Functional varying coefficient models for longitudinal data

By DAMLA ŞENTÜRK

Department of Statistics, Penn State University, University Park, PA 16802, USA

dsenturk@stat.psu.edu

And HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, CA 95616, USA

mueller@wald.ucdavis.edu

May 2010

Author's Footnote:

Damla Şentürk is Assistant Professor, Department of Statistics, Pennsylvania State University, University Park, PA 16801, U.S.A, email: `dsenturk@stat.psu.edu`. Hans-Georg Müller is Professor, Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, U.S.A, email: `mueller@wald.ucdavis.edu`.

This research was supported in part by a grant from the National Science Foundation (DMS-08-06199). We are grateful to two referees and an Associate Editor for their constructive comments and careful reading.

Summary

The proposed functional varying coefficient model provides a versatile and flexible analysis tool for relating longitudinal responses to longitudinal predictors. Specifically, this approach provides a novel representation of varying coefficient functions through suitable auto- and cross-covariances of the underlying stochastic processes, which is particularly advantageous for sparse and irregular designs, as often encountered in longitudinal studies. Existing methodology for varying coefficient models is not adapted to such data. The proposed approach extends the customary varying coefficient models to a more general setting, in which not only current but also recent past values of the predictor time course may have an impact on the current value of the response time course. The influence of past predictor values is modeled by a smooth history index function, while the effects on the response are described by smooth varying coefficient functions. The resulting estimators for varying coefficient and history index functions are shown to be asymptotically consistent for sparse designs. In addition, prediction of unobserved response trajectories from sparse measurements on a predictor trajectory is obtained, along with asymptotic pointwise confidence bands. The proposed methods perform well in simulations, especially when compared with commonly used local polynomial smoothing methods for varying coefficient models, and are illustrated with longitudinal primary biliary liver cirrhosis data. The data application and detailed assumptions and proofs can be found in online Supplemental Material.

Some key words: Functional data analysis; History index; Local least squares; Repeated measurements; Smoothing; Sparse design.

1. Introduction

In this paper we propose two innovations for varying coefficient models (Cleveland, Grosse and Shyu, 1991; Hastie and Tibshirani, 1993) in longitudinal studies. First, a new representation for varying coefficient functions is introduced that relates a response process to a predictor process. This representation is particularly advantageous when one has only noisy and intermittent measurements available for the trajectories of these processes for a sample of subjects, a common situation in longitudinal studies. Second, we extend the standard framework of varying coefficient models, where the current value of a response process is modeled in dependence on the current value of a predictor process, by including the effect of recent past values of the predictor through a smooth history index function. While in functional linear models with both predictors and responses as random functions (Ramsay and Dalzell, 1991; Yao, Müller and Wang, 2005b), it is assumed that past, present and future values of the predictor process influence current response, only the current value of the predictor process affects the current response in varying coefficient models. The assumption that the recent past of the predictor process (but not the future or the distant past) has an affect on current responses is plausible in many applications where responses are driven by recent trends in predictors.

The historical functional linear model of Malfait and Ramsay (2003), the functional regression evolution of Müller and Zhang (2005) and the generalized varying coefficient model of Şentürk and Müller (2008) are other models intermediate between functional and varying coefficient linear models, as these models also include the effect of past values of the predictor process on current response. In contrast to these approaches, the proposed *functional varying coefficient model* provides a parsimonious and intuitive balance by introducing a *history index*, which serves to convey the effects of the recent past of the predictor on current response. In this model, the current value of the response process $Y(t)$ at time t depends on the recent

history of the predictor process X in a sliding window of length Δ,

$$E\{Y(t)|X(t)\} = \beta_0(t) + \beta_1(t) \int_0^{\Delta} \gamma(u)X(t-u)du, \tag{1}$$

for $t \in [\Delta, T]$ with a suitable $T > 0$. The *history index function* $\gamma(\cdot)$ in (1) defines the history index factor at $\beta_1(t)$, by quantifying the influence of the recent history of the predictor values on the response. The varying coefficient function $\beta_1(\cdot)$ represents the magnitude of this influence as a function of time. Functions γ, β_1 and the intercept function β_0 are assumed to be smooth. For identifiability, we assume that $\gamma(\cdot)$ is normalized by requiring that $\int_0^{\Delta} \gamma^2(u)du = 1$ and that $\gamma(0) > 0$ for identifiability, which is no real restriction, as $\{-\beta_1(t)\}\{-\gamma(u)\} = \beta_1(t)\gamma(u)$.

An assumption implicit in this model is that the history index function γ itself does not change over time, leading to a clear separation of time effects encoded in β_1 and history effects encoded in γ, thus decomposing the functional regression of Y on X into these two easily interpretable one-dimensional component functions. We propose an estimation algorithm for functional varying coefficient models that is geared towards addressing a commonly encountered challenge for longitudinal data, namely the irregularity of the subject-specific measurement times and the varying number of measurements available for each subject. We build on previous approaches of functional analysis that address the problem of sparse designs and noisy measurements (James, Hastie and Sugar 2000; Yao et al. 2005a,b).

Once $\gamma(\cdot)$ has been estimated, (1) reduces to a varying coefficient model. Note that even if $\gamma(\cdot)$ is assumed to be known, obtaining the predictors of the reduced varying coefficient model, i.e., $\int_0^{\Delta} \gamma(u)X(t-u)du$, may not be straightforward, due to the sparsity of the measurements for the predictor process in the history window $[t-\Delta, t]$, which renders numerical integration infeasible. Hence, we use functional estimation tools to develop a novel estimation procedure for this second step, which also turn out to be of value for the common version of varying coefficient models

relating $Y(t)$ to $X(t)$, especially when the design is sparse. A review of available estimation methods is provided by Fan and Zhang (2008). These methods include polynomial spline (Huang, Wu and Zhou 2004) and smoothing spline (Chiang, Rice and Wu 2001) approaches, as well as local polynomial smoothing (Fan and Zhang 2000; Wu and Chiang 2000).

In contrast to our proposal, most of these methods face severe challenges for sparse designs. While there are well established results for incorporating within subject correlation to increase efficiency for parametric models, much less work exists for nonparametric approaches (Lin and Carroll 2001; Qu and Li 2006; Fan, Huang and Li 2007). In these approaches, the varying coefficient estimation is usually based on local polynomial smoothing, and hence may suffer in the case of sparse designs. In Liang, Wu and Carroll (2003), an interesting mixed effects varying coefficient model coupled with a finite-dimensional mixed effects spline model for the predictors is proposed, addressing both the dependency in the predictor process and adjusting for measurement errors in a flexible parametric approach, while our approach is nonparametric and takes full advantage of the functional nature of the variables. The key for our functional approach is to target the covariance structure of X and cross-covariance structure of X and Y; estimates of these covariance surfaces behave well even under sparse designs.

In section 2, our estimation procedures are introduced, initially within the framework of established varying coefficient models. The proposed estimates for fitting functional varying coefficient models with history effect are described in Section 3. Results on consistency and inference are also presented in Sections 2 and 3, respectively, and numerical results can be found in Section 4. Details describing an application to longitudinal primary biliary liver cirrhosis data are provided in online Supplemental Material, where one can also find a technical appendix with assumptions and proofs of the asymptotic results.

2. Functional Approach to Varying Coefficient Models

2.1 Data and model

The observed data consist of square integrable random predictor and response trajectories (X_i, Y_i), $i = 1, \ldots, n$, which are realizations of the smooth random processes (X, Y), defined on a finite and closed interval domain $[0, T]$. The smooth random processes have unknown smooth mean functions $\mu_X(t) = EX(t)$, $\mu_Y(t) = EY(t)$, and (auto-)covariance functions $G_{XX}(s,t) = \mathrm{cov}\{X(s), X(t)\}$, $G_{YY}(s,t) = \mathrm{cov}\{Y(s), Y(t)\}$, for $s, t \in [0, T]$. Under mild conditions, one has orthogonal expansions for the covariances in terms of eigenfunctions ϕ_m and ψ_k with non-increasing eigenvalues ρ_m and λ_k,

$$G_{XX}(s,t) = \sum_m \rho_m \phi_m(s)\phi_m(t), \quad G_{YY}(s,t) = \sum_k \lambda_k \psi_k(s)\psi_k(t), \quad \text{for } s, t \in [0, T].$$

We now describe what we mean by sparse (or longitudinal) designs.

(SD) For the i-th subject one has a random number N_i of repeated measurements $U_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, $V_{ij} = Y_i(T_{ij}) + \epsilon_{ij}$, $j = 1, \ldots, N_i$, obtained at i.i.d. random measurement times $T_{i1}, \ldots, T_{iN_i}$, where $\varepsilon_{ij}, \epsilon_{ij}$ are zero mean finite variance i.i.d. errors. The N_i are assumed to be i.i.d and N_i, $T_{ij}, \varepsilon_{ij}, \epsilon_{ij}, X_i, Y_i$ are assumed to be mutually independent. The observed data are then $U_{ij} = \mu_X(T_{ij}) + \sum_m \xi_{im}\phi_m(T_{ij}) + \varepsilon_{ij}$, $V_{ij} = \mu_Y(T_{ij}) + \sum_k \zeta_{ik}\psi_k(T_{ij}) + \epsilon_{ij}$, where ξ_{im}, ζ_{ik} are sequences of uncorrelated mean zero functional principal components with second moments equal to the eigenvalues ρ_m and λ_k, respectively.

The representations in (SD) follow from the Karhunen-Loève expansion (see, e.g., Ash and Gardner, 1975). The estimators proposed below can be easily modified to cover the case where measurements for predictor and response processes are available at different time points T_{ij} and S_{ij}, respectively. One common reason for differences in measurement times between X and Y are missing values.

Now consider the standard longitudinal varying coefficient model

$$E\{Y(t)|X(t)\} = \beta_0(t) + \beta_1(t)X(t), \tag{2}$$

where it is assumed that the varying coefficient functions β_0, β_1 are smooth. With centralized predictor and response trajectories, i.e., $Y^C(t) = Y(t) - \mu_Y(t)$ and $X^C(t) = X(t) - \mu_X(t)$, model (2) can be rewritten as $E\{Y^C(t)|X(t)\} = \beta_1(t)X^C(t)$ with $\beta_0(t) = \mu_Y(t) - \beta_1(t)\mu_X(t)$. Fixing $t = t_0$, utilizing the linear relationship between $X(t)$ and $Y(t)$ at $t = t_0$ and using the smoothness of β_1, a standard method for fitting varying coefficient models (Fan and Zhang, 2008) is to minimize $\sum_i \sum_j (V_{ij}^C - \theta U_{ij}^C)^2 K_h(T_{ij} - t_0)$ with respect to θ, obtaining $\hat{\beta}_1(t_0) = \hat{\theta}$.

This estimate involves a kernel function K, chosen as a symmetric probability density, and a bandwidth h, where $K_h(\cdot) = K(\cdot/h)/h$. The minimization given above corresponds to utilizing local constant fits. Another common form of the estimator based on local linear fits is given in more detail in Appendix A.3. This approach to fitting varying coefficient models for longitudinal data does not take full advantage of the functional nature of the underlying data and may be biased or inefficient in the case of sparse and noise-corrupted measurements U_{ij}, V_{ij} that are encountered in many longitudinal studies. Our functional methods are addressing such data and therefore of interest for the classical varying coefficient model.

2.2 Functional approach

Our starting point is the following population least squares representation,

$$\beta_1(t) = \text{argmin}_\theta\{E(Y^C(t) - \theta X^C(t))^2\} = \frac{\text{cov}\{X(t), Y(t)\}}{\text{var}\{X(t)\}} = \frac{G_{XY}(t,t)}{G_{XX}(t,t)}, \tag{3}$$

$$\text{with} \quad G_{XY}(s,t) = \text{cov}\{X(s), Y(t)\} = \sum_{m=1}^{\infty}\sum_{k=1}^{\infty} E(\xi_m \zeta_k)\phi_m(s)\psi_k(t),$$

the cross-covariance function between X and Y.

In a first step we target the mean functions for the predictor and response processes by smoothing the aggregated data (T_{ij}, U_{ij}) and (T_{ij}, V_{ij}), $i = 1, \ldots, n$,

$j = 1, \ldots, N_i$, respectively, with local linear fitting, say. After obtaining the estimated mean functions $\hat{\mu}_X$ and $\hat{\mu}_Y$, we compute the raw covariances of X and the raw cross-covariances between X and Y, based on all observations from the same subject by $G_{X,i}(T_{ij}, T_{i\ell}) = \{U_{ij} - \hat{\mu}_X(T_{ij})\}\{U_{i\ell} - \hat{\mu}_X(T_{i\ell})\}$ and $G_{XY,i}(T_{ij}, T_{i\ell}) = \{V_{ij} - \hat{\mu}_Y(T_{ij})\}\{U_{i\ell} - \hat{\mu}_X(T_{i\ell})\}$, $i = 1, \ldots, n$, $j, \ell = 1, \ldots, N_i$, respectively. The raw covariances and cross-covariances are then fed into a two-dimensional local least squares smoothing algorithm to arrive at the final estimates $\widehat{G}_{XY}$ and $\widehat{G}_{XX}$, where special care needs to be taken around the diagonal; explicit forms of the covariance surface estimators are given in Appendix A.3 (compare also Yao et al. 2005ab).

