Time-varying Additive Models for Longitudinal Data

Xiaoke Zhang[1], Byeong U. Park[2], and Jane-Ling Wang[1,*]

[1]University of California, Davis
[2]Seoul National University

January 4, 2012

Abstract

Additive model is an effective dimension reduction model that provides flexibility to model the relation between a response variable and key covariates. The literature is largely developed to scalar response and vector covariates. In this paper, more complex data is of interest, where both the response and covariates may be functions. A functional additive model is proposed together with a new smooth backfitting algorithm to estimate the unknown regression functions, whose components are time-dependent additive functions of the covariates. Due to the sampling plan, such functional data may not be completely observed as measurements may only be collected intermittently at discrete time points. We develop a uniform platform and efficient approach that can cover both dense and sparse functional data and the needed theory for statistical inference. The oracle properties of the component functions are also established.

Key Words: Smooth backfitting, local linear smoothing, oracle property, functional data, longitudinal data.

1 Introduction

Additive model is an effective approach to avoid the curse of dimensionality that also retains flexibility to model the relation between a response and its covariates. Let Y

[*]The research of Jane-Ling Wang is supported in part by NSF grant DMS-09-06813.

be a univariate response variable with d linearly independent covariates $(Z_1, \cdots, Z_d)^\top = \mathbf{Z}$. The traditional additive model takes the form

$$E(Y|\mathbf{Z}) = \mu_0 + \sum_{k=1}^{d} \mu_k(Z_k), \tag{1}$$

where $\mu_0 = E(Y)$ and $\mu_k(\cdot)$ is a smooth component function for all k.

Several algorithms have been proposed to estimate the component functions, including ordinary backfitting (Buja et al. (1989)), marginal integration (Linton and Nielsen (1995)), smooth backfitting (Mammen et al. (1999)) and regression splines (Stone (1985)). Oracle property is of interest for additive models, that is, one can achieve the same asymptotic property for a particular component as if all the other component functions were known. With strict constraints on the dependence between covariates, Opsomer and Ruppert (1997) and Opsomer (2000) showed that the ordinary backfitting method works but cannot achieved the oracle bias. While the marginal integration method cannot attain the oracle property, two improved versions by Linton (1997) and Fan et al. (1998) can achieve the oracle bias and variance. However, all marginal integration approaches suffer from the curse of dimensionality. This motivated Mammen et al. (1999) to propose the smooth backfitting algorithm, which can overcome the curse of dimensionality and achieve the oracle bias and variance when a local linear estimate is employed in the smooth backfitting algorithm. Spline approaches usually lack tractable bias and variance expressions. Wang and Yang (2007) and Wang and Yang (2009) overcome this theoretical hurdle by adding a second stage to backfit the spline estimator with a local smoother so that the two-stage spline-backfitted kernel estimator are asymptotically tractable and retain the oracle property for both the bias and variance.

All the aforementioned discussion is for univariate and independent response data with vector covariates. In this paper we are interested in modeling functional or longitudinal response data with possibly functional or longitudinal covariates. Let $Y(t)$ denote the response function and $\mathbf{Z}(t) = (Z_1(t), \cdots, Z_d(t))^\top$ its functional covariates defined on an interval, which is assumed to be $[0, 1]$ for simplicity. Some extensions of equation (1) or its generalized version have been considered for functional and longitudinal data and they take the form:

$$E(Y(t)|\mathbf{Z}) = E(Y(t)|\mathbf{Z}(t)) = \mu_0 + \sum_{k=1}^{d} \mu_k(Z_k(t)), \tag{2}$$

where $\mathbf{Z}$ is the vector of functional covariates. This line of work includes Berhane and Tibshirani (1998), Lin and Zhang (1999), Carroll et al. (2009), You and Zhou (2007), and Xue et al. (2010).

A closer look at equation (2) reveals that the effect of time on the component functions are only through the values of the functional or longitudinal covariates at that time, so these component functions are time independent. More specifically, let m be the regression function defined by $m(t, \mathbf{z}) = E(Y(t)|\mathbf{Z}(t) = \mathbf{z})$ where $\mathbf{z} = (z_1, \ldots, z_d)^\top$. Equation (2) is equivalent to

$$m(t, \mathbf{z}) = \mu_0 + \sum_{k=1}^{d} \mu_k(z_k). \tag{3}$$

This could be restrictive or should at least be validated. We adopt a dynamic approach that allows the component functions to vary with time and this leads to two-dimensional component functions and the following additive model:

$$m(t, \mathbf{z}) = \mu_0(t) + \sum_{k=1}^{d} \mu_k(t, z_k), \tag{4}$$

where $\mu_0(\cdot)$ is the overall mean function and $\mu_k(\cdot, \cdot)$ are the component functions. It is assumed that μ_k, $0 \le k \le d$, are smooth functions.

