Bayesian Lasso for Semiparametric Structural Equation Models

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

\textsuperscript{1}Department of Biostatistics, University of North Carolina at Chapel Hill, USA
\textsuperscript{2}Department of Psychology, University of North Carolina at Chapel Hill, USA

\textit{*email: rguo@bios.unc.edu}

\textit{**email: hzhu@bios.unc.edu}

\textit{***email: symiin@email.unc.edu}

\textit{****email: ibrahim@bios.unc.edu}
Web Appendix A: Review of Lasso type Penalization Methods

Consider the linear regression model \( y = \mu 1_n + X\beta + \epsilon \), where \( y \) is an \( n \times 1 \) response vector, \( X \) is a \( n \times p \) standardized design matrix, \( \epsilon \sim N(0, \sigma^2 I) \) is the error vector, \( \beta \) is a vector of unknown regression coefficients corresponding to the given covariates, and is usually of main inferential interest in the model. Estimation of \( \beta \) is based on ordinary least squares (OLS), which minimizes the residual of sum of squares (RSS): \( (y - \mu 1_n - X\beta)^T (y - \mu 1_n - X\beta) \).

Although the OLS estimates have low bias, they typically have a large variance and poor prediction accuracy. Penalization methods have been used to improve the prediction accuracy by shrinking some of the components of \( \beta \). Ridge regression shrinks \( \beta \) by imposing an \( L_2 \) penalty to the RSS, i.e.,

\[
\hat{\beta} = \arg\min_{\beta} \left\{ (y - \mu 1_n - X\beta)^T (y - \mu 1_n - X\beta) + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}.
\]

However, ridge regression shrinks very slowly and does not shrink any coefficient to zero, and hence the model is not easy to interpret. Tibshirani (1996) proposed the least absolute shrinkage and selection operator (Lasso) by introducing an \( L_1 \) penalty to the RSS, i.e.,

\[
\hat{\beta} = \arg\min_{\beta} \left\{ (y - \mu 1_n - X\beta)^T (y - \mu 1_n - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}.
\]

Notice that the intercept term \( \mu \) is not penalized, otherwise the results would depend on the origin chosen for \( y \). By the nature of the \( L_1 \) penalty, the Lasso can shrink the coefficients corresponding to the less important covariates exactly to 0. Thus, the Lasso procedure automatically serves as a model selection criterion and no additional model selection measures (e.g., AIC, BIC, Bayes factor, etc.) need to be computed.

As mentioned in Tibshirani (1996), by specifying a Laplace prior for \( \beta \) in the linear model, the Lasso estimator is actually equivalent to the posterior mode. Park and Casella (2008) further examined this in detail and proposed a Bayesian framework for Lasso. They proposed
a conditional Laplace prior for $[\beta | \sigma^2]$ of the form

$$
\pi(\beta | \sigma^2) = \prod_{j=1}^{p} \lambda e^{-|\beta_j|/\sqrt{\sigma^2}}/(2\sqrt{\sigma^2}),
$$

where the Laplace distribution with form $ae^{-a|z|}/2$ can be represented as a scale mixture of normals with an exponential mixing density

$$
ae^{-a|z|}/2 = \int_{0}^{\infty} \{e^{-z^2/(2s)}/(\sqrt{2\pi s})\} (a^2 e^{-a^2 s/2}) ds, a > 0.
$$

Therefore, the full model can be formulated in a hierarchical representation,

$$
y|\mu, X, \beta, \sigma^2 \sim N_n(\mu 1_n + X\beta, \sigma^2 I_n)$$

$$
\beta|\sigma^2, \tau_1^2, \tau_2^2, ..., \tau_p^2 \sim N_p(0_p, \sigma^2 D_\tau), \quad D_\tau = diag(\tau_1^2, ..., \tau_p^2)
$$

$$
\tau_1^2, ..., \tau_p^2 \sim \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2/2} d\tau_j^2
$$

$$
\sigma^2 \sim \pi(\sigma^2) d\sigma^2, \quad s^2 > 0,
$$

where an improper prior $\pi(\sigma^2) = 1/\sigma^2$ or any inverse-gamma prior can be used for $\sigma^2$. The intercept $\mu$ can be given an independent flat prior. Based on this hierarchical representation, the Bayesian Lasso can be conveniently implemented in a more complex Bayesian model by simply adding extra steps to the Gibbs sampler.

The Lasso/Bayesian Lasso method provides a very efficient way for simultaneous parameter estimation and model selection, which has drawn increasing attention. There has been a large amount of literature in this area from both a frequentist (e.g., Leng et al., 2004; Zou and Hastie, 2005; Yuan and Lin, 2006; Zou, 2006; Huang et al., 2008; Wang and Leng, 2008; Zhang and Huang, 2008; Zhao et al., 2009; etc.) and Bayesian framework (e.g., Park and Casella, 2008; Kyung et al., 2010; Li and Lin, 2010; Leng et al., 2010; Hans, 2010; Bornn et al., 2010; Griffin and Brown, 2010; etc.). There are various generalizations of the Lasso/Bayesian Lasso discussed in the above listed articles due to limitations of the Lasso (see the review in Kyung et al., 2010). In particular, Zou (2006) considered the consistency issue and proposed
to use the adaptive Lasso (aLasso):

$$\hat{\beta} = \arg \min_{\beta} \left\{ (y - \mu_1 n - X\beta)^T (y - \mu_1 n - X\beta) + \sum_{j=1}^{p} \lambda_j |\beta_j| \right\},$$

which uses different regularization parameters for the coefficients so that larger $\lambda_j$ can be imposed for those unimportant covariates. Zou (2006) suggested to use some preliminary estimates of $\beta$ such as the OLS estimates $\hat{\beta}_0^j$ to weight $\lambda_j = \lambda / \hat{\beta}_0^j$. See also Huang and Zhang (2008) for more on the aLasso method. The Bayesian version of aLasso is discussed in Griffin and Brown (2010) and Leng et al. (2010). Yuan and Lin (2006) proposed the group lasso (gLasso) and claimed that gLasso does the variable selection at the group level, which is defined as

$$\hat{\beta} = \arg \min_{\beta} \left\{ (y - \mu_1 n - X\beta)^T (y - \mu_1 n - X\beta) + \lambda \sum_{g=1}^{G} ||\beta_g||_G \right\},$$

where $G$ is the number of groups of covariates, $\beta_g$ is the vector of regression coefficients for the covariates in group $g$, and $||\beta||_G = (\beta^T \beta)^{1/2}$. The Bayesian version (BgLasso) can be found in Kyung et al. (2010), where the BgLasso prior can also be represented as a scale of mixture of normals, and hence the model can also be represented in a hierarchical fashion. In fact, our proposed method can be easily extended to the BgLasso version by only modifying the MCMC steps for updating $\Lambda_\omega$ and $\tau_{\Lambda_\omega}$ according to Kyung et al. (2010). Further, gLasso can be extended to the adaptive version (agLasso) as given in Wang and Leng (2008). More details and variants of these methods can be found in the cited papers and the references therein.

