Statistica Sinica **21** (2011), 1831-1856
doi:http://dx.doi.org/10.5705/ss.2009.328

VARYING COEFFICIENT MODELS FOR SPARSE NOISE-CONTAMINATED LONGITUDINAL DATA

Damla Şentürk and Danh V. Nguyen

University of California, Los Angeles and University of California, Davis

Abstract: In this paper we propose a varying coefficient model for sparse longitudinal data that allows for error-prone time-dependent variables and time-invariant covariates. We develop a new estimation procedure, based on covariance representation techniques, that enables effective borrowing of information across all subjects in sparse and irregular longitudinal data observed with measurement error, a challenge for which there is no current adequate solution. Sparsity is addressed via a functional analysis approach that considers the observed longitudinal data as noise contaminated realizations of a random process that produces smooth trajectories. This approach allows for estimation based on pooled data, borrowing strength from all subjects, in targeting the mean functions and auto- and cross-covariances to overcome sparse noisy designs. The resulting estimators are shown to be uniformly consistent. Consistent prediction for the response trajectories are also obtained via conditional expectation under Gaussian assumptions. Asymptotic distributions of the predicted response trajectories are derived, allowing for construction of asymptotic pointwise confidence bands. Efficacy of the proposed method is investigated in simulation studies and compared to the commonly used local polynomial smoothing method. The proposed method is illustrated with a sparse longitudinal data set, examining the age-varying relationship between calcium absorption and dietary calcium. Prediction of individual calcium absorption curves as a function of age are also examined.

Key words and phrases: Functional data analysis, local least squares, measurement error, repeated measurements, smoothing, sparse design.

1. Introduction

Varying coefficient models (Cleveland, Grosse, and Shyu (1991); Hastie and Tibshirani (1993)) are extensions of parametric regression models that have attracted many applications in diverse scientific research areas in the last fifteen years. An example is the modeling of the time-varying relationship between virologic response and immunologic status (as measured by viral load and CD4+ status) and other covariates in AIDS clinical studies. As recently reviewed by Fan and Zhang (2008), estimation in varying coefficient models for longitudinal data is based on three main approaches: polynomial splines (Huang, Wu, and Zhou (2002, 2004)), smoothing splines (Hoover et al. (1998); Chiang, Rice, and Wu (2001)), and perhaps the most natural approach of all, local polynomial

smoothing (Wu, Chiang, and Hoover (1998); Hoover et al. (1998); Fan and Zhang (2000) Wu and Chiang (2000)). Qu and Li (2006) proposed penalized spline and quadratic inference functions for incorporating the correlation structure into the estimation.

Although these approaches may be more effective for densely measured longitudinal data, sparse longitudinal data combined with measurement error poses unique unresolved challenges. Here sparsity refers to the irregular measurement times between subjects and the availability of only a few observed repetitions per subject in longitudinal designs. With measurement error, estimates of the varying coefficient functions with existing methods are biased. Measurement error is typical in studies of dietary intake (Carroll et al. (2006)), such as individual calcium absorption and dietary calcium in adult women population as considered in Section 4. Of particular interest is estimation of the age-varying relationship between calcium absorption and dietary calcium and other baseline measures, such as body surface area. In addition to inherent measurement errors, the data is sparse with subject ages ranging from 39 to 58, and due to dropouts and missed visits, about 40% of the subjects with only one or two measurements. For estimation of the varying coefficient function and prediction of individual calcium absorption curves, an effective strategy for pooling information across subjects is needed.

In this paper we take a functional analysis approach to multiple varying coefficient modeling of noise-contaminated sparse longitudinal data, and we develop a new estimation method that allows for both cross-sectional (time-invariant) and longitudinal predictors. The main idea is to view the observed longitudinal data as a noise contaminated realization of a stochastic process that produces smooth trajectories; this allows for pooling of information across subjects in order to strengthen the estimation. We note that functional data analysis has been extended to sparse longitudinal data in the context of functional regression models, by Yao, Müller, and Wang (2005a). More recently, Şentürk and Müller (2010) considered estimation in *functional* varying coefficient models with one covariate process that incorporates a history index. However, as the authors point out, their estimation approach is also useful for univariate varying coefficient models relating a longitudinal response process to a single longitudinal predictor process. The authors represent the varying coefficient functions using auto- and cross-covariances of the underlying stochastic processes that are then estimated based on all the data. We utilize similar representations for the varying coefficient functions and propose an estimation procedure for multiple predictor processes, which may include cross-sectional and longitudinal covariates. Several important distinctions of the current proposal from Şentürk and Müller's methodology are as follows. First, the current proposal is designed for multiple predictor processes. Second, cross-sectional predictors are included in the functional analysis

approach proposed. We note that incorporation of cross-sectional predictor variables is not common in functional linear models and functional data analysis. We develop the estimation method to accommodate these features and study its theoretical and finite sample properties. The proposed estimation procedure enables a novel way of incorporating the within-subject correlation and the handling of sparse noise-contaminated longitudinal designs, which leads to improved finite sample performance relative to the commonly used local polynomial smoothing methods for varying coefficient models.

In the next section we describe the proposed estimation procedure for the multiple varying coefficient model with time-dependent and time-invariant covariates in its full generality, and provide uniform consistency of the proposed estimators. Consistent predictors of response trajectories, via conditional expectation obtained under Gaussian assumptions, are given in Section 3, along with their asymptotic distributions and asymptotic pointwise confidence bands. In Section 4 the method is illustrated with the aforementioned data set, where we examine the age-varying relationship between calcium absorption and dietary calcium. Simulation studies, including comparisons with local polynomial smoothing, and concluding remarks follow in Section 5 and 6, respectively. Technical assumptions and proofs are given in the Appendix.

2. Estimation in Multiple Varying Coefficient Models

2.1. Sparse data and model representation

Consider the observed data, consisting of p time-dependent and q time-independent predictors along with a time-dependent response. The q time-independent predictors Z_{gi}, $i = 1, \ldots, n$, $g = 1, \ldots, q$ are assumed to have finite variance. The time-dependent predictors X_{ri} and response Y_i, $i = 1, \ldots, n$, $r = 1, \ldots, p$, are square integrable random realizations of the smooth random processes X_r and Y, respectively, both defined on a finite and closed interval domain $[0, T]$. Predictor and response processes X and Y have smooth mean functions $\mu_{X_r}(t) = EX_r(t)$, $\mu_Y(t) = EY(t)$, and (auto-)covariance functions $G_{X_rX_r}(s,t) = \text{Cov}\ \{X_r(s), X_r(t)\}$, $G_{YY}(s,t) = \text{Cov}\,\{Y(s), Y(t)\}$, for $s, t \in [0, T]$ and $r = 1, \ldots, p$. Orthogonal expansions of the covariances, $G_{X_rX_r}(s,t) = \sum_{m=1}^{\infty} \rho_{rm}\phi_{rm}(s)\phi_{rm}(t)$ and $G_{YY}(s,t) = \sum_{k=1}^{\infty} \lambda_k\psi_k(s)\psi_k(t)$ for $s, t \in [0, T]$ and $r = 1, \ldots, p$, follow under mild conditions, where ϕ_{rm} and ψ_k denote the eigenfunctions with nonincreasing eigenvalues ρ_{rm} and λ_k. The sparse design (SD), after Şentürk and Müller (2010), can formally be described as follows.

(SD) For the i-th subject one has a random number N_i of repeated measurements on the rth time-dependent predictor $X_{rij} = X_{ri}(T_{ij}) + \varepsilon_{rij}$, and

on the response $Y_{ij} = Y_i(T_{ij}) + \epsilon_{ij}$, $j = 1, \ldots, N_i$, obtained at i.i.d. random time points $T_{i1}, \ldots, T_{iN_i}$, where $\varepsilon_{rij}, \epsilon_{ij}$ are zero mean, finite variance i.i.d. measurement errors. The N_i are assumed to be i.i.d., with $N_i, T_{ij}, \varepsilon_{rij}, \epsilon_{ij}$ mutually independent, and independent of the underlying processes X_{ri}, Y_i, as well as Z_{gi}. Hence the predictor and response observations can be represented as $X_{rij} = \mu_{X_r}(T_{ij}) + \sum_{m=1}^{\infty} \xi_{rim}\phi_{rm}(T_{ij}) + \varepsilon_{rij}$, $Y_{ij} = \mu_Y(T_{ij}) + \sum_{k=1}^{\infty} \zeta_{ik}\psi_k(T_{ij}) + \epsilon_{ij}$, where ξ_{rim}, ζ_{ik} are uncorrelated, mean zero functional principal component scores with second moments equal to the eigenvalues ρ_{rm} and λ_k, respectively.

The representations in (SD) follow from the Karhunen-Loève expansion (see Ash and Gardner (1975)), and we assume $\sum_m \rho_{rm} < \infty$ and $\sum_k \lambda_k < \infty$.

Consider the multiple varying coefficient model

$$E\{Y(t) \mid X_1(t), \ldots, X_p(t), Z_1, \ldots, Z_q\} = \beta_0(t) + \sum_{r=1}^{p} \beta_r(t) X_r(t) + \sum_{g=1}^{q} \alpha_g(t) Z_g, \tag{2.1}$$

where the varying coefficient functions, $\beta_r(t)$ and $\alpha_g(t)$, are assumed to be smooth functions. Note that for each fixed t, (2.1) reduces to a standard linear model. Centering the predictor and response trajectories, i.e., $X_r^C(t) = X_r(t) - \mu_{X_r}(t)$, $Z_g^C = Z_g - E(Z_g)$, and $Y^C(t) = Y(t) - \mu_Y(t)$, we can express (2.1) as

$$E\{Y^C(t) \mid X_1(t), \ldots, X_p(t), Z_1, \ldots, Z_q\} = \sum_{r=1}^{p} \beta_r(t) X_r^C(t) + \sum_{g=1}^{q} \alpha_g(t) Z_g^C.$$

Note that, alternatively, $\beta_0(t)$ can be given as $\mu_Y(t) - \sum_r \beta_r(t)\mu_{X_r}(t) - \sum_g \alpha_g(t) E(Z_g)$.

2.2. Local linear smoothing

A standard method for fitting varying coefficient models (Fan and Zhang (2008)) is local polynomial smoothing. For instance, local linear fitting would minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} K\left(\frac{T_{ij}-t}{h}\right)\left[Y_{ij}^C - \sum_{r=1}^{p}\{\theta_{r,0}+\theta_{r,1}(t-T_{ij})\}X_{rij}^C - \sum_{g=1}^{q}\{\gamma_{g,0}+\gamma_{g,1}(t-T_{ij})\}Z_{gi}^C\right]^2, \tag{2.2}$$

with respect to $\theta_{r,0}, \theta_{r,1}, \gamma_{g,0}, \gamma_{g,1}$, leading to $\hat{\beta}_r(t) = \hat{\theta}_{r,0}$ and $\hat{\alpha}_g(t) = \hat{\gamma}_{g,0}$ (Hoover et al. (1998)). The minimization at (2.2) requires a specified kernel function $K(\cdot)$ that corresponds to a symmetric probability density function associated with a bandwidth h. Şentürk and Müller (2010) point out that local polynomial smoothing does not take advantage of the functional nature of the underlying

processes. In other words, while (2.2) involves each repeated observation taken on a subject, it does not involve the cross product terms between the repetitions that would correspond to the underlying covariance structure. They also point out that local polynomial smoothing is biased in the case of sparse and noise-corrupted measurements. We will demonstrate that this bias is also present in the multiple varying coefficient model, through simulations.