This covariance estimation step eliminates the effect of the noise contamination of the observations, which is especially a problem for the predictor processes; and through pooling of the data across subjects it overcomes the problems associated with the sparseness of the design. In a subsequent step, we obtain the least squares estimates for the varying coefficient functions (3) as

$$\hat{\beta}_1(t) = \widehat{G}_{XY}(t,t)/\widehat{G}_{XX}(t,t), \quad \hat{\beta}_0(t) = \hat{\mu}_Y(t) - \hat{\beta}_1(t)\hat{\mu}_X(t). \tag{4}$$

This novel estimation procedure makes it possible to handle the noise contamination in both predictor and response processes, as well as the sparsity of the longitudinal data, and also to incorporate additional information that is inherent in the underlying covariance structure in the estimation step.

2.3 Uniform consistency and prediction

For details about the assumptions needed for the following results, see Appendix A.1. The proofs are provided in Appendix A.2 (see online Supplemental Material).

THEOREM 1. *Under Assumptions (A) in the Appendix, the varying coefficient function estimators (4) for the longitudinal varying coefficient model (2) satisfy*

$$\sup_{t \in [0,T]} |\hat{\beta}_r(t) - \beta_r(t)| = O_p\left\{\frac{1}{\sqrt{n}}\left(\frac{1}{h_1 h_2} + \frac{1}{h_X^2}\right)\right\}, \qquad \text{for } r = 0, 1.$$

Here bandwidths h_1, h_2 are used in the two-dimensional smoothing step of the raw covariances to obtain the cross-covariance function $\widehat{G}_{XY}$ and h_X is used in the two-dimensional smoothing step to obtain the covariance surface $\widehat{G}_{XX}$, as described explicitly in Appendix A.3. We note that this result holds under possibly very sparse conditions with, say, only two longitudinal measurements available per subject.

Aiming to predict the response trajectory Y^* for a new subject from a sparse predictor trajectory X^*, we find from (2),

$$E\{Y^*(t)|X^*(t)\} = \mu_Y(t) + \beta_1(t)\sum_{m=1}^{\infty}\xi_m^*\phi_m(t), \tag{5}$$

where $\xi_m^* = \int_0^T\{X^*(t) - \mu_X(t)\}\phi_m(t)dt$ is the mth functional principal component of X^*. The quantities $\mu_Y(t)$, $\beta_1(t)$ and $\phi_m(t)$ in (5) can be estimated from the data. For estimation of $\phi_m(t)$ and ρ_m, a nonparametric functional principal component analysis step is employed on the estimated auto-covariance surface $\widehat{G}_{XX}$; details can be found in Appendix A.3. Following Yao et al. (2005a), we invoke Gaussian assumptions for the estimation of ξ_m^* for the case of sparse designs.

For the jth measurement $X_j^* = X^*(T_j^*)$ of the predictor function X^* at time T_j^*, $j = 1, \ldots, N^*$, for a random number of total measurements N^*, and the observed data $U_j^* = X_j^* + \varepsilon_j^*$, in accordance with the data model (SD) in Section 2.1, assume that the functional principal components ξ_m^* and the measurement errors ε_j^* are jointly Gaussian. Then the best prediction of the scores ξ_m^*, given the observations $\boldsymbol{U}^* = (U_1^*, \ldots, U_{N^*}^*)$ and their number N^* and locations $\boldsymbol{T}^* = (T_1^*, \ldots, T_{N^*}^*)^{\mathrm{T}}$, is

$$\tilde{\xi}_m^* = \rho_m \boldsymbol{\phi}_m^{*^{\mathrm{T}}} \boldsymbol{\Sigma}_{\boldsymbol{U}^*}^{-1}(\boldsymbol{U}^* - \boldsymbol{\mu}_X^*). \tag{6}$$

In (6), $\tilde{\xi}_m^* = \rho_m \boldsymbol{\phi}_m^{*^{\mathrm{T}}} \boldsymbol{\Sigma}_{\boldsymbol{U}^*}^{-1}(\boldsymbol{U}^* - \boldsymbol{\mu}_X^*)$. Here, $\boldsymbol{\mu}_X^* = \{\mu_X(T_1^*), \ldots, \mu_X(T_{N^*})\}^{\mathrm{T}}$, $\boldsymbol{\phi}_m^* = \{\phi_m(T_1^*), \ldots, \phi_m(T_N^*)\}^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{U^*} = \mathrm{cov}(\boldsymbol{U}^*|N^*, \boldsymbol{T}^*)$, with (j,ℓ)th entry $(\Sigma_{U^*})_{j,\ell} = G_{XX}(T_{ij}, T_{i\ell}) + \sigma_X^2\delta_{j\ell}$ with $\delta_{j\ell} = 1$ if $j = \ell$ and 0 if $j \neq \ell$. To estimate the principal components ξ_m^*, we substitute into (6) the estimates of $\boldsymbol{\mu}_X^*$, ρ_m and

$\boldsymbol{\phi}_m^*$ that are based on the entire data, leading to $\hat{\xi}_m^* = \hat{\rho}_m \hat{\boldsymbol{\phi}}_m^{*\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}^*}^{-1}(\boldsymbol{U}^* - \hat{\boldsymbol{\mu}}_X^*)$, where $(\widehat{\Sigma}_{U^*})_{j,\ell} = \widehat{G}_{XX}(T_{ij}, T_{i\ell}) + \hat{\sigma}_X^2 \delta_{j\ell}$. For more details on obtaining $\hat{\sigma}_X^2$, see Appendix A.3. The predicted trajectories are

$$\widehat{Y}_M^*(t) = \hat{\mu}_Y(t) + \hat{\beta}_1(t) \sum_{m=1}^{M} \hat{\xi}_m^* \hat{\phi}_m(t). \tag{7}$$

THEOREM 2. *Under Assumptions (A), (C1), (C2), (C3a) in the Appendix, given N^* and $\boldsymbol{T}^*$, for all $t \in [0,T]$, predicted response trajectories in the varying coefficient model (2) satisfy*

$$\lim_{n\to\infty} \widehat{Y}_M^*(t) = \widetilde{Y}^*(t) \qquad \textit{in probability},$$

for the target trajectory $\widetilde{Y}^*(t) = \mu_Y(t) + \beta_1(t) \sum_{m=1}^{\infty} \tilde{\xi}_m^* \phi_m(t)$, as the number M of included eigencomponents satisfies $M = M(n) \to \infty$ as $n \to \infty$.

The number M of included eigenfunctions can be chosen by various criteria, including the Akaike information criterion (AIC), with more details given in Appendix A.3, or fraction of variance explained.

2.4 Asymptotic pointwise confidence bands for response trajectories

We next construct asymptotic confidence bands for predicted response trajectories Y^* (7), given $\boldsymbol{U}^*, N^*$ and $\boldsymbol{T}^*$. With $\boldsymbol{\xi}_*^M = (\xi_1^*, \ldots, \xi_M^*)^{\mathrm{T}}$, $\tilde{\boldsymbol{\xi}}_*^M = (\tilde{\xi}_1^*, \ldots, \tilde{\xi}_M^*)^{\mathrm{T}}$, where $\tilde{\xi}_m^*$ is as defined in (6), and the $M \times N^*$ matrix $\boldsymbol{H} = \mathrm{cov}(\boldsymbol{\xi}_*^M, \boldsymbol{U}^* | N^*, \boldsymbol{T}^*) = (\rho_1 \phi_1^*, \ldots, \rho_M \phi_M^*)^{\mathrm{T}}$, we observe that $\tilde{\boldsymbol{\xi}}_*^M = \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{U}^*}^{-1}(\boldsymbol{U}^* - \boldsymbol{\mu}_X^*)$ and $\mathrm{cov}(\tilde{\boldsymbol{\xi}}_*^M | N^*, \boldsymbol{T}^*) = \mathrm{cov}(\tilde{\boldsymbol{\xi}}_*^M, \boldsymbol{\xi}_*^M | N^*, \boldsymbol{T}^*) = \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{U}^*}^{-1}\boldsymbol{H}^{\mathrm{T}}$. Hence, $\mathrm{cov}(\tilde{\boldsymbol{\xi}}_*^M - \boldsymbol{\xi}_*^M | N^*, \boldsymbol{T}^*) = \mathrm{cov}(\boldsymbol{\xi}_*^M | N^*, \boldsymbol{T}^*) - \mathrm{cov}(\tilde{\boldsymbol{\xi}}_*^M | N^*, \boldsymbol{T}^*) = \boldsymbol{D} - \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{U}^*}^{-1}\boldsymbol{H}^{\mathrm{T}} \equiv \boldsymbol{\Omega}_M$, where $\boldsymbol{D} = \mathrm{diag}(\rho_1, \ldots, \rho_M)$. Therefore, under Gaussian assumptions, given N^* and $\boldsymbol{T}^*$, one has $\tilde{\boldsymbol{\xi}}_*^M - \boldsymbol{\xi}_*^M \sim \mathrm{N}(0, \boldsymbol{\Omega}_M)$.

Let $\widehat{\boldsymbol{\Omega}}_M = \widehat{\boldsymbol{D}} - \widehat{\boldsymbol{H}}\widehat{\boldsymbol{\Sigma}}_{\widetilde{X}^*}^{-1}\widehat{\boldsymbol{H}}^{\mathrm{T}}$, where $\widehat{\boldsymbol{D}} = \mathrm{diag}(\hat{\rho}_1, \ldots, \hat{\rho}_M)$ and $\widehat{\boldsymbol{H}} = (\hat{\rho}_1 \hat{\phi}_1^*, \ldots, \hat{\rho}_M \hat{\phi}_M^*)^{\mathrm{T}}$, and set $\boldsymbol{\phi}_{tM} = \{\beta_1(t)\phi_1(t), \ldots, \beta_1(t)\phi_M(t)\}^{\mathrm{T}}$ for $t \in [0,T]$, with $\hat{\boldsymbol{\phi}}_{tM}$ as the corresponding estimate. The following result facilitates the construction of pointwise confidence intervals for the mean response $E\{Y^*(t) | X^*(t)\}$ at predictor level X^*.

THEOREM 3. *Under Assumptions (A), (C1), (C2), (C3a), (C4a) in the Appendix, given N^* and $\boldsymbol{T}^*$, for all $t \in [0,T]$, $x \in \mathbb{R}$, predicted response trajectories in the varying coefficient model (2) satisfy*

$$\lim_{n\to\infty} P\left[\frac{\widehat{Y}^*_M(t) - E\{Y^*(t)|X^*(t)\}}{\hat{\omega}_{tM}} \le x\right] = \Phi(x),$$

where $\omega_{tM} = \boldsymbol{\phi}_{tM}^{\mathrm{T}} \boldsymbol{\Omega}_M \boldsymbol{\phi}_{tM}$, $\hat{\boldsymbol{\omega}}_{tM} = \hat{\boldsymbol{\phi}}_{tM}^{\mathrm{T}} \widehat{\boldsymbol{\Omega}}_M \hat{\boldsymbol{\phi}}_{tM}$ and $\Phi(\cdot)$ denotes the Gaussian cdf and $M(n) \to \infty$ as $n \to \infty$.

It follows from Theorem 3 that, ignoring truncation bias resulting from truncation at M in $\widehat{Y}^*_M$, the $(1-\alpha)100\%$ asymptotic pointwise confidence interval for $E\{Y^*(t)|X^*(t)\}$ is given by $\widehat{Y}^*_M(t) \pm \Phi(1-\alpha/2)\sqrt{\hat{\omega}_{tM}}$.

3. FITTING OF THE MODEL WITH HISTORY INDEX

To study functional varying coefficient models with history index, we observe that at each fixed time point t, the model in (1) reduces to a functional linear model between the scalar response $Y(t)$ and the functional predictor $X(s)$, $t-\Delta \le s \le t$.

3.1 Estimation of the history index function

Writing the model as

$$\begin{aligned} E\{Y^C(t)|X^C(s), s \in [t-\Delta, t]\} &= \beta_1(t) \int_0^\Delta \gamma(s) X^C(t-s) ds \\ &= \int_0^\Delta \alpha_t(s) X^C(t-s) ds, \end{aligned} \tag{8}$$

with regression parameter functions $\alpha_t(s) = \beta_1(t)\gamma(s)$, the functions $\alpha_t(s)$ contain the factor $\gamma(s)$ for each t. Owing to $\int_0^\Delta \gamma^2(s)\, ds = 1$, for each fixed time point t

$$\gamma(s) = \frac{\alpha_t(s)}{\{\int_0^\Delta \alpha_t^2(s)\, ds\}^{1/2}}. \tag{9}$$

Even though the estimator of $\alpha_t(s)$, obtained at a single time point t, suffices to obtain the history index function $\gamma(s)$ via (9), improved finite sample behavior

and stability of resulting estimators can be obtained by averaging the representation (9) over an equidistant grid of time points $(t_1, \ldots, t_R)$ in $[\Delta, T]$, pertaining to

$$\gamma(s) = \frac{\sum_{r=1}^{R} \alpha_{t_r}(s)}{[\int_0^\Delta \{\sum_{r=1}^{R} \alpha_{t_r}(s)\}^2 \, ds]^{1/2}}. \tag{10}$$

Here number of time points,R, typically would be small. Note that once the history index function γ is recovered, model (1) reduces to a varying coefficient model where a variant of the proposed estimation procedure in Section 2 can be applied.