**Web Appendix B: Full Conditional Distributions and MCMC Implementation**

In this section, we give details regarding the main steps of our MCMC algorithm summarized in Section 3.4 for sampling from the joint posterior distribution $p(\Omega, \theta | Y, X, C)$. This algorithm can be easily extended to the case of ordered categorical variables and the additional steps required for this case are also provided. Since our examples use a spline basis with
constant basis functions, we present the MCMC algorithm for the case with the $\beta_0$ term. For the case without $\beta_0$, we just exclude the sampling step for $\beta_0$ and set it equal to zero in the rest of the implementation steps below.

B.1 Updating $\Omega$

Due to the independence among the observations, we only need to generate from the full conditional $p(\omega_i|y_i, x_i, c_i, \theta)$ for each observation $i \in \{1, 2, ..., n\}$, and $p(\Omega|Y, X, C, \theta) = \prod_{i=1}^{n} p(\omega_i|y_i, x_i, c_i, \theta)$. The full conditional distribution of $\omega_i$ is given by

$$p(\omega_i|y_i, x_i, c_i, \theta) \propto p(y_i|c_i, \omega_i, \theta_1)p(\eta_i|x_i, \xi_i, \theta_2)p(\xi_i|\theta_2)$$

$$\propto \exp \left\{ -\frac{1}{2} (\eta_i - \beta_0 - \Lambda_\omega g_{\omega_i})^T \Psi_\zeta^{-1} (\eta_i - \beta_0 - \Lambda_\omega g_{\omega_i}) - \frac{1}{2} \xi_i^T \Phi^{-1} \xi_i ight\} ,$$

where $g_{\omega_i} = (\eta_i^T, H(x_i, \xi_i)^T)^T$. Sampling from (1) can be done via the Metropolis-Hastings (MH) algorithm. To update $\omega_i$ at the $j$-th MCMC iteration, a new candidate $\omega_i^*$ is proposed from a proposal distribution $q(\omega_i^*|\sigma_\omega^2) \sim N(\omega_i^{(j)}, \sigma_\omega^2 \Sigma_\omega)$ given the previous value $\omega_i^{(j)}$, where

$$\Sigma_\omega^{-1} = \Lambda^T \Psi^{-1} \Lambda + \begin{pmatrix} \Pi_0^T \Psi_\zeta^{-1} \Pi_0 & -\Pi_0^T \Psi_\zeta^{-1} B \Delta_H \\ -\Delta_H^T B^T \Psi_\zeta^{-1} \Pi_0 & \Phi^{-1} + \Delta_H^T B^T \Psi_\zeta^{-1} B \Delta_H \end{pmatrix}$$

with $\Delta_H = \partial H(x_i, \xi_i)/\partial \xi_i|_{\xi_i=0}$. Once the spline basis or parametric structural equation is determined, computing $\Delta_H$ is straightforward. The proposed $\omega_i^*$ is accepted with probability

$$\alpha_{\omega_i} = \min \left\{ 1, \frac{p(\omega_i^*|Y, X, C, \theta)}{p(\omega_i^{(j)}|Y, X, C, \theta)} \right\} .$$

The parameter $\sigma_\omega^2$ is set to control the acceptance rate to be around 25% (Gelman et. al., 1995). For example, in our simulation study, we used $\sigma_\omega^2 = 2.5$ to achieve an acceptance rate around 25%.
B.2 Updating $\theta_1$

Recall that $\theta_1 = (\Lambda_y, \Phi_\epsilon)$. The structural component of $\Lambda_y$ is usually pre-specified and some of the parameters are fixed for identifiability reasons. We specify a prior only for the unknown parameters $\Lambda^*_yk$ for the $k$-th row of $\Lambda_y$. If $r_{yk} > 0$, define $y^*_k = (y^*_k1, \ldots, y^*_kn)^T$ with

$$y^*_{ki} = y_{ki} - \sum_{j=1}^{r+q} \lambda_{ykj}g_{pij}(1-m_{kj}),$$

where $g_{pij}$ is the $j$-th element of $g_{pi} = (c_i, \omega_i)^T$, $i = 1, \ldots, n$. Let $G_y = (C^T, \Omega^T)^T$ be an $(r + q) \times n$ matrix. For those rows with a nonzero $r_{yk}$, let $G_{yk}$ denote the submatrix of $G_y$ such that the $j$-th row with $m_{kj} = 0$ is deleted. That is, $G_{yk}$ is an $r_{yk} \times n$ matrix. Using a conjugate prior for $\{\Lambda^*_yk, \psi_{\epsilon k}\}$, i.e., $\Lambda^*_yk \mid \psi_{\epsilon k} \sim N_{r_{yk}}(\mu^*_yk, \psi_{\epsilon k}H^*_yk)$, the full conditional distribution of $\Lambda^*_yk$ is $N_{r_{yk}}(\mu_{yk}, \psi_{\epsilon k}H_{yk})$ where $H_{yk} = (H_{0yk}^{-1} + G_{yk}G_{yk}^T)^{-1}$ and $\mu_{yk} = H_{yk}(H_{0yk}^{-1}\mu^*_yk + G_{yk}y^*_k)$. Under the conjugate prior $\psi_{\epsilon k} \sim IG(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$, the posterior distribution of $\psi_{\epsilon k}$ is $IG(\alpha_{\epsilon k}, \beta_{\epsilon k})$ with

$$\alpha_{\epsilon k} = \alpha_{0\epsilon k} + n/2$$

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(y^*_k y^*_k + \mu^*_yk H_{0yk}^{-1} \mu^*_yk - \mu^*_yk H_{yk}^{-1} \mu_{yk}).$$

The hyperparameters $\mu^*_yk$, $H^*_yk$, $\alpha_{0\epsilon k}$, and $\beta_{0\epsilon k}$ can be specified based on available prior information, or from similar studies in the existing literature. For example, in our simulation studies, we used $H^*_yk = 0.25I$ as in Lee and Zhu (2000) and Lee and Tang (2006). If prior information is not available, we can specify a diffuse proper prior for $\Lambda^*_yk$ with large $H^*_yk$, say, a diagonal matrix with large diagonal elements, such as $H^*_yk = 50I$. The choice of the hyperparameter $H^*_yk$ does not affect the results too much. Similarly, if prior information is not available for $\psi_{\epsilon k}$, a non-informative prior can be used with small values for $\alpha_{0\epsilon k}$ and $\beta_{0\epsilon k}$, say $\alpha_{0\epsilon k} = 1$ and $\beta_{0\epsilon k} = 0.1$. Furthermore, Lee (2007) suggests that if the sample is large enough, we can use part of the data to conduct an auxiliary Bayesian estimation using non-informative priors and then utilize the rest to do the analysis using the obtained initial Bayesian estimates from the auxiliary study as the hyperparameter values. More discussion
about choosing hyperparameters in the context of Bayesian SEMs can be found in Lee (2007) and the references therein.

