Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

Functional random effect time-varying coefficient model for longitudinal data[†]

Jeng-Min Chiou[a,*], Yanyuan Ma[b] and Chih-Ling Tsai[c]

Received 27 August 2012; Accepted 2 October 2012

We propose a functional random effect time-varying coefficient model to establish the dynamic relationship between the response and predictor variables in longitudinal data. This model allows us not only to interpret time-varying covariate effects, but also to depict random effects via time-varying profiles that are characterized by functional principal components. We develop the functional profiling-backfitting method to estimate model components, which includes the profiling and backfitting procedures via a set of least squares type estimating equations. Asymptotic properties of the resulting estimator are obtained. Furthermore, we investigate the finite sample performance of the proposed method through simulation studies and present an application to primary biliary cirrhosis data. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: backfitting; functional data analysis; random effects model; varying coefficient; weighting

1 Introduction

In longitudinal or functional observations, functional data analysis (Ramsay & Dalzell, 1991) has been widely used to explore characteristics of the functional data and study the relationship between functional random variables. To model functional data with the i-th response random process, $Z_i(t)$, one often considers the following structure:

$$Z_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t), \tag{1}$$

where $Z_i(t)$ lies in a space of square integrable random functions, $t \in \mathcal{T}$ which is a real compact interval, $\mu(t)$ is a mean function, $\{\phi_k(t), k = 1, 2, \ldots\}$ is a set of orthonormal basis functions, and the random coefficients ξ_{ik} are uncorrelated, with mean zero and finite variance for each k. When the response process is observed subject to random noise, Yao et al. (2005a) employed the conditional regression approach together with the tool of functional data analysis to estimate unknown parameters. Detailed properties and applications of model (1) can be found in Ramsay & Dalzell (1991), Ramsay & Silverman (2005), Hall & Hosseini-Nasab (2006), Sentürk & Müller (2010), among others.

In practice, one may observe both response and predictor processes, $Y_i(t)$ and $X_i(t)$. As a result, it is natural to establish the relationship between these two processes. An attractive approach is to consider the varying coefficient model,

[a]Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan
[b]Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
[c]Graduate School of Management, University of California, Davis CA95616, USA & College of Management, National Taiwan University, Taipei 106, Taiwan
*Email: jmchiou@stat.sinica.edu.tw

[†]Supporting information may be found in the online version of this article.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

$$Y_i(t) = X_i^T(t)\beta(t) + \tilde{\varepsilon}_i(t) \quad (2)$$

where $X_i(t)$ is a $L \times 1$ covariate vector, $\beta(t)$ is a $L \times 1$ regression coefficient function, and $\tilde{\varepsilon}_i(t)$ is the random error with mean zero and constant variance. Some useful references for model (2) include Brumback & Rice (1998), Hoover et al. (1998), Wu et al. (1998), Fan & Zhang (2000), Lin & Ying (2001), Huang et al. (2002), and Wang et al. (2008). When covariates do not vary with t, model (2) becomes a classical varying coefficient model (see, e.g., Hastie & Tibshirani, 1993; Fan & Zhang, 1999; Wang & Xia, 2009).

In addition to the varying coefficient model, Faraway (1997) and Shen & Faraway (2004) considered a functional regression model by replacing $X_i^T(t)\beta(t)$ and $\tilde{\varepsilon}_i(t)$ in model (2) with $X_i^T\beta(t)$ and $\sum_{k=1}^{\infty}\xi_{ik}\phi_k(t)$, respectively. Their approach motivates us to propose the following model,

$$Y_i(t) = X_i^T(t)\beta(t) + \sum_{k=1}^{\infty}\xi_{ik}\phi_k(t) + \varepsilon_i(t), \quad (3)$$

where $\varepsilon_i(t)$ is the random error with mean zero and constant variance σ^2. Accordingly, the random response $Y_i(t)$ can be viewed as encompassing the time-varying coefficient effect $X_i^T(t)\beta(t)$ and the random effect $\sum_{k=1}^{\infty}\xi_{ik}\phi_k(t)$ of the ith response at t. Thus, we name (3) as the functional random effect time-varying coefficient model, and both the functional data model (1) and the varying coefficient model (2) are special cases of this model.

As far as we are aware, there has not been research on functional data with both random effects and time-varying coefficients taken into account simultaneously. The proposed functional random effect time-varying coefficient model first uses covariates to explain the variation in the observed response process, then employs functional principal component analysis (FPCA) to capture most of the unexplained variation, and finally leaves the remaining variation to disturbance errors. Since this hybrid model combines functional principal component analysis and varying coefficient approach, exploring its usefulness is important in functional data analysis.

In the rest of the article, we develop respectively a functional backfitting procedure and a profiling procedure to estimate the unknown parameters in Section 2. We show the asymptotic properties of the resulting estimates in Section 3. Simulation results and an empirical example are presented to illustrate the finite sample performance of the proposed estimation approach in Section 4. All the technical details are collected in the Appendix.

2 Estimation

2.1. Preliminary setting

Consider longitudinal or functional observations $\{(X_{ij}, T_{ij}, Y_{ij}), j = 1, \ldots, m_i, i = 1, \ldots, n\}$, where $Y_{ij} = Y_i(T_{ij})$ is the observed response at time T_{ij}, and $X_{ij} \in \mathcal{R}^L$ is the covariate vector that may or may not be time variant. When it is not time variant, X_{ij} equals a common vector X_i for all j; when it is time variant, $X_{ij} \equiv X_i(T_{ij})$ is the ith covariate vector observed at time T_{ij}. The longitudinal observations for each subject can either be sparsely or densely sampled. In the sparse situation, the m_i's are bounded from above by a constant m_0, i.e., $m_i < m_0, i = 1, \ldots, n$. In the dense situation, the m_i's are bounded from below by m_0, i.e., $m_i > m_0, i = 1, \ldots, n$; this is a typical observation pattern in functional data structures. In both situations, we assume the observation times, i.e., T_{ij}'s, are random and remain in a fixed compact interval $\mathcal{T}$.

When the basis functions are determined through the covariance function of $Z_i(t)$, model (1) corresponds to the Karhunen-Loève expansion of a random function, but with mean $X_i^T(t)\beta(t)$ (see, e.g., Ash & Gardner, 1975). Let $v(s,t) = \mathrm{cov}\{Z_i(s), Z_i(t)\}$ for $s, t \in \mathcal{T}$ and define the covariance operator A_v as a linear integral operator with kernel v by $(A_v f)(t) = \int_{\mathcal{T}} f(s)v(s,t)ds$, where f is a square integral function from $\mathcal{T}$ to the real line. Assume the operator A_v

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

has a sequence of orthonormal eigenfunctions $\{\phi_k\}$ with the corresponding eigenvalues $\{\lambda_k\}$, where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots$, satisfying $(A_v\phi_k)(t) = \lambda_k\phi_k(t)$ with $\int_{\mathcal{T}} \phi_k(t)\phi_l(t)dt = \delta_{kl}$, where δ_{kl} is 1 if $k = l$ and is 0 otherwise. Then, the covariance kernel can be represented as $v(s,t) = \sum_{k=1}^{\infty} \lambda_k\phi_k(s)\phi_k(t)$, and the coefficients $\xi_{ik} = \int_{\mathcal{T}} \left\{Z_i(t) - X_i^\top(t)\beta(t)\right\} \phi_k(t)dt$ are uncorrelated random variables with mean zero and variance λ_k.