Such a model is a natural extension of the conventional additive model (1) by allowing equation (1) to hold for any given time t through time dynamic bivariate component functions. It can also be viewed as a dimension reduction approach for the model in Jiang and Wang (2010) to accommodate multiple covariates through an additive structure. The new model not only enjoys the merit of dimension reduction but also captures the time-dynamic relation of functional and longitudinal covariates and responses. Moreover, since model (3) is nested within model (4), the latter can be used for model checking of model (3) or other submodels.

To estimate the component functions in model (4), we adopt the smooth backfitting algorithm with a local linear smoother. Here, we need to extend the smooth backfitting algorithm by Mammen et al. (1999) to two dimensional setting and this triggers some technical challenges as unlike model (1) or (2) the mean function is not additive in t, hence the resulted additive space of the mean function may not be a closed subspace of $L^2([0, 1])$. Moreover, the algorithm involves higher dimensional smoothing. Despite these challenges, we establish some uniform results in t and covariate values, and obtain the oracle bias and variance together with joint asymptotic normality of the component functions.

The paper is organized as follows: Section 2 introduces the sampling plan and the model together with the estimation approach. Asymptotic properties of the estimator are presented in Section 3. A plug-in method for bandwidth choice is introduced in Section 4. Simulation results and a real data application are respectively shown in Section 5 and Section 6. Discussions and conclusions are in Section 7.

2 Methodology

The functional response $Y(t)$ and covariates, $\mathbf{Z}(t) = (Z_1(t), \cdots, Z_d(t))^\top$, we consider in this paper are regarded as L^2-stochastic processes on an interval, which we assume to be $[0, 1]$ for convenience. Let $\mathbf{Z}_i$ denote i.i.d. copies of $\mathbf{Z}$ and Y_i i.i.d. copies of Y, so $\mathbf{Z}_i$ and Y_i are all L^2-stochastic processes on $[0, 1]$. Based on equation (4), the model for $Y_i(t)$ after adding a stochastic component $w_i(t)$ is:

$$Y_i(t) = \mu_0(t) + \sum_{k=1}^{d} \mu_k(t, Z_{ik}(t)) + w_i(t), \tag{5}$$

where w_i's are i.i.d. copies of a L^2-stochastic processes w on $[0, 1]$ with $E(w(t)|\mathbf{Z}(t)) = 0$ and correlation may exists between $w(s)$ and $w(t)$ at any points s and t .

2.1 Model

Instead of observing the entire covariate and response processes, in longitudinal studies they are observed intermittently at discrete time points. Specifically, for the ith subject $i = 1, \cdots, n$, we observe $T_{ij} : j = 1, \cdots, N_i$, and the response may further be subject to random noises e_{ij}, which are i.i.d. copies of e. Thus, for the ith subject, we observe $\{(T_{ij}, \mathbf{Z}_{ij}, Y_{ij}) : i = 1, \cdots, N_i\}$, where $\mathbf{Z}_i(T_{ij}) = (Z_{ij1}, Z_{ij2}, \ldots, Z_{ijd})^\top$ and $Y_{ij} = Y_i(T_{ij}) + e_{ij}$. It follows from (5) that the observations Y_{ij}, T_{ij} and $\mathbf{Z}_{ij}$ admit the following longitudinal additive model:

$$Y_{ij} = \mu_0(T_{ij}) + \sum_{k=1}^{d} \mu_k(T_{ij}, Z_{ijk}) + w_{ij} + e_{ij}, \quad 1 \le j \le N_i;\ 1 \le i \le n, \tag{6}$$

where $w_{ij} = w_i(T_{ij})$ and $E(w_{ij}|T_{ij}, \mathbf{Z}_{ij}) = 0$.

We assume that T_{ij} are independent of the process $\mathbf{Z}_i$. We also assume that $N_i, i = 1, \ldots, n$, are i.i.d. copies of a random variable N but our technical arguments are valid for non-stochastic N_i with obvious modification. Conditioning on $N_1, \ldots, N_n$, the time points T_{ij} are i.i.d. copies of a random variable T. Therefore, $\mathbf{Z}_{ij}$ are identically distributed as $\mathbf{Z}(T)$ across all i and j, independent across the n subjects, but correlated within each subject.