To obtain a representation of the functions α_t in the functional linear regression model (8), consider processes $Z_t(s) = X^C(t-s)$, $s \in [0, \Delta]$, with autocovariance function $G_t(s_1, s_2) = G_{XX}(t - s_1, t - s_2)$ for $s_1, s_2 \in [0, \Delta]$ and orthonormal expansion into eigenfunctions ϕ_{tm} and eigenvalues ρ_{tm} given by $G_t(s_1, s_2) = \sum_m \rho_{tm}\phi_{tm}(s_1)\phi_{tm}(s_2)$. Expanding $\alpha_t(s) = \sum_m \alpha_{tm}\phi_{tm}(s)$, $s \in [0, \Delta]$, with suitable expansion coefficients α_{tm}, $m \geq 1$, and observing the representation $Z_t(s) = \sum_m \xi_{tm}\phi_{tm}(s)$, $s \in [0, \Delta]$, with random coefficients $\xi_{tm} = \int_0^\Delta Z_t(s)\phi_{tm}(s)$, one finds that minimizing the expected squared deviation $E\{Y^C(t) - \int_0^\Delta \alpha_t(s)X^C(t-s)ds\}^2$ corresponds to finding the values for α_{tm}, $m = 1, 2, \ldots$, satisfying $(d/d\alpha_{tm})[E\{Y^C(t) - \sum_m \alpha_{tm}\xi_{tm}\}^2] = 0$, $m = 1, 2, \ldots$. A straightforward calculation shows that the solutions are

$$\alpha_t'(s) = \sum_m \alpha_{tm}'\phi_{tm}(s), \qquad \alpha_{tm}' = \frac{1}{\rho_{tm}} \int_0^\Delta G_{XY}(t-s, t)\phi_{tm}(s) \, ds. \tag{11}$$

Given a fixed time t, reversing the time order of the data for all subjects that are observed in the window $[t - \Delta, t]$, and then performing a functional principal component analysis, yields estimates $\hat{\phi}_{tm}$, $\hat{\rho}_{tm}$ of the eigenfunctions and eigenvalues ϕ_{tm}, ρ_{tm} of processes Z_t. Estimates $\widehat{G}_t$ of covariance surfaces G_t are obtained as described in Appendix A.3. Utilizing the estimates $\widehat{G}_{XY}$, as used before in (4), and applying numerical integration, we obtain the estimates

$$\hat{\alpha}_{tm} = \frac{1}{\hat{\rho}_{tm}} \int_0^\Delta \widehat{G}_{XY}(t-s, t)\hat{\phi}_{tm}(s) \, ds, \;\; \hat{\alpha}_t(s) = \sum_{m=1}^{M_t} \hat{\alpha}_{tm}\hat{\phi}_{tm}(s). \tag{12}$$

One then proceeds to obtain estimates (12) for the equidistant grid of time points $(t_1, \ldots, t_R)$, leading to a series of estimates $\{\hat{\alpha}_{t_1}(s), \ldots, \hat{\alpha}_{t_R}(s)\}$, each of which targets a multiple of $\gamma(s)$, namely $\{\beta_1(t_1)\gamma(s), \ldots, \beta_1(t_R)\gamma(s)\}$. We apply (10) to combine these estimates, making use of the identifiability conditions $\int_0^\Delta \gamma^2(u)du = 1$ and $\gamma(0) > 0$, obtaining the estimated history index function

$$\hat{\gamma}(s) = \frac{\sum_{r=1}^R \hat{\alpha}_{t_r}(s)}{\sqrt{\int_0^\Delta \{\sum_{r=1}^R \hat{\alpha}_{t_r}(u)\}^2 du}}(-1)^{\mathcal{I}}, \tag{13}$$

where the integral is obtained numerically. Here $\mathcal{I}$ is the indicator function for the event $\{\sum_{r=1}^R \hat{\alpha}_{t_r}(0)\}/\sqrt{\int_0^\Delta \{\sum_{r=1}^R \hat{\alpha}_{t_r}(u)\}^2 du} < 0$, or equivalently, $\sum_{r=1}^R \hat{\alpha}_{t_r}(0) < 0$.

3.2 Estimating the varying coefficient function

Once $\hat{\gamma}(s)$ has been obtained, the remaining unknown component in model (1) is the varying coefficient function β_1. Defining $\widetilde{X}(t) = \int_0^\Delta \gamma(s)X^C(t-s)ds$, model (8) can be interpreted as a varying coefficient model with predictor process $\widetilde{X}$. In principle, this leads to a straightforward procedure to obtain β_1 by replacing $X(t)$ with $\widetilde{X}(t)$ and applying the methods developed in Section 2. However, because of the sparsity of the available data, the numerical integration involved in estimating $\widetilde{X}(t)$ will often not yield good approximations in sparse longitudinal settings.

This motivates a simpler approach for the sparse case that avoids the estimation of $\widetilde{X}(t)$ separately for each subject. From (8), we find by conditioning

$$\begin{aligned}
\mathrm{cov}\{X(t), Y(t)\} &= \mathrm{cov}[E\{X^C(t)|X\}, E\{Y^C(t)|X\}] + E[\mathrm{cov}(X^C(t), Y^C(t)|X] \\
&= \beta_1(t) \int_0^\Delta \gamma(s)\mathrm{cov}\{X(t-s), X(t)\}\, ds, \\
\mathrm{cov}\{X(t), \widetilde{X}(t)\} &= \int_0^\Delta \gamma(s)\mathrm{cov}\{X(t-s), X(t)\}\, ds,
\end{aligned}$$

and therefore $\beta_1(t) = G_{XY}(t,t)/\int_0^\Delta \gamma(s)G_{XX}(t-s,t)ds$. This leads to estimates

$$\begin{aligned}
\hat{\beta}_1(t) &= \widehat{G}_{XY}(t,t)/\int_0^\Delta \hat{\gamma}(s)\widehat{G}_{XX}(t-s,t)\, ds, \\
\hat{\beta}_0(t) &= \hat{\mu}_Y(t) - \hat{\beta}_1(t)\int_0^\Delta \hat{\gamma}(s)\hat{\mu}_X(s)\, ds,
\end{aligned} \tag{14}$$

using the estimates for $G_{XY}, G_{XX}, \mu_X, \mu_Y$ that are described in Section 2.

3.3 Asymptotic consistency

The following result provides uniform convergence for the history index function γ and the varying coefficient functions β_0 and β_1 in the history index model (8).

THEOREM 4. *Under assumptions (A) and (B) in the Appendix, the functional varying coefficient function estimators (13), (14) satisfy*

$$\lim_{n\to\infty} \sup_{s\in[0,\Delta]} |\hat{\gamma}(s) - \gamma(s)| = 0, \qquad \lim_{n\to\infty} \sup_{t\in[\Delta,T]} |\hat{\beta}_r(t) - \beta_r(t)| = 0,$$

in probability for $r = 0, 1$.

The rates of convergence depend on specific properties of processes X and Y, on how the M_t increase with sample size, and on the bandwidth sequences that are used in the various smoothing steps. We mention here in passing that similar results also hold, under suitable regularity conditions, when Δ is replaced by a consistent estimate $\hat{\Delta}$ with $\hat{\Delta} \to \Delta$. One can see this from the definition of the functions α_t as minimizers of a least squares problem, see (8)-(11). For two choices Δ, Δ', for the non-overlapping part of the domains, $\int_{\min(\Delta,\Delta')}^{\max(\Delta,\Delta')} \alpha_t(s) X^C(t-s)ds = O_P(|\Delta - \Delta'|)$, which implies the same bound for the overlapping part of the domains, due to the minimum property, and these bounds are uniform in t, from which we may conclude results such as $\sup_t |\int_0^{\hat{\Delta}} \hat{\gamma}(s)X(t-s)ds - \int_0^{\Delta} \gamma(s)X(t-s)ds| = O_P(|\Delta - \hat{\Delta}|)$.

3.4 Prediction and asymptotic pointwise confidence bands for response trajectories

Prediction of the response trajectory Y^* for a new subject based on the functional linear model in (8) is obtained from the conditional expectation

$$E\{Y^*(t)|X^*(s), s \in [t-\Delta, t]\} = \mu_Y(t) + \beta_1(t) \int_0^{\Delta} \gamma(s) {X^*}^C(t-s)ds,$$

$$\text{with } \gamma(s) = \frac{\sum_{r=1}^{R} \sum_{m'=1}^{\infty} \alpha'_{rm'} \phi_{rm'}(s)}{[\int_0^{\Delta} \{\sum_{r=1}^{R} \alpha_{t_r}(s)\}^2 ds]^{1/2}} \quad \text{and} \quad {X^*}^C(t-s) = \sum_{m=1}^{\infty} \xi^*_{tm} \phi_{tm}(s),$$

by (10) and (11). Here we denote dependence on t_r by only one subscript r for simplicity, i.e., for example $\alpha'_{rm'} \equiv \alpha'_{t_r m'}$ and $\phi_{rm'}(s) \equiv \phi_{t_r m'}(s)$. Then

$$E\{Y^*(t)|X^*(s), s \in [t-\Delta, t]\}$$
$$= \mu_Y(t) + \frac{\beta_1(t) \sum_{m=1}^{\infty} \xi^*_{tm} \sum_{r=1}^{R} \sum_{m'=1}^{\infty} \alpha'_{rm'} \int_0^\Delta \phi_{tm}(s)\phi_{rm'}(s)}{[\int_0^\Delta \{\sum_{r=1}^{R} \alpha_{t_r}(s)\}^2 ds]^{1/2}}. \tag{15}$$

As in Section 2.4, we assume that the local functional principal components ξ^*_{tm} and the measurement errors are jointly Gaussian. With $Z^*_{tj} = Z^*_t(T_{tj})$ denoting the jth measurement for the predictor function $Z^*_t(s) = X^{*\mathrm{C}}(t-s), s \in [0, \Delta]$, at time T_{tj}, for a random number of measurements N^*_t, i.e., $j = 1, \ldots, N^*_t$, and with $\widetilde{Z}^*_{tj}$ denoting the noise contaminated version of Z^*_{tj} and $\widetilde{\boldsymbol{Z}}^*_t = (\widetilde{Z}^*_{t1}, \ldots, \widetilde{Z}^*_{tN^*_t})$, one finds that the best prediction of the scores ξ^*_{tm}, conditional on $\widetilde{\boldsymbol{Z}}^*_t$, N^*_t and $\boldsymbol{T}^*_t = (T^*_{t1}, \ldots, T_{tN^*_t})^{\mathrm{T}}$, is given by

$$\tilde{\xi}^*_{tm} = \rho_{tm} \boldsymbol{\phi}^{*\mathrm{T}}_{tm} \boldsymbol{\Sigma}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \widetilde{\boldsymbol{Z}}^*_t, \tag{16}$$

$\boldsymbol{\phi}^*_{tm} = \{\phi_{tm}(T^*_{t1}), \ldots, \phi_{tm}(T^*_{tN^*_t})\}^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{t\widetilde{\boldsymbol{Z}}^*} = \mathrm{cov}(\widetilde{\boldsymbol{Z}}^*_t | N^*_t, \boldsymbol{T}^*_t)$. The quantities in (16) can be estimated from the entire data, analogously to the description in Section 2.4, leading to $\hat{\xi}^*_{tm} = \hat{\rho}_{tm} \hat{\boldsymbol{\phi}}^{*\mathrm{T}}_{tm} \widehat{\boldsymbol{\Sigma}}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \widetilde{\boldsymbol{Z}}^*_t$, and the predicted trajectories

$$\widehat{Y}^*_{\mathcal{M},M_t}(t) = \hat{\mu}_Y(t) + \frac{\hat{\beta}_1(t) \sum_{m=1}^{M_t} \hat{\xi}^*_{tm} \sum_{r=1}^{R} \sum_{m'=1}^{M_r} \hat{\alpha}'_{rm'} \int_0^\Delta \hat{\phi}_{tm}(s)\hat{\phi}_{rm'}(s)}{[\int_0^\Delta \{\sum_{r=1}^{R} \hat{\alpha}_{t_r}(s)\}^2 ds]^{1/2}}, \tag{17}$$

where $\mathcal{M} = \sum_{r=1}^{R} M_r$. The following result establishes convergence of estimated response trajectories to the target function $\widetilde{Y}^*(t) = \mu_Y(t) + \{\beta_1(t) \sum_{m=1}^{\infty} \tilde{\xi}^*_{tm}$ $\sum_{r=1}^{R} \sum_{m'=1}^{\infty} \alpha'_{rm'} \int_0^\Delta \phi_{tm}(s)\phi_{rm'}(s)\}/[\int_0^\Delta \{\sum_{r=1}^{R} \alpha_{t_r}(s)\}^2 ds]^{1/2}$.

