

A Penalized Spline Approach to Functional Mixed Effects Model Analysis

Huaihou Chen and Yuanjia Wang*

Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 W168th Street, New York, New York 10032, U.S.A.

**email*: yw2016@columbia.edu

SUMMARY. In this article, we propose penalized spline (P-spline)-based methods for functional mixed effects models with varying coefficients. We decompose longitudinal outcomes as a sum of several terms: a population mean function, covariates with time-varying coefficients, functional subject-specific random effects, and residual measurement error processes. Using P-splines, we propose nonparametric estimation of the population mean function, varying coefficient, random subject-specific curves, and the associated covariance function that represents between-subject variation and the variance function of the residual measurement errors which represents within-subject variation. Proposed methods offer flexible estimation of both the population- and subject-level curves. In addition, decomposing variability of the outcomes as a between- and within-subject source is useful in identifying the dominant variance component therefore optimally model a covariance function. We use a likelihood-based method to select multiple smoothing parameters. Furthermore, we study the asymptotics of the baseline P-spline estimator with longitudinal data. We conduct simulation studies to investigate performance of the proposed methods. The benefit of the between- and within-subject covariance decomposition is illustrated through an analysis of Berkeley growth data, where we identified clearly distinct patterns of the between- and within-subject covariance functions of children's heights. We also apply the proposed methods to estimate the effect of antihypertensive treatment from the Framingham Heart Study data.

KEY WORDS: Functional random effects; Multilevel functional data; Semiparametric longitudinal data analysis.

1. Introduction

Longitudinal designs are routinely implemented in biomedical research studies. Comprehensive presentations on parametric methods for analyzing longitudinal data can be found in, for example, Diggle, Liang, and Zeger (2002). In some applications, in addition to modeling a mean function, modeling a covariance function of the subject-specific processes is of scientific interest. For example, in genetic studies the covariance of related subjects within a family represents genetic information. This function is used to compute heritability (ratio of genetic variance and total trait variance), which quantifies the genetic effect on a trait (Khoury, Beaty, and Cohen, 1993). In other applications, although not of direct scientific interest, accurate estimation of a covariance function leads to efficiency gain in estimating the population mean function and fixed effects parameters (Fan, Huang, and Li, 2007).

In practice, concerns on model misspecification for parametric methods may call for more flexible nonparametric or semiparametric approaches. In the context of longitudinal data analysis, Diggle and Verbyla (1998) provided nonparametric estimation of covariance structure by using local polynomials to smooth various moment estimators of the variance and covariance functions. Wu and Pourahmadi (2003) and Huang et al. (2006) proposed nonparametric estimators for large covariance matrix via Cholesky decomposition that are guaranteed to be positive definite. In the context of functional data analysis, Guo (2002) considered functional mixed

effects models and introduced a Kalman filtering algorithm to handle large matrices in the mixed model representation of smoothing splines that may be computationally challenging. Crainiceanu et al. (2007) proposed Bayesian penalized spline (P-spline) to model variance function of heteroscedastic errors nonparametrically and provided a spatially adaptive smoothing parameter for the population mean function. Krafty et al. (2008) dealt with a varying coefficient model and pursued a smoothing spline-based approach with an iterative reweighted least square procedure to fit the model. Rice and Wu (2001) used regression spline-based methods and treated subject-specific curves to be nonparametric random curves. Fan et al. (2007) proposed a semiparametric method to estimate the error covariance function where the variance function is modeled nonparametrically with local polynomials and the correlation function is modeled parametrically. In the context of spatial smoothing, Wood, Jiang, and Tanner (2002) considered spatial process as a mixture of smoothing splines to achieve spatial adaptivity. Alternatively, functional principal components are used to reduce dimensionality and model a covariance function. Methods along this line include Ramsay and Silverman (2005, Chapter 8–10) for univariate data; Yao, Müller, and Wang (2005), Yao and Lee (2006), and Kauermann and Wegener (2009) for longitudinal data (or sparse functional data); and Di et al. (2009) and Staicu, Crainiceanu, and Carroll (2010) for multilevel functional data. To alleviate computational burden, Durbán

et al. (2005) pursued a simple P-spline (O'Sullivan, 1986; Eilers and Marx, 1996) approach to fit subject-specific curves, which expresses these curves as a linear combination of truncated polynomial spline basis with random coefficients and specifies a simplified parametric covariance matrix for the basis coefficients.

In this article, we present methods for functional mixed effects models that decompose longitudinal or functional outcomes as a sum of several terms: an unspecified population mean function, covariates with time-varying coefficients, functional subject-specific random effects, and residual measurement error process. Using P-splines, we propose nonparametric estimation of the population mean function, varying coefficient, subject-specific curves and the associated covariance function, which represents between-subject variation, and the variance function of the residual measurement errors, which represents within-subject variation. The proposed model and methods maintain flexibility in modeling both the population- and subject-level curves. In addition, decomposing variability of the outcomes as a between- and within-subject source is useful in identifying the dominant variance component therefore optimally modelling a covariance function. The benefit of such decomposition is illustrated through an analysis of the Berkeley growth data (Tuddenham and Snyder, 1954), where we identified clearly distinct patterns of the between- and within-subject covariance. Both estimated covariance functions satisfy the positive semidefinite constraint.

All nonparametric components of our model are estimated through P-spline, which is considered as a reduced rank smoother. P-spline was originally proposed by O'Sullivan (1986) and has gained popularity since Eilers and Marx (1996) and Ruppert, Wand, and Carroll (2003). A comprehensive review of P-spline can be found in Ruppert et al. (2003) and Ruppert, Wand, and Carroll (2009). With a P-spline basis expansion of the random subject-specific curves, the dimensionality of the covariance matrix of the random basis coefficients is reduced due to the moderate number of knots, leading to computational advantage. Theoretical work has shown that P-spline as a low rank approximation can be asymptotically as efficient as full rank estimators such as smoothing splines (Li and Ruppert, 2008; Claeskens, Kivobokova, and Opsomer, 2009).

This article distinguishes from the functional principal components based and longitudinal data analysis based approaches in the literature (e.g., Diggle and Verbyla, 1998; Yao et al., 2005; Yao and Lee, 2006; Di et al., 2009; Kauermann and Wegener, 2009) in that we do not require a moment estimator or a surface smoother of the covariance function/matrix before smoothing, and the dimensionality is reduced through using moderate number of knots instead of using reduced number of principal components. The proposed methods improve upon the regression spline-based approaches (Rice and Wu, 2001) that are sensitive to number and location of the knots through imposing a penalty on the spline coefficients to control overfitting and achieve smooth fit. This article also distinguishes from Durbán et al. (2005) by allowing a general unstructured covariance matrix for random spline basis of the subject-specific curves and allowing the residual error variance to be modeled nonparametrically. Compared to the

local polynomial- or kernel-based approaches (e.g., Fan et al., 2007) the proposed methodologies allow for easy incorporation of the fixed and random effects.

