Supplementary Information for **Bayesian Semiparametric Inference in Longitudinal Metabolomics Data: The EarlyBird Study**

Abhra Sarkar[1],
Ornella Cominetti[2],
Ivan Montoliu[2],
Joanne Hosking[3],
Jonathan Pinkney[3],
Francois-Pierre Martin[2],
David B. Dunson[4]

[1]Department of Statistics and Data Sciences, University of Texas at Austin,
2317 Speedway D9800, Austin, TX 78712-1823, USA
[2]Nestlé Research,
CH-1015 Lausanne, Switzerland
[3]Plymouth University Peninsula Schools of Medicine and Dentistry
Plymouth PL6 8BT, UK
[4]Department of Statistical Science, Duke University,
Box 90251, Durham, NC 27708-0251, USA

Supplementary Information presents a general modeling and inference framework for complex multivariate longitudinal data with ignorable missingness, designed to simplify the imputation tasks for both missing response and covariate values. Supplementary Information also details the MCMC algorithm used to sample from the posterior, post-processing model selection and variable selection procedures, a simulation study evaluating the proposed method in synthetic settings, and a few additional figures summarizing the results of the EarlyBird application and the simulation studies. The results for the simulation study correspond to the dataset that produced the median average RMSE using our method.

S.1 General Modeling Framework

The goal is to estimate the longitudinal evolution of a continuous response y and its relationship with p continuous covariates $\{x_j\}_{j=1}^{p}$. These variables are measured for n subjects at T time points. Let $\mathbf{y} = \{y_{it}\}_{i=1,t=1}^{n,T}$ denote the response values and $\mathbf{x} = \{x_{ijt}\}_{i=1,j=1,t=1}^{n,p,T}$ comprise measurements on the covariates. Both $\mathbf{y}$ and $\mathbf{x}$ involve missing values which need to be imputed. Let $\mathbf{y}_{\text{obs}}$ and $\mathbf{y}_{\text{mis}}$ denote the observed and the missing values in $\mathbf{y}$, respectively, and $\mathbf{m}_y$ denote the corresponding missingness indicators so that $m_{y,it} = 1$ if y_{it} is observed and $m_{y,it} = 0$ otherwise. Likewise, we define $\mathbf{x}_{\text{obs}}$, $\mathbf{x}_{\text{mis}}$ and $\mathbf{m}_x$. Also, let $\mathbf{m} = (\mathbf{m}_y, \mathbf{m}_x)$.

To address the imputation challenges presented by the missing values $\mathbf{x}_{\text{mis}}$, we consider a joint probability model for $(\mathbf{y}, \mathbf{x}, \mathbf{m})$. In addition to $(\mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m})$, time-invariant baseline predictors $\mathbf{x}_{\text{bl}}$ of arbitrary data types but with no missing values may be observed. We avoid placing a probability model on $\mathbf{x}_{\text{bl}}$ and for simplicity in notation, ignore such variables in the following exposition.

Let the joint distribution $p(\mathbf{y}, \mathbf{x}, \mathbf{m}|\boldsymbol{\omega})$ be indexed by the parameter $\boldsymbol{\omega}$. Let $p_0(\boldsymbol{\omega})$ denote a prior on $\boldsymbol{\omega}$. Inference on $\boldsymbol{\omega}$ may then be based on samples drawn from the posterior

$$p(\boldsymbol{\omega}|\mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) = \int\int p(\boldsymbol{\omega}, \mathbf{y}_{\text{mis}}, \mathbf{x}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) d\mathbf{y}_{\text{mis}} d\mathbf{x}_{\text{mis}}.$$

Sampling strategies can be greatly simplified by imputing the missing values $\mathbf{y}_{\text{mis}}$ and $\mathbf{x}_{\text{mis}}$ and working with the joint posterior $p(\boldsymbol{\omega}, \mathbf{y}_{\text{mis}}, \mathbf{x}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m})$ instead. Algorithm 1 describes a general strategy for sampling from such a joint posterior, where the missing values $(\mathbf{y}_{\text{mis}}, \mathbf{x}_{\text{mis}})$ are imputed from their full conditional

$$p(\mathbf{y}_{\text{mis}}, \mathbf{x}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) = p(\mathbf{y}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}, \mathbf{m}) p(\mathbf{x}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}).$$

Next, assume that the missing values are MAR. That is, $p(\mathbf{m}|\boldsymbol{\omega}, \mathbf{y}, \mathbf{x}) = p(\mathbf{m}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}})$. Also, assume $\boldsymbol{\omega}$ can be decomposed as $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ such that $\boldsymbol{\theta}$ fully characterizes the distributions of $(\mathbf{y}, \mathbf{x})$ whereas $\boldsymbol{\psi}$ separately collects the parameters characterizing the missing data mechanism given the observed values. Also, let them have independent priors. That is,

$$\begin{aligned} &p_0(\boldsymbol{\omega}) = p_0(\boldsymbol{\theta}) p_0(\boldsymbol{\psi}), \quad p(\mathbf{x}|\boldsymbol{\omega}) = p(\mathbf{x}|\boldsymbol{\theta}), \\ &p(\mathbf{y}|\boldsymbol{\omega}, \mathbf{x}) = p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}), \quad p(\mathbf{m}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}) = p(\mathbf{m}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}). \end{aligned}$$

The full conditional $p(\boldsymbol{\omega}|\mathbf{y}, \mathbf{x}, \mathbf{m})$ then factorizes as

$$p(\boldsymbol{\omega}|\mathbf{y}, \mathbf{x}, \mathbf{m}) \propto p_0(\boldsymbol{\omega}) p(\mathbf{y}, \mathbf{x}, \mathbf{m}|\boldsymbol{\omega}) = \{p_0(\boldsymbol{\theta}) p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta})\} \{p_0(\boldsymbol{\psi}) p(\mathbf{m}|\boldsymbol{\psi}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}})\}.$$

Likewise, we get

$$p(\mathbf{y}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}, \mathbf{m}) = p(\mathbf{y}_{\text{mis}}|\boldsymbol{\theta}, \mathbf{y}_{\text{obs}}, \mathbf{x}),$$

$$p(\mathbf{x}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) \propto p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}, \mathbf{x})p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}, \mathbf{x}_{\text{obs}}).$$

Here, $\boldsymbol{\psi}$ is not involved in any other step except in its own full conditional. Therefore, if interest lies primarily in inference on $\boldsymbol{\theta}$ the parameter $\boldsymbol{\psi}$ that characterizes the missingness mechanism can be ignored. Additionally, it is often natural to assume $\boldsymbol{\theta} = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_x)$ with

$$p_0(\boldsymbol{\theta}_y, \boldsymbol{\theta}_x) = p_0(\boldsymbol{\theta}_y)p_0(\boldsymbol{\theta}_x), \quad p(\mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta}_x), \quad p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}) = p(\mathbf{y}|\boldsymbol{\theta}_y, \mathbf{x}).$$

This leads to the factorization

$$p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}) \propto \{p_0(\boldsymbol{\theta}_y)p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_y)\}\{p_0(\boldsymbol{\theta}_x)p(\mathbf{x}|\boldsymbol{\theta}_x)\}.$$

The parameters $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$ may then be separately updated. Conjugate priors for $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$ will lead to closed form expressions for $p(\boldsymbol{\theta}_x|\mathbf{x})$ and $p(\boldsymbol{\theta}_y|\mathbf{y}, \mathbf{x})$, further facilitating posterior computation. Also, the full conditionals for $\mathbf{y}_{\text{mis}}$ and $\mathbf{x}_{\text{mis}}$ are simplified as

$$\begin{aligned} &p(\mathbf{y}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}, \mathbf{m}) = p(\mathbf{y}_{\text{mis}}|\boldsymbol{\theta}_y, \mathbf{y}_{\text{obs}}, \mathbf{x}), \\ &p(\mathbf{x}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) \propto p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}_y, \mathbf{x})p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}_x, \mathbf{x}_{\text{obs}}). \end{aligned}$$

Focusing on updating the parameters $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$, and imputing the missing values $\mathbf{y}_{\text{mis}}$ and $\mathbf{x}_{\text{mis}}$, but ignoring the missingness mechanism, the MCMC sampler thus takes the form described in Algorithm 2.

In this article, we additionally assume that the model parametrizations are such that

$$p(\mathbf{x}|\boldsymbol{\theta}_x) = p(\mathbf{x}_{\text{obs}}|\boldsymbol{\theta}_x)p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}_x), \tag{S.1}$$

$$p(\mathbf{y}|\boldsymbol{\theta}_y, \mathbf{x}) = p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}_y, \mathbf{x})p(\mathbf{y}_{\text{mis}}|\boldsymbol{\theta}_y, \mathbf{x}). \tag{S.2}$$

See Figure S.1a. The full conditionals of $\mathbf{y}_{\text{mis}}$ and the $\mathbf{x}_{\text{mis}}$ then get simplified to

$$\begin{aligned} &p(\mathbf{y}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}, \mathbf{m}) = p(\mathbf{y}_{\text{mis}}|\boldsymbol{\theta}_y, \mathbf{x}), \\ &p(\mathbf{x}_{\text{mis}}|\boldsymbol{\omega}, \mathbf{y}_{\text{obs}}, \mathbf{x}_{\text{obs}}, \mathbf{m}) \propto p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}_y, \mathbf{x})p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}_x). \end{aligned}$$

