

A Bayesian Regression Model for Multivariate Functional Data

BY ORI ROSEN

*Department of Mathematical Sciences, University of Texas at El Paso,
El Paso, Texas 79968, U.S.A.*

ori@math.utep.edu

AND WESLEY K. THOMPSON

*Department of Psychiatry, University of California San Diego,
La Jolla, California, U.S.A.*

wes.stat@gmail.com

ABSTRACT

In this paper we present a model for the analysis of multivariate functional data with unequally spaced observation times that may differ among subjects. Our method is formulated as a Bayesian mixed-effects model in which the fixed part corresponds to the mean functions, and the random part corresponds to individual deviations from these mean functions. Covariates can be incorporated into both the fixed and the random effects. The random error term of the model is assumed to follow a multivariate Ornstein-Uhlenbeck process. For each of the response variables, both the mean and the subject-specific deviations are estimated via low-rank cubic splines using radial basis functions. Inference is performed via Markov chain Monte Carlo methods.

1 Introduction

The term functional data analysis describes nonparametric analyses of longitudinal data which focus on the curves themselves as the basic unit of data. Some of the goals

of functional data analysis include exploring individual variation of curves from an overall mean function, and modeling the dependence of the curves on covariates. The mean function, as well as the subject-specific functions are estimated nonparametrically. In this paper we propose a method for analyzing multivariate functional data with unequally spaced observation times that may differ among subjects. It is assumed that all variables are observed at the same time points. Fitting a regression model with a multivariate response may be done by either fitting a separate regression for each of the response variables or by fitting a single regression with all response variables simultaneously. The latter may be advantageous if the error terms corresponding to each variable are correlated. Thus, fewer observations may be required to obtain reliable nonparametric function estimates compared to fitting each regression separately and ignoring the correlation. This has been shown to be the case in seemingly unrelated regression (see for example Smith and Kohn, 2000).

Our method is formulated as a Bayesian mixed-effects model in which the fixed part corresponds to the mean functions, and the random part corresponds to individual deviations from these mean functions. Covariates can be incorporated into both the fixed and the random effects. The random error term of the model is assumed to follow a first order continuous-time multivariate autoregression, also known as a multivariate Ornstein-Uhlenbeck process. For each of the response variables, both the mean and the subject-specific deviations are estimated via low-rank cubic splines using radial basis functions. Inference is performed via Markov chain Monte Carlo methods.

Our model is closest in spirit to the functional mixed effects model of Guo (2002), where the fixed and random effects are modeled by cubic smoothing splines. However, Guo's model accommodates only a univariate response variable, and does not allow correlated error terms. It can be fit either via standard mixed effects software or by

Kalman filtering. Inference and model selection are based on a generalized maximum likelihood ratio test. Baladandayuthapani et al. (2008) have proposed a Bayesian model for spatially correlated functional data analysis. The smoothing technique they use is similar to ours, but their emphasis is on spatial correlation rather than on temporal correlation. Smith and Kohn (2000) consider multivariate nonparametric regression using the seemingly unrelated regression approach. They show that if the error terms of the regression equations are correlated, better nonparametric estimates of the regression functions are obtained by accounting for this correlation compared to fitting separate regressions ignoring the correlation. It is noted that Smith and Kohn (2000) consider multivariate nonparametric regression, not functional data analysis. In functional data analysis, each individual subject has its own function which needs to be estimated for each variable. Smith and Kohn (2000) only estimate a single function for each variable.

The Ornstein-Uhlenbeck process has been used before in various contexts. Unlike most diffusion processes, its transition density is available in closed form, which results in a closed-form expression for the likelihood function. Jones (1993), Chapter 8, uses a state-space approach to parameter estimation. Sy et al. (1997) present a model for multivariate repeated measures which allows unequally spaced observations by using the multivariate integrated Ornstein-Uhlenbeck process. The fixed and random effects in their model have parametric forms. Markov chain Monte Carlo methods for inference on the Ornstein-Uhlenbeck process (univariate or multivariate) parameters have also been proposed. A recent review of estimation for discretely observed diffusion processes is given in Beskos et al. (2006). Golightly and Wilkinson (2006) discuss Bayesian inference for nonlinear multivariate diffusions. A number of authors have assumed a common spacing between the observed times. Blackwell (2003) takes this common spacing to be the most frequently occurring interval between observations. De la Cruz-

Mesía and Marshall (2003, 2006) discuss the univariate Ornstein-Uhlenbeck process and take the common spacing to be the average time difference between two consecutive observations.

The example for application of our methodology is taken from a recent psychiatric study comparing psychotherapy to pharmacotherapy carried out at the University of Pittsburgh and the University of Pisa, Italy (Frank, et al., 2008). This study sought differential baseline predictors of response to these two forms of treatment of major depression. Here, we examine the interaction effect of treatment group with Lifetime Depressive Spectrum symptoms (LDS; Cassano et al, 1997) in 252 patients entering the study in an acutely depressive episode. Levels of depression are determined by the clinician-administered Hamilton Rating Scale for Depression (HRSD) and the Quick Inventory for Depression Self-report (QIDS). These two scales were given to patients at baseline and again roughly weekly over the course of each subject's acutely depressive episode.

Our main contribution in this paper is accommodating multivariate functional data including covariates and accounting for correlation across variables and time using smoothing techniques in combination with modeling the error term via the multivariate Ornstein-Uhlenbeck process.

The rest of the paper is organized as follows. In Section 2, we describe our model, the prior distributions and the sampling scheme. Section 3 provides results of a simulation study. Section 4 discusses an application, and Section 5 ends with a brief discussion.

2 The model, priors and sampling scheme

2.1 The model

Suppose $\mathbf{y}_i(t_{ij})$ is a $p \times 1$ vector of response variables on subject i at time t_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m_i$, and consider the model

$$\mathbf{y}_i(t_{ij}) = X_{ij}\boldsymbol{\mu}(t_{ij}) + Z_{ij}\mathbf{g}_i(t_{ij}) + \boldsymbol{\delta}_i(t_{ij}) . \quad (1)$$

In (1), $\boldsymbol{\mu}(t) = (\boldsymbol{\mu}'_1(t), \dots, \boldsymbol{\mu}'_p(t))'$ and $\mathbf{g}_i(t) = (\mathbf{g}'_{i1}(t), \dots, \mathbf{g}'_{ip}(t))'$, where $\boldsymbol{\mu}_k(t) = (\mu_{k1}(t), \dots, \mu_{kr}(t))'$ and $\mathbf{g}_{ik}(t) = (g_{ik1}(t), \dots, g_{iks}(t))'$ are an $r \times 1$ vector of fixed functions, and an $s \times 1$ vector of random functions, respectively, for $k = 1, \dots, p$. Associated with $\boldsymbol{\mu}(t_{ij})$ is an $r \times 1$ covariate vector \mathbf{x}_{ij} , and with $\mathbf{g}_i(t_{ij})$ – an $s \times 1$ covariate vector \mathbf{z}_{ij} , such that $X_{ij} = I_p \otimes \mathbf{x}'_{ij}$ and $Z_{ij} = I_p \otimes \mathbf{z}'_{ij}$, where I_p is a $p \times p$ identity matrix, and \otimes denotes the Kronecker product. We have assumed here that the p response variables share the same covariates.