In our simulation studies, we examined different hyperparameter values and the posterior estimates are quite robust with respect to different hyperparameter specifications (see Appendix D for our sensitivity analysis).

B.3 Updating $\theta_2$ and the Regularization Parameters

The Bayesian Lasso prior is used for the structural equation parameters. To sample $\theta_2 = (\Lambda_\omega, \Psi_\zeta, \Phi)$, we also need to update the $\tau$’s and the regularization parameters associated with the Lasso prior.

First, we examine the intercept term $\beta_0$, which is not penalized by Lasso. Under a uniform improper prior, the full conditional of $\beta_0$ is $N(\mu_{\beta_0}, \Sigma_{\beta_0})$ with $\Sigma_{\beta_0} = n^{-1}\Psi_\zeta$ and $\mu_{\beta_0} = n^{-1}\sum_{i=1}^n(\eta_i - \Lambda_\omega g_{\omega i})$.

Let $G_\omega = (g_{\omega 1}, ..., g_{\omega n})$. The full conditional for the $h$-th row of $\Lambda_\omega$, $h = 1, ..., q_1$ is

$$[\Lambda_{\omega h}|\Omega, X, \psi_\zeta h, \tau_{\Lambda_{\omega h}}] \sim N_{q_1+N_H}(\mu_{\Lambda_{\omega h}}, \psi_\zeta h \Sigma_{\Lambda_{\omega h}}), \tag{2}$$

where

$$\mu_{\Lambda_{\omega h}} = \Sigma_{\Lambda_{\omega h}} G_\omega (\eta_h - \beta_{0h}1_n), \quad \Sigma_{\Lambda_{\omega h}} = (G_\omega G_\omega^T + D_{\omega h}^{-1})^{-1},$$

and $N_H$ is the number of non-constant spline basis functions of $H(x_i, \xi_i)$.

The full conditional distributions for the elements of $\tau_{\Lambda_{\omega h}} = (\tau_{\Pi h 1}^2, ..., \tau_{\Pi h q_1}^2, \tau_{B h 1}^2, ..., \tau_{B h N_H}^2)^T$ are conditionally independent with

$$\begin{bmatrix} \tau_{\Pi h j}^{-1} \mid \Pi_h, \psi_\zeta \end{bmatrix} \sim \text{Inv-Gaussian}(\mu_{\Pi_h}', \lambda_{\Pi_h}), \quad \mu_{\Pi_h}' = \frac{\lambda_{\Pi_h}}{\Pi_{h j}} \sqrt{\psi_\zeta}, \quad \lambda_{\Pi_h} = \lambda_{\Pi_h}^2;$$

$$\begin{bmatrix} \tau_{B h j}^{-1} \mid B_h, \psi_\zeta \end{bmatrix} \sim \text{Inv-Gaussian}(\mu_{B_h}', \lambda_{B_h}), \quad \mu_{B_h}' = \frac{\lambda_{B_h}}{B_{h j}} \sqrt{\psi_\zeta}, \quad \lambda_{B_h} = \lambda_{B_h}^2. \tag{3}$$
The full conditional of $\psi_{\zeta h}$ follows an inverse gamma distribution \( \text{IG}(\alpha_{\zeta h}, \beta_{\zeta h}) \) with

\[
\alpha_{\zeta h} = \alpha_{0\zeta h} + (n + 1 + q_1 + N_H)/2,
\]

\[
\beta_{\zeta h} = \beta_{0\zeta h} + \frac{1}{2} \left\{ (\eta_h - \beta_{0h} 1_n - G^T_\omega A_{wh})(\eta_h - \beta_{0h} 1_n - G^T_\omega A_{wh})^T + A^T_{wh} D^{-1}_{\omega h} A_{\omega h} \right\}.
\]

If $\beta_0 = 0$, $\alpha_{\zeta h} = \alpha_{0\zeta h} + (n + q_1 + N_H)/2$.

For $\Phi$, we use an inverse Wishart prior, i.e., $\pi(\Phi) \sim \text{IW}_{q_2}(R_0, \rho_0)$. Then, the full conditional distribution of $\Phi$ is

\[
[\Phi | \Omega] \sim \text{IW}_{q_2}(R, n + \rho_0),
\]  

where $R = \Omega_2 \Omega_2^T + R_0$ with $\Omega_2 = (\xi_1, \ldots, \xi_n)$.

Conjugate priors are used for the regularization parameters. For example, we may use $\lambda^2_{\Pi h} \sim \text{Gamma}(r_{0\Pi}, \delta_{0\Pi})$, and $\lambda^2_{B h} \sim \text{Gamma}(r_{0B}, \delta_{0B})$. Thus, it is easy to derive the full conditional distributions for $(\lambda^2_{\Pi h}, \lambda^2_{B h})$:

\[
\lambda^2_{\Pi h} | \tau_{\Pi h} \sim \text{Gamma}(q_1 + r_{0\Pi}, \sum_{j=1}^{q_1} \tau^2_{\Pi h,j} + \delta_{0\Pi}),
\]

\[
\lambda^2_{B h} | \tau_{B h} \sim \text{Gamma}(N_H + r_{0B}, \sum_{j=1}^{N_H} \tau^2_{B h,j} + \delta_{0B}).
\]

The hyperparameters involved here are $\alpha_{0\zeta h}, \beta_{0\zeta h}, \Phi_0, \rho_0, r_{0\Pi}, \delta_{0\Pi}, r_{0B}, \text{ and } \delta_{0B}$, which can be specified following similar guidelines as discussed in Appendix B.2 by either using prior information if available, suggested values from previous studies, or diffuse priors. For example, diffuse priors can be obtained by specifying small values for $\alpha_{0\zeta h}$ and $\beta_{0\zeta h}$ (e.g., $\alpha_{0\zeta h} = 1.0$, $\beta_{0\zeta h} = 0.1$), a diagonal matrix for $R_0$ with small diagonal elements (e.g., $R_0 = 0.02I$), small $\rho_0$ (e.g., $\rho_0 = 2$), and small values for $r_{0\Pi}, \delta_{0\Pi}, r_{0B}, \text{ and } \delta_{0B}$ similar to those of $\alpha_{0\zeta h}$ and $\beta_{0\zeta h}$.