Finally, with $\boldsymbol{\theta}_x = \cup_{i,t}\boldsymbol{\theta}_{it,x}$ and $\boldsymbol{\theta}_y = \cup_{i,t}\boldsymbol{\theta}_{it,y}$, we let

$$p(\mathbf{x}|\boldsymbol{\theta}_x) = \textstyle\prod_{i=1}^n \prod_{t=1}^T \{p(\mathbf{x}_{it,\text{obs}}|\boldsymbol{\theta}_{it,x})p(\mathbf{x}_{it,\text{mis}}|\boldsymbol{\theta}_{it,x})\}, \tag{S.3}$$

$$p(\mathbf{y}|\boldsymbol{\theta}_y, \mathbf{x}) = \textstyle\prod_{i=1}^n \prod_{t=1}^T \left[\{p(y_{it,\text{obs}}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it})\}^{m_{y,it}}\{p(y_{it,\text{mis}}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it})\}^{(1-m_{y,it})}\right]. \tag{S.4}$$

See Figure S.1b. Equations (S.3) and (S.4) directly imply (S.1) and (S.2), respectively. The conditional independence relationships encoded in (S.3) and (S.4) further greatly simplify the tasks of imputing the missing values $\mathbf{y}_{\text{mis}}$ and $\mathbf{x}_{\text{mis}}$. The missing values $\mathbf{y}_{\text{mis}}$, for instance, can

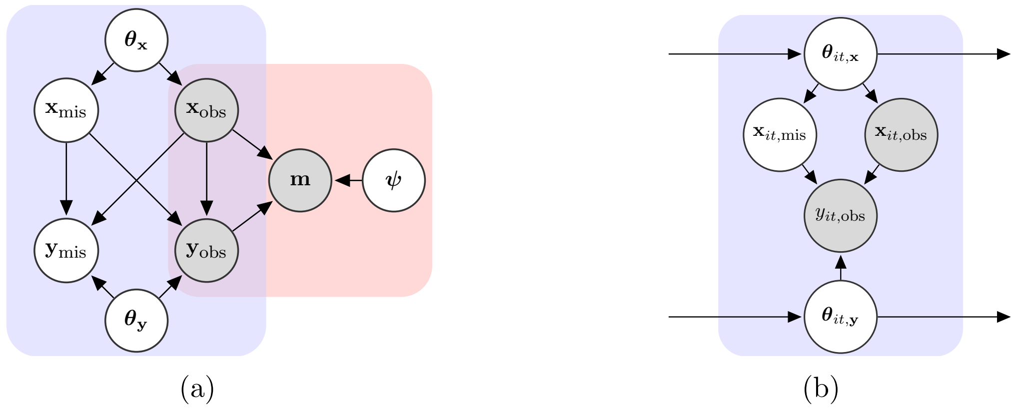

Figure S.1: Graphs showing the dependence-independence relationships in models to which Algorithm 3 applies. In (a), the shaded and the unshaded nodes represent observed and latent variables, respectively. The missingness mechanism, shown in the red rectangle, depends only on the observed values and hence can be ignored while inferring about the model of primary interest, shown in the blue rectangle. To be amenable to Algorithm 3, the model of interest must additionally be made up of slices, one for each individual i at each time point t. The longitudinal dependencies between successive slices are captured via dependencies in the underlying parameters $\boldsymbol{\theta}_{it,x}$ and $\boldsymbol{\theta}_{it,y}$. Within each slice, the observed and the missing covariate values are also independent of each other conditional on $\boldsymbol{\theta}_{it,x}$. Graph (b) focuses on one such slice, assuming, without loss of generality, that the response is observed.

be imputed by generating $y_{it,\mathrm{mis}}$ independently from $p(y_{it,\mathrm{mis}}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it})$. Also, the distribution used to impute the missing values $\mathbf{x}_{\mathrm{mis}}$ now factorizes as

$$p(\mathbf{x}_{\mathrm{mis}}|\boldsymbol{\theta}_y, \boldsymbol{\theta}_x, \mathbf{y}_{\mathrm{obs}}, \mathbf{x}_{\mathrm{obs}}) \propto \prod_{i=1}^{n} \prod_{t=1}^{T} [\{p(y_{it,\mathrm{obs}}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it,\mathrm{obs}}, \mathbf{x}_{it,\mathrm{mis}})\}^{m_{y,it}} p(\mathbf{x}_{it,\mathrm{mis}}|\boldsymbol{\theta}_{it,x})].$$

This allows $\mathbf{x}_{it,\mathrm{mis}}$ to be imputed independently of each other using efficient Metropolis-Hastings (MH) steps with proposal distribution $p(\mathbf{x}_{it,\mathrm{mis}}|\boldsymbol{\theta}_{it,x})$. Letting $\mathbf{x}_{it,\mathrm{mis}}^{prop}$ and $\mathbf{x}_{it,\mathrm{mis}}^{curr}$ denote the proposed and current values of $\mathbf{x}_{it,\mathrm{mis}}$, respectively, the acceptance probability of an MH move is given by

$$\min\left\{1, \frac{p(y_{it}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it,\mathrm{obs}}, \mathbf{x}_{it,\mathrm{mis}}^{prop})}{p(y_{it}|\boldsymbol{\theta}_{it,y}, \mathbf{x}_{it,\mathrm{obs}}, \mathbf{x}_{it,\mathrm{mis}}^{curr})}\right\}.$$

The MCMC sampling strategy outlined in Algorithm 2 then takes the form of Algorithm 3.

Generally speaking, Algorithm 3 exploits conditional independence relationships between the covariate and response values both within and across time points (Figure S.1). Models that directly associate covariate or response values at successive times points, such as vector auto-regressive models etc, are thus not amenable to Algorithm 3. Different classes of latent variable models, such as hidden Markov models, latent factor models, etc., can, however, be adapted to build longitudinal models that satisfy the assumptions of Algorithm 3. One such model is developed in the main paper.

Algorithm 1

1: Update $\boldsymbol{\omega}$ by sampling from $p(\boldsymbol{\omega}|\mathbf{y},\mathbf{x},\mathbf{m})$.
2: Impute $(\mathbf{y}_{\text{mis}},\mathbf{x}_{\text{mis}})$ by sampling from $p(\mathbf{y}_{\text{mis}}|\boldsymbol{\omega},\mathbf{y}_{\text{obs}},\mathbf{x},\mathbf{m})p(\mathbf{x}_{\text{mis}}|\boldsymbol{\omega},\mathbf{y}_{\text{obs}},\mathbf{x}_{\text{obs}},\mathbf{m})$.

Algorithm 2

1: Update $(\boldsymbol{\theta}_y,\boldsymbol{\theta}_x)$ by sampling from
   (a) $p(\boldsymbol{\theta}_x|\mathbf{x}) \propto p_0(\boldsymbol{\theta}_x)p(\mathbf{x}|\boldsymbol{\theta}_x)$,
   (b) $p(\boldsymbol{\theta}_y|\mathbf{y},\mathbf{x}) \propto p_0(\boldsymbol{\theta}_y)p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}_y)$,
2: Impute $(\mathbf{y}_{\text{mis}},\mathbf{x}_{\text{mis}})$ by sampling from $p(\mathbf{y}_{\text{mis}}|\boldsymbol{\theta}_y,\mathbf{y}_{\text{obs}},\mathbf{x})$. Then, impute $\mathbf{x}_{\text{mis}}$ by sampling from $p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}_y,\boldsymbol{\theta}_x,\mathbf{y}_{\text{obs}},\mathbf{x}_{\text{obs}}) \propto p(\mathbf{y}_{\text{obs}}|\boldsymbol{\theta}_y,\mathbf{x})p(\mathbf{x}_{\text{mis}}|\boldsymbol{\theta}_x,\mathbf{x}_{\text{obs}})$.

Algorithm 3

1: Update $(\boldsymbol{\theta}_y,\boldsymbol{\theta}_x)$ by sampling from
   (a) $p(\boldsymbol{\theta}_x|\mathbf{x}) \propto p_0(\boldsymbol{\theta}_x)p(\mathbf{x}|\boldsymbol{\theta}_x)$,
   (b) $p(\boldsymbol{\theta}_y|\mathbf{y},\mathbf{x}) \propto p_0(\boldsymbol{\theta}_y)p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}_y)$.
2: For $i=1,\ldots,n$ and $t=1,\ldots,T$, impute $y_{it,\text{mis}}$ by sampling from $p(y_{it,\text{mis}}|\boldsymbol{\theta}_{it,y},\mathbf{x}_{it})$. Then, impute $\mathbf{x}_{it,\text{mis}}$ by first proposing a new value $\mathbf{x}_{it,\text{mis}}^{prop} \sim p(\mathbf{x}_{it,\text{mis}}|\boldsymbol{\theta}_{it,x})$. If $y_{it}=y_{it,\text{mis}}$ is missing, replace the current value $\mathbf{x}_{it,\text{mis}}^{curr}$ by $\mathbf{x}_{it,\text{mis}}^{prop}$ with probability 1. If $y_{it}=y_{it,\text{obs}}$ is observed, replace the current value $\mathbf{x}_{it,\text{mis}}^{curr}$ by $\mathbf{x}_{it,\text{mis}}^{prop}$ with probability

$$\min\left\{1, \frac{p(y_{it,\text{obs}}|\boldsymbol{\theta}_{it,y},\mathbf{x}_{it,\text{obs}},\mathbf{x}_{it,\text{mis}}^{prop})}{p(y_{it,\text{obs}}|\boldsymbol{\theta}_{it,y},\mathbf{x}_{it,\text{obs}},\mathbf{x}_{it,\text{mis}}^{curr})}\right\}.$$