Before proceeding to specify the $p \times 1$ vector of random errors, $\boldsymbol{\delta}_i(t_{ij})$, we give an example which is a special case of model (1). Suppose $p = 2$, $r = 2$ and $s = 1$, with \mathbf{x}_{ij} taking values in $\{(1 \ 0)', (1 \ 1)'\}$, and $\mathbf{z}_{ij} = 1$ with corresponding functions $\boldsymbol{\mu}_1(t) = (\mu_{11}(t), \mu_{12}(t))'$, $\boldsymbol{\mu}_2(t) = (\mu_{21}(t), \mu_{22}(t))'$, $g_{i1}(t)$ and $g_{i2}(t)$. In this case, model (1) reduces to

$$\begin{aligned} y_{i1}(t_{ij}) &= \mu_{11}(t_{ij}) + x_{ij2} \cdot \mu_{12}(t_{ij}) + g_{i1}(t_{ij}) + \delta_{i1}(t_{ij}) \\ y_{i2}(t_{ij}) &= \mu_{21}(t_{ij}) + x_{ij2} \cdot \mu_{22}(t_{ij}) + g_{i2}(t_{ij}) + \delta_{i2}(t_{ij}) , \end{aligned}$$

where x_{ij2} , the second entry of \mathbf{x}_{ij} , can take the values 0 or 1. In this example, there are two groups of subjects – control ($x_{ij2} = 0$) and treatment ($x_{ij2} = 1$). Each group has its own mean curve for each of the two variables, and individual deviations from these curves are accommodated by the random functions $g_{i1}(t)$ and $g_{i2}(t)$, $i = 1, \dots, n$.

The error term in model (1) is assumed to follow a multivariate Ornstein-Uhlenbeck process. More specifically, $\boldsymbol{\delta}_i(t)$ satisfies the stochastic differential equation

$$d\boldsymbol{\delta}_i(t) = -A\boldsymbol{\delta}_i(t) + Bd\mathbf{W}_i(t) ,$$

where A and B are $p \times p$ matrices of full rank common to all $i = 1, \dots, n$, and $\mathbf{W}_i(t)$ is the p -dimensional Wiener process. Three properties of the Ornstein-Uhlenbeck process (Gardiner, 1983 , pp. 110–111) which will be useful in what follows are

1. The Ornstein-Uhlenbeck process will be stationary provided the eigenvalues of A have positive real parts.
2. The solution Σ to the matrix equation $A\Sigma + \Sigma A' = BB'$ is the stationary variance-covariance matrix of the process.
3. In the stationary state, the covariance of $\boldsymbol{\delta}_i(t)$ and $\boldsymbol{\delta}_i(s)$, for $s < t$, is

$$\text{Cov}(\boldsymbol{\delta}_i(t), \boldsymbol{\delta}_i(s)) = \exp\{-A(t - s)\}\Sigma . \quad (2)$$

Let $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ for $j = 1, \dots, m_i$, where $t_{i0} = 0$. The transition density of the Ornstein-Uhlenbeck process is given by

$$p(\boldsymbol{\delta}_i(t_{ij}) | \boldsymbol{\delta}_i(t_{i,j-1}), \Delta t_{ij}) \propto |\Omega_{\Delta t_{ij}}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}'_{t_{ij}} \Omega_{\Delta t_{ij}}^{-1} \boldsymbol{\gamma}_{t_{ij}}\right\} , \quad (3)$$

where $\boldsymbol{\gamma}_{t_{ij}} = \boldsymbol{\delta}_i(t_{ij}) - \exp(-A\Delta t_{ij})\boldsymbol{\delta}_i(t_{i,j-1})$ and $\Omega_{\Delta t_{ij}} = \Sigma - \exp(-A\Delta t_{ij})\Sigma \exp(-A'\Delta t_{ij})$.

The functions $\mu_{kl}(t)$ and $g_{ikm}(t)$, $k = 1, \dots, p$, $l = 1, \dots, r$, $m = 1, \dots, s$, $i = 1, \dots, n$ are modeled as cubic splines using low-rank radial basis functions (French et al., 2001, Ruppert et al., 2003). In Section 2.2 we review briefly nonparametric function estimation.

2.2 Nonparametric function estimation

The functions $\mu_{kl}(t)$ and $g_{ikm}(t)$ are estimated nonparametrically. In this section we explain the basis function approach which is used in turn in Section 2.3 to estimate these functions. For simplicity, we focus on scatterplot smoothing, with observations (x_i, y_i) , $i = 1, \dots, n$. The description in this section is based on French et al. (2001), Ruppert et al. (2003) and Crainiceanu et al. (2005). Consider the model

$$y_i = f(x_i) + \epsilon_i,$$

where $E(\epsilon_i) = 0$, $i = 1, \dots, n$, and f is an unknown smooth function. A linear spline basis function can be expressed as $(x - \kappa)_+ = \max(0, x - \kappa)$, where κ is a knot. Any linear combination of linear spline basis functions $1, x, (x - \kappa_1)_+, \dots, (x - \kappa_K)_+$ is a piecewise linear function with knots at $\kappa_1, \dots, \kappa_K$. The function f may thus be expressed as

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^K u_k (x - \kappa_k)_+, \quad (4)$$

where the u_k s are the coefficients of the basis functions. We comment later in this section on the value of K .

To understand how the spline model (4) can be used for fitting a nonparametric curve to data, consider Figure 1, which displays in the top panel whip-shaped data similar to the example in Ruppert et al. (2003). The left half of the data exhibits linear behavior while curvature is apparent on the right-hand side. The bottom panel presents the basis functions used in the spline model. In particular, the equally-spaced knots are $0.50, 0.55, 0.60, \dots, 0.95$. Comparing the two panels of Figure 1, it is quite easy to see that a linear combination of the basis functions in the bottom panel should be able to capture the data structure in the top panel.

In general, any structure can be accommodated by placing basis functions at additional

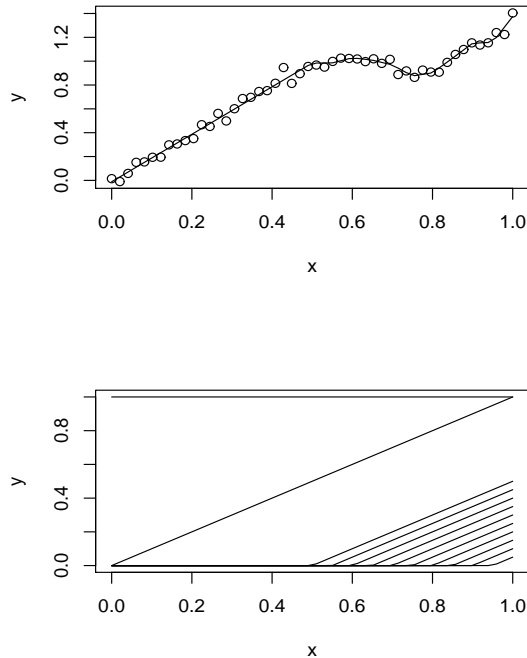


Figure 1: Top: data with fitted curve. Bottom: the basis functions

knots. To automate the process, two main approaches are commonly taken. One approach is automatic knot selection which can be carried out via Bayesian variable selection. Specifically, a large number of knots are placed at either equally spaced locations or at specific percentiles of the covariate, and an indicator variable is attached to each knot (see for example, Thompson and Rosen, 2008). The indicator value is 1 if a knot is to be retained at a given location or 0 if the knot should be removed from that location. In a Bayesian MCMC procedure, the indicator variables are sampled from at each iteration. The other approach is to retain all the knots but to constrain their influence. This can be accomplished by penalized spline regression or equivalently by using a linear mixed effects model formulation. In this paper, we use the latter approach in a Bayesian framework. In both approaches, the value of K is not crucial, as long as it is not too small. Typically, 30–40 knots are sufficient for medium-sized

datasets. Instead of the linear spline representation (4), we use in this paper the low rank thin-plate spline representation

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^K u_k |x - \kappa_k|^3. \quad (5)$$

Using cubic radial basis functions tends to result in a more aesthetically appealing fit,

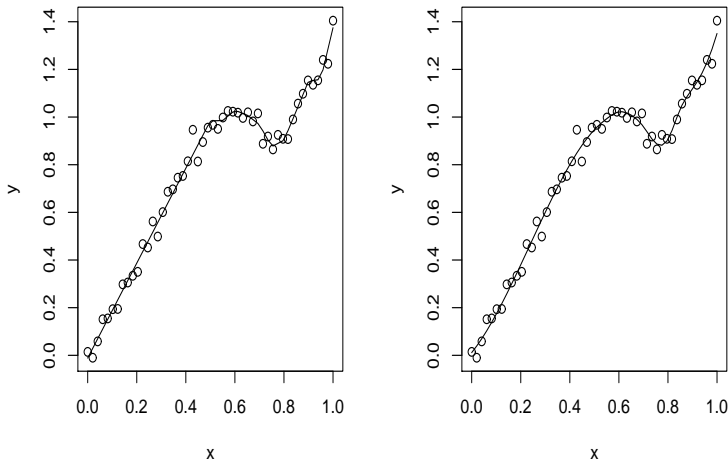


Figure 2: The data with a fitted curve. Left: truncated lines basis. Right: Cubic radial basis functions.

