

Variable Selection in the Functional Linear Concurrent Model

Jeff Goldsmith^{1,*} and Joseph E. Schwartz^{2,3}

¹Department of Biostatistics, Mailman School of Public Health, Columbia University
**jeff.goldsmith@columbia.edu*

²Department of Medicine, Columbia University Medical Center

³Department of Psychiatry and Behavioral Sciences, Stony Brook University

April 1, 2016

Abstract

We propose methods for variable selection in the context of modeling the association between a functional response and concurrently observed functional predictors. This data structure, and the need for such methods, is exemplified by our motivating example: a study in which blood pressure values are observed throughout the day, together with measurements of physical activity, location, posture, attitude, and other quantities that may influence blood pressure. We estimate the coefficients of the concurrent functional linear model using variational Bayes and jointly model residual correlation using functional principal components analysis. Latent binary indicators partition coefficient functions into included and excluded sets, incorporating variable selection into the estimation framework. The proposed methods are evaluated in real-data analyses, and are implemented in a publicly available R package.

Key Words: Functional data; Longitudinal data; Spline smoothing; Variational Bayes.

1 Introduction

Wearable devices and other portable, unobtrusive monitors have made it possible to concurrently collect many data streams on study participants in parallel. In this and similar data settings, the functional linear concurrent model (FLCM) provides a useful framework for understanding the

association between the functional response and functional covariates at a specific time. This model can be written

$$Y_i(t) = \beta_0(t) + \sum_{k=1}^p X_{ik}(t)\beta_k(t) + \delta_i(t), \quad (1)$$

and has been previously studied in the context of a relatively small number of predictors p (Ramsay and Silverman, 2005; Fan and Zhang, 2008; Şentürk and Nguyen, 2011). However, the presence of a large number of data streams – a context that is becoming increasingly common – necessitates the incorporation of variable selection methods into the estimation of model (1). This problem has not yet been addressed, and the development of such methods is the main contribution of this manuscript.

Our approach begins with the expansion of coefficient functions $\beta_k(t)$ in model (1) in terms of a spline basis. Sparsity is induced by setting all spline coefficients for a given coefficient function to zero, at least approximately; to do so we use tools from Bayesian variable selection, in particular by specifying spike-and-slab priors for groups of spline coefficients. To model correlation in residual curves $\delta_i(t)$, we use a functional principal components expansion. We jointly estimate all model parameters using a computationally efficient variational Bayes algorithm and choose the tuning parameter, the non-zero “spike” prior variance, using cross validation. All methods are implemented in the R package `vbvs.concurrent`, publicly available as a GitHub repository.

As noted above, the estimation of parameters in the FLCM has been the subject of several previous papers, although none of these have focused on variable selection. Indeed, variable selection is the subject of a small but growing literature in functional data analysis. Gertheiss et al. (2013) developed variable selection tools for the linear scalar-on-function regression model; more recently, Barber et al. (2016) and Chen et al. (2016) have proposed methods for function-on-scalar regression models. Each of these made use of group penalties (e.g. group lasso or group MCP) for spline coefficients and, by recasting the functional regression model as a standard linear model, could apply off-the-shelf software for estimation and penalization. Because of the form of the least-

squares estimate of the FLCM, it is not obvious that this strategy is applicable in the current setting. Moreover, [Chen et al. \(2016\)](#) noted the importance of accounting for residual correlation when performing variable selection in a functional response model and used a “pre-whitening” approach to address this issue. In contrast, we jointly model the residual correlation structure and the regression coefficients in a Bayesian framework.

Variational Bayes algorithms have enjoyed some popularity in the functional data analysis literature, largely due to the computation efficiency of these approaches in comparison to Monte Carlo sampling and, in some cases, frequentist estimation methods. [Goldsmith et al. \(2011\)](#) and [McLean et al. \(2013\)](#) developed variational algorithms for scalar-on-function regressions, and [Goldsmith and Kitago \(2016\)](#) used variational Bayes for function-on-scalar regression. [van der Linde \(2008\)](#) and [van der Linde \(2009\)](#) used variational approximations in a Bayesian approach to functional principal components analysis; this work is relevant in particular to our residual decomposition, although it did not address the incorporation of covariates in the mean structure. Recently, [Earls and Hooker \(2016\)](#) used variational Bayes to combine a factor analysis related to Bayesian FPCA with the registration of features across curves.

Our method for variable selection is related to a recent EM-based approach to variable selection in the standard linear model ([Ročková and George, 2014](#)). In contrast to previous spike-and-slab normal mixture formulations, in which the spike distribution is considered to be a point mass at zero ([George and McCulloch, 1993, 1997](#)), [Ročková and George \(2014\)](#) posited a continuous spike prior. Doing so allowed the derivation of a closed-form EM algorithm and introduced the spike prior variance as a tuning parameter whose adjustment results in a sequence of progressively sparser models. Here, we adapt this basic approach to the context of variable selection in the FLCM and estimate parameters using variational Bayes rather than an EM algorithm.

Our presentation of methods in Section 3 will be general, but we first motivate our work by a discussion of data collected in the Masked Hypertension Study in Section 2. An analysis of these data is presented in Section 4. We end with a discussion in Section 5.

2 The Masked Hypertension Study

Elevated blood pressure, or hypertension, is associated with increased risk of several common diseases, most notably diabetes, myocardial infarction, and stroke. Hypertension diagnoses have historically been based solely on blood pressure measurements taken in the clinic setting, but a growing body of work suggests that out-of-clinic blood pressure may provide a more complete assessment of cardiovascular health. Indeed, the term “masked hypertension” was coined to describe individuals with normal clinic blood pressure but elevated ambulatory or daytime blood pressure ([Pickering et al., 2002](#)); in the Masked Hypertension Study, roughly 15% of subjects with normal clinic blood pressure had masked hypertension ([Shimbo et al., 2012](#)).

Understanding daytime blood pressure level depends on frequent measurement as a subject engages in normal activities, and ambulatory blood pressure (ABP) monitoring is used to this end. Briefly, participants in the Masked Hypertension Study were fitted with a mobile cuff that measured blood pressure every 28 minutes over a 24-hour monitoring period; the relatively sparse measurement is attributable to the requirement that subjects stop activities to ensure an accurate reading and the requirement, for this study, that they complete a 2-minute diary entry immediately after each reading. Figure 1 shows systolic blood pressure (SBP) measurements for 40 participants taken between 10:00am and 10:00pm; clear differences between subjects exist, but there is also substantial within-subject variability over the course of the day.