Conformity of the Model in the Main Paper to the General Framework Described Above. According to the model for the covariates proposed in the main paper, given $\boldsymbol{\theta}_x=(\boldsymbol{\mu}_x,\mathbf{b}_x,\boldsymbol{\Lambda},\boldsymbol{\eta},\boldsymbol{\sigma}_u^2)$, where $\boldsymbol{\mu}_x=\{\boldsymbol{\mu}_x\}_{t=1}^T$, $\mathbf{b}_\mathbf{x}=\{b_{x,i}\}_{i=1}^n$, $\boldsymbol{\Lambda}=\{\boldsymbol{\Lambda}_t\}_{t=1}^T$, $\boldsymbol{\eta}=\{\boldsymbol{\eta}_{it}\}_{i=1,t=1}^{n,T}$ with $\boldsymbol{\eta}_{it}^{\mathrm{T}}=[\eta_{it1},\ldots,\eta_{itq_t}]$ and $\boldsymbol{\sigma}_u^2=\{\boldsymbol{\sigma}_{u,t}^2\}_{t=1}^T$, we have

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}_x) &= \textstyle\prod_{i=1}^n\prod_{t=1}^T\{p(\mathbf{x}_{it,\text{obs}}|\boldsymbol{\theta}_x)p(\mathbf{x}_{it,\text{mis}}|\boldsymbol{\theta}_x)\} \\ &= \textstyle\prod_{i=1}^n\prod_{t=1}^T\left\{\prod_{j=1}^p \text{Normal}(x_{ijt}|\mu_{x,jt}+b_{x,ij}+\boldsymbol{\lambda}_{jt}^{\mathrm{T}}\boldsymbol{\eta}_{it},\sigma_{u,jt}^2)\right\}. \end{aligned}$$

When the latent factors are integrated out, the components of $\mathbf{x}_{it}$ would be correlated with each other and across time points and such factorization would not be possible. The latent factor formulation, therefore, ensures that the conditional independence relationships encoded in (S.3) and depicted in Figure S.1b, which simplify the imputation task, are satisfied.

Likewise, for the model for the response, given $\boldsymbol{\theta}_y=(\boldsymbol{\alpha},\mathbf{b}_y,\boldsymbol{\beta},\sigma_v^2)$ and $\mathbf{x}$, we have

$$p(\mathbf{y}|\boldsymbol{\theta}_y,\mathbf{x})=\textstyle\prod_{i=1}^n\prod_{t=1}^T \text{Normal}(y_{it}|\mu_{y,t}+b_{y,i}+\mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta},\sigma_v^2).$$

The conditional independence relationships encoded in (S.4), which simplify the imputation task, are thus satisfied.

S.2 Posterior Inference

S.2.1 Prior Hyper-parameters and MCMC Initializations

The hyper-parameters of the inverse-gamma priors on the different variance parameters were all set as $(a_\sigma, b_\sigma) = (1, 1)$. The hyper-parameters for the MGPS priors were chosen as $a_{\lambda,1} = a_{\alpha,1} = 1, a_{\lambda,h} = a_{\alpha,h} = 3, h \geq 2$ which make the distributions of $\tau_{\lambda,th}^{-1}$ and $\tau_{\alpha,h}^{-1}$ stochastically increasing with increasing h on large intervals around 0 (Durante, 2017), facilitating greater shrinkage as h increases. We set $\nu_\lambda = 1$. The mean process $\boldsymbol{\mu}_{x,t}$ was initialized at $\mathbf{0}$ for all t. We initialized all regression coefficients and all random effects at 0 and all variance parameters at 1. All λ_{tjh} and $\eta_{i\ell}$ were also initialized at 0. The parameters $\phi_{\lambda,tjh}$, $\tau_{\lambda,th}$, $\delta_{\lambda,t\ell}$, $\tau_{\alpha,k}$, $\delta_{\alpha,k}$, $\psi_{\beta,j}$, $\phi_{\beta,j}$ and τ_β were all initialized at 1. While constructing the predictor matrix $\mathbf{P}_{s,t} = [\mathbf{p}_{s,1}, \ldots, \mathbf{p}_{s,12}]^{\mathrm{T}}$, where $\mathbf{p}_{s,t}^{\mathrm{T}} = [1, t, \ldots, t^s]$, we transformed each time point $t = 1, \ldots, 12$ to $\widetilde{t} = 2(t - 6.5)$ and then standardized the columns of $\mathbf{P}_{s,t}$.

S.2.2 Sampling from the Posterior

In what follows, $\boldsymbol{\zeta}$ denotes a generic variable that collects all other parameters of the model except the ones being updated, including observed data $(\mathbf{y}_{\mathrm{obs}}, \mathbf{x}_{\mathrm{obs}}, \mathbf{m})$.

1. **Updating the model parameters:** The parameters characterizing the response generating process are updated as

$$
\begin{aligned}
(\boldsymbol{\alpha}|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_s(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha),\\
(\boldsymbol{\beta}|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta),\\
(b_{y,i}|\boldsymbol{\zeta}) &\sim \mathrm{Normal}(\mu_{y,b,i}, \sigma_{y,b,i}^2),\\
(\sigma_v^2|\boldsymbol{\zeta}) &\sim \text{Inv-Ga}\left(a_{v,\sigma} + nN/2, b_{v,\sigma} + \textstyle\sum_{i=1}^n \mathbf{r}_{v,i}^{\mathrm{T}}\mathbf{r}_{v,i}/2\right),
\end{aligned}
$$

where $\boldsymbol{\mu}_\alpha = \boldsymbol{\Sigma}_\alpha \mathbf{P}^{\mathrm{T}}\mathbf{r}_\alpha$, $\boldsymbol{\Sigma}_\alpha = (\sigma_v^{-2} n\mathbf{P}^{\mathrm{T}}\mathbf{P} + D_\alpha^{-1})^{-1}$, $\mathbf{P}^{\mathrm{T}} = [\mathbf{p}_{s,t}, \ldots, \mathbf{p}_{s,t}]$, $\mathbf{D}_\alpha^{-1} = \mathrm{diag}\{\tau_{\alpha,1}, \ldots, \tau_{\alpha,s}\}$, $\mathbf{r}_\alpha = \sum_i(\mathbf{y}_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - b_{y,i}\mathbf{1})$, $\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta\left(\sigma_v^{-2}\sum_i \mathbf{x}_i\mathbf{r}_{\beta,i} + \mathbf{D}_\beta^{-1}\boldsymbol{\mu}_{0,\beta}\right)$, $\boldsymbol{\Sigma}_\beta = \left(\sigma_v^{-2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} + \mathbf{D}_\beta^{-1}\right)^{-1}$, $\mathbf{D}_\beta^{-1} = \mathrm{diag}\{\psi_{\beta,1}\phi_{\beta,1}^2\tau_\beta, \ldots, \psi_{\beta,p}\phi_{\beta,p}^2\tau_\beta\}$, $\mathbf{r}_{\beta,i} = (\mathbf{y}_i - \boldsymbol{\mu}_{y,i} - b_{y,i}\mathbf{1})$, $\mu_{y,b,i} = \sigma_{y,b,i}^2(\sigma_v^{-2}\sum_t r_{y,b,it})$, $\sigma_{y,b,i}^2 = (n\sigma_v^{-2} + \sigma_{y,b}^{-2})^{-1}$, $r_{y,b,it} = (y_{it} - \mu_{y,it} - \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta})$, $\mathbf{r}_{v,i} = (\mathbf{y}_i - \boldsymbol{\mu}_{y,i} - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - b_{y,i}\mathbf{1})$. The associated prior parameters are updated as

$$
\begin{aligned}
(\delta_{\alpha,h}|\boldsymbol{\zeta}) &\sim \mathrm{Ga}\{a_{\alpha,h} + (s-h+1)/2, 1 + \textstyle\sum_{\ell=h}^s \tau_{\alpha,\ell}^{(h)}\alpha_\ell^2/2\},\\
(\psi_{\beta,j}|\boldsymbol{\zeta}) &\sim \mathrm{giG}(1/2, \beta_j^2\phi_{\beta,j}^{-1}\tau_\beta^{-2}, 1),\\
(\tau_\beta|\boldsymbol{\zeta}) &\sim \mathrm{giG}(pa_\beta - p, 2\textstyle\sum_j |\beta_j|\,\phi_{\beta,j}^{-1}, 1),\\
\phi_{\beta,j} &= T_{\beta,j}/\textstyle\sum_j T_{\beta,j} \text{ with } (T_{\beta,j}|\boldsymbol{\zeta}) \sim \mathrm{giG}(a_\beta - 1, 2\,|\beta_j|\,, 1),\\
(\sigma_{y,b}^2|\boldsymbol{\zeta}) &\sim \text{Inv-Ga}\left(a_{y,b,\sigma} + n/2, b_{y,b,\sigma} + \textstyle\sum_{i=1}^n b_{y,i}^2/2\right).
\end{aligned}
$$

Here $\mathrm{giG}(\lambda,\chi,\psi)$ denotes a generalized inverse Gaussian distribution with pdf proportional to $x^{\lambda-1}\exp\{-0.5(\chi/x+\psi x)\}$.