compared to that of the truncated-line basis, and may lead to faster convergence of the MCMC algorithm. The penalized spline approach prevents overfitting by adding a roughness penalty. Specifically, the minimization criterion is

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \frac{1}{\lambda} \boldsymbol{\theta}' D \boldsymbol{\theta}, \quad (6)$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, u_1, \dots, u_K)'$, λ is the smoothing parameter and D is a known positive-definite penalty matrix. For thin-plate splines, the matrix D is given by

$$D = \begin{pmatrix} 0_{2 \times 2} & 0_{2 \times K} \\ 0_{K \times 2} & \Omega_K \end{pmatrix},$$

where the (k, l) th element of Ω_K is $|\kappa_k - \kappa_l|^3$. From the structure of the matrix D it is clear that only the u_k s are penalized. Let $\mathbf{y} = (y_1, \dots, y_n)'$, $X = [1 \ x_i]_{1 \leq i \leq n}$,

$Z_K = [|x_i - \kappa_1|^3 \dots |x_i - \kappa_K|^3]_{1 \leq i \leq n}$. Dividing (6) by σ_ϵ^2 and expressing the penalty term explicitly as a function of Ω_K results in

$$\frac{1}{\sigma_\epsilon^2} \|\mathbf{y} - X\boldsymbol{\beta} - Z_K \mathbf{u}\|^2 + \frac{1}{\lambda \sigma_\epsilon^2} \mathbf{u}' \Omega_K \mathbf{u}, \quad (7)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\mathbf{u} = (u_1, \dots, u_K)'$ are considered fixed and random parameters, respectively. The solution to (7) is equal to the best linear unbiased predictor (BLUP) in the linear mixed model

$$\mathbf{y} = X\boldsymbol{\beta} + Z_K \mathbf{u} + \boldsymbol{\epsilon}, \quad \text{Cov}(\mathbf{u}) = \sigma_u^2 (\Omega_K^{-1/2}) (\Omega_K^{-1/2})', \quad (8)$$

where $\Omega_K^{-1/2}$ is based on the singular value decomposition. Note that Ω_K is not a positive definite matrix so it is not a proper covariance matrix; however, French et al. (2001) show that the smooth fit is not affected by this fact. Let $\mathbf{b} = \Omega_K^{1/2} \mathbf{u}$ and $Z = Z_K \Omega_K^{-1/2}$, then the mixed model (8) is equivalent to

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}, \quad \text{Cov} \begin{pmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{pmatrix} = \begin{pmatrix} \sigma_b^2 I_K & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 I_n \end{pmatrix}. \quad (9)$$

In a Bayesian framework, prior distributions need to be placed on all the model parameters.

2.3 Estimating $\mu_{kl}(t)$ and $g_{ikm}(t)$

To estimate the functions $\mu_{kl}(t)$ and $g_{ikm}(t)$, $k = 1, \dots, p$, $l = 1, \dots, r$, $m = 1, \dots, s$, $i = 1, \dots, n$, we use the basis function approach described in Section 2.2. In particular, let $\kappa_1, \dots, \kappa_K$ be K knots obtained as sample quantiles of t_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m_i$, and let $\Lambda_K = [|\kappa_k - \kappa_{k'}|^3]_{1 \leq k, k' \leq K}$ be a $K \times K$ matrix. Let $\boldsymbol{\phi}'_{ij} = (1 \quad t_{ij})$, $\boldsymbol{\xi}'_{ij} = (|t_{ij} - \kappa_1|^3, \dots, |t_{ij} - \kappa_K|^3)$, and $\boldsymbol{\psi}'_{ij} = \boldsymbol{\xi}'_{ij} \Lambda_K^{-1/2}$, where $\Lambda_K^{1/2}$ is obtained via the singular value decomposition. The vectors $\boldsymbol{\phi}_{ij}$ and $\boldsymbol{\psi}_{ij}$ are basis functions evaluated at t_{ij} and

are used to model the linear part and the nonlinear part, respectively, of the fixed and random functions. In particular, $\mu_{kl}(t)$ and $g_{ikm}(t)$ can be evaluated at t_{ij} by

$$\mu_{kl}(t_{ij}) = \boldsymbol{\phi}'_{ij} \boldsymbol{\beta}_{kl} + \boldsymbol{\psi}'_{ij} \mathbf{v}_{kl} \quad \text{and} \quad g_{ikm}(t_{ij}) = \boldsymbol{\phi}'_{ij} \mathbf{w}_{ikm} + \boldsymbol{\psi}'_{ij} \mathbf{u}_{ikm} \quad (10)$$

for $k = 1, \dots, p$, $l = 1, \dots, r$, $m = 1, \dots, s$ and $i = 1, \dots, n$. In (10), $\boldsymbol{\beta}_{kl}$, \mathbf{v}_{kl} , \mathbf{w}_{ikm} and \mathbf{u}_{ikm} are unknown parameter vectors.

2.4 Priors on the basis function coefficients and the variance components

We place the following prior distributions on $\boldsymbol{\beta}_{kl}$, \mathbf{v}_{kl} , \mathbf{w}_{ikm} and \mathbf{u}_{ikm} , $k = 1, \dots, p$, $l = 1, \dots, r$, $m = 1, \dots, s$ and $i = 1, \dots, n$.

1. $\boldsymbol{\beta}_{kl} \stackrel{ind}{\sim} N(\mathbf{0}, \sigma_{\boldsymbol{\beta}_{kl}}^2 I_2)$, where I_2 is a 2×2 identity matrix, and $\sigma_{\boldsymbol{\beta}_{kl}}^2$ is a large known value.
2. $\mathbf{v}_{kl} \stackrel{ind}{\sim} N(\mathbf{0}, \sigma_{\mathbf{v}_{kl}}^2 I_K)$, where I_K is a $K \times K$ identity matrix, and K is the number of knots.
3. $\mathbf{w}_{ikm} \stackrel{ind}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{w_{km0}}^2, \sigma_{w_{km1}}^2))$.
4. $\mathbf{u}_{ikm} \stackrel{ind}{\sim} N(\mathbf{0}, \sigma_{\mathbf{u}_{ikm}}^2 I_K)$.

Similar prior distributions on the coefficients of the basis functions were used by Durbán et al. (2005). Note that the variances of the elements of \mathbf{w}_{ikm} are different while those of the elements of \mathbf{u}_{ikm} are all the same. This is merely for computational convenience to avoid an additional $(K - 1)$ parameters. The priors on the variance components $\sigma_{\mathbf{v}_{kl}}^2$, $k = 1, \dots, p$, $l = 1, \dots, r$, are independent inverse gamma distributions with densities

$$p(\sigma_{\mathbf{v}_{kl}}^2) \propto (\sigma_{\mathbf{v}_{kl}}^2)^{-(a_1+1)} \exp(-b_1/\sigma_{\mathbf{v}_{kl}}^2),$$

where a_1 and b_1 are known small values reflecting vague knowledge on σ_{vkl}^2 . The priors on $\sigma_{w_{km0}}^2$, $\sigma_{w_{km1}}^2$ and $\sigma_{u_{km}}^2$ are similar inverse gamma distributions. Recently, a number of authors (see for example, Gelman, 2006) have proposed alternative prior distributions for variance components which may exhibit superior behavior to that of inverse gamma distributions. However, Zhao et al. (2006) reported good performance of inverse gamma priors in the case of nonparametric regression, provided the hyperparameters are not too small. Specifically, hyperparameter values of 0.01 worked well, whereas values of 0.001 behaved erratically.