To study parameter estimators, we take the same approach of Yao et al. (2005b) and Hall & Vial (2006), and assume that the local effect signal can be identified up to the first leading M components in (1), while those beyond are either degenerated or confounded with random errors. Then, we propose the functional profiling-backfitting (FPB) method given below to estimate the regression coefficient function $\beta(t)$ in (3). In the estimation method, we consider the weights $w_{ij}(t_0,h)$'s which are centered at the given time t_0 with the bandwidth h. Common choices for the weights include the Nadaraya-Watson weights $w_{ij}(t_0,h) = K\left\{(T_{ij}-t_0)/h\right\}/\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left\{(T_{ij}-t_0)/h\right\}$, the Gasser-Müller weights $w_{ij}(t_0,h) = \frac{1}{h}\int_{\tau_{ij}^-}^{\tau_{ij}^+} K\{(t_0-u)/h\}du$, where τ_{ij}^- is the middle point between T_{ij} and the previous observation time, and τ_{ij}^+ is the middle point between T_{ij} and the next observation time, and the local-linear weights

$$w_{ij}(t_0,h) = \frac{K\left\{(T_{ij}-t_0)/h\right\}\left\{S_{n,2}-(T_{ij}-t_0)S_{n,1}\right\}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left\{(T_{ij}-t_0)/h\right\}\left\{S_{n,2}-(T_{ij}-t_0)S_{n,1}\right\}},$$

where $S_{n,m} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left\{(T_{ij}-t_0)/h\right\}(T_{ij}-t_0)^m$ (see, e.g., Fan & Gijbels, 1996).

2.2. Functional profiling-backfitting estimator

Considering that the functional random effect time-varying coefficient model (3) contains both parametric and nonparametric components, we devise two different strategies accordingly. For estimating the smooth function $\beta(t)$, at any fixed ξ_{ik}'s, we construct a locally weighted least squares estimate of $\beta(t_0)$ by minimizing $\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}(t_0,h)\left\{Y_{ij} - X_{ij}^\top\beta(t_0) - \sum_{k=1}^{M}\xi_{ik}\phi_k(T_{ij})\right\}^2$ with respect to $\beta(t_0)$. On the other hand, at any fixed $\beta(t)$ function, we construct the least squares estimator of ξ_{ik}'s by minimizing $\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{Y_{ij} - X_{ij}^\top\beta(T_{ij}) - \sum_{k=1}^{M}\xi_{ik}\phi_k(T_{ij})\right\}^2$ with respect to ξ_{ik}. The above strategy yields the following estimating equations that need to be solved simultaneously.

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}(t_0,h)X_{ij}\left\{Y_{ij} - X_{ij}^\top\beta(t_0) - \sum_{k=1}^{M}\xi_{ik}\phi_k(T_{ij})\right\} = 0, \tag{4}$$

$$\sum_{j=1}^{m_i}\phi_k(T_{ij})\left\{Y_{ij} - X_{ij}^\top\beta(T_{ij}) - \sum_{l=1}^{M}\xi_{il}\phi_l(T_{ij})\right\} = 0, \tag{5}$$

where $w_{ij}(t_0,h)$'s are the weights centering at t_0 with the bandwidth h as given in Section 2.1., (4) holds for $t_0 = T_{ij}, j = 1,\ldots,m_i, i = 1,\ldots,n$, and (5) holds for $k = 1,\ldots,M, i = 1,\ldots,n$.

To solve these equations, we employ the profiling and backfitting procedures (Hastie & Tibshirani, 1990). In the profiling procedure, we solve (4) and obtain $\beta(t_0)$ as a function of ξ, denoted by $\hat\beta(t_0,\xi)$, where $\xi = \left(\xi_1^\top,\ldots,\xi_n^\top\right)^\top$ and $\xi_i = (\xi_{i1},\cdots,\xi_{iM})^\top$, then we plug $\hat\beta(t_0,\xi)$ back into (5) to solve for ξ. In the backfitting procedure, we solve for $\beta(t_0)$ from (4) while treating ξ as a known quantity, and then solve for ξ from (5) while treating $\beta(t_0)$ as a known vector; iterate until convergence.

To explicitly express the resulting estimators, let $\hat B_i$ be an $m_i \times Lm_i$ block diagonal matrix with the jth block diagonal being $\hat\beta(T_{ij})^\top$, $\boldsymbol{\phi}_i = (\phi_{1i},\ldots,\phi_{Mi})^\top$, and $\mathcal{A}_i$ be an $M\times M$ matrix with its (k,l) entry $\mathcal{A}_{ikl} = \phi_{ki}^\top\phi_{li}$. Furthermore, let $\tilde Y_i = (Y_{i1},\ldots,Y_{im_i})^\top$ and $\tilde X_i = (X_{i1},\ldots,X_{im_i})$ be an $m_i\times 1$ vector and an $L\times m_i$ matrix, respectively.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid Dissemination of Statistics Research

Following the profiling and backfitting procedures, we have that

$$\hat{\beta}(t_0) = \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h) \tilde{X}_i^{\top} \right\}^{-1} \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h) \left(\tilde{Y}_i - \boldsymbol{\phi}_i^{\top} \hat{\xi}_i \right) \right\} \quad \text{and} \tag{6}$$

$$\hat{\xi}_i = \mathcal{A}_i^{-1} \boldsymbol{\phi}_i (\tilde{Y}_i - \hat{B}_i \mathbf{X}_i) \ . \tag{7}$$

Hence, we term estimators $\hat{\beta}(t_0)$ and $\hat{\xi}_i$, respectively, as the *functional profiling-backfitting estimator* (FPBE) of $\beta(t_0)$ and ξ_i. In addition, when the smoothness of the coefficient functions for different covariates is different, we could use the specific bandwidth h_r to solve for the rth component of $\beta(t_0)$ to obtain

$$\hat{\beta}_r(t_0) = e_r^{\top} \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h_r) \tilde{X}_i^{\top} \right\}^{-1} \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h_r) \left(\tilde{Y}_i - \boldsymbol{\phi}_i^{\top} \hat{\xi}_i \right) \right\}. \tag{8}$$

After convergence, we keep the estimated $\hat{\xi}_i$ as the final estimate of ξ_i, and then take one additional step in (6) with a nonparametric optimal bandwidth $h = O(N^{-1/5})$, where $N = \sum_{i=1}^{n} m_i$, to obtain the final estimate $\hat{\beta}(t_0)$.

In practice, the eigenfunction $\boldsymbol{\phi}_i(t)$ is unknown. Hence, we first solve (6) for $\hat{\beta}(t_0)$ by setting $\xi_{ik} = 0$ as the initial estimates. The covariance function estimate $\hat{v}(s,t)$ is obtained via the two-dimensional smoothing of the raw covariances $\left\{ \left(Y_{ij} - X_{ij}^{\top} \hat{\beta}(T_{ij}) \right) \left(Y_{il} - X_{il}^{\top} \hat{\beta}(T_{il}) \right), 1 \leqslant j \neq l \leqslant m_i, 1 \leqslant i \leqslant n \right\}$. To obtain the estimated eigenfunction $\hat{\phi}_k(t)$, we simply find the solution of the eigenequation, $\int_{\mathcal{T}} \hat{v}(s,t) \hat{\phi}_k(s) ds = \hat{\lambda}_k \hat{\phi}_k(t)$, subject to $\int_{\mathcal{T}} \hat{\phi}_k(t)^2 dt = 1$, with $\hat{\lambda}_k$ being the estimated eigenvalue in a descending order for $k = 1, \ldots, M$. Subsequently, we solve (7) for $\hat{\xi}_i$ by plugging in the estimated $\hat{\beta}$ into $\hat{B}_i$. The profiling and backfitting procedure iterates until the convergence criterion is met. That is,

$$\sum_{r=1}^{L} \left\| \hat{\beta}_r^{(\tau)} - \hat{\beta}_r^{(\tau-1)} \right\| \Big/ \left\| \hat{\beta}_r^{(\tau)} \right\| < c, \tag{9}$$

where τ is the iteration indicator and c is a preset threshold value.