Previous analyses of ambulatory blood pressure have mostly focused on average daytime (or nighttime) blood pressure ([Shimbo et al., 2012](#); [Abdalla et al., 2015](#)). However, individual ambulatory blood pressure readings are thought to depend on the specific context at the time of measurement, which changes throughout the day. Blood pressure may vary, for example, based on recent physical activity, mood, posture, location, activity type, and other variables. In the Masked Hypertension Study, these covariates were assessed using accelerometers to quantify the intensity of activity and ecological momentary assessments, brief diary entries completed on a pre-programmed electronic diary (Palm Pilot Tungsten 3) immediately after each blood pressure reading, to quantify

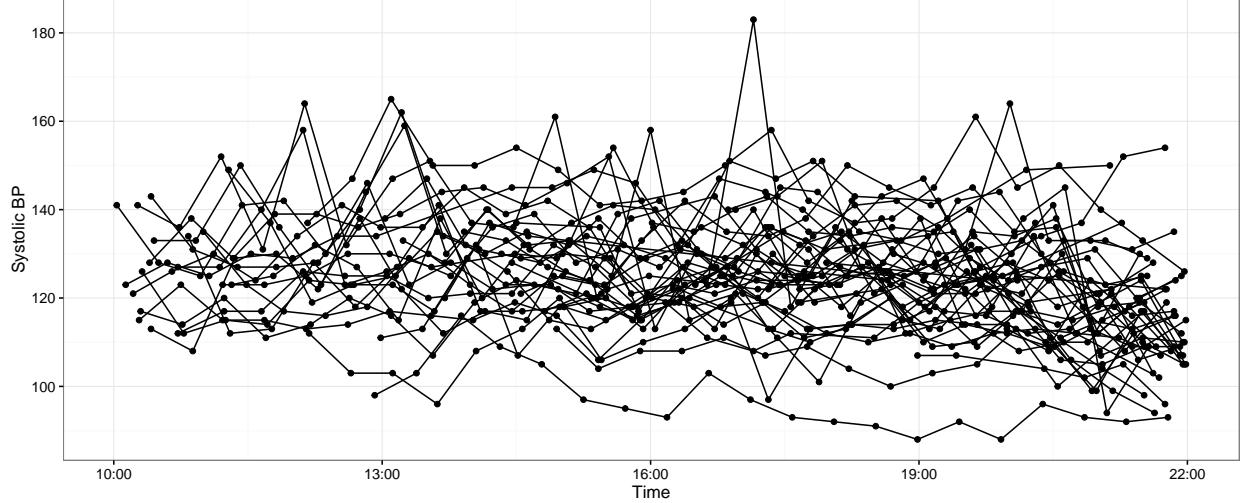


Figure 1: Observed ambulatory systolic blood pressure data for 40 subjects between 10:00am and 10:00pm.

other variables.

The potential for time-specific dependence of blood pressure on recent activity and other contextual variables was one factor in the design of the Masked Hypertension Study and motivates our use of the FLCM. The number of these variables requires selection methods to prevent overfitting and provide accurate predictions.

3 Methods

Our goal is to fit the functional linear concurrent model while inducing sparsity in the estimated coefficient functions. As part of our strategy, we also model residuals using a functional principal components expansion. That is, we expand the residuals $\delta_i(t)$ in (1) in terms of a shared basis to obtain

$$Y_i(t) = \sum_{k=1}^p X_{ik}(t)\beta_k(t) + \sum_{k=1}^{K_\Psi} c_{ik}\psi_k(t) + \epsilon_i(t). \quad (2)$$

In (2), the $\psi_k(t)$ are the functional principal component basis functions (FPCs) and the c_{ik} are the random subject-specific loadings; together, these account for correlation within residual curves.

The $\epsilon_i(t)$ is thus a mean-zero uncorrelated error curve, which we will assume to have a constant variance σ_ϵ^2 . The intercept $\beta_0(t)$ is omitted from (2) because both predictor and response curves will be centered prior to estimation (see Section 3.3 for details).

Model (2) is conceptual in that predictor and response curves are observed on a finite grid which is, in our case, sparse and irregular across subjects. Nonetheless, it is useful for framing our approach and emphasizing the functional nature underlying the observed data. In Section 3.1 we detail the model specification for observed data, while in Sections 3.2 and 3.3 we describe our fitting algorithm and practical considerations, respectively.

3.1 Model specification for observed data

We specify our model with sparse and irregular data in mind, although the following applies equally to curves measured on a dense grid shared across subjects. For subject i , $1 \leq i \leq I$, assume that the response and predictor curves are observed at time points $\mathbf{t}_i = \{t_{i1}, \dots, t_{iJ_i}\}$. Let functions evaluated at \mathbf{t}_i denote $J_i \times 1$ vectors of those functions on the observed time points (e.g. $Y_i(\mathbf{t}_i) = [Y_i(t_{i1}), \dots, Y_i(t_{iJ_i})]^T$ and $\beta_k(\mathbf{t}_i) = [\beta_k(t_{i1}), \dots, \beta_k(t_{iJ_i})]^T$).

Coefficient functions $\beta_k(t)$ in (2) are expanded in terms of a fixed spline basis $\Theta(t)$ made up of K_Θ functions $\theta_1(t), \dots, \theta_{K_\Theta}(t)$. Let $\Theta(\mathbf{t}_i)$ be the $J_i \times K_\Theta$ spline evaluation matrix on the grid \mathbf{t}_i ; then $\beta_k(\mathbf{t}_i) = \Theta(\mathbf{t}_i)\mathbf{b}_k^\beta$ where \mathbf{b}_k^β is the vector of spline coefficients for the k th coefficient function. For the fixed effects in (2) evaluated on \mathbf{t}_i , substitution yields

$$\begin{aligned} \sum_{k=1}^p X_{ik}(\mathbf{t}_i) \cdot \beta_k(\mathbf{t}_i) &= \sum_{k=1}^p X_{ik}(\mathbf{t}_i) \cdot [\Theta(\mathbf{t}_i)\mathbf{b}_k^\beta] \\ &= \sum_{k=1}^p [(X_{ik}(\mathbf{t}_i) \otimes \mathbf{1}_{K_\Theta}) \cdot \Theta(\mathbf{t}_i)] \mathbf{b}_k^\beta \\ &\equiv \sum_{k=1}^p \mathbf{X}_{ik}^* \mathbf{b}_k^\beta \end{aligned}$$

where “.” is the element-wise product, “ \otimes ” is the Kronecker product, and $\mathbf{1}_K$ is a length K column vector with each entry equal to 1. Defining $\mathbf{X}_i^* = [\mathbf{X}_{i1}^* | \dots | \mathbf{X}_{ip}^*]$ and $\mathbf{b}^\beta = \left[(\mathbf{b}_1^\beta)^T | \dots | (\mathbf{b}_{K_\Theta}^\beta)^T \right]^T$,

we have that

$$\sum_{k=1}^p X_{ik}(\mathbf{t}_i) \cdot \beta_k(\mathbf{t}_i) = \mathbf{X}_i^* \mathbf{b}^\beta. \quad (3)$$

Lastly, we note that $\beta(\mathbf{t}_i)$, the $J_i \times p$ matrix of coefficient functions evaluated on the grid \mathbf{t}_i , is given by $\Theta(\mathbf{t}_i) \mathbf{B}^\beta$ where \mathbf{B}^β is the matrix of spline coefficients with k th column equal to \mathbf{b}_k^β .

Variable selection is induced through our prior specification on the \mathbf{b}_k^β . For each k , our spike-and-slab prior is

$$\mathbf{b}_k^\beta \sim N[0, (1 - \gamma_k)v_0\sigma_\epsilon^2 \mathbf{I}_{K_\Theta} + \gamma_k v_1\sigma_\epsilon^2 \mathbf{I}_{K_\Theta}]$$

where γ_k is the latent binary inclusion indicator, v_0 and v_1 control the spike and slab variances, respectively, and \mathbf{I}_K is a $K \times K$ identity matrix. Each γ_k is assigned a Bernoulli prior with probability p_γ ; in turn, p_γ is assigned a Beta[.5, .5] prior. Choices for v_0 and v_1 are discussed in Section 3.3.

