Supplementary Materials: Bayesian Sensitivity Analysis of Dynamic Factor Analysis Models with Nonparametric Prior and Possible Nonignorable Missingness

A. Gibbs Sampling Algorithm for Estimating Dynamic Factor Analysis Models with Nonparametric Prior and Possible Nonignorable Missingness

To implement the Gibbs sampler for estimating the possibly nonlinear DFA model with the characteristics proposed in the paper, *Bayesian Sensitivity Analysis of Dynamic Factor Analysis Models with Nonparametric Prior and Possible Nonignorable Missingness*, we start with initial values $\{\boldsymbol{\mu}_Z^{(0)}, \boldsymbol{\Psi}_Z^{(0)}, \alpha^{(0)}, \boldsymbol{\pi}^{(0)}, \boldsymbol{Z}^{(0)}, \boldsymbol{L}^{(0)}, \boldsymbol{H}^{(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{Y}_{\text{mis}}^{(0)}, \boldsymbol{\tau}^{(0)}, \boldsymbol{\varphi}^{(0)}\}$. At the $(\kappa+1)$th iteration with current values $\{\boldsymbol{\mu}_Z^{(\kappa)}, \boldsymbol{\Psi}_Z^{(\kappa)}, \alpha^{(\kappa)}, \boldsymbol{\pi}^{(\kappa)}, \boldsymbol{Z}^{(\kappa)}, \boldsymbol{L}^{(\kappa)}, \boldsymbol{H}^{(\kappa)}, \boldsymbol{\theta}^{(\kappa)}, \boldsymbol{Y}_{\text{mis}}^{(\kappa)}, \boldsymbol{\tau}^{(\kappa)}, \boldsymbol{\varphi}^{(\kappa)}\}$,

(a) Generate $\boldsymbol{\mu}_Z^{(\kappa+1)}$ from $p(\boldsymbol{\mu}_Z|\boldsymbol{Z}^{(\kappa)}, \boldsymbol{\Psi}_Z^{(\kappa)})$;

(b) Generate $\boldsymbol{\Psi}_Z^{(\kappa+1)}$ from $p(\boldsymbol{\Psi}_Z|\boldsymbol{Z}^{(\kappa)}, \boldsymbol{\mu}_Z^{(\kappa+1)})$;

(c) Generate $\alpha^{(\kappa+1)}$ from $p(\alpha|\boldsymbol{\pi}^{(\kappa)})$;

(d) Generate $(\boldsymbol{\pi}^{(\kappa+1)}, \boldsymbol{Z}^{(\kappa+1)})$ from $p(\boldsymbol{\pi}, \boldsymbol{Z}|\boldsymbol{L}^{(\kappa)}, \boldsymbol{\mu}_Z^{(\kappa+1)}, \boldsymbol{\Psi}_Z^{(\kappa+1)}, \alpha^{(\kappa+1)}, \boldsymbol{H}^{(\kappa)}, \boldsymbol{\theta}^{(\kappa)})$;

(e) Generate $\boldsymbol{L}^{(\kappa+1)}$ from $p(\boldsymbol{L}|\boldsymbol{\pi}^{(\kappa+1)}, \boldsymbol{Z}^{(\kappa+1)}, \boldsymbol{\theta}^{(\kappa)}, \boldsymbol{H}^{(\kappa)})$;

(f) Generate $(\boldsymbol{\tau}^{(\kappa+1)}, \boldsymbol{Y}_{(1)\text{obs}}^{(\kappa+1)})$ from $p(\boldsymbol{\tau}, \boldsymbol{Y}_{(1)\text{obs}}|\boldsymbol{\theta}^{(\kappa)}, \boldsymbol{H}^{(\kappa)}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{C})$;

(g) Generate $\boldsymbol{H}^{(\kappa+1)}$ from $p(\boldsymbol{H}|\boldsymbol{Y}_{\text{mis}}^{(\kappa)}, \boldsymbol{Y}_{(1)\text{obs}}^{(\kappa+1)}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{\theta}^{(\kappa)}, \boldsymbol{b}^{(\kappa+1)})$;

(h) Generate $\boldsymbol{\theta}^{(\kappa+1)}$ from $p(\boldsymbol{\theta}|\boldsymbol{Y}_{\text{mis}}^{(\kappa)}, \boldsymbol{Y}_{(1)\text{obs}}^{(\kappa+1)}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{H}^{(\kappa+1)}, \boldsymbol{b}^{(\kappa+1)})$;

(i) Generate $\boldsymbol{Y}_{\text{mis}}^{(\kappa+1)}$ from $p(\boldsymbol{Y}_{\text{mis}}|\boldsymbol{\theta}^{(\kappa+1)}, \boldsymbol{H}^{(\kappa+1)}, \boldsymbol{r}, \boldsymbol{\varphi}^{(\kappa)}, \boldsymbol{C}, \boldsymbol{U}_{\text{obs}})$;

(j) Generate $\boldsymbol{\varphi}^{(\kappa+1)}$ from $p(\boldsymbol{\varphi}|\boldsymbol{Y}_{\text{mis}}^{(\kappa+1)}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})$.

Nextm we describe each of these full conditional distributions in turn.

A.1 Steps (a)—(e) Conditional Distributions Related to the Non–parametric Components

The main idea behind efficient sampling of the non–parametric components is to recast the definition of $\boldsymbol{b}_i$ in terms of the latent variable L_i $(i = 1, \ldots, n)$, which records the cluster membership of $\boldsymbol{b}_i$ such that $\boldsymbol{b}_i = \boldsymbol{Z}_{L_i}$. The base distribution in the present context was defined to be a n_b–variate normal distribution with mean vector $\boldsymbol{\mu}_Z$ and covariance matrix $\boldsymbol{\Psi}_Z$. Conjugate prior distributions were specified for $\boldsymbol{\mu}_Z$, $\boldsymbol{\Psi}_Z$ and α. To explore the posterior in relation to the non-parametric components, we sample $(\boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \alpha)$ by means of the blocked Gibbs sampler to encourage mixing of the Markov chain. That is, Gibbs sampling of the nonparametric components was regrouped into five subsidiary steps—or *blocks*, involving sampling from the conditional distributions $p(\boldsymbol{\mu}_Z|\boldsymbol{Z}, \boldsymbol{\Psi}_Z)$, $p(\boldsymbol{\Psi}_Z|\boldsymbol{Z}, \boldsymbol{\mu}_Z)$, $p(\alpha|\boldsymbol{\pi})$, $p(\boldsymbol{\pi}, \boldsymbol{Z}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \alpha, \boldsymbol{H}, \boldsymbol{\theta})$ and $p(\boldsymbol{L}|\boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{H})$. These five conditional distributions are summarized below.