Remark 1

To obtain the functional profiling-backfitting estimator, we take into account the functional data feature via functional principal component analysis. In addition, the estimating equations (4) do not involve the inverse of the covariance function in the weighting scheme, which eases the estimation complexity. Moreover, if the random errors ε_i are normally distributed, then FPBE becomes the maximum likelihood estimator.

In practice, the bandwidths h_r are often unknown. Hence, we follow the leave-one-subject-out cross validation procedure (Rice & Silverman, 1991) to select the optimal bandwidth $\mathbf{h} \equiv (h_1, \ldots, h_L)^{\top}$ by minimizing

$$\mathrm{CV}(\mathbf{h}) = \sum_{i=1}^{n} \left(\tilde{Y}_i - \hat{B}_{i,-i} \mathbf{X}_i - \sum_{k=1}^{M} \hat{\xi}_{ik} \hat{\phi}_{ki} \right)^{\top} \left(\tilde{Y}_i - \hat{B}_{i,-i} \mathbf{X}_i - \sum_{k=1}^{M} \hat{\xi}_{ik} \hat{\phi}_{ki} \right),$$

where $\hat{B}_{i,-i}$ is the estimator of B_i with the ith observation being left out, B_i is an $m_i \times Lm_i$ block diagonal matrix with the jth block diagonal being $\beta(T_{ij})^{\top}$, $\mathbf{X}_i = \mathrm{vec}(\tilde{X}_i) = \left(X_{i1}^{\top}, \ldots, X_{im_i}^{\top} \right)^{\top}$ is an $Lm_i \times 1$ vector, and $\hat{\phi}_{ki}$ and $\hat{\xi}_{ik}$ are the estimators of $\phi_{ki} = \left\{ \phi_k(T_{i1}), \ldots, \phi_k(T_{im_i}) \right\}^{\top}$, and ξ_{ik}, respectively.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

Remark 2

When the observations are dense, (5) is a sample version of

$$\langle \phi_k(t), Y_i - X_i^\top \beta - \sum_{l=1}^{M} \xi_{il}\phi_l \rangle = \langle \phi_k, Y_i - X_i^\top \beta \rangle - \xi_{ik} = 0,$$

where the inner product of two integrable functions $\tilde{f}$ and $\tilde{g}$ is defined as $\langle \tilde{f}, \tilde{g} \rangle = \int_{\mathcal{T}} \tilde{f}(t)\tilde{g}(t)dt$. As a result, the functional principle component score is $\xi_{ik} = \langle \phi_k, Y_i - X_i^\top \beta \rangle$. Under the assumption that the functional principle component scores ξ_{ik} and the random errors $\varepsilon_i(t)$ are jointly Gaussian, we can then adopt the conditional expectation approach of Yao et al. (2005b) to obtain the best prediction of the functional principal component score for the ith subject. That is,

$$\check{\xi}_i = \Lambda \boldsymbol{\phi}_i \Sigma_i^{-1} (\tilde{Y}_i - \hat{B}_i \mathbf{X}_i) \ , \tag{10}$$

where $\Sigma_i = \boldsymbol{\phi}_i^\top \Lambda \boldsymbol{\phi}_i + \sigma^2 I$, and $\Lambda = diag(\lambda_k)$ for $k = 1, \dots, M$. If $\sigma^2 = 0$, we are able to show that $\check{\xi}_i$ in (10) and $\hat{\xi}_i$ in (7) are identical.

2.3. A local generalized least squares estimator

For the sake of comparison, we adopt the idea from Chen & Jin (2005) and consider the local generalized least squares estimating equation given below; thus, the resulting estimates are simple to compute.

$$\sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h)^{1/2} V_i^{-1} W_i(t_0, h)^{1/2} \left(\tilde{Y}_i - \tilde{X}_i^\top \beta(t_0) \right) = 0,$$

where $W_i(t_0, h)$ is an $m_i \times m_i$ diagonal matrix whose jth diagonal element is equal to $w_{ij}(t_0, h)$, V_i is the covariance matrix of $\tilde{Y}_i$ with the (j, l) element $(V_i)_{jl} = \sum_{k=1}^{M} \lambda_k \phi_k(T_{ij}) \phi_k(T_{il}) + \sigma^2 \delta_{jl}$, where $\lambda_k = \text{var}(\xi_{ik})$. Accordingly, we have that

$$\hat{\beta}^*(t_0) = \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h)^{1/2} V_i^{-1} W_i(t_0, h)^{1/2} \tilde{X}_i^\top \right\}^{-1} \left\{ \sum_{i=1}^{n} \tilde{X}_i W_i(t_0, h)^{1/2} V_i^{-1} W_i(t_0, h)^{1/2} \tilde{Y}_i \right\}. \tag{11}$$

We refer to the above estimator as the *local generalized least squares estimator* (LoGLSE) of $\beta(t_0)$. Analogous to the extension from (6) to (8), we can obtain the rth component estimate $\hat{\beta}_r^*(t_0)$ with an individual bandwidth h_r.

To handle the typically unknown covariance matrix V_i in (11), a simple procedure is to estimate V_i borrowing the technique from Yao et al. (2005a). Specifically, set $V_i = I$ in (11) and obtain an initial estimate of $\beta(t)$, say $\check{\beta}(t)$. Then obtain the covariance estimate $\hat{v}^*(s, t)$ via the two-dimensional smoothing of the raw covariances $\left\{ \left(Y_{ij} - X_{ij}^\top \check{\beta}(T_{ij}) \right) \left(Y_{il} - X_{il}^\top \check{\beta}(T_{il}) \right), 1 \leqslant j \neq l \leqslant m_i, 1 \leqslant i \leqslant n \right\}$. Let $|\mathcal{T}|$ be the interval length between $a = \inf\{t : t \in \mathcal{T}\}$ and $b = \sup\{t : t \in \mathcal{T}\}$. Accordingly, the covariance estimate $\hat{V}_i$ is obtained, where its (j, l) off-diagonal and jth diagonal elements are $\hat{v}^*(T_{ij}, T_{il})$ and $\hat{v}^*(T_{ij}, T_{ij}) + \hat{\sigma}^2$, respectively, and $\hat{\sigma}^2 = 2 \int_{a+|\mathcal{T}|/4}^{b-|\mathcal{T}|/4} \{\hat{\gamma}(t) - \hat{v}^*(t, t)\} dt / |\mathcal{T}|$ if $\hat{\sigma}^2 > 0$, and $\hat{\sigma}^2 = 0$, otherwise. Here, $\hat{\gamma}(\cdot)$ is the local linear smoother using the diagonal elements $\left\{ \left(Y_{ij} - X_{ij}^\top \check{\beta}(T_{ij}) \right)^2, 1 \leqslant j \leqslant m_i, 1 \leqslant i \leqslant n \right\}$. Although LoGLSE is naturally linked to the longitudinal data analysis, it does not fully utilize the functional data information.

3 Theoretical properties

Before we demonstrate asymptotic properties of the functional profiling-backfitting and local generalized least squares estimators, we present the following conditions.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid Dissemination of Statistics Research

(C1) The kernel function K is symmetric, bounded, and it is positive in $(-1, 1)$ and vanishes outside $(-1, 1)$. In addition, its second moment is $C_2 = \int u^2 K(u)du > 0$.
(C2) The probability density function of T_{ij}, $f(t)$, is positive and bounded on $\mathcal{T}$. Furthermore, $f(t)$ and $\beta(t)$ have smooth first and second derivatives ($f'(t)$ and $f''(t)$) and ($\beta'(t)$ and $\beta''(t)$), respectively.
(C3) There exist positive constants c_1 and c_2 such that $c_1 < h/h_r < c_2$ for $r = 1, \dots, L$.
(C4) For the covariate $X_i(T)$, $\max_i \sup_t ||X_i(t)||_\infty$ is bounded.
(C5) Let $m_0 = \min(m_1, \dots, m_n)$ and $m_a = \max(m_1, \dots, m_n)$. During the backfitting procedure of FPBE, h satisfies $h \to 0$, $Nh^2 \to \infty$, and $Nh^4 \to 0$ as $m_0 \to \infty$ and m_a/m_0 is bounded. Furthermore, we assume that $h = O(N^{-1/5})$ in the final estimation step of $\beta(t)$.
(C6) The covariance V_i is strictly positive definite. In addition, the entries of V_i and V_i^{-1} are uniformly bounded for $i = 1, \dots, n$.
(C7) There exists a positive constant C such that $m_i < C$ for $i = 1, \dots, n$. In addition, h satisfies $nh^4 \to \infty$ and $h \to 0$ as $n \to \infty$.