2.3. Proposed estimation procedure

The proposed approach utilizes the functional nature of the covariate processes. We take

$$G_{YX_r}(s,t) = \mathrm{Cov}\{Y(s), X_r(t)\} = \sum_{m=1}^{\infty}\sum_{k=1}^{\infty} E(\xi_{rm}\zeta_k)\phi_{rm}(s)\psi_k(t),$$

$$G_{X_rX_{r'}}(s,t) = \mathrm{Cov}\{X_r(s), X_{r'}(t)\} = \sum_{m=1}^{\infty}\sum_{m'=1}^{\infty} E(\xi_{rm}\xi_{r'm'})\phi_{rm}(s)\phi_{r'm'}(t),$$

$$G_{YZ_g}(t) = \mathrm{Cov}\{Y(t), Z_g\} = \sum_{k=1}^{\infty} E(\zeta_k Z_g)\psi_k(t),$$

$$G_{X_rZ_g}(t) = \mathrm{Cov}\{X_r(t), Z_g\} = \sum_{m=1}^{\infty} E(\xi_{rm} Z_g)\phi_{rm}(t),$$

and $G_{Z_gZ_{g'}} = \mathrm{Cov}(Z_g, Z_{g'})$. From (2.1),

$$G_{YX_{r'}}(t,t) = \sum_{r=1}^{p}\beta_r(t)G_{X_rX_{r'}}(t,t) + \sum_{g=1}^{q}\alpha_g(t)G_{X_{r'}Z_g}(t),$$

$$G_{YZ_{g'}}(t) = \sum_{r=1}^{p}\beta_r(t)G_{X_rZ_{g'}}(t) + \sum_{g=1}^{q}\alpha_g(t)G_{Z_gZ_{g'}},$$

for $r' = 1,\ldots,p$ and $g' = 1,\ldots,q$. Then the varying coefficient functions of interest are

$$[\beta_1(t),\ldots\beta_p(t),\alpha_1(t),\ldots,\alpha_q(t)]^{\mathrm{T}} = \mathcal{X}_t^{-1}\Xi_t, \tag{2.3}$$

where

$$\mathcal{X}_t = \left[\begin{array}{ccc|ccc} G_{X_1X_1}(t,t) & \ldots & G_{X_1X_p}(t,t) & G_{X_1Z_1}(t) & \ldots & G_{X_1Z_q}(t) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ G_{X_1X_p}(t,t) & \ldots & G_{X_pX_p}(t,t) & G_{X_pZ_1}(t) & \ldots & G_{X_pZ_q}(t) \\ \hline G_{X_1Z_1}(t) & \ldots & G_{X_pZ_1}(t) & G_{Z_1Z_1} & \ldots & G_{Z_1Z_q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ G_{X_1Z_q}(t) & \ldots & G_{X_pZ_q}(t) & G_{Z_1Z_q} & \ldots & G_{Z_qZ_q} \end{array}\right] \tag{2.4}$$

and $\Xi_t = [G_{YX_1}(t,t), \ldots, G_{YX_p}(t,t), G_{YZ_1}(t), \ldots, G_{YZ_q}(t)]^{\mathrm{T}}$. Estimation of the varying coefficient functions in (2.3) involves first obtaining estimates of the auto- and cross-covariances in $\mathcal{X}_t$ and Ξ_t, and then using the plug-in estimator $\widehat{\mathcal{X}}_t^{-1}\widehat{\Xi}_t$. The special case of $p = 1$ and $q = 0$ is considered by Şentürk and Müller (2010), based on the relations $\beta_1(t) = G_{YX}(t,t)/G_{XX}(t,t)$. The proposal here is adopted for multiple predictor processes in addition to cross-sectional predictors. The estimation algorithm is given as follows.

1. ***Mean functions.*** Estimate the mean functions for the predictor and response processes by smoothing the aggregated data (T_{ij}, X_{rij}) and (T_{ij}, Y_{ij}) for $j = 1, \ldots, N_i$, and $i = 1, \ldots, n$, with local linear fitting. Denote the estimated mean functions by $\hat{\mu}_{X_r}$ and $\hat{\mu}_Y$.
2. ***Raw covariances.*** Compute the raw covariances of X_r and Z_g, and the raw cross-covariances between (Y, X_r), (X_r, Z_g), and (Y, Z_g), based on all observations from the same subject, as

$$\begin{aligned} G_{X_rX_{r'},i}(T_{ij}, T_{i\ell}) &= \{X_{rij} - \hat{\mu}_{X_r}(T_{ij})\}\{X_{r'i\ell} - \hat{\mu}_{X_{r'}}(T_{i\ell})\}, \\ G_{Z_gZ_{g'},i} &= (Z_{gi} - \bar{Z}_g)(Z_{g'i} - \bar{Z}_{g'}), \\ G_{YX_r,i}(T_{ij}, T_{i\ell}) &= \{Y_{ij} - \hat{\mu}_Y(T_{ij})\}\{X_{ri\ell} - \hat{\mu}_{X_r}(T_{i\ell})\}, \\ G_{X_rZ_g,i}(T_{ij}) &= \{X_{rij} - \hat{\mu}_{X_r}(T_{ij})\}\{Z_{gi} - \bar{Z}_g\} \text{ and} \\ G_{YZ_g,i}(T_{ij}) &= \{Y_{ij} - \hat{\mu}_Y(T_{ij})\}\{Z_{gi} - \bar{Z}_g\}, \end{aligned}$$

   for $j, \ell = 1, \ldots, N_i$ and $i = 1, \ldots, n$.
3. ***Smoothed covariances.*** (3A) The estimates of the two-dimensional auto- and cross-covariances, $\widehat{G}_{X_rX_{r'}}$ and $\widehat{G}_{YX_r}$, are obtained by feeding the corresponding two-dimensional raw covariances, $G_{X_rX_{r'},i}$ and $G_{YX_r,i}$, from Step 2, into a two dimensional local least squares smoothing algorithm. (3B) The estimates of the one-dimensional cross-covariances, $\widehat{G}_{X_rZ_g}$ and $\widehat{G}_{YZ_g}$, are obtained by feeding the corresponding one-dimensional raw cross-covariances, $G_{X_rZ_g,i}$ and $G_{YZ_g,i}$, into a one-dimensional local polynomial smoothing algorithm. In addition the variance estimator $\widehat{G}_{Z_gZ_{g'}}$ is given as $n^{-1}\sum_{i=1}^n G_{Z_gZ_{g'},i}$.
4. ***Plug-in estimator.*** Estimators for the varying coefficient functions are obtained by using plug-in estimators for $\mathcal{X}_t$ and Ξ_t: $[\hat{\beta}_1(t), \ldots \hat{\beta}_p(t), \hat{\alpha}_1(t), \ldots, \hat{\alpha}_q(t)]^{\mathrm{T}} = \widehat{\mathcal{X}}_t^{-1}\widehat{\Xi}_t$. The estimate of the intercept function can be given as $\hat{\beta}_0(t) = \hat{\mu}_Y(t) - \sum_{r=1}^p \hat{\beta}_r\hat{\mu}_{X_r}(t) - \sum_{g=1}^q \hat{\alpha}_g(t)\bar{Z}_g$, as noted in Section 2.1.

We note that this estimation procedure differs from the standard methods for fitting varying coefficient models that do not take advantage of the covariance structure of underlying processes. Using this structure in the estimation process makes it possible to handle the sparsity of the longitudinal data, but also

allows for incorporating additional information that is inherent in the underlying covariance structure in the estimation step.

Remark 1. For estimation of the auto-covariances $G_{X_rX_r}$, the diagonal of the raw auto-covariance matrix is removed before the two-dimensional smoothing step, in order to eliminate the effects of measurement error on the longitudinal predictors. This covariance estimation step (inspired by the approach in Yao, Müller, and Wang (2005a,b)) achieves two major objectives. First, it eliminates the effect of the noise contamination on the longitudinal observations. Second, through pooling of the data across subjects, it overcomes the problems associated with the sparseness of the design. In addition, to guarantee that the estimator of $G_{X_rX_r}$ is nonnegative definite, we propose an adjusted estimator in which we exclude the negative estimates of the eigenvalues and corresponding eigenfunctions in the functional principal component decomposition of the covariance function. More precisely, a nonparametric functional principal component analysis step employed on the smooth estimate of the auto-covariance surface yields estimators for $\phi_{rm}(t)$ and ρ_{rm}; details are described in Appendix A.1. Then $\widehat{G}_{X_rX_r}$ is given as $\sum_{m:\hat{\rho}_{rm}>0}^{M_r} \hat{\rho}_{rm}\hat{\phi}_{rm}(s)\hat{\phi}_{rm}(t)$. The number M_r of included eigenfunctions can be chosen by a one-curve-leave-out cross-validation, the Akaike information criterion (AIC), fraction of variance explained, or similar criteria.

Remark 2. Explicit forms of all one- and two-dimensional smoothing estimators are assembled in Appendix A.1.

2.4. Uniform consistency

The proposed estimators for the varying coefficient functions are uniformly consistent, as summarized in Theorem 1. The assumptions and proof can be found in the Appendix.

Theorem 1. *Under Assumptions* (A) *in the Appendix, the varying coefficient function estimators satisfy* $\sup_{t\in[0,T]} |\hat{\beta}_r(t)-\beta_r(t)| = O_p(\tau_n)$ *and* $\sup_{t\in[0,T]} |\hat{\alpha}_g(t) - \alpha_g(t)| = O_p(\tau_n)$, *for* $r = 0,\dots,p$, *and* $g = 1,\dots,q$, *where* $\tau_n = n^{-1/2}(\sum_{r=1}^{p}(h_{r1}h_{r2})^{-1} + \sum_{r=1}^{p}\sum_{r'=1}^{p}(h_{X_r}h_{X_{r'}})^{-1})$.

In the expression for τ_n, the bandwidths used in the two-dimensional smoothing step of the raw covariances to obtain the cross-covariance function $\widehat{G}_{YX_r}$ are h_{r1} and h_{r2}. Similarly, the corresponding bandwidths used in the two-dimensional smoothing step to obtain the cross-covariance surface $\widehat{G}_{X_rX_{r'}}$ are denoted by h_{X_r} and $h_{X_{r'}}$. Details are given in Appendix A.1. The bandwidths are required to converge to zero, and to satisfy some other restrictions as outlined in Appendix A.2.