The parameters characterizing the covariate generating process are updated as

$$\begin{aligned}
(\boldsymbol{\mu}_{x,t}|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_p(\boldsymbol{\mu}_{x,\mu,t},\boldsymbol{\Sigma}_{x,\mu,t}),\\
(\boldsymbol{\eta}_i|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_q(\boldsymbol{\mu}_{\eta,i},\boldsymbol{\Sigma}_{\eta,i}),\\
(\boldsymbol{\lambda}_{tj}|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_q(\boldsymbol{\mu}_{\lambda,tj},\boldsymbol{\Sigma}_{\lambda,tj}),\\
(\mathbf{b}_{x,i}|\boldsymbol{\zeta}) &\sim \mathrm{MVN}_p(\boldsymbol{\mu}_{x,b,i},\boldsymbol{\Sigma}_{x,b,i}),\\
(\sigma^2_{u,jt}|\boldsymbol{\zeta}) &\sim \text{Inv-Ga}(a_{u,\sigma}+n/2, b_{u,\sigma}+\textstyle\sum_i r^2_{u,ijt}/2).
\end{aligned}$$

where $\boldsymbol{\mu}_{x,\mu,t}=\boldsymbol{\Sigma}_{x,\mu,t}(\sum_i \boldsymbol{\Sigma}_{x,t}^{-1}\mathbf{r}_{x,\mu,it}+\boldsymbol{\Delta}_{\epsilon,t}^{-1}\boldsymbol{\mu}_{x,t-1}+\boldsymbol{\Delta}_{\epsilon,t}^{-1}\boldsymbol{\mu}_{x,t+1})$, $\boldsymbol{\Sigma}_{x,\mu,t}=(n\boldsymbol{\Delta}_{u,t}^{-1}+2\boldsymbol{\Delta}_{\epsilon,t}^{-1})^{-1}$, $\mathbf{r}_{x,\mu,t}=(\mathbf{x}_{it}-\mathbf{b}_{x,i}-\boldsymbol{\Lambda}_t\boldsymbol{\eta}_i)$, $\boldsymbol{\mu}_{\eta,i}=\boldsymbol{\Sigma}_{\eta,i}\sum_t \boldsymbol{\Lambda}_t^{\mathrm{T}}\boldsymbol{\Delta}_{u,t}^{-1}\mathbf{r}_{\eta,it}$, $\boldsymbol{\Sigma}_{\eta,i}=(\sum_t\boldsymbol{\Lambda}_t^{\mathrm{T}}\boldsymbol{\Delta}_{u,t}^{-1}\boldsymbol{\Lambda}_t+\mathbf{I}_q)^{-1}$, $\mathbf{r}_{\eta,it}=(\mathbf{x}_{it}-\boldsymbol{\mu}_{x,t}-\mathbf{b}_{x,i})$, $\boldsymbol{\mu}_{\lambda,tj}=\boldsymbol{\Sigma}_{\lambda,tj}\sigma_{u,jt}^{-2}\sum_i r_{\lambda,ijt}\boldsymbol{\eta}_i$, $\boldsymbol{\Sigma}_{\lambda,tj}=(\sigma_{u,jt}^{-2}\sum_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}+\mathbf{D}_{\lambda,tj}^{-1})^{-1}$, $r_{\lambda,ijt}=(x_{ijt}-b_{x,ij})$, $\mathbf{D}_{\lambda,tj}^{-1}=\mathrm{diag}\{\phi_{\lambda,tj1}\tau_{\lambda,t1},\ldots,\phi_{\lambda,tjq}\tau_{\lambda,tq}\}$, $\boldsymbol{\mu}_{x,b,i}=\boldsymbol{\Sigma}_{x,b,i}\left(\sum_t\boldsymbol{\Sigma}_{x,t}^{-1}\mathbf{r}_{x,b,it}\right)$, $\boldsymbol{\Sigma}_{x,b,i}=(\sum_t\boldsymbol{\Delta}_{u,t}^{-1}+\boldsymbol{\Sigma}_{x,b}^{-1})^{-1}$, $\mathbf{r}_{x,b,it}=(\mathbf{x}_{it}-\boldsymbol{\mu}_{x,t}-\boldsymbol{\Lambda}_t\boldsymbol{\eta}_i)$, and $r_{u,ijt}=(x_{ijt}-\mu_{x,jt}-b_{x,it}-\boldsymbol{\lambda}_{tj}^{\mathrm{T}}\boldsymbol{\eta}_i)$.

The associated prior parameters are updated as

$$\begin{aligned}
(\phi_{\lambda,tjh}|\boldsymbol{\zeta}) &\sim \mathrm{Ga}\{(\nu_\lambda+1)/2,(\nu_\lambda+\tau_{\lambda,th}\lambda^2_{jh})/2\},\\
(\delta_{\lambda,th}|\boldsymbol{\zeta}) &\sim \mathrm{Ga}\{a_{\lambda,h}+p(q-h+1)/2, 1+\textstyle\sum_{\ell=1}^q\tau^{(h)}_{\lambda,t\ell}\sum_{j=1}^p\phi_{\lambda,tj\ell}\lambda^2_{tj\ell}/2\},\\
(\sigma^2_{x,b,j}|\boldsymbol{\zeta}) &\sim \text{Inv-Ga}\left(a_{x,b,\sigma}+n/2, b_{x,b,\sigma}+\textstyle\sum_{i=1}^n b^2_{x,ij}/2\right),\\
(\sigma^2_{\epsilon,j}|\boldsymbol{\zeta}) &\sim \text{Inv-Ga}\left\{a_{\epsilon,\sigma}+nN/2, b_{\epsilon,\sigma}+\textstyle\sum_{i=1}^n\sum_{t=1}^T(\mu_{x,jt}-\mu_{x,j,t-1})^2/2\right\},
\end{aligned}$$

where $\tau^{(h)}_{\lambda,t\ell}=\prod_{s=1,s\neq h}^{\ell}\delta_{\lambda,ts}$.

2. **Imputing $(\mathbf{y}_{\mathbf{mis}},\mathbf{x}_{\mathbf{mis}})$:** The missing values in $\mathbf{y}$ are imputed by sampling each $y_{it,\mathrm{mis}}$ independently from

$$(y_{it,\mathrm{mis}}|\boldsymbol{\zeta})\sim\mathrm{Normal}(y_{it,\mathrm{mis}}|\mu_{y,t}+b_{y,i}+\mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta},\sigma_v^2).$$

To impute $\mathbf{x}_{\mathrm{mis}}$, we first propose candidates by sampling the missing values $x^{prop}_{ijt,\mathrm{mis}}$ independently from

$$(x^{prop}_{ijt,\mathrm{mis}}|\boldsymbol{\zeta})\sim\mathrm{Normal}(x_{ijt,\mathrm{mis}}|\mu_{x,jt}+b_{x,i}+\boldsymbol{\lambda}_{jt}^{\mathrm{T}}\boldsymbol{\eta}_i,\sigma^2_{u,jt}).$$

Let $\mathbf{x}_{it,curr}$ and $\mathbf{x}_{it,prop}$ denote the values of $\mathbf{x}_{it}$ with common observed component $\mathbf{x}_{it,\mathrm{obs}}$, and current and proposed missing components $\mathbf{x}^{curr}_{it,\mathrm{mis}}$ and $\mathbf{x}^{prop}_{it,\mathrm{mis}}$, respectively. If y_{it} is missing, $\mathbf{x}^{curr}_{it,\mathrm{mis}}$ is replaced by $\mathbf{x}^{prop}_{it,\mathrm{mis}}$ with probability 1. If y_{it} is observed, $\mathbf{x}^{curr}_{it,\mathrm{mis}}$

is replaced by $\mathbf{x}_{it,\text{mis}}^{prop}$ with probability

$$\min\left\{1, \frac{\text{Normal}(y_{it,\text{obs}}|\mu_{y,t} + b_{y,i} + \mathbf{x}_{it,prop}^{\mathrm{T}}\boldsymbol{\beta}, \sigma_v^2)}{\text{Normal}(y_{it,\text{obs}}|\mu_{y,t} + b_{y,i} + \mathbf{x}_{it,curr}^{\mathrm{T}}\boldsymbol{\beta}, \sigma_v^2)}\right\}.$$

We implemented the model in R. In simulation experiments described in the Supplementary Information and the EarlyBird analysis discussed in the main paper, 3000 MCMC iterations with the initial 1000 discarded as burn-in produced stable estimates of model components and parameters, with diagnostics indicative of fast convergence and good mixing. With $n = 130$ subjects and $T = 12$ time points, on an ordinary desktop, $3,000$ MCMC iterations took approximately 20 minutes to run.

S.2.3 Post-processing Model Selection

To compare model fits between competing Bayesian models, we estimated the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) and log pseudo marginal likelihood (LPML) (Geisser and Eddy, 1979). Smaller values of DIC and larger values of LPML are indicative of better model fits. In describing these two measures and their posterior sample based estimates below, we assume first the data set consist of response values $\mathbf{y} = \{y_i\}_{i=1}^n$ only, with $f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta})$.

Letting $D(\boldsymbol{\theta}) = -2\ \text{log}f(\mathbf{y}|\boldsymbol{\theta})$ denote the deviance, the DIC is defined as $DIC = 2\mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}D(\boldsymbol{\theta}) - D(\mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}\boldsymbol{\theta})$. Based on M samples $\{\boldsymbol{\theta}_m\}_{m=1}^M$ drawn from the posterior $\overline{D}(\boldsymbol{\theta})$, the DIC can be estimated as $DIC \mathrel{\widehat{=}} 2\overline{D}(\boldsymbol{\theta}) - D(\overline{\boldsymbol{\theta}})$ where $\overline{D}(\boldsymbol{\theta}) = \sum_{m=1}^M D(\boldsymbol{\theta}_m)/M$ and $\overline{\boldsymbol{\theta}} = \sum_{m=1}^M \boldsymbol{\theta}_m/M$. LPML is defined as $LPML = \sum_{i=1}^n \log\ (CPO_i)/n$ with the conditional predictive ordinates CPO_i given by

$$CPO_i = f(y_i|\mathbf{y}_{-i}) = \left\{\int \frac{1}{f(y_i|\boldsymbol{\theta})} f(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}\right\}^{-1} = \left\{\mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}} \frac{1}{f(y_i|\boldsymbol{\theta})}\right\}^{-1}.$$

Based on the posterior samples, CPO_i can be estimated as $\left\{\sum_{m=1}^M f(y_i|\boldsymbol{\theta}^{(m)})^{-1}/M\right\}^{-1}$.