Condition (C1) is a standard requirement for the kernel function. Condition (C2) is the smoothness condition on the design density and the regression coefficient function. Furthermore, the time domain is restricted to be on a compact set. Moreover, the boundedness and positivity of $f(t)$ avoid a situation where all the observed times are clumped together or where no observations are obtained in a certain interval. Condition (C3) ensures that the oscillations of coefficient functions do not differ dramatically. This is not a stringent condition and is only used to simplify the proof. Condition (C4) assures the boundedness of the covariate functions. Condition (C5) is used for investigating the large sample properties of FPBE. Condition (C6) assures existence of V_i^{-1} and Condition (C7) is needed for studying the asymptotic properties of LoGLSE.

Theorem 1
Under regularity conditions (C1) to (C5), we have that (i) $\sqrt{N}(\hat{\xi} - \xi)$ converges to a multivariate normal distribution with mean zero and variance $\mathcal{F}^{-1}\Sigma\mathcal{F}^{-T}$ as $N \to \infty$, where $\mathcal{F}$ and Σ are defined in the Appendix. (ii) The functional profiling-backfitting estimator $\hat{\beta}(t_0)$ is asymptotically normally distributed with bias and variance of orders $O(h^2)$ and $O\{1/(Nh)\}$, respectively.

Theorem 2
Assume conditions (C1) to (C4), (C6) and (C7) hold. If $N \to \infty$, then $U^{-1/2}\{\hat{\beta}^*(t_0) - \beta(t_0) - G\}$ converges to the standard multivariate normal distribution, where the rth element of the bias G is

$$G_r = h_r^2 C_2 \left[\beta_r''(t_0)/2 + \beta_r'(t_0) f'(t_0)/f(t_0)\right] + O\left\{h_r^4 + (Nh_r)^{-1/2}\right\} = O(h^2),$$

the (r, s) element of U is $U_{r,s} = \sum_{i=1}^n e_r^\top A(h_r)^\top \tilde{X}_i W_i^{1/2}(h_r)^{1/2} V_i^{-1} W_i^{1/2}(h_r) V_i W_i^{1/2}(h_s) V_i^{-1} W_i^{1/2}(h_s) \tilde{X}_i^\top A(h_s) e_s = O\{1/(Nh)\}$, and $A(h) = \left\{\sum_{i=1}^n \tilde{X}_i W_i(t_0, h)^{1/2} V_i^{-1} W_i(t_0, h)^{1/2} \tilde{X}_i^\top\right\}^{-1}$.

The proof of Theorem 1 is given in the Appendix, while the proof of Theorem 2 is omitted for saving space and can be obtained from authors. Theorem 1 shows the asymptotic distributions of the FPB estimators $\hat{\xi}_i$ and $\hat{\beta}(t_0)$, and Theorem 2 demonstrates the asymptotic distribution of the LoGLS estimator $\hat{\beta}(t_0)$. It is noteworthy that the biases of $\hat{\beta}(t_0)$ and $\hat{\beta}^*(t_0)$ have the same order $O(h^2)$. Hence, it is of interest to examine their finite sample performance, which is presented in the next section.

Stat

Functional random effects model

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

4 Numerical studies

4.1. Simulation

To compare the functional profiling-backfitting estimator (FPBE) and the local generalized least squares estimator (LoGLSE), we consider the following simulation settings. The recording times $\{t_{ij}; i = 1, \ldots, n, j = 1, \ldots, m_i\}$ are sampled from the set $\{t_0, t_1, \ldots, t_m\}$ with equally spaced time intervals, where $t_0 = 1$, $t_m = 10$, $m = 50$, and the number of observations for each subject, m_i, follows a discrete uniform distribution. Four scenarios are considered for the distributions of m_i along with the sample size n: (N1) $Uniform(5, 10)$ and $n = 100$; (N2) $Uniform(5, 10)$ and $n = 500$; (N3) $Uniform(30, 40)$ and $n = 100$; and (N4) $Uniform(30, 40)$ and $n = 500$. As a result, (N1) and (N2) represent sparse designs, while (N3) and (N4) serve as dense designs. Then, we consider two regression functions with $L = 1$ and $L = 2$, respectively: (F1) $\beta(t) = \sin(t)$ and $X_i = 1$; and (F2) $\beta_1(t) = t$, $\beta_2(t) = \sin(t)$, $X_{1i} = 1$, and $X_{2i} = (i/n)^2$. Furthermore, the eigenfunctions are $\phi_1(t) = -\sqrt{2/T}\cos(\pi t/T)$ and $\phi_2(t) = \sqrt{2/T}\sin(\pi t/T)$. Their corresponding random coefficients ξ_{ik} $(k = 1, 2)$ are generated from (R1) a normal distribution of $N(0, \lambda_k)$; and (R2) a mixture normal distribution of $1/2N(\sqrt{\lambda_k/2}, \lambda_k/2) + 1/2N(-\sqrt{\lambda_k/2}, \lambda_k/2)$. Moreover, $\lambda_1 = 10$ and $\lambda_2 = 5$. Finally, the random errors ϵ_{ij} are simulated from (E1) $N(0, 1)$; and (E2) $N(0, 0.01)$.

To compute regression parameter estimates, we set c in (9) to 0.005 and the resulting estimates often converge in 2 or 3 iterations. For the sake of simplicity, we employ the same bandwidth to estimate different components of the regression parameters. Furthermore, the number of random components M used in the computation of FPBE is selected via a data-adaptive approach so that the proportion of total variance being explained reaches 90%. To assess the performance of different estimates, we conduct 500 Monte Carlo realizations. In the ℓth realization, we compute three performance measures: bias, variance, and unweighted average squared error (Fan & Zhang, 2000). Specifically, they are $BIAS_\ell = (1/m)\sum_{k=1}^{L}\sum_{j=1}^{m}\left|\beta_k(t_j) - \hat{\beta}_k^{(\ell)}(t_j)\right|$, $VAR_\ell = (1/m)\sum_{k=1}^{L}\sum_{j=1}^{m}\left\{\hat{\beta}_k^{(\ell)}(t_j) - \bar{\beta}_k(t_j)\right\}^2$, and $UASE_\ell = (1/4m)\sum_{j=1}^{m}\sum_{k=1}^{L}\left\{\beta_k(t_j) - \hat{\beta}_k^{(\ell)}(t_j)\right\}^2$, respectively, where $\bar{\beta}_k(t_j) = (1/500)\sum_{\ell=1}^{500}\hat{\beta}_k^{(\ell)}(t)$. Accordingly, we report the sample means of bias and variance as well as the sample median of the average squared error obtained from the 500 realizations.