Current literature studies the asymptotic properties of P-spline estimator obtained from univariate data (a single measurement for each subject). Li and Ruppert (2008) examined the asymptotics of a P-spline estimator with B-spline basis and first- or second-order penalty assuming the number of knots is relatively large. Kauermann, Krivobokova, and Fahrmeir (2009) studied asymptotics of P-spline estimator allowing for generalized nonnormal outcomes. Claeskens et al. (2009) obtained two asymptotic scenarios of the P-spline estimator and showed the asymptotic bias and variance for each scenario with univariate data. In this article, we examine the asymptotic properties of the P-spline nonparametric population mean function estimated with longitudinal data. We show that under appropriate assumptions, the order of the bias and the variance term is the same as for the univariate data as shown in Claeskens et al. (2009).

The remaining part of the article is structured as follows. Section 2 proposes methods to estimate the population mean function and the within-subject heteroscedastic error variance function nonparametrically. Section 3 develops methods for a wider class of functional mixed effects models with varying coefficients, random unspecified subject-specific curves, and heteroscedastic measurement errors. In Section 4, we show the asymptotic bias, variance, and normality of the P-spline estimator with longitudinal data. In Section 5, we conduct two simulation studies to investigate performance of the proposed methods and apply them to analyze the Berkeley growth data and the Framingham Heart Study data. In Section 6, we discuss possible extensions.

2. Semiparametric Estimation of the Within-Subject Variation

In this section, we account for heteroscedastic within-subject errors while estimating the population mean function and the residual variance function nonparametrically. Let i index subjects and let j index visits. A useful model for longitudinal data analysis is a partially linear mixed effects model,

$$y_{ij} = \mu(t_{ij}) + x_{ij}^T \beta_0 + z_{ij}^T b_i + \epsilon_{ij}(t_{ij}),$$

$$\epsilon_i \sim N(0, V_i^{\frac{1}{2}} R_i(\theta) V_i^{\frac{1}{2}}), \quad V_i = \text{diag}\{\sigma^2(t_{i1}), \dots, \sigma^2(t_{i,m_i})\}, \quad (1)$$

where $\mu(t)$ is a nonparametric population mean function, x_{ij} is a $p_x \times 1$ vector of covariates and β_0 is the associated parameter vector, b_i are *i.i.d.* random effect vectors following $N(0, D)$, z_{ij} are the associated design vectors, and the vectors of heteroscedastic measurement errors $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{i,m_i})^T$ are assumed to be independent of the random effects, and their variance function, $\sigma^2(t)$, will be modeled nonparametrically, and $R_i(\theta)$ is a parametric correlation matrix such as AR-1 (first-order autoregressive) or compound symmetry with θ as the vector of unknown parameters. When $\mu(t)$ takes a linear form in model (1), it reduces to a linear mixed effects model with heteroscedastic errors. When $\mu(t)$ has a known nonlinear form such as exponential, the model (1) reduces to a nonlinear mixed effects model with heteroscedastic errors.

In practice it may not be easy to model the population mean and the error variance function parametrically. For example, the Berkeley growth data that we analyze in Section 5.2 clearly illustrates a nonlinear trend of the mean function and the variance function of children's heights, which are not straightforward to be specified parametrically. In the Web Appendix, we provide methods to estimate these functions nonparametrically by P-splines and present a likelihood-based smoothing parameter selection approach to choose multiple smoothing parameters.

3. Functional Mixed Effects Model and Nonparametric Estimation of the Between-Subject Variation

3.1 Model and Proposed Methods

In this section, we propose methods for a wider class of functional mixed effects models where we accommodate covariates with varying coefficients and in addition to heteroscedastic errors, we accommodate functional subject-specific random effects. To be specific, consider

$$\begin{aligned} y_{ij} &= x_{ij}^T \beta_0 + \mu(t_{ij}) + w_{ij} \beta(t_{ij}) + \nu_i(t_{ij}) + \epsilon_{ij}(t_{ij}), \\ \nu_i(t) &\sim W(0, \gamma), \quad \epsilon_i \sim N(0, V_i^{\frac{1}{2}} R_i(\theta) V_i^{\frac{1}{2}}), \\ V_i &= \text{diag}\{\sigma^2(t_{i1}), \dots, \sigma^2(t_{i, m_i})\}, \end{aligned} \quad (2)$$

where $\nu_i(t)$ are functional subject-specific random effects assumed to be independent, $W(0, \gamma)$ is a Gaussian process with covariance function $\gamma(s, t)$, and the residuals ϵ_{ij} are again assumed to have nonparametric variance $\sigma^2(t)$. When $\beta(t) = 0$ and $\nu_i(t)$ has a parametric form, model (2) reduces to model (1). The model (2) can handle nonparametric population mean function, varying coefficients and unspecified subject-specific curves with an unspecified covariance function; therefore, one obtains flexible estimation of both the population- and subject-level curves.

Assume that the population mean, time-varying coefficient, functional random effects, and heteroscedastic error variance functions can be approximated as

$$\begin{aligned} \mu(t) &= B_\mu(t) \beta_\mu, \quad \beta(t) = B_c(t) \beta_c, \\ \nu_i(t) &= B_\nu(t) \xi_i, \quad \text{and} \quad \log \sigma^2(t) = B_\sigma(t) \eta, \end{aligned}$$

where $B_\mu(t)$, $B_c(t)$, $B_\nu(t)$, and $B_\sigma(t)$ are row vectors of basis functions for the mean, varying coefficient, subject-specific curves, and error variance function with possible different order and different number of knots; β_μ , β_c , and η are the associated basis coefficients; and ξ_i are vectors of random subject-specific basis coefficients. Because the functional random effects $\nu_i(t)$ are approximated by a linear combination of spline basis with random coefficients, the between-subject covariance function can be approximated by

$$\gamma(s, t) = B_\nu(s) \Omega B_\nu^T(t), \quad \text{where } \Omega = \text{cov}(\xi_i).$$

Let $B_c^i = (w_{i1} B_c^T(t_{i1}), \dots, w_{im_i} B_c^T(t_{im_i}))^T$, $X_i = (x_i, B_\mu^i, B_c^i)$, $Z_i = (B_\nu^T(t_{i1}), \dots, B_\nu^T(t_{im_i}))^T$ and $\beta = (\beta_0^T, \beta_\mu^T, \beta_c^T)^T$. Then the model (2) can be re-written as

$$Y_i = X_i \beta + Z_i \xi_i + \epsilon_i, \quad \xi_i \sim N(0, \Omega), \quad \text{and} \quad \epsilon_i \sim N(0, V_i^{\frac{1}{2}} R_i V_i^{\frac{1}{2}}).$$

Direct maximization of the penalized marginal likelihood of the above model is a difficult nonconvex problem. However, we can treat ξ_i as missing data and employ the expectation-maximization algorithm. Define the penalized joint log likelihood of Y_i and ξ_i as

$$\begin{aligned} &\sum_{i=1}^n \left\{ (Y_i - X_i \beta - Z_i \xi_i)^T (V_i^{\frac{1}{2}} R_i V_i^{\frac{1}{2}})^{-1} \right. \\ &\quad \times (Y_i - X_i \beta - Z_i \xi_i) + \xi_i^T \Omega^{-1} \xi_i \left. \right\} + \lambda_\mu \beta_\mu^T P_\mu \beta_\mu \\ &\quad + \lambda_c \beta_c^T P_c \beta_c + \lambda_\eta \eta^T P_\eta \eta + \lambda_\nu \sum_{i=1}^m \xi_i^T P_\nu \xi_i, \end{aligned} \quad (3)$$

where λ_μ , λ_c , λ_ν , and λ_η are smoothing parameters and P_μ , P_c , P_ν , and P_η are penalty matrices depending on the chosen basis. For example, for the p th-order truncated polynomial basis with K knots, the penalty matrix is $\text{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$. The first three penalty terms in (3) control the smoothness of the fitted population mean, varying coefficient, and error variance functions. The last penalty term controls smoothness of the fitted subject-specific curves. It is motivated by the assumption that the random effects are realizations of a Gaussian process with smooth covariance function. Similar penalty was used in Krafty et al. (2008) for smoothing splines and in Wu and Zhang (2006).