3. Prediction of Response Trajectories

Of interest in our example, in addition to estimation of the varying coefficient functions, is the prediction of calcium absorption trajectories as a function of age, based on dietary calcium (X^*) and body surface area (Z^*). More generally, prediction of an individual response trajectory Y^*, based on a new subject's sparse observations from the longitudinal predictor trajectories $X_1^*, \ldots, X_p^*$ and the cross-sectional predictors $Z_1^*, \ldots, Z_q^*$, is of interest. In this section, we provide consistent predictors of individual response trajectories and provide their asymptotic distribution for the construction of asymptotic pointwise confidence intervals.

From (2.1), the predicted response trajectory would be obtained through the conditional expectation

$$\begin{aligned} &E\{Y^*(t)|X_1^*(t), \ldots X_p^*(t), Z_1^*, \ldots, Z_q^*\} \\ &\quad = \mu_Y(t) + \sum_{r=1}^{p} \beta_r(t) \sum_{m=1}^{\infty} \xi_{rm}^* \phi_{rm}(t) + \sum_{g=1}^{q} \alpha_g(t) Z_g^{C^*}, \end{aligned} \tag{3.1}$$

where $\xi_{rm}^* = \int_0^T \{X_r^*(t) - \mu_{X_r}(t)\}\phi_{rm}(t)dt$ is the mth functional principal component score of X_r^*. To estimate the predicted trajectory in (3.1), we note that estimates of $\mu_Y(t)$, $\beta_r(t)$ and $\alpha_g(t)$ were given in the previous section. Also, the nonparametric functional principal component analysis step employed in the estimating the auto-covariance surface $\widehat{G}_{X_rX_r}$ yields estimators for $\phi_{rm}(t)$ and ρ_{rm}; details are in the Appendix A.1. Thus, the only term remaining in (3.1) that requires estimation is ξ_{rm}^*.

Estimation of $\xi_{rm}^* = \int_0^T \{X_r^*(t) - \mu_{X_r}(t)\}\phi_{rm}(t)dt$ is a challenging problem since the integral cannot be approximated feasibly from the sparse trajectory $X_r^*(t)$. However, estimation is feasible under a Gaussian framework, following the novel work of Yao, Müller, and Wang (2005b). More precisely, let $X_{rj}^* = X_r^*(T_j)$ be the jth measurement for the predictor function X_r^* and $\widetilde{X}_{rj}^* = X_{rj}^* + \varepsilon_{rj}^*$ be the observed noise-contaminated version of it at time T_j^*, for a random number of total measurements N^*, $j = 1, \ldots, N^*$. Let $\widetilde{X}_r^* = (\widetilde{X}_{r1}^*, \ldots, \widetilde{X}_{rN^*}^*)^{\mathrm{T}}$. Assume that the functional principal component scores ξ_{rm}^*, the measurement errors ε_{rj}^*, and Z_g^* are jointly Gaussian. Then the predicted ξ_{rm}^* is given as the best linear prediction conditional on the $(N^*p + q) \times 1$ observation vector $U^* = (\widetilde{X}_1^{*\mathrm{T}}, \ldots, \widetilde{X}_p^{*\mathrm{T}}, Z_1^*, \ldots, Z_q^*)$, N^*, and locations of the observations $T^* = (T_1^*, \ldots, T_{N^*}^*)^{\mathrm{T}}$, namely

$$\tilde{\xi}_{rm}^* = H_{rm}^{*\mathrm{T}} \Sigma_{U^*}^{-1} (U^* - \mu_U^*). \tag{3.2}$$

In (3.2) $\mu_U^* = (\mu_{X_1}^{*\mathrm{T}}, \ldots, \mu_{X_p}^{*\mathrm{T}}, \mu_{Z_1}, \ldots, \mu_{Z_q})$ is the $(N^*p+q)\times 1$ mean vector with $\mu_{X_r}^* = \{\mu_{X_r}(T_1^*), \ldots, \mu_{X_r}(T_{N^*}^*)\}^{\mathrm{T}}$,

$$H_{rm}^{*\mathrm{T}} = \left\{ \sum_{m'=1}^{\infty} \rho_{rm,1m'} \phi_{1m'}^{*\mathrm{T}}, \ldots, \sum_{m'=1}^{\infty} \rho_{rm,pm'} \phi_{pm'}^{*\mathrm{T}}, E(\xi_{rm} Z_1), \ldots, E(\xi_{rm} Z_q) \right\} \tag{3.3}$$

is the $(N^*p+q)\times 1$ covariance vector with $\rho_{rm,r'm'} = \mathrm{Cov}\,(\xi_{rm}, \xi_{r'm'})$ and $\phi_{rm}^* = \{\phi_{rm}(T_1^*), \ldots, \phi_{rm}(T_{N^*}^*)\}^{\mathrm{T}}$. The $(N^*p+q)\times(N^*p+q)$ covariance matrix Σ_{U^*} in (3.2) is

$$\Sigma_{U^*} = \left[\begin{array}{ccc|ccc} \widetilde{G}_{X_1X_1} & \ldots & \widetilde{G}_{X_1X_p} & \widetilde{G}_{X_1Z_1} & \ldots & \widetilde{G}_{X_1Z_q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{G}_{X_1X_p} & \ldots & \widetilde{G}_{X_pX_p} & \widetilde{G}_{X_pZ_1} & \ldots & \widetilde{G}_{X_pZ_q} \\ \hline \widetilde{G}_{X_1Z_1} & \ldots & \widetilde{G}_{X_pZ_1} & G_{Z_1Z_1} & \ldots & G_{Z_1Z_q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{G}_{X_1Z_q} & \ldots & \widetilde{G}_{X_pZ_q} & G_{Z_1Z_q} & \ldots & G_{Z_qZ_q} \end{array}\right],$$

where the $N^*\times 1$ vector $\widetilde{G}_{X_rZ_g} = \mathrm{Cov}\,(\widetilde{X}_r^*, Z_g^* \mid N^*, T^*)$, the scalars $G_{Z_gZ_{g'}}$ are as defined Section 2.3, and the $N^*\times N^*$ covariance matrix $\widetilde{G}_{X_rX_{r'}} = \mathrm{Cov}\,(\widetilde{X}_r^*, \widetilde{X}_{r'}^* \mid N^*, T^*)$ for $r \neq r'$, and $\widetilde{G}_{X_rX_r} = \mathrm{Cov}\,(\widetilde{X}_r^* \mid N^*, T^*)$ with the (j,ℓ)th entry $(\widetilde{G}_{X_rX_r})_{j,\ell} = G_{X_rX_r}(T_j^*, T_\ell^*) + \mathrm{var}(\varepsilon_r)\delta_{j\ell}$ where $\delta_{j\ell} = 1$ if $j = \ell$ and 0 if $j \neq \ell$.

Estimators of μ_U^*, ϕ_{rm}^*, $\widetilde{G}_{X_rX_r'}$, $\widetilde{G}_{X_rZ_g}$, $G_{Z_gZ_{g'}}$, and $\widetilde{G}_{X_rX_r}$, based on the entire data with $(\widehat{\widetilde{G}}_{X_rX_r})_{j,\ell} = \widehat{G}_{X_rX_r}(T_j^*, T_\ell^*) + \widehat{\mathrm{var}}(\varepsilon_r)\delta_{j\ell}$, are then substituted in (3.2) to obtain a plug-in estimator for ξ_{rm}^*. Explicit forms of the estimator of the variance of the measurement errors, $\widehat{\mathrm{var}}(\varepsilon_r)$, are given in Appendix A.1. The covariance $\rho_{rm,r'm'}$ can be estimated via $\int \widehat{G}_{X_rX_{r'}}(s,t)\hat{\phi}_{rm}(s)\hat{\phi}_{r'm'}(t)dsdt$, and $E(\xi_{rm}Z_g)$ can be estimated by $\int \widehat{G}_{X_rZ_g}(t)\hat{\phi}_{rm}(t)dt$ using estimates of $G_{X_rX_{r'}}$, $G_{X_rZ_g}$, and $\phi_{rm}(t)$. Finally, $\widehat{H}_{rm}^{*\mathrm{T}} = \{\sum_{m'=1}^{M_1}\hat{\rho}_{rm,1m'}\hat{\phi}_{1m'}^{*\mathrm{T}}, \ldots, \sum_{m'=1}^{M_p}\hat{\rho}_{rm,pm'}\hat{\phi}_{pm'}^{*\mathrm{T}}\}$, where the numbers $M_1, \ldots, M_p$ of included eigenfunctions can be chosen by one-curve-leave-out cross-validation, the Akaike information criterion (AIC) or the fraction of variance explained. This leads to $\hat{\xi}_{rm}^* = \widehat{H}_{rm}^{*\mathrm{T}}\widehat{\Sigma}_{U^*}^{-1}(U^* - \hat{\mu}_U^*)$. Hence, the predicted trajectories are

$$\widehat{Y}_{\mathcal{M}}^*(t) = \hat{\mu}_Y(t) + \sum_{r=1}^{p}\hat{\beta}_r(t)\sum_{m=1}^{M_r}\hat{\xi}_{rm}^*\hat{\phi}_{rm}(t) + \sum_{g=1}^{q}\hat{\alpha}_g(t)Z_g^{C^*}, \tag{3.4}$$

where $\mathcal{M} = \sum_{r=1}^{p} M_r$. We look to the consistency of the prediction $\widehat{Y}_{\mathcal{M}}^*(t)$ for the target trajectory $\widetilde{Y}^*(t) = \mu(t) + \sum_{r=1}^{p}\beta_r(t)\sum_{m=1}^{\infty}\tilde{\xi}_{rm}^*\phi_{rm}(t) + \sum_{g=1}^{q}\alpha_g(t)Z_g^{C^*}$.

Theorem 2. *Under Assumptions* (A) *and* (B) *in the Appendix, given* N^* *and* T^**, for all* $t \in [0,T]$*,* $\lim_{n\to\infty} \widehat{Y}^*_{\mathcal{M}}(t) = \widetilde{Y}^*(t)$*, in probability. Here the number* $M_r = M_r(n)$ *of eigen-components included in the eigen-decomposition of* X^*_r *for* $r = 1,\ldots,p$*, and hence* $\mathcal{M}$*, all tend to infinity as* $n \to \infty$*.*

Next we consider construction of asymptotic confidence bands for the response trajectory Y^*, given the observed sparse and noisy data. For $M_1,\ldots,M_p \geq 1$, let $\xi_r^{*^{M_r}} = (\xi^*_{r1},\ldots,\xi^*_{rM_r})^{\mathrm{T}}$ for $r = 1,\ldots,p$, and $\xi^{*^{\mathcal{M}}} = (\xi_1^{*^{M_1^{\mathrm{T}}}},\ldots,\xi_p^{*^{M_p^{\mathrm{T}}}})^{\mathrm{T}}$. Quantities $\tilde{\xi}_r^{*^{M_r}}$ and $\tilde{\xi}^{*^{\mathcal{M}}}$ are defined similarly. Under the Gaussian assumption, given N^* and T^*, $\tilde{\xi}^{*^{\mathcal{M}}} - \xi^{*^{\mathcal{M}}} \sim \mathrm{N}(0,\Omega_{\mathcal{M}})$, where the normality, covariance matrix $\Omega_{\mathcal{M}}$, and its plug-in estimator $\widehat{\Omega}_{\mathcal{M}}$ are derived in the proof of Theorem 3 in Appendix A.2. Define $\phi_{t\mathcal{M}} = \{\beta_1(t)\phi_{11}(t),\ldots,\beta_1(t)\phi_{1M_1}(t),\ldots,\beta_p(t)\phi_{p1}(t),\ldots,\beta_p(t)\phi_{pM_p}(t)\}^{\mathrm{T}}$ for $t \in [0,T]$, and let $\hat{\phi}_{t\mathcal{M}}$ be its estimate obtained from the data. We establish the asymptotic distribution of the predicted trajectories $\widehat{Y}^*_{\mathcal{M}}(t) = \hat{\mu}_Y(t) + \hat{\phi}^{\mathrm{T}}_{t\mathcal{M}}\hat{\xi}^{*^{\mathcal{M}}} + \sum_{g=1}^{q} \hat{\alpha}_g(t) Z_g^{C^*}$, conditional on N^* and T^*.