FPC basis functions $\psi_k(t)$ in (2), like coefficient functions $\beta_k(t)$, are expanded in terms of a fixed spline basis; for notational convenience we use the same basis $\Theta(t)$. We let $\psi_k(\mathbf{t}_i) = \Theta(\mathbf{t}_i) \mathbf{b}_k^\psi$ where \mathbf{b}_k^ψ is the vector of spline coefficients for the k th FPC basis function. Substituting this expansion for the FPC basis functions in (2) yields

$$\begin{aligned} \sum_{k=1}^{K_\Psi} c_{ik} \psi_k(\mathbf{t}_i) &= \sum_{k=1}^{K_\Psi} c_{ik} [\Theta(\mathbf{t}_i) \mathbf{b}_k^\psi] \\ &= [\mathbf{c}_i^T \otimes \Theta(\mathbf{t}_i)] \mathbf{b}^\psi \end{aligned} \quad (4)$$

where $\mathbf{c}_i = [c_{i1}, \dots, c_{iK_\Psi}]^T$ and $\mathbf{b}^\psi = [(b_1^\psi)^T | \dots | (b_{K_\Psi}^\psi)^T]^T$. In contrast to the situation for fixed effects, in which the predictor functions are known and spline coefficients must be estimated, here both \mathbf{c}_i and \mathbf{b}^ψ are unknown. A useful expression equivalent to that in (4) emphasizing that the \mathbf{c}_i must be estimated is $[\Theta(\mathbf{t}_i) \mathbf{B}^\psi] \mathbf{c}_i$, where \mathbf{B}^ψ is the matrix of spline coefficients with k th column equal to \mathbf{b}_k^ψ .

The prior specification for parameters in the residual FPC decomposition draws on probabilistic

and Bayesian PCA for non-functional data ([Tipping and Bishop, 1999](#); [Bishop, 1999](#)). In particular, scores \mathbf{c}_i are given independent standard Normal priors $\mathbf{c}_i \sim N[0, \mathbf{I}_{K_\Psi}]$ and, for each k , spline coefficients are given Normal priors $\mathbf{b}_k^\psi \sim N[0, \sigma_{\psi_k}^2 \mathbf{I}_{K_\Theta}]$. Variances $\sigma_{\psi_k}^2$ and σ_ϵ^2 are modeled using uninformative inverse gamma priors. Thus, our complete model specification is

$$\begin{aligned}
Y_i(\mathbf{t}_i) &\sim N\left[\mathbf{X}_i^* \mathbf{b}^\beta + [\mathbf{c}_i^T \otimes \Theta(\mathbf{t}_i)] \mathbf{b}^\psi, \sigma_\epsilon^2 \mathbf{I}_{J_i}\right] \text{ for subjects } 1 \leq i \leq I; \\
\mathbf{b}_k^\beta &\sim N\left[0, (1 - \gamma_k)v_0 \sigma_\epsilon^2 \mathbf{I}_{K_\Theta} + \gamma_k v_1 \sigma_\epsilon^2 \mathbf{I}_{K_\Theta}\right] \text{ for } 1 \leq k \leq p \\
\gamma_k &\sim \text{Bernoulli}[p] \text{ for } 1 \leq k \leq p \\
p_\gamma &\sim \text{Beta}[.5, .5] \\
\mathbf{b}_k^\psi &\sim N\left[0, \sigma_{\psi_k}^2 \mathbf{I}_{K_\Theta}\right] \text{ for } 1 \leq k \leq K_\Psi \\
\mathbf{c}_i &\sim N[0, \mathbf{I}_{K_\Psi}] \text{ for subjects } 1 \leq i \leq I \\
\sigma_{\psi_k}^2 &\sim \text{IG}[.5, .5] \text{ for } k = 1 \dots K_\Psi \\
\sigma_\epsilon^2 &\sim \text{IG}[.5, .5]. \tag{5}
\end{aligned}$$

3.2 Variational algorithm

To estimate the parameters in (5) we implement a variational Bayes algorithm as an alternative to sampler-based estimation. This approach is motivated by three factors: *i* the general accuracy of variational algorithms, at least for posterior modes, to estimate model parameters *ii* the computational efficiency of variational algorithms, and *iii* the emphasis on prediction accuracy rather than on posterior inference, making accurate estimation sufficient for many cases. We also argue that traditional Markov chain Monte Carlo implement in a Gibbs sampler, for example, would perform poorly for the proposed model: the mean structure and FPC decomposition of the residual curves in the proposed model are highly correlated, which is expected to lead to poor mixing. Some possible directions for future work to address these concerns are noted in the discussion.

For a detailed introduction to variational Bayes, see [Ormerod and Wand \(2010\)](#) and [Bishop \(2006, Chapter 10\)](#); here be briefly review the general methodology. Variational Bayes methods

seek an approximation $q(\phi)$ to the full posterior $p(\phi|\mathbf{y})$ for parameter vector ϕ and data \mathbf{y} . To this end, q is restricted to a class of functions that are more tractable than the full posterior distribution. In our algorithm, we assume that the posterior factors such that $q(\phi) = \prod_{l=1}^L q_l(\phi_l)$, and each q_l is a parametric density function. Within this class, we wish to choose the element q^* that minimizes the Kullback-Leibler distance from $p(\phi|\mathbf{y})$. It can be shown that the optimal component densities q_l^* are given by

$$q_l^*(\phi_l) \propto \exp [E_{\phi_{-l}} \log p(\mathbf{y}, \phi)] \propto \exp [E_{\phi_{-l}} \log p(\phi_l | \text{rest})]$$

where $\text{rest} \equiv \{\mathbf{y}, \phi_1, \dots, \phi_{l-1}, \phi_{l+1}, \dots, \phi_L\}$ is the collection of all remaining parameters and the observed data.

The parameters of the q_l^* are updated iteratively and deterministically using expressions involving the current estimates of the remaining parameters and the data. This suggests an algorithm of the following form:

1. Initialize parameter estimates;
2. Update the mean and variance of \mathbf{b}^β ;
3. For $1 \leq k \leq p$, update the mean of γ_k ; update p_γ ;
4. Update the mean and variance of \mathbf{b}^ψ ;
5. For $i \leq i \leq I$, update the mean and variance of \mathbf{c}_i ;
6. For $1 \leq k \leq p$, update $\sigma_{\psi_k}^2$; update σ_ϵ^2 ;
7. Repeat steps 2-6 until convergence

The complete variational algorithm is given in the Supplementary Materials. We initialize parameters so that $\mathbf{b}^\beta = 0$, $\gamma_k = 1$ for all k , $\mathbf{b}^\psi = 0$, and c_{ik} is drawn from a standard Normal for all i and k . Variational algorithms are often monitored for convergence through a lower bound on the

divergence between the true posterior and the variational approximation; we have found that monitoring convergence through estimated coefficients is also reasonable, as is using a fixed (relatively large) number of iterations.

3.3 Practical concerns

3.3.1 Centering and scaling

Similarly to non-functional variable selection settings, we recommend that predictor functions are centered to have mean zero and scaled to have variance one over their domain. Additionally, we center the response function so that the intercept can be omitted from (2). To achieve this we take a two-stage approach, implemented for each predictor separately. First, we estimate the mean curve using the average of the k nearest neighbors, with k selected to include a relatively large number of observations. Second, we subtract the mean from each curve and square the difference, and smooth the squared difference curves using the same k nearest neighbors method. These steps give estimates of the mean and variance curves, and can be used to standardize predictors over the full time domain. Although this does not account for correlations within a subject, it avoids model assumptions that can result in a negative variance curve, particularly for sparse and irregular datasets.

