Web-based Supplementary Materials for "Bayesian Latent Factor Regression for Functional and Longitudinal Data" by Silvia Montagna, Surya T. Tokdar, Brian Neelon, and David B. Dunson

SILVIA MONTAGNA, SURYA T. TOKDAR, and DAVID B. DUNSON
Department of Statistical Science, Duke University, Durham, NC 27708, U.S.A.

BRIAN NEELON
Children's Environmental Health Initiative, Nicholas School of the Environment, Duke University, Durham, NC 27708, U.S.A.

Web Appendix A

Choosing the number of latent factors k adaptively

The number of latent factors, k, is tuned as the sampler progresses, with adaptations designed to satisfy the diminishing adaptation condition in Theorem 5 of Roberts and Rosenthal (2007). Following Bhattacharya and Dunson (2011), we adapt with probability $p(t) = \exp\{\alpha_0 + \alpha_1 t\}$, with t denoting the t-th iteration and α_0, α_1 chosen so that adaptation occurs around every 10 iterations at the beginning of the chain and then decreases in frequency exponentially fast. In our application, we set $\alpha_0 = -1$ and $\alpha_1 = -5 \times 10^{-4}$. At every iteration, a random number u_t is sampled from a uniform distribution Unif(0,1), and adaptation occurs if $u_t \leq p(t)$. Whenever adaptation occurs, we count the columns of $\mathbf{\Lambda}$ having all elements in some pre-specified neighborhood of zero. We can intuitively assume that the factors corresponding to such columns have a negligible contribution, therefore we discard these columns of $\mathbf{\Lambda}$ and continue the sampler with a reduced number of factors, which also helps save computing time. Otherwise, if the number of such columns drops to zero we may be missing important factors, therefore we add a column to the loadings. The other parameters are modified accordingly and, when a factor is added, the new parameters are sampled from their prior distributions.

Refer to Bhattacharya and Dunson (2011) for further details on the adaptive Gibbs sampler.

References

[1] Bhattacharya, A. and Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika* **98,** 291–306.

[2] Roberts, G.O. and Rosenthal, J.S. (2007). Coupling and ergodicity of adaptive MCMC. *Journal of Applied Probability* **44,** 458–475.

Web Appendix B

MCMC algorithm for the latent factor regression model

We provide a detailed description of the MCMC algorithm used to update from the posterior distributions of the parameters based on the priors given in Section 2.1. The sampler cycles through the following steps:

- Step 1 Update of $\boldsymbol{\Lambda}$: sample $\lambda_{jh}, \delta_1, \delta_h, \phi_{jh}$ from the following posteriors:

  1. Denote the jth row of $\boldsymbol{\Lambda}_{k^*}$ (the loading matrix $\boldsymbol{\Lambda}$ truncated to $k^* << p$) by $\boldsymbol{\lambda}_j$; then the $\boldsymbol{\lambda}_j$'s have independent conditionally conjugate posteriors given by
  $$\pi(\boldsymbol{\lambda}_j|-) \sim N_{k^*}((\mathbf{D}_j^{-1} + \sigma_j^{-2}\boldsymbol{\eta}'\boldsymbol{\eta})^{-1}\boldsymbol{\eta}'\sigma_j^{-2}\boldsymbol{\theta}^{(j)}, (\mathbf{D}_j^{-1} + \sigma_j^{-2}\boldsymbol{\eta}'\boldsymbol{\eta})^{-1})$$
  with $\mathbf{D}_j^{-1} = \text{diag}(\phi_{j1}\tau_1, \ldots, \phi_{jk}\tau_{k^*})$, $\boldsymbol{\eta}' = [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_{k^*}]$ and $\boldsymbol{\theta}^{(j)} = (\theta_{j1}, \ldots, \theta_{jn})$, for $j = 1, \ldots, p$.
  2. Sample ϕ_{jh} from
  $$\pi(\phi_{jh}|-) \sim \text{Gamma}\left(\frac{v+1}{2}, \frac{v}{2} + \frac{\tau_h\lambda_{jh}^2}{2}\right)$$
  3. Sample δ_1 from
  $$\pi(\delta_1|-) \sim \text{Gamma}\left(a_1 + \frac{pk^*}{2}, 1 + \frac{1}{2}\sum_{l=h}^{k^*}\tau_l^{(1)}\sum_{j=1}^{p}\phi_{jl}\lambda_{jl}^2\right)$$
  4. Sample δ_h from
  $$\pi(\delta_h|-) \sim \text{Gamma}\left(a_2 + \frac{p^*}{2}(k-h+1), 1 + \frac{1}{2}\sum_{l=1}^{k^*}\tau_l^{(h)}\sum_{j=1}^{p}\phi_{jl}\lambda_{jl}^2\right)$$
  for $h \geq 2$, where $\tau_l^{(h)} = \prod_{t=1,t\neq h}^{l}\delta_t$ for $h = 1, \ldots, p$.