Block 1. Posterior samples of $[\boldsymbol{\mu}_Z|\boldsymbol{Z},\boldsymbol{\Psi}_Z]$ for the specified prior $p(\boldsymbol{\mu}_Z) \stackrel{D}{=} \mathrm{N}_{n_b}(\boldsymbol{\mu}_{Z_0},\boldsymbol{\Psi}_{\mu_Z})$ can be obtained by sampling from

$$p(\boldsymbol{\mu}_Z|\boldsymbol{Z},\boldsymbol{\Psi}_Z) \quad \sim \quad \mathrm{N}_{n_b}(\boldsymbol{\mu}_\mu,\boldsymbol{\Sigma}_\mu), \tag{1}$$

where $\boldsymbol{\Sigma}_\mu = (G\boldsymbol{\Psi}_Z^{-1} + \boldsymbol{\Psi}_{\mu_Z}^{-1})^{-1}$ and $\boldsymbol{\mu}_\mu = \boldsymbol{\Sigma}_\mu(\boldsymbol{\Psi}_{\mu_Z}^{-1}\boldsymbol{\mu}_{Z_0} + \boldsymbol{\Psi}_Z^{-1}\sum_{g=1}^{G}\boldsymbol{Z}_g)$.

Block 2. For $j = 1,\ldots, n_b$, each of the diagonal elements of $\boldsymbol{\Psi}_Z$ given $\boldsymbol{Z}$ and $\boldsymbol{\mu}_Z$ for the specified prior $p(\psi_{z_j}^{-1}) \stackrel{D}{=} \Gamma(c_1,c_2)$ is distributed as

$$p(\psi_{z_j}^{-1}|\boldsymbol{Z},\boldsymbol{\mu}_Z) \stackrel{i.i.d}{\sim} \mathrm{Gamma}(c_1 + \frac{G}{2}, c_2 + \frac{1}{2}\sum_{g=1}^{G}(u_{g_j} - \mu_{z_j})^2), \tag{2}$$

where u_{g_j} is the jth element of the values in $\boldsymbol{Z}$ associated with point mass (or cluster) g and μ_{z_j} is the jth element of $\boldsymbol{\mu}_Z$.

Block 3. Following the derivations detailed elsewhere (Ishwaran & Zarepour, 2000; Ishwaran & James, 2001; Lee et al., 2007), the conditional distribution $(\alpha|\boldsymbol{\pi})$ corresponding to prior $p(\alpha) \stackrel{D}{=} \Gamma(a_1,a_2)$ can be shown to be

$$p(\alpha|\boldsymbol{\pi}) \quad \sim \quad \mathrm{Gamma}(a_1 + G - 1, a_2 - \sum_{g=1}^{G-1}\log(1-\nu_g^*)), \tag{3}$$

where ν_g^* is a random weight sampled from the beta distribution and it is sampled within Block 4.

Block 4. As $\boldsymbol{\pi}$ and α are independent given $(\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{H})$, the distribution $(\boldsymbol{\pi},\boldsymbol{Z}|\boldsymbol{L},\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z,\alpha,\boldsymbol{\theta},\boldsymbol{H})$ is proportional to $p(\boldsymbol{\pi}|\boldsymbol{L},\alpha)p(\boldsymbol{Z}|\boldsymbol{L},\boldsymbol{\mu}_Z,\boldsymbol{\Psi}_Z,\boldsymbol{\theta},\boldsymbol{H})$. Thus, the conditional distribution can be decomposed into two independent components to be derived separately.

Conditional distribution $p(\boldsymbol{\pi}|\boldsymbol{L},\alpha)$. It can be shown that the conditional distribution $(\boldsymbol{\pi}|\boldsymbol{L}, \alpha)$ conforms to a generalized Dirichlet distribution as

$$p(\boldsymbol{\pi}|\boldsymbol{L}, \alpha) \quad \sim \quad \mathfrak{g}(a_1^*, b_1^*, \cdots, a_{G-1}^*, b_{G-1}^*), \tag{4}$$

where $a_g^* = 1 + d_g$, $b_g^* = \alpha + \sum_{j=g+1}^{G} d_j$ for $g = 1, \ldots, G-1$, and d_g is the number of L_is (and thus individuals) whose value equals to g. Sampling from the conditional distribution $(\boldsymbol{\pi}|\boldsymbol{L}, \alpha)$ can be accomplished as follows. First, ν_g^* is drawn from a Beta(a_g^*, b_g^*) distribution. Subsequently, π_g is obtained for $g = 1, \ldots, G$ as

$$\pi_1 = \nu_1^*, \ \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g \text{ and } \pi_g = \prod_{j=1}^{g-1} (1 - \nu_j^*) \nu_g^* \text{ for } g \neq 1 \text{ or } G. \tag{5}$$

Conditional distribution $p(\boldsymbol{Z}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H})$. Let $L_1^*, \cdots, L_d^*$ be the d unique L_i values (i.e., unique number of "clusters"), $\boldsymbol{Z}^L = (\boldsymbol{Z}_{L_1^*}, \cdots, \boldsymbol{Z}_{L_d^*})$, and let $\boldsymbol{Z}^{[L]}$ be components in $\boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_G)$ other than $\boldsymbol{Z}^L$. Then

$$p(\boldsymbol{Z}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H}) = p(\boldsymbol{Z}^{[L]}|\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) p(\boldsymbol{Z}^L|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H}),$$

where $p(\boldsymbol{Z}^{[L]}|\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z)$ is simply the n_b–variate normal distribution, $\mathrm{N}_{n_b}(\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z)$, and

$$p(\boldsymbol{Z}^L|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H}) = \Pi_{g=1}^d p(\boldsymbol{Z}_{L_g^*}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H}).$$

It can be shown that the conditional distribution $p(\boldsymbol{Z}_{L_g^*}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H})$ is non–standard and cannot be derived directly via Gibbs sampling. Specifically, $p(\boldsymbol{Z}_{L_g^*}|\boldsymbol{L}, \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\theta}, \boldsymbol{H}) \propto p(\boldsymbol{Z}_{L_g^*}|\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \prod_{\{i: L_i = L_g^*\}} p(\boldsymbol{\eta}_i|\boldsymbol{b}_i =$

$\boldsymbol{Z}_{L_g^*}, \boldsymbol{\theta}_\eta)$, in which $p(\boldsymbol{\eta}_i|\boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}, \boldsymbol{\theta}_\eta)$ is given by

$$\begin{cases} p(\boldsymbol{\eta}_{i0}) \prod_{t=1}^{T_i} p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}, \boldsymbol{\theta}_\eta) & \text{if } \boldsymbol{\eta}_{i0} \text{ is stochastic,} \\ \prod_{t=1}^{T_i} p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}, \boldsymbol{\theta}_\eta) & \text{otherwise.} \end{cases} \tag{6}$$