Given the variance components Ω , V_i , and R_i , we minimize the joint penalized likelihood (3) with respect to β and ξ_i to obtain

$$\begin{aligned} \hat{\beta} &= \left(\sum_{i=1}^n X_i^T \hat{\Sigma}_i^{-1} X_i + P_{\lambda_\mu, \lambda_c} \right)^{-1} \left(\sum_{i=1}^n X_i^T \hat{\Sigma}_i^{-1} Y_i \right), \\ \hat{\xi}_i &= \hat{\Omega}_{\lambda_\nu}^* Z_i^T \hat{\Sigma}_i^{-1} (Y_i - X_i \hat{\beta}), \end{aligned} \quad (4)$$

where $\hat{\Sigma}_i = Z_i \Omega_{\lambda_\nu}^* Z_i^T + V_i^{\frac{1}{2}} R_i V_i^{\frac{1}{2}}$, $\hat{\Omega}_{\lambda_\nu}^* = (\hat{\Omega}^{-1} + \lambda_\nu P_\nu)^{-1}$, and $P_{\lambda_\mu, \lambda_c} = \text{diag}(0_{p_x}, \lambda_\mu P_\mu, \lambda_c P_c)$, where p_x is the column dimension of X_i . The estimation of the between-subject variance component Ω is through restricted maximum likelihood, which yields

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\xi}_i \hat{\xi}_i^T + \hat{\Omega}_{\lambda_\nu}^* - \hat{\Omega}_{\lambda_\nu}^* Z_i^T M_i Z_i \hat{\Omega}_{\lambda_\nu}^* \right\}, \quad (5)$$

with $M_i = \hat{\Sigma}_i^{-1} - \hat{\Sigma}_i^{-1} X_i (\sum_{i=1}^n X_i^T \hat{\Sigma}_i^{-1} X_i + P_{\lambda_\mu, \lambda_c})^{-1} X_i \hat{\Sigma}_i^{-1}$.

To summarize, we use the following algorithm to estimate parameters in (2). Assuming working independent residuals (WI) with constant variance, we can obtain initial value $\hat{\beta}_{(0)}$. Let $\Omega_{(0)} = \text{diag}\{1, \dots, 1\}$, $\lambda_\nu = 1$, $\Omega_{(0)}^* = (\Omega_{(0)}^{-1} + \lambda_\nu P_\nu)^{-1}$, and $\hat{\xi}_{i(0)} = \Omega_{(0)}^* Z_i^T \hat{\Sigma}_{i(0)}^{-1} (Y_i - X_i \hat{\beta}_{(0)})$. We repeat the following step 1 and step 2 until convergence is reached.

Step 1. Use methods introduced in Section 2 and the Web Appendix to estimate η and θ , which are associated with the within-subject covariance function.

Step 2. Calculate the expectation-maximization algorithm based estimators (4) and (5).

There are four smoothing parameters, λ_μ , λ_c , λ_ν , and λ_σ , involved in the estimation. A crossvalidation-based approach would be computationally intensive. It is also complicated

to carry out information-criteria-based model selection due to difficulties in defining degrees of freedom. We choose the smoothing parameters by a likelihood-based approach as described in the Web Appendix.

After the convergence is reached, the estimated nonparametric population-level curve is $\widehat{\mu}(t) = B_\mu(t)\widehat{\beta}_\mu$, and the predicted nonparametric subject-level curve for the individual i is

$$\widehat{s}_i(t) = x_i^T(t)\widehat{\beta}_0 + B_\mu(t)\widehat{\beta}_\mu + w_i(t)B_c(t)\widehat{\beta}_c + B_\nu(t)\widehat{\xi}_i.$$

Furthermore, the estimated between-subject covariance function is

$$\widehat{\gamma}(s, t) = B_\nu(s)\widehat{\Omega}B_\nu^T(t). \tag{6}$$

3.2 Testing the Varying Coefficients

In some applications, one may be interested in testing whether the varying coefficient changes with time, that is, the hypothesis

$$H_0 : \beta(t) = \beta^* \text{ for any } t \text{ vs. } H_1 : \beta(t) \neq \beta^* \text{ for some } t.$$

Due to the nonstandard distribution of the likelihood ratio test under the null hypothesis reported in Crainiceanu and Ruppert (2004a, 2004b), we compute p -value of the likelihood ratio test based on bootstrap resampling. Specifically, let

$$\widehat{\epsilon}_i = Y_i - x_i\widehat{\beta}_0 - B_\mu(t_i)\widehat{\beta}_\mu - w_i B_c(t_i)\widehat{\beta}_c - Z_i\widehat{\xi}_i$$

be the residuals obtained under H_1 , and let

$$Y_i^{(b)} = x_i\widehat{\beta}_0^{H_0} + B_\mu(t_i)\widehat{\beta}_\mu^{H_0} + w_i\widehat{\beta}_c + Z_i\widehat{\xi}_i^{H_0} + \widehat{\epsilon}_i, \quad i = 1, \dots, n,$$

denote the b th pseudoutcome under H_0 , where $\widehat{\beta}_0^{H_0}, \widehat{\beta}_\mu^{H_0}, \widehat{\beta}_c$, and $\widehat{\xi}_i^{H_0}$ are the corresponding estimators obtained under the null hypothesis. We resample the data $Y_i^{(b)}$ from the above model \mathcal{B} times, and compute the likelihood ratio test with each copy of the \mathcal{B} samples. We then compute the p -value of the test based on the empirical distribution of the bootstrapped likelihood ratio statistics. A similar procedure was used in Huang, Wu, and Zhou (2002).

4. Asymptotic Properties

In this section, we examine the asymptotic convergence rate of the bias and variance of the estimated population mean function and examine the asymptotic normality. These results are closely related to those obtained in Claeskens et al. (2009) and Zhu, Fung, and He (2008). Assume that the range of the variable t_{ij} is $[a, b]$, with $-\infty < a < b < \infty$. We will first consider the estimator with B-spline basis, and then extend the results to the truncated polynomial basis by a transformation of the two sets of basis functions (the latter results are presented in the Web Appendix).

4.1 Preliminary

Let $a = \tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1} = b$. In addition, define p knots $\tau_{-p} = \tau_{-p+1} = \dots = \tau_{-1} = \tau_0$ and another set of p knots $\tau_{K+1} = \tau_{K+2} = \dots = \tau_{K+p+1}$. Denote the B-spline basis functions as $N(t) = \{N_{-p,p+1}(t), \dots, N_{K,p+1}(t)\}$, let $N = \{N^T(t_{11}), \dots, N^T(t_{nm})\}^T$ and let $\Sigma = \text{diag}\{V, \dots, V\}$. We allow the covariance of Y_i to be unstructured, assume it is known and does not change across subjects. As described in

Section 2, the population mean function is obtained by minimizing

$$(Y - N\beta_\mu)^T \Sigma^{-1} (Y - N\beta_\mu) + \lambda \int_a^b [(N(t)\beta_\mu)^{(q)}]^2 dt, \tag{7}$$

where the penalty is the integrated squared q th-order derivative of the B-spline function and is assumed to be finite.