The conditional independence relationships encoded in our proposed latent factor based model make adjusting these measures to our longitudinal setting, including incorporating covariates, relatively straightforward. Instead of $f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta})$, we now use $f(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{t=1}^T f(y_{it}|\boldsymbol{\theta}_y, \mathbf{x}_{it}) f(\mathbf{x}_{it}|\boldsymbol{\theta}_x)$. Exact adjustments for missing data, however, involve complicated integrals. Instead, at each MCMC iteration, we plugged in the imputed values of the missing data to favor computational simplicity.

S.2.4 Post-processing Variable Selection

Under the continuous shrinkage priors, posterior samples of $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ are never exactly zero. We can further process the posterior estimates using loss function-based criteria to obtain a sparse $\boldsymbol{\gamma}$ with exact zeros to facilitate variable selection. We adapt to the general recipe prescribed in Hahn and Carvalho (2015).

Let $\widehat{z}$ be a generic notation for the posterior mean of any variable z. Consider the problem of predicting a future data matrix $\widetilde{\mathbf{y}}$ at some fixed design tensor $\widetilde{\mathbf{x}}$. A plausible, convenient choice for $\widetilde{\mathbf{x}}$ is given by the design tensor constructed from $\mathbf{x}_{\rm obs}, \widehat{\mathbf{x}}_{\rm mis}$. The predictive distribution of $\widetilde{\mathbf{y}}$ is given by $p(\widetilde{\mathbf{y}}|\widetilde{\mathbf{x}}, \mathbf{y}_{\rm obs}, \mathbf{x}_{\rm obs}) = \int p(\widetilde{\mathbf{y}}|\widetilde{\mathbf{x}}, \boldsymbol{\theta}_y)p(\boldsymbol{\theta}_y|\mathbf{y}_{\rm obs}, \mathbf{x}_{\rm obs})d\boldsymbol{\theta}_y$, where $\boldsymbol{\theta}_y = (\boldsymbol{\gamma}, \sigma^2_{y,b}, \sigma^2_v)$. An optimal $\boldsymbol{\gamma}$ is then obtained as $\boldsymbol{\gamma}_{\rm opt} = \arg\min_{\widetilde{\boldsymbol{\gamma}}} \mathcal{L}_0(\widetilde{\boldsymbol{\gamma}})$, where $\mathcal{L}_0(\widetilde{\boldsymbol{\gamma}}) = \int \mathcal{L}_0(\widetilde{\mathbf{y}}, \widetilde{\boldsymbol{\gamma}})p(\widetilde{\mathbf{y}}|\widetilde{\mathbf{x}}, \mathbf{y}_{\rm obs}, \mathbf{x}_{\rm obs})d\widetilde{\mathbf{y}}$ is the expected value of a loss function $\mathcal{L}_0(\widetilde{\mathbf{y}}, \widetilde{\boldsymbol{\gamma}})$, the expectation taken with respect to the posterior predictive distribution based on the observed data. Evaluating the integral and dropping the constant terms, for $\mathcal{L}_0(\widetilde{\mathbf{y}}, \widetilde{\boldsymbol{\gamma}}) = \lambda \sum_j 1\{\widetilde{\gamma}_j \neq 0\} + (nT)^{-1} \sum_{i,t}(\widetilde{\mathbf{x}}_{it}^{\rm T}\widetilde{\boldsymbol{\gamma}} - \widetilde{y}_{it})^2$, where the first term penalizes the inclusion of too many covariates and the second term quantifies the squared error prediction loss of the predictor defined by $\boldsymbol{\gamma}$, we get $\mathcal{L}_0(\widetilde{\boldsymbol{\gamma}}) = \lambda \sum_j 1\{\widetilde{\gamma}_j \neq 0\} + (nT)^{-1} \sum_{i,t}(\widetilde{\mathbf{x}}_{it}^{\rm T}\widetilde{\boldsymbol{\gamma}} - \widetilde{\mathbf{x}}_{it}^{\rm T}\widehat{\boldsymbol{\gamma}})^2$. Finally, the computationally intractable ℓ_0 penalty is replaced by locally adaptive ℓ_1 approximations as $\mathcal{L}_1(\widetilde{\boldsymbol{\gamma}}) = \lambda \sum_j |\widetilde{\gamma}_j| \,/\, |\widehat{\gamma}_j| + (nT)^{-1} \sum_{i,t}(\widetilde{\mathbf{x}}_{it}^{\rm T}\widetilde{\boldsymbol{\gamma}} - \widetilde{\mathbf{x}}_{it}^{\rm T}\widehat{\boldsymbol{\gamma}})^2$. Minimizing $\mathcal{L}_1(\widetilde{\boldsymbol{\gamma}})$ using the `lars` algorithm (Efron *et al.*, 2004) produces a sequence of sparse solutions $\boldsymbol{\gamma}_\lambda$ for a range of λ values. The excess error for such a sparse estimator $\boldsymbol{\gamma}_\lambda$ is given by

$$\psi_\lambda = \sqrt{(nT)^{-1} \textstyle\sum_{i,t}(\widetilde{\mathbf{x}}_{it}^{\rm T}\boldsymbol{\gamma}_\lambda - \widetilde{\mathbf{x}}_{it}^{\rm T}\boldsymbol{\gamma})^2 + \sigma^2_{y,b} + \sigma^2_v} - \sqrt{\sigma^2_{y,b} + \sigma^2_v}.$$

From the posterior samples of $\boldsymbol{\gamma}, \sigma^2_{y,b}, \sigma^2_v$, we can then obtain samples of ψ_λ and construct 90% credible intervals of ψ_λ for different λ. Finally, we report a single sparse estimate $\boldsymbol{\gamma}_\lambda$ that corresponds to the smallest model whose credible interval contains $E\psi_0$ for the fully saturated model.

S.3 Simulation Experiments

Simulation experiments are designed to evaluate the performance of our method in assessing various aspects of the data generating processes, including the selection of the true predictors. The design closely mimics the EarlyBird dataset analyzed in the main paper. We considered $n = 129$ subjects, $T = 12$ time points, $p_{\mathrm{bl}} = 4$ baseline predictors, and $p = 88$ continuous covariates evolving over time as the response. The first baseline predictor was binary, the remaining ones were continuous. The binary predictor comprised $n_1 = 92$ ones and the rest $n_0 = 37$ zeros, mimicking the variable 'gender of the subjects' in the EarlyBird study. The continuous baseline predictors are generated by sampling them independently from Normal$(0, 1)$ distributions. Let $\widehat{z}$ denote the posterior mean of any variable z for the EarlyBird data. The time varying predictors are simulated as $\mathbf{x}_{it} = \widehat{\boldsymbol{\mu}}_{x,t} + \widehat{\mathbf{b}}_{x,i} + \boldsymbol{\xi}_{it}$ with $\boldsymbol{\xi}_{it} \sim \mathrm{MVN}_p(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_{x,t})$. We then generated y according to the model $y_{it} = \sum_{\ell=0}^{5} \widehat{\alpha}_\ell f(t, \ell) + \widehat{b}_{y,i} + \mathbf{x}_{it}^{\mathrm{T}} \widehat{\boldsymbol{\beta}} + v_{it}$, where $f(t, \ell)$ is a normalized version of t^ℓ obtained by first transforming each time point $t = 1, \ldots, 12$ to $\widetilde{t} = 2(t - 6.5)$ and then standardizing the resulting columns of $\widetilde{t}^\ell$. The errors v_{it} were generated as $v_{it} \sim \mathrm{Normal}(0, \widehat{\sigma}_v^2)$. We then applied our method to estimate the regression coefficients $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$.

We evaluated the performances of our latent factor based method (BSP-LF), its adaptation where the covariance matrices of the covariates were restricted to all be diagonal (BSP-Diag) and the lme method in estimating $\boldsymbol{\gamma}$ in terms of their root mean squared errors $\mathrm{RMSE}(\gamma_j) = \{\widehat{\sigma}_{\gamma_j}^2 + (\widehat{\gamma}_j - \gamma_j)^2\}^{1/2}$. Of the 92 total components of $\boldsymbol{\gamma}$, the proposed BSP-LF method outperforms the lme method across all but 10 components, often very significantly (Figure S.2). The BSP-LF method also outperformed the BSP-Diag method for 70 components of $\boldsymbol{\gamma}$ (Figure S.2). The plots of 90% confidence intervals obtained by the lme method (Figure S.2, third row) and 90% posterior credible regions obtained by the proposed BSP-LF method (Figure S.2, fourth row) and the restrictive BSP-Diag method (Figure S.2, fifth row) provide a general idea of the efficacy of the three methods. Figure S.3 shows the average absolute bias in estimating the missing values in the response and the covariate values using the three methods. The proposed BSP-LF method significantly outperforms the alternatives.

Additional figures summarizing the results of the simulation studies are presented in the Supplementary Information (Figures S.4, S.5, S.6, S.7, S.8, S.10, S.12). Comparison between the results for the EarlyBird study and results for the simulated datasets again shows our method's numerical stability and reproducibility (compare, for instance, the parameter estimates in panels (a) and (c) in Figure 5 and estimates in the fourth row in Figure S.2).