2.2 Smooth Backfitting

To estimate the component functions μ_0 and μ_k in the longitudinal additive model (6), we need to develop a time-dynamic version of the smooth backfitting technique by Mammen et al. (1999). Throughout the paper we denote the joint density of $(T, \mathbf{Z}(T))$ by $p(t, \mathbf{z})$, and that of $(T, Z_k(T))$ by $p_k(t, z_k)$. Let $p(\cdot|t)$ and $p_k(\cdot|t)$ be the densities of $\mathbf{Z}(t)$ and $Z_k(t)$, respectively, and p_T be the density of T, then $p(t, \mathbf{z}) = p_T(t)p(\mathbf{z}|t)$ and $p_k(t, z_k) = p_T(t)p_k(z_k|t)$. Note that the additive components μ_k in model (6) are not identifiable, so we impose the following constraints:

$$\int \mu_k(t, z_k)p_k(z_k|t)\, dz_k = 0 \;\text{ for each } t, \quad 1 \le k \le d. \tag{7}$$

2.2.1 Objective function

Writing $Y^* = m(T, \mathbf{Z}(T)) + w(T) + e$, where w is the stochastic component of the response process Y defined right after (5) and e is the random noise defined shortly before (6), we identify the functions μ_k, $0 \le k \le d$, by minimizing $E[Y^* - \theta_0(T) - \sum_{k=1}^d \theta_k(T, Z_k(T))]^2$. This can be done by minimizing, for each t,

$$\begin{aligned} &E\left[\left(Y^* - \theta_0(T) - \sum_{k=1}^d \theta_k(T, Z_k(T))\right)^2 \middle| T = t\right] \qquad (8)\\ &= \int E\left[\left(Y^* - \theta_0(T) - \sum_{k=1}^d \theta_k(T, Z_k(T))\right)^2 \middle| T = t, \mathbf{Z}(T) = \mathbf{z}\right] p(\mathbf{z}|t)\, d\mathbf{z}. \end{aligned}$$

A kernel-type estimator of the conditional mean inside of the integration is given by

$$\hat{p}(t, \mathbf{z})^{-1} \cdot \frac{1}{\mathcal{N}_s} \sum_{i=1}^n \sum_{j=1}^{N_i} \left[Y_{ij} - \theta_0(T_{ij}) - \sum_{k=1}^d \theta_k(T_{ij}, Z_{ijk})\right]^2 K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}), \tag{9}$$

where $\mathcal{N}_s = \sum_{i=1}^n N_i$, $K_{\mathbf{h}}(\cdot;\cdot)$ is a multivariate kernel function to be defined below, and $\hat{p}(t, \mathbf{z})$ is the kernel density estimator of $p(t, \mathbf{z})$ defined by

$$\hat{p}(t, \mathbf{z}) = \mathcal{N}_s^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij}).$$

To define the multivariate kernel function $K_{\mathbf{h}}(\cdot;\cdot)$, let K denote a univariate kernel function, and for a bandwidth h define a boundary corrected kernel $K_h(\cdot;\cdot)$, as in Mammen et al. (1999) and Yu et al. (2008), by

$$K_h(u; v) = \frac{K((u - v)/h)/h}{\int_{\mathcal{I}}[K((x - v)/h)/h]dx} \cdot I(u, v \in \mathcal{I}),$$

where $\mathcal{I}$ is the support of the random variable that kernel smoothing is applied to. With a slight abuse of notation, we write $K_{\mathbf{h}}(t, \mathbf{z}; s, \mathbf{u}) = K_{h_0}(t; s) \prod_{k=1}^{d} K_{h_k}(z_k; u_k)$ for a bandwidth vector $\mathbf{h} = (h_0, h_1, \ldots, h_d)$. Also, we define $K_{h_0,h_k}(t, z_k; s, u_k) = \int K_{\mathbf{h}}(t, \mathbf{z}; s, \mathbf{u})\, d\mathbf{z}_{-k}$ which equals $K_{h_0}(t; s)K_{h_k}(z_k; u_k)$, where $\mathbf{z}_{-k}$ for a given vector denotes the vector $(z_1, \ldots, z_d)^\top$ without its kth component.