THEOREM 5. *Under Assumptions (A), (B), (C2), (C3b) in the Appendix, given N^*_t and $\boldsymbol{T}^*_t$, for all $t \in [\Delta, T]$, estimated response trajectories in the functional varying coefficient model (8) satisfy*

$$\lim_{n\to\infty} \widehat{Y}^*_{\mathcal{M},M_t}(t) = \widetilde{Y}^*(t), \qquad \textit{in probability.}$$

For constructing pointwise asymptotic confidence intervals for the mean response, let $\hat{\boldsymbol{\xi}}^{M_t}_{*,t} = (\xi^*_{t1}, \ldots, \xi^*_{tM_t})^{\mathrm{T}}$ and define $\tilde{\boldsymbol{\xi}}^{M_t}_{*,t}$ analogously. Further let the $M_t \times N^*_t$ matrix $\boldsymbol{H}_t = \mathrm{cov}(\boldsymbol{\xi}^{M_t}_{*,t}, \widetilde{\boldsymbol{Z}}^*_t | N^*_t, \boldsymbol{T}^*_t) = (\rho_{t1}\phi^*_{t1}, \ldots, \rho_{tM_t}\phi^*_{tM_t})^{\mathrm{T}}$. Since $\tilde{\boldsymbol{\xi}}^{M_t}_{*,t} = \boldsymbol{H}_t \boldsymbol{\Sigma}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \widetilde{\boldsymbol{Z}}^*_t$, $\mathrm{cov}(\tilde{\boldsymbol{\xi}}^{M_t}_{*,t} | N^*_t, \boldsymbol{T}^*_t) = \mathrm{cov}(\tilde{\boldsymbol{\xi}}^{M_t}_{*,t}, \boldsymbol{\xi}^{M_t}_{*,t} | N^*_t, \boldsymbol{T}^*_t) = \boldsymbol{H}_t \boldsymbol{\Sigma}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \boldsymbol{H}^{\mathrm{T}}_t$. Hence with similar arguments as in Section 2.4, given N^*_t and $\boldsymbol{T}^*_t$, $\tilde{\boldsymbol{\xi}}^{M_t}_{*,t} - \boldsymbol{\xi}^{M_t}_{*,t} \sim \mathrm{N}(0, \boldsymbol{\Omega}_{tM_t})$, where $\boldsymbol{\Omega}_{tM_t} = \boldsymbol{D}_t - \boldsymbol{H}_t \boldsymbol{\Sigma}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \boldsymbol{H}^{\mathrm{T}}_t$ with $\boldsymbol{D} = \mathrm{diag}(\rho_{t1}, \ldots, \rho_{tM_t})$.

We next establish the asymptotic distribution of $[\widehat{Y}^*_{\mathcal{M},M_t}(t) - E\{Y^*(t) | X^*(s), s \in [t-\Delta, t]\}]$. Let $\widehat{\boldsymbol{\Omega}}_{tM_t} = \widehat{\boldsymbol{D}}_t - \widehat{\boldsymbol{H}}_t \widehat{\boldsymbol{\Sigma}}^{-1}_{t\widetilde{\boldsymbol{Z}}^*} \widehat{\boldsymbol{H}}^{\mathrm{T}}_t$, where $\widehat{\boldsymbol{D}}_t = \mathrm{diag}(\hat{\rho}_{t1}, \ldots, \hat{\rho}_{tM_t})$ and $\widehat{\boldsymbol{H}}_t = (\hat{\rho}_{t1}\hat{\phi}^*_{t1}, \ldots, \hat{\rho}_{tM_t}\hat{\phi}^*_{tM_t})^{\mathrm{T}}$. Define $\boldsymbol{\phi}_{t\mathcal{M}} = \beta_1(t)\{\sum_{r=1}^{R}\sum_{m'=1}^{M_r} \alpha'_{rm'} \int_0^{\Delta} \phi_{t1}(s)\phi_{rm'}(s), \ldots, \sum_{r=1}^{R}\sum_{m'=1}^{M_r} \alpha'_{rm'} \int_0^{\Delta} \phi_{tM_t}(s)\phi_{rm'}(s)\}^{\mathrm{T}} / [\int_0^{\Delta}\{\sum_{r=1}^{R} \alpha_{t_r}(s)\}^2 ds]^{1/2}$ and let $\hat{\boldsymbol{\phi}}_{t\mathcal{M}}$ be its estimate obtained from the data. The following result provides the asymptotic distribution for estimated predicted trajectories, where $\widehat{Y}^*_{\mathcal{M},M_t}(t) = \hat{\mu}_Y(t) + \hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t\mathcal{M}} \hat{\boldsymbol{\xi}}^{M_t}_{*,t}$.

THEOREM 6. *Under Assumptions (A), (B), (C2), (C3b), (C4b), (C5) in the Appendix, given N^*_t and $\boldsymbol{T}^*_t$, for all $t \in [\Delta, T]$, $x \in \mathbb{R}$, the estimated predicted response trajectories $\widehat{Y}^*_{\mathcal{M},M_t}(t)$ (17) in the varying coefficient model (8) satisfy*

$$\lim_{n\to\infty} P\left[\frac{\widehat{Y}^*_{\mathcal{M},M_t}(t) - E\{Y^*(t) | X^*(s), s \in [t-\Delta, t]\}}{\hat{\omega}_{t\mathcal{M},M_t}} \le x\right] = \Phi(x),$$

where $\omega_{t\mathcal{M},M_t} = \boldsymbol{\phi}^{\mathrm{T}}_{t\mathcal{M}} \boldsymbol{\Omega}_{tM_t} \boldsymbol{\phi}_{t\mathcal{M}}$, $\hat{\omega}_{t\mathcal{M},M_t} = \hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t\mathcal{M}} \widehat{\boldsymbol{\Omega}}_{tM_t} \hat{\boldsymbol{\phi}}_{t\mathcal{M}}$ are as defined above.

As a consequence, the $(1-\alpha)100\%$ asymptotic pointwise confidence interval for $E\{Y^*(t) | X^*(s), s \in [t-\Delta, t]\}$ is given by $\widehat{Y}^*_{\mathcal{M},M_t}(t) \pm \Phi(1-\alpha/2)\sqrt{\hat{\omega}_{t\mathcal{M},M_t}}$.

4. Numerical Results

4.1 Choice of Lag Δ and Application to Primary Biliary Liver Cirrhosis Data

The proposed functional varying coefficient index model with history index was applied to longitudinal measurements made on patients with primary biliary cirrhosis, regressing prothrombin time on albumin levels. The data come from a sparse longitudinal design, with the number of repeated measurements per subject ranging

between 1 and 8. Our analysis includes the fitting and interpretation of the proposed model in the framework of a longitudinal study and a comparison with the functional linear model, applied to the same data. Details can be found in the online Supplemental Material.

For the data-adaptive selection of the lag parameter Δ in model (1), we choose the minimizer of the absolute prediction error

$$\hat{\Delta} = \text{argmin}_{\Delta} \sum_{i=1}^{n} \sum_{j=1}^{N_i} |Y_i(T_{ij}) - \widehat{Y}^*_{i\mathcal{M},M_{T_{ij}},\Delta}(T_{ij})|, \tag{18}$$

where $\widehat{Y}^*_{i\mathcal{M},M_{T_{ij}},\Delta}$ is the predicted trajectory for the i-th subject, employing lag Δ and the estimated scores in (17).

4.2 Simulation Results

We report here the results of two simulation studies. The goal of the first study was to assess the performance of the functional approach (4) for the standard varying coefficient model (without history component), as described in Section 2. This approach is especially aimed at fitting varying coefficient models for situations with sparse and noisy data. We compare its performance in such settings with standard kernel linear smoothing. A second simulation study was designed to assess the performance of the proposed estimators (13), (14) for functional varying coefficient models with history component. Both simulations employed 500 Monte Carlo runs.

The scenario for the first simulation reflects very sparse designs. The number of measurements per subject was randomly chosen with equal probability from $\{3, 4, 5\}$ for each of $n = 100$ subjects. The locations T_{ij} of the measurements for the i-th subject were assumed to be uniformly distributed on $[0, T]$, with $T = 10$. The simulated predictor process X was generated with mean function $\mu_X(t) = t + \sin(t)$, and covariance function constructed from two eigenfunctions, $\phi_1(t) = \cos(\pi t/10)/\sqrt{5}$ and $\phi_2(t) = \sin(\pi t/10)/\sqrt{5}$, for $0 \leq t \leq 10$ and two eigenvalues, $\rho_1 = 2$ and $\rho_2 = 1$, respectively. The functional principal components ξ_{im} $(m = 1, 2)$ were

generated from $\mathcal{N}(0, \rho_m)$. Following (2), the response trajectories were generated from $Y_i(t) = \beta_0(t) + \beta_1(t)X_i(t) + W_i(t)$, where $\beta_0(t) = t$, $\beta_1(t) = \sin(\pi t/10)$ and the mean zero process $W_i(t)$ corresponds to the part of $Y_i(t)$ that is not explained by the predictor $X_i(t)$. Specifically, the W_i were constructed from the same two eigenfunctions as used for $X(t)$, with Gaussian functional principal components generated with eigenvalues $\rho = 0.1$ for both eigenfunctions. The measurements of both predictor and response trajectories were assumed to be contaminated with measurement errors, accordingly the measurements are $U_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, $V_{ij} = Y_i(T_{ij}) + \epsilon_{ij}$, where $\varepsilon_{ij}, \epsilon_{ij}$ are i.i.d. zero mean Gaussian errors, both with variances 0.1. Bandwidths for the smoothing of mean functions and auto- and cross-covariance surfaces were chosen by generalized cross-validation.

We compared the proposed functional approach with kernel linear smoothing, described in Appendix A.3, with evaluation criteria mean absolute deviation error (MADE) and weighted average squared error (WASE), defined as, respectively,

$$\text{MADE} = \frac{1}{2T}\sum_{r=0}^{1}\frac{\int|\beta_r(t) - \hat{\beta}_r(t)|dt}{\{\text{range}(\beta_r)\}}, \quad \text{WASE} = \frac{1}{2T}\sum_{r=0}^{1}\frac{\int\{\beta_r(t) - \hat{\beta}_r(t)\}^2dt}{\{\text{range}^2(\beta_r)\}},$$

where range(β_r) is the range of the function $\beta_r(t)$ and $T = 10$. We also considered unweighted average squared error (UASE), defined in the same way as WASE, but without weights in the denominator. Boxplots of the ratios of the values of MADE, WASE and UASE of the proposed method over the kernel linear smoothing approach are given in Figure 1 (a). Figure 1 (d) contains the boxplots from another simulation scenario with the same setup, except for irregular but dense (non-sparse) measurement times, with total number of repeated measurements generated uniformly from $\{20, \ldots, 30\}$. The boxplots indicate that the proposed estimators lead to much improved finite sample performance in both sparse and dense cases, likely due to the fact that the proposed method adjusts for noise contaminated measurements and takes advantage of information inherent in the underlying correlation structure of

the longitudinal processes. The cross-sectional mean varying coefficient function estimates obtained by averaging over the Monte Carlo runs are provided in Figures 1 (b, c) and (e, f) for both simulation scenarios, respectively.

In a second simulation scenario we studied the finite sample performance of the proposed estimators for functional varying coefficient models with history index. Fifty trajectories were generated, along with a random number of measurements for each trajectory, chosen with equal probability from $\{1, \ldots, 8\}$, with uniformly distributed locations on the domain $[0, 10]$. The time lag in model (1) was $\Delta = 5$, corresponding to half the length of the domain of the measurement locations. Predictor processes X were simulated with the same mean function as in the first simulation and covariance function constructed from eigenfunctions $\phi_1(t) = \sin(\pi t/5)/\sqrt{5}$ and $\phi_2(t) = \cos(\pi t/5)/\sqrt{5}$, for $0 \leq t \leq 10$, with eigenvalues $\rho_1 = 10$ and $\rho_2 = 5$, respectively. The functional principal components ξ_{im} $(m = 1, 2)$ were generated from $\mathcal{N}(0, \rho_m)$, and the response trajectories according to $Y_i(t) = \beta_0(t) + \beta_1(t) \int_0^\Delta \gamma(u) X_i(t-u) du + W_i(t)$, with $\beta_0(t) = t^2/2$, $\beta_1(t) = 5 \sin(\pi t/10)$, and $\gamma(u) = \sqrt{2/5} \cos(\pi u/5)$, satisfying the identifiability conditions $\int \gamma^2(u) du = 1$ and $\gamma(0) > 0$. As in the first simulation, $W_i(t)$ was generated as a zero mean Gaussian process with the same eigenfunctions as X and eigenvalues 0.5 and 0.3, respectively, and both observed predictor and response trajectories were assumed to be contaminated with Gaussian measurement errors with variances 0.3; smoothing bandwidths were chosen by generalized cross-validation.