  The sampling begins with a very conservative choice of k^*, which is then automatically selected within the adaptive Gibbs sampler as described in Bhattacharya and Dunson (2011).

- Step 2 Update of $\sigma_j^2, j = 1, \ldots, p$: denoting as σ_j^{-2} the diagonal elements of $\boldsymbol{\Sigma}^{-1}$, sample σ_j^{-2} from conditionally independent posteriors
$$\pi(\sigma_j^{-2}|-) \sim \text{Gamma}\left(\frac{n}{2} + a_\sigma, b_\sigma + \frac{\sum_{i=1}^{n}(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)^2}{2}\right)$$

- Step 3 Update of φ^{-2}: sample φ^{-2} from
$$\pi(\varphi^{-2}|-) \sim \text{Gamma}\left(\frac{N}{2} + a_\varphi, b_\varphi + \frac{\sum_{j=1}^{N}(y_j - \boldsymbol{\Theta}_j)^2}{2}\right)$$
where N denotes the total number of observations, $\mathbf{y}$ is a column vector which stacks the measurements for all women, $\mathbf{y} = (y_{1,t_{1,1}}, \ldots, y_{n,t_{n,n_n}})'$, and $\boldsymbol{\Theta}$ is a $N \times 1$ column vector which stacks the scores for all subjects, $\boldsymbol{\Theta} = \{\mathbf{B}_i\boldsymbol{\theta}_i, \ldots, \mathbf{B}_n\boldsymbol{\theta}_n\}'$, where each $\mathbf{B}_i\boldsymbol{\theta}_i$ has dimension $n_i \times 1$ with n_i the number of measurements for subject i.

- Step 4: Update of $\boldsymbol{\beta}$ and ω elements:

  4-a) Given the prior $\omega_{lj} \sim \text{Gamma}(1/2, 1/2)$, $l = 1, \ldots, r$ and $j = 1, \ldots, k$, sample ω_{lj} from the full conditional posterior

$$\pi(\omega_{lj}|-) \sim \text{Gamma}\left(1, \frac{1}{2}\left(1 + \beta_{lj}^2\right)\right)$$

  4-b) Sample the jth column of the matrix of coefficients $\boldsymbol{\beta}$ from the full conditional posterior

$$\pi(\boldsymbol{\beta}_j|-) \sim N\left(\left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}' + \mathbf{E}^{-1}\right)^{-1}\tilde{\mathbf{X}}\boldsymbol{\eta}'_{\cdot j}, \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}' + \mathbf{E}^{-1}\right)^{-1}\right)$$

  with matrix $\mathbf{E}$ corresponding to $\mathbf{E} = \text{Diag}(\omega_{lj}^{-1})$, $l = 1, \ldots, r$ and $j = 1, \ldots, k$.

- Step 5 Update of $\boldsymbol{\eta}_i$: marginalizing out $\boldsymbol{\theta}_i$, the model can be rewritten as

$$\begin{aligned}\mathbf{y}_i &= \mathbf{B}_i\boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{B}_i\boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(0, \varphi^2\mathbf{I}_{n_i}), \quad \boldsymbol{\zeta}_i \sim N_p(0, \boldsymbol{\Sigma}) \\ &= \mathbf{B}_i\boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\alpha}_i^*, \quad \boldsymbol{\alpha}_i^* \sim N(0, \varphi^2\boldsymbol{I}_{n_i} + \mathbf{B}_i\boldsymbol{\Sigma}\mathbf{B}_i')\end{aligned}$$