Next, we approximate the unknown functions θ_0 and θ_k in (9) locally by polynomials. For simplicity and inspired by its excellent practical and theoretical properties in various regression settings, we adopt linear polynomials and this leads to local linear smoothing. Thus, we approximate $\theta_0(T_{ij})$ and $\theta_k(T_{ij}, Z_{ijk})$ in the bracket at (9) by $\theta_{0,0}(t) + \theta_{0,1}(t)(T_{ij} - t)/h_0$ and $\theta_{k,0}(t, z_k) + \theta_{k,1}(t, z_k)(T_{ij} - t)/h_0 + \theta_{k,2}(t, z_k)(Z_{ijk} - z_k)/h_k$, respectively, where $\theta_{0,1}/h_0$ and $\theta_{k,1}/h_0$, respectively, are the (partial) derivatives of $\theta_{0,0}$ and $\theta_{k,0}$ with respect to the first argument, and $\theta_{k,2}/h_k$ are the partial derivatives of θ_k with respect to the second argument. These considerations lead to the following estimator of the left hand side of (8):

$$\hat{p}_T(t)^{-1} \int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[Y_{ij} - \theta_{0,0}(t) - \theta_{0,1}(t) \left(\frac{T_{ij} - t}{h_0} \right) - \sum_{k=1}^{d} \theta_{k,0}(t, z_k) \right.$$
$$\left. - \sum_{k=1}^{d} \theta_{k,1}(t, z_k) \left(\frac{T_{ij} - t}{h_0} \right) - \sum_{k=1}^{d} \theta_{k,2}(t, z_k) \left(\frac{Z_{ijk} - z_k}{h_k} \right) \right]^2 K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})\, d\mathbf{z} \quad (10)$$

where $\hat{p}_T$ is the kernel density estimator of p_T defined by

$$\hat{p}_T(t) = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_{h_0}(t; T_{ij}). \quad (11)$$

We consider minimizing, for each t, the integrated kernel-weighted squared error in (10) over the space of tuples of functions $\boldsymbol{\theta} = (\theta_{0,0}, \theta_{0,1}; \{\theta_{k,0}, \theta_{k,1}, \theta_{k,2}\}_{k=1}^{d})^\top$ such that θ_0 and $\theta_{0,1}$ are constants and $\theta_{k,0}$, $\theta_{k,1}$ and $\theta_{k,2}$ for $1 \le k \le d$ are univariate functions satisfying the constraints

$$\int \left[\theta_{k,0}(z_k)\hat{p}_k(z_k|t) + \theta_{k,1}(z_k)\hat{p}_{k,1}(z_k|t) + \theta_{k,2}(z_k)\hat{p}_{k,2}(z_k|t) \right] dz_k = 0. \quad (12)$$

For a given triplet $(\theta_{k,0}, \theta_{k,1}, \theta_{k,2})$ satisfying (7), the adjusted version $(\theta^*_{k,0}, \theta_{k,1}, \theta_{k,2})$ satisfies the constraint (12), where $\theta^*_{k,0}(z_k) = \theta_{k,0}(z_k) - \int [\theta_{k,0}(z_k)\hat{p}_k(z_k|t) + \theta_{k,1}(z_k)\hat{p}_{k,1}(z_k|t) + \theta_{k,2}(z_k)\hat{p}_{k,2}(z_k|t)]\, dz_k$. The constraints (12) are regarded as empirical versions of those in

(7) as we will detail in the next section. Here and below, define

$$\hat{p}_k(t, z_k) = \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk}), \tag{13}$$

$$\hat{p}_{k,1}(t, z_k) = \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\frac{T_{ij} - t}{h_0}\right) K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk}), \tag{14}$$

$$\hat{p}_{k,2}(t, z_k) = \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\frac{Z_{ijk} - z_k}{h_k}\right) K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk}) \tag{15}$$

and

$$\hat{p}_k(z_k|t) = \hat{p}_k(t, z_k)/\hat{p}_T(t) \quad \hat{p}_{k,1}(z_k|t) = \hat{p}_{k,1}(t, z_k)/\hat{p}_T(t) \quad \hat{p}_{k,2}(z_k|t) = \hat{p}_{k,2}(t, z_k)/\hat{p}_T(t) \tag{16}$$

where $\hat{p}_T(t)$ is defined in (11).

Minimization of (10) cannot be done inside of the integration with respect to $\mathbf{z}$ since it would not produce a solution with additive structure. Thus, integration with respect to $\mathbf{z}$ is the key element of the proposed method.