In Figure 2, the cross-sectional medians of the estimated functional varying coefficient functions are presented along with the 5% and 95% cross-sectional percentiles, overlaid with the true coefficient functions. The displayed functions from the functional varying coefficient model fit are obtained by fixing Δ at the true value 5. Also displayed are the varying coefficient functions obtained by fitting the varying coefficient model given in equation (2), with kernel linear smoothing as outlined

in Appendix A.3 (see online Supplemental Material). Boxplots of the ratios of the values of MADE, WASE and UASE of the proposed method over the corresponding values for kernel linear smoothing for estimating the varying coefficient functions β_0 and β_1 are given in Figure 2 (d). These ratios are presented under two scenarios, where in the first scenario Δ is fixed at 5 and hence is not estimated, and in the second scenario Δ is selected data-adaptively, according to the proposed lag selection criterion (18). This second scenario is found to lead to slightly improved performance of the estimators in terms of the error ratios. Overall, the standard varying coefficient model is clearly unsuited to target the varying coefficient functions, due to the presence of a history index, while the proposed method is seen to work well even for very sparse data.

We separately evaluated the performance of the absolute prediction error criterion (18) for choosing the lag Δ, by adopting the second simulation scenario and a true lag parameter $\Delta = 5$. The percentages of the chosen values for Δ were found to be $(2, 2, 26, 32, 38)$ and $(0, 2, 70, 10, 18)$, among the choices $\Delta = (3, 4, 5, 6, 7)$, for simulations with sample sizes $n = 50$ and $n = 200$, respectively. So the criterion leads to the correct choice of Δ more frequently, as sample size increases. In a second simulation for the case where the true lag parameter is $\Delta = 0$, corresponding to the special case of a regular varying coefficient model, we found the percentages of chosen lag values to be $(89, 7, 4)$ and $(100, 0, 0)$, among the choices $\{0, 1, 2\}$, for sample sizes 50 and 200, respectively. We conclude that this criterion satisfactorily distinguishes between varying coefficient models that require a history index and those that do not.

While the proposed estimation procedures in both simulation scenarios yield significant gains for the modeling of sparse error-prone longitudinal data, the local linear smoothing approach is computationally simpler. The computational effort for the proposed method is comparable to that for fitting a functional linear model. For

sample sizes $n = 200$ and $n = 50$, computing times for fitting the respective models are $(0.4, 150, 179)$ and $(0.1, 43, 52)$ seconds for (local linear smoothing, proposed estimation algorithm and fitting a functional linear model), respectively. Here, an equidistant grid of $R = 15$ time points is used in obtaining the proposed history function estimator in comparisons. Using fewer grid points down to just one point preserves the consistency properties and accelerates computation.

5. Concluding Remarks

The functional data analysis perspective suggests two useful extensions of varying coefficient modelling. The first extension provides an alternative and improved way of fitting standard varying coefficient models, especially for the sparse and noise-contaminated data case that is commonly encountered in longitudinal studies. The second extension generalizes the varying coefficient model by introducing a history index, defined by a smooth history index function, summarizing the effect of the predictor function on the response function in a window prior to current time.

The proposed functional varying coefficient model incorporates predictor effects on the response in a parsimonious and easily interpretable fashion through the history index function and the proposed representation of varying coefficient functions through suitable auto- and cross-covariances (1) enables the user to directly incorporate information residing in the underlying covariance structure; (2) allows for estimation in sparse designs; (3) easily handles additional measurement errors in both predictors and responses. These three features lead to improved finite sample performance for sparsely and densely sampled, error-prone longitudinal data.

Note that the theoretical results of Section 3, notably Theorem 4, are obtained under the assumption of a fixed lag Δ. Future extensions could include further study of lag estimators $\hat{\Delta}$ and of models with time-varying lags $\Delta(t)$.

6. Supplemental Materials

Two supplements are provided in the online Supplemental Materials:

Supplement I: Application to Longitudinal Primary Biliary Liver Cirrhosis Data. This is a description of the data application of the proposed varying coefficient model with history index, as briefly described in Section 4.1, including figures.

Supplement II: Technical Appendix. This appendix provides the assumptions for the asymptotic results, the proofs of the theorems and details about the estimation procedures, describing the smoothing steps and the functional AIC.

References

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

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001), "Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables," *Journal of the American Statistical Association* 96, 605–619.

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

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.

Hastie, T. and Tibshirani, R. (1993), "Varying coefficient models," *Journal of the Royal Statistical Society* B , 55, 757–796.

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

James, G., Hastie, T. J. and Sugar, C. A. (2000), "Principal component models for sparse functional data," *Biometrika*, 87, 587–602.

Liang, H., Wu, H. and Carroll, R. J. (2003), "The relationship between virologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error," *Biostatistics*, 4, 297–312.

Lin, X. and Carroll, R. J. (2001), "Semiparametric regression for clustered data using generalized estimating equations, *Journal of the American Statistical Association* , 96, 1045–1056.

Malfait, N. and Ramsay, J. O. (2003), "The historical functional linear model, *Canadian Journal of Statistics*, 31, 115–128.

Müller, H. G. and Zhang, Y. (2005), "Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories," *Biometrics*, 61, 1064–1075.

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

Ramsay, J. O. and Dalzell, C. J. (1991), "Some tools for functional data analysis," *Journal of the Royal Statistical Society* B , 53, 539–572.

Şentürk, D. and Müller, H. G. (2008), "Generalized varying coefficient models for longitudinal data," *Biometrika*, 95, 653–666.

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

Yao, F., Müller, H. G. and Wang, J. L. (2005a), "Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association* 100, 577–590.

Yao, F., Müller, H. G. and Wang, J. L. (2005b), "Functional linear regression analysis for longitudinal data," *Annals of Statistics* , 3, 2873–2903.

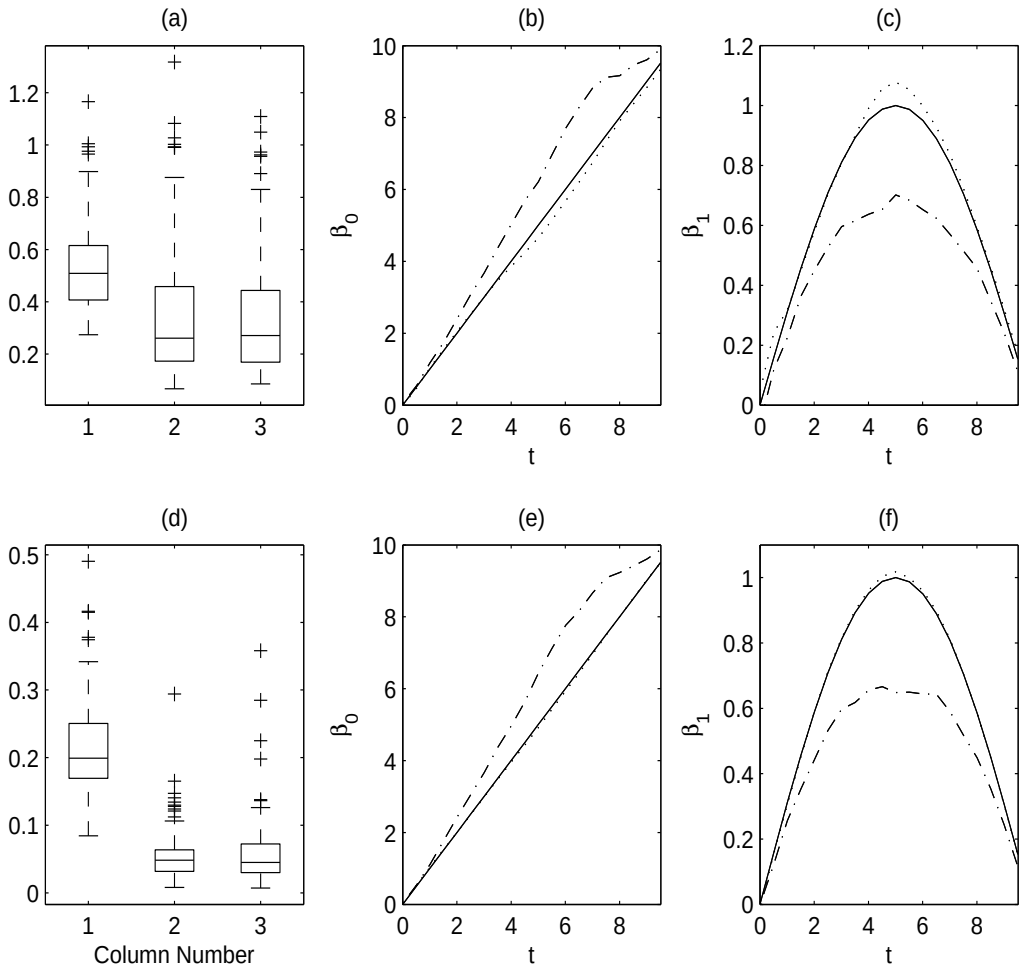

Figure 1: (a, d) Boxplots for the ratios of error measures for proposed estimates for varying coefficient models over kernel linear smoothing, for MADE, WASE and UASE (shown from left to right) for (a) sparse longitudinal data, and (d) irregular non-sparse longitudinal data. Quotients smaller than correspond to cases where the proposed functional approach is superior. The box plots are based on ratios obtained from 500 Monte Carlo runs. (b, c, e, f) The cross-sectional mean curves of the proposed estimates (dotted) and of kernel linear smoothing (dash-dotted) overlaying the true varying coefficient functions (solid) for sparse error-prone measurements (b, c), and irregular non-sparse error prone measurements (e, f).

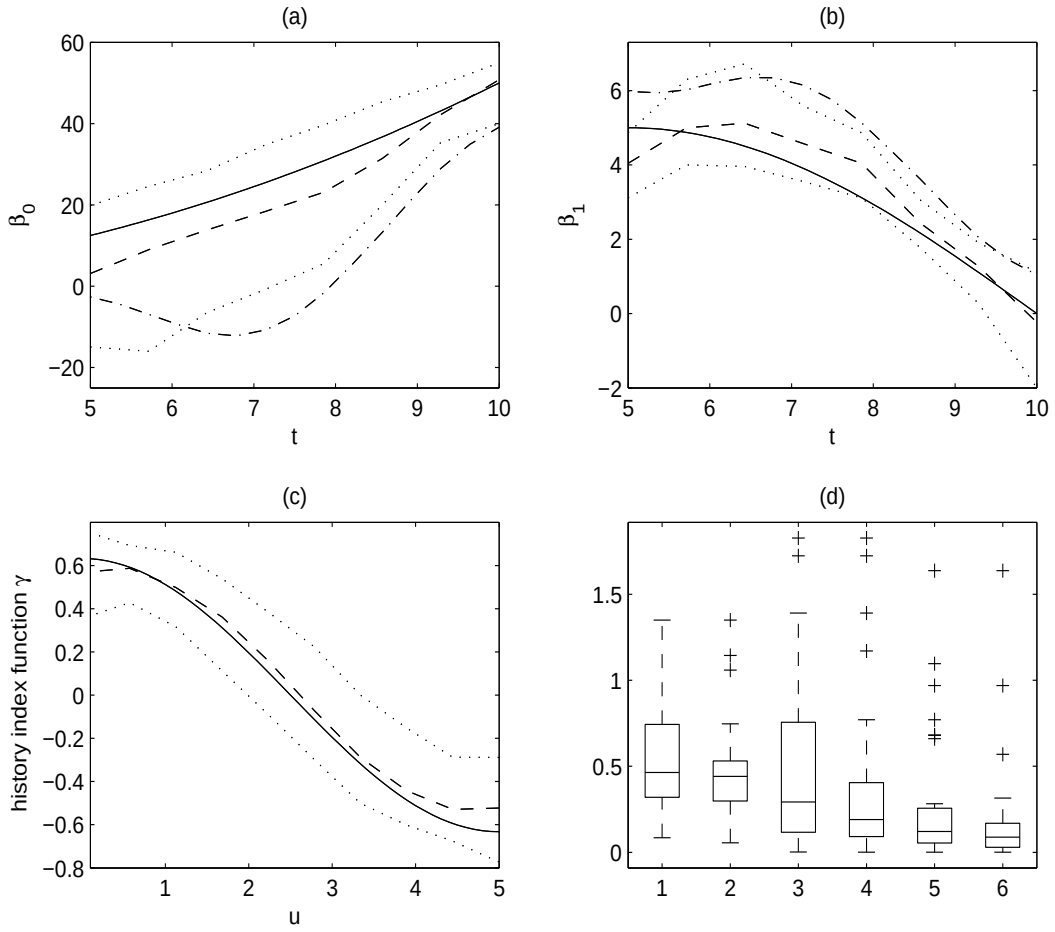

Figure 2: (a) The cross-sectional median curves of the proposed estimates (dashed) along with 5% and 95% cross-sectional percentiles (dotted) overlaying the true varying coefficient function β_0 (solid) from the second simulation scenario described in Section 4.2 for a functional varying coefficient model with lag $\Delta = 5$. Also displayed are the cross-sectional median curves for standard varying coefficient model fits, using kernel linear smoothing (dash-dotted). (b) The cross-sectional median curves for the varying coefficient function β_1. (c) The cross-sectional median curves for the history function γ. (d) Boxplots for the ratios of error measures for proposed estimates with a fixed Δ in comparison to the standard kernel linear smoothing approach (for MADE – 1, WASE – 3, UASE – 5), and for proposed estimates with estimated Δ (for MADE – 2, WASE – 4, UASE – 6).