An alternative is to center at the subject level; doing so would model the effect of changes in predictors relative to each subject's mean and variance. How to best implement this for sparse data is not obvious, and our efforts in this direction for our data did not yield results meaningfully different from those using the above approach.

3.3.2 Tuning parameters

Our variational algorithm estimates coefficient functions for fixed values of v_0 and v_1 , which control the spike and slab variances. We fix v_1 to be relatively large to avoid unnecessary penalization on selected coefficients. Allowing v_0 to have a small but positive value, rather than setting the spike

variance to zero, helps to absorb negligible nonzero coefficient functions into the spike distribution. We treat v_0 as a tuning parameter and choose its value via five-fold cross validation in which subjects, rather than observations within subjects, are partitioned into training and validation sets.

The values of K_Θ and K_Ψ , the number of spline basis functions in our expansions and the number of FPCs in our decomposition of the residual curves, are also fixed outside of our variational algorithm. Because we don't impose smoothness penalization, the value of K_Θ acts as an implicit tuning parameter. In our analyses, similar results were found for $K_\Theta = 5$ and $K_\Theta = 10$; the effect of this choice should be carefully evaluated in future applications. Similarly, we explore a range of possible values for K_Ψ . A useful strategy in practice is to initially use a large value of K_Ψ to allow the examination of a scree plot, and then to consider smaller values that nonetheless explain a large proportion of the residual variance.

3.3.3 Rotation of FPCs

The Bayesian FPC analysis, like similar approaches, does not explicitly introduce orthogonality in the estimated basis functions. Although the estimation approach is valid even without these constraints, the interpretation of the FPC analysis is more straightforward when FPCs are orthogonal. For this reason, we suggest that estimated basis functions be rotated into an equivalent orthonormal space at the convergence of the variational algorithm.

3.3.4 Implementation

Lastly, we note that all methods are implemented in the R package `vbvs.concurrent`, installable from GitHub using the following code:

```
install.packages("devtools")
devtools::install_github("jeff-goldsmit/vbvs.concurrent")
```

This package contains code for fitting the proposed model, as well as a model that uses variational

Bayes without variable selection to fit the FLCM. Helper functions for implementing five-fold cross validation to select v_0 , extracting coefficients, and making predictions are also included.

4 Numerical results

In this section we undertake analyses of the Masked Hypertension Study, introduced in Section 2. The dataset consists of 563 subjects with at least 10 ambulatory blood pressure measurements taken approximately every half-hour over the period 10:00am to 10:00pm in a single day. While both systolic and diastolic measurements were taken, we focus here on systolic BP as the outcome of interest. Additional variables were collected using ecological momentary assessments and accelerometers. In particular, we have measures of mood (excited; frustrated; happy; tired; angry; anxious), alcohol and caffeine consumption, activity type (relaxing, working, doing chores, commuting, exercising, having a meal, or none of these), current exertion level, current pain rating, posture (sitting or moving, with reference category standing), location (at work or other, with reference category at home), and physical activity in the first, second, third, fourth, and fifth minute preceding the blood pressure measurement. Because completion of the ecological momentary assessment requires the subject to be awake, and because the accelerometer was removed during sleep, we restrict our analysis to observations between 10:00am and 10:00pm. In addition to the time-varying covariate functions, we include baseline systolic and diastolic blood pressures (the mean of nine readings taken by a trained technician over three visits to the clinic). These are included as covariate functions that are constant over time.

We use the methods described in Section 3 to fit the FLCM with variable selection, and have two main objectives. First, in Section 4.1, we focus on the ability of the proposed methods to improve prediction accuracy, particularly in small samples. Second, in Section 4.2, we perform an analysis of the complete dataset and interpret the results. Throughout, we set $K_\Theta = 5$, $K_\psi = 2$, and $v_1 = 100$; similar results are obtained for other choices of these values. For reference, we also compare to a variational Bayes algorithm analogous to that in Section 3 except that it omits the variable selection

step. That is, we also fit the FLCM using only a “slab” prior on regression coefficients to assess the impact of variable selection on predictive performance and coefficient function estimation; this approach is roughly equivalent to unpenalized estimation of the FLCM.

4.1 Predictive performance

To illustrate the effect of variable selection on prediction accuracy in the FCLM, especially for small sample sizes, and to identify which variables are selected as important predictors of ambulatory blood pressure in the Masked Hypertension Study, we randomly partition the complete data into training and validation sets of size 100 and 463, respectively. We then fit model 1 to the training data using the methods described in Section 3 using five-fold cross validation to choose v_0 ; we also fit a variational approach without variable selection for comparison. Fitted values $\hat{Y}_i(t_{ij})$ for subjects i in the validation set are computed from the regression coefficients estimated from the training data and compared to observed values. This process, beginning with the partitioning into training and validation sets and ending with the computation of fitted values for the validation set, is repeated 250 times. For each iteration, we record which variables are selected, the estimated coefficients $\hat{\beta}_k(t)$, and the average prediction mean squared error $APMSE = \frac{1}{\sum_i J_i} \sum_i \sum_{j \in J_i} (Y_i(t_{ij}) - \hat{Y}_i(t_{ij}))^2$ for subjects i in the validation set.

Figure 2 shows the frequency with which each possible variable is selected across the 250 replications. As expected, clinical SBP was selected as relevant in almost all replications. Surprisingly, however, other variables are never or only rarely selected, suggesting that the variation observed in Figure 1 is either measurement error or biological variation not directly attributable to measure covariates.

The left panel of Figure 3 shows the APMSE for each replication. In addition to the variational Bayes algorithms with and without variable selection (abbreviated VBVS and VB, respectively), we include two other models for comparison. First, we include an “intercept only” model; because both predictor and response curves have been centered, this model sets $\hat{Y}_i(t_{ij}) = 0$ for all i and j . Second, based on the results in Figure 2, the “Clinical SBP Only” model fits the FLCM with only

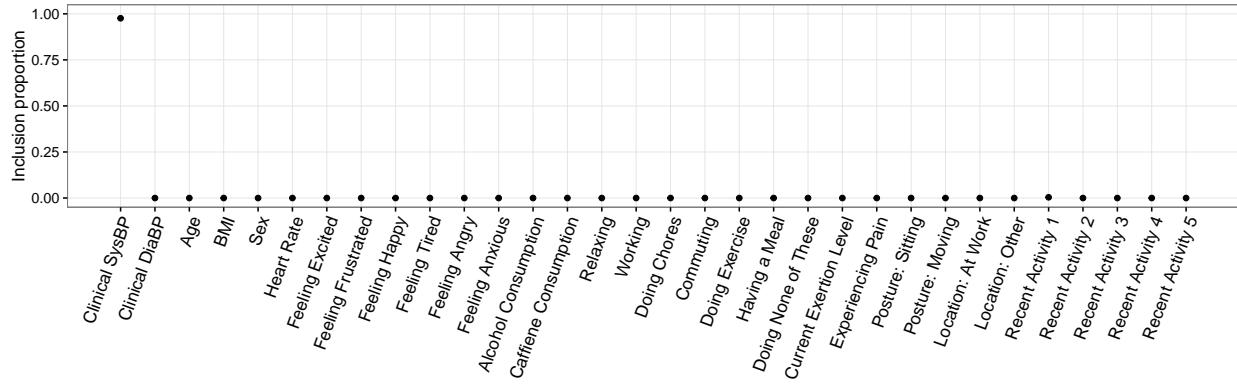


Figure 2: Frequency with which potential covariates were selected by the proposed method.

clinical SBP as a predictor; because there is a single predictor, this model is fit with no variable selection. The VBVS approach substantially outperforms the intercept only model in terms of median APMSE, because clinical SBP usefully predicts ambulatory SBP. The VBVS approach also substantially outperforms the VB method, emphasizing the detrimental effect of including unimportant covariates on prediction accuracy. The VBVS approach slightly outperforms the “Clinical SBP Only” model even though, in most cases, the VBVS method only selects clinical SBP as a predictor. We speculate that this slight improvement stems from the heavy ridge penalization on non-selected covariates, rather than their complete omission from the fitted model. Lastly, we note that the outlying APMSE for the VBVS method are the result of cases in which no variables are selected.