2.2.2 Solution

Let $(\hat{\mu}_{0,0}(t),\ \hat{\mu}_{0,1}(t);\ \{\hat{\mu}_{k,0}(t,\cdot), \hat{\mu}_{k,1}(t,\cdot), \hat{\mu}_{k,2}(t,\cdot)\}_{k=1}^{d})^\top = \hat{\boldsymbol{\theta}}$ where

$$\begin{aligned}\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \Bigg[Y_{ij} - \theta_{0,0}(t) - \theta_{0,1}(t)\left(\frac{T_{ij} - t}{h_0}\right) - \sum_{k=1}^{d} \theta_{k,0}(t, z_k) \\ - \sum_{k=1}^{d} \theta_{k,1}(t, z_k)\left(\frac{T_{ij} - t}{h_0}\right) - \sum_{k=1}^{d} \theta_{k,2}(t, z_k)\left(\frac{Z_{ijk} - z_k}{h_k}\right)\Bigg]^2 K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})\, d\mathbf{z},\end{aligned} \tag{17}$$

subject to the constraints (12). The estimators $\hat{\mu}_{0,0}$ and $\hat{\mu}_{0,1}$ are in fact the intercept and slope of the local linear smoothers obtained by regressing Y_{ij} onto T_{ij}. Thus, they can be obtained independently of $\hat{\mu}_{k,0}$, $\hat{\mu}_{k,1}$, $\hat{\mu}_{k,2}$. To see this, we note that, by the score equation of (17) and the constraints (12), $\hat{\mu}_{0,0}$ and $\hat{\mu}_{0,1}$ satisfy

$$\int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[Y_{ij} - \hat{\mu}_{0,0}(t) - \hat{\mu}_{0,1}(t)\left(\frac{T_{ij} - t}{h_0}\right)\right] K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})\, d\mathbf{z} = 0$$

$$\int \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[Y_{ij} - \hat{\mu}_{0,0}(t) - \hat{\mu}_{0,1}(t)\left(\frac{T_{ij} - t}{h_0}\right)\right]\left(\frac{T_{ij} - t}{h_0}\right) K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})\, d\mathbf{z} = 0$$

for all t. Since $\int K_{\mathbf{h}}(t, \mathbf{z}; T_{ij}, \mathbf{Z}_{ij})\, d\mathbf{z} = K_{h_0}(t; T_{ij})$, this implies

$$(\hat{\mu}_{0,0}(t), \hat{\mu}_{0,1}(t)) = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \frac{1}{\mathcal{N}_s} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[Y_{ij} - \beta_0(t) - \beta_1(t)\left(\frac{T_{ij} - t}{h_0}\right)\right]^2 K_{h_t}(t; T_{ij}). \tag{18}$$

To find $\hat{\mu}_{k,0}(t,\cdot)$, $\hat{\mu}_{k,1}(t,\cdot)$ and $\hat{\mu}_{k,2}(t,\cdot)$, we minimize

$$\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left[Y_{ij}-\hat{\mu}_{0,0}(t)-\hat{\mu}_{0,1}(t)\left(\frac{T_{ij}-t}{h_0}\right)-\sum_{k=1}^{d}\theta_{k,0}(z_k)\right.$$
$$\left.-\sum_{k=1}^{d}\theta_{k,1}(z_k)\left(\frac{T_{ij}-t}{h_0}\right)-\sum_{k=1}^{d}\theta_{k,2}(z_k)\left(\frac{Z_{ijk}-z_k}{h_k}\right)\right]^2 K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}.$$

Define $\mathbf{v}_{ij}(t,\mathbf{z}) = (1,(T_{ij}-t)/h_0,(Z_{ij1}-z_1)/h_1,\ldots,(Z_{ijd}-z_d)/h_d)^\top$ and $\mathbf{v}_{ijk}(t,z_k) = (1,(T_{ij}-t)/h_0,(Z_{ijk}-z_k)/h_k)^\top$ for $1\le k\le d$. Note that $\mathbf{v}_{ijk}(t,z_k)=\mathbf{J}_k^\top\mathbf{v}_{ij}(t,\mathbf{z})$, where

$$\mathbf{J}_k=(\mathbf{1}_1,\mathbf{1}_2,\mathbf{1}_{k+2}),\quad \mathbf{1}_l=(\mathbf{0}_{l-1}^\top,1,\mathbf{0}_{d+2-l}^\top)^\top, \tag{19}$$