Supplemental Material

Supplement I: Application to Longitudinal Primary Biliary Liver Cirrhosis Data

We analyze longitudinal measurements made on primary biliary cirrhosis (a chronic liver disease) patients that participated in a clinical trial at the Mayo Clinic in the ten year interval from January 1974 to May 1984 (as discussed in appendix D of Fleming and Harrington, 1991, data available at http://lib.stat.cmu.edu/datasets/pbcseq). Multiple laboratory measurements were taken at visits scheduled at six months, one year, and annually thereafter post diagnosis. Actual measurement times however vastly differed from the scheduled times and also the number of available repeated measurements differed across subjects. Thus the data is sparse and irregular. We consider the longitudinal regression relation between prothrombin time (PT, in seconds) as response and serum albumin levels (ALB, in mg/dl) as predictor and include in our analysis the measurements made on 194 female patients on the interval 0 to 2000 days. For both PT and ALB, the available number of repeated measurements per subject ranges between 1 and 8, with a median of 5. The included measurements are shown in Figure 3, overlaid with smooth estimates of the respective mean functions, which for ALB indicate a decreasing and for PT an increasing trend, as the disease progresses.

The classical functional linear regression model

$$E\{Y(t)|X\} = \mu_Y(t) + \int_0^{2000} \beta(s,t)X^C(s)ds, \tag{19}$$

of Y (PT) versus X (ALB), fitted by the methods proposed for sparse longitudinal data in Yao et al. (2005b), serves as a starting point. The estimate of the bivariate regression surface β is shown in Figure 4, along with slices through the regression surface for a sequence of response times $t \in (1400, 2000)$ and predictor times $s \in (t - 1400, t)$, i.e., for the 1400 days preceding the respective response time t. Going from PT at $t = 0$ to PT at $t = 2000$, the regression surface β indicates a dependency of the PT outcome on the ALB time course trend that is seen to steadily reverse from positive to negative. While the initial positive dependence of PT near time 0 on the long-term trend in ALB is clearly a feature of this functional linear model, it remains hard to interpret, as contrary to the implications of this model, in reality future

values of ALB cannot causally influence PT at current time. Thus this positive dependence must be viewed as a purely statistical association; a more negative slope in the ALB trend might be associated with a higher initial ALB level, which could help explain the phenomenon of slope reversal in the dependence of PT on ALB. We conclude that functional linear regression, due to its emphasis on regressing entire time courses on each other, is often not fully appropriate for longitudinal data.

The slices through the regression surface shown in the right panel of Figure 2 indicate a decaying negative dependency of the late time values of PT on 1400 days of prior measurements of ALB. The shapes of these slices suggest that the proposed functional varying coefficient model with its product-type form provides an adequate approximative modeling approach, where the common dependence feature of PT on recent past ALB levels is captured by the history index function, and the change in the magnitude of this dependence by the varying coefficient function.

We next fit the proposed functional varying coefficient model with $\Delta = 1400$ days. The lag Δ was chosen by minimizing absolute prediction error

$$\sum_{i=1}^{n} \sum_{j=1}^{N_i} |Y_i(T_{ij}) - \widehat{Y}^*_{i\mathcal{M},M_{T_{ij}},\Delta}(T_{ij})|, \tag{20}$$

where the preliminary range of Δ values considered was 600 to 1400 days. For more details, we refer to Section 4.1 of the article. In (20), the sum is taken over $T_{ij} \in (\max \Delta, T)$, where predictions of the response trajectories $\widehat{Y}_{i\mathcal{M},M_t,\Delta}(T_{ij})$ are obtained as described in (17), employing lag Δ. The resulting estimated history index and varying coefficient functions are displayed in Figure 5, along with 90% bootstrap percentile confidence intervals. Bootstrap confidence intervals are constructed from 500 bootstrap samples, generated by resampling subjects. Bandwidths for the smoothing of the cross-sectional mean functions, the covariance surfaces G_t, G_{XX} and the cross-covariance surface G_{XY} were selected by generalized cross-validation. The numbers of included eigenfunctions for the local expansions in (12) were chosen by AIC, where the explicit form of the criterion used is given in Appendix A.3. An equidistant grid of $R = 15$ time points was used to obtain the history function estimator via (13).

Also shown in Figure 5 are varying coefficient model fits obtained by kernel linear smoothing, as described in Appendix A.3. The estimated varying coefficient functions from both approaches are displayed in Figures 5 (a) and (b). We find that PT and ALB have a significant negative relationship, which becomes more negative

as time progresses. The estimated history index function γ displayed in Figure 5 (c) indicates a substantial longer term effect of past ALB levels on current PT, with maximal effect around 900 days prior to current time, while the effects of both more recent and more distant ALB levels are slightly smaller. This suggests that this relationship cannot be adequately modeled with a standard varying coefficient model, where only current ALB is assumed to have an effect on PT, so that the model with history index is clearly preferrable.

We further compare the predicted response trajectories from the proposed functional varying coefficient model with those obtained from the functional linear model (19). Predicted PT trajectories on the domain $[1400, 2000]$ for four randomly selected subjects are shown in Figure 6, along with estimated pointwise asymptotic bands, obtained with the methods in Section 3.4, also including predicted trajectories from the functional linear model. The proposed model appears to be considerably more flexible as compared to the functional linear model, which utilizes predictor information from the entire ALB trajectories, and therefore suffers from the above discussed difficulties in interpreting the effect that future predictor values have on past responses. The varying coefficient model with history index emerges as preferred model.

Supplement II: Technical Appendix

A.1. Assumptions

A common set of assumptions needed for all results is listed under (A), assumption (B) is needed for the asymptotic results for estimators in Section 3, and assumptions (C) for the asymptotic results on predicted response trajectories.

The data (T_{ij}, U_{ij}) and (T_{ij}, V_{ij}), $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, are assumed to have the same distribution as $(\mathcal{T}, U)$ and $(\mathcal{T}, V)$, with joint densities $g_1(t, x)$ and $g_2(t, y)$. Assume also that the observation times T_{ij} are i.i.d. with marginal densities $f_{\mathcal{T}}(t)$. Let T_1 and T_2 be i.i.d. T and U_1 and U_2 be the repeated measurements of X made on the same subject at times T_1 and T_2, similarly define V_1, V_2. The predictor (and response) measurements made on the same subject at different times are allowed to be dependent. Assume $(T_{ij}, T_{i\ell}, U_{ij}, U_{i\ell})$, $1 \leq j \neq \ell \leq N_i$, is identically distributed as (T_1, T_2, U_1, U_2) with joint density function $g_X(t_1, t_2, u_1, u_2)$ and analogously for $(T_{ij}, T_{i\ell}, V_{ij}, V_{i\ell})$ with identical joint density function $g_Y(t_1, t_2, v_1, v_2)$. The following regularity conditions are assumed on $f_{\mathcal{T}}(t)$, $g_1(t, u)$, $g_2(t, v)$, $g_X(t_1, t_2, u_1, u_2)$ and $g_Y(t_1, t_2, v_1, v_2)$.

(A1) Let p_1, 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 $[0, T]$, $(dp/dt^p) g_1(t, u)$ and $(dp/dt^p) g_2(t, v)$ exist and are continuous on $[0, T] \times \mathbb{R}$, and $\{dp/(dt_1^{p_1} dt_2^{p_2})\} g_X(t_1, t_2, u_1, u_2)$ and $\{dp/(dt_1^{p_1} dt_2^{p_2})\} g_Y(t_1, t_2, v_1, v_2)$ exist and are continuous on $[0, T]^2 \times \mathbb{R}^2$.

(A2) The number of measurements N_i made on the ith subject is a random variable such that $N_i \overset{iid}{\sim} N$, where N is a positive discrete random variable with $P(N > 1) > 0$. The observation times and measurements are assumed to be independent of the number of observations for any subset $J_i \subset \{1, \ldots, N_i\}$ and for all $i = 1, \ldots, n$, i.e., $\{T_{ij}, V_{ij}, U_{ij} : j \in J_i\}$ is independent of N_i.

Let $K_1(\cdot)$ and $K_2(\cdot, \cdot)$ be the nonnegative univariate and bivariate kernel functions that are used in the smoothing steps for the mean functions μ_X, μ_Y, covariance surface G_{XX}, local covariance surfaces G_t and cross-covariance surface G_{XY} (see Appendix A.3 for the definition of the smoothers). Assume that K_1 and K_2 are compactly supported densities with zero means and finite variances. Let b_X, b_Y be the bandwidths used for estimating μ_X, μ_Y, h_X be the bandwidth for estimating G_{XX}

and G_t and h_1, h_2 be the bandwidths for estimating G_{XY}, where all bandwidths depend on n. Explicit forms of the estimators are given in Appendix A.3. Also define the Fourier transformations of $K_1(u)$ and $K_2(u,v)$ by $\kappa_1(t) = \int e^{-iut} K_1(u)du$ and $\kappa_2(t,s) = \int e^{-(iut+ivs)} K_2(u,v)dudv$, which are required to satisfy assumption (A3).

(A3) The Fourier transform $\kappa_1(t)$ is absolutely integrable, i.e., $\int |\kappa_1(t)|dt < \infty$ and $\kappa_2(t,s)$ is absolutely integrable, i.e., $\int\int |\kappa_2(t,s)|dtds < \infty$. As $n \to \infty$, the following is assumed about the bandwidths: $b_X \to 0$, $b_Y \to 0$, $nb_X^4 \to \infty$, $nb_Y^4 \to \infty$, $nb_X^6 < \infty$ and $nb_Y^6 < \infty$; $h_X \to 0$, $nh_X^6 \to \infty$ and $nh_X^8 < \infty$; $h_1/h_2 \to 1$, $nh_1^6 \to \infty$ and $nh_1^8 < \infty$.

(A4) Assume that U and V have finite fourth moments.

(B) The number of included eigenfunctions from local eigen-decompositions in (12) and (17), M_t or $M_{t_r} \equiv M_r$ when $t = t_r$ introduced in Sections 3.1 and 3.4 are integer valued sequences that depend on sample size n such that $\inf_{t\in[\Delta,T]} M_t(n) \to \infty$ and both $\inf_{t\in[\Delta,T]} M_t(n)$ and $\sup_{t\in[\Delta,T]} M_t(n)$ satisfy the rate conditions given in assumption $(B5)$ of Yao et al. (2005b). In addition $\sum_{m=1}^{\infty} |\alpha_{tm}^* \phi_{tm}(s)|$ is continuous in s for all $t \in [\Delta, T]$ where α_{tm}^* is defined in (11) and the function $\sum_{m=1}^{M_t} \alpha_{tm}^* \phi_{tm}(s)$ converges absolutely to $\sum_{m=1}^{\infty} \alpha_{tm}^* \phi_{tm}(s)$, as $M_t \to \infty$.

(C1) The number M of included eigenfunctions in (7) is an integer valued sequences that depends on sample size n such that it satisfies the rate conditions given in assumption $(B5)$ of Yao et al. (2005b)

(C2) The number and locations of the measurements for a subject or cluster remain unaltered as the sample size $n \to \infty$.

(C3a) For all $1 \le i \le n$, $m \ge 1$ and $1 < \ell < N_i$, the functional principal components ξ_{im} and the measurement errors $\epsilon_{i\ell}$ in (SD) are jointly Gaussian.

(C3b) For all $1 \le i \le n$, $m \ge 1$, $t \in [\Delta, T]$ and $1 \le \ell \le N_{it}$, the functional principal components ξ_{itm} and the measurement errors $\epsilon_{it\ell}$ are jointly Gaussian.

(C4a) For all $t \in [\Delta, T]$, there exists a continuous positive definite function ω_t such that ω_{tM} (see Theorem 3) satisfies $\omega_{tM} \to \omega_t$ as $M \to \infty$.

(C4b) For all $t \in [\Delta, T]$, there exists a continuous positive definite function ω_t such that $\omega_{t\mathcal{M},M_t}$ (see Theorem 6) satisfies $\omega_{t\mathcal{M},M_t} \to \omega_t$ as $M_t, M_1, \ldots, M_R \to \infty$.

(C5) $\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\{E(\xi_m\zeta_k)\}^2/(\lambda_k\rho_m) < \infty$.

A.2. Proofs

Proof of Theorem 1. Uniform consistency of $\widehat{G}_X(s,t)$, $\hat{\mu}_X(t)$ and $\hat{\mu}_Y(t)$ follow from Theorem 1 of Yao et al. (2005a), where $\sup_{t\in[0,T]}|\hat{\mu}_X - \mu_X| = O_p\{1/(\sqrt{n}b_X)\}$ and $\sup_{t\in[0,T]}|\hat{\mu}_Y - \mu_Y| = O_p\{1/(\sqrt{n}b_Y)\}$. Uniform consistency of $\widehat{G}_{XY}(s,t)$ follows from Lemma $A1$ of Yao et al. (2005b). Combining these results completes the proof.