Theorem 3. *Under Assumptions* (A)*,* (B)*, and* (C) *in the Appendix, given* N^* *and* T^**, for all* $t \in [0,T]$*,* $x \in \mathbb{R}$*, the prediction for the response trajectory satisfies*

$$\lim_{n\to\infty} P\left[\frac{\widehat{Y}^*_{\mathcal{M}}(t) - E\{Y^*(t) \mid X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\}}{\hat{\omega}_{t\mathcal{M}}} \leq x\right] = \Phi(x),$$

where $\omega_{t\mathcal{M}} = \phi^{\mathrm{T}}_{t\mathcal{M}}\Omega_{\mathcal{M}}\phi_{t\mathcal{M}}$*,* $\hat{\omega}_{t\mathcal{M}} = \hat{\phi}^{\mathrm{T}}_{t\mathcal{M}}\widehat{\Omega}_{\mathcal{M}}\hat{\phi}_{t\mathcal{M}}$*, and* $\Phi(\cdot)$ *denotes the Gaussian cdf, and* M_r*,* $r = 1,\ldots,p$*, and hence* $\mathcal{M}$*, all tend to infinity as* $n \to \infty$*.*

Hence, ignoring bias resulting from truncation at $M_1,\ldots,M_p$ in $\widehat{Y}^*_{\mathcal{M}}$, the $(1-\alpha)100\%$ asymptotic pointwise confidence interval for $E\{Y^*(t) \mid X_1^*(t),\ldots,X_p^*(t), Z_1^*,\ldots,Z_q^*\}$ is given by $\widehat{Y}^*_{\mathcal{M}}(t) \pm \Phi(1-\alpha/2)\sqrt{\hat{\omega}_{t\mathcal{M}}}$.

4. Application to Sparse Dietary Calcium Absorption Data

In a study of calcium deficiency, Heaney et al. (1989) showed a complex inverse relation between calcium intake and calcium absorption, where age and body surface area are among the variables that affect calcium absorption efficiency. We examine the age-varying coefficient regression of calcium absorption on intake and body surface area via the analysis of data from a longitudinal study on factors affecting calcium absorption (Davis (2002, p.336)). Longitudinal measurements were taken on absorption and intake among others, where repeated measurement per subject were taken in roughly five-year intervals. We analyze the data where patient ages are between 39 and 58, yielding 182 subjects with 1 to 4 repeated measurements per subject. The data is sparse and irregular, due to measurement times between subjects that differed vastly, and the number of

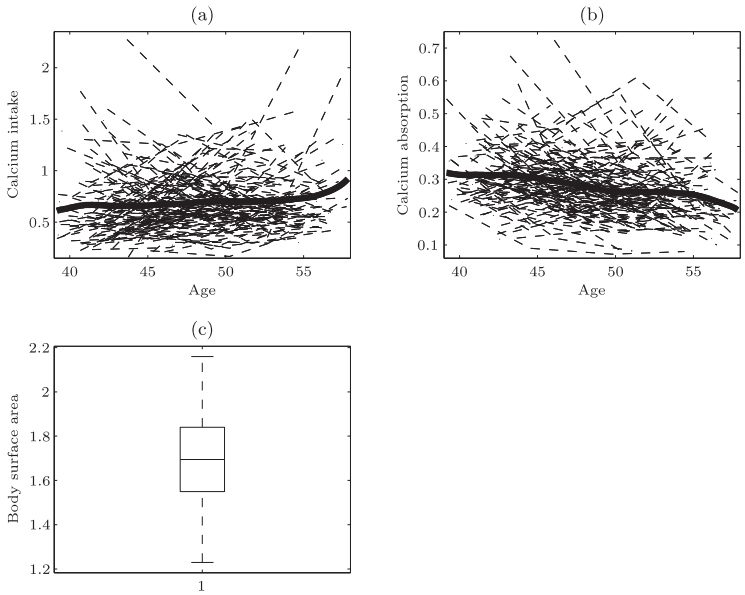

Figure 1. (a) Observed individual trajectories (dashed) and the smoothed estimate of the mean function $\hat{\mu}_X$ (thick solid) for calcium intake. (b) Observed individual trajectories (dashed) and the smoothed estimate $\hat{\mu}_Y$ of the mean function for calcium absorption (thick solid). (c) Boxplot of baseline body surface area values of the 182 female patients.

total repetitions per subject is small. Figure 1 displays the observed individual trajectories of calcium intake $X_1(\text{age})$ and absorption $Y(\text{age})$, along with the corresponding mean functions $\hat{\mu}_X$ and $\hat{\mu}_Y$. An increasing trend with age is observed for intake and a decreasing trend is observed for absorption.

We fit the age varying coefficient model

$$E\{Y(\text{age}) \mid X_1(\text{age}), Z_1\} = \beta_0(\text{age}) + \beta_1(\text{age})X_1(\text{age}) + \alpha_1(\text{age})Z_1$$

of Y (calcium absorption) on X_1 (calcium intake) and Z_1 (baseline body surface area; see Figure 1) using the proposed estimation procedure and kernel linear smoothing, as described in Section 2.2. The resulting estimated varying coefficient functions from both methods are displayed in Figure 2, along with 90% bootstrap percentile confidence intervals. Bootstrap confidence intervals were constructed from 500 bootstrap samples, generated by resampling subjects. Bandwidths for the smoothing of the cross-sectional mean functions, the covariance functions ($G_{X_1Z_1}$ and G_{YZ_1}), the covariance surface ($G_{X_1X_1}$), and the cross-covariance surface (G_{YX_1}) were selected by generalized cross-validation.

The estimated varying coefficient functions from both approaches, displayed in Figure 2, suggest a significant negative relationship between calcium intake

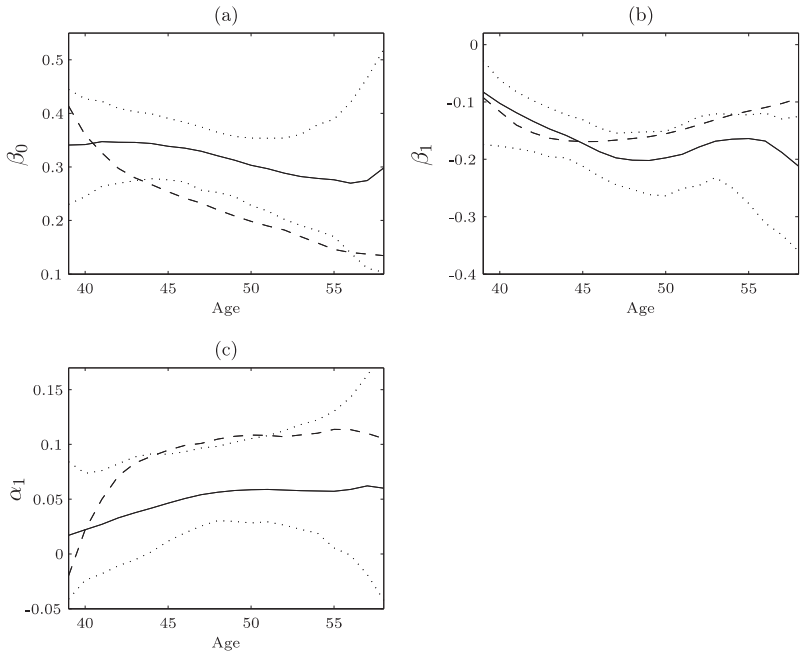

Figure 2. (a) Estimated varying coefficient function β_0(age) from the proposed varying coefficient model fit (solid) along with 90% bootstrap confidence intervals (dotted) for the calcium absorption data. Estimated functions from the varying coefficient model fits using kernel linear smoothing (dashed) are also displayed. (b) Estimated varying coefficient function β_1(age), the slope function of calcium intake, from both fits, along with 90% bootstrap confidence intervals. (c) Estimated varying coefficient function α_1(age), the slope function of the cross-sectional variable body surface area, from both fits, along with 90% bootstrap confidence intervals.

and absorption. While the inverse relationship between intake and absorption is declining with age (Figure 2b), especially after age 45 in the kernel linear fit, such a decline is not observed in the proposed fit. Both methods indicate a significant positive effect of baseline body surface area on absorption, for ages between 43 and 55. While the effect slowly becomes positive with age, in both methods, the effect estimated by the kernel linear fit is much larger in magnitude. These differences in the estimated varying coefficients for the kernel linear method are attributed to its estimation bias because of potential measurement error and for lack of efficiency since it does not incorporate the underlying correlation structure into the estimation. These issues are further investigated in the simulation studies of Section 5. The estimated positive and negative relations of calcium absorption with intake and with body surface area, respectively, are consistent with earlier findings (Heaney et al. (1989)).

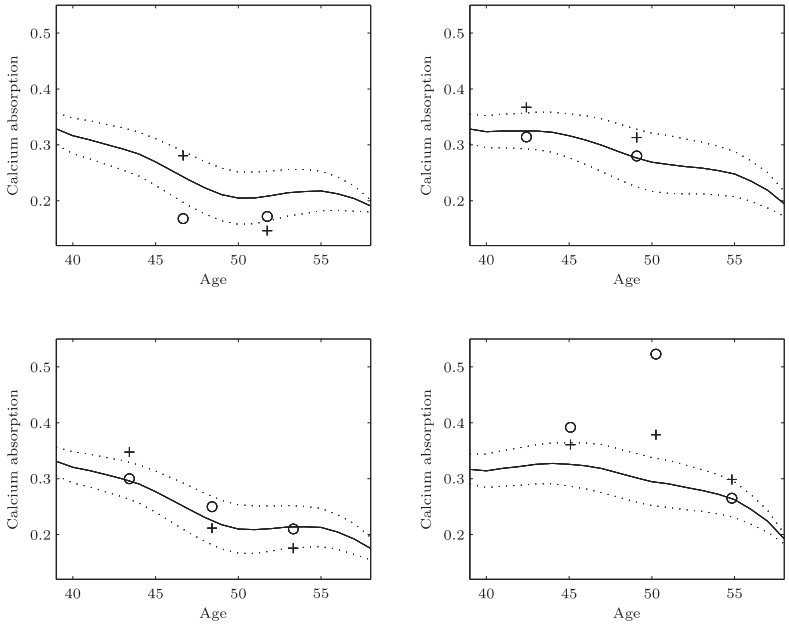

Figure 3. Observed values (circles) for calcium absorption (not used for prediction), predicted curves (solid) and 90% pointwise confidence bands (dotted), for four randomly selected patients, where bands and predicted curves are based on one-curve-leave-out analysis. Also displayed (+) are predicted values from the kernel linear smoothing estimation.