Let R denote a matrix with elements $R_{ij} = \int_a^b N_{j,p+1-q}(t)N_{i,p+1-q}(t)dt$, for $i, j = -p + q, \dots, K$ and let Δ_q denote a difference operator. The penalty term can be rewritten as $\lambda\beta_\mu^T \Delta_q^T R \Delta_q \beta_\mu$. Let $D_q = \Delta_q^T R \Delta_q$, the fitted population mean function can be expressed as a ridge regression estimator with weighted least squares

$$\widehat{\mu} = N(N^T \Sigma^{-1} N + \lambda D_q)^{-1} N^T \Sigma^{-1} Y,$$

with $\widehat{\mu} = \{\widehat{\mu}(t_{11}), \dots, \widehat{\mu}(t_{nm})\}^T$. A regression spline estimator is the solution to (7) ignoring the penalty term, that is,

$$\widehat{\mu}_{\text{reg}} = N(N^T \Sigma^{-1} N)^{-1} N^T \Sigma^{-1} Y.$$

Denote $C^{p+1}[a, b] = \{\mu : \mu \text{ has } p + 1 \text{ continuous derivatives. Under the assumptions A1, (A-1) in A2, and A3 stated in the Web Appendix, and } \mu \in C^{p+1}[a, b], \text{ Zhu et al. (2008) obtained the approximation bias and variance for } \widehat{\mu}_{\text{reg}} \text{ as}$

$$\begin{aligned} E\widehat{\mu}_{\text{reg}}(t) - \mu(t) &= b_a(t, p + 1) + o(\delta^{p+1}), \\ \text{Var}\{\widehat{\mu}_{\text{reg}}(t)\} &= \frac{1}{n} N(t)G^{-1}N^T(t) + o((n\delta)^{-1}), \end{aligned}$$

where $\widehat{\mu}_{\text{reg}}(t) = N(t)(N^T \Sigma^{-1} N)^{-1} N^T \Sigma^{-1} Y$, $G = (g_{ij})$, and $\Sigma^{-1} = (\sigma^{st})$ with

$$\begin{aligned} g_{ij} &= \sum_{s \neq t}^m \int_a^b \int_a^b N_i(x)\sigma^{st} N_j(y)\rho_{st}(x, y) dx dy \\ &+ \sum_{s=1}^m \int_a^b N_i(x)\sigma^{ss} N_j(x)\rho_s(x) dx, \end{aligned}$$

where ρ_s and ρ_{st} are defined in the Web Appendix. The approximation bias is

$$\begin{aligned} b_a(t, p + 1) &= -\frac{\mu^{(p+1)}(t)}{(p + 1)!} \\ &\times \sum_{i=0}^K I(\tau_i \leq t < \tau_{i+1}) \delta_i^{p+1} B_{p+1}\left(\frac{t - \tau_i}{\delta_i}\right), \end{aligned}$$

with $B_{p+1}(t)$ as the $(p + 1)$ th Bernoulli polynomial (Barrow and Smith, 1978). These results will be used to derive the asymptotic properties of the P-spline estimator. The asymptotic results are in the sense of keeping the number of measurements per subject fixed and letting the number of subjects go to infinity.

4.2 Asymptotic Properties for P-Spline Estimator with B-Spline Basis

Denote $K_q = \lambda K^{2q}/n$ and $\widehat{\mu}(t) = N(t)(N^T \Sigma^{-1} N + \lambda D_q)^{-1} N^T \Sigma^{-1} Y$.

THEOREM 1: (1) Under assumptions A1, (A-1) in A2, A3, $K_q = o(1)$, and $\mu(\cdot) \in C^{p+1}[a, b]$, the following statements hold

$$E(\widehat{\mu}(t)) - \mu(t) = b_a(t, p + 1) + b_\lambda(t, \Sigma) + o(\delta^{p+1}) + o(\lambda n^{-1} \delta^{-q}),$$

$$\begin{aligned} \text{Var}(\widehat{\mu}(t)) &= \frac{1}{n} N(t) \left(G + \frac{\lambda}{n} D_q \right)^{-1} \\ &\quad \times G \left(G + \frac{\lambda}{n} D_q \right)^{-1} N^T(t) + o(n^{-1} \delta^{-1}), \end{aligned}$$

and for $K \sim n^{1/(2p+3)}$ and $\lambda = O(n^{(p+2-q)/(2p+3)})$, the optimal rate for mean squared error (MSE), $n^{-(2p+2)/(2p+3)}$, is attained by the P-spline estimator.

(2) Under assumptions A1, (A-2) in A2, A3, $K_q = O(1)$ and $\mu(\cdot) \in W^q[a, b] = \{\mu : \mu \text{ has } q-1 \text{ absolutely continuous derivatives, } \int_a^b \{\mu^{(q)}(x)\}^2 dx < \infty\}$ the Sobolev space of order q , the following statements hold

$$E(\widehat{\mu}(x)) - \mu(x) = b_a(t, q) + b_\lambda(t, \Sigma) + o(\delta^q) + o((\lambda/n)^{1/2}),$$

$$\begin{aligned} \text{Var}(\widehat{\mu}(x)) &= \frac{1}{n} N(t) \left(G + \frac{\lambda}{n} D_q \right)^{-1} G \left(G + \frac{\lambda}{n} D_q \right)^{-1} N^T(t) \\ &\quad + o(n^{-1} (\lambda/n)^{-1/2q}), \end{aligned}$$

and for $\lambda \sim n^{1/(2q+1)}$ and $K \sim n^{1/(2q+1)}$, the optimal rate for MSE, $n^{-2q/(2q+1)}$, is attained by the P-spline estimator.

The proof of the theorem is given in the Web Appendix.

Remark 1: For the both scenarios in the Theorem 1, the shrinkage bias $b_\lambda(t, \Sigma) = -\frac{\lambda}{n} N(t) (G + \frac{\lambda}{n} D_q)^{-1} D_q \beta$ depends on Σ through G .

Remark 2: Theorem 1 holds for both fixed designs and random designs. The asymptotic approximation bias does not depend on the design distribution. The asymptotic shrinkage bias depends on the design distribution through G .

Remark 3: Under different conditions, Theorem 2 in Claeskens et al. (2009) obtained the same rate for the bias and the variance with $m = 1$ and $\Sigma = \sigma^2 I_n$, i.e., the univariate case.

Remark 4: The above theorem suggests that the asymptotic properties of the P-spline estimator are closer to the regression spline estimator when the number of knots is small ($K_q = o(1)$) while its asymptotic properties are closer to the smoothing spline estimators when the number of knots is large ($K_q = O(1)$). This observation is also noted in Claeskens et al. (2009) for independent data.