Proof of Theorem 2. For fixed M, let $\widetilde{Y}^*_M(t) = \mu_Y(t) + \beta_1(t)\sum_{m=1}^{M}\tilde{\xi}^*_m\phi_m(t)$ and recall that $\widetilde{Y}^*(t) = \mu_Y(t) + \beta_1(t)\sum_{m=1}^{\infty}\tilde{\xi}^*_m\phi_m(t)$. From

$$|\widehat{Y}^*_M(t) - \widetilde{Y}^*(t)| \le |\widehat{Y}^*_M(t) - \widetilde{Y}^*_M(t)| + |\widetilde{Y}^*_M(t) - \widetilde{Y}^*(t)|,$$

it follows from Lemma 3 of Yao et al. (2005b) that $\widetilde{Y}^*_M(t) \xrightarrow{p} \widetilde{Y}^*(t)$ as $M \to \infty$ and $n \to \infty$ and from Theorem 1 there that $\sup_{t\in[0,T]}|\hat{\mu}_Y(t) - \mu_Y(t)| = O_p\{1/\sqrt{n}b_Y)\}$. Theorem 1 of Section 2.3, Theorem 3 and (17) of Yao et al. (2005b) and Slutsky's Theorem imply $\sup_{t\in[0,T]}|\hat{Y}^*_M(t) - \widetilde{Y}^*_M(t)| \to 0$ as $n \to \infty$ and Theorem 2 follows.

Proof of Theorem 3. According to the proof of Theorem 2, $\lim_{n\to\infty}\sup_{t\in[0,T]}|\widehat{Y}^*_M(t) - \widetilde{Y}^*_M(t)| \to 0$. Under the Gaussian assumption, it is shown in Section 2.4 that for a fixed $M \ge 1$, $\tilde{\boldsymbol{\xi}}^M_* - \boldsymbol{\xi}^M_* \sim \mathcal{N}(0, \boldsymbol{\Omega}_M)$. It then follows that $[\widehat{Y}^*_M(t) - E_M\{Y^*(t)|X^*(t)\}] \xrightarrow{D} Z_M \sim \mathcal{N}(0, \omega_{tM})$, for $E_M\{Y^*(t)|X^*(t)\} = \mu_Y(t) + \beta_1(t)\sum_{m=1}^{M}\xi^*_m\phi_m(t)$, since $\widehat{Y}^*_M(t) - E_M\{Y^*(t)|X^*(t)\} = \widehat{Y}^*_M(t) - \widetilde{Y}^*_M(t) + \widetilde{Y}^*_M(t) - E_M\{Y^*(t)|X^*(t)\}$. Under Assumption (C4a), letting $M \to \infty$ leads to $Z_M \xrightarrow{D} Z \sim \mathcal{N}(0, \omega_t)$. From the Karhunen-Loéve Theorem, $|\mathrm{E}_M\{Y^*(t)|X^*(t)\} - E\{Y^*(t)|X^*(t)\}| \xrightarrow{P} 0$, as $M \to \infty$. Hence $\lim_{M\to\infty}\lim_{n\to\infty}[\widehat{Y}^*_M(t) - E\{Y^*(t)|X^*(t)\}] \overset{D}{=} Z$. From Theorem 1 of Section 2.3 and Theorems 1 and 2 of Yao et al. (2005b), it follows that $\hat{\omega}_{tM} \xrightarrow{P} \omega_{tM}$ as $n \to \infty$, and therefore that $\lim_{M\to\infty}\lim_{n\to\infty}\hat{\omega}_{tM} = \omega_t$. Theorem 3 follows by Slutsky's Theorem.

Proof of Theorem 4. To prove uniform consistency of $\hat{\gamma}(s)$, note that

$$\begin{aligned}
&\sup_{s\in[0,\Delta]}|\hat{\alpha}_t(s) - \alpha_t(s)|\sup_s\left|\sum_{m=1}^{M}\{\hat{\alpha}_{tm}\hat{\phi}_{tm}(s) - \alpha^*_{tm}\phi_{tm}(s)\}\right| \\
&+ \sup_s\left|\sum_{m=1}^{M}\alpha^*_{tm}\phi_{tm}(s) - \sum_{m=1}^{\infty}\alpha^*_{tm}\phi_{tm}(s)\right| + \sup_s\left|\sum_{m=1}^{\infty}\alpha^*_{tm}\phi_{tm}(s) - \sum_{m=1}^{\infty}\alpha_{tm}\phi_{tm}(s)\right| \\
&\equiv Q_1(n) + Q_2(n) + Q_3(n),
\end{aligned}$$

for $t \in [\Delta, T]$. One has $Q_1(n) \xrightarrow{P} 0$ as $n \to \infty$ from the uniform convergence of $\widehat{G}_{XY}$ given in Lemma $A1$ of Yao et al. (2005b) and Theorem 2 there implies the uniform convergence of $\hat{\phi}_{tm}$ and $\hat{\rho}_{tm}$. Assumption (B) and the fact that α^*_{tm} are the least squares estimators of α_{tm}, respectively, then lead to $Q_2(n) = o_p(1)$ and $Q_3(n) = o_p(1)$, and the uniform consistency of $\hat{\alpha}_t(s)$ to the uniform consistency of $\hat{\gamma}(s)$. The uniform consistency of $\hat{\beta}_1(t)$ and $\hat{\beta}_0(t)$ follows analogously.

Proof of Theorem 5. For fixed $\mathcal{M}$ and M_t, let $\widetilde{Y}^*_{\mathcal{M},M_t}(t) = \mu_Y(t) + \{\beta_1(t)\sum_{m=1}^{M_t}\tilde{\xi}^*_{tm}$ $\sum_{r=1}^{R}\sum_{m'=1}^{M_r}\alpha'_{rm'}\int_0^\Delta \phi_{tm}(s)\phi_{rm'}(s)\}/[\int_0^\Delta\{\sum_{r=1}^{R}\alpha_{t_r}(s)\}^2 ds]^{1/2}$. Similar to the proof of Theorem 2, note that

$$|\widehat{Y}^*_{\mathcal{M},M_t}(t) - \widetilde{Y}^*(t)| \le |\widehat{Y}^*_{\mathcal{M},M_t}(t) - \widetilde{Y}^*_{\mathcal{M},M_t}(t)| + |\widetilde{Y}^*_{\mathcal{M},M_t}(t) - \widetilde{Y}^*(t)|.$$

By similar arguments as in the proof of Lemma 3 of Yao et al. (2005b), $\widetilde{Y}^*_{\mathcal{M},M_t}(t) \xrightarrow{P} \widetilde{Y}^*(t)$ as $M_t, M_1, \ldots, M_R \to \infty$ and $n \to \infty$, and Theorem 4 of Section 3.3, Theorems 1 and 3 of Yao et al. (2005b) and Slutsky's Theorem imply $\sup_{t\in[\Delta,T]}|\widehat{Y}^*_{\mathcal{M},M_t}(t) - \widetilde{Y}^*_{\mathcal{M},M_t}(t)| \to 0$ as $n \to \infty$, whence Theorem 5 follows.

Proof of Theorem 6. Similar to the proof of Theorem 3, where the decomposition $\widehat{Y}^*_{\mathcal{M},M_t}(t) - E_{\mathcal{M},M_t}\{Y^*(t)|X^*(s), s \in [t-\Delta, t]\} = \widehat{Y}^*_{\mathcal{M},M_t}(t) - \widetilde{Y}^*_{\mathcal{M},M_t}(t) + \widetilde{Y}^*_{\mathcal{M},M_t}(t) - E_{\mathcal{M},M_t}\{Y^*(t)|X^*(s), s \in [t-\Delta, t]\}$ is utilized with $E_{\mathcal{M},M_t}\{Y^*(t)|X^*(s, s \in [t-\Delta, t])\} = \mu_Y(t) + \{\beta_1(t)\sum_{m=1}^{M_t}\xi^*_{tm}\sum_{r=1}^{R}\sum_{m'=1}^{M_r}\alpha'_{rm'}\int_0^\Delta \phi_{tm}(s)\phi_{rm'}(s)\}$ $/[\int_0^\Delta\{\sum_{r=1}^{R}\alpha_{t_r}(s)\}^2 ds]^{1/2}$. Convergence of $\hat{\omega}_{t\mathcal{M},M_T}$ to $\omega_{t\mathcal{M},M_t}$ follows from Lemma 1 of Yao et al. (2005b) and Theorem 4 of Section 3.3.

A.3 Estimation Procedures

We provide here more explicit versions for the proposed mean and covariance estimators and for the standard local linear smoothing procedure for varying coefficient models. Define local linear scatterplot smoothers for $\mu_X(t)$ through minimizing $\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\{(T_{ij}-t)/b_X\}\{U_{ij} - \eta_0 - \eta_1(t - T_{ij})\}^2$, with respect to η_0, η_1, leading to $\hat{\mu}_X(t) = \hat{\eta}_0(s)$, analogously for the mean function $\mu_Y(t)$ of Y.

Let $G_{X,i}(T_{ij}, T_{i\ell}) = \{U_{ij} - \hat{\mu}_X(T_{ij})\}\{U_{i\ell} - \hat{\mu}_X(T_{i\ell})\}$, and define the local linear surface smoother for $G_{XX}(s,t)$ through minimizing

$$\sum_{i=1}^{n}\sum_{1\le j\ne\ell\le N_i} K_2\left(\frac{T_{ij}-s}{h_X}, \frac{T_{i\ell}-t}{h_X}\right)[G_{X,i}(T_{ij}, T_{i\ell}) - f\{\eta, (s,t), (T_{ij}, T_{i\ell})\}]^2,$$

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 $\widehat{G}_{XX}(s,t) = \hat{\eta}_0(s,t)$. For estimation of σ^2_X, we use the procedure

described in Yao et al. (2005a). Eigenfunctions and eigenvalues are obtained as solutions $(\hat{\rho}_m, \hat{\phi}_m)_{m\geq 1}$ of the eigenequations, numerically obtained by discretization.

Estimates of the covariance surfaces G_t, noise contaminated versions $\Sigma_{t\widetilde{Z}}$ and local eigenfunctions ϕ_{tm} and eigenvalues ρ_{tm} are obtained as follows. Since Δ does not depend on n, optimal bandwidths for the local linear surface smoothers for G_t are of the same order as h_X, and for simplicity were chosen to equal h_X. Let $G_{t,i}(S_{ij}, S_{i\ell}) = G_{X,i}(T_{ij}, T_{i\ell})$ where $S_{ij} = t - T_{ij}$, $S_{i\ell} = t - T_{i\ell}$, $t \in [\Delta, T]$, $T_{ij}, T_{i\ell} \in [t-\Delta, t]$ and $S_{ij}, S_{i\ell} \in [0, \Delta]$. Define the local linear surface smoother for $G_t(s_1, s_2) = G_{XX}(t - s_1, t - s_2)$ for $t \in [\Delta, T]$, $s_1, s_2 \in [0, \Delta]$, through minimizing

$$\sum_{i=1}^{n} \sum_{j\neq\ell} K_2\left(\frac{S_{ij} - s_1}{h_X}, \frac{S_{i\ell} - s_2}{h_X}\right)[G_{t,i}(S_{ij}, S_{i\ell}) - f\{\eta, (s_1, s_2), (S_{ij}, S_{i\ell})\}]^2,$$

with respect to $\eta = (\eta_0, \eta_1, \eta_2)$, yielding $\widehat{G}_t(s_1, s_2) = \hat{\eta}_0(s_1, s_2)$. The local error variance σ^2_{tX} and the local eigenfunction and eigenvalue estimators are obtained analogously to the global estimates, and estimation of G_t and of local error variances σ^2_{tX} yields estimates of the noise contaminated local covariance surface $\Sigma_{t\widetilde{Z}}$. Let $G_{XY,i}(T_{ij}, T_{i\ell}) = \{V_{ij} - \hat{\mu}_Y(T_{ij})\}\{U_{i\ell} - \hat{\mu}_X(T_{i\ell})\}$. The local linear surface smoother for the cross-covariance surface $G_{XY}(s, t)$ is obtained through minimizing

$$\sum_{i}^{n} \sum_{j=1}^{N_i} \sum_{\ell=1}^{N_i} K_2\left(\frac{T_{ij} - s}{h_1}, \frac{T_{i\ell} - t}{h_2}\right)[G_{XY,i}(T_{ij}, T_{i\ell}) - f\{\eta, (s, t), (T_{ij}, T_{i\ell})\}]^2,$$

with respect to $\eta = (\eta_0, \eta_1, \eta_2)$, leading to $\widehat{G}_{XY}(s, t) = \hat{\eta}_0(s, t)$.

The local linear smoothing estimators targeting the varying coefficient functions used for comparison with the proposed estimators minimize

$$\sum_{i=1}^{n} \sum_{j=1}^{N_i} K\left(\frac{T_{ij} - t}{h}\right)\left[V_{ij} - \sum_{r=0}^{1}\{\theta_{r,0} + \theta_{r,1}(t - T_{ij})\}U_{ij}\right]^2,$$

with respect to $\theta_{0,0}, \theta_{0,1}, \theta_{1,0}, \theta_{1,1}$, leading to $\hat{\beta}_0(t) = \hat{\theta}_{0,0}(t)$ and $\hat{\beta}_1(t) = \hat{\theta}_{1,0}(t)$.