From Equation (6), it can be noted that multiplication involving the density $p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta)$ results in a conditional density that is non–normal and non–standard due to the nonlinearity of $f_t(.)$ and the fact that $\boldsymbol{Z}_{L_g^*}$ is random, as opposed to fixed within this sampling step. Instead, we adopt a MH step as follows. At the jth iteration with a current value $\boldsymbol{Z}_{L_g^*}^{(j)}$, a new candidate $\boldsymbol{Z}_{L_g^*}$ is generated from the normal distribution $\mathrm{N}(\boldsymbol{Z}_{L_g^*}^{(j)}, \sigma_b^2 \boldsymbol{\Omega}_b)$, where $\boldsymbol{\Omega}_b = (\boldsymbol{\Psi}_Z^{-1} + \sum_{\{i:L_i=L_g^*\}} \sum_{t=1}^{T_i} \boldsymbol{\Delta}_{bit}^T \boldsymbol{\Psi}_\zeta^{-1} \boldsymbol{\Delta}_{bit})^{-1}$ and $\boldsymbol{\Delta}_{bit} = \partial \boldsymbol{\eta}_{it} / \partial \boldsymbol{b}_i^T |_{\boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}^{(j)}}$. The new $\boldsymbol{Z}_{L_g^*}$ is accepted with probability

$$\min \left\{ 1, \frac{p(\boldsymbol{Z}_{L_g^*}|\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \prod_{\{i:L_i=L_g^*\}} \prod_{t=1}^{T_i} p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}, \boldsymbol{\theta}_\eta)}{p(\boldsymbol{Z}_{L_g^*}^{(j)}|\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \prod_{\{i:L_i=L_g^*\}} \prod_{t=1}^{T_i} p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i = \boldsymbol{Z}_{L_g^*}^{(j)}, \boldsymbol{\theta}_\eta)} \right\}. \tag{7}$$

The variance σ_b^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

Block 5. The conditional distribution $(L_i|\boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{H})$ is given by

$$(L_i|\boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{H}) \overset{\text{i.i.d}}{\sim} \text{Multinomial}(\pi_{ig}^*, g = 1, \cdots, G), \tag{8}$$

where π_{ig}^* is proportional to $(\pi_g p(\boldsymbol{\eta}_i|\boldsymbol{b}_i = \boldsymbol{Z}_g, \boldsymbol{\theta}_\eta))$ and π_g ($g = 1, \ldots$, G) are available from step (i.e., block) 4 summarized in Equation (5).

A.2 Step (f): Conditional Distribution $p(\boldsymbol{\tau}, \boldsymbol{Y}_{(1)\text{obs}}|\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{C})$

We use an improper prior for the threshold parameters, namely, $\boldsymbol{\tau}_k \sim c_\tau \mathbf{1}$. To sample $\boldsymbol{\tau}$ and $\boldsymbol{Y}_{(1)\text{obs}}$, we first note that $p(\boldsymbol{\tau}_k, \boldsymbol{Y}_{(1)k,\text{obs}}|\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{Y}^*_{(1)k}, \boldsymbol{C}) = p(\boldsymbol{\tau}_k|\boldsymbol{Y}^*_{(1)k}, \boldsymbol{\theta}, \boldsymbol{H}) p(\boldsymbol{Y}_{(1)k,\text{obs}}|\boldsymbol{\tau}_k, \boldsymbol{Y}^*_{(1)k}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{C})$, where

$$p(\boldsymbol{\tau}_k|\boldsymbol{Y}^*_{(1)k}, \boldsymbol{\theta}, \boldsymbol{H}) \quad \propto \prod_{i=1}^{n} \prod_{t=1}^{T_i} \left(\Phi(\frac{\tau_{k,u_{itk}} - A_k^T c_{it} - \Lambda_k^T \eta_{it}}{\psi_{\epsilon k}^{1/2}}) - \Phi(\frac{\tau_{k,u_{itk}-1} - A_k^T c_{it} - \Lambda_k^T \eta_{it}}{\psi_{\epsilon k}^{1/2}}) \right) \tag{9}$$

$$p(y_{itk}|\boldsymbol{\tau}_k, \boldsymbol{Y}^*_{(1)k}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{c}_k) \quad = \quad \mathrm{N}(\boldsymbol{A}_k^T \boldsymbol{c}_{it} + \boldsymbol{\Lambda}_k^T \boldsymbol{\eta}_{it}, \psi_{\epsilon k}) I_{(\tau_{k,u_{itk}-1}, \tau_{k,u_{itk}}]}(y_{itk}),$$

where $\boldsymbol{\tau}_k = (\tau_{k1}, \ldots, \tau_{k,b_k-1})$, $\boldsymbol{Y}_{(1)k,\text{obs}} = \{y_{itk} : i = 1, \ldots, t = 1, \ldots, T_i\}$ in which y_{itk} is the kth component of $\boldsymbol{y}_{(1)it,\text{obs}}$ corresponding to $\boldsymbol{U}_{it,\text{obs}}$, $\boldsymbol{Y}^*_{(1)k} = \{\boldsymbol{u}_{itk} : i = 1, \ldots, t = 1, \ldots, T_i\}$ in which u_{itk} is the kth component of $\boldsymbol{U}_{it}$ corresponding to $\boldsymbol{U}_{it,\text{obs}}$, $\boldsymbol{A}_k^T$ and $\boldsymbol{\Lambda}_k^T$ are the kth row vectors of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$, respectively. To generate observations from the non–standard and complex joint conditional density of $\boldsymbol{\tau}_k$ and $\boldsymbol{Y}_{(1)k,\text{obs}}$, the following MH step is embedded within the Gibbs sampler. Specifically, a vector of thresholds $(\tau_{k2}, \ldots, \tau_{k,b_k-2})$ is first generated from the truncated normal distribution

$$\tau_{kw} \sim \mathrm{N}(\tau_{kw}^{(j)}, \sigma^2_{\tau_k}) I_{(\tau_{k,w-1}, \tau^{(j)}_{k,w+1})}(\tau_{kw}), \qquad \text{for } w = 2, \ldots, b_k - 2, \tag{10}$$

where $\tau_{kw}^{(j)}$ denotes the value of τ_{kw} at the jth iteration of the Gibbs sampler and $\sigma^2_{\tau_k}$ is a preassigned constant. As mentioned earlier, the values of the first ($w = 1$) and last ($w = b_k - 1$) thresholds are fixed for identification purpose. Each new draw of τ_{kw} is then retained with acceptance probability min(1,R_k), where

$$R_k = \prod_{w=2}^{b_k-2} \frac{\Phi[(\tau_{k,w+1}^{(j)} - \tau_{k,w}^{(j)})/\sigma_{\tau_k}] - \Phi[(\tau_{k,w-1} - \tau_{k,w}^{(j)})/\sigma_{\tau_k}]}{\Phi[(\tau_{k,w+1} - \tau_{k,w})/\sigma_{\tau_k}] - \Phi[(\tau_{k,w-1}^{(j)} - \tau_{k,w})/\sigma_{\tau_k}]} \times$$