2.5 Priors on the Ornstein-Uhlenbeck parameters

The Ornstein-Uhlenbeck process parameters are the matrix A and the matrix $C = BB'$. Both matrices consist of parameters which are constrained to satisfy certain conditions. In particular, as mentioned in Section 2, the stationarity condition requires the real parts of the eigenvalues of A to be positive. Also, the matrix C is required to be symmetric and positive definite. Imposing the constraints directly on the elements of these matrices would be difficult. Instead, we first express each of these matrices in an appropriate decomposition and then place prior distributions on the parameters of the decomposition factors. This is a much easier task, as the factor parameters are either unconstrained or constrained to be nonnegative. To place a prior on A , we express it as $A = S\Psi S^{-1}$, where S is a matrix of linearly independent eigenvectors, and Ψ is a diagonal matrix of real positive eigenvalues. This parameterization, used also by Sy et al. (1997) for the bivariate Ornstein-Uhlenbeck process, satisfies the stationarity condition mentioned above for the Ornstein-Uhlenbeck process. Aït-Sahalia (2008) discusses identifiability related to A and expresses it as a lower triangular matrix with positive diagonal elements. Kessler and Rahbek (2004) discuss identifiability issues in

the case of equidistant observation times. The matrix S is parameterized as $S = (s_{ij})$, $i, j = 1, \dots, p$, with unit diagonal elements. Independent $N(0, \sigma_a^2)$ priors are placed on the off-diagonal elements of S , and on the logarithms of the diagonal elements of Ψ . The matrix C is symmetric and positive definite. To place priors on its elements which satisfy the symmetry and positive definiteness, we first express the matrix C as a modified Cholesky factorization, $C = LDL'$, where L is unit lower triangular, and D is diagonal. This approach was used for example in Smith and Kohn (2002) and in Rosen and Stoffer (2007). The emphasis of Rosen and Stoffer (2007) is on estimation in the frequency domain for multivariate time series observed at equally spaced time points. The priors on the off-diagonal elements of L are taken to be independent $N(0, \sigma_L^2)$ with a fixed large value of σ_L^2 . The priors placed on $\log(D_i)$, where D_i is the i th diagonal element of D , are independent $N(0, \sigma_D^2)$ with a fixed large value of σ_D^2 .

2.6 The sampling scheme

Let $\boldsymbol{\theta}_k = (\boldsymbol{\beta}'_{k1}, \mathbf{v}'_{k1}, \dots, \boldsymbol{\beta}'_{kr}, \mathbf{v}'_{kr})'$, for $k = 1, \dots, p$, and let $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_p)'$. Similarly, let $\boldsymbol{\eta}_{ik} = (\mathbf{w}'_{ik1}, \mathbf{u}'_{ik1}, \dots, \mathbf{w}'_{iks}, \mathbf{u}'_{iks})'$, for $k = 1, \dots, p$ and $\boldsymbol{\eta}_i = (\boldsymbol{\eta}'_{i1}, \dots, \boldsymbol{\eta}'_{ip})'$ for $i = 1, \dots, n$. The sampling scheme consists of the following stages. More details are given in the Appendix.

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{\eta}_i$, $i = 1, \dots, n$, and the variance components by fitting p mixed effects models, for $k = 1, \dots, p$. Initialize A and C by maximizing numerically the log conditional joint posterior distribution of A and C .
2. Generate $\boldsymbol{\theta}$ from its full conditional posterior distribution, which is multivariate normal.
3. For each i , $i = 1, \dots, n$, generate $\boldsymbol{\eta}_i$ from its full conditional posterior distribution,

which is multivariate normal.

4. For $k = 1, \dots, p$, $l = 1, \dots, r$, $m = 1, \dots, s$, generate the variance components $\sigma_{v_{kl}}^2$, $\sigma_{w_{km0}}^2$, $\sigma_{w_{km1}}^2$ and $\sigma_{u_{km}}^2$ from their full conditional posterior distributions, which are inverse gamma.
5. Generate A from its full conditional posterior distribution. Since this distribution is not standard, we use a Metropolis step with a multivariate normal proposal density centered at the current value of A . The variance-covariance matrix of this normal proposal is based on the inverse of the estimated negative hessian of the log conditional posterior distribution.
6. Generate C from its full conditional posterior distribution using a Metropolis step.

3 Simulations

In this section, we explore by simulation the potential improvement in curve fitting when modeling the correlation structure of multivariate functional data rather than ignoring it. Specifically, we examine improvements in mean squared error for the individual subject-level functions. For this purpose we generated 100 datasets, with each dataset consisting of observations, without covariates, on $n = 50$ subjects. The number of observations per subject is $m_i = 2 + w_i$, where w_i is a Poisson random variable with expectation 5, giving an average of 7 observation times per subject. The observation times themselves were independently generated from a uniform distribution on the interval $[0, m_i]$. For each subject, there are $p = 3$ response variables with overall subject mean functions chosen to represent a variety of possible relationships. The first true mean function is $\mu_1(t) = 7 \sin(-.5t)$, which exhibits low frequency variation

on the range of t . Note that the second subscript on $\mu_1(t)$ was dropped, since there are no covariates in our simulation setting. The second true mean function is $\mu_2(t) = 10\phi(t; 1.5, .3) + 6\phi(t; 4, .6)$, where $\phi(t; a, b)$ is a univariate normal density with mean a and standard deviation b . The third true mean function is $\mu_3(t) = 2 \sin(-t)$, which has higher frequency oscillations on the range of t . Let $f_{ik}(t) = \mu_k(t) + g_{ik}(t)$, $k = 1, 2, 3$, be the individual subject functions, where we have again dropped the covariate subscript.

In particular,

$$\begin{aligned} f_{i1}(t) &= a_{i1} \sin(-.5t) + a_{i2} \\ f_{i2}(t) &= b_{i1}\phi(t; 1.5, .3) + b_{i2}\phi(t; 4, .6) \\ f_{i3}(t) &= c_{i1} \sin(-t) + c_{i2} , \end{aligned} \tag{11}$$

where $a_{i1} \sim N(7, .5)$, $a_{i2} \sim N(0, .2)$, $b_{i1} \sim N(10, .25)$, $b_{i2} \sim N(6, .25)$, $c_{i1} \sim N(2, .5)$ and $c_{i2} \sim N(0, .2)$. Here, $N(a, b)$ indicates the univariate normal distribution with mean a and standard deviation b . The observations $\mathbf{y}_i(t_{ij})$ were obtained by drawing the random coefficients a_{i1} , a_{i2} , b_{i1} , b_{i2} , c_{i1} , c_{i2} , evaluating the equations in (11) at time t_{ij} and adding $\boldsymbol{\delta}_i(t_{ij})$, which was in turn generated according to a multivariate Ornstein-Uhlenbeck error process with parameter values

$$A = \begin{pmatrix} 2 & -0.6 & -0.6 \\ -0.6 & 2 & -0.6 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 15 \end{pmatrix} .$$

These settings result in fairly noisy data with cross-correlations ranging from .25 to .47 among the three variables when evaluated at $\bar{\Delta}t_i = \frac{1}{m_i} \sum_{j=1}^{m_i} (t_{ij} - t_{i,j-1}) \approx 1$. The cross-correlation matrix is obtained from the cross-covariance (2), evaluated at $\bar{\Delta}t_i$. Plots of these mean functions along with one randomly generated dataset can be seen in Figure 3.