and $\mathbf{0}_l$ denotes the zero-vector of dimension l with obvious meaning for $\mathbf{0}_0$. Also, define $\hat{\boldsymbol{\mu}}_k(t,z_k) = (\hat{\mu}_{k,0}(t,z_k),\hat{\mu}_{k,1}(t,z_k),\mathbf{0}_{k-1}^\top,\hat{\mu}_{k,2}(t,z_k),\mathbf{0}_{d-k}^\top)^\top$ and $\hat{\boldsymbol{\mu}}_0(t) = (\hat{\mu}_{0,0}(t),\hat{\mu}_{0,1}(t),\mathbf{0}_d^\top)^\top$. Likewise, for a given triplet of univariate functions $(\theta_{k,0},\theta_{k,1},\theta_{k,2})$, let $\boldsymbol{\theta}_k(z_k) = (\theta_{k,0}(z_k),\theta_{k,1}(z_k),\mathbf{0}_{k-1}^\top,\theta_{k,2}(z_k),\mathbf{0}_{d-k}^\top)^\top$. Then, the solutions $\hat{\boldsymbol{\mu}}_k(t,\cdot)=\hat{\boldsymbol{\theta}}_k$ for $1\le k\le d$, where

$$\{\hat{\boldsymbol{\theta}}_k\}_{k=1}^d=\underset{\{\boldsymbol{\theta}_k\}_{k=1}^d}{\operatorname{argmin}}\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left[Y_{ij}-\left(\hat{\boldsymbol{\mu}}_0(t)+\sum_{k=1}^{d}\boldsymbol{\theta}_k(z_k)\right)^\top\mathbf{v}_{ij}(t,\mathbf{z})\right]^2 K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}. \tag{20}$$

They satisfy

$$\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left[Y_{ij}-\left(\hat{\boldsymbol{\mu}}_0(t)+\sum_{k=1}^{d}\hat{\boldsymbol{\mu}}_k(t,z_k)\right)^\top\mathbf{v}_{ij}(t,\mathbf{z})\right]$$
$$\times\mathbf{v}_{ijk}(t,z_k)K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}_{-k}=\mathbf{0}_3,\quad 1\le k\le d. \tag{21}$$

Let $\hat{\mathbf{m}}_k(t,z_k)=(\hat{m}_{k,0}(t,z_k),\ \hat{m}_{k,1}(t,z_k),\ \mathbf{0}_{k-1}^\top,\hat{m}_{k,2}(t,z_k),\mathbf{0}_{d-k}^\top)^\top$, where $\hat{m}_{k,0}(t,z_k)$, $\hat{m}_{k,1}(t,z_k)$ and $\hat{m}_{k,2}(t,z_k)$ denote the local linear estimators obtained by regressing Y_{ij} onto (T_{ij},Z_{ijk}), that is,

$$\hat{\mathbf{m}}_k(t,z_k)=\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(Y_{ij}-v_{ijk}^\top\boldsymbol{\beta}\right)^2 K_{h_0,h_k}(t,z;T_{ij},Z_{ijk}). \tag{22}$$

They solve the system of the equations

$$\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(Y_{ij}-\hat{\mathbf{m}}_k(t,z_k)^\top\mathbf{v}_{ij}(t,\mathbf{z})\right)\mathbf{v}_{ijk}(t,z_k)K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk})=\mathbf{0}_3.$$

Since $\int K_h(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}_{-k}=K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk})$ and $\hat{\mathbf{m}}_k(t,z_k)^\top\mathbf{v}_{ij}(t,\mathbf{z})$ depends only on t and z_k, the system of equations (21) is equivalent to

$$\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ijk}(t,z_k)\mathbf{v}_{ij}(t,\mathbf{z})^\top\left[\hat{\mathbf{m}}_k(t,z_k)-\hat{\boldsymbol{\mu}}_0(t)-\sum_{l=1}^{d}\hat{\boldsymbol{\mu}}_l(t,z_l)\right]$$
$$\times K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}_{-k}=\mathbf{0}_3,\quad 1\le k\le d. \tag{23}$$

Define

$$\hat{\mathbf{M}}(t,\mathbf{z}) = \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ij}(t,\mathbf{z})\mathbf{v}_{ij}(t,\mathbf{z})^\top K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}), \tag{24}$$

and for $k \neq l$,

$$\hat{\mathbf{M}}_{kl}(t,z_k,z_l) = \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ijk}(t,z_k)\mathbf{v}_{ijl}(t,z_l)^\top K_{h_0,h_k,h_l}(t,z_k,z_l;T_{ij},Z_{ijk},Z_{ijl}), \tag{25}$$

$$\hat{\mathbf{M}}_{kk}(t,z_k) = \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ijk}(t,z_k)\mathbf{v}_{ijk}(t,z_k)^\top K_{h_0,h_k}(t,z_k;T_{ij},Z_{ijk}). \tag{26}$$