$$\prod_{i=1}^{n}\prod_{t=1}^{T_i} \frac{\Phi[\psi_{\epsilon_k}^{-1/2}\{\tau_{k,u_{itk}} - \boldsymbol{A}_k^T\boldsymbol{c}_{it} - \boldsymbol{\Lambda}_k^T\boldsymbol{\eta}_{it}\}] - \Phi[\psi_{\epsilon_k}^{-1/2}\{\tau_{k,u_{itk}-1} - \boldsymbol{A}_k^T\boldsymbol{c}_{it} - \boldsymbol{\Lambda}_k^T\boldsymbol{\eta}_{it}\}]}{\Phi[\psi_{\epsilon_k}^{-1/2}\{\tau_{k,u_{itk}}^{(j)} - \boldsymbol{A}_k^T\boldsymbol{c}_{it} - \boldsymbol{\Lambda}_k^T\boldsymbol{\eta}_{it}\}] - \Phi[\psi_{\epsilon_k}^{-1/2}\{\tau_{k,u_{itk}-1}^{(j)} - \boldsymbol{A}_k^T\boldsymbol{c}_{it} - \boldsymbol{\Lambda}_k^T\boldsymbol{\eta}_{it}\}]}. \tag{11}$$

Once the threshold values have been determined, they are then used to generate new draws of y_{itk} using the MH algorithm as done in Step (9) on the basis of Equation (9).

A.3 Step (g): Conditional Distribution for Latent Variable Estimates, $p(\boldsymbol{H}|\boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{b})$

The conditional distribution from which posterior samples of the latent variable estimates are obtained can be derived as

$$p(\boldsymbol{H}|\boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{b}) = \prod_{i=1}^{n}\prod_{t=1}^{T_i} p(\boldsymbol{\eta}_{it}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}_{i,t+1}^{*}, \boldsymbol{y}_{it}, \boldsymbol{\theta}, \boldsymbol{b}_i)$$

where $\boldsymbol{H}_{i,t-1} = (\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{i,t-1})$ and $\boldsymbol{H}_{i,t+1}^{*} = (\boldsymbol{\eta}_{i,t+1}, \ldots, \boldsymbol{\eta}_{iT_i})$. According to the Gibbs sampler, random draws of $\boldsymbol{\eta}_i$ from $p(\boldsymbol{\eta}_i|\boldsymbol{Y}_i, \boldsymbol{\theta}, \boldsymbol{b}_i)$ are based on those of $\boldsymbol{\eta}_{it}$ from $p(\boldsymbol{\eta}_{it}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}_{i,t+1}^{*}, \boldsymbol{y}_{it}, \ \boldsymbol{\theta}, \boldsymbol{b}_i)$ for each time point. That is, for $i = 1, \ldots, n$:

$p(\boldsymbol{\eta}_{it}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}_{i,t+1}^{*}, \boldsymbol{y}_{it}, \boldsymbol{\theta}, \boldsymbol{b}_i) \sim$

$$\begin{cases} p(\boldsymbol{y}_{it}|\boldsymbol{\eta}_{it}, \boldsymbol{\theta}_y)p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta, \boldsymbol{\Psi}_\zeta)p(\boldsymbol{\eta}_{i,t+1}|\boldsymbol{\eta}_{it}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta, \boldsymbol{\Psi}_\zeta) & \text{for } t = 1, \ldots, T_i\text{-}1, \\ p(\boldsymbol{y}_{it}|\boldsymbol{\eta}_{it}, \boldsymbol{\theta}_y)p(\boldsymbol{\eta}_{it}|\boldsymbol{\eta}_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta, \boldsymbol{\Psi}_\zeta) & \text{for } t = T_i. \end{cases}$$

Note that we could obtain a standard conditional distribution for $t = T_i$ but not for $t < T_i$. Specifically, at $t = T_i$, the conditional distribution $p(\boldsymbol{\eta}_{iT_i}|\boldsymbol{H}_{i,T_i-1}, \boldsymbol{y}_{iT_i}, \boldsymbol{\theta}, \boldsymbol{b}_i)$

is given by $\boldsymbol{\eta}_{iT_i} \sim \mathrm{N}_q(\boldsymbol{U}^*_{it_i}, \boldsymbol{B}^*)$, where $\boldsymbol{B}^* = (\boldsymbol{\Psi}_\zeta^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{\Lambda})^{-1}$ and $\boldsymbol{U}^*_{it_i} = \boldsymbol{B}^*[\boldsymbol{\Psi}_\zeta^{-1} \boldsymbol{f}_{T_i}(\boldsymbol{\eta}_{i,T_i-1}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta) + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1}(\boldsymbol{y}_{iT_i} - \boldsymbol{A}\boldsymbol{c}_{it})]$. However, when $t < T_i$, multiplication involving the density $p(\boldsymbol{\eta}_{i,t+1}|\boldsymbol{\eta}_{it}, \boldsymbol{b}_i, \boldsymbol{\theta}_\eta)$ would result in a conditional density that is non–normal and non–standard. This is due directly to the nonlinearity of $\boldsymbol{f}_t(.)$ and the fact that $\boldsymbol{\eta}_{it}$ is random, as opposed to fixed, at each t. We adopted the following MH algorithm to sample observations from the posterior density $p(\boldsymbol{\eta}_{it}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}^*_{i,t+1}, \boldsymbol{y}_{it}, \boldsymbol{\theta}, \boldsymbol{b}_i)$. At the jth iteration with a current value $\boldsymbol{\eta}_{it}^{(j)}$, a new candidate $\boldsymbol{\eta}_{it}$ is generated from the normal distribution $N(\boldsymbol{\eta}_{it}^{(j)}, \sigma_\eta^2 \boldsymbol{\Omega}_\eta)$, where $\boldsymbol{\Omega}_\eta = (\boldsymbol{B}^{*-1} + \boldsymbol{\Delta}_{it}^T \boldsymbol{\Psi}_\zeta^{-1} \boldsymbol{\Delta}_{it})^{-1}$ and $\boldsymbol{\Delta}_{it} = \partial \boldsymbol{f}_{t+1}/\partial \boldsymbol{\eta}_{it}^T|_{\eta_{it}=0}$, and it is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\eta}_{it}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}^*_{i,t+1}, \boldsymbol{y}_{it}, \boldsymbol{\theta}, \boldsymbol{b}_i)}{p(\boldsymbol{\eta}_{it}^{(j)}|\boldsymbol{H}_{i,t-1}, \boldsymbol{H}^*_{i,t+1}, \boldsymbol{y}_{it}, \boldsymbol{\theta}, \boldsymbol{b}_i)}\right\}.$$

The variance σ_η^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

A.4 Step (h): Conditional Distributions for Parameters in $\boldsymbol{\theta}$

Assuming that the parameters in $\boldsymbol{b} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ are independent of those contained in $\boldsymbol{\theta}$, and that parameters in $\boldsymbol{\theta}_\eta$ are conditionally independent of those in $\boldsymbol{\theta}_y$, the conditional distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{b}) = p(\boldsymbol{\theta}_\eta, \boldsymbol{\Psi}_\zeta|\boldsymbol{H}, \boldsymbol{b})\ p(\boldsymbol{\theta}_y|\boldsymbol{Y}, \boldsymbol{H})$ is derived by computing the latter two densities separately for all the person–invariant parameters in the dynamic and measurement models.