The right panels of Figure 3 show the estimated coefficient functions for clinical SBP and current exertion level for the VB, VBVS, and Clinical SBP only models across all replications. Recall that predictor functions are centered and scaled, meaning that coefficient functions represent the change in systolic ABP for a one standard deviation increase in the predictor. Estimated coefficient functions for the effect of SBP are more variable for VB than for VBVS, illustrating the added noise in estimates when unimportant predictors are included; meanwhile, the estimates for the VBVS and Clinical SBP Only models are nearly indistinguishable. Estimates of the coefficient for exertion are near (but not exactly equal to) zero for the VBVS approach in each replication, in contrast

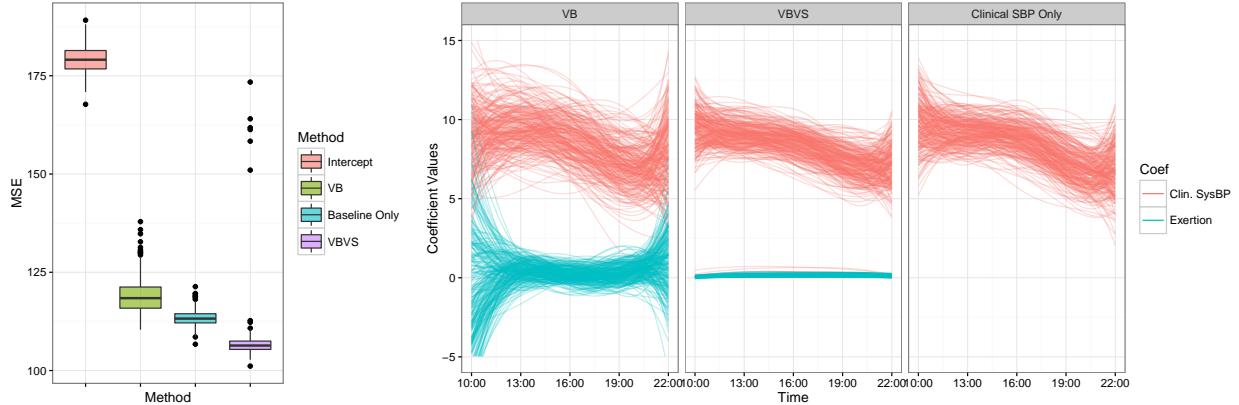


Figure 3: The left panel shows boxplots of the average prediction mean squared error for each of four approaches: intercept only, variational Bayes without variable selection (VB), variational Bayes with variable selection (VBVS), and a model with only clinical SBP as a predictor. The right panels show the estimated coefficient functions for clinical SBP and current exertion level for VB, VBVS, and Clinical SBP Only.

to the noisy estimates of the VB method. Across replications, the coefficient for clinical SBP is positive over the course of the day; it is roughly flat until the late afternoon, at which time the coefficient function decreases slightly. This dip may reflect a weakening of the association between clinic SBP, which is taken in the daytime, and ambulatory SBP readings in the evening.

4.2 Full data analysis

Our study of prediction accuracy in Section 4.1 established the importance of variable selection in the context of the FLCM, particularly for relatively small training sets. We now present results from the analysis of the complete Masked Hypertension Study data using the proposed approach; once again, we compare to a variational Bayes algorithm without variable selection to provide context. Fixed values for K_Θ , K_ψ and v_1 and the range of values for v_0 , chosen by five-fold cross validation, are as above.

Coefficient functions estimated using variational Bayes with and without variable selection are shown in the right and left panels of Figure 4. In both panels, we highlight the coefficient functions for clinical SBP and for current exertion level to aid in comparing these to estimates shown in Figure 3. In the VBVS analysis, only clinical SBP is selected as an important predictor; the estimate and

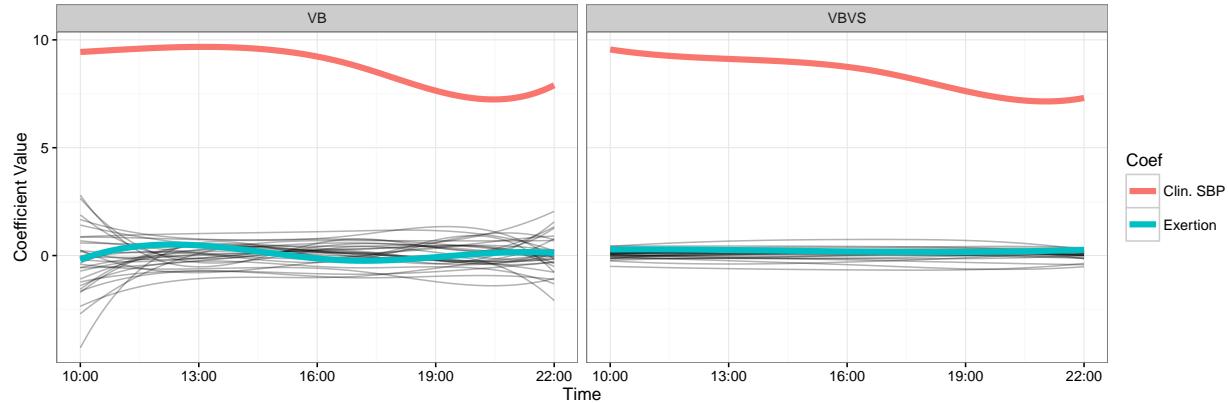


Figure 4: Coefficient functions for each covariate, estimated using variational Bayes without variable selection (“VB”, left panel) and with variable selection (“VBVS”, right panel), in the full data analysis. Coefficients for clinic SBP and Exertion are colored to correspond to Figure 3; other coefficients are shown in black.

its interpretation is similar to that in Figure 3. Remaining coefficient functions are shrunk toward zero over their full domain. There are two clear groups of estimated coefficient functions in the VB analysis: most are clustered near zero, while the coefficient for clinical SBP alone is far from zero. The comparison of these results with those for VBVS indicates the effect of the “spike” prior on the magnitude and variability of coefficient functions for non-selected covariates.