Next, we illustrate the prediction of calcium absorption trajectories, as described in Section 3, based on sparse longitudinal intake trajectories, along with baseline body surface area. The numbers of eigenfunctions used in the expansion of the predictor trajectories given in (3.4) were chosen by AIC; further details on these choices can be found in Yao, Müller, and Wang (2005a). The predicted trajectories for four randomly selected subjects are given in Figure 3. The trajectories show a decline in calcium absorption with age, which is the same pattern as observed in the estimated smooth mean function of Figure 1. Overlaying the predictions are 90% approximate confidence intervals, as proposed in Section 3, along with observed calcium absorption values and predictions obtained from kernel linear smoothing. Predictions and confidence intervals are obtained with the subject's predictor trajectory left out. The predicted values obtained from kernel linear smoothing are similar to those obtained from the proposed method; the average absolute prediction error from kernel linear smoothing and the proposed method are 0.0612 and 0.0643, respectively.

Note that even though the kernel linear smoothing estimation procedure cannot target the true underlying varying coefficient functions in the analysis

of sparse noise-contaminated longitudinal data, as will be demonstrated in the simulations of Section 5, it can produce predictions of the response values. This is a well-known phenomenon in nonparametric measurement error models, where in the nonparametric regression model $Y = g(X)+\epsilon$, the predictor X is measured with additive measurement error U yielding the observations $W = X + U$. Even though the nonparametric estimation of $g(\cdot)$ needs to adjust for the additive measurement error, the prediction of future response values can be obtained without adjusting for the additive measurement error via estimation of $E(Y \mid X+U) = E(Y \mid W)$. See Carroll et al. (2006), Carroll and Hall (1988), Stefanski and Carroll (1990), and Carroll, Delaigle, and Hall (2009) for further details and discussions on similar issues. Finally, we note an important distinction: although kernel linear smoothing and the proposed estimation methodology yield similar predictions, the proposed method has the distinct advantage of providing predicted response trajectories for the entire length of the study, while kernel linear smoothing and other methods can only provide pointwise predictions.

5. Simulation Studies

We assess the finite sample performance of the proposed estimation algorithm and compare its performance to that of kernel linear smoothing via three simulation studies. While the first set-up corresponds to highly sparse designs, the second reflects denser longitudinal designs; they both involve a varying coefficient model with one longitudinal and one cross-sectional covariate, similar to our data example. We also study a varying coefficient model with two longitudinal and two cross-sectional predictors in the third simulation for sparse longitudinal data. We report all results based on 500 Monte Carlo runs.

In the first study the number of measurements per subject was randomly chosen with equal probability from $\{1, 2, 3, 4\}$ for each of $n = 182$ subjects, similar to the calcium absorption data, to reflect sparse designs. The locations T_{ij} of the measurements for the i-th subject were generated uniformly from $[0, 10]$. The predictor process X was generated according to (SD) of Section 2.1 with mean function $\mu_X(t) = t + \sin(t)$, two eigenfunctions, $\phi_1(t) = \cos(\pi t/10)/\sqrt{5}$ and $\phi_2(t) = \sin(\pi t/10)/\sqrt{5}$, for $0 \le t \le 10$, and two eigenvalues, $\rho_1 = 2$ and $\rho_2 = 1$, respectively. The functional principal components ξ_{im} $(m = 1, 2)$ were generated from $\mathcal{N}(0, \rho_m)$, and the mean zero additive measurement error ε_{ij} were Gaussian with variance 0.2. The cross-sectional variable Z_1 was generated from $\mathcal{N}(0, 1)$; this is the marginal component from a bivariate normal distribution for (Z_{1i}, ξ_{i2}) with $\mathrm{Cov}\,(Z_1, \xi_2) = 0.3$, to allow for correlation between X_1 and Z_1. The response trajectories were generated from

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_{1i}(t) + \alpha_1(t)Z_{1i} + V_i(t), \tag{5.1}$$

according to (2.1), where $\beta_0(t) = 10\sin(\pi + t\pi/5)$, $\beta_1(t) = \sin(\pi t/10)$, $\alpha_1(t) = t/10$. The functional error V_i in (5.1) was constructed from the same two eigenfunctions as used for $X(t)$, with Gaussian functional principal components generated with eigenvalues $\rho_1 = 0.2$ and $\rho_2 = 0.1$. The observed measurements on the response were further contaminated with additive measurement errors according to $Y_{ij} = Y_i(T_{ij}) + \epsilon_{ij}$, the ϵ_{ij} i.i.d. zero mean Gaussian errors with variance 0.2. In the second simulation set-up, the variables were generated in the same way as in the first, except at still irregular but denser (non-sparse) measurement times, with the total number of repeated measurements generated uniformly from $\{5, \ldots, 15\}$. The average number of repeated measurements per subject was 10 for the dense case, compared to less than 3 observations per subject for the sparse case.

We compare the performance of the proposed estimation algorithm with the performance of kernel linear smoothing under the sparse and denser set-ups using mean absolute deviation error (MADE) and weighted average squared error (WASE),

$$\text{MADE} = \frac{1}{3T}\left[\sum_{r=0}^{1} \frac{\int |\beta_r(t) - \hat{\beta}_r(t)|dt}{\text{range}(\beta_r)} + \frac{\int |\alpha_1(t) - \hat{\alpha}_1(t)|dt}{\text{range}(\alpha_1)}\right], \text{ and}$$

$$\text{WASE} = \frac{1}{3T}\left[\sum_{r=0}^{1} \frac{\int \{\beta_r(t) - \hat{\beta}_r(t)\}^2 dt}{\text{range}^2(\beta_r)} + \frac{\int \{\alpha_1(t) - \hat{\alpha}_1(t)\}^2 dt}{\text{range}^2(\alpha_1)}\right],$$

where $T = 10$, $\text{range}(\beta_r)$ is the range of the function $\beta_r(t)$, and $\text{range}(\alpha_1)$ is defined similarly. We also consider the unweighted average squared error (UASE) to compare the estimators, where UASE is defined as is WASE, but without weights in the denominator. Bandwidths involved in smoothing of the mean functions and the auto- and cross-covariance surfaces were chosen by generalized cross-validation.

Results from sparse and denser simulation set-ups are given in Figures 4 and Figure 5, respectively. More specifically, plot (d) in both figures are boxplots of logarithms of the ratios of MADE, WASE, and UASE values of the proposed method over the kernel linear smoothing approach. The proposed estimators led to improved finite sample performance for both sparse and denser cases with respect to all three error criteria. More specifically, the proposed estimators had improved performance in $(85, 77, 71)\%$ of the Monte Carlo runs for sparse design according to (MADE, WASE, UASE) criteria respectively, while they led to improved performance in all Monte Carlo runs according to all three criteria in the case of denser design. This can be attributed to the fact that the proposed method adjusts for noise contaminated measurements and incorporates information inherent in the underlying correlation structure of the longitudinal processes.

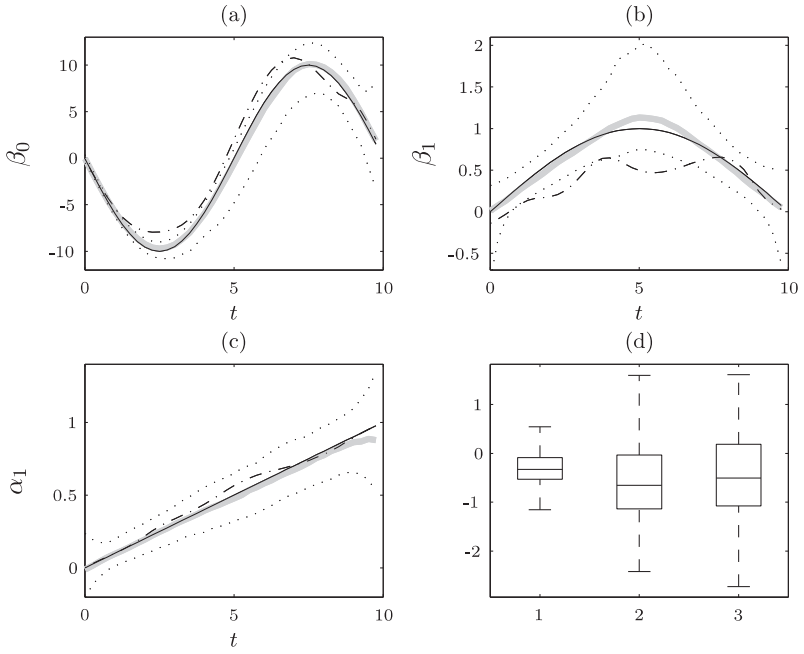

Figure 4. First simulation set-up given in (5.1) with highly sparse design: (a) The cross-sectional median curves of the proposed estimates (grey) along with 5% and 95% cross-sectional percentiles (dotted) overlaying the true varying coefficient function $\beta_0(t)$ (solid). Also displayed are the cross-sectional median curves from fits using kernel linear smoothing (dash-dotted). Similarly, for (b) $\beta_1(t)$ and (c) $\alpha_1(t)$. (d) Boxplots for the logarithm of the ratios of error measures (MADE, WASE and UASE) for proposed estimates over kernel linear smoothing. Values smaller than zero show that the proposed method is superior.