More guidance regarding the choices of hyperparameters for the regularization parameters can be found in Park and Casella (2008).
B.4 Sampling Ordered Categorical Data

If there are ordered categorical manifest variables, an additional MCMC step is needed. Suppose that $y_i = (y_{i(1)}^T, y_{i(2)}^T)^T$ for $i = 1, ..., n$, where $y_{i(1)}$ is an $l \times 1$ vector of manifest variables that cannot be directly observed and $y_{i(2)}$ is the part that can be observed directly. Although $y_{i(1)}$ cannot be observed directly, the observable ordered categorical variables $z_i = (z_{i1}, ..., z_{il})^T$ imply that the information in these underlying continuous variables is given by $z_{is} = j$, if $\alpha_{sj} \leq y_{is} < \alpha_{s,j+1}$, $j \in \{0, 1, ..., b_s\}$, $s = 1, ..., l$, where $-\infty = \alpha_{s0} < \alpha_{s1} < ... < \alpha_{sb_s} < \alpha_{s,b_s+1} = \infty$. Let $\alpha = \{\alpha_1, ..., \alpha_l\}$, where $\alpha_s = (\alpha_{s1}, ..., \alpha_{sb_s})^T$, $s = 1, ..., l$.

Let $Y = \{Y_1, Y_2\}$, $Y_1 = \{y_{1(1)}, ..., y_{1(n)}\}$, $Y_2 = \{y_{2(1)}, ..., y_{2(n)}\}$. Since $Y_1$ is a latent quantity, the joint posterior distribution of interest is $p(\Omega, \theta, Y_1 | Y_2, X, Z, C)$. Thus, an additional MCMC step is needed to sample from $p(\alpha, Y_{1(s)} | \Omega, Z, C, \theta)$.

Uniform improper priors are used for $\alpha_s$, $s = 1, ..., l$, i.e., $p(\alpha_s) = p(\alpha_{s2}, ..., \alpha_{sb_s}) \propto 1$. We note that $\alpha_{s1}$ and $\alpha_{sb_s}$ are fixed for identifiability purposes. The joint posterior $p(\alpha_s, Y_{1(s)} | z_s, \Omega, C, \theta_1)$ can be written as follows

$$p(\alpha_s, Y_{1(s)} | z_s, \Omega, C, \theta_1) = p(\alpha_s | z_s, \Omega, C, \theta_1)p(y_{1(s)} | \alpha_s, z_s, \Omega, C, \theta_1).$$

Since

$$p(y_{1(s)} | \alpha_s, z_s, \Omega, C, \theta_1) \sim \prod_{i=1}^n N(\Lambda_{gs}^T g_s, \psi_{ss}) 1_{(\alpha_{s,z_s}, \alpha_{s,z_s+1})}(y_{si}),$$

it can be shown that

$$p(\alpha_s | z_s, \Omega, C, \theta_1) \propto p(z_s | \alpha_s, \Omega, \theta)p(\alpha_s) \propto \prod_{i=1}^n \left[ \Phi \left\{ \psi_{cs}^{-1/2}(\alpha_{s,z_{si}} - \Lambda_{gs}^T g_{yi}) \right\} \Phi \left\{ \psi_{cs}^{-1/2}(\alpha_{s,z_{si}} - \Lambda_{gs}^T g_{yi}) \right\} \right],$$

where $\Phi \{ \cdot \}$ is the cumulative density function of the standard normal distribution. By (6) and (5), we obtain the joint posterior

$$p(\alpha_s, y_{1(s)} | z_s, \Omega, C, \theta_1) \propto \prod_{i=1}^n \phi \left\{ \psi_{cs}^{-1/2}(y_{si} - \Lambda_{gs}^T g_{yi}) 1_{(\alpha_{s,z_s}, \alpha_{s,z_s+1})}(y_{si}) \right\},$$

where $\phi \{ \cdot \}$ is the density function of the standard normal distribution. The MH algorithm
is used to sample from (7), as in Lee and Zhu (2000). It can be seen that the acceptance rate of the proposed \((\alpha^*_s, Y^{(1)s}_s)\) does not depend on the old or proposed value of \(Y^{(1)s}_s\). Thus, we only need to update \(Y^{(1)s}_s\) from (5) if \(\alpha^*_s\) is accepted. A new \(\alpha^*_s\) is proposed from
\[
\alpha^*_{s,zs} \sim N(\alpha^*_{s,zs-1}, \sigma_{\alpha}^2(\alpha^*_j s, zs)^{1/(\alpha^*_s z_s - 1, \alpha^*_j s,zs + 1)}(\alpha^*_s z_s)),
\]
which is accepted with probability \(\min\{1, r_{\alpha_s}\}\) where
\[
r_{\alpha_s} = \frac{p(\alpha^*_s | z_s, \Omega, C, \theta_1)p(\alpha^*_{s,j} | \alpha^*_s)}{p(\alpha^*_{s,j} | z_s, \Omega, C, \theta_1)p(\alpha^*_s | \alpha^*_{j s})}.
\]
It is straightforward to show that \(r_{\alpha_s}\) depends only on \(\alpha_s\) (see Lee and Zhu, 2000). Finally, \(\sigma_{\alpha}^2\) is chosen to obtain an acceptance rate around 25%.

