA. *Interaction and Nonlinear Effects in Structural Equation Modeling*. Mahway, NJ: Lawrence Erlbaum Associates; 1998.
- Skrondal, A.; Rabe-Hesketh, S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC; 2004.
- Song X-Y, Lu Z. Semiparametric latent variable models with Bayesian P-splines. *Journal of Computational and Graphical Statistics*. 2010 accepted, 1–19.
- Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*. 1996; 58:267–288.
- von Känel R, Kudielka BM, Preckel D, Hanebuth D, Herrmann-Lingen C, Frey K, Fischer JE. Opposite effect of negative and positive affect on stress procoagulant reactivity. *Physiological Behavior*. 2005; 86:61–68.

Figure 1.
Plot of EPSR values for the parameters in the simulation study. The horizontal dotted line is for EPSR=1.2.

I. Function Plots

II. Surface Plots

Figure 2.

I. Function plots: (a) True plot for $f_1(\xi_1)$; (b) Estimated plot for $f_1(\xi_1)$; (c) True plot for $f_2(\xi_2)$; (d) Estimated plot for $f_2(\xi_2)$. II. Surface plots: the left panel is the true surface with the contour plot, while the right panel is the estimated surface via the BLasso for SSEM with the contour plot. EV1 stands for the first exogenous variable ξ_1 , and EV2 for ξ_2 .

Figure 3.

Left Panel: Estimated surface plot of $F_1(PSS, PR)$ for PE with contour plot; Right Panel: Estimated curve of $f_1(PSS)$ for NE.

Table 1

Simulation Results of BLasso for SSEM, NSEM-I, and NSEM-II

Para.	True	SSEM			NSEM-I			NSEM-II		
		Est.	Bias (STD)	Est.	Est.	Bias (STD)	Est.	Est.	Bias (STD)	Est.
μ_4	0.36	0.3556	0.0044 (0.0119)	0.3720	0.0120 (0.0193)	0.3653	0.0053 (0.0399)	0.3653	0.0053 (0.0399)	0.3653
λ_{21}	0.36	0.3611	0.0011 (0.0127)	0.3645	0.0045 (0.0161)	0.3585	0.0015 (0.0153)	0.3585	0.0015 (0.0153)	0.3585
λ_{31}	0.36	0.3592	0.0008 (0.0203)	0.3675	0.0075 (0.0226)	0.3620	0.0020 (0.0212)	0.3620	0.0020 (0.0212)	0.3620
ϕ_{11}	1.0	1.0011	0.0011 (0.1270)	0.9151	0.0849 (0.1190)	0.9401	0.0599 (0.1135)	0.9401	0.0599 (0.1135)	0.9401
ϕ_{12}	0.25	0.2474	0.0026 (0.0635)	0.2576	0.0076 (0.0556)	0.2459	0.0041 (0.0487)	0.2459	0.0041 (0.0487)	0.2459
ϕ_{22}	1.0	0.9804	0.0196 (0.0869)	0.9251	0.0749 (0.1280)	0.9931	0.0069 (0.0812)	0.9931	0.0069 (0.0812)	0.9931
b_1	0.36	0.3066	0.0534 (0.0438)	-0.2020	0.5620 (0.0520)	0.3453	0.0147 (0.0357)	0.3453	0.0147 (0.0357)	0.3453
b_2	0.36	0.3611	0.0011 (0.0465)	0.6436	0.2836 (0.0641)	0.3565	0.0035 (0.0432)	0.3565	0.0035 (0.0432)	0.3565
b_3	0.0	0.0001	0.0001 (0.0350)	-0.9380	0.9380 (0.1121)	0.0042	0.0042 (0.0736)	0.0042	0.0042 (0.0736)	0.0042
γ_1	1.0 [†]	-	-	-0.6179	- (0.2302)	1.0095	0.0095 (0.2523)	1.0095	0.0095 (0.2523)	1.0095
γ_2	1.0 [†]	-	-	0.9256	- (0.1815)	0.9814	0.0186 (0.0797)	0.9814	0.0186 (0.0797)	0.9814
γ_3	0.0	-	-	-0.2718	- (0.2884)	-0.0447	0.0447 (0.1145)	-0.0447	0.0447 (0.1145)	-0.0447
λ_B	-	3.2651	- (0.0229)	3.9995	- (0.1274)	4.6699	- (0.1283)	4.6699	- (0.1283)	4.6699
RMSE($\hat{\theta}$)		0.5883		1.8549		0.1800		0.1800		0.1800

- indicates unavailable entry

[†] indicates true value for NSEM-II only

Para.: unknown parameter; Notice that those less important parameters or parameters with similar results ($\Psi_{\mathcal{E}}$, $\Psi_{\mathcal{G}}$, and the rest unknown parameters in $\mathbf{A}_m$ and $\mathbf{\Lambda}$) are not reported here to save space