The AIC criterion used in the selection of M, the number of eigenfunctions included in (7), is given by

$$\begin{aligned} AIC(M) &= \sum_{i=1}^{n}\left\{\frac{1}{2\hat{\sigma}^2_X}\left(\widetilde{U}_i - \hat{\mu}_{X_i} - \sum_{m=1}^{M}\hat{\xi}_{im}\hat{\phi}_{im}\right)^{\mathrm{T}}\left(\widetilde{U}_i - \hat{\mu}_{X_i} - \sum_{m=1}^{M}\hat{\xi}_{im}\hat{\phi}_{im}\right)\right. \\ &+ \left.\frac{N_i}{2}\log(2\pi) + \frac{N_i}{2}\log\hat{\sigma}^2_X\right\} + M, \end{aligned}$$

where $\widetilde{U}_i = (U_{i1}, \ldots, U_{iN_i})^{\mathrm{T}}$, $\hat{\mu}_{X_i} = \{\hat{\mu}_X(T_{i1}), \ldots, \hat{\mu}_X(T_{iN_i})\}^{\mathrm{T}}$, $\hat{\phi}_{im} = \{\hat{\phi}_m(T_{i1}), \ldots, \hat{\phi}_m(T_{iN_i})\}^{\mathrm{T}}$ and $\hat{\xi}_{im}$ is as defined in Section 2.4. The number of eigenfunctions included in the local expansions of Sections 3.1 and 3.4, M_t or M_r when $t = t_r$, are chosen analogously by $AIC(M_t)$, where in the above definition N_i, $\hat{\sigma}_X^2$, $\widetilde{U}_i$, $\hat{\mu}_{Xi}$, $\hat{\phi}_{im}$ and $\hat{\xi}_{im}$ are replaced by their local counterparts at t.

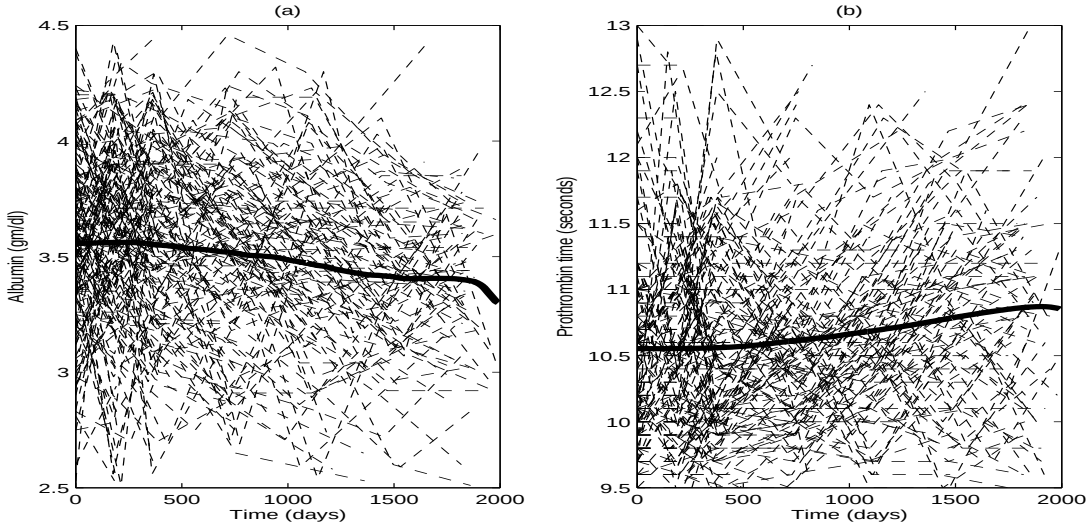

Figure 3: (a) Observed individual trajectories (dashed) and the smoothed estimate of the mean function $\hat{\mu}_X$ (thick solid) for serum albumin levels. (b) Observed individual trajectories (dashed) and the smoothed estimate $\hat{\mu}_Y$ of the mean function for prothrombin time (thick solid).

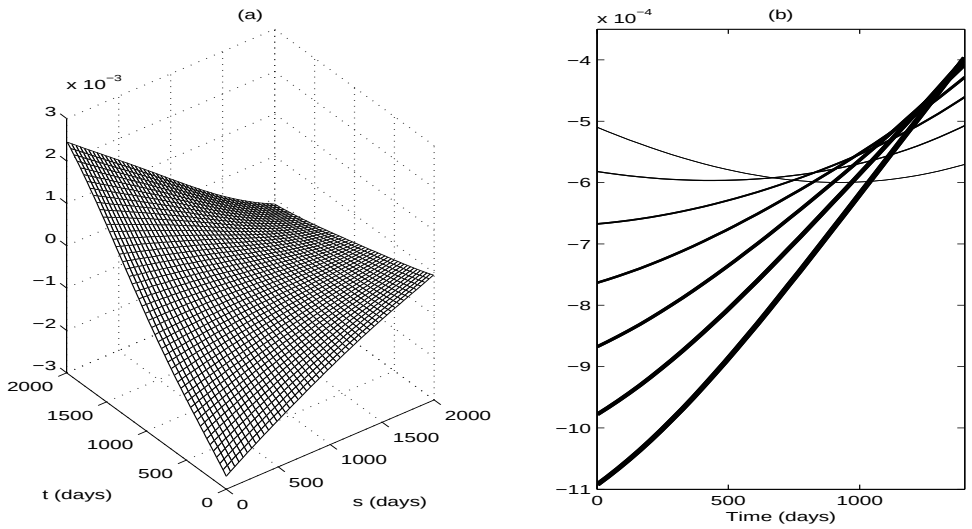

Figure 4: (a) Estimated bivariate regression surface $\beta(s,t)$ for the functional linear model in (19), regressing prothrombin time at time t on albumin level at time s (for centered trajectories). (b) Slices through the estimated bivariate regression surface in (a) for a sequence of time points $t = 1400$ (thinnest) , $t = 1500, \ldots, t = 2000$ (thickest) and $s \in (t - 1400, t)$.

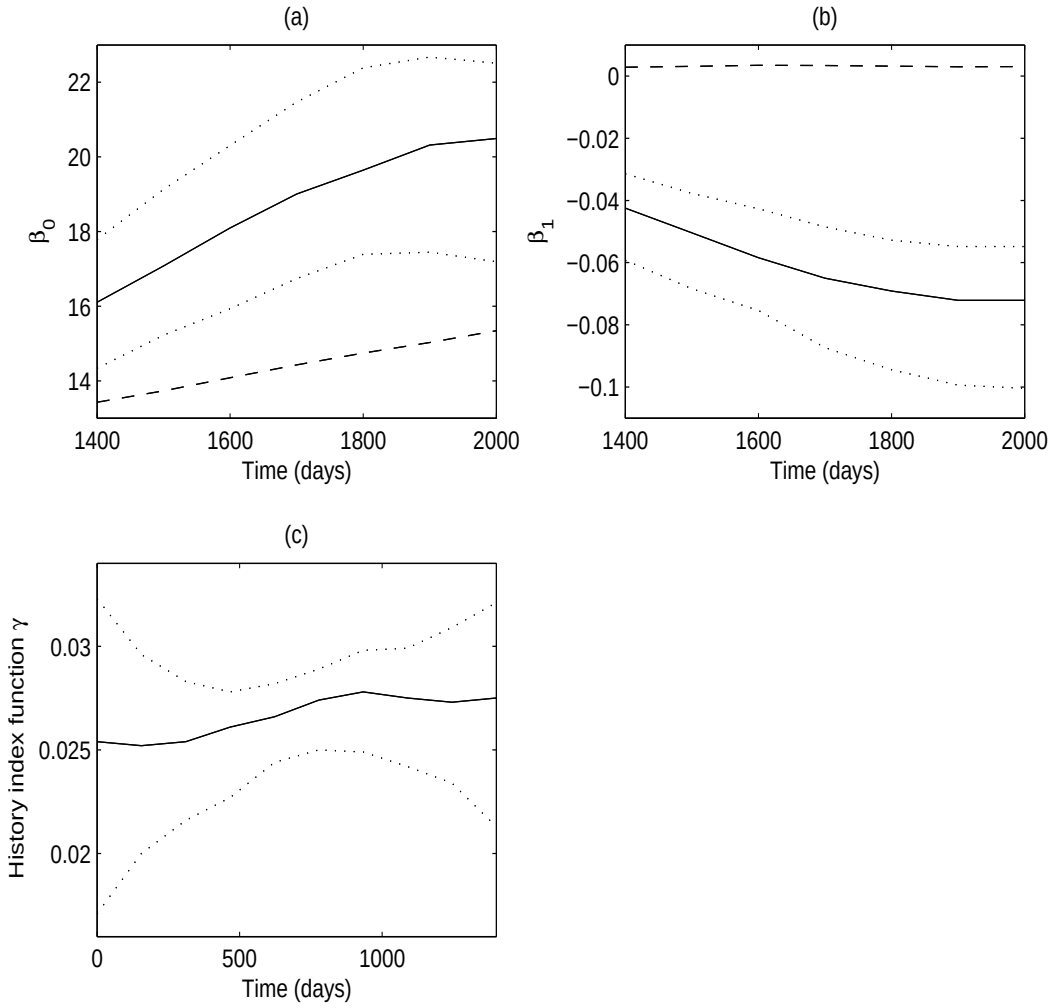

Figure 5: (a) Estimated varying coefficient function β_0 from the functional varying coefficient model fit (solid) with $\Delta = 1400$ days, along with 90% bootstrap confidence intervals (dotted) for the functional varying coefficient approach, for the primary biliary cirrhosis data. Estimated functions from the varying coefficient model fits using kernel linear smoothing (dashed) are also displayed. (b) Estimated varying coefficient function β_1 from both fits, along with 90% bootstrap confidence intervals. (c) Estimated history index function γ, along with 90% bootstrap confidence intervals.

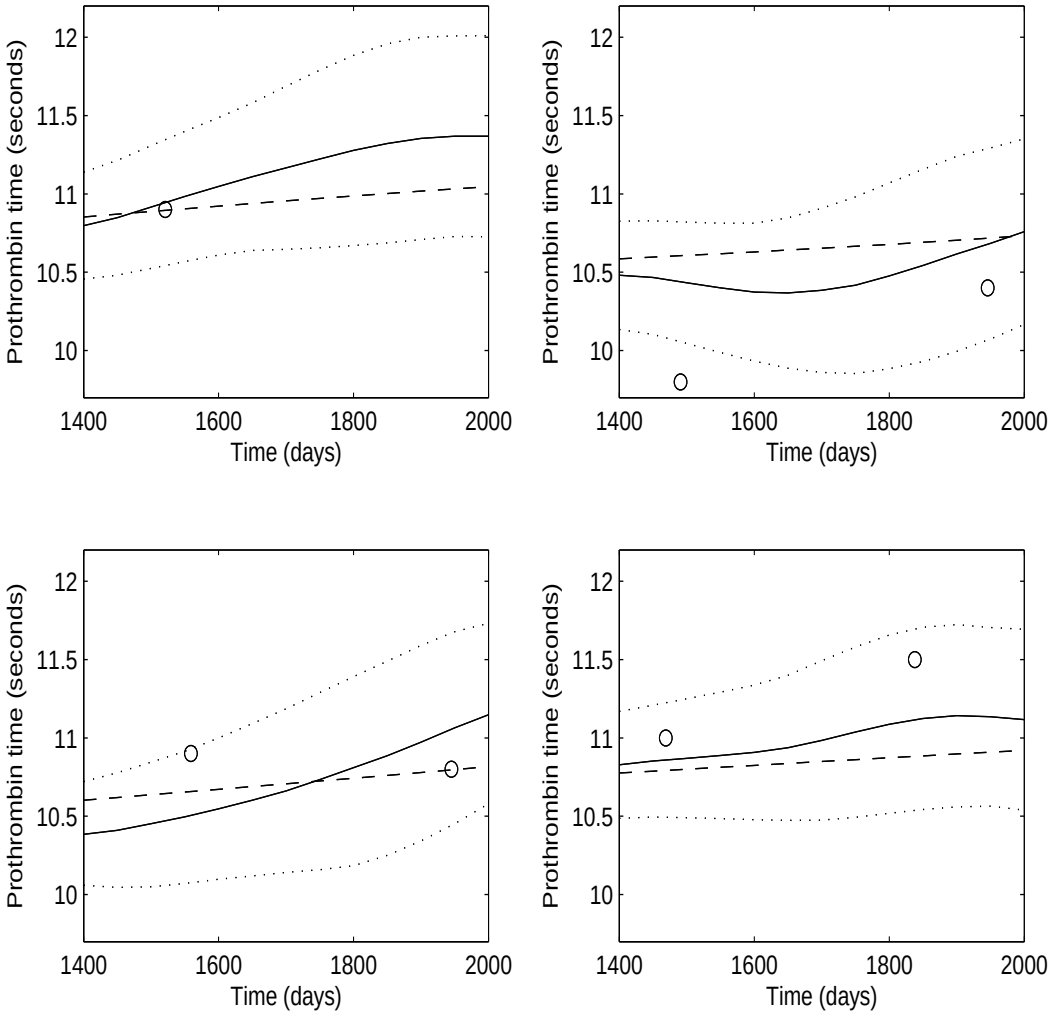

Figure 6: Observed values (circles) for prothrombin times (not used for prediction), predicted curves (solid) and 95% pointwise confidence bands (dotted), for four randomly selected patients, where bands and predicted curves are based on one-curve-leave-out analysis. Also displayed (dashed) are predicted curves from the fitted functional linear model given in (19).