The estimated varying coefficient functions based on the proposed method and the kernel linear approach are provided in Figure 4 (a)-(c) and Figure 5 (a)-(c) for both simulation scenarios. Displayed are the cross-sectional medians of the estimated varying coefficient functions for the proposed method and the kernel linear method together with estimated functions corresponding to the 5% and 95% cross-sectional percentiles for the proposed method. The kernel linear smoothing fits deviate from the underlying true functions for both sparse and denser simulation set-ups. The bias is especially apparent in the estimation of $\beta_1(t)$ (e.g. see Figure 5(b)), and is also apparent in the estimation of $\beta_0(t)$. The (median) estimated functions of the proposed method target the corresponding true functions closely for both simulation scenarios, and for the dense data case note that they essentially coincide with the true functions. This is not surprising, since from the sparse to the dense case, there is an average of 4-fold increase in

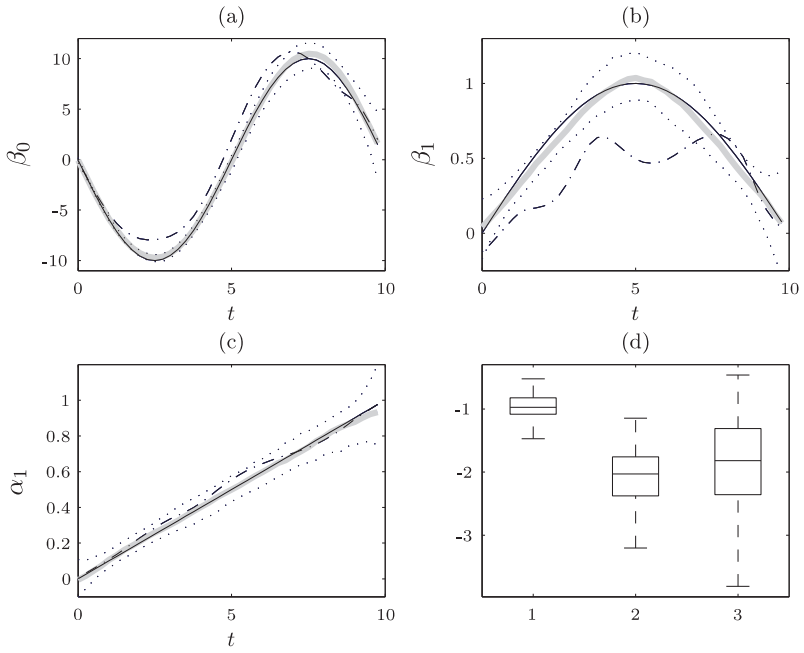

Figure 5. Second simulation set-up given in (5.1) with denser design: (a) The cross-sectional median curves of the proposed estimates (grey) along with 5% and 95% cross-sectional percentiles (dotted) overlaying the true varying coefficient function $\beta_0(t)$ (solid). Also displayed are the cross-sectional median curves from fits using kernel linear smoothing (dash-dotted). Similarly, for (b) $\beta_1(t)$ and (c) $\alpha_1(t)$. (d) Boxplots for the logarithm of the ratios of error measures (MADE, WASE and UASE) for proposed estimates over kernel linear smoothing. Values smaller than zero show that the proposed method is superior.

the number of repeated observations per subject. However, the bias of estimated functions via the kernel linear method remains, due to the measurement error.

In addition, we studied the coverage level of the proposed asymptotic confidence intervals for the predicted response trajectories given in Section 3 under the sparse set-up of the first simulation. Pointwise confidence intervals were constructed at a grid of time points at the 95% level, with coverage levels averaged over number of subjects in 100 Monte Carlo runs. The estimated coverage levels are given in Figure 6, for sample sizes $n = 182, 400$, and 1,000. A boundary effect is observed in the estimated coverage levels given over time, where the coverage level approaches the targeted 95% for the middle time range, and the region for the boundary effect gets smaller with increasing sample size. For example, excluding boundary regions (time 0-1 and 9-10), the coverage is between 83% and 96% for the time region 1-9 for $n = 1{,}000$.

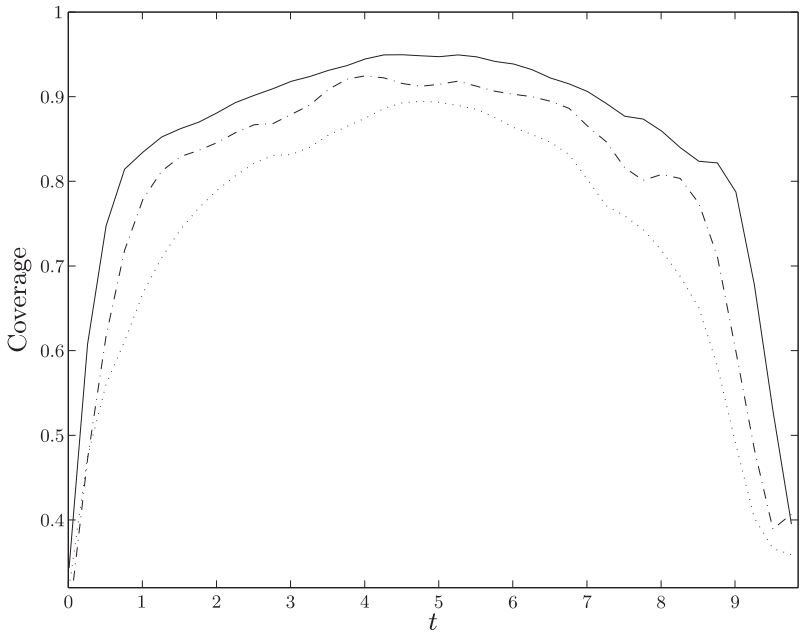

Figure 6. Estimated coverage levels for the 95% asymptotic confidence intervals of the predicted response trajectory proposed in Section 3 under the sparse design of the first simulation for $n = 182$ (dotted), $n = 400$ (dash-dotted) and $n = 1{,}000$ (solid).

For the third simulation, the number of measurements per subject were randomly chosen with equal probability from $\{4, 5, 6, 7, 8\}$ for each of $n = 400$ subjects; the locations T_{ij} of the measurements for the i-th subject were generated uniformly from $[0, 10]$. The first longitudinal predictor process X_1 was generated with the same mean function and eigenbasis as in the first two simulations, with $\rho_1 = 1$ and $\rho_2 = 1$, while the second longitudinal predictor X_2 was generated with mean function $\mu_{X_2}(t) = -(t-5)^2/2$, two basis functions, $\phi_1(t) = (t-5)/5$ and $\phi_2(t) = 3((t-5)/5)^2 - 1)/2$, for $0 \le t \le 10$, and with mean zero variance 1 Gaussian coefficients. The mean zero additive measurement error ε_{rij} was taken to be Gaussian with variance 0.2 for both predictor processes. In order to allow for correlations between the two longitudinal and two cross-sectional predictors, the two cross-sectional variables Z_1 and Z_2 were generated from Gaussian distributions with means 1 and 2, variances 2 and 2, respectively; these are marginal components from a six-dimensional multivariate normal vector containing the two random coefficients of the two longitudinal predictors and the two cross-sectional

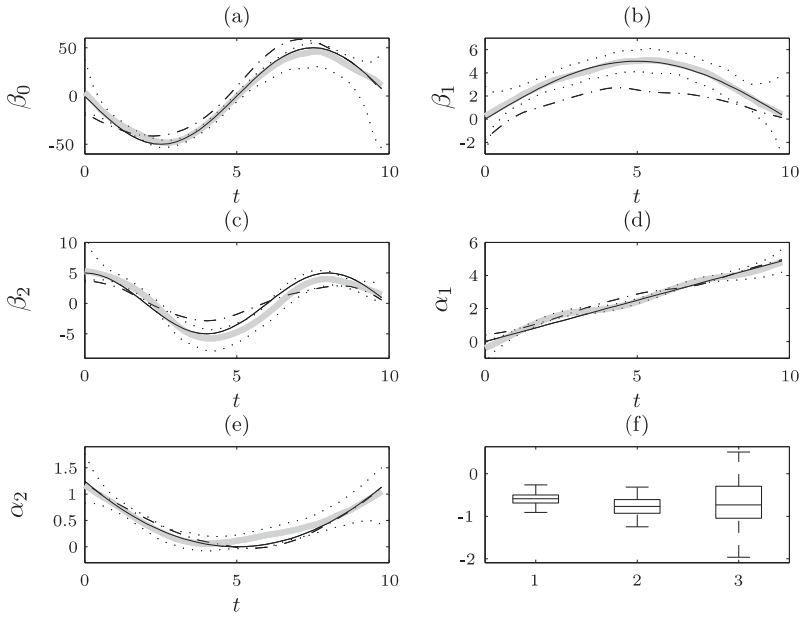

Figure 7. Third simulation set-up given in (5.2): (a) The cross-sectional median curves of the proposed estimates (grey) along with 5% and 95% cross-sectional percentiles (dotted) overlaying the true varying coefficient function $\beta_0(t)$ (solid). Also displayed are the cross-sectional median curves from fits using kernel linear smoothing (dash-dotted). Similarly, for (b) $\beta_1(t)$ (c) β_2 (d) $\alpha_1(t)$ (e) $\alpha_2(t)$. (f) Boxplots for the logarithm of the ratios of error measures (MADE, WASE and UASE) for proposed estimates over kernel linear smoothing. Values smaller than zero show that the proposed method is superior.

predictors, i.e. $(\xi_{1i1}, \xi_{1i2}, \xi_{2i1}, \xi_{2i2}, Z_{1i}, Z_{g2})$. The 6×6 covariance matrix was

$$\begin{bmatrix} 1 & 0 & 0.2 & 0 & 0.2 & 0.3 \\ 0 & 1 & 0 & 0.2 & 0.4 & 0 \\ 0.2 & 0 & 1 & 0 & 0 & 0.1 \\ 0 & 0.2 & 0 & 1 & 0.5 & 0 \\ 0.2 & 0.4 & 0 & 0.5 & 2 & 0.2 \\ 0.3 & 0 & 0.1 & 0 & 0.2 & 2 \end{bmatrix}.$$

The response trajectories were generated from

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_{1i}(t) + \beta_2(t)X_{2i}(t) + \alpha_1(t)Z_{1i} + \alpha_2(t)Z_{2i} + V_i(t), \quad (5.2)$$

according to (2.1), where $\beta_0(t) = 50\sin(\pi + t\pi/5)$, $\beta_1(t) = 5\sin(\pi t/10)$, $\beta_2(t) = 5\cos(\pi t/4)$, $\alpha_1(t) = t/2$, and $\alpha_2(t) = (t-5)^2/20$. The functional error V_i in (5.2)

was constructed from the same two eigenfunctions as used for $X_1(t)$, with Gaussian functional principal components generated with eigenvalues $\rho_1 = 0.2$ and $\rho_2 = 0.1$. The additive measurement error on the response was generated as zero mean Gaussian with variance 0.2. Boxplots of logarithms of the ratios of MADE, WASE, and UASE values of the proposed method over the kernel linear smoothing approach, along with cross-sectional medians and 5% and 95% percentiles of the estimated varying coefficient functions based on the proposed method and the medians of the kernel linear approach, are provided in Figure 7. The proposed method performs better than the kernel linear smoothing according to all three criterion in 95% of the total Monte Carlo runs. The superior performance of the proposed method can be seen especially in the estimated β_0 and the two slope varying coefficient functions β_1 and β_2 corresponding to longitudinal predictors measured with additive measurement error. The estimated median curves for the kernel linear smoother deviate from the true curves, not being able to handle measurement error in covariates.

6. Discussion

In this work we propose a multiple varying coefficient model in the context of highly sparse longitudinal data, where the longitudinal response and predictor processes are measured with error. Our estimation algorithm is applicable to cases with possibly different time grids for each longitudinal covariate of a subject, whereas most longitudinal regression methods are not. In our estimation procedure, observed measurements with missing pairs also contribute to the estimation due to the unique representation of the varying coefficient functions and the component-wise nature of the algorithm. This could lead to favorable estimation properties in case of missing covariates, especially for sparse designs where imputation via smoothing techniques are not feasible.