Table I presents three performance measures across four sample sizes ((N1) to (N4)), two functional forms ((F1) and (F2)), two random coefficient structures ((R1) and (R2)), and two types of random errors ((E1) and (E2)). It indicates that FPBE outperforms LoGLSE in all 32 combinations. This can be explained from the following two aspects: (i) LoGLSE does not fully take into account the functional data information; (ii) LoGLSE requires taking the inverse of the estimated covariance matrices, which can be numerically unstable. Comparing (N1) to (N2) and (N3) to (N4), it is not surprising that the bias, variance, and unweighted mean squared errors decrease as the sample size n increases. Analogous results can be found by increasing the number of observations per subject (see (N1) versus (N3) and (N2) versus (N4), respectively, for sparse versus dense designs). It is also of interest to note that there are only minor differences between the performance of (R1) and (R2), indicating that the results are robust to the distributions of random coefficients. In conclusion, the finite sample performance of the estimator FPBE is consistent with theoretical findings. Since FPBE is superior to LoGLSE, we also recommend using the functional profiling-backfitting estimator in practical analysis.

4.2. Application to a biliary cirrhosis study

To further demonstrate the usefulness of the proposed functional profiling-backfitting method, we consider data from patients with primary biliary cirrhosis, collected in the ten year interval from January 1974 to May 1984 by the Mayo Clinic; see also the appendix of Fleming & Harrington (1991). The data are available at

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table I. Comparison of the FPB and LoGLS methods under four scenarios. (FPB: functional profiling-backfitting method; LoGLS: local generalized least squares method; BIAS: sample mean of $BIAS_\ell$; VAR: sample mean of VAR_ℓ; UASE: median of $UASE_\ell$.)

			F1				F2			
			R1		R2		R1		R2	
			FPB	LoGLS	FPB	LoGLS	FPB	LoGLS	FPB	LoGLS
E1	N1	BIAS	0.1617	0.1719	0.1672	0.1782	0.6317	0.7001	0.6139	0.6934
		VAR	0.0377	0.0471	0.0408	0.0499	0.5359	0.7009	0.5077	0.6748
		UASE	0.0094	0.0108	0.0098	0.0112	0.1182	0.1565	0.1126	0.1486
	N2	BIAS	0.0775	0.0798	0.0802	0.0828	0.2718	0.3132	0.2693	0.3127
		VAR	0.0071	0.0091	0.0077	0.0098	0.0884	0.1376	0.0878	0.1363
		UASE	0.0022	0.0023	0.0023	0.0024	0.0222	0.0321	0.0223	0.0305
	N3	BIAS	0.1142	0.1165	0.1218	0.1238	0.3925	0.4494	0.3901	0.4555
		VAR	0.0192	0.0203	0.0221	0.0233	0.1837	0.2466	0.1839	0.2601
		UASE	0.0038	0.0040	0.0042	0.0046	0.0386	0.0513	0.0380	0.0499
	N4	BIAS	0.0564	0.0562	0.0605	0.0607	0.1871	0.2061	0.1910	0.2125
		VAR	0.0037	0.0039	0.0045	0.0047	0.0388	0.0500	0.0400	0.0523
		UASE	0.0010	0.0010	0.0011	0.0011	0.0090	0.0113	0.0098	0.0113
E2	N1	BIAS	0.1226	0.1605	0.1312	0.1685	0.4010	0.6220	0.4021	0.6286
		VAR	0.0207	0.0403	0.0245	0.0456	0.1953	0.5680	0.1919	0.5638
		UASE	0.0046	0.0093	0.0051	0.0100	0.0395	0.1180	0.0394	0.1261
	N2	BIAS	0.0609	0.0772	0.0652	0.0803	0.1891	0.2886	0.1901	0.2899
		VAR	0.0037	0.0085	0.0046	0.0093	0.0359	0.1163	0.0367	0.1194
		UASE	0.0012	0.0022	0.0013	0.0022	0.0084	0.0257	0.0091	0.0258
	N3	BIAS	0.1064	0.1110	0.1148	0.1187	0.3349	0.4217	0.3425	0.4352
		VAR	0.0164	0.0181	0.0197	0.0213	0.1313	0.2398	0.1328	0.2456
		UASE	0.0031	0.0036	0.0035	0.0040	0.0249	0.0452	0.0267	0.0449
	N4	BIAS	0.0529	0.0541	0.0569	0.0576	0.1689	0.1933	0.1721	0.1958
		VAR	0.0031	0.0034	0.0038	0.0041	0.0295	0.0457	0.0305	0.0443
		UASE	0.0008	0.0009	0.0010	0.0010	0.0065	0.0090	0.0070	0.0096

http://lib.stat.cmu.edu/datasets/pbcseq. It is known that the serum albumin and prothrombin time are two commonly used indicators for liver function, among others. In advanced liver disease, the level of the serum albumin is reduced for insufficient production of proteins. Hence, it is not surprising to find a good correlation between abnormalities in coagulation measured by the prothrombin time and the degree of liver dysfunction. In the study of liver failure treatment in rats (Cai et al., 2002), improvements in prothrombin time and serum albumin level are included as indicators to show the effectiveness of transplantation of immortalized hepatocytes.

In this study, we explore the time-varying relationship between the two commonly used biomarkers of liver disease: prothrombin time (PT, in seconds) as the response variable, and serum albumin level (ALB, in mg/dl) as the predictor variable. The times of observation for each subject are irregularly spaced and sparsely distributed across 2000 days, and the visits of each subject range from one to sixteen days (see Figure 1). Four subjects are excluded from the analysis as they are obvious outliers. The observed trajectories and the corresponding mean profiles of PT and ALB

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

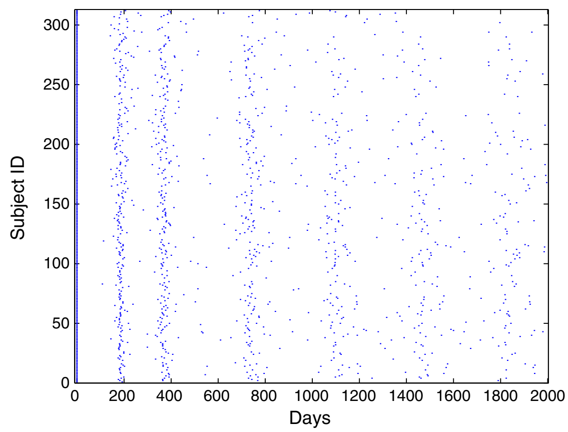

Figure 1. Observed time points of each subject.

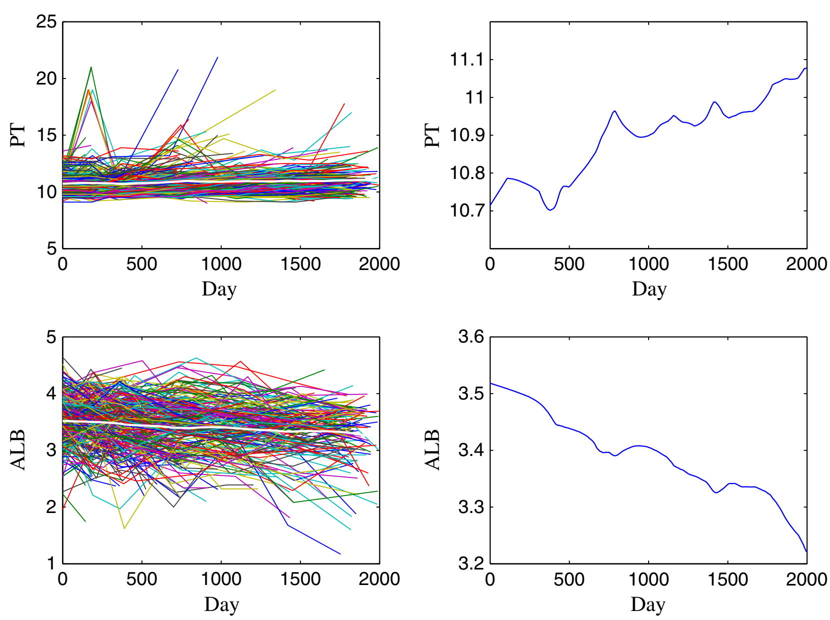

Figure 2. Trajectories of PT (upper left panel) and ALB (lower left panel) and their corresponding mean functions (right panels).

are presented in Figure 2. They indicate that PT has a general increasing trend, while ALB shows a decreasing trend in disease progression. Since the analysis is based on 272 female and 36 male patients, the indicator variable of gender effect (SEX) is also included. Accordingly, there are three explanatory variables: X_{0i} is the constant 1; $X_{1i}(t)$ denotes each individual's time-dependent ALB; and $X_{2i}(t) = 1$ if the subject is female, and $X_{2i}(t) = 0$ otherwise. Their associated regression coefficient functions are $\beta_0(t)$, $\beta_1(t)$, and $\beta_2(t)$, and the bandwidths used to estimate them are chosen by the aforementioned cross-validation method.