  Thus, sample $\boldsymbol{\eta}_i$ from the full conditional posterior

$$\begin{aligned}\pi(\boldsymbol{\eta}_i|-) &\sim N(\mathbf{A}^{-1} \times \mathbf{C}, \mathbf{A}^{-1}) \\ \mathbf{A} &= \boldsymbol{\Lambda}'\mathbf{B}_i'(\varphi^2\boldsymbol{I}_{n_i} + \mathbf{B}_i\boldsymbol{\Sigma}\mathbf{B}_i')^{-1}\mathbf{B}_i\boldsymbol{\Lambda} + \boldsymbol{I}_k \\ \mathbf{B} &= \boldsymbol{\beta}'\boldsymbol{x}_i + \boldsymbol{\Lambda}'\mathbf{B}_i'(\varphi^2\mathrm{I}_{n_i} + \mathbf{B}_i\boldsymbol{\Sigma}\mathbf{B}_i')^{-1}\mathbf{y}_i\end{aligned}$$

- Step 6 Update of $\boldsymbol{\theta}_i$: sample $\boldsymbol{\theta}_i$ from conditionally independent posteriors

$$\begin{aligned}\pi(\boldsymbol{\theta}_i|-) \sim N_p((\varphi^{-2}\mathbf{B}_i'\mathbf{B}_i + \boldsymbol{\Sigma}^{-1})^{-1}(\varphi^{-2}\mathbf{B}_i'\mathbf{y}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{\eta}_i), \\ (\varphi^{-2}\mathbf{B}_i'\mathbf{B}_i + \boldsymbol{\Sigma}^{-1})^{-1})\end{aligned}$$

Web Appendix C

Setting parameters

To facilitate the routine implementation of the proposed method, the Matlab codes for the LFRM and its joint modeling extensions (see Section 3) are available at the *Biometrics* website on Wiley Online Library.

To implement our methodology, one has to choose the hyperparameters for the priors in Section 2.1 and the parameters ν in (3) and p in (2). Likely, the most daunting task is the choice of the bandwidth ν, that we fixed to 4 as a reasonable default value to ensure smooth trajectories. In general, one can not obtain curves bumpier than the resolution determined by the bandwidth, thus the choice of ν requires careful sensitivity analysis to identify a value which induces the desired level of smoothness for the trajectories. Applications with trajectories not having the same level of smoothness everywhere would require spatially adaptive smoothness, which can potentially be achieved by choosing a pre-specified finite dictionary of different bandwidths and then allowing the kernels to have varying unknown bandwidths via a griddy Gibbs sampler.

The basis function representation in (2) requires the choice of a truncation p. In general,

one can include a rich, pre-specified set of basis functions ($p \approx 10, 20$ or larger) since the model allows automatic shrinkage and effective removal of basis coefficients not needed to characterize any of the curves under study, thus effectively induces basis selection. In our blood pressure application, the choice $p = 10$ ensured sufficiently many equally-spaced kernels to capture a high variety of smooth trajectory shapes.

Other parameters that need to be determined in the MCMC algorithm include υ, a_1, a_2 in (8)-(9). As remarked in Section 2.1, a choice $a_2 > 1$ induces stochastically increasing τ_h in (8), which favors more shrinkage as the column index increases. We set $\upsilon = 5$ and $a_1 = a_2 = 1.5$, but our sensitivity analyses showed robustness to different choices of these hyperparameters. Furthermore, one has to choose a_σ and b_σ, which are the inverse-gamma hyperparameters values for σ_j^2, and a_φ and b_φ, which are the inverse-gamma hyperparameter values for the measurement error variance φ^2. Our suggestion is to fix a mean and variance for the inverse-gamma priors and solve for the hyperparameters.

Alternatively to the Cauchy prior in (11), one could choose a Gaussian prior distribution for the $\boldsymbol{\beta}$ coefficients but this leads to poorer performance if a subsample of women has very sparse measurements. This occurrence is common when dealing with longitudinal data, which often consist of few and sparse measurements per subject, and it is verified in the blood pressure data where a group of women has few observations, usually located in the second half of the pregnancy. For this group of women, the prior becomes more influential and the intercept is pulled closer to zero than for women with more observations, resulting in an undesired low MAP trajectory estimate at early pregnancy.