**B.5 Sampling the Unknown Scale Parameters for the Latent Variables**

If there are no empirical results regarding the range of the latent variables, free scale parameters are introduced to determine the knots as discussed in Section 3.3. In this case, an additional MCMC step for the scale parameters is required. We follow Song and Lu (2010) to assign a prior to \(s = (s_1, ..., s_{q_2})\) as follows:
\[
\pi(s | t) \propto \prod_{j=1}^{q_2} \frac{1}{(2\pi t_j)^{K_j/2}} \exp \left\{ -\frac{1}{2t_j} \sum_{k=1}^{K_j} (\ln |s_j k|) \right\},
\]
where \(k_k\) are the percentiles of a \(N(0, 1)\) distribution and \(t = (t_1, ..., t_{q_2})\) is a vector of unknown parameters which can be assigned diffuse inverse gamma priors. The full conditional distribution of \(s\) under this prior is given by:
\[
\pi(s | \cdot) \propto \exp \left[ -\frac{1}{2} \sum_{i=1}^{n} \{\eta_i - \beta_0 - \Lambda_\omega g_{\omega i}(s)\}^T \Psi_\zeta^{-1} \{\eta_i - \beta_0 - \Lambda_\omega g_{\omega i}(s)\} - \frac{1}{2} \sum_{j=1}^{q_2} \sum_{k=1}^{K_j} \frac{(\ln |s_j k|)^2}{t_j} \right],
\]
where \(g_{\omega i}\) now depends on the scale parameters \(s\) since the spline basis functions depend on the knots. The MH algorithm is needed to sample from this full conditional distribution.

The new candidate is proposed by \(s^*_j = v \cdot s^{(j)}_j\), where the density function of the variable \(v\) is \(p(v) \propto 1 + 1/v\), \(v \in [1/c, c]\) with tuning parameter \(c > 1\). As shown in Song and Lu...
(2010), the proposal density of $s_j$ is symmetric. Thus, the acceptance rate of $s_j$ is then

$$\min \left\{ 1, \frac{\pi(s_j^*|\cdot)}{\pi(s_j^{(j)}|\cdot)} \right\},$$

where $\pi(s_j|\cdot)$ is the full conditional distribution of $s_j$, which can be derived from (8). See Song and Lu (2010) for more details.

**Web Appendix C: Evaluation of MCMC Convergence**

The convergence of the MCMC chains can be evaluated by calculating the Estimated Potential Scale Reduction (EPSR) values (Gelman, 1996; Lee, 2007). EPSR is calculated based on parallel MCMC chains generated from different starting values of the unknown parameters. As suggested by Gelman (1996), convergence of the MCMC procedure is achieved if all of the EPSR values for the parameters are less than 1.2. We refer to Gelman (1996) for more details about EPSR. In this paper, we calculated the EPSR values for both our simulation and real data analysis to illustrate the convergence using three parallel MCMC chains with different initial values. In addition, we also test the convergence using Geweke (Geweke, 1992) and Heidelberger-Welch (Heidelberger and Welch, 1983) diagnostics. The convergence can be further observed from the trace plots as shown below.

**C.1 Simulation**

The EPSR values for all the parameters in the simulation study were calculated and plotted against the iteration number in Figure 1 of the paper. We see from Figure 1 that the MCMC procedure converged in only a few hundred iteration and the EPSR values were all less than 1.2 and close to 1.0. The convergence is further confirmed by Geweke and Heidelberger-Welch diagnostics. Geweke diagnostic suggests satisfactory convergence with a burn-in around 3,000 in terms of small magnitudes of corresponding z-scores for all the unknown parameters. Besides, all the parameters passed the Heidelberger-Welch stationary test, and the results suggest a burn-in around 1,000. Part of the diagnostic results is reported in Table 1 for
illustration. Therefore MCMC chains of 10,000 using the first half as burn-in suffice for this model.

[Table 1 about here.]

To further look at convergence of the chains, we present trace plots in Figure 1 and see from these plots that the MCMC chains have indeed converged.

[Figure 1 about here.]

C.2 ADID Study

We also calculated the EPSR values for the parameters in the ADID study. The plots of the EPSR values against the iteration number are displayed in Figure 2. It can be seen from the plot that the MCMC procedure converged very fast after about 800 iterations, whereafter the EPSR values were all less than 1.2 and close to 1.0. Geweke and Heidelberger-Welch diagnostics were also conducted to test the convergence, both of which suggest satisfactory convergence with a burn-in around 1500. The results are presented in Table 2. Therefore, an MCMC chain of size 5,000 with the first half used as a burn-in in our ADID study suffices for inference. The trace plots are shown in Figure 3 and these plots indicate excellent convergence.

[Figure 2 about here.]

[Table 2 about here.]

[Figure 3 about here.]

Web Appendix D: Sensitivity Analysis for Simulation Study

To assess the sensitivity of the Bayesian estimates to the choice of hyperparameters, we conducted the sensitivity analysis following Lee (2007) by repeating our simulation study using two sets of hyperparameters (Prior I and Prior II, perturbations of the reported prior
inputs in the paper). We perturb the hyperparameters as follows, (I) $\alpha_{0k} = \alpha_{0\xi k} = 2.5$, $\beta_{0k} = \beta_{0\xi k} = 0.5$, $r_{0B} = 10.0$, $\delta_{0B} = 0.1$ (Prior I); (II) fix $\mu_{0yk}$ at the half of the true values, $\rho_0 = 2$, $H_{0yk} = 25I$, $r_{0B} = 10.0$, $\delta_{0B} = 0.1$ (Prior II).

The representative estimation results are presented in Table 3. As we can see, similar results to those reported in the paper were obtained by using Priors I and II. Thus, the posterior estimates are not very sensitive to the choice of the hyperparameters.

[Table 3 about here.]

**Web Appendix E: Simulation Results in the Presence of Interaction**

The estimation results for our simulation study with a true interaction effect are given below. As we can see from Table 4, the unknown non-spline parameters are estimated accurately with small biases and standard deviations. Moreover, the spline estimates imply that there is a strong interaction effect between the two exogenous variables. Figure 4 further demonstrates that our method reasonably identifies the underlying structure among latent variables.

[Table 4 about here.]

[Figure 4 about here.]

**Web Appendix F: More Information about ADID Data**

The subset sample of the ADID study used in the paper consists of 271 participants (182 women, 89 men) between the ages of 18 and 86 years old. 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 during which the participants rated their momentary feelings over 30 consecutive days. Only the baseline trait measures collected during the laboratory session are used in the present analysis. The two endogenous variables,
PE and NE, were measured using items from the Positive Affect and Negative Affect Schedule (Watson et al., 1988) and remaining portions of the affect circumplex (Larsen and Diener, 1992; Russell, 1980) on a scale of 1 (never) to 4 (very often). Three latent exogenous variables used in the data analysis include individuals’ scores on the EC subscale of the Trait Meta-Mood scale (Salovey et al., 1995), the PSS (Cohen et al., 1983) and the participants’ PR, as measured using the ego-resilience scale (Block and Kremen, 1996).