Our model was fitted four times for each dataset, once for each univariate outcome separately (thereby ignoring across-variable correlation) and then to all three outcomes simultaneously. The sampling scheme was run for 10,000 iterations per dataset, with

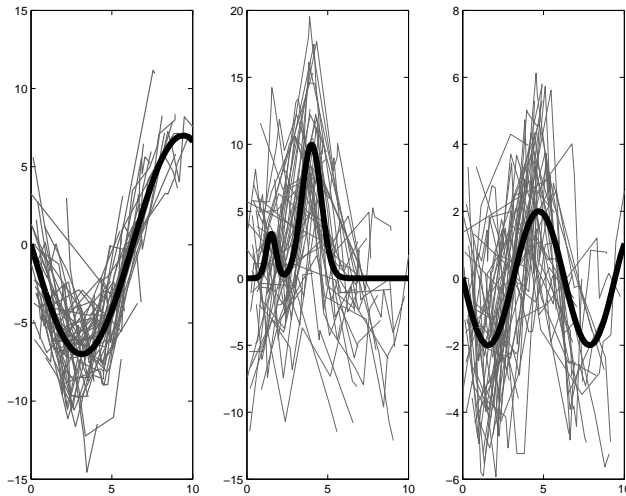


Figure 3: Mean functions (heavy lines) and data from one randomly generated dataset (light lines) with 50 individual subjects.

a burn-in period of 5,000. Median estimates \hat{A} and \hat{C} of the A and C matrices across all 100 multivariate fits were given by

$$\hat{A} = \begin{pmatrix} 2.91 & -0.72 & -1.31 \\ -0.38 & 2.29 & -0.63 \\ 0.02 & -0.07 & 3.03 \end{pmatrix} \quad \text{and} \quad \hat{C} = \begin{pmatrix} 18.82 & -0.81 & -1.67 \\ -0.81 & 16.81 & -0.19 \\ -1.67 & -0.19 & 19.59 \end{pmatrix}.$$

To assess the quality of the resulting estimates of the three mean functions, we calculated the average squared difference between the function estimates and the true mean functions at the unique observation times, $t_1 < \dots < t_M$. For the k th function this was computed by

$$MSE_k^{(1)} = \frac{1}{M} \sum_{m=1}^M (\hat{\mu}_k(t_m) - \mu_k(t_m))^2,$$

where $\hat{\mu}_k(\cdot)$ is the fitted mean function for the k th response variable. This was done for all three univariate fits, as well as for the joint trivariate fit. Boxplots of the resulting $MSE_k^{(1)}$, $k = 1, 2, 3$, are displayed in Figure 4. These boxplots show that the separate univariate fittings and the joint multivariate fitting resulted in little difference in the mean squared error for the first variable but lower mean squared error for the second

two. For the univariate fits, the median estimates for $MSE_k^{(1)}$ were .244, 1.19, and .206 for $k = 1, 2, 3$, respectively. For the multivariate fits, the corresponding median estimates of $MSE_k^{(1)}$ were .225, 1.03, and .188, for $k = 1, 2, 3$, respectively. Paired

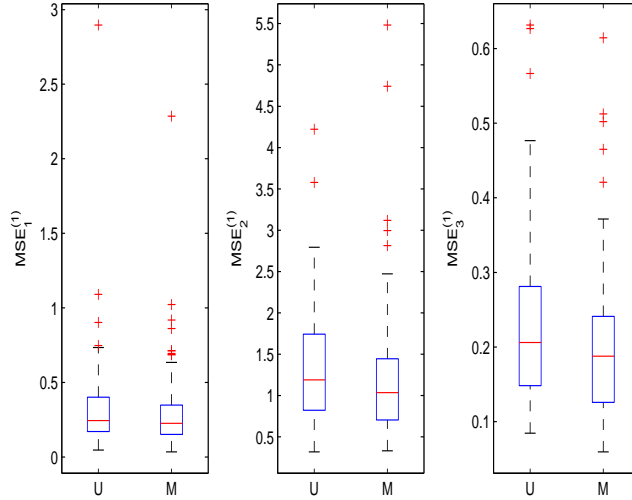


Figure 4: Boxplots of $MSE_k^{(1)}$ for the posterior means $\hat{\mu}_k(\cdot)$, $k = 1, 2, 3$, based on 100 simulated samples. Univariate and multivariate fits are denoted by U and M, respectively.

t-tests of the $\log MSE_k^{(1)}$ showed no significant difference of $\log MSE_1^{(1)}$ corresponding to the univariate and multivariate fits ($t = 1.0$, $p = 0.16$) but that $\log MSE_2^{(1)}$ and $\log MSE_3^{(1)}$ were significantly lower for the multivariate fits ($t = 2.3$, $p = 0.01$ and $t = 3.4$, $p < .0005$, respectively). To assess the quality of the fitted individual subject functional estimates, we calculated the average squared difference between the true individual subject functions and their estimates from the model at the measured observation times. For the k th variable, this mean squared error for the individual subject functions was computed by

$$MSE_k^{(2)} = \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^{m_i} (\hat{f}_{ik}(t_{ij}) - f_{ik}(t_{ij}))^2,$$

where $m = \sum_{i=1}^n m_i$ and $\hat{f}_{ik}(\cdot)$ is the fitted function for the i th subject's k th response.

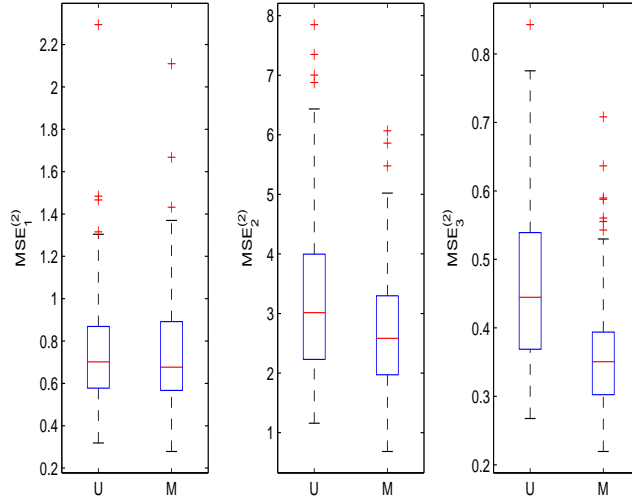


Figure 5: Boxplots of $MSE_k^{(2)}$ for the posterior means $\hat{f}_{ik}(\cdot)$, $k = 1, 2, 3$, $i = 1, \dots, 25$, based on 100 simulated samples. Univariate and multivariate fits are denoted by U and M, respectively.

Boxplots of the resulting mean squared errors for the individual subject functions are displayed in Figure 5. The mean squared errors for the subject functions show a similar pattern as for the overall mean. The median mean squared errors for each of the three outcome variables for the univariate fits were .702, 3.02, and .445, respectively. The median values for the corresponding multivariate fits were .676, 2.584, and .351. Thus, there was a 15% – 20% reduction in mean squared error for the individual functions for the last two variables when accounting for the multivariate covariance among them. Again, there was no significant difference of $\log MSE_1^{(2)}$ between the multivariate and univariate fits ($t = 0.33$, $p = 0.39$) but the multivariate fits had significantly lower $\log MSE_k^{(2)}$ for $k = 2, 3$ ($t = 4.6$, $p < 0.0005$, $t = 9.2$, $p < 0.0005$, respectively). One possible reason why the first variable exhibits no improvement in mean squared errors is that the low-frequency variation of the corresponding mean function renders it easier to fit and hence it is less important to borrow information across variables.

4 Application

As described in Section 1, we apply our methodology to the results of a randomized clinical trial conducted at the University of Pittsburgh and the University of Pisa, Italy (Frank et al., 2008). Despite decades of clinical trial experience in major depression, there is only limited understanding of which patients with major depressive disorder respond better to psychotherapy or to pharmacotherapy. This clinical trial compares the effects of psychotherapy (129 subjects) vs. pharmacotherapy (123 subjects). For clarity, Figure 6 shows the trajectories corresponding to 25 subjects only. We limit the current analysis to the first 12 weeks after baseline, at which about 95% of the subjects were still on study. Our methodology, which allows for nonlinear estimation of time courses, can accommodate the subject trajectories which are clearly nonlinear. In addition, our methodology accounts for the possibility of nonlinear effects of baseline covariates over time. Of particular interest is the identification of baseline subject characteristics which differentially predict treatment response in the two groups. The treatment response was change over time in two depression scales, the clinician-administered Hamilton Rating Scale for Depression (HRSD) and the Quick Inventory of Depression - Self Report (QIDS). These measures were collected more than once per week on average, though there was variation both within and between patients in the actual timing and number of measurements, with a mean of 11.2 measurement times per subject over the course of treatment. The HRSD scores ranged from 0–31 with a median of 10, and the QIDS scores ranged from 0–26 with a median of 6. In both measures, higher values indicate more depressive symptoms. Both measures were log transformed and standardized before running the analyses. A Lifetime Depression Spectrum (LDS) score was assessed on each patient at baseline; this gives an omnibus measure of depressive symptomatology over a patient’s lifetime (Cassano, et al., 1997).