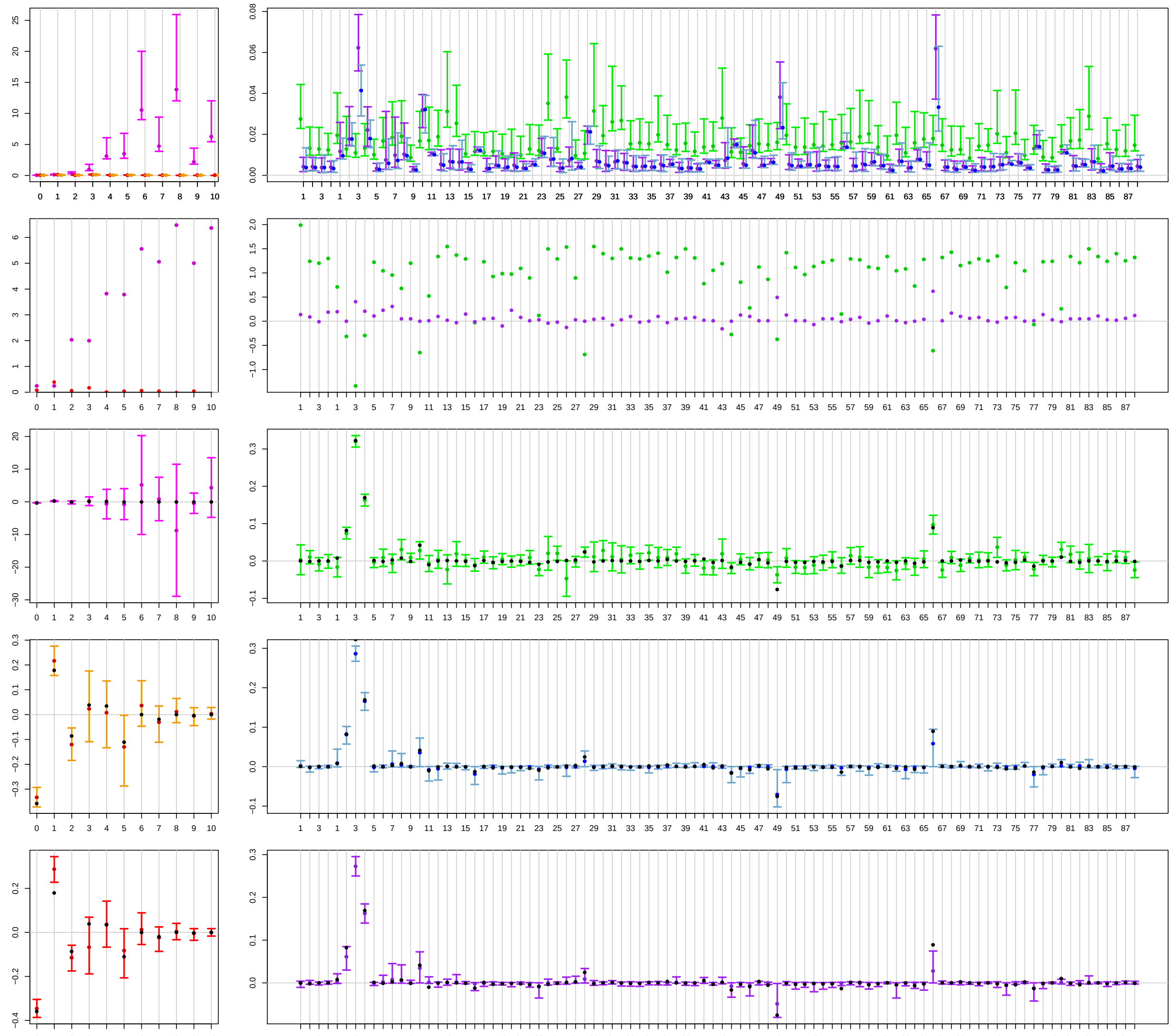

Figure S.2: Results for the simulation study: The top row shows the 5%, 50% and 95% quantiles of RMSEs for estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the lme method (magenta and green), our proposed BSP-LF method (orange and blue), and its diagonal covariance matrix adaptation BSP-Diag (red and purple). The second row shows, the logarithms of the ratios of the median RMSEs for the lme method and the BSP-LF method (in magenta and green), and the logarithms of the ratios of the median RMSEs for the BSP-Diag method and the BSP-LF method (in red and purple). The third row shows the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and their 90% confidence intervals obtained by the lme method for the dataset that corresponds to the median average RMSE. The fourth and fifth rows show respectively the estimated posterior means of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and their 90% credible intervals obtained by the BSP-LF method and the BSP-Diag method for the datasets that correspond to the corresponding median average RMSE. In the last three rows, the true values are shown as black circles.

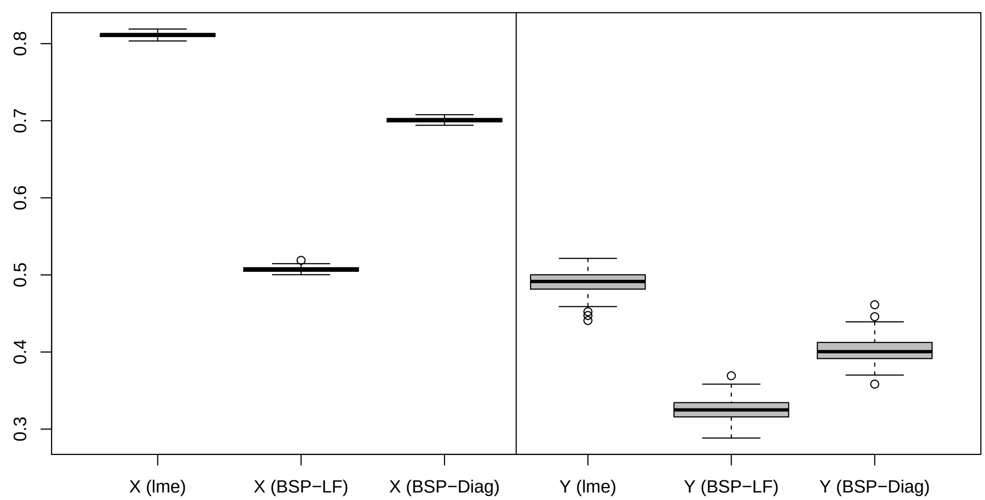

Figure S.3: Results for the simulation study: boxplots showing the average absolute bias in imputing the missing values in the covariates (X) and the response (Y) using the lme method, the proposed Bayesian semiparametric latent factor based model (BSP-LF), and a related sub-model with diagonal covariance matrices for the covariates (BSP-Diag).

S.4 Additional Figures

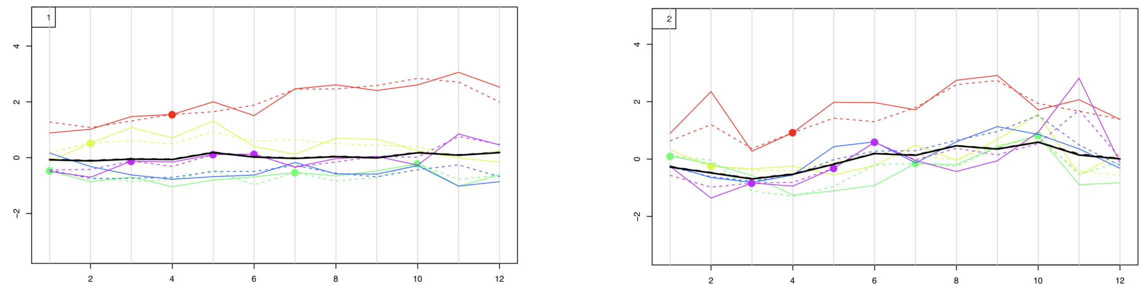

Figure S.4: Results for the simulation study: Observed (solid lines) and fitted (dotted lines) trajectories for the first 2 time-varying predictors for 5 randomly selected subjects superimposed over time-specific sample means across all subjects (solid black line) and the corresponding fitted values (dotted black line). The bullets represent mean imputed missing values.

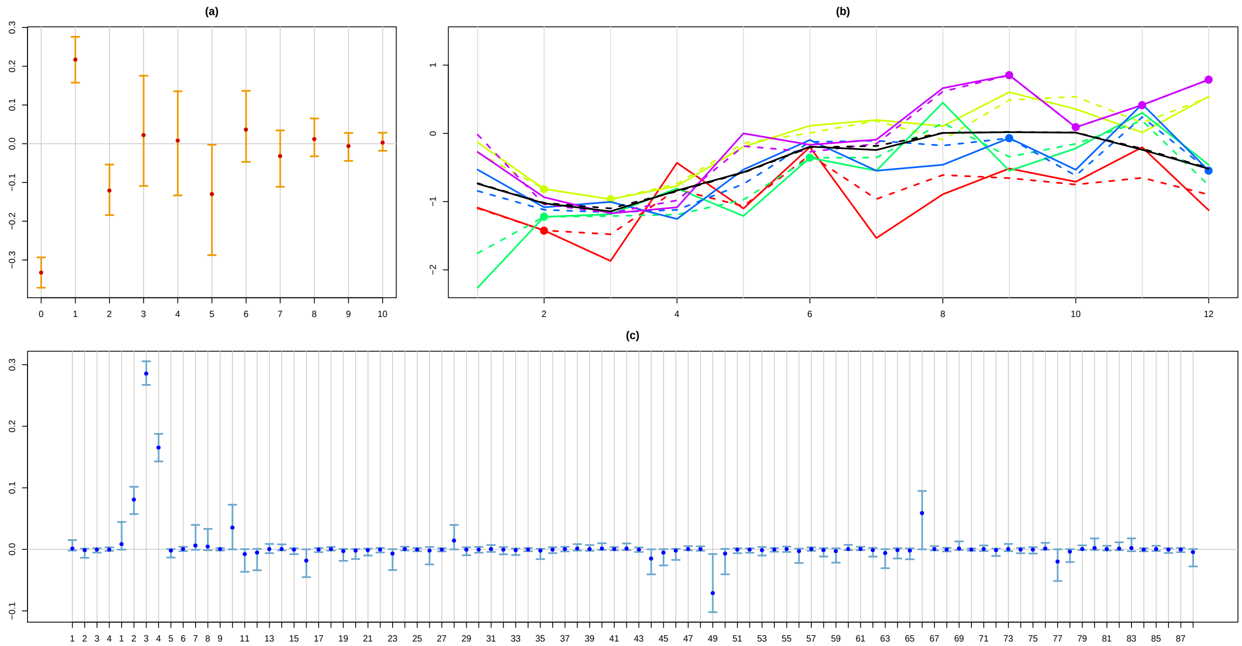

Figure S.5: Results for the simulation study: (a) Estimated posterior means of components of $\boldsymbol{\alpha}$ and their 90% credible intervals. (b) Observed (solid lines) and fitted (dotted lines) trajectories of y for 5 randomly selected subjects super-imposed over time-specific sample means across all subjects (solid black line) and the corresponding fitted values (dotted black line). The bullets represent mean imputed missing values. (c) Estimated posterior means of components of $\boldsymbol{\beta}$ and their 90% credible intervals. The first 4 components correspond to anthropometric baseline predictors. The remaining 88 components correspond to the time-varying predictors, comprising 10 clinical variables and 78 metabolites.