Generally, higher PSS is known to be associated with lower PE and higher NE (von Känel, 2005). There is, however, evidence that certain protective personality traits, such as PR and EC, may help counteract the adverse effect of stress (Ong and Allaire, 2005). In particular, PR has been found to be associated with quicker physiological recovery from negative affect (Souza et al., 2007; Tugade et al., 2009), and more positive evaluations of the environment (Arce et al., 2009). In a similar vein, individuals with higher clarity of emotional experience have been reported to have a lower tendency to experience depression, show decreased within-person variance in emotion and exhibit more rapid return to baseline following negative events (Wilkowski and Robinson, 2008). Additionally, individuals with high levels of emotional clarity also tend to show increased psychological resilience (Salovey, 1995) and higher life satisfaction (Extremera et al., 2009). Taken together, findings from the affect literature have suggested some general linkages among PSS, PR and EC on PE and NE. The precise structural relationships among them and the extent to which such relationships are nonlinear remain unclear. In addition, findings in the aging literature have suggested that older adults tend to show selective retention of positive stimuli (Charles et al., 2003; Mikels and Larkin, 2005) and report fewer negative emotional experiences (Birditt and Fingerman, 2003; Gross and Levenson, 1997) than do younger adults.

In our analysis, three “parcels” were created for each latent variable through item parceling.
More information about item parceling can be found in Cattell (1956, 1974), Kishton and Widaman (1994), and Bandalos and Finney (2001).

References


Figure 1. Trace plots of selected parameters ($\lambda_{21}, \psi_\varepsilon, b_2, \psi_\zeta, \beta_{24}, \lambda_B$) in the simulation study
Figure 2. EPSR values for the parameters in the ADID study. The horizontal dotted line is for EPSR=1.2.
Figure 3. Trace plots of selected parameters ($\lambda_{93}, \psi_{\epsilon 2}, b_{12}, \psi_{\zeta 1}, \beta_{1.22}, \lambda_B$) in the ADID study.
Figure 4. Left panel: true surface plot \( f(\xi_1, \xi_2) \) with the contour plot; Right panel: estimated surface plot via the Bayesian Lasso for SSEM with the contour plot. EV1 stands for the first exogenous variable \( \xi_1 \), and EV2 for \( \xi_2 \).
### Table 1

*MCMC Diagnostic Results for Simulation Study*

<table>
<thead>
<tr>
<th>Para.</th>
<th>z-score</th>
<th>Start</th>
<th>p-value</th>
<th>mean</th>
<th>halfwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_4$</td>
<td>-0.9197</td>
<td>1001</td>
<td>0.3917</td>
<td>0.3218</td>
<td>0.0075</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>-1.1490</td>
<td>1001</td>
<td>0.1702</td>
<td>0.3616</td>
<td>0.0025</td>
</tr>
<tr>
<td>$\mu_9$</td>
<td>-0.9587</td>
<td>1</td>
<td>0.3259</td>
<td>0.3800</td>
<td>0.0032</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>-0.3002</td>
<td>1001</td>
<td>0.9330</td>
<td>0.3499</td>
<td>0.0009</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>0.9793</td>
<td>1</td>
<td>0.0530</td>
<td>0.3501</td>
<td>0.0017</td>
</tr>
<tr>
<td>$\lambda_{83}$</td>
<td>0.3780</td>
<td>1</td>
<td>0.7045</td>
<td>0.3925</td>
<td>0.0051</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>-0.8222</td>
<td>1</td>
<td>0.4641</td>
<td>0.4991</td>
<td>0.0070</td>
</tr>
<tr>
<td>$\psi_{41}$</td>
<td>1.9587</td>
<td>1001</td>
<td>0.5715</td>
<td>0.2893</td>
<td>0.0112</td>
</tr>
<tr>
<td>$\psi_{49}$</td>
<td>0.5227</td>
<td>1</td>
<td>0.4030</td>
<td>0.3249</td>
<td>0.0024</td>
</tr>
<tr>
<td>$b_1$</td>
<td>-0.6693</td>
<td>1001</td>
<td>0.7717</td>
<td>0.3994</td>
<td>0.0039</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-1.1703</td>
<td>1001</td>
<td>0.5546</td>
<td>-0.0390</td>
<td>0.0032</td>
</tr>
<tr>
<td>$b_3$</td>
<td>1.0201</td>
<td>1</td>
<td>0.9301</td>
<td>0.4062</td>
<td>0.0033</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-0.5693</td>
<td>1</td>
<td>0.5043</td>
<td>-0.2354</td>
<td>0.0164</td>
</tr>
<tr>
<td>$\beta_{15}$</td>
<td>-1.0815</td>
<td>1</td>
<td>0.3950</td>
<td>-0.0465</td>
<td>0.0066</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>-0.7653</td>
<td>1</td>
<td>0.3294</td>
<td>0.2280</td>
<td>0.0120</td>
</tr>
<tr>
<td>$\beta_{25}$</td>
<td>0.9021</td>
<td>1</td>
<td>0.3280</td>
<td>0.2790</td>
<td>0.0151</td>
</tr>
<tr>
<td>$\psi_{5}$</td>
<td>-1.0543</td>
<td>1001</td>
<td>0.1143</td>
<td>0.4140</td>
<td>0.0059</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-1.4823</td>
<td>1001</td>
<td>0.0598</td>
<td>0.9636</td>
<td>0.0270</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.8826</td>
<td>1</td>
<td>0.1603</td>
<td>0.2591</td>
<td>0.0048</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.3072</td>
<td>1</td>
<td>0.7911</td>
<td>0.8080</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

Para.: unknown parameter; Notice that only part of the results are reported here for illustration.

z-score: the test statistic of Geweke diagnostic.