THEOREM 2: Assume $K^{2p+3} \sim n$, $\lambda = O(K^{p-q+2})$, and $h > 0, C > 0$, such that $\sup_{i,j} E|\epsilon_{ij}|^{2+h} \leq C$. Then

$$\frac{\widehat{\mu}(t) - \mu(t) - b_a(t, p + 1) - b_\lambda(t, \Sigma)}{\sqrt{\text{Var}(\widehat{\mu}(t))}} \longrightarrow N(0, 1)$$

in distribution, as $n \rightarrow \infty$.

Remark 5: Under the assumptions of this theorem, $K_q = \lambda K^{2q}/n = O(K^{p-q+2} K^{2q}/n) = O(n^{(p+q+2)/(2p+3)}/n) = O(n^{-\frac{p-q+1}{2p+3}}) = o(1)$. Hence, the asymptotic normality addresses the first scenario in Theorem 1.

In the Web Appendix, we present similar asymptotic properties for P-spline estimator with truncated polynomial basis.

5. Numerical Results

5.1 Simulation Studies

Simulation Study I. Our first simulation study examines performance of the semiparametric estimator of the within-subject covariance presented in Section 2. We compared the proposed P-spline estimator with three other alternatives: (1) Regression spline estimator (R-spline) for both mean and variance function; (2) P-spline estimator for the mean function when assuming WI with constant variance; and (3) P-spline estimator for the mean function when assuming a correctly specified parametric model for the covariance function of the residuals (Parametric). Two simulation scenarios were considered. In the first model, we generated data from

$$y_{ij} = \sin(2\pi t_{ij}) + b_i + \epsilon_{ij}(t_{ij}),$$

where the variance function of the residuals was $\text{Var}\{\epsilon_{ij}(t)\} = \exp(3t)$, and the correlation structure was AR-1 with autoregressive parameter $\rho = 0.6$. The number of subjects $n = 200$ and the number of repeated measurements per subject $m = 10$ with probability of missing equals to 0.1. Hence the number of repeated measurements can differ across subjects. The covariates t_{ij} were generated from a uniform distribution, $U(0, 1)$. The random effects b_i were generated independently from a standard normal distribution.

In the second simulation model, we used $\mu(t) = 7 - 16t + 30t^2 - 15t^3$ and $\sigma^2(t) = 10\sqrt{t}$ and all the other settings were the same as the first case.

We conducted 200 simulation runs. To evaluate performance of the estimated nonparametric functions, the MSEs were calculated over grid points $\{0.05, 0.06, \dots, 0.95\}$ for each simulated dataset. The MSEs were then averaged across the 200 simulated datasets to obtain the average MSE (AMSE). Table 1 summarizes the simulation results. The AMSE_μ and AMSE_σ are the corresponding AMSEs of $\mu(t)$ and $\sigma^2(t)$. The RMSE_μ and RMSE_σ are the ratios of AMSE of the proposed P-spline estimators $\widehat{\mu}(t)$ and $\widehat{\sigma}^2(t)$ over other estimators. The RMSE_μ of the proposed method over assuming WI was around 0.85 for both simulation models, which suggests efficiency gain of estimating mean function by properly accounting for the within-subject covariance by the proposed semiparametric estimator. The RMSE_μ of the proposed P-spline estimator over the regression spline was about 0.95. The corresponding AMSE_σ for the variance function of the proposed over the regression spline was about 0.90 for both simulation models, which shows the proposed method to be also more efficient (10% reduction in AMSE) in estimating the variance function. To compare with the parametric approach assuming the functional form of the variance function to be known, we note that the RMSE_μ of the proposed over the parametric approach was slightly over one indicating low efficiency loss in adopting the proposed semiparametric approach to estimate variance functions.

Table 1
Simulation results based on model 1,200 replications

Method	$RMSE_{\mu}^{\dagger}$	$AMSE_{\mu}$	$RMSE_{\sigma}^{\dagger}$	$AMSE_{\sigma}$	Method	$RMSE_{\mu}^*$	$AMSE_{\mu}$
Case I							
P-spline	1	0.0317	1	0.637	Parametric	1.001	0.0315
R-spline	0.948	0.0335	0.901	0.708	WI	0.829	0.0381
Case II							
P-spline	1	0.0281	1	0.611	Parametric	1.005	0.028
R-spline	0.946	0.0297	0.889	0.687	WI	0.849	0.033

*RMSE: The ratio of AMSE between the proposed method and other methods.

Simulation Study II. Our second simulation study examines methods proposed for the functional mixed effects model with varying coefficients and nonparametric random subject-specific curves in Section 3. We generated data from the model

$$Y_{ij} = \mu(t_{ij}) + \beta(t_{ij}) * trt_i + b_{i0} + b_{i1} * \nu(t_{ij}) + \epsilon_{ij}(t_{ij}),$$

where we considered two simulation scenarios. In the first scenario, we specified

$$\mu(t) = 2 \sin(2\pi t), \quad \beta(t) = \frac{1}{3} \log t,$$

$$\nu(t) = 1.5 \exp\{-10(t - 0.8)^2\}, \quad \sigma^2(t) = \exp(t).$$

The random coefficients b_{i0} and b_{i1} were sampled from $N(0, 4)$ and $N(0, 1)$, respectively. The measurement errors $\epsilon_{ij}(t_{ij})$ were generated independently from $N(0, \sigma^2(t_{ij}))$. The group indicators, trt_i , were generated from Bernoulli distribution with probability 0.6. The total number of subjects $n = 200$ while the repeated measurements within each subject $m = 10$ with probability 0.15 of being missing. The measurement time points were generated from $U(0, 1)$.

In the second scenario, we specified

$$\mu(t) = 2 \exp\{\sin(4t)\}, \quad \beta(t) = \sqrt{t},$$

$$\nu(t) = 0.7 \exp(t), \quad \sigma^2(t) = \exp\{-5(t - 0.1)^2\},$$

$n = 100$, and $m = 20$ with a missing probability of 0.15. All the other settings were the same as the first case.

The simulation results are summarized in Table 2. Again the $AMSE_{\mu}$, $AMSE_{\beta}$, $AMSE_{\sigma}$, and $AMSE_{\gamma}$ are the corresponding AMSEs of $\mu(t)$, $\beta(t)$, $\sigma^2(t)$, and $\gamma(t, t)$, respectively. The RMSEs are the ratios of the AMSE of the proposed method over other methods. Similar to the first simulation study, we compared the proposed estimators to regression spline (R-spline), P-spline assuming WI and P-spline assuming a correctly specified parametric model for the subject-specific random effects covariance and residual effects variance (Parametric). The efficiency gains of the proposed method for estimating the mean and varying coefficient function were about 15% compared to assuming WI in both simulation scenarios, which is nonignorable. For estimating the mean function, in the first scenario, the proposed method performed better than the regression spline in terms of AMSE, while in the second scenario their performance was similar. The $AMSE_{\sigma}$ for estimating the covariance function was 30% lower for the P-spline compared to regression spline in the first scenario and 17% lower in the second scenario. Analogous to the simulation study I, the differences in AMSE between the parametric approach assuming a correctly specified subject-specific ran-

Table 2
Simulation results based model 2,200 replications

	$RMSE_{\mu}^*$	$AMSE_{\mu}$	$RMSE_{\beta}^*$	$AMSE_{\beta}$
Case I				
P-spline	1	0.0749	1	0.120
R-spline	0.952	0.0787	0.932	0.129
WI	0.800	0.0935	0.841	0.143
Parametric	1.003	0.0746	0.988	0.122
Case II				
P-spline	1	0.144	1	0.270
R-spline	0.998	0.144	0.991	0.273
WI	0.834	0.173	0.888	0.304
Parametric	1.007	0.143	1.003	0.269
	$RMSE_{\sigma}^*$	$AMSE_{\sigma}$	$RMSE_{\gamma}^*$	$AMSE_{\gamma}$
Case I				
P-spline	1	0.111	1	0.606
R-spline	0.701	0.158	0.991	0.611
Case II				
P-spline	1	0.0048	1	0.624
R-spline	0.831	0.0058	0.995	0.627

*RMSE: The ratio of AMSE between the proposed method and other methods.

dom effects covariance and residual effects variance and the proposed method were small in both simulation cases.