Parameters in the Dynamic Model. At the dynamic level, the only parametric posterior distribution associated with $p(\boldsymbol{\theta}_\eta, \boldsymbol{\Psi}_\zeta|\boldsymbol{H}, \boldsymbol{b})$ is that of $p(\boldsymbol{\Psi}_\zeta|\boldsymbol{H}, \boldsymbol{b})$. We used a q–dimensional inverse Wishart distribution as the conjugate prior for the process noise covariance matrix, $\boldsymbol{\Psi}_\zeta$, i.e., $p(\boldsymbol{\Psi}_\zeta) \sim \mathrm{IW}_q(\rho_0, \boldsymbol{\Psi}_{\zeta_0})$, thus yielding

$$p(\boldsymbol{\Psi}_\zeta|\boldsymbol{H},\boldsymbol{b}) \sim \mathrm{IW}_q\left(\sum_{i=1}^{n} T_i + \rho_0, \boldsymbol{R}_\eta + \boldsymbol{\Psi}_{\zeta_0}\right),$$

where $\boldsymbol{R}_\eta = \sum_{i=1}^{n}\sum_{t=1}^{T_i}[\boldsymbol{\eta}_{it} - \boldsymbol{f}_t(\boldsymbol{\eta}_{i,t-1},\boldsymbol{b}_i,\boldsymbol{\theta}_\eta)][\boldsymbol{\eta}_{it} - \boldsymbol{f}_t(\boldsymbol{\eta}_{i,t-1},\boldsymbol{b}_i,\boldsymbol{\theta}_\eta)]^T$.

Parameters in the Measurement Model. Following the work of many others (e.g., Lindley and Smith, 1972; Shi and Lee, 1998; Lee and Zhu, 2000), we specified the following conjugate priors for the distributions of $\psi_{\epsilon k}^{-1}$ and $\boldsymbol{\Lambda}_{yk}|\psi_{\epsilon k}$:

$$p(\boldsymbol{\Lambda}_{yk}|\psi_{\epsilon k}) \overset{D}{=} \mathrm{N}_{s+q}(\boldsymbol{\Lambda}_{0yk}, \psi_{\epsilon k}\boldsymbol{H}_{0\Lambda_{yk}}), \quad p(\psi_{\epsilon k}^{-1}) \overset{D}{=} \Gamma(\alpha_{0\epsilon k}, \beta_{0\epsilon k}), \tag{12}$$

where $\boldsymbol{\Lambda}_{yk}^T$ represents the kth row vector of $\boldsymbol{\Lambda}_y = (\boldsymbol{A}, \boldsymbol{\Lambda})$ for $k = 1, \ldots, p$. The components $\boldsymbol{\Lambda}_{0yk}$, $\boldsymbol{H}_{0\Lambda_{yk}}$, $\alpha_{0\epsilon k}$ and $\beta_{0\epsilon k}$ are all hyperparameters whose values are assumed to be known. Thus, we have

$$p(\psi_{\epsilon k}^{-1}|\boldsymbol{Y},\boldsymbol{H}) \sim \mathrm{Gamma}(\frac{1}{2}\sum_{i=1}^{n} T_i + \alpha_{0\epsilon k}, \beta_{\epsilon k}), \quad p(\boldsymbol{\Lambda}_{yk}|\psi_{\epsilon k},\boldsymbol{Y},\boldsymbol{H}) \sim \mathrm{N}[\boldsymbol{v}_k, \psi_{\epsilon k}\boldsymbol{\Upsilon}_k],$$

where $\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\sum_{i=1}^{n}\sum_{t=1}^{T_i} y_{itk}^2 - \boldsymbol{v}_k^T\boldsymbol{\Upsilon}_k^{-1}\boldsymbol{v}_k + \boldsymbol{\Lambda}_{0yk}^T\boldsymbol{H}_{0\Lambda_{yk}}^{-1}\boldsymbol{\Lambda}_{0yk})$, $\boldsymbol{\Upsilon}_k = (\boldsymbol{H}_{0\Lambda_{yk}}^{-1} + \sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{v}_{it}^*\boldsymbol{v}_{it}^{*T})^{-1}$, $\boldsymbol{v}_k = \boldsymbol{\Upsilon}_k(\sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{v}_{it}^* y_{itk} + \boldsymbol{H}_{0\Lambda_{yk}}^{-1}\boldsymbol{\Lambda}_{0yk})$ and $\boldsymbol{v}_{it}^* = (\boldsymbol{c}_{it}^T, \boldsymbol{\eta}_{it}^T)^T$.

Particularly, when $\boldsymbol{c}_{it} = 1$ and $\boldsymbol{A} = \boldsymbol{\mu}$, we consider the following conjugate priors for the distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}_k$:

$$p(\boldsymbol{\mu}|\boldsymbol{\mu}^0,\boldsymbol{\Sigma}_0) \sim \mathrm{N}(\boldsymbol{\mu}^0,\boldsymbol{\Sigma}_0), \quad p(\boldsymbol{\Lambda}_k|\psi_{\epsilon k}) \sim \mathrm{N}_q(\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\boldsymbol{H}_{0\Lambda_k}),$$

where $\boldsymbol{\Lambda}_k^T$ is the kth row vector of $\boldsymbol{\Lambda}$ for $k = 1, \ldots, p$, and the components $\boldsymbol{\mu}^0$, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Lambda}_{0k}$, $\boldsymbol{H}_{0\Lambda_k}$ are all hyperparameters whose values are assumed to be known. Then, we have

$$p(\boldsymbol{\mu}|\boldsymbol{Y},\boldsymbol{H},\boldsymbol{\Lambda},\boldsymbol{\Psi}_\epsilon) \sim \mathrm{N}(\boldsymbol{\mu}_\mu, \boldsymbol{\Omega}_\mu),$$

$$p(\psi_{\epsilon k}^{-1}|\boldsymbol{Y},\boldsymbol{H}) \sim \text{Gamma}(\frac{1}{2}\sum_{i=1}^{n} T_i + \alpha_{0\epsilon k}, \beta_{\epsilon k}), \quad p(\boldsymbol{\Lambda}_k|\psi_{\epsilon k},\boldsymbol{Y},\boldsymbol{H}) \sim \text{N}[\boldsymbol{v}_k, \psi_{\epsilon k}\boldsymbol{\Upsilon}_k],$$

where $\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\sum_{i=1}^{n}\sum_{t=1}^{T_i}(y_{itk}-\mu_k)^2 - \boldsymbol{v}_k^T\boldsymbol{\Upsilon}_k^{-1}\boldsymbol{v}_k + \boldsymbol{\Lambda}_{0k}^T\boldsymbol{H}_{0\Lambda_k}^{-1}\boldsymbol{\Lambda}_{0k})$, $\boldsymbol{\Upsilon}_k = (\boldsymbol{H}_{0\Lambda_k}^{-1} + \sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{\eta}_{it}\boldsymbol{\eta}_{it}^T)^{-1}$, $\boldsymbol{v}_k = \boldsymbol{\Upsilon}_k(\sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{\eta}_{it}(y_{itk}-\mu_k) + \boldsymbol{H}_{0\Lambda_k}^{-1}\boldsymbol{\Lambda}_{0k})$.