As for the bivariate probit model in Section 3.2, we chose normal and multivariate normal priors for the additional model parameters. The prior for the intercept on the latent indicator of preeclampsia, α_1, was set to be $\alpha_1 \sim \mathrm{N}(\Phi^{-1}(0.12), 0.25)$, whereas the prior on the intercept for the latent indicator of low birth weight, α_2, was set to correspond to $\alpha_2 \sim \mathrm{N}(\Phi^{-1}(0.082), 0.25)$. The hyperprior mean for α_1 was set to be moderately high provided that the proportion of preeclamptic women in the sample is over twice the typical incidence range of 5-8%, and that of α_2 was chosen to correspond to the national average. Finally, $\boldsymbol{\gamma}_1 \sim \mathrm{N}_k(\boldsymbol{\mu}_{\gamma,1}, \boldsymbol{\Sigma}_{\gamma,1})$ and $\boldsymbol{\gamma}_2 \sim \mathrm{N}_k(\boldsymbol{\mu}_{\gamma,2}, \boldsymbol{\Sigma}_{\gamma,2})$, with $\boldsymbol{\mu}_{\gamma,1} = \boldsymbol{\mu}_{\gamma,2} = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\gamma,1} = \boldsymbol{\Sigma}_{\gamma,2} = \mathbf{I}_k$. We repeated the analysis for a variety of these hyperparameter values (i.e., with the variance multiplied by 2 and divided by 2, etc.), but no noticeable differences were found in the results.

Finally, we examined the joint model of birth weight, gestational age at delivery and blood pressure. Specifically, we chose $\boldsymbol{\mu}_\gamma^* = \mathbf{0}$ and $\boldsymbol{\Sigma}_\gamma^* = \mathbf{I}_k$ as mean and covariance matrix of the multivariate normal prior distribution for $\boldsymbol{\gamma}$, whereas $\nu_h^* = 4$, and $\boldsymbol{V}_h = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ were chosen as the hyperparameter values of the inverse-Wishart distribution on $\boldsymbol{\Sigma}_h, h = 0, 1$. Our sensitivity analyses showed that results were robust to different choices of these parameters. Finally, we applied an EM algorithm MLE using a two-component mixture of bivariate normals to the data (without including covariate information) to determine the hyperparameters $\boldsymbol{\mu}_0^h$ and $\boldsymbol{\Sigma}_{\mu 0}^h$, the mean and covariance matrices of the two Gaussian mixture components. We obtained

$$\boldsymbol{\mu}_0^1 = \begin{pmatrix} \mu_{0g}^1 \\ \mu_{0b}^1 \end{pmatrix} = \begin{pmatrix} 36 \\ 2.57 \end{pmatrix}, \boldsymbol{\mu}_0^2 = \begin{pmatrix} \mu_{0g}^2 \\ \mu_{0b}^2 \end{pmatrix} = \begin{pmatrix} 39 \\ 3.30 \end{pmatrix}, \boldsymbol{\Sigma}_{\mu 0}^1 = \begin{pmatrix} 7.66 & 1.37 \\ 1.37 & 0.35 \end{pmatrix}, \boldsymbol{\Sigma}_{\mu 0}^2 = \begin{pmatrix} 1.34 & 0.19 \\ 0.19 & 0.22 \end{pmatrix}.$$

Web Appendix D

Simulation experiment

To evaluate the performance of our model and to compare it with related methods, we considered a simulation example. To make the simulated data more realistic and interpretable we based them on the Healthy Pregnancy, Healthy Baby (HPHB) study, assuming $n = 200$ and with the true parameters set equal to the posterior means from the real data analysis (see Section 4 of the paper). We generated samples of gestational age (in weeks) and birth weight (in Kg) from a two-component mixture of bivariate normal distributions with true means set equal to $\boldsymbol{\mu}_1 = (34.54, 2.27)'$ and $\boldsymbol{\mu}_2 = (38.17, 3.50)'$ and covariance matrices

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1.516 & 0.261 \\ 0.261 & 1.235 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1.212 & 0.185 \\ 0.185 & 1.221 \end{pmatrix}$$

We standardized time to the $[0, 1]$ interval, $t_{ij} \in [0, 1]$, and set $b_1(t_{ij}) = 1$ and $b_{l+1}(t_{ij}) = \exp\{-4||t_{ij} - \psi_l||^2\}, l = 1, \ldots, 9$, with ψ_l's equally spaced kernel locations in $[0, 1]$ and $p = 10$.