After fitting model (3), we apply a heuristic Akaike information criterion (AIC) given below to jointly select the predictor variables (L) and the number of principal components (M). To this end, we consider the conditional pseudo-Gaussian

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid Dissemination of Statistics Research

likelihood function,

$$\ell(L,M)=\sum_{i=1}^{n}\left[-\frac{m_i}{2}\log(2\pi)-\frac{m_i}{2}\log(\hat{\sigma}^2)-\frac{2}{\hat{\sigma}^2}\sum_{j=1}^{m_i}\left\{Y_{ij}-\sum_{l=1}^{L}X_{li}(t_{ij})\hat{\beta}_l(t_{ij})-\sum_{k=1}^{M}\hat{\xi}_{ik}\hat{\phi}_k(t_{ij})\right\}^2\right],$$

where $\hat{\sigma}^2$ is the estimated error variance. Given $M=M_0$, define $\mathrm{AIC}_1(L;M_0)=-\ell(L,M_0)+2L$. Then, obtain the optimal set of predictors by $\tilde{L}(M_0)=\arg\min_L \mathrm{AIC}_1(L;M_0)$. Analogously, given $L=L_0$, define $\mathrm{AIC}_2(M;L_0)=-\ell(L_0,M)+2M$. Subsequently, select the optimal M by $M^*=\arg\min_M \mathrm{AIC}_2(M;\tilde{L}(M))$ and then obtain $L^*=\tilde{L}(M^*)$. In addition to AIC, the Bayesian information criterion (BIC) can be developed via the same procedure by replacing the penalty terms $2L$ and $2M$ in AIC_1 and AIC_2 with $(\log N)L$ and $(\log N)M$, respectively, where $N=\sum_{i=1}^{n}m_i$. Table II shows that both AIC and BIC yield the same model, in which case the resulting optimal number of predictor variables and components are $L^*=2$ and $M^*=4$, respectively. Thus, the SEX predictor variable is omitted.

Based on the best fitted model, Figure 3 depicts the estimated regression coefficient functions and their corresponding 95% confidence intervals. Since the asymptotic variance of $\hat{\beta}(t_0)$ in Theorem 1 is not easy to compute for constructing confidence intervals, we apply a resampling scheme to obtain the variance estimate. Following the estimate of $\hat{\beta}(t_0)$

Table II. Selection of predictor variable (L) and principal component (M) via AIC and BIC criteria, respectively.

	AIC			BIC		
No. Comp.	$L=1$ $(I)^\dagger$	$L=2$ $(II)^\ddagger$	$L=3$ $(III)^\S$	$L=1$ (I)	$L=2$ (II)	$L=3$ (III)
$M=1$	3890.0	3831.6	3838.3	3895.3	3842.2	3854.2
$M=2$	3772.8	3726.8	3737.3	3778.2	3737.4	3753.3
$M=3$	3657.8	3644.7	3659.6	3663.1	3655.4	3675.5
$M=4$	3656.4	**3643.9**	3656.9	3661.7	**3654.5**	3672.8
$M=5$	3655.4	3650.4	3662.8	3660.8	3661.0	3678.7
$M=6$	3654.2	3651.3	3665.5	3659.5	3661.9	3681.4

$^\dagger I$: $Y_i(t)=\beta_0(t)+\sum_{k=1}^{M}\xi_{ik}\phi_k(t)$;
$^\ddagger II$: $Y_i(t)=\beta_0(t)+\beta_1(t)X_{i1}(t)+\sum_{k=1}^{M}\xi_{ik}\phi_k(t)$;
$^\S III$: $Y_i(t)=\beta_0(t)+\beta_1(t)X_{i1}(t)+\beta_2(t)X_{i2}+\sum_{k=1}^{M}\xi_{ik}\phi_k(t)$;
where $Y_i(t)=$PT, $X_{i1}=$ALB, and $X_{i2}=$SEX.

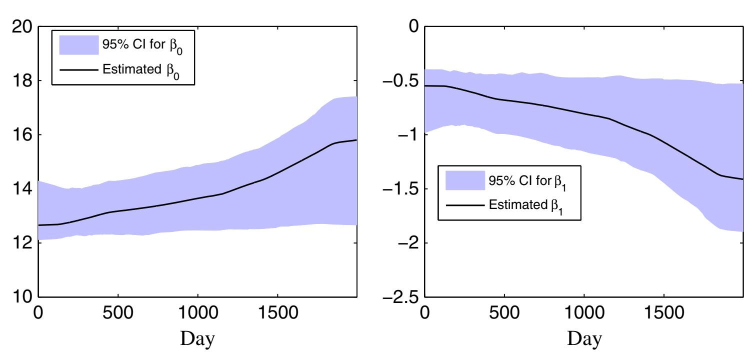

Figure 3. Estimated functions of the intercept (left) and the regression coefficient of ALB (right) with their corresponding 95% confidence intervals.

in (7), we have that

$$\text{cov}\left\{\hat{\beta}(t_0)\right\} = \Gamma_1(t_0,h)^{-1}\left\{\sum_{i=1}^{n}\tilde{X}_i W_i(t_0,h)\ \text{cov}\left(\tilde{Y}_i - \boldsymbol{\phi}_i^{\top}\hat{\xi}_i\right)\ W_i(t_0,h)\tilde{X}_i^{\top}\right\}\Gamma_1(t_0,h)^{-1}, \tag{12}$$

where $\Gamma_1(t_0,h) = \sum_{i=1}^{n}\tilde{X}_i W_i(t_0,h)\tilde{X}_i^{\top}$. Then, resample the observed curves $\{\tilde{Y}_i\}$ with replacement to obtain bootstrap samples, and subsequently use the resampled curves to obtain the estimate $\widehat{\text{cov}}\left(\tilde{Y}_i - \boldsymbol{\phi}_i^{\top}\hat{\xi}_i\right)$. Afterwards, we plugin the covariance estimate into (12) to obtain $\widehat{\text{cov}}(\hat{\beta}(t_0))$. The ALB covariate effect is clearly time-varying rather than a constant as $\hat{\beta}_1(t)$ is decreasing with time. It is of interest to note that the shapes of the estimated curves $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ look similar, but in the opposite direction. To interpret them, let $\hat{\mu}_Y(t)$ and $\hat{\mu}_{X_1}(t)$ be the estimated mean functions of PT and ALB (see the left panels of Figure 2). Since $\hat{\beta}_0(t) \approx \hat{\mu}_Y(t) - \hat{\beta}_1(t)\hat{\mu}_{X_1}(t)$, the intercept term is the mean of PT adjusted by the mean of ALB with the scale $|\hat{\beta}_1(t)|$. As a result, $\hat{\beta}_0(t)$ exhibits a general increasing trend for PT in the progression of liver disease. In contrast, the estimated regression function $\hat{\beta}_1(t)$ shows a decreasing trend. Hence, there is a negative correlation between PT and ALB across the 2000 days. Accordingly, a one-unit (mg/dl) decrease in ALB results in an average delay in PT of $|\hat{\beta}_1(t)|$ seconds, which depends on the progression time.