Acknowledgement

We are extremely grateful to two anonymous referees, an associate editor and the Editor for helpful remarks that improved the paper. Support for this work includes the National Institute of Health grants UL1RR024922, RL1AG032119, and RL1AG032115 and grant UL1 RR024146 from the National Center for Research Resources.

Appendix

A.1. Details on estimation procedures

Explicit forms of the proposed mean and covariance estimators, functional principal components decompositions, and measurement error variance estimators are given as follows. The local linear scatterplot smoother for $\mu_{X_r}(t)$ is

obtained by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\left(\frac{T_{ij}-t}{b_{X_r}}\right)\{X_{rij}-\eta_0-\eta_1(t-T_{ij})\}^2,$$

with respect to η_0 and η_1, leading to $\hat{\mu}_X(t)=\hat{\eta}_0$. Other one-dimensional smoothing estimators of the proposed estimation algorithm, namely $\hat{\mu}_Y(t)$, $\widehat{G}_{X_rZ_g}$ and $\widehat{G}_{YZ_g}$ can be obtained similarly.

For the two-dimensional smoothers, recall $G_{X_rX_r',i}(T_{ij},T_{i\ell})=\{X_{rij}-\hat{\mu}_{X_r}(T_{ij})\}\{X_{r'i\ell}-\hat{\mu}_{X_{r'}}(T_{i\ell})\}$, and take the local linear surface smoother for $G_{X_rX_{r'}}(s,t)$ to minimize

$$\sum_{i=1}^{n}\sum_{1\le j,\ell\le N_i} K_2\left(\frac{T_{ij}-s}{h_{X_r}},\frac{T_{i\ell}-t}{h_{X_{r'}}}\right)[G_{X_rX_{r'},i}(T_{ij},T_{i\ell})-f\{\eta,(s,t),(T_{ij},T_{i\ell})\}]^2, \tag{5.3}$$

where $f\{\eta,(s,t),(T_{ij},T_{i\ell})\}=\eta_0+\eta_1(s-T_{ij})+\eta_2(t-T_{i\ell})$, with respect to $\eta=(\eta_0,\eta_1,\eta_2)$, yielding $\widehat{G}_{X_rX_{r'}}(s,t)=\hat{\eta}_0$. The two-dimensional smoother in the estimation of G_{YX_r} is found similarly.

For the estimation of the auto-covariance $G_{X_rX_r}$, the second sum in (5.3) is taken over $1\le j\ne\ell\le N_i$ to leave the noise contaminated diagonal raw covariance elements out of the smoothing procedure. For the eigen-decomposition of the auto-covariance surface $G_{X_rX_r}$, the eigenquations, $\int_0^T \widehat{G}^*_{X_rX_r}(s,t)\hat{\phi}_{rm}(s)ds=\hat{\rho}_{rm}\hat{\phi}_{rm}(t)$ are solved under orthonormal constraints on the eigenfunctions, where $\widehat{G}^*_{X_rX_r}$ is the smooth estimator of the covariance function. To arrive at the auto-covariance estimator, we exclude the negatively estimated eigenvalues and corresponding eigenfunctions in the functional principal component decomposition of the covariance function, i.e., $\widehat{G}_{X_rX_r}=\sum_{m:\hat{\rho}_{rm}>0}^{M_r}\hat{\rho}_{rm}\hat{\phi}_{rm}(s)\hat{\phi}_{rm}(t)$.

In the original smoothing estimator of the auto-covariance surface leading to $\widehat{G}^*_{X_rX_r}$, for estimation of $\mathrm{var}(\varepsilon_r)$, a local quadratic component is fit orthogonal to the diagonal of $G_{X_rX_r}$ and a local linear component is fit in the direction of the diagonal, resulting in a surface estimate; the diagonal is denoted by $G_r(s)$. In addition, a separate local linear smoother is fit only to the diagonal values $\{G_{X_rX_r}(t,t)+\mathrm{var}(\varepsilon_r)\}$, denoted by $\widehat{V}_X(t)$. The estimator of $\mathrm{var}(\varepsilon_r)$ is taken as the difference between the above two smoothing estimators for the diagonal terms, $\widehat{\mathrm{var}}(\varepsilon_r)=(2/T)\int_{T/4}^{3T/4}\{\widehat{V}(s)-G_r(t)\}dt$ if $\mathrm{var}(\varepsilon_r)>0$, and $\mathrm{var}(\varepsilon_r)=0$ otherwise.

A.2. Assumptions and proofs

Assumptions (A1–A6) are needed for all three theorems, (B1–B2) are needed for Theorem 2 and Theorem 3, while (C) is only needed for Theorem 3.

(A1) The cross-sectional predictors Z_{gi} are iid for $i = 1, \ldots, n$, with $\text{var}(Z_{gi}) > 0$ for $g = 1, \ldots, q$.

(A2) The covariance matrices $\mathcal{X}_t$ defined in (2.4) are nonsingular for $t \in [0, T]$.

The longitudinal predictor and response trajectories (T_{ij}, X_{rij}) and (T_{ij}, Y_{ij}), $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, $r = 1, \ldots, p$, are assumed to have the same distribution as $(\mathcal{T}, X_r)$ and $(\mathcal{T}, Y)$, with joint densities $g_r(t, x)$ and $h(t, y)$. The observation times T_{ij} are i.i.d. with density $f_{\mathcal{T}}(t)$. Let T_1 and T_2 be i.i.d. T and X_{r1} and X_{r2} be the repeated measurements of X_r made on the same subject at times T_1 and T_2, and assume $(T_{ij}, T_{i\ell}, X_{rij}, X_{ri\ell})$, $1 \leq j \neq \ell \leq N_i$, is distributed as $(T_1, T_2, X_{r1}, X_{r2})$, with joint density function $g_{X_r X_r}(t_1, t_2, x_1, x_2)$. It is analogously assumed that the response measurements $(T_{ij}, T_{i\ell}, Y_{ij}, Y_{i\ell})$, $1 \leq j \neq \ell \leq N_i$, are identically distributed with joint density function $g_{YY}(t_1, t_2, y_1, y_2)$. The following regularity conditions are assumed on $f_{\mathcal{T}}(t)$, $g_r(t, x)$, $h(t, y)$, $g_{X_r X_r}(t_1, t_2, x_1, x_2)$, and $g_{YY}(t_1, t_2, y_1, y_2)$.

(A3) Let p_1 and p_2 be integers with $0 \leq p_1, p_2 \leq p = p_1 + p_2 = 2$. The derivative $(dp/dt^p) f_{\mathcal{T}}(t)$ exists and is continuous on $[0, T]$ with $f_{\mathcal{T}}(t) > 0$ on $[0, T]$, $(dp/dt^p) g_r(t, x)$ and $(dp/dt^p) h(t, y)$ exist and are continuous on $[0, T] \times \mathbb{R}$, and $\{dp/(dt_1^{p_1} dt_2^{p_2})\} g_{X_r X_r}(t_1, t_2, x_1, x_2)$ and $\{dp/(dt_1^{p_1} dt_2^{p_2})\} g_{YY}(t_1, t_2, y_1, y_2)$ exist and are continuous on $[0, T]^2 \times \mathbb{R}^2$.

(A4) The numbers of measurements made on subjects are i.i.d. N, where N is discrete with $P(N > 1) > 0$. Observation times and measurements are independent of the number of observations for any subset $J_i \in \{1, \ldots, N_i\}$ and for all $i = 1, \ldots, n$.

Let $K_1(\cdot)$ be the nonnegative, mean zero, finite variance, compactly supported kernel function used in estimating μ_{X_r}, μ_Y, G_{YZ_g}, and $G_{X_r Z_g}$, and $K_2(\cdot, \cdot)$ be the bivariate kernel function with similar properties used in estimating the covariance surfaces $G_{X_r X_{r'}}$ and $G_{Y X_r}$. Explicit forms for the estimators of these quantities are given in Appendix A.3.

(A5) The Fourier transform $\kappa_1(t) = \int e^{-iut} K_1(u) du$ of $K_1(u)$ and $\kappa_2(t, s) = \int e^{-(iut + ivs)} K_2(u, v) du dv$ of $K_2(u, v)$ are absolutely integrable.

Let b_{X_r}, b_Y be the bandwidths used for estimating $\hat{\mu}_{X_r}$ and $\hat{\mu}_Y$, $(h_{X_r}, h_{X_{r'}})$ be the bandwidths for estimating $\widehat{G}_{X_r X_{r'}}$, (h_{r1}, h_{r2}) be the bandwidths for obtaining $\widehat{G}_{Y X_r}$, h_g for obtaining $\widehat{G}_{Y Z_g}$, and h_{rg} for $\widehat{G}_{X_r Z_{g'}}$, where all bandwidths depend on n.

(A6) As $n \to \infty$, $b_{X_r} \to 0$, $b_Y \to 0$, $h_g \to \infty$ and $h_{rg} \to \infty$, $n b_{X_r}^4 \to \infty$, $n b_Y^4 \to \infty$, $n h_g^4 \to \infty$, $n h_{rg}^4 \to \infty$ and $n b_{X_r}^6 < \infty$, $n b_Y^6 < \infty$, $n h_g^6 < \infty$ and $n h_{rg}^6 < \infty$. Without loss of generality $h_{X_r}/h_{X_{r'}} \to 1$, $h_{r1}/h_{r2} \to 1$ and $n h_{X_r}^6 \to \infty$, $n h_{r1}^6 \to \infty$ and $n h_{X_r}^8 < \infty$ and $n h_{r1}^8 < \infty$.

(A7) The fourth moments of Y and X, centered at $\mu_Y(t)$ and $\mu_X(t)$, are finite.
(A8) The number of included eigenfunctions in (3.4), $M_1, \ldots, M_p$, are integer-valued sequences that depend on sample size n such that $\inf_{t\in[0,T]} M_r(n) \to \infty$ and both $\inf_{t\in[0,T]} M_r(n)$ and $\sup_{t\in[0,T]} M_r(n)$ satisfy the rate conditions given in Assumption $(B5)$ of Yao, Müller, and Wang (2005a).
(A9) The autocovariance operator A_{G_r} generated by the continuously differentiable covariance function $G_{X_rX_r}(s,t)$ is positive definite.
(B1) The number and locations of the measurements for a subject or cluster remain unaltered as the sample size $n \to \infty$.
(B2) For all $1 \le i \le n$, $m \ge 1$, $1 \le r \le p$, $1 \le g \le q$, and $1 < j < N_i$, the functional principal component scores ξ_{rim}, the cross-sectional predictors Z_{gi}, and the measurement errors ϵ_{rij} in (SD) are jointly Gaussian.
(C) There exists a continuous positive definite function ω_t such that $\omega_{t\mathcal{M}}$, as defined in Theorem 3, satisfies $\omega_{t\mathcal{M}} \to \omega_t$ as $M_1, \ldots, M_p \to \infty$.