To implement our Bayesian analysis, we chose a Gamma$(0.5, 0.25)$ prior distribution with mean 2 for the diagonal elements of $\boldsymbol{\Sigma}^{-1}$, and we placed a Gamma(0.5, 0.2) with mean 2.5 on φ^{-2}. The gamma hyperparameter for ϕ_{jh} was set to be $\upsilon = 5$, $a_1 = a_2 = 1.5$ in (8)-(9) and a Cauchy prior was induced on the matrix of coefficients $\boldsymbol{\beta}$ (11). We chose $k = 4$ as the starting number of factors, and we adapted k according to the procedure described in Bhattacharya and Dunson (2011). The MCMC algorithm was run for 25,000 iterations including a 5,000 iterations burn-in, and collected every 5th sample to thin the chain and reduce the autocorrelation in the posterior samples. Based on the examination of traceplots of function values at a variety of time locations and for different subjects, the sampler appeared to converge rapidly and to mix efficiently.

The average of the estimated number of factors was 11.37 corresponding to $k_{\text{true}} = 11$, and with empirical 95% credible interval given by $[9, 13]$. The estimated posterior mean of $\boldsymbol{\mu}_1$ was (34.37, 2.35) and the estimated posterior mean for $\boldsymbol{\mu}_2$ was $(37.91, 3.42)$ respectively, with corresponding 95% credible intervals containing the true values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The estimates of the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ were

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 1.413 & 0.429 \\ 0.429 & 1.105 \end{pmatrix} \text{ and } \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 1.098 & 0.265 \\ 0.265 & 1.152 \end{pmatrix}$$

with 95% credible intervals containing the true values of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

The left panels of Figure 1 show the data, true curves and estimates under the LFRM for three randomly selected subjects. In general, estimates are very close to the true curves even when data are sparse, as for subject 122, and the true curves are always enclosed in the credible bounds.

We then obtained a smooth estimator of the covariance operator and its corresponding eigenfunctions as described by Crainiceanu and Goldsmith (2010). In contrast to the LFRM, FPCA does not allow to learn about the representation size k, thus we need to estimate the dimension of the functional space. As a fast alternative to cross-validation, we decided to retain a number of eigenfunctions such that the cumulative percentage of explained variance was greater than 90% and the explained variance by any single subsequent component was less than 5%. Therefore, we retained the first $k = 4$ eigenfunctions and obtained $\boldsymbol{\Lambda}$ as the least squares estimate of

$$\boldsymbol{\Psi} = \mathbf{B}^* \times \boldsymbol{\Lambda} \tag{1}$$

with $\boldsymbol{\Psi}$ denoting here the matrix of eigenfunctions and $\mathbf{O}_i \times \mathbf{B}^* = \mathbf{B}_i$, $\mathbf{B}_i$ denoting the design matrix for subject i and $\mathbf{O}_i$ representing an $(n_i \times T)$ matrix with column j equal to a column of 1's if subject i was measured at time j, $j = 1, \ldots, T$ (T denotes the number of unique time locations). We then repeated the analysis fitting the LFRM with $\boldsymbol{\Lambda}$ and the number of factors $k = 4$ fixed. We will denote this procedure as two-stage FPCA approach. Estimates are shown in the right panels of Figure 1. We can notice some deviations of the estimated curves from the true curves along the course of the entire pregnancy, with very wide confidence intervals at early pregnancy when typically no or few measurements are observed and when data are more sparse, as for subject 122. Notice also that for subject 8 the true curve is no longer enclosed within the credible bounds at delivery. The analysis was repeated retaining $k_{true} = 11$ eigenfunctions, but this did not lead to any significant improvement in the performance.

Under the two-stage FPCA approach, the estimated posterior mean of $\boldsymbol{\mu}_1$ was $(34.26, 2.31)$ and the estimated posterior mean for $\boldsymbol{\mu}_2$ was $(37.81, 3.39)$ respectively, with corresponding 95% credible intervals containing the true values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The estimates of the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ were

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 1.249 & 0.383 \\ 0.383 & 1.087 \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 1.227 & 0.305 \\ 0.305 & 1.165 \end{pmatrix}$$

To assess the predictive performance, we repeated the analysis holding out and predicting the MAP measurements collected after the $30th$ week of gestation for 100 randomly selected women having at least 1 observation in the first 30 weeks and 1 observation after the $30th$ week. Also, we fitted "baseline" LFRM and two-stage FPCA approach with no covariates setting $\boldsymbol{\eta}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_k)$. Results are reported in Table 1 together with the computing time in seconds per hundred of iterations. The high values of the predictive errors are not surprising given the presence of many outliers in the MAP measurements that are hard to predict. The LFRM leads to better predictive performance than the two-stage FPCA approach both with and without covariates. However, we notice that the predictive errors do not decrease with the incorporation of covariate information: this seems to suggest that in our blood pressure application the outcomes are predominantly learned from the random deviations rather than from the covariates. In fact, the MAP measurements are affected by great variability that makes them hard to predict despite the available covariate information.