The computing time to fit the model by the proposed algorithm depends on the number of subjects and number of observations per subject. For example, for the first scenario in simulation II, the average running time for 100 repetitions with 100 subjects and 10 observations per subject on a Dell desktop with 2.67 GHz CPU and 4GB RAM was 1.42 minutes. We present the computing time for other sample sizes in the Web Appendix.

5.2 Data Examples

Example I. We applied proposed methods to analyze the Berkeley Growth Study (Tuddenham and Snyder, 1954) data, a long-term investigation of children’s developmental characteristics conducted by the California Institute of Child Welfare. There were 93 subjects examined, including 39 boys and 54 girls. The heights of the children were measured at each of the scheduled times. There were four measurements by a child’s first birthday followed by annual measurements from 2 to 8 years, and then biannual measurements until the end of age 18.

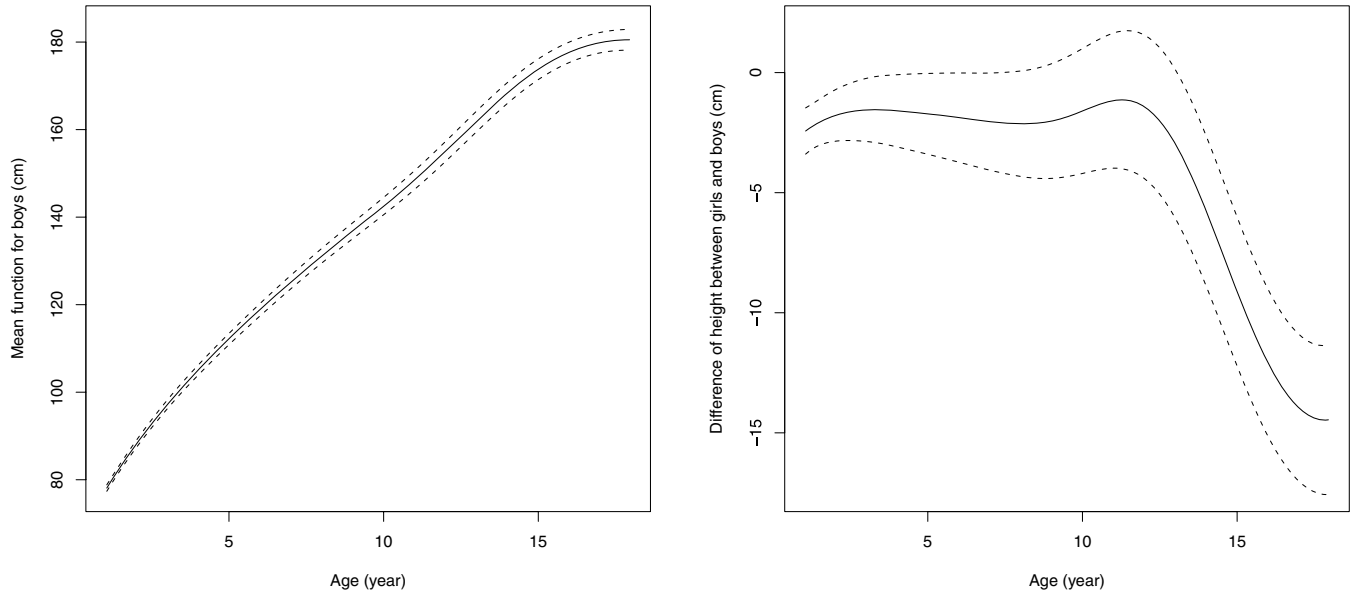


Figure 1. Estimated population mean function for boys $\mu(t)$ (left panel), varying coefficient $\beta(t)$ (right panel) and their 95% confidence bands.

Let y_{ij} be the height of subject i measured at occasion j , and let t_{ij} be the corresponding age. We fitted the model

$$y_{ij} = \mu(t_{ij}) + \text{sex}_i \times \beta(t_{ij}) + \nu_i(t_{ij}) + \epsilon_{ij}(t_{ij}),$$

$$i = 1, \dots, 93, \quad j = 1, \dots, 31,$$

where $\mu(t)$ was the mean height function for the boys, $\beta(t)$ was the height difference between girls and boys over time, and $\nu_i(t)$ were the random subject-specific deviations from their respective population mean function for boys and girls.

We used quadratic truncated polynomial splines for the mean, varying coefficient and variance functions and linear splines for the random subject-specific curves. The estimated varying coefficient and mean functions and the associated 95% confidence bands were plotted in Figure 1. The mean function for boys increased rapidly and then slowed down after age 16. The varying coefficient function $\beta(t)$ decreased quickly after age 12, while for the remaining time it was close to a constant. On average, the girls were shorter than the boys by about 2 cm under the age of 12. After age 12, the difference between boys and girls increased quickly. At the age of 18, the maximum difference of about 14 cm was reached, with boys being taller. Also note that during ages 10 and 12, there was a visible bump of the difference between the boys and girls corresponding to the first period of puberty of girls coming 2 years earlier than boys.

Using the bootstrap test introduced in Section 3.2, we tested whether the varying coefficient function was a constant. We simulated $B = 100$ bootstrap samples. The observed likelihood ratio test statistic was $T = \log L_{H_1} - \log L_{H_0} = 3840$. Based on the simulated empirical distribution of T under the null hypothesis, the p -value < 0.01 . Therefore, we observed significant evidence that the height difference between boys and girls varies across time. This can also be seen from the pointwise 95% confidence interval for $\beta(t)$.

The estimated covariance function $\gamma(s, t)$ of the subject-specific curves is plotted on the left panel of Figure 2. We can see that there was considerable variation of the subject-specific curves around their mean function, indicating substantial between-subject variation of the height growth patterns across children compared to the within-subject variation. The between-subject variation increased with age. The estimated standard deviation function $\sigma(t)$ of the residual measurement errors is shown in the right panel of Figure 2. It is evident that the variance function is not a constant. There was a decreasing trend of the variance function suggesting improvement of the precision of height measurements as a child grows. It is conceivable that height measurements for newborns are more variable than teenagers. The magnitude of $\sigma^2(t)$ is much smaller compared to $\gamma(t, t)$, suggesting that the dominant variance component of the variation in children's heights is the between-subject source.