After the response PT is adjusted by the estimated regression coefficient functions (named adjusted-PT), Figure 4 (left panel) displays its estimated covariance surface. The associated leading eigenfunctions are presented in Figure 4 (right panel), which depicts variation directions of random effects after adjusting the ALB time-varying effects. Each of the first four eigenfunctions, respectively, explains 78.02%, 17.25%, 3.19%, and 0.32% of total variability in adjusted-PT. The first eigenfunction reveals a general decreasing pattern in terms of variations of adjusted-PT over disease progression time, similar to the overall mean level of ALB. The second eigenfunction shows a contrast between early and late disease progression. The third eigenfunction reflects additional variations that are in a complementary direction to the second eigenfunction. The fourth eigenfunction, while explaining a very small portion of total variation, catches the additional trend, especially on the right tail near the 2000th day. In summary, these four random components capture the subject effect resulting from individual patient differences. Consequently, the functional random effects model not only establishes the time-varying relationship between PT and ALB, but also explores the characteristic patterns and variation in the data. Since the trend in ALB is more steady than that in PT and there exists a relationship between them, ALB is likely to be a better biomarker in monitoring the progression of primary biliary cirrhosis.

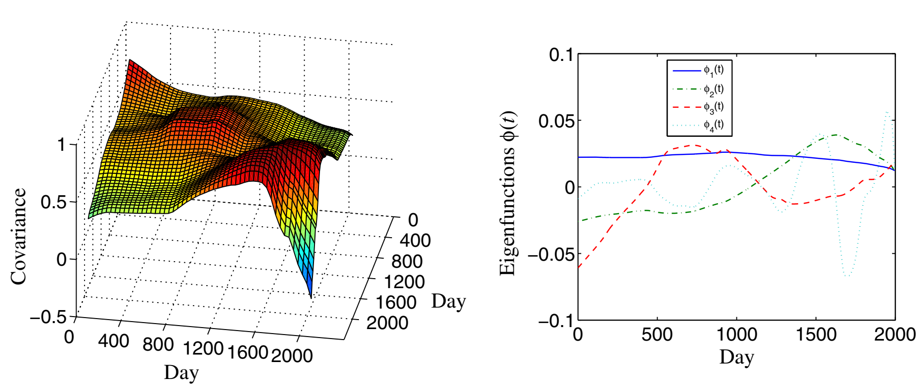

Figure 4. Estimated covariance function and its corresponding leading eigenfunctions.

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid
Dissemination of Statistics Research

Appendix: Proof of Theorem 1

To save space, we do not provide those conditions that are implicitly used in the derivations of the proof, and they can be found in Claeskens & van Keilegom (2003) and Chen et al. (2003). Denote $\mathcal{Y}_{ij} = (X_{ij}, T_{ij}, Y_{ij})$ and $\mathcal{Y}_{\cdot j} = (\mathcal{Y}_{1j} \ldots, \mathcal{Y}_{nj})$ for $j = 1, \ldots, m_a$, where $\mathcal{Y}_{ij} = 0$ if $j > m_i$. Then, let $\mathcal{L}_{\xi,j} = \mathcal{L}_\xi(\mathcal{Y}_{\cdot j}, \xi, \beta)$ be the $nM \times 1$ vector with the $\{(i-1)M+k\}$th component $\phi_k(T_{ij}) \left\{Y_{ij} - X_{ij}^T\beta(T_{ij}) - \sum_{l=1}^M \xi_{il}\phi_l(T_{ij})\right\}$ if $j \leqslant m_i$ and 0 if $j > m_i$. In addition, let $\Psi_{\beta,ij} = \Psi_\beta(\mathcal{Y}_{ij}, \xi, \beta) = X_{ij}\left\{Y_{ij} - X_{ij}^T\beta(t) - \sum_{k=1}^M \xi_{ik}\phi_k(T_{ij})\right\}$, which is an $L \times 1$ vector. It is noteworthy that $E\left\{\mathcal{L}_\xi(\mathcal{Y}_{\cdot j}, \xi, \beta)\right\} = 0$ and $E\left\{\Psi_\beta(\mathcal{Y}_{ij}, \xi, \beta)\right\} = 0$. Moreover, the estimating equations in (4) and (5) can be written as

$$\sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij}(t_0, h)\Psi_\beta(\mathcal{Y}_{ij}, \xi, \beta) = 0 \quad \text{and} \quad \sum_{i=1}^n \sum_{j=1}^{m_a} \mathcal{L}_\xi(\mathcal{Y}_{\cdot j}, \xi, \beta) = 0,$$

respectively, except adding a redundant $\sum_{i=1}^n$ on the left-hand side of the second equation to facilitate the proof. For the sake of simplicity, we sometimes omit the subscripts ij or $\cdot j$. Due to condition (C3), we also consider a common bandwidth h in the rest of the proof.

To show asymptotic results, we define $\mathcal{L}_{\xi\xi}$ as the partial derivative of $\mathcal{L}_\xi$ with respect to ξ, $\mathcal{L}_{\xi\beta}$ as the partial derivative of $\mathcal{L}_\xi$ with respect to β, $\Psi_{\beta\beta}$ as the partial derivative of Ψ_β with respect to β, and $\Psi_{\beta\xi}$ as the partial derivative of Ψ_β with respect to ξ. With the argument $(\bullet)$ being $\{\mathcal{Y}, \xi, \beta(T)\}$, we also define $\Omega(T) = E\left\{\Psi_{\beta\beta}(\bullet)|T\right\}$, $\mathcal{U}(T) = E\left\{\mathcal{L}_{\xi\beta}(\bullet)|T\right\}\Omega(T)^{-1}$, and $J_\xi(T) = -\Omega(T)^{-1}E\left\{\Psi_{\beta\xi}(\bullet)|T\right\}$.

Using standard expansion, it can be shown that

$$-\mathcal{F}N^{1/2}(\hat{\xi} - \xi) = N^{-1/2}\sum_{i=1}^n \sum_{j=1}^{m_a}\left[\mathcal{L}_{\xi,j}(\bullet) + \mathcal{L}_{\xi\beta,j}(\bullet)\left\{\hat{\beta}(T_{ij}, \xi) - \beta(T_{ij})\right\}\right] + o_p(1), \tag{A.1}$$

where $\mathcal{F} = E\left[\mathcal{L}_{\xi\xi}\{\mathcal{Y}, \xi, \beta(T)\} + \mathcal{L}_{\xi\beta}\{\mathcal{Y}, \xi, \beta(T)\}J_\xi(T, \xi)\right]$. From the local estimating equation and using conditions (C1), (C2) and (C4), we have that

$$\hat{\beta}(t_0, \xi) - \beta(t_0) = (h^2/2)\beta''(t_0) - N^{-1}\sum_{i=1}^n \sum_{j=1}^{m_a} w_{ij}(t_0, h)\Omega(t_0)^{-1}\Psi_\beta\left\{\mathcal{Y}_{ij}, \xi, \beta(T_{ij})\right\}/f(t_0) + o_p(N^{-1/2}). \tag{A.2}$$

Substituting (A.2) into (A.1) and then employing conditions (C3) and (C5), we obtain that

$$-\mathcal{F}N^{1/2}(\hat{\xi} - \xi) = N^{-1/2}\sum_{i=1}^n \sum_{j=1}^{m_a} \mathcal{L}_\xi\left\{\mathcal{Y}_{\cdot j}, \xi, \beta(T_{ij})\right\} - \mathcal{U}(T_{ij})\Psi_\beta\left\{\mathcal{Y}_{ij}, \xi, \beta(T_{ij})\right\} + o_p(1). \tag{A.3}$$

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

Hence, $N^{1/2}(\hat{\xi}-\xi)$ is asymptotically normally distributed with mean zero and covariance matrix $\mathcal{F}^{-1}\Sigma\mathcal{F}^{-\top}$, where $\Sigma = \text{cov}\left\{\mathcal{L}_\xi(\bullet) - \Psi_\beta(\bullet)\mathcal{U}(T)\right\}$. We next demonstrate the asymptotic results of $\hat{\beta}(t_0)$.