Table 1

Mean square predictive error (MSPE), predictive average absolute bias (PAAB) and predictive maximum absolute bias (PMAB) for the simulated data with the LFRM and the two-stage FPCA approach fitted with and without covariates, respectively. The computing time in seconds is per hundred of iterations.

	LFRM		two-stage FPCA	
	Covariates	No Covariates	Covariates	No Covariate
MSPE	70.02	69.31	73.45	74.11
PAAB	6.63	6.61	6.80	6.83
PMAB	27.74	27.50	28.63	28.72
Comp. Time	40	21	34	15

Figure 2 shows the estimated joint distribution of gestational age (in weeks) and birth weight (in Kg) for subjects 8 and 46 in the simulation example under the LFRM, along with corresponding contour plots. The true values of gestational age at delivery and birth weight correspond to

(38.82, 3.47) and (33.51, 1.41) for subject 8 and subject 46, respectively. The joint distribution is bimodal, with the two components of the Gaussian mixture clearly distinct, and with the joint model assigning higher mass to the true component each subject belongs to, that is, the second component for subject 8 and the first component for subject 46. The posterior probability of being in component 1 is 0.3057 for subject 8, and increases to 0.6025 for subject 46. Analogous results are obtained with the two-stage FPCA approach, with posterior probabilities of being in component 1 being equal to 0.2938 and 0.5592 for subjects 8 and 46, respectively.

The analysis was repeated under different choices of the hyperparameter values and initial number of factors for the LFRM. The results were robust, with no noticeable differences in the conclusions.

References

[1] Crainiceanu C. and Goldsmith J. (2010). Bayesian functional data analysis using WinBUGS. *Journal of Statistical Software* **32,** 1–33.

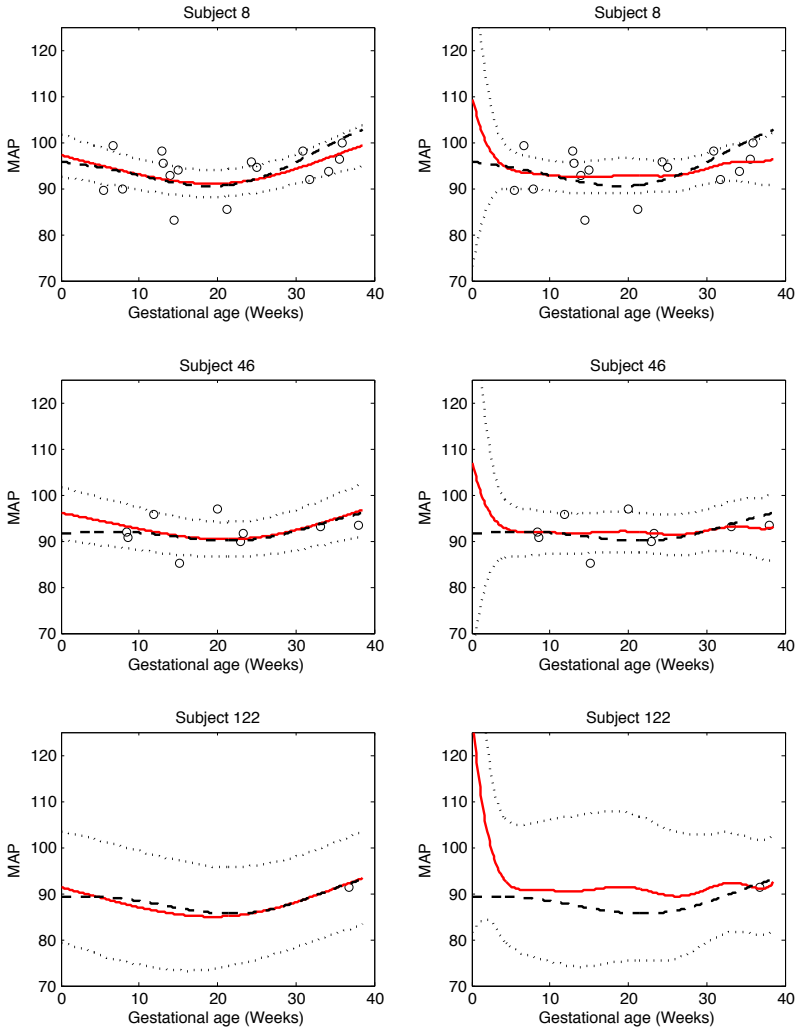

Figure 1: Data and function estimates for 3 subjects in the simulation example under the LFRM (left panels) and two-stage FPCA (right panels). The true functions are represented with dashed lines, the posterior means are solid lines, and the dotted lines are 95% pointwise credible intervals.