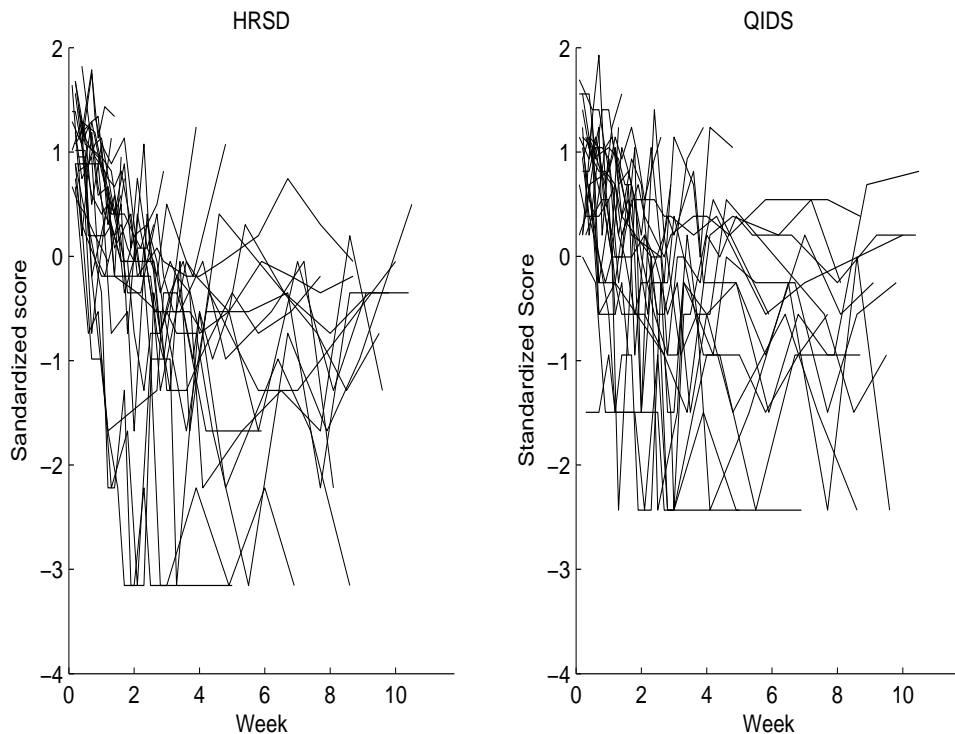


Figure 6: HRSD subject trajectories (left panel) and QIDS subject trajectories (right panel) for acutely depressed subjects. For clarity, only 25 subject trajectories are displayed. Scores are standardized to have zero mean and unit variance. Trajectories were truncated at 12 weeks.

In this example, we considered the LDS score to be a pre-treatment covariate with potentially differential effects on treatment outcomes for the two treatment groups. To explore this possibility, treatment group, LDS score, and their interaction were entered as time-varying fixed effects into our model with responses HRSD and QIDS entered as bivariate dependent variables. A time-varying random intercept was also included in the model. In the notation of Section 2, $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3}, x_{ij4})'$ and $z_{ij} = 1$, where $x_{ij1} = 1$, x_{ij2} is a group indicator (equal to 1 if subject i received psychotherapy and zero otherwise), x_{ij3} is the i th subject's LDS score, and $x_{ij4} = x_{ij2}x_{ij3}$. The sampling scheme described in Section 2.6 was run for 10,000 iterations with a burn-in period of

5,000 iterations. The estimated parameters of the Ornstein-Uhlenbeck process are

$$\hat{A} = \begin{pmatrix} 5.75 & -3.88 \\ -4.40 & 7.04 \end{pmatrix} \quad \text{and} \quad \hat{C} = \begin{pmatrix} 3.03 & 0.05 \\ 0.05 & 3.50 \end{pmatrix}.$$

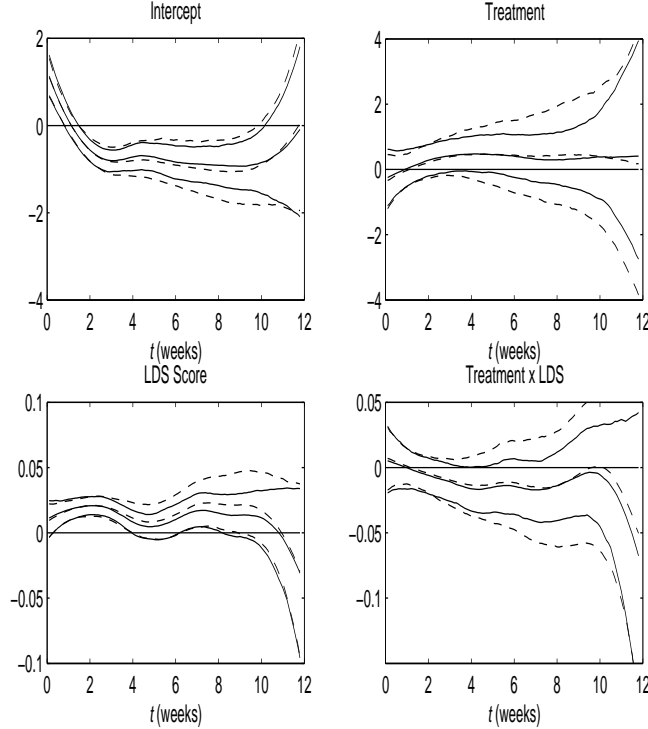


Figure 7: Time-varying functional coefficients for HRSD responses. The solid lines are $\hat{\mu}_{1l}(t)$, $1 \leq l \leq 4$, and their corresponding pointwise 95% credible intervals. The dashed lines are the analogous estimates and credible intervals corresponding to the univariate fits. Upper left panel: $\hat{\mu}_{11}(t)$. Upper right panel: $\hat{\mu}_{12}(t)$. Lower left panel: $\hat{\mu}_{13}(t)$. Lower right panel: $\hat{\mu}_{14}(t)$.

The estimated time-varying functional coefficients $\hat{\mu}_k(t) = (\hat{\mu}_{k1}(t), \hat{\mu}_{k2}(t), \hat{\mu}_{k3}(t), \hat{\mu}_{k4}(t))'$, $k = 1, 2$, for the HRSD and QIDS responses are plotted in figures 7 and 8, respectively. Solid lines correspond to the multivariate fits; for comparison, the univariate fits appear in dashed lines. As can be seen in these plots, the multivariate fits show little evidence for a treatment group effect on HRSD, but evidence for a slight difference in the QIDS