Combining (A.2) and (A.3), in conjunction with conditions (C3) and (C5), we obtain that

$$\begin{aligned}
\hat{\beta}(t_0,\hat{\xi}) - \beta(t_0) &= \left\{\hat{\beta}(t,\hat{\xi}) - \hat{\beta}(t_0,\xi)\right\} + \left\{\hat{\beta}(t_0,\xi) - \beta(t_0)\right\} \\
&= \hat{\beta}_\xi(t_0,\xi)(\hat{\xi}-\xi) + \left\{\hat{\beta}(t_0,\xi) - \beta(t_0)\right\} + o_p(N^{-1/2}) \\
&= -\hat{\beta}_\xi(t_0,\xi)\mathcal{F}^{-1}N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_a}\left\{\mathcal{L}_{\xi,j}(\bullet) - \mathcal{U}(T_{ij})\Psi_{\beta,ij}(\bullet)\right\} \\
&\quad + (h^2/2)\beta''(t_0) - N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}(t_0,h)\Omega(t)^{-1}\Psi_{\beta,ij}(\bullet)/f_T(t_0) + o_p(N^{-1/2}) \\
&= (h^2/2)\beta''(t_0) - N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}(t_0,h)\Omega(t)^{-1}\Psi_{\beta,ij}(\bullet)/f_T(t_0) + O_p(N^{-1/2}),
\end{aligned}$$

where $\hat{\beta}_\xi(t_0,\xi)$ is the first derivative matrix of $\hat{\beta}(t_0,\xi)$ respect to ξ. Taking the expectation from the above equation, together with the fact that the final estimation of $\beta(t_0)$ is evaluated under $h = O(N^{-1/5})$, leads to $\hat{\beta}(t_0)$ having the bias $\beta''(t_0)h^2/2 + o(h^2)$. Applying the standard approach, we can further show that it is asymptotically normally distributed with the following variance:

$$\begin{aligned}
N^{-1}\text{var}\left\{w_{ij}(t_0,h)\Omega(t_0)^{-1}\Psi_{\beta,ij}(\bullet)/f_T(t_0)\right\} &= N^{-1}E\left\{w_{ij}^2(t_0,h)\Omega(t_0)^{-1}\Psi_{\beta,ij}(\bullet)\Psi_{\beta,ij}(\bullet)^T\Omega(t_0)^{-T}/f_T^2(t_0)\right\} \\
&= N^{-1}\Omega(t_0)^{-1}E\left\{w_{ij}^2(t_0,h)\Psi_{\beta,ij}(\bullet)\Psi_{\beta,ij}(\bullet)^T\right\}\Omega(t_0)^{-T}/f_T^2(t_0) \\
&= O\left\{(Nh)^{-1}\right\}.
\end{aligned}$$

Hence, the proof is complete. □

Acknowledgements

This research was partially supported by grants from the Taiwan National Science Council (98-2118-M-001-019-MY3), Academia Sinica (99-ASIS-02), US National Science Foundation and National Institute of Stroke and Neurological Disorders.

References

Ash, RB & Gardner, MF (1975), *Topics in Stochastic Processes*, Academic Press, New York.

Cai, J, Ito, M, Nagata, H, Westerman, KA, LaFleur, D, Chowdhury, JR, Leboulch, P & Fox, IJ (2002), 'Treatment of liver failure in rats with end-stage cirrhosis by transplantation of immortalized hepatocytes', *Hepatology*, **36**, 386–394.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

The ISI's Journal for the Rapid Dissemination of Statistics Research

Brumback, BA & Rice, JA (1998), 'Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion)', *Journal of the American Statistical Association*, **93**, 961–994.

Chen, X, Linton, O & van Keilegom, I (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica*, **71**, 1591–1608.

Chen, K & Jin, Z (2005), 'Local polynomial regression analysis of clustered data', *Biometrika*, **92**, 59–74.

Claeskens, G & van Keilegom, I (2003), 'Bootstrap confidence bands for regression functions and their derivatives', *The Annals of Statistics*, **31**, 1852–1884.

Fan, J & Gijbels, I (1996), *Local Polynomial Modelling and its Applications*, Chapman & Hall, London.

Fan, J & Zhang, JT (2000), 'Two-step estimation of functional linear models with applications to longitudinal data', *Journal of the Royal Statistical Society, Series B*, **62**, 303–322.

Fan, J & Zhang, W (1999), 'Statistical estimation in varying coefficient models', *The Annals of Statistics*, **27**, 1491–1518.

Faraway, JJ (1997), 'Regression analysis for a functional response', *Technometrics*, **39**, 254–261.

Fleming, TR & Harrington, DP (1991), *Counting Processes and Survival Analysis*, John Wiley & Sons, Chichester.

Hall, P & Hosseini-Nasab, M (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society, Series B*, **68**, 109–126.

Hall, P & Vial, C (2006), 'Assessing the finite dimensionality of functional data', *Journal of the Royal Statistical Society, Series B*, **68**, 689–705.

Hastie, T & Tibshirani, R (1990), *Generalized Additive Models*, Chapman & Hall, New York.

Hastie, T & Tibshirani, R (1993), 'Varying coefficient models (with discussion)', *Journal of the Royal Statistical Society, Series B*, **55**, 757–796.

Hoover, DR, Rice, JA, Wu, CO & Yang, LP (1998), 'Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data', *Biometrika*, **85**, 809–822.

Huang, JZ, Wu, CO & Zhou, L (2002), 'Varying-coefficient models and basis function approximations for analysis of repeated measurements', *Biometrika*, **89**, 111–128.

Lin, DY & Ying, Z (2001), 'Semiparametric and nonparametric regression analysis for longitudinal data', *Journal of the American Statistical Association*, **96**, 103–126.

Ramsay, JO & Dalzell, CJ (1991), 'Some tools for functional data analysis', *Journal of the Royal Statistical Society, Series B*, **53**, 539–572.

Ramsay, JO & Silverman, BW (2005), *Functional Data Analysis*, Springer, New York.

Rice, JA & Silverman, BW (1991), 'Estimating the mean and covariance structure nonparametrically when data are curves', *Journal of the Royal Statistical Society, Series B*, **53**, 233–243.

Sentürk, D & Müller, HG (2010), 'Functional varying coefficient models for longitudinal data', *Journal of the American Statistical Association*, **105**, 1256–1264.

Shen, Q & Faraway, J (2004), 'An F test for linear models with functional responses', *Statistica Sinica*, **14**, 1239–1257.

Wang, H & Xia, C (2009), 'Shrinkage estimation of the varying coefficient model', *Journal of the American Statistical Association*, **104**, 747–757.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.10

Wang, L, Li, H & Huang, JZ (2008), 'Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements', *Journal of the American Statistical Association*, **103**, 1556–1569.

Wu, CO, Chiang, CT & Hoover, DR (1998), 'Asymptotic confidence regions for kernel smoothing of a varying coefficient models with longitudinal data', *Journal of the American Statistical Association*, **93**, 1388–1402.

Yao, F, Müller, HG & Wang, JL (2005a), 'Functional linear regression analysis for longitudinal data', *The Annals of Statistics*, **33**, 2873–2903.

Yao, F, Müller, HG & Wang, JL (2005b), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association*, **100**, 577–590.