Example II. In this example, we applied the proposed method to analyze the Framingham Heart Study longitudinal systolic blood pressure (SBP) data. The Framingham Heart Study is a large ongoing prospective study of risk factors for cardiovascular disease with the third generation data collected between 2002 and 2005 (Splansky et al., 2007). We analyzed subjects with ages ranging between 30 and 75. There were 190 independent subjects with 2406 observations. For each subject, their SBP, body mass index (BMI), and antihypertensive treatment status (trt) were measured over time. Sex was a baseline covariate coded as one for females and negative one for males. We centered the covariate BMI.

Let y_{ij} be the SBP of subject i measured at occasion j , and t_{ij} be the corresponding age. We fitted the model

$$y_{ij} = \alpha_1 \text{sex}_i + \alpha_2 \text{BMI}_{ij} + \mu(t_{ij})$$

$$+ \beta(t_{ij}) \times \text{trt}_{ij} + \nu_i(t_{ij}) + \epsilon_{ij}(t_{ij}),$$

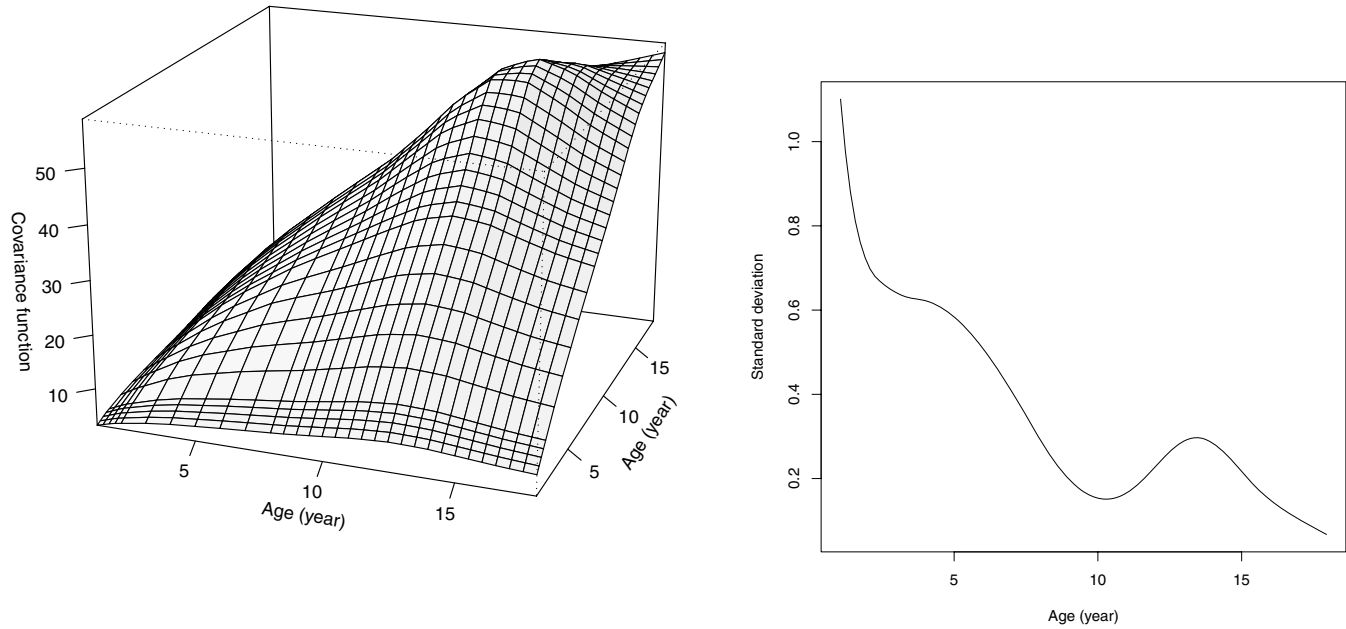


Figure 2. Estimated between-subject variation $\gamma(s, t)$ (left panel) and within-subject variation $\sigma(t)$ (right panel).

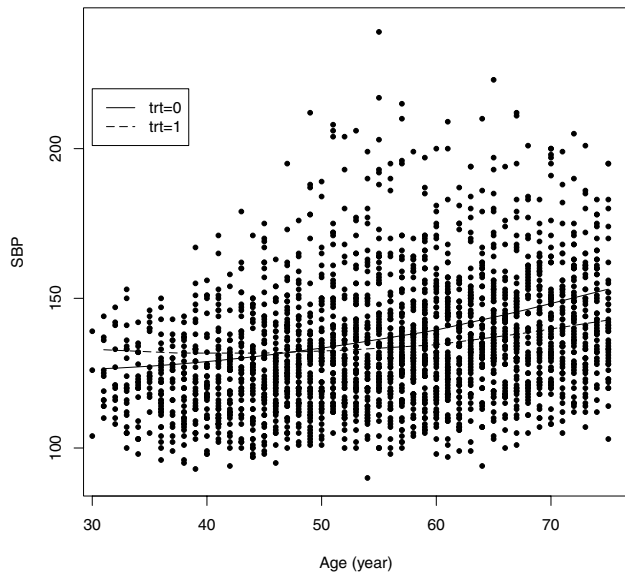


Figure 3. Observed SBP values (dots) and estimated mean SBP function for treated ($\mu(t) + \beta(t)$, $\text{trt} = 1$) and untreated subjects ($\mu(t)$, $\text{trt} = 0$).

where $\mu(t)$ was the population mean SBP function, $\beta(t)$ was the time varying effect of the antihypertensive treatment, and $\nu_i(t)$ were the random subject-specific deviations from the population mean function. We used linear truncated polynomial spline basis for the mean function, varying coefficient and random subject-specific curves; and quadratic spline for the variance function. We show the estimated mean SBP function for subjects with and without treatment in Figure 3. The population mean SBP increased with age, and taking antihyper-

tensive treatment reduced the SBP over time. For example, the mean SBP was 128.8 (95% CI: [126.5, 131.1]) at age 40 and then increased to 139.3 (95% CI: [136.5, 142.1]) at age 60.

We first tested whether the effect of antihypertensive is zero and the test was found to be significant. Using the bootstrap procedure in Section 3.2, we then tested whether there is any time-varying treatment effect (i.e., $H_0 : \beta(t) = \beta^*$). The observed log-likelihood ratio test statistic $T = 972$ with p -value = 0.02 based on $\mathcal{B} = 100$ bootstrap samples. Therefore, we observe significant evidence that the effect of antihypertensive treatment was nonzero and it varied with time. The square root of the estimated variance function, $\hat{\sigma}^2(t)$, and between-subject covariance function $\hat{\gamma}(s, t)$ are plotted in Figure 4. The variance function appeared to be nonlinear, with a change of the rate of increase at around age 50 and 68.

6. Discussion

In this article, we propose flexible estimation of population- and subject-level curves in a class of functional mixed effects models with varying coefficients. We also propose nonparametric estimation of the between-subject covariance and semi-parametric estimation of the within-subject covariance, which are useful descriptive tools to examine the outcome variability over time. When parsimony is desirable, these functions can be used to design reasonable parametric structures for the covariance of the outcomes. It is easy to see that the estimated covariance functions satisfy the positive semidefinite constraint. Furthermore, taking into account the covariance function improves efficiency in estimating the population mean function and the varying coefficients. The relative efficiency of estimating the covariance function with more subjects or more observations per subject depends on complexity of the functions $\sigma^2(t)$ and $\gamma(s, t)$.