Our proposed approach jointly models the coefficient functions using a spike-and-slab prior and residual correlation using FPCA. While our emphasis in this manuscript is on the former, the results of the FPC analysis are useful in understanding the structure of the variability that is unexplained by covariate functions. To illustrate these results for the Masked Hypertension Study, Figure 5 shows $\pm\sqrt{\lambda_k}\psi_k(t)$ for $k = 1$ and 2 in the left and right panels, respectively. The first FPC is largely a mean shift, indicating that some subjects have uniformly higher or lower ambulatory SBP than expected based on their covariates. Those who are higher show a masked hypertension effect in which ambulatory BP is higher than expected based on clinic BP, while those who are lower show a white-coat effect in which the clinic setting induces higher BP than is observed in ambulatory settings. The second FPC highlights distinct time-of-day effects: subjects with high scores for FPC 2 have greater than expected ambulatory SBP in the morning and afternoon and lower than expected SBP in the evening. Thus the second FPC may differentiate those who have

more home stress than work stress and vice versa.

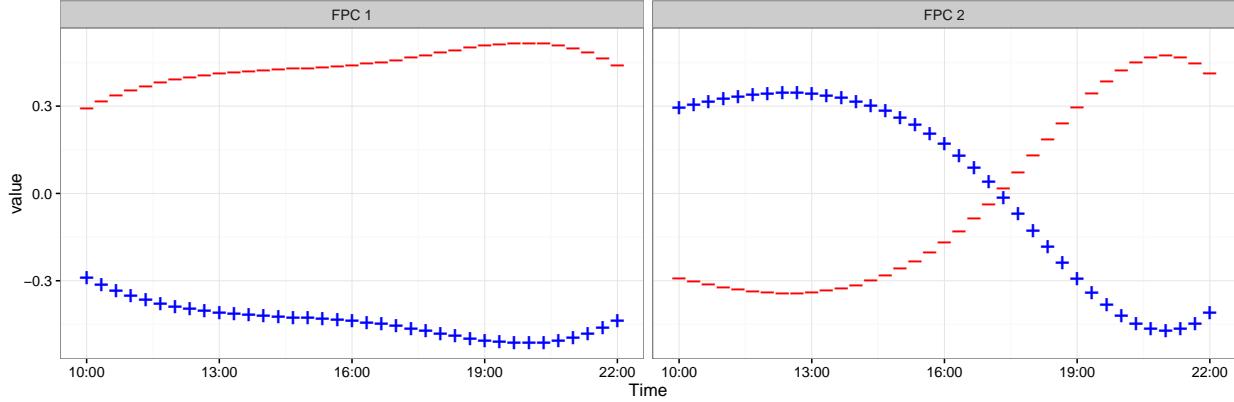


Figure 5: Illustration of the effects of FPC basis functions. Panels are $\pm\sqrt{\lambda_k}\psi_k(t)$ for $k = 1$ (left) and $k = 2$ (right).

5 Concluding remarks

We have developed methods for variable selection in the functional linear concurrent model; although we are motivated by a specific application, the methods are general and broadly useful. There are several key contributions to the functional data analysis literature: the use of Bayesian variable selection methods, in contrast to the currently-used group lasso or similar approaches; jointly modeling the regression coefficients with sparsity and the residual covariance structure; support for both densely- and sparsely-sampled functional responses; and the availability of robust, user-friendly code in an R package.

In our application, it is perhaps surprising that so few of the collected variables were selected as predictors of systolic blood pressure: these variables were recorded specifically because such associations were anticipated. There are several implications to this finding that are worth considering. First, we acknowledge that measurement of relevant variables may be imperfect. As with all surveys, the ecological momentary assessment relies on self-report; for variables like mood or anger, there may be individual-level differences in the rating of these emotions. Multiple parallel data streams also require careful alignment of recording devices; it is possible that the accelerometer

and ambulatory blood pressure monitor were mis-aligned for some subjects, although steps were taken to prevent this. Second, we note that the subject-level patterns uncovered by the FPCA decomposition of the residuals indicates subject-level effects that are unexplained by baseline variables or covariate functions, and there remains substantial unexplained variation in ambulatory blood pressure readings. Whether this is natural biological variation, measurement error, or the result of other concurrent covariates is unclear. Lastly, we emphasize that although only clinical SBP was selected as a predictor of ambulatory SBP, the proposed approaches were necessary to make this finding. We also acknowledge that it is possible that some EMA variables are significantly associated with ambulatory BP, but the effects are sufficiently small enough to omitted in our prediction-focused analysis.

Several future directions for statistical work are possible. Adaptations of our methods for other functional regression settings, such as function-on-scalar and scalar-on-function, are possible. In function-on-scalar regression specifically, our joint modeling approach may have important benefits over the “pre-whitening” approach of [Chen et al. \(2016\)](#): extensions to multilevel functional responses appears more straightforward in the Bayesian setting. Including an explicit penalty on the smoothness of coefficient functions for selected covariates would be similar to estimation strategies without variable selection; however, how to include both smoothness and sparsity constraints in our current framework is not clear. The development of a sampling-based Bayesian approach is also reasonable, although it should be undertaken with some care. In [Goldsmith et al. \(2015\)](#), the authors pose a function-on-scalar regression with an FPCA decomposition of residual curves and use a Hamiltonian Monte Carlo sampler ([Hoffman and Gelman, 2011](#)), implemented in [Stan](#) ([Stan Development Team, 2013](#)); a similar implementation may be appropriate here. Alternative approaches to modeling residual correlation can also be considered; in [Goldsmith and Kitago \(2016\)](#) a Wishart prior models this correlation, and Gibbs sampling worked well. Whether a similar approach is suitable for higher-dimensional curves and sparse data is unclear. Lastly, a mixture of the concurrent functional model and the historic functional model ([Malfait and Ramsay, 2003](#)) could be relevant for including the effect of recent physical activity on current blood pressure, although

based on our findings for the Masked Hypertension Study such methodological developments may not be necessary.

6 Acknowledgments

The first author's research was supported in part by Award R01HL123407 from the National Heart, Lung, and Blood Institute and by Award R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering. The second author's research, including the collection of all data for the Masked Hypertension Study, was supported by Award P01HL047540 from the National Heart, Lung, and Blood Institute.

References

- Abdalla, M., Goldsmith, J., Muntner, P., Diaz, K. M., Reynolds, K., Schwartz, J. E., and Shimbo, D. “Is Isolated Nocturnal Hypertension A Reproducible Phenotype?” *American journal of hypertension*, hpy058 (2015).
- Barber, R. F., Reimherr, M., and Schill, T. “The Function-on-Scalar LASSO with Applications to Longitudinal GWAS.” *Under Review* (2016).
- Bishop, C. M. “Bayesian PCA.” *Advances in Neural Information Processing Systems*, 382–388 (1999).
- . *Pattern Recognition and Machine Learning*. New York: Springer (2006).
- Chen, Y., Goldsmith, J., and Ogden, T. “Variable Selection in Function-on-Scalar Regression.” *Stat* (2016).
- Earls, C. and Hooker, G. “Combining Functional Data Registration and Factor Analysis.” *Journal of Computational and Graphical Statistics* (2016).
- Fan, J. and Zhang, W. “Statistical methods with varying coefficient models.” *Statistics and its Interface*, 1:179 (2008).
- George, E. I. and McCulloch, R. E. “Variable selection via Gibbs sampling.” *Journal of the American Statistical Association*, 88:881–889 (1993).
- . “Approaches for Bayesian Variable Selection.” *Statistica Sinica*, 7:339–373 (1997).
- Gertheiss, J., Goldsmith, J., Crainiceanu, C., and Greven, S. “Longitudinal Scalar-on-Functions Regression with Application to Tractography Data.” *Biostatistics*, 14:447–461 (2013).