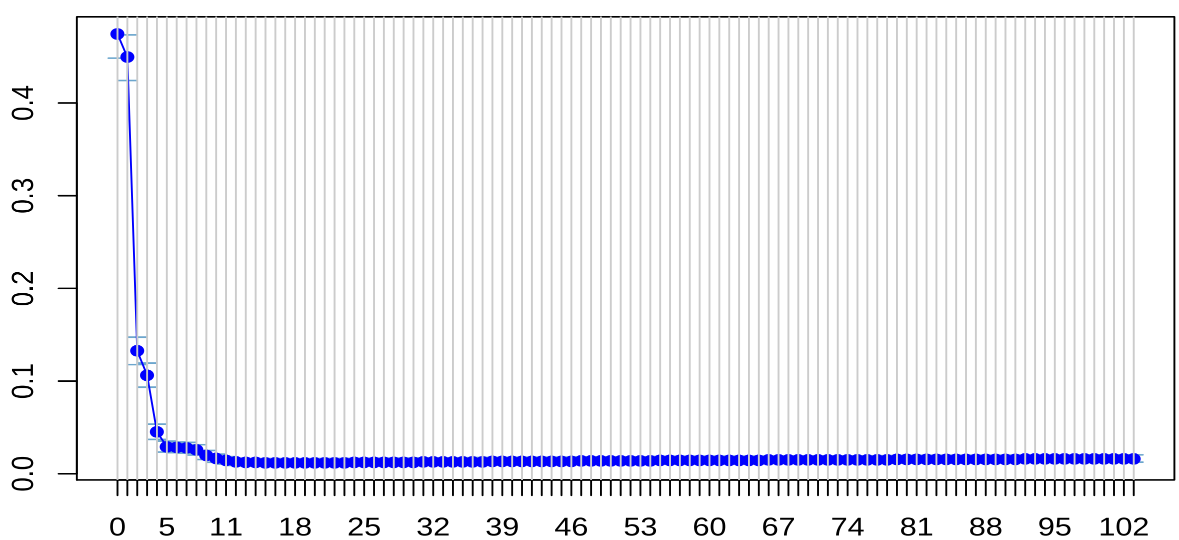

Figure S.6: Results for the simulation study: Model size vs the corresponding excess error ψ_λ. See Section S.2.4 in the Supplementary Information for additional details.

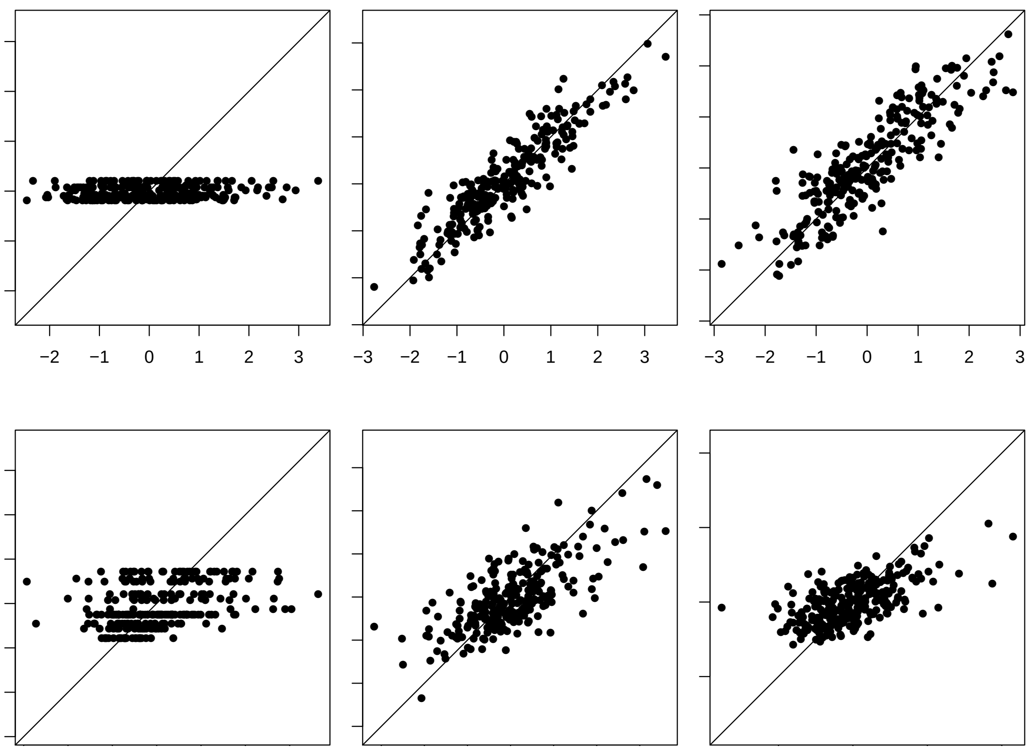

Figure S.7: Results for the simulation study: scatterplots showing the true missing values and the corresponding imputed values first 2 time-varying predictors for the lme method (left panel), the proposed Bayesian semiparametric latent factor based method (BSP-LF, middle panel), and its adaptation with diagonal covariance matrices for the covariates (BSP-Diag, right panel). In each case, the results correspond to a dataset that corresponds to the median average RMSE in estimating the regression coefficients.

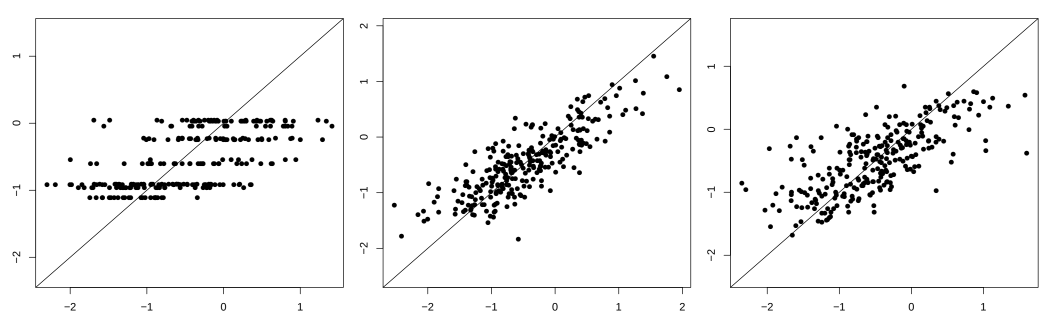

Figure S.8: Results for the simulation study: scatterplots showing the true missing values and the corresponding imputed values for the response y for the lme method (left panel), the proposed Bayesian semiparametric latent factor based method (BSP-LF, middle panel), and its adaptation with diagonal covariance matrices for the covariates (BSP-Diag, right panel). In each case, the results correspond to a dataset that corresponds to the median average RMSE in estimating the regression coefficients.

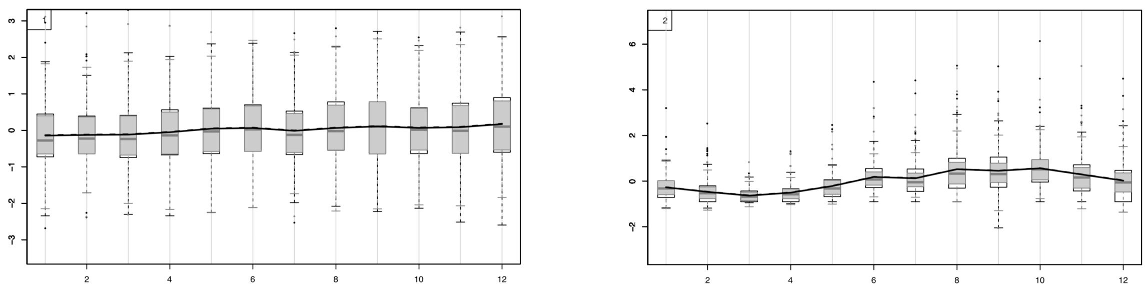

Figure S.9: Results for the EarlyBird study: Boxplots of observed values and imputed missing values (black) and fitted values (gray) of the first 2 time-varying predictors across time points super-imposed over time-specific sample means across all subjects (solid black line) and the corresponding fitted values (dotted black line).