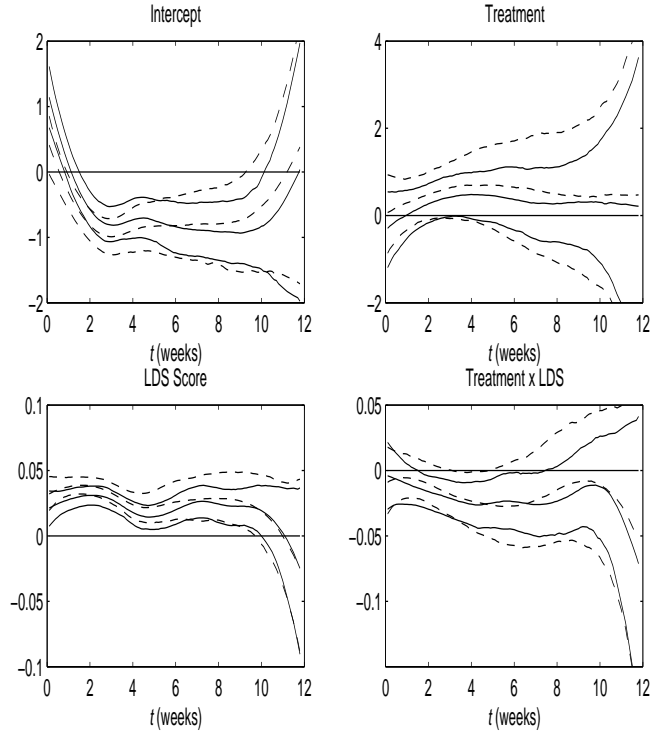


Figure 8: Time-varying functional coefficients for QIDS responses. The solid lines are $\hat{\mu}_{2l}(t)$, $1 \leq l \leq 4$, and their corresponding 95% credible intervals. The dashed lines are the analogous estimates and credible intervals corresponding to the univariate fits. Upper left panel: $\hat{\mu}_{21}(t)$. Upper right panel: $\hat{\mu}_{22}(t)$. Lower left panel: $\hat{\mu}_{23}(t)$. Lower right panel: $\hat{\mu}_{24}(t)$.

at around 3 weeks. However, there is a significant effect of LDS score on both outcomes, so that higher lifetime depression spectrum predicts worse outcomes over the first eight weeks or so. The interaction term is insignificant for HRSD and marginally significant in the 2–8 week time period for the QIDS responses. The effect of the interaction on responses is that LDS score is less predictive of poor QIDS response in the psychotherapy group than in the pharmacotherapy group. In general, the pointwise 95% credible intervals are wider for the univariate fits. While the functional coefficient estimates were substantially similar, the interaction coefficient for both models were not significant, i.e., the pointwise 95% credible intervals contained zero for the entire

time course.

5 Discussion

In this paper we have devised a regression model appropriate for multivariate functional responses with unequally-spaced observation times. Efficiency may be gained by fitting a single regression with all response variables simultaneously, as opposed to fitting regression models for each functional response separately. This is especially true if the error terms corresponding to each variable are correlated. In our formulation, the random error terms of the model were assumed to follow a multivariate Ornstein-Uhlenbeck process. Through this formulation we were able to extend the seemingly unrelated regression framework to the unequally-spaced multivariate functional data context.

The model we proposed uses a Bayesian mixed-effects approach, where the fixed part corresponds to the mean functions, and the random part corresponds to individual deviations from these mean functions. Covariates were allowed as either fixed or random effects. For each of the response variables, both the mean and the subject-specific deviations were estimated via low-rank cubic splines using radial basis functions. Thus mean and subject-specific deviation from the mean were allowed to vary smoothly as a function of time. Inference was performed via Markov chain Monte Carlo methods.

We demonstrated the improvement in efficiency that is possible by using this model in simulations which show mean squared-error is lower for the full multivariate algorithm compared to fitting each of the functional responses univariately, thereby ignoring the across-variable correlation. This seems especially important when the mean functions are wiggly, so that borrowing information across multiple responses becomes more

important.

Finally, the utility of this methodology was demonstrated by application to a real-life psychiatric dataset looking at the relationship of multiple depression measures over time in a clinical trial. Here, using a multivariate approach resulted in narrower posterior confidence bands.

We plan future research to extend the multivariate functional model to mixed discrete and continuous functional outcome data. We also plan to develop methods for the joint analysis of multivariate functional data and time-to-event data.

Acknowledgements

We thank the referees for their helpful comments, which greatly improved the paper. We also thank Dr. Ellen Frank, University of Pittsburgh, Department of Psychiatry, for use of the example data. The first author was supported in part by RCMI grant 5G12 RR008124 from the NIH and by NSF grants DMS-0706752 and DMS-0804140. The second author was supported by NIH grant K25 MH076981-01 and NSF grant DMS-0904825.

Appendix

Starting values for θ , η_i s and σ^2 :

Let $D_{\phi_{ij}}^r = I_r \otimes \phi'_{ij}$, where \otimes denotes the Kronecker product, and I_r is an $r \times r$ identity matrix. Define $D_{\psi_{ij}}^r$, $D_{\phi_{ij}}^s$ and $D_{\psi_{ij}}^s$ similarly.

and

$$\chi_i(t_{ij}, t_{i,j-1}) = \chi_{ij} - \exp(-A\Delta t_{ij})\chi_{i,j-1} .$$

Let $G = \text{blockdiag}(\sigma_{\beta_{11}}^{-2}I_2, \sigma_{v_{11}}^{-2}I_K, \dots, \sigma_{\beta_{1r}}^{-2}I_2, \sigma_{v_{1r}}^{-2}I_K, \dots, \sigma_{\beta_{p1}}^{-2}I_2, \sigma_{v_{p1}}^{-2}I_K, \dots, \sigma_{\beta_{pr}}^{-2}I_2, \sigma_{v_{pr}}^{-2}I_K)$.

Then,

$$[\boldsymbol{\theta}|\boldsymbol{\eta}, A, C, \boldsymbol{\sigma}^2, \mathbf{y}] \sim N(\boldsymbol{\mu}_\theta, \Sigma_\theta) ,$$

where $\boldsymbol{\sigma}^2 = (\sigma_{\beta_{kl}}^2, \sigma_{v_{kl}}^2, \sigma_{w_{km0}}^2, \sigma_{w_{km1}}^2, \sigma_{u_{km}}^2)'$ for $k = 1, \dots, p, l = 1, \dots, r, m = 1, \dots, s$,

$$\Sigma_\theta = [G + \sum_{i=1}^n \sum_{j=1}^{m_i} \chi'_i(t_{ij}, t_{i,j-1})\Omega_{\Delta t_{ij}}^{-1}\chi_i(t_{ij}, t_{i,j-1})]^{-1}$$

and

$$\boldsymbol{\mu}'_\theta = [\sum_{i=1}^n \sum_{j=1}^{m_i} \zeta'_i(t_{ij}, t_{i,j-1})\Omega_{\Delta t_{ij}}^{-1}\chi_i(t_{ij}, t_{i,j-1})]\Sigma_\theta .$$

In the expression for $\Omega_{\Delta t_{ij}}$, when $p = 2$, the stationary variance of the Ornstein-Uhlenbeck process is given by

$$\Sigma = \frac{\det(A)C + [A - \text{tr}(A)I]C[A - \text{tr}(A)I]'}{2\text{tr}(A)\det(A)} ,$$

see Gardiner (1983). For $p = 1$, this reduces to $\Sigma = C/(2A)$. For $p > 2$, Σ can be obtained numerically by Matlab's `lyap` function, for example.

Generating $\boldsymbol{\eta}_i$:

Let

$$\mathbf{q}_i(t_{ij}, t_{i,j-1}) = \mathbf{y}_i(t_{ij}) - \chi_{ij}\boldsymbol{\theta} - \exp(-A\Delta t_{ij})[\mathbf{y}_i(t_{i,j-1}) - \chi_{i,j-1}\boldsymbol{\theta}]$$

and

$$G_{wu} = \text{blockdiag}(\Sigma_{w_{11}}^{-1}, \sigma_{u_{11}}^{-2}I_K, \dots, \Sigma_{w_{1s}}^{-1}, \sigma_{u_{1s}}^{-2}I_K, \dots, \Sigma_{w_{p1}}^{-1}, \sigma_{u_{p1}}^{-2}I_K, \dots, \Sigma_{w_{ps}}^{-1}, \sigma_{u_{ps}}^{-2}I_K) ,$$

where $\Sigma_{w_{km}} = \text{diag}(\sigma_{w_{km0}}^2, \sigma_{w_{km1}}^2)$. Then

$$[\boldsymbol{\eta}_i|\boldsymbol{\theta}, A, C, \boldsymbol{\sigma}^2, \mathbf{y}] \sim N(\boldsymbol{\mu}_{\eta_i}, \Sigma_{\eta_i}) ,$$

where

$$\Sigma_{\eta_i} = \left[\sum_{j=1}^{m_i} E_i'(t_{ij}, t_{i,j-1}) \Omega_{\Delta t_{ij}}^{-1} E_i(t_{ij}, t_{i,j-1}) + G_{wu} \right]^{-1}$$

and

$$\boldsymbol{\mu}'_{\eta_i} = \left[\sum_{j=1}^{m_i} \mathbf{q}'_i(t_{ij}, t_{i,j-1}) \Omega_{\Delta t_{ij}}^{-1} E_i(t_{ij}, t_{i,j-1}) \right] \Sigma_{\eta_i} .$$