- Goldsmith, J. and Kitago, T. “Assessing Systematic Effects of Stroke on Motor Control using Hierarchical Function-on-Scalar Regression.” *Journal of the Royal Statistical Society: Series C*, 65:215–236 (2016).
- Goldsmith, J., Wand, M. P., and Crainiceanu, C. M. “Functional Regression via Variational Bayes.” *Electronic Journal of Statistics*, 5:572–602 (2011).
- Goldsmith, J., Zipunnikov, V., and Schrack, J. “Generalized multilevel function-on-scalar regression and principal component analysis.” *Biometrics*, 71:344–353 (2015).
- Hoffman, M. D. and Gelman, A. “The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo.” *arXiv preprint arXiv:1111.4246* (2011).
- Malfait, N. and Ramsay, J. O. “The historical functional linear model.” *Canadian Journal of Statistics*, 31:115–128 (2003).
- McLean, M. W., Scheipl, F., Hooker, G., Greven, S., and Ruppert, D. “Bayesian Functional Generalized Additive Models for Sparsely Observed Covariates.” *Under Review* (2013).
- Ormerod, J. and Wand, M. P. “Explaining Variational Approximations.” *The American Statistician*, 64:140–153 (2010).
- Pickering, T. G., Davidson, K., Gerin, W., and Schwartz, J. E. “Masked hypertension.” *Hypertension*, 40:795–796 (2002).
- Ramsay, J. O. and Silverman, B. W. *Functional Data Analysis*. New York: Springer (2005).
- Ročková, V. and George, E. I. “EMVS: The EM approach to bayesian variable selection.” *Journal of the American Statistical Association*, 109:828–846 (2014).
- Şentürk, D. and Nguyen, D. V. “Varying coefficient models for sparse noise-contaminated longitudinal data.” *Statistica Sinica*, 21:1831–1856 (2011).
- Shimbo, D., Newman, J. D., and Schwartz, J. E. “Masked hypertension and prehypertension: diagnostic overlap and interrelationships with left ventricular mass: the Masked Hypertension Study.” *American journal of hypertension*, 25:664–671 (2012).
- Stan Development Team. *Stan Modeling Language User’s Guide and Reference Manual, Version 1.3* (2013).
URL <http://mc-stan.org/>
- Tipping, M. E. and Bishop, C. “Probabilistic Principal Component Analysis.” *Journal of the Royal Statistical Society: Series B*, 61:611–622 (1999).
- van der Linde, A. “Variational Bayesian Functional PCA.” *Computational Statistics and Data Analysis*, 53:517–533 (2008).

—. “A Bayesian latent variable approach to functional principal components analysis with binary and count.” Advances in Statistical Analysis, 93:307–333 (2009).

Supplementary Material for: Variable Selection in the Functional Linear
Concurrent Model

Jeff Goldsmith and Joseph E. Schwartz

A Variational algorithm

In this Section we provide the complete variational algorithm for approximate Bayesian inference for our proposed variable selection approach to the functional linear concurrent model (FLCM). For completeness, we recall that our conceptual model is

$$Y_i(t) = \sum_{k=1}^p X_{ik}(t)\beta_k(t) + \sum_{k=1}^{K_\Psi} c_{ik}\psi_k(t) + \epsilon_i(t).$$

and that our model specification for observed data is

$$\begin{aligned} Y_i(\mathbf{t}_i) &\sim N \left[\mathbf{X}_i^* \mathbf{b}^\beta + [\mathbf{c}_i^T \otimes \Theta(\mathbf{t}_i)] \mathbf{b}^\psi, \sigma_\epsilon^2 \mathbf{I}_{J_i} \right] \text{ for subjects } 1 \leq i \leq I; \\ \mathbf{b}^\beta &\sim N [0, \sigma_\epsilon^2 D] \text{ for } 1 \leq k \leq p \\ \gamma_k &\sim \text{Bernoulli}[p] \text{ for } 1 \leq k \leq p \\ p_\gamma &\sim \text{Beta}[.5, .5] \\ \mathbf{b}_k^\psi &\sim N [0, \sigma_{\psi_k}^2 \mathbf{I}_{K_\Theta}] \text{ for } 1 \leq k \leq K_\Psi \\ \mathbf{c}_i &\sim N [0, \mathbf{I}_{K_\Psi}] \text{ for subjects } 1 \leq i \leq I \\ \sigma_{\psi_k}^2 &\sim \text{IG} [.5, .5] \text{ for } k = 1 \dots K_\Psi \\ \sigma_\epsilon^2 &\sim \text{IG} [.5, .5]. \end{aligned} \tag{A.1}$$

We note that our prior for \mathbf{b}^β is not separated for individual vectors k . Instead, we define the overall covariance $D = \text{diag}((1 - \gamma_k)v_0 + \gamma_k v_1) \otimes \mathbf{I}_{K_\Theta}$

Our variational approach assumes that the posterior distribution

$$p\left(\mathbf{b}^\beta, \gamma_1, \dots, \gamma_p, p_\gamma, \mathbf{b}^\psi, c_1, \dots, c_I, \sigma_{\psi_1}^2, \dots, \sigma_{\psi_{K_\Psi}}^2, \sigma_\epsilon^2 | Y\right)$$

can be approximated using

$$\begin{aligned} q\left(\mathbf{b}^\beta, \gamma_1, \dots, \gamma_p, p_\gamma, \mathbf{b}^\psi, c_1, \dots, c_I, \sigma_{\psi_1}^2, \dots, \sigma_{\psi_{K_\Psi}}^2, \sigma_\epsilon^2 | Y\right) \\ = q(\mathbf{b}^\beta)q(\gamma_1, \dots, \gamma_p)q(p_\gamma)q(\mathbf{b}^\psi)q(c_1, \dots, c_I)q(\sigma_{\psi_1}^2, \dots, \sigma_{\psi_{K_\Psi}}^2, \sigma_\epsilon^2) \end{aligned}$$

where the functions q are distinguished by their argument. The additional factorizations

$$\begin{aligned} q(\gamma_1, \dots, \gamma_p) &= \prod_{k=1}^p q(\gamma_k) \\ q(c_1, \dots, c_I) &= \prod_{i=1}^I q(c_i) \\ q(\sigma_{\psi_1}^2, \dots, \sigma_{\psi_{K_\Psi}}^2, \sigma_\epsilon^2) &= \left(\prod_{k=1}^{K_\Psi} q(\sigma_{\psi_k}^2) \right) q(\sigma_\epsilon^2) \end{aligned}$$

follow from the conditional independence of these parameters. Thus our final factorization is

$$q(\mathbf{b}^\beta) \left(\prod_{k=1}^p q(\gamma_k) \right) q(p_\gamma)q(\mathbf{b}^\psi) \left(\prod_{i=1}^I q(c_i) \right) \left(\prod_{k=1}^{K_\Psi} q(\sigma_{\psi_k}^2) \right) q(\sigma_\epsilon^2).$$