Est.: corresponding posterior mean

STD: standard deviation of the estimate calculated based on 100 MCMC chains

Table 2

Estimation of Spline Coefficients for SSEM

Spline Para.	Est. (STD)	Spline Para.	Est. (STD)
β_{12}	-0.2633* (0.1248)	$\beta_{32}^{(12)}=0$	0.0100 (0.0389)
β_{13}	-0.0029 (0.0662)	$\beta_{33}^{(12)}=0$	0.0288 (0.0238)
β_{14}	-0.0195 (0.0452)	$\beta_{34}^{(12)}=0$	0.0381 (0.0249)
β_{15}	-0.1024 (0.0885)	$\beta_{35}^{(12)}=0$	0.0589 (0.0452)
β_{22}	0.1222 (0.0668)	$\beta_{42}^{(12)}=0$	-0.0026 (0.0381)
β_{23}	0.1830* (0.0551)	$\beta_{43}^{(12)}=0$	0.0168 (0.0227)
β_{24}	0.1936* (0.0534)	$\beta_{44}^{(12)}=0$	0.0276 (0.0255)
β_{25}	0.2740* (0.1297)	$\beta_{45}^{(12)}=0$	0.0689 (0.0466)
$\beta_{22}^{(12)}=0$	0.0429 (0.0552)	$\beta_{52}^{(12)}=0$	-0.0199 (0.0510)
$\beta_{23}^{(12)}=0$	-0.0679 (0.0526)	$\beta_{53}^{(12)}=0$	-0.0173 (0.0387)
$\beta_{24}^{(12)}=0$	-0.0517 (0.0571)	$\beta_{54}^{(12)}=0$	-0.0027 (0.0407)
$\beta_{25}^{(12)}=0$	-0.0111 (0.0813)	$\beta_{55}^{(12)}=0$	0.0454 (0.0666)

* marks selected relatively large values indicating the importance of the corresponding term

Spline Para.: the unknown parameter corresponding to the spline basis

Est.: corresponding posterior mean

STD: standard deviation of the estimate calculated based on 100 MCMC chains

Table 3

Estimation Results of ADID Study

Para.	SASTCALIS		Non-spline		SSEM					
	Est.	Para.	Est.	Para.	Spline Parameters for PE					
					Est.	Para.	Est.	Para.	Est.	Para.
μ_7	2.0648	β_{01}	2.0078	β_{01}	2.8761	$\beta_{1,22}^{(12)}$	0.0309	$\beta_{1,22}^{(13)}$	0.1109*	
λ_{21}	0.9492	$\beta_{1,12}$	0.9865	$\beta_{1,12}$	-0.1209*	$\beta_{1,23}^{(12)}$	-0.0253	$\beta_{1,23}^{(13)}$	-0.0505	
λ_{31}	0.9625	$\beta_{1,13}$	0.7993	$\beta_{1,13}$	-0.0533	$\beta_{1,24}^{(12)}$	-0.0122	$\beta_{1,24}^{(13)}$	-0.0277	
$\lambda_{14,5}$	0.8138	$\beta_{1,14}$	0.7107	$\beta_{1,14}$	-0.0537	$\beta_{1,25}^{(12)}$	0.0035	$\beta_{1,25}^{(13)}$	-0.0403	
$\lambda_{15,5}$	1.2906	$\beta_{1,15}$	1.0117	$\beta_{1,15}$	-0.0138	$\beta_{1,32}^{(12)}$	-0.0033	$\beta_{1,32}^{(13)}$	0.1037*	
$\psi_{\zeta 1}$	0.0414	$\beta_{1,22}$	0.0824	$\beta_{1,22}$	-0.0291	$\beta_{1,33}^{(12)}$	-0.0105	$\beta_{1,33}^{(13)}$	0.0583	
$\psi_{\zeta 2}$	0.0220	$\beta_{1,23}$	0.0563	$\beta_{1,23}$	0.0073	$\beta_{1,34}^{(12)}$	-0.0023	$\beta_{1,34}^{(13)}$	0.0552	
ϕ_{11}	0.2877	$\beta_{1,24}$	0.2402	$\beta_{1,24}$	0.0009	$\beta_{1,35}^{(12)}$	0.0044	$\beta_{1,35}^{(13)}$	-0.0137	
ϕ_{12}	-0.0009	$\beta_{1,25}$	-0.0011	$\beta_{1,25}$	0.0018	$\beta_{1,42}^{(12)}$	0.0252	$\beta_{1,42}^{(13)}$	0.0454	
ϕ_{33}	0.1032	$\beta_{1,32}$	0.1392	$\beta_{1,32}$	0.1047*	$\beta_{1,43}^{(12)}$	-0.0130	$\beta_{1,43}^{(13)}$	-0.0021	
$\gamma_{11}^{*}(b_{11})$	-0.1587	$\beta_{1,33}$	-0.0352	$\beta_{1,33}$	0.0651	$\beta_{1,44}^{(12)}$	-0.0118	$\beta_{1,44}^{(13)}$	0.0104	
$\gamma_{12}^{*}(b_{12})$	-1.5263	$\beta_{1,34}$	0.0031	$\beta_{1,34}$	0.0201	$\beta_{1,45}^{(12)}$	-0.0034	$\beta_{1,45}^{(13)}$	-0.0265	
$\gamma_{13}^{*}(b_{13})$	0.4283	$\beta_{1,35}$	-0.0030	$\beta_{1,35}$	-0.0357	$\beta_{1,52}^{(12)}$	0.0195	$\beta_{1,52}^{(13)}$	0.0037	

SSEM					
SASTCALIS			Non-spline		
Para.	Est.	Para.	Est.	Para.	Est.
Spline Parameters for PE					
$\gamma_{21}^{*}(b_{21})$	0.1584	0.0105		$\beta_{1,53}^{(12)}$	0.0001
					$\beta_{1,53}^{(13)}$
$\gamma_{22}^{*}(b_{22})$	0.4682	-0.002		$\beta_{1,54}^{(12)}$	0.0013
					$\beta_{1,54}^{(13)}$
$\gamma_{23}^{*}(b_{23})$	-0.0516	-0.0180		$\beta_{1,55}^{(12)}$	-0.0016
					$\beta_{1,55}^{(13)}$

γ 's are the structural coefficients corresponding to the exogenous latent variables for the model used in SAS TCALIS, while b 's are the structural coefficients corresponding to the covariates in SSEM.

* marks selected relatively large values indicating the importance of the corresponding term

Para.: the unknown parameter; Notice that less important parameters or parameters with similar results (Ψ_e and the rest parameters in $A_{\eta\eta}$, Λ and Φ) are not reported here to save space

Est.: corresponding posterior mean