Then, the system of the equations (23) can be expressed as the following *smooth backfitting equation*: for $1 \leq k \leq d$,

$$\hat{\boldsymbol{\mu}}_k(t,z_k) = \hat{\mathbf{m}}_k(t,z_k) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{l\neq k}^{d}\int \hat{\mathbf{M}}_{kk}(t,z_k)^{-1}\hat{\mathbf{M}}_{kl}(t,z_k,z_l)\hat{\boldsymbol{\mu}}_l(t,z_l)\,dz_l, \tag{27}$$

where, with a slight abuse of notation, $\hat{\boldsymbol{\mu}}_k(t,z_k)$, $\hat{\mathbf{m}}_k(t,z_k)$ and $\hat{\boldsymbol{\mu}}_0(t)$ are the three-dimensional vector obtained by deleting $(d-1)$ zero entries in their definitions.

Projection Interpretation: We now discuss a projection interpretation of the smooth backfitting estimators $\hat{\boldsymbol{\mu}}_k(t,\cdot)$ and how the constraints (12) are motivated. Let $\hat{\mathbf{m}}(t,\mathbf{z})$ be the full-dimensional local linear estimator that solves

$$\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(Y_{ij} - \boldsymbol{\theta}^\top\mathbf{v}_{ij}(t,\mathbf{z})\right)\mathbf{v}_{ij}(t,\mathbf{z})K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij}) = 0, \tag{28}$$

for each fixed $(t,\mathbf{z})$. It can be shown that it satisfies

$$\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ijk}(t,z_k)\mathbf{v}_{ij}(t,\mathbf{z})^\top\left[\hat{\mathbf{m}}(t,\mathbf{z}) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{l=1}^{d}\hat{\boldsymbol{\mu}}_l(t,z_l)\right]K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}_{-k} = \mathbf{0}_3,$$

for all $1 \leq k \leq d$. This implies that $\hat{\boldsymbol{\mu}}_k(t,\cdot)$, for $1 \leq k \leq d$, minimizes

$$\begin{aligned}&\int \frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left[\left(\hat{\mathbf{m}}(t,\mathbf{z}) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{k=1}^{d}\boldsymbol{\theta}_k(z_k)\right)^\top \mathbf{v}_{ij}(t,\mathbf{z})\right]^2 K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\,d\mathbf{z}\\ &= \int\left(\hat{\mathbf{m}}(t,\mathbf{z}) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{k=1}^{d}\boldsymbol{\theta}_k(z_k)\right)^\top\hat{\mathbf{M}}(t,\mathbf{z})\left(\hat{\mathbf{m}}(t,\mathbf{z}) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{k=1}^{d}\boldsymbol{\theta}_k(z_k)\right)d\mathbf{z}\end{aligned} \tag{29}$$

over $\boldsymbol{\theta}_k$, where $\boldsymbol{\theta}_k$ take the form $\boldsymbol{\theta}_k(z_k) = (\theta_{k,0}(z_k), \theta_{k,1}(z_k), \mathbf{0}_{k-1}^\top, \theta_{k,2}(z_k), \mathbf{0}_{d-k}^\top)^\top$ for some univariate functions $\theta_{k,0}$, $\theta_{k,1}$ and $\theta_{k,2}$. $\hat{\mathbf{M}}(t,\mathbf{z})$ is defined in (24). This projection interpretation underscores the appeal of the objective function (8), as we illustrate below.

The full-dimensional estimator $\hat{\mathbf{m}}(t,\mathbf{z})$ aims at $\mathbf{m}(t,\mathbf{z}) = (m(t,\mathbf{z}), m_1(t,\mathbf{z}), m_{1,2}(t,\mathbf{z}), \ldots, m_{d,2}(t,\mathbf{z}))^\top$, where $m_1(t,\mathbf{z})/h_0$ is the partial derivative of $m(t,\mathbf{z})$ with respect to t, and $m_{k,2}(t,\mathbf{z})/h_k$ is its partial derivative with respect to z_k. From model (4), $m_1(t,\mathbf{z})/h_0 = \partial\mu_0(t)/\partial t + \sum_{k=1}^d \partial\mu_k(t,z_k)/\partial t$ and $m_{k,2}(t,\mathbf{z})/h_k = \partial\mu_k(t,z_k)/\partial z_k$. Since $E(Y|T = t, \mathbf{Z}(t) = \mathbf{z}) = m(t,\mathbf{z})$ and by (7), minimization of (8) is equivalent to minimization of