A.5 Step (i): Conditional Distribution $p(\boldsymbol{Y}_{\text{mis}}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{r},\boldsymbol{\varphi},\boldsymbol{C},\boldsymbol{U}_{\text{obs}})$

Since $\boldsymbol{y}_{it}$ are mutually independent for $i=1,\ldots,n$ and $t=1,\ldots,T_i$, $\boldsymbol{y}_{it,\text{mis}}$ are also independent of each other for $i=1,\ldots,n$ and $t=1,\ldots,T_i$. In addition, $\boldsymbol{\Psi}_\epsilon$ is assumed to be a diagonal matrix. Thus, $\boldsymbol{y}_{it,\text{mis}}$ is also independent of $\boldsymbol{y}_{it,\text{obs}}$, and we have $p(\boldsymbol{y}_{it,\text{mis}}|\boldsymbol{c}_{it},\boldsymbol{\eta}_{it},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{r}_{it},\boldsymbol{\theta}_y,\boldsymbol{\varphi}) \overset{\text{i.i.d}}{\sim} N(\boldsymbol{A}_{it,\text{mis}}\boldsymbol{c}_{it} + \boldsymbol{\Lambda}_{it,\text{mis}}\boldsymbol{\eta}_{it}, \boldsymbol{\Psi}_{\epsilon it,\text{mis}}) \times p(\boldsymbol{r}_{it}|\boldsymbol{y}_{(2)it},\boldsymbol{y}_{(1)it,\text{mis}}, \boldsymbol{U}_{it,\text{obs}},\boldsymbol{\varphi})$, where $\boldsymbol{A}_{it,\text{mis}}$ is a subvector of $\boldsymbol{A}$ with components corresponding to the missing components in $\boldsymbol{y}_{it,\text{mis}}$, $\boldsymbol{\Lambda}_{it,\text{mis}}$ is a submatrix of $\boldsymbol{\Lambda}$ with rows corresponding to the missing components in $\boldsymbol{y}_{it,\text{mis}}$, and $\boldsymbol{\Psi}_{\epsilon it,\text{mis}}$ is a submatrix of $\boldsymbol{\Psi}_\epsilon$ with rows and columns corresponding to the missing componetns in $\boldsymbol{y}_{it,\text{mis}}$. The conditional density is non–normal and non–standard due to the presence of $p(\boldsymbol{r}_{it}|\boldsymbol{y}_{(2)it},\boldsymbol{y}_{(1)it,\text{mis}},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{\varphi})$. As in some of the other steps, the MH algorithm is employed to draw observations from the posterior density $p(\boldsymbol{y}_{it,\text{mis}}|\boldsymbol{c}_{it},\boldsymbol{\eta}_{it},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{r}_{it},\boldsymbol{\theta}_y,\boldsymbol{\varphi})$ with the following steps. At the jth iteration with a current value $\boldsymbol{y}_{it,\text{mis}}^{(j)}$, a new candidate $\boldsymbol{y}_{it,\text{mis}}$ is generated from the normal distribution $N(\boldsymbol{y}_{it,\text{mis}}^{(j)}, \sigma_y^2\boldsymbol{U}_y)$, where $\boldsymbol{U}_y = (\boldsymbol{\Psi}_{\epsilon it,\text{mis}}^{-1} + \boldsymbol{\Delta}_y)^{-1}$ and $\boldsymbol{\Delta}_y = \frac{\partial^2 \log\{\text{pr}(\boldsymbol{r}_{it}|\boldsymbol{y}_{(2)it},\boldsymbol{y}_{(1)it,\text{mis}},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{\varphi})\}}{\partial\boldsymbol{y}_{it,\text{mis}}\partial\boldsymbol{y}_{it,\text{mis}}^T|_{y_{it,\text{mis}}=0}}$, and it is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{y}_{it,\text{mis}}|\boldsymbol{c}_{it},\boldsymbol{\eta}_{it},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{r}_{it},\boldsymbol{\theta}_y,\boldsymbol{\varphi})}{p(\boldsymbol{y}_{it,\text{mis}}^{(j)}|\boldsymbol{c}_{it},\boldsymbol{\eta}_{it},\boldsymbol{U}_{it,\text{obs}},\boldsymbol{r}_{it},\boldsymbol{\theta}_y,\boldsymbol{\varphi})}\right\}.$$

The variance σ_y^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

A.6 Step (j): Conditional Distribution $p(\boldsymbol{\varphi}|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})$

It follows from missingness mechanism Equation (6) and prior for $\boldsymbol{\varphi}$, $p(\boldsymbol{\varphi}) \stackrel{D}{=} N_d(\boldsymbol{\varphi}^0, \boldsymbol{H}_{\varphi}^0)$, that the conditional distribution $p(\boldsymbol{\varphi}|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})$ is proportional to

$$\exp\{\sum_{i=1}^{n}\sum_{t=1}^{T_i}[(\sum_{j=1}^{p} r_{itj})\boldsymbol{\varphi}^T\boldsymbol{x}_{it}^* - p\log(1+\exp(\boldsymbol{\varphi}^T\boldsymbol{x}_{it}^*))] - \frac{1}{2}(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{H}_{\varphi}^0)^{-1}(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)\},$$

where $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^T, \boldsymbol{\varphi}_2^T)^T$, and $\boldsymbol{x}_{it}^* = (\boldsymbol{x}_{it}^T, \boldsymbol{x}_{i,t-1}^T)^T$. It is easily seen that the conditional density is non-normal and non-standard. Again, the MH algorithm is adopted to sample observations from the posterior density $p(\boldsymbol{\varphi}|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})$ as follows. At the jth iteration with a current value $\boldsymbol{\varphi}^{(j)}$, a new candidate $\boldsymbol{\varphi}$ is generated from the normal distribution $N(\boldsymbol{\varphi}^{(j)}, \sigma_{\varphi}^2\boldsymbol{U}_{\varphi})$, where $\boldsymbol{U}_{\varphi} = (\frac{p}{4}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{x}_{it}^*\boldsymbol{x}_{it}^{*T} + (\boldsymbol{H}_{\varphi}^0)^{-1})^{-1}$, and it is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\varphi}|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})}{p(\boldsymbol{\varphi}^{(j)}|\boldsymbol{Y}_{\text{mis}}, \boldsymbol{U}_{\text{obs}}, \boldsymbol{Y}_{(2)\text{obs}}, \boldsymbol{r})}\right\}.$$