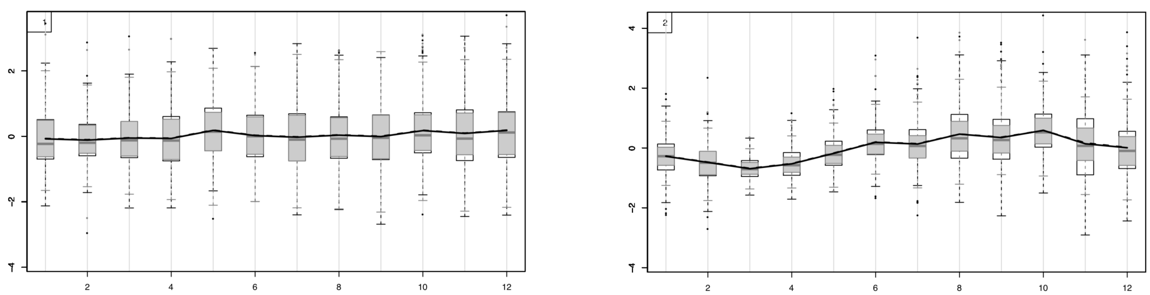

Figure S.10: Results for the simulation study: Boxplots of observed values and imputed missing values (black) and fitted values (gray) of the first 2 time-varying predictors across time points super-imposed over time-specific sample means across all subjects (solid black line) and the corresponding fitted values (dotted black line).

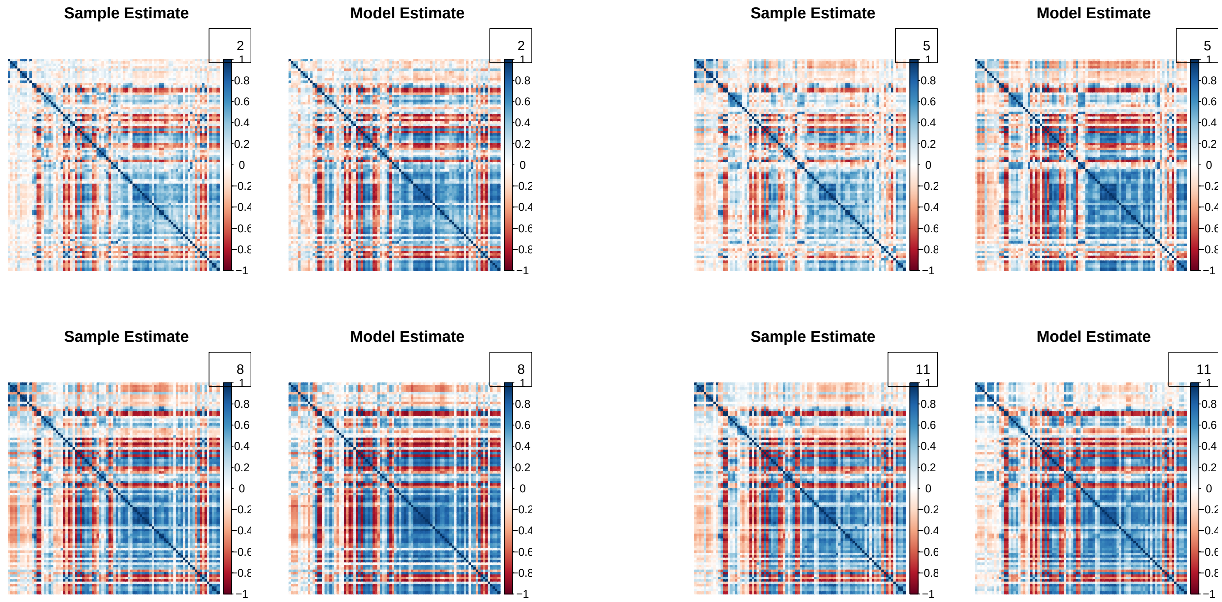

Figure S.11: Results for the EarlyBird study: Plots of empirical correlations between time-varying predictors based on observed and imputed missing values (sample estimates) and the corresponding model estimates for time-points 2, 5, 8 and 11.

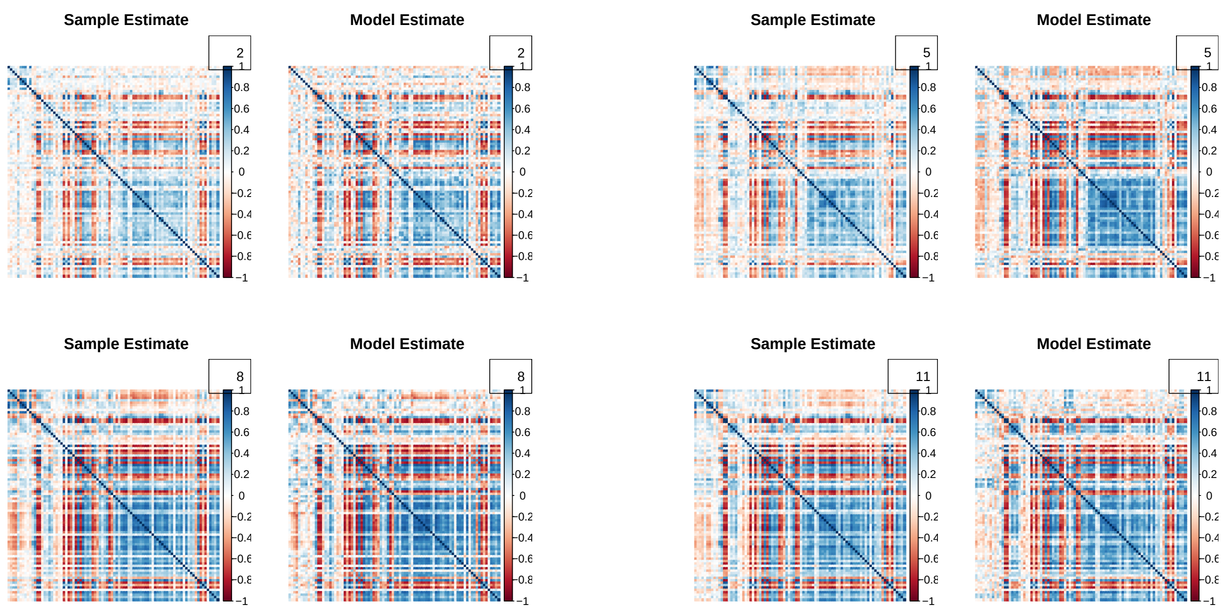

Figure S.12: Results for the simulation study: Plots of empirical correlations between time-varying predictors based on observed and imputed missing values (sample estimates) and the corresponding model estimates for time-points 2, 5, 8 and 11.

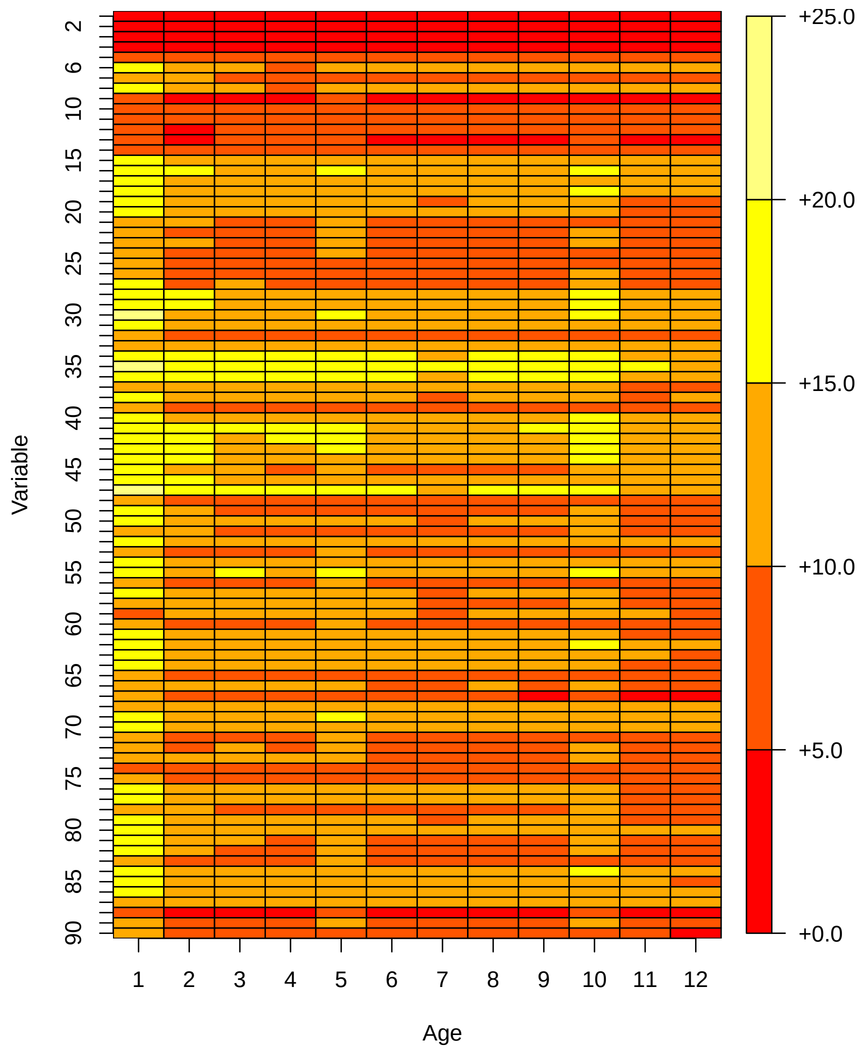

Figure S.13: Results for the EarlyBird study: $\log_2(\text{VIF})$ for the 92 time-varying predictors.

References

Durante, D. (2017). A note on the multiplicative gamma process. *Statistics & Probability Letters*, **122**, 198–204.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, **32**, 407–499.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.

Hahn, P. R. and Carvalho, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association*, **110**, 435–448.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**, 583–639.