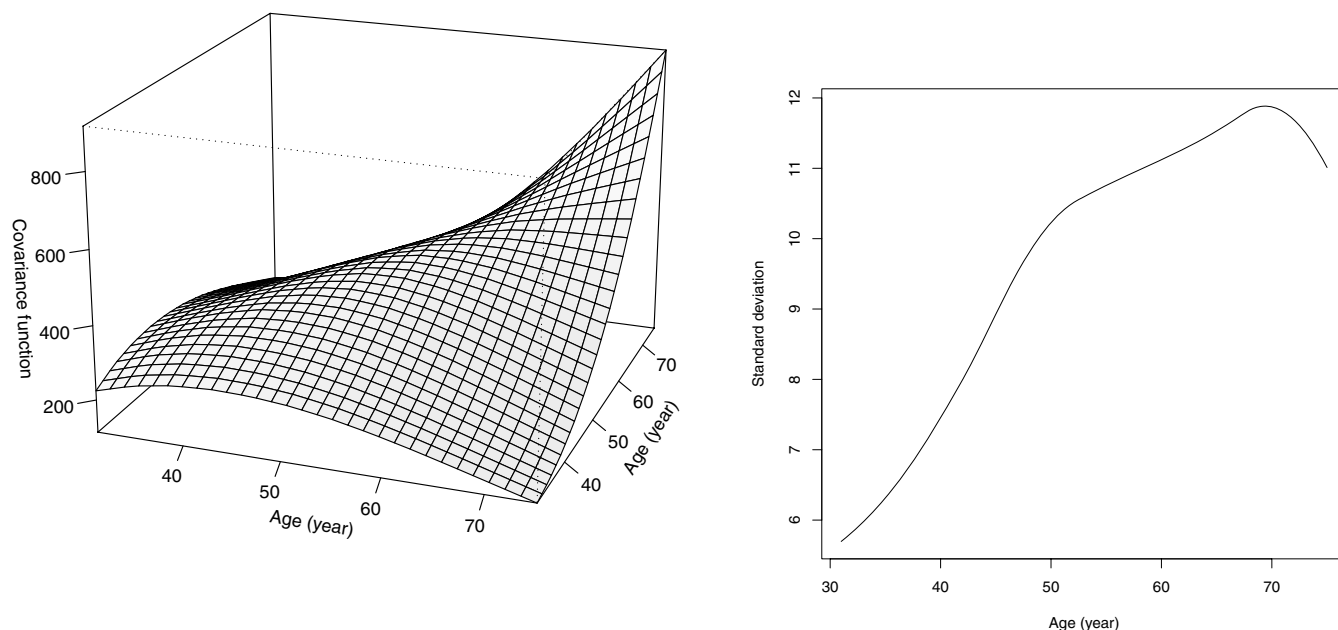


Figure 4. Estimated between-subject variation $\gamma(s, t)$ (left panel) and within-subject variation $\sigma(t)$ (right panel).

In the model (2), we assumed that the covariance function $\gamma(s, t)$ of the subject-specific curves is the same for all subjects. It is possible that $\gamma(s, t)$ differs across groups of subjects. For example, the covariance function for boys and girls may be different in the Berkeley Growth data. It is easy to accommodate such extension through the proposed P-spline methods by including an interaction between the basis functions and a covariate. In addition, adding parametric random effects to the model (2) is also straightforward.

The asymptotic theories for P-spline estimator are underdeveloped until very recently. In this article, we have extended the asymptotic bias and variance results in Claeskens et al. (2009) for univariate data to the longitudinal data case and we show the asymptotic normality for one of the asymptotic scenarios. The convergence rates obtained are consistent with those in Claeskens et al. (2009). Although the sample size required for the asymptotics to be an accurate approximation may be large, these results suggest that the P-spline estimator can be asymptotically as efficient as other smoothing techniques such as smoothing splines when the number of knots increases with sample size at a proper rate.

7. Supplementary Materials

Web Appendix A, referenced in Sections 2, 3, 4, and 5.1, is available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

ACKNOWLEDGEMENTS

The Framingham data were obtained from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine (Contract N01-HC-25195). YW's research is supported by NIH grant AG031113-01A2. The authors thank the associate editor and two anonymous re-

viewers for their constructive and helpful comments on this manuscript.

REFERENCES

- Barrow, D. L. and Smith, P. W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by splines with variable knots. *Quarterly of Applied Mathematics* **36**, 293–304.
- Claeskens, G., Kivobokova, T., and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–544.
- Crainiceanu, C. and Ruppert, D. (2004a). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statistica Sinica* **14**, 713–729.
- Crainiceanu, C. and Ruppert, D. (2004b). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* **66**, 165–185.
- Crainiceanu, C., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive Bayesian P-splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.
- Di, C., Crainiceanu, C., Caffo, B. S., and Punjabi, N. M. (2009). Multi-level functional principal component analysis. *Annals of Applied Statistics* **3**, 458–488.
- Diggle, P. and Verbyla, A. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**, 401–415.
- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. Oxford, U.K.: Oxford University Press.
- 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.
- Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines. *Statistical Science* **11**, 89–121.
- Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semi-parametric estimation of covariance function. *Journal of the American Statistical Association* **102**, 632–641.
- Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128.

- Huang, J., Wu, C., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.
- Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalized normal likelihood. *Biometrika* **93**, 85–98.
- Kauermann, G. and Wegener, M. (2009). Functional variance estimation using penalized splines with principal components analysis. *Statistics and Computing*, doi: 10.1007/s11222-009-9156-5.
- Kauermann, G., Krivobokova, T., and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society, Series B* **71**, 487–503.
- Khoury, M., Beaty, H., and Cohen, B. (1993). *Fundamentals of Genetic Epidemiology*. New York: Oxford University Press.
- Krafty, R., Gimotty, P., Holtz, D., Coukos, G., and Guo, W. (2008). Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. *Biometrics* **64**, 1023–1031.
- Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415–436.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (c/r: P519-527). *Statistical Science* **1**, 502–518.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd edition. New York: Springer.
- Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.
- Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge, U.K.: Cambridge University Press.
- Ruppert, D., Wand, M., and Carroll, R. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics* **3**, 1193–1256.
- Splansky, G. L., Corey, D., Yang, Q., Atwood, L. D., Cupples, L. A., Benjamin, E. J., D’Agostino, R. B., Fox, C. S., Larson, M. G., Murabito, J. M., O’Donnell, C. J., Vasan, R. S., Wolf, P. A., and Levy, D. (2007). The third generation cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: Design, recruitment, and initial examination. *American Journal of Epidemiology* **165**, 1328–1335.
- Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **8**, 1–29.
- Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publication in Child Development* **1**, 183–364.
- Wood, S., Jiang, W., and Tanner, M. (2002). Bayesian mixture of splines for spatially adaptive nonparametric regression. *Biometrika* **89**, 513–528.
- Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis Mixed-Effects Modeling Approaches*. Hoboken, New Jersey: Wiley.
- Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831–844.
- Yao, F. and Lee, T. C. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B* **68**, 3–25.
- Yao, F., Müller, H. G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.
- Zhu, Z., Fung, W. K., and He, X. (2008). On the asymptotics of marginal regression splines with longitudinal data. *Biometrika* **95**, 907–917.

Received March 2010. Revised September 2010.

Accepted September 2010.