B. Hyperparameter Specification

The hyperparameter values of the prior distributions were specified as follows. For the priors in Step (8)–sampling from the condition distributions of the parameters in $\boldsymbol{\theta}$—we set $\boldsymbol{\mu}^0$ to a 8×1 vector of zeros and $\boldsymbol{\Sigma}_0$ to $0.5\boldsymbol{I}_8$ corresponding to hyperparameters in specifying the prior distribution of $\boldsymbol{\mu}$ (i.e., $\boldsymbol{A}$ in Equation (**??**)). For the unknown parameters λ_{kj} in the factor loading matrix $\boldsymbol{\Lambda}$, we set $\lambda_{0kj} = 0.8$ for $k = 2, 3, 4, 6, 7, 8$ and $j = 1, 2$, $H_{0\lambda_k}$ to 1.0. For the conjugate priors of the measurement error variances, we set $\alpha_{0\epsilon k}$ to 8 and $\beta_{0\epsilon k}$ to 10 to yield variance values that were relatively large and diffuse. For the priors of the dynamic parameters in $\boldsymbol{\theta}_\eta$, we set ρ_0 to 10 and $\boldsymbol{\Psi}_{\zeta_0}$ to $(\rho_0 - q - 1)\boldsymbol{R}_0^{-1}$, where $\boldsymbol{R}_0$ is the true value of $\boldsymbol{\Psi}_\zeta$.

With respect to the hyperparameters for the DP prior for $\boldsymbol{b}_i$, the following specifications were used. Based on the acceptance rates for the MH step for drawing posterior samples from $p(\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \boldsymbol{\pi}, \boldsymbol{Z}, \boldsymbol{L}|\boldsymbol{\theta}, \boldsymbol{H})$, we specified c_1 to 10, and c_2 to be 5.0 and 8.0 for the first two and last two elements of $\boldsymbol{b}_i$, respectively; and the diagonal elements in $\boldsymbol{\Psi}_{\mu_Z}$ to be 5.0. We set $\mu_{Z_{0j}}$ to .15 for $j = 1$ and 2 (i.e., corresponding to b_{11i} and b_{22i}) and to -.15 for $j = 3$ and 4 (i.e., corresponding to b_{12i} and b_{21i}). In terms of hyperparameters for the base distribution of α, we set a_1 to 10 and a_2 to 2 to yield large values of α (and consequently, more unique $\boldsymbol{Z}_i$ values) to capture some of the more subtle individual differences in these dynamic parameters.

In the MH steps (i.e., steps 4, 6, 7, 9 and 10), a diffuse prior was specified for the threshold parameters (see Step 6), so c_τ can be set to any arbitrary constant value without affecting the resultant posterior distributions of the threshold parameters. We further take $\sigma^2_{\tau 1} = 0.0017$, $\sigma^2_{\tau 2} = 0.0014$, $\sigma^2_{\tau 3} = 0.0016$, $\sigma^2_{\tau 4} = 0.0015$, $\sigma^2_{\tau 5} = \sigma^2_{\tau 6} = \sigma^2_{\tau 7} = 0.0014$, $\sigma^2_{\tau 8} = 0.0013$, $\sigma^2_y = 3.0$ and $\sigma^2_\varphi = 3.8$, giving the average acceptance rates 0.322, 0.276, 0.261, 0.286, 0.297, 0.296, 0.274, 0.290, 0.292 and 0.297, respectively.

C. Uninformative Extension to Illustrative Example I

This is a slight extension to Illustrative Example I (the coin toss example) described in the paper. Consider an uninformative special case of the beta prior, $p(\theta) \sim \text{Beta}(1, 1)$, in which θ is assumed to have uniform probabilities for all values on the interval of [0, 1]. The sensitivity of any subsequent modeling results to the prior may be illustrated by perturbing the prior: $p(\theta) \sim \text{Beta}(1, 1)$ via the perturbation scheme: $p(\theta|\omega) \sim \text{Beta}(\omega_1, \omega_2)$, where $\omega = (\omega_1, \omega_2)^T$. In this case, $\omega^0 = (1, 1)^T$ represents no perturbation. The perturbed likelihood function is given by

$$p(\theta|\omega) = \frac{\theta^{\omega_1 - 1}(1-\theta)^{\omega_2 - 1}}{B(\omega_1, \omega_2)},$$

where $B(.)$ denotes the beta function. The first- and second-order partial derivatives with respect to ω are given by

$$\frac{\partial \log p(\theta|\omega)}{\partial \omega_1} = \log\theta - \psi(\omega_1) + \psi(\omega_1 + \omega_2), \quad \frac{\partial \log p(\theta|\omega)}{\partial \omega_2} = \log(1-\theta) - \psi(\omega_2) + \psi(\omega_1 + \omega_2),$$

$$\frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_1^2} = -\dot{\psi}(\omega_1) + \dot{\psi}(\omega_1 + \omega_2), \quad \frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_2^2} = -\dot{\psi}(\omega_2) + \dot{\psi}(\omega_1 + \omega_2),$$

$$\frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_1 \partial \omega_2} = \dot{\psi}(\omega_1 + \omega_2),$$

where $\psi(x)$ is the digamma function (the logarithmic derivative of the gamma function), $\dot{\psi}(x)$ is the first-order partial derivative of $\psi(x)$. From properties of the digamma function, we have

$$\frac{\partial \log p(\theta|\omega)}{\partial \omega_1}|_{\omega^0} = \log\theta + 1, \quad \frac{\partial \log p(\theta|\omega)}{\partial \omega_2}|_{\omega^0} = \log(1-\theta) + 1,$$

$$E\left\{-\frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_1^2}|_{\omega^0}\right\} = 1,\ E\left\{-\frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_2^2}|_{\omega^0}\right\} = 1,\ E\left\{-\frac{\partial^2 \log p(\theta|\omega)}{\partial \omega_1 \partial \omega_2}|_{\omega^0}\right\} = \dot{\psi}(2),$$

which yields $G(\omega^0) = \begin{pmatrix} 1 & \dot{\psi}(2) \\ \dot{\psi}(2) & 1 \end{pmatrix}$. In this case, the two perturbations are not orthogonal to each other, and the same amount of perturbation is administered in the first two directions.