Start: the chain starts from this number is remained

p-value: the p-value of the stationary test

mean: the mean estimate using the remained chain.

halfwidth: half the width of the 95% confidence interval for the mean.
Table 2
MCMC Diagnostic Results for ADID Study

<table>
<thead>
<tr>
<th>Para.</th>
<th>Geweke z-score</th>
<th>Heidelberg and Welch</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_7$</td>
<td>-0.4855</td>
<td>Start 1</td>
<td>0.5254</td>
<td>2.0150</td>
</tr>
<tr>
<td>$\mu_8$</td>
<td>-0.9248</td>
<td>Start 1</td>
<td>0.5354</td>
<td>1.9054</td>
</tr>
<tr>
<td>$\mu_9$</td>
<td>-1.2156</td>
<td>Start 1</td>
<td>0.8723</td>
<td>3.0675</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>-0.1708</td>
<td>Start 1</td>
<td>0.4443</td>
<td>0.9602</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>-0.7172</td>
<td>Start 1</td>
<td>0.4483</td>
<td>0.7781</td>
</tr>
<tr>
<td>$\lambda_{14,5}$</td>
<td>1.1050</td>
<td>Start 502</td>
<td>0.4702</td>
<td>0.7047</td>
</tr>
<tr>
<td>$\lambda_{15,5}$</td>
<td>1.5120</td>
<td>Start 502</td>
<td>0.9148</td>
<td>0.9906</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.6646</td>
<td>Start 502</td>
<td>0.7778</td>
<td>0.0774</td>
</tr>
<tr>
<td>$\psi_{19}$</td>
<td>-1.7176</td>
<td>Start 1</td>
<td>0.2036</td>
<td>0.1408</td>
</tr>
<tr>
<td>$\psi_{15}$</td>
<td>0.4013</td>
<td>Start 1</td>
<td>0.6649</td>
<td>0.2364</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>1.8786</td>
<td>Start 1</td>
<td>0.2713</td>
<td>-0.0347</td>
</tr>
<tr>
<td>$b_{23}$</td>
<td>0.8377</td>
<td>Start 1502</td>
<td>0.4071</td>
<td>-0.0188</td>
</tr>
<tr>
<td>$\beta_{1,12}$</td>
<td>-1.5188</td>
<td>Start 1</td>
<td>0.2441</td>
<td>-0.1167</td>
</tr>
<tr>
<td>$\beta_{1,32}^{(13)}$</td>
<td>-1.6427</td>
<td>Start 1</td>
<td>0.1652</td>
<td>0.1084</td>
</tr>
<tr>
<td>$\beta_{2,12}$</td>
<td>-1.5261</td>
<td>Start 502</td>
<td>0.4010</td>
<td>0.0458</td>
</tr>
<tr>
<td>$\beta_{2,22}^{(13)}$</td>
<td>0.9891</td>
<td>Start 1</td>
<td>0.3476</td>
<td>-0.0328</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>1.0826</td>
<td>Start 1</td>
<td>0.5171</td>
<td>0.0677</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.6660</td>
<td>Start 502</td>
<td>0.7658</td>
<td>0.2405</td>
</tr>
<tr>
<td>$\phi_{13}$</td>
<td>0.8309</td>
<td>Start 502</td>
<td>0.0518</td>
<td>-0.0603</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>-1.7499</td>
<td>Start 502</td>
<td>0.8635</td>
<td>0.1398</td>
</tr>
</tbody>
</table>

Para.: unknown parameter; Notice that only part of the results are reported here for illustration.
z-score: the test statistic of Geweke diagnostic.
Start: the chain starts from this number is remained
$p$-value: the $p$-value of the stationary test
mean: the mean estimate using the remained chain.
halfwidth: half the width of the 95% confidence interval for the mean.
## Simulation Results of BLasso for SSEM using Priors I and II

<table>
<thead>
<tr>
<th>Para.</th>
<th>True</th>
<th>Est.</th>
<th>Bias (STD)</th>
<th>Est.</th>
<th>Bias (STD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_4 )</td>
<td>0.36</td>
<td>0.3861</td>
<td>0.0261 (0.0573)</td>
<td>0.4010</td>
<td>0.0410 (0.0490)</td>
</tr>
<tr>
<td>( \lambda_{21} )</td>
<td>0.36</td>
<td>0.3523</td>
<td>0.0077 (0.0081)</td>
<td>0.3519</td>
<td>0.0081 (0.0099)</td>
</tr>
<tr>
<td>( \lambda_{31} )</td>
<td>0.36</td>
<td>0.3578</td>
<td>0.0022 (0.0195)</td>
<td>0.3643</td>
<td>0.0043 (0.0189)</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>1.0</td>
<td>0.9903</td>
<td>0.0097 (0.0698)</td>
<td>1.0425</td>
<td>0.0425 (0.0763)</td>
</tr>
<tr>
<td>( \phi_{12} )</td>
<td>0.25</td>
<td>0.2692</td>
<td>0.0192 (0.0595)</td>
<td>0.2588</td>
<td>0.0088 (0.0705)</td>
</tr>
<tr>
<td>( \phi_{22} )</td>
<td>1.0</td>
<td>1.0137</td>
<td>0.0137 (0.1155)</td>
<td>0.9814</td>
<td>0.0186 (0.0518)</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>0.36</td>
<td>0.3389</td>
<td>0.0211 (0.0190)</td>
<td>0.3387</td>
<td>0.0213 (0.0199)</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>0.36</td>
<td>0.3224</td>
<td>0.0376 (0.0430)</td>
<td>0.3457</td>
<td>-0.0143 (0.0217)</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>0.0</td>
<td>0.0550</td>
<td>0.0550 (0.0600)</td>
<td>-0.0390</td>
<td>0.0390 (0.0677)</td>
</tr>
<tr>
<td>( \beta_{12} )</td>
<td>-</td>
<td>-0.2490</td>
<td>- (0.1118)</td>
<td>-0.1681</td>
<td>- (0.0712)</td>
</tr>
<tr>
<td>( \beta_{22} )</td>
<td>-</td>
<td>0.1265</td>
<td>- (0.0776)</td>
<td>0.1082</td>
<td>- (0.0751)</td>
</tr>
<tr>
<td>( \beta_{33}^{(12)} )</td>
<td>0.0</td>
<td>0.0344</td>
<td>0.0344 (0.0200)</td>
<td>0.0335</td>
<td>0.0335 (0.0233)</td>
</tr>
<tr>
<td>( \beta_{34}^{(12)} )</td>
<td>0.0</td>
<td>0.0388</td>
<td>0.0388 (0.0177)</td>
<td>0.0457</td>
<td>0.0457 (0.0190)</td>
</tr>
<tr>
<td>( \lambda_B )</td>
<td>-</td>
<td>6.2253</td>
<td>- (0.5018)</td>
<td>6.8052</td>
<td>- (0.2941)</td>
</tr>
</tbody>
</table>

- indicates unavailable entry

Para.: unknown parameter; Notice that those less important parameters or parameters with similar results (\( \Psi_\epsilon, \Psi_\zeta \), and the rest unknown parameters in \( A_m \) and \( \Lambda \)) are not reported here to save space

Est.: corresponding posterior mean

STD: standard deviation of the estimate calculated based on 100 MCMC chains
Table 4
Simulation Results of Bayesian Lasso for SSEM with Interaction