Generating σ^2 :

$$\sigma_{v_{kl}}^2 | v_{kl} \sim IG(K/2 + a_1, b_1 + \frac{1}{2} \mathbf{v}'_{kl} \mathbf{v}_{kl})$$

for $k = 1, \dots, p, l = 1, \dots, r$.

$$\begin{aligned} \sigma_{w_{km0}}^2 | \{\mathbf{w}_{ikm0}\}_{i=1, \dots, n} &\sim IG(n/2 + a_2, b_2 + \frac{1}{2} \sum_{i=1}^n w_{ikm0}^2) \\ \sigma_{w_{km1}}^2 | \{\mathbf{w}_{ikm1}\}_{i=1, \dots, n} &\sim IG(n/2 + a_3, b_3 + \frac{1}{2} \sum_{i=1}^n w_{ikm1}^2) \\ \sigma_{u_{km}}^2 | \{\mathbf{u}_{ikm}\}_{i=1, \dots, n} &\sim IG\left(\frac{nK}{2} + a_4, b_4 + \frac{1}{2} \sum_{i=1}^n \mathbf{u}'_{ikm} \mathbf{u}_{ikm}\right) \end{aligned}$$

for $k = 1, \dots, p, m = 1, \dots, s$.

Starting values for A and C :

Starting values for A and C are obtained by numerically maximizing the conditional posterior

$$p(A, C | \boldsymbol{\theta}_0, \{\boldsymbol{\eta}_{i0}\}, \mathbf{y}) \propto \prod_{i=1, \dots, n} \prod_{j=1}^{m_i} |\Omega_{\Delta t_{ij}}|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}'_{t_{ij}} \Omega_{\Delta t_{ij}}^{-1} \boldsymbol{\gamma}_{t_{ij}} \right\} \times p(A) \times p(C) ,$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_{i0}, i = 1, \dots, n$, are the starting values for the basis function coefficients.

Note that $\boldsymbol{\gamma}_{t_{ij}}$ depends on $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_{i0}, i = 1, \dots, n$, through $\boldsymbol{\delta}_i(t_{ij})$ (expression (A1)).

Generating the Ornstein-Uhlenbeck process parameters:

To generate A , note that

$$p(A | C, \boldsymbol{\theta}, \{\boldsymbol{\eta}_i\}, \mathbf{y}) \propto \prod_{i=1, \dots, n} \prod_{j=1}^{m_i} |\Omega_{\Delta t_{ij}}|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}'_{t_{ij}} \Omega_{\Delta t_{ij}}^{-1} \boldsymbol{\gamma}_{t_{ij}} \right\} \times p(A) . \quad (\text{A2})$$

Since (A2) is not a standard distribution, we use a Metropolis step to generate A . The proposal distribution is multivariate normal centered at the current value of A with a

variance-covariance matrix equal to the inverse of the negative Hessian of the log of (A2) evaluated numerically at the mode. This variance-covariance matrix is computed once, conditional on the starting values for the other parameters, and is then fixed throughout the sampling scheme. To increase the acceptance rate, this variance-covariance matrix is multiplied by $5.76/p$, as proposed in Gelman et al. (2004), page 306. More generally, when using a normal proposal distribution centered at the current point, Gelman et al. (2004) suggest using $c^2\Sigma$ as the covariance matrix of that proposal distribution. Among this class of proposal densities, the most efficient one has scale $c \approx 2.4/\sqrt{p}$. The acceptance probability is

$$\min \left\{ 1, \frac{p(A^p|C, \boldsymbol{\theta}, \{\boldsymbol{\eta}_i\}_{i=1, \dots, n}, \mathbf{y})}{p(A^c|C, \boldsymbol{\theta}, \{\boldsymbol{\eta}_i\}_{i=1, \dots, n}, \mathbf{y})} \right\} .$$

The matrix C is generated by first generating the matrices L and D via a Metropolis step based on

$$p(L, D|A, \boldsymbol{\theta}, \{\boldsymbol{\eta}_i\}, \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^{m_i} |\Omega_{\Delta t_{ij}}|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}'_{t_{ij}} \Omega_{\Delta t_{ij}}^{-1} \boldsymbol{\gamma}_{t_{ij}} \right\} \times p(L) \times p(D) .$$

An iterate for C is then given by $C = LDL'$.

References

- Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics*, **36**, 906–937.
- Baladandayuthapani, V., Mallick, B.K., Young Hong, M., Lupton, J.R., Turner, N.D. and Carroll, R.J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, **64**, 64–73.
- Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. Royal Statistical Society B*, **68**, 333–382.

- Blackwell, P.G. (2003). Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika*, **90**, 613–627.
- Durbán, M., Harezlak, J., Wand, M.P. and Carroll, R.J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–1167.
- Cassano, G.B., Michelini, S., Shear, M.K., Coli, E., Maser, J.D., Frank, E. (1997). The panic-agoraphobic spectrum: a descriptive approach to the assessment and treatment of subtle symptoms. *American Journal of Psychiatry*, **154**, 27–38.
- De la Cruz-Mesía, R. and Marshall, G. (2003). A Bayesian approach for nonlinear regression models with continuous errors. *Communications in Statistics, Theory and Methods*, **32**, 1631–1646.
- De la Cruz-Mesía, R. and Marshall, G. (2006). Nonlinear random effects models with continuous time autoregressive errors. *Statistics in Medicine*, **25**, 1471–1484.
- Frank, E., Cassano, G.B., Rucci, P., Fagiolini, A., Maggi, L., Kraemer, H.C., Kupfer, D.J., Pollock, B., Bies, R., Nimgaonkar, V., Pilkonis, P., Shear, M.K., Thompson, W.K., Grochocinski, V.J., Scocco, P., Battenfield, J., and Forgiione, R.N. (2008). Addressing the challenges of a cross-national investigation: lessons from the Pittsburgh-Pisa study of treatment-relevant phenotypes of unipolar depression. *Clinical Trials*, under revision.
- French, J.L., Kammann, E.E. and Wand, M.P. (2001). Comment on “Semiparametric Nonlinear Mixed-effects Models and Their Application” by C. Ke and Y. Wang. *J. American Statistical Association*, **96**, 1285–1288.
- Gardiner, C.W. (1983). *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences*, Springer-Verlag, Berlin.

- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1**, 515–534.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*, Second Edition, Chapman & Hall/CRC, Boca Raton.
- Golightly, A. and Wilkinson, D.J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, **16**, 323–338.
- Guo, W. (2002). Functional mixed effects models. *Biometrics*, **58**, 121–128.
- Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*, Chapman & Hall/CRC, Boca Raton.
- Kessler, M. and Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic gaussian diffusions. *Statistical Inference for stochastic Processes*, **7**, 137–151.
- Rosen, O. and Stoffer, D.S. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, **94**, 335–345.
- Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.
- Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression. *J. Econometrics*, **98**, 257–281.
- Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. American Statistical Association*, **97**, 1141–1153.
- Sy, J.P., Taylor, J.M.G. and Cumberland, W.G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics*, **53**, 542–555.

Thompson, W.K. and Rosen, O. (2008). A Bayesian model for sparse functional data.

Biometrics, **64**, 54–63. Web appendices are available at <http://www.tibs.org/biometrics>.

Zhao, Y., Staudenmayer, J., Coull, B.A. and Wand, M.P. (2006). General design

Bayesian generalized linear mixed models. *Statistical Science*, **21**, 35–51.