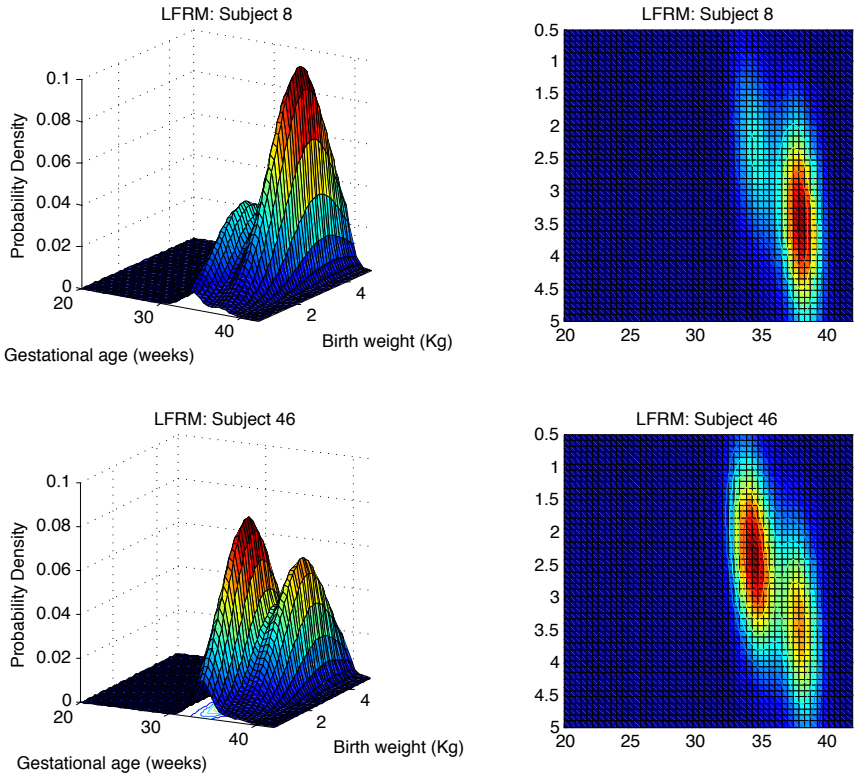

Figure 2: LFRM-estimated joint distribution of gestational age (weeks) and birth weight (Kg) and contour plot for subjects 8 and 46 in the simulation example.

1 Web Figure 1

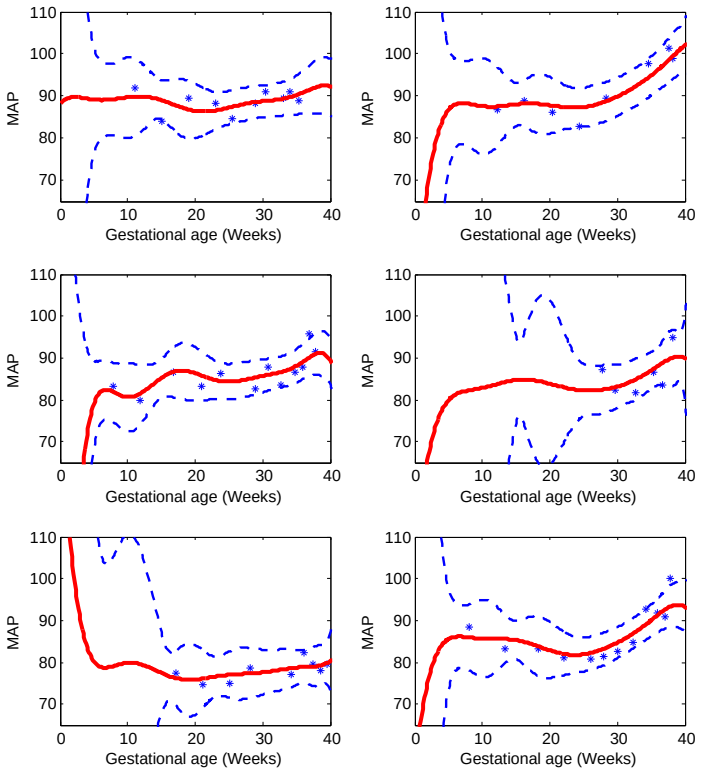

Figure 3: MAP function estimates for 6 randomly selected women under the two-stage FPCA approach. The posterior means are solid lines and dashed lines are 95% pointwise credible intervals. The x-axis scale is time in weeks starting at the estimated day of ovulation.