The score vectors summarizing changes in the Bayes factor and posterior means of $d(\theta)$ with respect to the perturbations, and Hessian matrix showing the curvatures in the ϕ-divergence function with respect to the perturbations, are denoted respectively as:

$$\nabla_{\mathrm{BF}} = E_\theta \begin{pmatrix} \log\theta + 1 \\ \log(1-\theta) + 1 \end{pmatrix}, \quad \nabla_{\mathrm{M_d}} = E_\theta \begin{pmatrix} (\log\theta + 1) d(\theta) \\ \{\log(1-\theta) + 1\} d(\theta) \end{pmatrix},$$

$$H_\phi = \ddot{\phi}(1) \left(E_\theta \begin{pmatrix} \log\theta + 1 \\ \log(1-\theta) + 1 \end{pmatrix}^{\otimes 2} - \begin{pmatrix} E_\theta(\log\theta) + 1 \\ E_\theta\{\log(1-\theta)\} + 1 \end{pmatrix}^{\otimes 2} \right)$$

where $\boldsymbol{q}^{\otimes 2} = \boldsymbol{q}\boldsymbol{q}^T$ for any vector $\boldsymbol{q} \in \mathbb{R}^{\mathcal{T}}$, and E_θ represents expectation taken with respect to the posterior distribution of θ (i.e., $p(\theta|z_1, z_2) \sim \mathrm{Beta}(z_1 + z_2 + 1, 2N - z_1 - z_2 + 1)$). It is impossible to obtain closed forms of the above three equations. However, MCMC approximations may be used to compute them as:

$$\nabla_{\mathrm{BF}} \approx \frac{1}{K_1} \sum_{\kappa=K_0+1}^{K_0+K_1} \begin{pmatrix} \log(\theta^{(\kappa)}) + 1 \\ \log(1-\theta^{(\kappa)}) + 1 \end{pmatrix}, \nabla_{\mathrm{M_d}} \approx \frac{1}{K_1} \sum_{\kappa=K_0+1}^{K_0+K_1} \begin{pmatrix} \{\log(\theta^{(\kappa)}) + 1\} d(\theta^{(\kappa)}) \\ \{\log(1-\theta^{(\kappa)}) + 1\} d(\theta^{(\kappa)}) \end{pmatrix},$$

$$H_\phi \approx \ddot{\phi}(1) \tfrac{1}{K_1} \sum_{\kappa=K_0+1}^{K_0+K_1} \left\{ \begin{pmatrix} \log(\theta^{(\kappa)}) + 1 \\ \log(1-\theta^{(\kappa)}) + 1 \end{pmatrix} - \tfrac{1}{K_1} \sum_{\kappa'=K_0+1}^{K_0+K_1} \begin{pmatrix} \log(\theta^{(\kappa')}) + 1 \\ \{\log(1-\theta^{(\kappa')})\} + 1 \end{pmatrix} \right\}^{\otimes 2}, \tag{13}$$

respectively, where $\{\theta^{(\kappa)} : \kappa = K_0 + 1, \cdots, K_0 + K_1\}$ are the observations generated from the posterior distribution of θ;, K_0 denotes the number of burn-in iterations, and K_1 denotes the number of additional iterations after burn-in.

It can be seen that the local influence measures vary as functions of the sampled values of $\theta^{(k)}$. Under such an uninformative beta prior, modeling results would likely not be overly sensitive to local perturbations to the prior in most scenarios. However, when the number of trials, N, is small, and the posterior distribution resembles the prior distribution closely, even slight perturbations to the prior may be influential in changing model fit (as revealed through the Bayes factor), as well as characteristics of the posterior distributions. Alternatively, if extreme proportions of successes are observed, the shape and the mean of the posterior distribution of θ would also differ from those associated with the "uniform-like" prior distribution. In this case, local perturbations to the prior may also yield relatively large values of local influence in some regions. Matlab code for this illustrative example can be downloaded as supplementary materials on the journal website.

D. Prior and Hyperparameter Choices for the Real Data Application

Similar to the prior and hyperparameter choices used in our simulation studies, a diffuse prior was specified for all of the threshold parameters. We also set $\boldsymbol{\mu}^0 = \tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_8$ in specifying prior of $\boldsymbol{\mu}$, $\boldsymbol{\Psi}_{\zeta_0} = (\rho_0 - q - 1)\tilde{\boldsymbol{\Psi}}_\zeta$ with $\rho_0 = 10$ in specifying prior of $\boldsymbol{\Psi}_\zeta$, and let the unknown components λ_{jk} of the factor loading matrix $\boldsymbol{\Lambda}$ to be $\tilde{\lambda}_{jk}$ and $H_{0\lambda_{jk}} = 1.0$ ($k = 1$ and $j = 2, 3$ and 4; $k = 2$ and $j = 6, 7$ and 8), where $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Psi}}_\zeta$ and $\tilde{\lambda}_{jk}$ are the auxiliary Bayesian estimates of $\boldsymbol{\mu}$, $\boldsymbol{\Psi}_\zeta$ and λ_{jk} obtained from non-informative prior inputs. For the conjugate priors of the measurement error variances, we set $\alpha_{0\epsilon j}$ to 8 and $\beta_{0\epsilon j}$ to 10 to yield variance values that were relatively large and diffuse for $j = 1, \ldots, 8$. We set $\mu_{Z_{0j}}$ to 0.05 for $j = 1$ and 2 (i.e., corresponding to b_{11i} and b_{22i}) and to -0.01 for $j = 3$ and 4 (i.e., corresponding to b_{12i} and b_{21i}), where $\mu_{Z_{0j}}$ is the jth component of the hyperparameter $\boldsymbol{\mu}_{Z_0}$; and $\Psi_{\mu_{Zj}}$ were set 0.01 for $j = 1, 2, 3, 4$ (i.e., corresponding to b_{11i}, b_{22i}, b_{12i} and b_{21i}), where $\Psi_{\mu_{Zj}}$ is the jth component of the hyperparameter $\boldsymbol{\Psi}_{\mu_Z}$. Furthermore, we set c_1 to 10, c_2 corresponding to b_{11i} and b_{22i} to be 5.0, and c_2 corresponding to b_{12i} and b_{22i} to be 2.0. With respect to hyperparameters for the prior distribution of α, we set a_1 to 10.0 and a_2 to 0.01 to yield large values of α to capture some of the more subtle individual differences in these dynamic parameters. In the Gibbs sampler, we set $\sigma^2_{\tau j} = 0.00012$ for $j = 1, \ldots, 3$, $\sigma^2_{\tau 4} = 0.00013$, $\sigma^2_{\tau 5} = 0.00032$, $\sigma^2_{\tau 6} = 0.0003$, $\sigma^2_{\tau 7} = 0.00018$, $\sigma^2_{\tau 8} = 0.0002$, $\sigma^2_y = 9.0$ and $\sigma^2_\varphi = 14.5$, which gave an average acceptance rate of 0.328.

References

Lee, S.-Y. and Zhu, H. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, 53:209–232.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:1–42.

Shi, J.-Q. and Lee, S.-Y. (1998). Using factor analysis to estimate parameters. *British Journal of Mathematical and Statistical Psychology*, 51:233–252.