Proof of Theorem 1. Uniform consistency of $\hat{\mu}_{X_r}(t)$ and $\hat{\mu}_Y(t)$ follow from Theorem 1 of Yao, Müller, and Wang (2005b), and that of $\widehat{G}_{X_rZ_g}(t)$ and $\widehat{G}_{YZ_g}(t)$ can be shown similarly for $r = 1, \ldots, p$, $g = 1, \ldots, q$. The properties of A_{G_r} in (A9) imply that the ρ_{rm} are all positive. Hence $\sum_{m:\hat{\rho}_{rm}>0}^{M_r} \hat{\rho}_{rm}\hat{\phi}_{rm}(s)\hat{\phi}_{rm}(t)$ and $\sum_m^{M_r} \hat{\rho}_{rm}\hat{\phi}_{rm}(s)\hat{\phi}_{rm}(t)$ are asymptotically equivalent since $|\hat{\rho}_{rm} - \rho_{rm}| = O_p\{1/(\sqrt{n}h_{X_r}^2)\}$, by Theorem 2 of Yao, Müller, and Wang (2005b). The uniform consistency of $\widehat{G}_{X_rX_r}(s,t)$ follows from uniform consistency of the eigenvalue and eigenfunction estimators shown in Theorem 2 of Yao, Müller, and Wang (2005b). For the rate conditions on M_r, we refer the reader to assumption $(B5)$ of Yao, Müller, and Wang (2005b), and note that further details on theoretical properties of functional principal component analysis can be found in Silverman (1996), Hall and Hosseini-Nasab (2009), and Hall, Müller, and Wang (2006). Uniform consistency of the cross-covariance estimators $\widehat{G}_{X_rX_{r'}}(s,t)$ and $\widehat{G}_{YX_r}(s,t)$ follow from Lemma $A1$ of Yao, Müller, and Wang (2005a). Combining these results implies uniform consistency of $\hat{\mathcal{X}}_t$ and $\hat{\Xi}_t$, and Theorem 1 follows.

Proof of Theorem 2. For fixed $M_1, \ldots, M_p$ and $\mathcal{M}$, let $\widetilde{Y}^*_{\mathcal{M}}(t) = \mu_Y(t) + \sum_{r=1}^p \beta_r(t) \sum_{m=1}^{M_r} \tilde{\xi}^*_{rm}\phi_{rm}(t) + \sum_{g=1}^q \alpha_g(t)Z_g^{C^*}$, and recall that $Y^*(t) = \mu_Y(t) + \sum_{r=1}^p \beta_r(t) \sum_{m=1}^{\infty} \tilde{\xi}^*_{rm}\phi_{rm}(t) + \sum_{g=1}^q \alpha_g(t)Z_g^{C^*}$. Since $|\widehat{Y}^*_{\mathcal{M}}(t) - \widetilde{Y}^*(t)| \le |\widehat{Y}^*_{\mathcal{M}}(t) - \widetilde{Y}^*_{\mathcal{M}}(t)| + |\widetilde{Y}^*_{\mathcal{M}}(t) - \widetilde{Y}^*(t)|$, it follows, as in Lemma 3 of Yao, Müller, and Wang (2005b), that $\widetilde{Y}^*_{\mathcal{M}}(t) \xrightarrow{p} \widetilde{Y}^*(t)$ as $M_1, \ldots, M_p \to \infty$ and $n \to \infty$. The uniform consistency of $\hat{\mu}_Y(t)$ follows from Theorem 1 of Yao, Müller, and Wang (2005b). Hence, using Theorem 1 of Section 2.4, Theorem 3, (17) of Yao, Müller, and Wang (2005b) and Slutsky's Theorem, it follows that $|\hat{Y}^*_{\mathcal{M}}(t) - \widetilde{Y}^*_{\mathcal{M}}(t)| \to 0$ as $n \to \infty$ and Theorem 2 follows.

Proof of Theorem 3. Write $E_{\mathcal{M}}\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\} = \mu_Y(t) + \sum_{r=1}^p \beta_r(t) \sum_{m=1}^{M_r} \xi_{rm}^* \phi_{rm}(t) + \sum_{g=1}^q \alpha_g(t) Z_g^{C^*}$, and note that

$$\begin{aligned}&\widehat{Y}_{\mathcal{M}}^*(t) - E_{\mathcal{M}}\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\}\\ &\quad = \widehat{Y}_{\mathcal{M}}^*(t) - \widetilde{Y}_{\mathcal{M}}^*(t) + \widetilde{Y}_{\mathcal{M}}^*(t) - E_{\mathcal{M}}\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\}.\end{aligned}$$

It follows from the proof of Theorem 2 that $\lim_{n\to\infty} \sup_{t\in[0,T]} |\widehat{Y}_{\mathcal{M}}^*(t) - \widetilde{Y}_{\mathcal{M}}^*(t)| \to 0$.

Let the $\mathcal{M}\times(N^*p+q)$ matrix $H = \mathrm{Cov}\,(\xi^{*^{\mathcal{M}}}, U^* \mid N^*, T^*) = [H_{11}^*,\ldots,H_{1M_1}^*, \ldots,H_{p1}^*,\ldots,H_{pM_r}^*]^{\mathrm{T}}$, where $H_{rm}^{*^{\mathrm{T}}}$ is as defined in (3.3). Since $\tilde{\xi}^{*^{\mathcal{M}}} = H\Sigma_{U^*}^{-1}(U^* - \mu_U^*)$, $\mathrm{Cov}\,(\tilde{\xi}^{*^{\mathcal{M}}}|N^*,T^*) = \mathrm{Cov}\,(\tilde{\xi}^{*^{\mathcal{M}}}, \xi^{*^{\mathcal{M}}}|\ N^*,T^*) = H\Sigma_{U^*}^{-1}H^{\mathrm{T}}$. Hence, $\mathrm{Cov}\,(\tilde{\xi}^{*^{\mathcal{M}}} - \xi^{*^{\mathcal{M}}}|N^*,T^*) = \mathrm{Cov}\,(\xi^{*^{\mathcal{M}}}|N^*,T^*) - \mathrm{Cov}\,(\tilde{\xi}^{*^{\mathcal{M}}}|N^*,T^*) = D - H\Sigma_{U^*}^{-1}H^{\mathrm{T}} \equiv \Omega_{\mathcal{M}}$, where $D = \mathrm{Cov}\,(\xi^{*^{\mathcal{M}}}|N^*,T^*)$ is the $\mathcal{M}\times\mathcal{M}$ matrix with (r,r')th partition, an $M_r \times M_{r'}$ matrix $(D)_{rr'} = D_{rr'} = \mathrm{Cov}\,(\xi_r^{*^{M_r}}, \xi_{r'}^{*^{M_{r'}}}|N^*,T^*)$. Let $\widehat{\Omega}_{\mathcal{M}} = \widehat{D} - \widehat{H}\widehat{\Sigma}_{U^*}^{-1}\widehat{H}^{\mathrm{T}}$, where $\widehat{D}$, $\widehat{H} = (\widehat{H}_{11}^*,\ldots,\widehat{H}_{1M_1}^*,\ldots,\widehat{H}_{p1}^*,\ldots,\widehat{H}_{pM_r}^*)^{\mathrm{T}}$, and $\widehat{\Sigma}_{U^*}$ are estimated based on the entire data with $\widehat{H}_{rm}^{*^{\mathrm{T}}}$, $\widehat{E}(\xi_{rm}Z_g)$ and $\hat{\xi}_{rm,r'm'}$ as defined in Section 3. It follows that, under the Gaussian assumption for a fixed $M_1,\ldots,M_p \geq 1$, $\tilde{\xi}^{*^{\mathcal{M}}} - \xi^{*^{\mathcal{M}}} \sim \mathcal{N}(0,\Omega_{\mathcal{M}})$. Hence,

$$\widehat{Y}_{\mathcal{M}}^*(t) - E_{\mathcal{M}}\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\} \xrightarrow{D} Z_{\mathcal{M}} \sim \mathcal{N}(0,\omega_{t\mathcal{M}}).$$

Under Assumption (C), letting $M_1,\ldots,M_p \to \infty$ leads to $Z_{\mathcal{M}} \xrightarrow{D} Z \sim \mathcal{N}(0,\omega_t)$. From the Karhunen-Loéve Theorem, we have

$$\begin{aligned}&|\mathrm{E}_{\mathcal{M}}\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\}\\ &\qquad -E\{Y^*(t)|X_1^*(t),\ldots,X_p^*(t),Z_1^*,\ldots,Z_q^*\}| \xrightarrow{P} 0\end{aligned}$$

as $M_1,\ldots,M_p \to \infty$. Hence, $\lim_{M_1,\ldots,M_p\to\infty} \lim_{n\to\infty}[\widehat{Y}_{\mathcal{M}}^*(t) - E\{Y^*(t)|X_1^*(t),\ldots, X_p^*(t),Z_1^*,\ldots,Z_q^*\}] \stackrel{D}{=} Z$. From Theorem 1 of Section 2.4 and Lemma 1 of Yao, Müller, and Wang (2005a), it follows that $\lim_{M_1,\ldots,M_p\to\infty} \lim_{n\to\infty} \hat{\omega}_{t\mathcal{M}} = \omega_t$. Theorem 3 follows by Slutsky's Theorem.

References

Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press, New York.

Carroll, R. J., Delaigle, A. and Hall, P. (2009). Nonparametric prediction in measurement error models. *J. Amer. Statist. Assoc.* **104**, 993-1003.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd edition. Chapman and Hall CRC Press, Baco Raton.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* **96**, 605-619.

Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. In *Statistical Models in S* (Edited by J. M. Chambers and T. J. Hastie), 309-376. Wadsworth & Brooks, Pacific Grove.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715-731.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface* **1**, 179-195.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *J. Roy. Statist. Soc. Ser. B* **68**, 109-126.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493-1517.

Hastie, T. and Tibshirani, R. (1993). Varying coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

Heaney, R. P., Recker, R. R., Stegman, M. R. and Moy, A. J. (1989). Calcium absorption in women: relationships to calcium intake, estrogen status, age. *J. Bone and Mineral Research* **4**, 469-475.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111-128.

Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14**, 763-788.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying coefficient models with longitudinal data. *Biometrics* **62**, 379-391.

Şentürk, D. and Müller, H. G. (2010). Functional varying coefficient models for longitudinal data. *J. Amer. Statist. Assoc.* **105**, 1256-1264.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24**, 1-24.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* 21, 165-184.

Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10,** 433-456.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388-1402.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **3**, 2873-2903.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Department of Biostatistics, University of California, Los Angeles, California 90095, U.S.A.
E-mail: dsenturk@ucla.edu

Division of Biostatistics, University of California, Davis, California 95616, U.S.A.
E-mail: ucdnguyen@ucdavis.edu

(Received December 2009; accepted August 2010)