In our model specification, the $q(\cdot)$ functions have parametric exponential family forms. Updates for the parameters of these densities are derived using

$$q_l^*(\phi_l) \propto \exp [E_{\phi_{-l}} \log p(\mathbf{y}, \phi)] \propto \exp [E_{\phi_{-l}} \log p(\phi_l | \text{rest})]$$

where $\text{rest} \equiv \{\mathbf{y}, \phi_1, \dots, \phi_{l-1}, \phi_{l+1}, \dots, \phi_L\}$ is the collection of all remaining parameters and the observed data; see e.g. [Goldsmith et al. \(2011\)](#) and [Goldsmith and Kitago \(2016\)](#) for a thorough illustration of this process. Notationally, we let $\mu_{q(\phi)}$ and $\Sigma_{q(\phi)}$ indicate the mean and variance of

the density $q(\phi)$; for inverse gamma densities, the parameters of interest are $A_{q(\phi)}$ and $B_{q(\phi)}$. The derivation of optimal density parameters in our model yields the iterative algorithm below:

Initialize:

- $\mu_{q(\gamma_k)} = 1$ for all k ;
- $\mu_{q(D^{-1})} = \text{diag}\left(\frac{1-\mu_{q(\gamma)}}{v_0} + \frac{\mu_{q(\gamma)}}{v_1}\right) \otimes \mathbf{I}_{K_\Theta}$;
- $\mu_{q(b^\psi)} = 0$;
- $\Sigma_{q(c_i)} = \mathbf{I}_{k_\Psi}$ for all i ;
- $\mu_{q(c_{ik})}$ as a draw from a standard Normal for all i and k ;
- $A_{q(p_\gamma)} = .5 + p$ and $B_{q(p_\gamma)} = .5$
- $B_{q(\sigma_{\psi_k}^2)} = 1$ for all k
- $B_{q(\sigma_e^2)} = 1$

Cycle until convergence:

- $\Sigma_{q(b^\beta)} \leftarrow [\mu_{q(1/\sigma_e^2)} (\sum_i X_i^{*T} X_i^* + \mu_{q(D^{-1})})]^{-1}$
- $\mu_{q(b^\beta)} \leftarrow \mu_{q(1/\sigma_e^2)} \Sigma_{q(b^\beta)} \left[\sum_i X_i^{*T} (Y_i(\mathbf{t}_i) - \Theta(\mathbf{t}_i) \mu_{q(B^\psi)} \mu_{q(c_i)}) \right]$
- For $1 \leq k \leq p$:
 - $\mu_{q(\gamma_k)} \leftarrow \left(\frac{v_1}{v_0}\right)^{K_\Theta/2} \exp \left\{ \frac{-1}{2} \left(\frac{1}{v_1} - \frac{1}{v_0} \right) \mu_{q(1/\sigma_e^2)} \mu_{q(b_k^\beta)}^T \mu_{q(b_k^\beta)} + (\mu_{q(\ln(p))} - \mu_{q(\ln(1-p))}) \right\}$
- $\mu_{q(D^{-1})} \leftarrow \text{diag}\left(\frac{1-\mu_{q(\gamma)}}{v_0} + \frac{\mu_{q(\gamma)}}{v_1}\right) \otimes \mathbf{I}_{K_\Theta}$
- $A_{q(p)} \leftarrow .5 + \sum_k \mu_{q(\gamma_k)}$ and $B_{q(p)} \leftarrow .5 + p - \sum_k \mu_{q(\gamma_k)}$
- $\Sigma_{q(b^\psi)} \leftarrow [\mathbf{I}_{K_\Theta} \otimes \text{diag}(\mu_{q(1/\sigma_{\psi_k}^2)}) + \mu_{q(1/\sigma_e^2)} \sum_i (E(\mathbf{c}_i \mathbf{c}_i^T) \otimes \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i))]^{-1}$
- $\mu_{q(b^\psi)} \leftarrow \mu_{q(1/\sigma_e^2)} \Sigma_{q(b^\psi)} \sum_i \left[(\mu_{q(c_i)} \otimes \Theta(\mathbf{t}_i)) (Y_i(\mathbf{t}_i) - X_i^* \mu_{q(B^\beta)} \mu_{q(c_i)}) \right]$
- For $1 \leq i \leq I$:
 - $\Sigma_{q(c_i)} \leftarrow [\mathbf{I}_{K_\Psi} + \mu_{q(1/\sigma_e^2)} E(B^{\psi T} \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i) B^\psi)]^{-1}$
 - $\mu_{q(c_i)} \leftarrow \mu_{q(1/\sigma_e^2)} \Sigma_{q(c_i)} \mu_{q(B^\psi)}^T \Theta^T(\mathbf{t}_i) (y_i(\mathbf{t}_i) - X_i^* \mu_{q(B^\beta)})$
- For $1 \leq k \leq K_\Psi$:
 - $B_{q(\sigma_{\psi_k}^2)} \leftarrow .5 + .5 \left[\mu_{q(b_k^\psi)}^T \mu_{q(b_k^\psi)} + \text{tr}(\Sigma_{q(b_k^\psi, b_k^\psi)}) \right]$

- $b_{q(\sigma_\epsilon^2)} \leftarrow .5 + .5 \left\{ \sum_i \left(Y_i(\mathbf{t}_i) - X_i^* \mu_{q(\mathbf{b}^\beta)} - \Theta(\mathbf{t}_i) \mu_{q(\mathbf{B}^\psi)} \mu_{q(\mathbf{c}_i)} \right)^T \left(Y_i(\mathbf{t}_i) - X_i^* \mu_{q(\mathbf{b}^\beta)} - \Theta(\mathbf{t}_i) \mu_{q(\mathbf{B}^\psi)} \mu_{q(\mathbf{c}_i)} \right) \right. \\ \left. + tr [X_i^{*T} X_i^* \Sigma_{q(\mathbf{c}_i)}] + tr [E(B^{\psi T} \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i) B^\psi) E(\mathbf{c}_i \mathbf{c}_i^T)] \right\}$
-

In this algorithm, the following expressions are useful:

$$\begin{aligned} \mu_{q(1/\sigma_\epsilon^2)} &= \frac{.5 + \frac{\sum_i J_i}{2}}{B_q(1/\sigma_\epsilon^2)} \\ \mu_{q(1/\sigma_{\psi_k}^2)} &= \frac{.5 + \frac{K_\Theta}{2}}{B_q(1/\sigma_{\psi_k}^2)} \text{ for all } k \\ \mu_{q(\log(p_\gamma))} &= \ln(.5 + p - \sum_k \mu_{q(\gamma_k)}) - \eta(.5 + \sum_k \mu_{q(\gamma_k)}) \\ \mu_{q(\log(1-p_\gamma))} &= \ln(.5 + \sum_k \mu_{q(\gamma_k)}) - \eta(.5 + p - \sum_k \mu_{q(\gamma_k)}) \end{aligned}$$

where $\eta(\cdot)$ is the digamma function. Additionally,

$$E(\mathbf{c}_i \mathbf{c}_i^T) = \mu_{q(\mathbf{c}_i)} \mu_{q(\mathbf{c}_i)}^T + \Sigma_{q(\mathbf{c}_i)}$$

and

$$E(B^{\psi T} \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i) B^\psi) = M + \mu_{q(\mathbf{B}^\psi)}^T \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i) \mu_{q(\mathbf{B}^\psi)}$$

where M is $K_\Theta \times K_\Theta$ matrix with i, j th entry $tr \left(\Sigma_{q(b_i^\psi, b_j^\psi)} \Theta^T(\mathbf{t}_i) \Theta(\mathbf{t}_i) \right)$.