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_4$</td>
<td>0.36</td>
<td>0.3574</td>
<td>0.0026 (0.0089)</td>
<td>$\beta_0$</td>
<td>-2.0449 (0.3202)</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>0.36</td>
<td>0.3659</td>
<td>0.0059 (0.0267)</td>
<td>$\beta_{12}$</td>
<td>-0.2729* (0.2412)</td>
</tr>
<tr>
<td>$\mu_6$</td>
<td>0.36</td>
<td>0.3513</td>
<td>0.0087 (0.0252)</td>
<td>$\beta_{13}$</td>
<td>0.4886* (0.1409)</td>
</tr>
<tr>
<td>$\mu_7$</td>
<td>0.36</td>
<td>0.3746</td>
<td>0.0146 (0.0556)</td>
<td>$\beta_{14}$</td>
<td>0.2678* (0.1127)</td>
</tr>
<tr>
<td>$\mu_8$</td>
<td>0.36</td>
<td>0.3661</td>
<td>0.0061 (0.0350)</td>
<td>$\beta_{15}$</td>
<td>0.0381 (0.0204)</td>
</tr>
<tr>
<td>$\mu_9$</td>
<td>0.36</td>
<td>0.3763</td>
<td>0.0163 (0.0345)</td>
<td>$\beta_{22}$</td>
<td>0.0392 (0.0690)</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.36</td>
<td>0.3585</td>
<td>0.0015 (0.0175)</td>
<td>$\beta_{23}$</td>
<td>0.1622 (0.1015)</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>0.36</td>
<td>0.3653</td>
<td>0.0053 (0.0399)</td>
<td>$\beta_{24}$</td>
<td>0.2067* (0.0850)</td>
</tr>
<tr>
<td>$\lambda_{52}$</td>
<td>0.36</td>
<td>0.3733</td>
<td>0.0133 (0.0352)</td>
<td>$\beta_{25}$</td>
<td>0.2630* (0.1453)</td>
</tr>
<tr>
<td>$\lambda_{62}$</td>
<td>0.36</td>
<td>0.3665</td>
<td>0.0065 (0.0311)</td>
<td>$\beta^{(12)}_{22}$</td>
<td>0.0950 (0.1205)</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>0.36</td>
<td>0.3755</td>
<td>0.0155 (0.0482)</td>
<td>$\beta^{(12)}_{23}$</td>
<td>-0.1073 (0.0794)</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>0.36</td>
<td>0.3702</td>
<td>0.0102 (0.0403)</td>
<td>$\beta^{(12)}_{24}$</td>
<td>-0.0862 (0.0793)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.36</td>
<td>0.3743</td>
<td>0.0143 (0.0312)</td>
<td>$\beta^{(12)}_{25}$</td>
<td>-0.0803 (0.0295)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.36</td>
<td>0.3535</td>
<td>0.0065 (0.0218)</td>
<td>$\beta^{(12)}_{32}$</td>
<td>-0.1709 (0.1008)</td>
</tr>
<tr>
<td>$\psi_{13}$</td>
<td>0.36</td>
<td>0.3518</td>
<td>0.0082 (0.0234)</td>
<td>$\beta^{(12)}_{33}$</td>
<td>0.0902 (0.0869)</td>
</tr>
<tr>
<td>$\psi_{14}$</td>
<td>0.36</td>
<td>0.3683</td>
<td>0.0083 (0.0206)</td>
<td>$\beta^{(12)}_{34}$</td>
<td>0.0824 (0.0843)</td>
</tr>
<tr>
<td>$\psi_{15}$</td>
<td>0.36</td>
<td>0.3576</td>
<td>0.0024 (0.0281)</td>
<td>$\beta^{(12)}_{35}$</td>
<td>0.1366 (0.0835)</td>
</tr>
<tr>
<td>$\psi_{16}$</td>
<td>0.36</td>
<td>0.3660</td>
<td>0.0060 (0.0281)</td>
<td>$\beta^{(12)}_{42}$</td>
<td>-0.2343* (0.0757)</td>
</tr>
<tr>
<td>$\psi_{17}$</td>
<td>0.36</td>
<td>0.3976</td>
<td>0.0376 (0.0634)</td>
<td>$\beta^{(12)}_{43}$</td>
<td>0.0208 (0.0747)</td>
</tr>
<tr>
<td>$\psi_{18}$</td>
<td>0.36</td>
<td>0.3543</td>
<td>0.0057 (0.0235)</td>
<td>$\beta^{(12)}_{44}$</td>
<td>-0.2333 (0.0954)</td>
</tr>
<tr>
<td>$\psi_{19}$</td>
<td>0.36</td>
<td>0.3546</td>
<td>0.0054 (0.0226)</td>
<td>$\beta^{(12)}_{45}$</td>
<td>0.1108 (0.0972)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.36</td>
<td>0.3611</td>
<td>0.0011 (0.0558)</td>
<td>$\beta^{(12)}_{52}$</td>
<td>-0.2379* (0.1932)</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.36</td>
<td>0.3705</td>
<td>0.0105 (0.0499)</td>
<td>$\beta^{(12)}_{53}$</td>
<td>-0.0431 (0.0840)</td>
</tr>
<tr>
<td>$b_3$</td>
<td>0.36</td>
<td>0.3051</td>
<td>0.0051 (0.0783)</td>
<td>$\beta^{(12)}_{54}$</td>
<td>0.0048 (0.0958)</td>
</tr>
<tr>
<td>$\psi_{9}$</td>
<td>0.36</td>
<td>0.3089</td>
<td>0.0209 (0.0321)</td>
<td>$\beta^{(12)}_{55}$</td>
<td>0.1181 (0.1865)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>1.0</td>
<td>0.9846</td>
<td>0.0154 (0.1066)</td>
<td>$\gamma_{11}$</td>
<td>0.0048 (0.0958)</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.25</td>
<td>0.2101</td>
<td>0.0399 (0.0539)</td>
<td>$\gamma_{12}$</td>
<td>0.0048 (0.0958)</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>1.0</td>
<td>0.9440</td>
<td>0.0560 (0.1199)</td>
<td>$\gamma_{12}$</td>
<td>0.0048 (0.0958)</td>
</tr>
<tr>
<td>$\lambda_B$</td>
<td>–</td>
<td>2.6357</td>
<td>– (0.1296)</td>
<td>$\gamma_{12}$</td>
<td>0.0048 (0.0958)</td>
</tr>
</tbody>
</table>

RMSE(\(\hat{f}\)) 0.6845

- indicates unavailable entry
* marks selected relatively large values indicating the importance of the corresponding term
Non-spline Para.: the unknown non-spline parameter
Spine 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