2 Web Figure 2

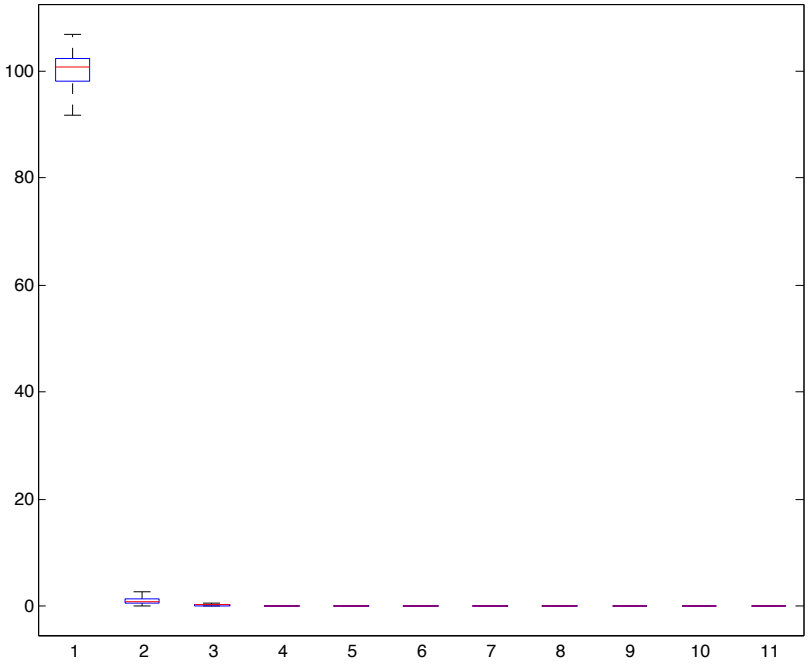

Figure 4: Side-by-side boxplot of the norms of the posterior mean estimates of the columns of the factor loading matrix $\boldsymbol{\Lambda}$.

The boxplots show a decay of the norms from $\boldsymbol{\lambda}_1$ to $\boldsymbol{\lambda}_{k^*}$, as expected by the structure induced by the MGPS prior on $\boldsymbol{\Lambda}$. Note that $k^* = 11$ corresponds to the posterior mean number of factors. Therefore, the first few important factors are loaded heavily and significantly contribute to the estimated of $\boldsymbol{\theta}_i$. Although the norms of the remaining factors appear equal to zero, none of the factor loadings is exactly equal to zero, but shrunk towards zero by the MGPS prior.

3 Web Figure 3

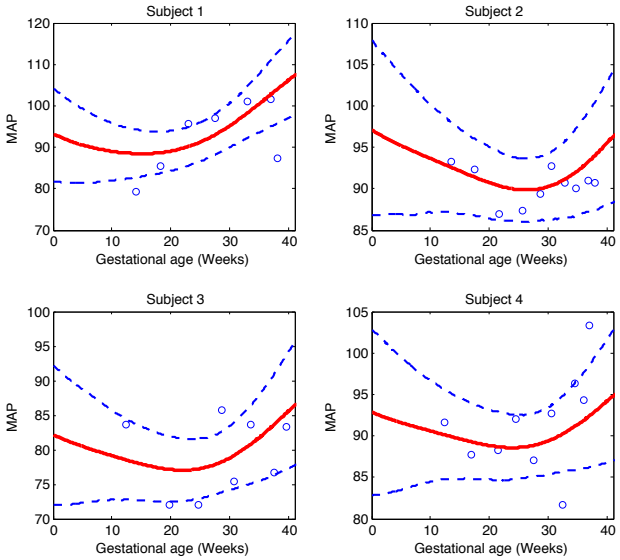

Figure 5: MAP function estimates at the $35th$ week for four subjects in the test set. The posterior means are solid lines and dashed lines are 95% pointwise credible intervals. The x-axis scale is time in weeks starting at the estimated day of ovulation.