$$\int \left(\mathbf{m}(t,\mathbf{z}) - \boldsymbol{\mu}_0(t) - \sum_{k=1}^d \boldsymbol{\theta}_k(z_k)\right)^\top \mathbf{P}(t,\mathbf{z}) \left(\mathbf{m}(t,\mathbf{z}) - \boldsymbol{\mu}_0(t) - \sum_{k=1}^d \boldsymbol{\theta}_k(z_k)\right) d\mathbf{z}, \tag{30}$$

where

$$\mathbf{P}(t,\mathbf{z}) = p(\mathbf{z}|t)\mathbf{D}, \tag{31}$$

and $\mathbf{D}$ is a $d+2$ by $d+2$ diagonal matrix with the first diagonal entry 1 and the other diagonal entries $\int u^2 K(u) du$.

Thus, the true functions $\boldsymbol{\mu}_k(t,\cdot)$ are obtained by projecting the full-dimensional mean function $(\mathbf{m}(t,\cdot) - \boldsymbol{\mu}_0(t))$ onto the space of additive functions, $\mathcal{H}^t$, equipped with a norm $\|\cdot\|_{*(t)}$ defined by $\|\mathbf{f}\|^2_{*(t)} = \int \mathbf{f}(\mathbf{z})^\top \mathbf{P}(t,\mathbf{z})\mathbf{f}(\mathbf{z})\, d\mathbf{z}$, while their estimators $\hat{\boldsymbol{\mu}}_k(t,\cdot)$ are obtained by projecting the full-dimensional estimator $(\hat{\mathbf{m}}(t,\cdot) - \hat{\boldsymbol{\mu}}_0(t))$ onto the space of additive functions, $\mathcal{H}^t(\hat{\mathbf{M}})$, equipped with a norm $\|\cdot\|_{\hat{\mathbf{M}}(t)}$ defined by $\|\mathbf{f}\|^2_{\hat{\mathbf{M}}(t)} = \int \mathbf{f}(\mathbf{z})^\top \hat{\mathbf{M}}(t,\mathbf{z})\mathbf{f}(\mathbf{z})\, d\mathbf{z}$. For rigorous definitions of the two spaces, see (38) and (39) in the proof of Theorem 1 in Appendix.

Note that $\hat{\mathbf{M}}(t,\mathbf{z})$, $\hat{\mathbf{m}}(t,\mathbf{z})$ and $\hat{\boldsymbol{\mu}}_0(t)$ approximate, respectively, $p_T(t)\mathbf{P}(t,\mathbf{z})$, $\mathbf{m}(t,\mathbf{z})$ and $\boldsymbol{\mu}_0(t)$. Thus the projection induced in equation (29) is an empirical version of the projection induced in equation (30), since $p_T(t)$ does not involve z. Therefore $\hat{\boldsymbol{\mu}}_k(t,\cdot)$ should be an approximate of $\boldsymbol{\mu}_k(t,\cdot)$.

The constraints (12) are nothing else than the requirement that $\hat{\boldsymbol{\mu}}_k(t,\cdot)$ are perpendicular, in $\mathcal{H}^t(\hat{\mathbf{M}})$, to the vector $(1, \mathbf{0}_{d+1}^\top)^\top$. That is

$$\langle (1, \mathbf{0}_{d+1}^\top)^\top, \hat{\boldsymbol{\mu}}_k(t,\mathbf{z})\rangle_{\hat{\mathbf{M}}(t)} = \int (1, \mathbf{0}_{d+1}^\top)\hat{\mathbf{M}}(t,\mathbf{z})\hat{\boldsymbol{\mu}}_k(t,\mathbf{z})\, d\mathbf{z} = 0.$$

These are empirical versions of the constraints (7), which are equivalent to

$$\begin{aligned}\langle (1, \mathbf{0}_{d+1}^\top)^\top, \hat{\boldsymbol{\mu}}_k(t,\mathbf{z})\rangle_{*(t)} &= \int (1, \mathbf{0}_{d+1}^\top)\mathbf{P}(t,\mathbf{z})\boldsymbol{\mu}_k(t,z_k)\, d\mathbf{z} \\ &= \int (1, \mathbf{0}_{d+1}^\top)\boldsymbol{\mu}_k(t,z_k) p(\mathbf{z}|t)\, d\mathbf{z} = 0.\end{aligned}$$

2.2.3 Algorithm

Based on the system of the integral equations in (27), the smooth backfitting algorithm starts with the estimates, $\hat{\mathbf{m}}_k(t, z_k)$, $\hat{\boldsymbol{\mu}}_0(t)$, $\hat{p}_k(t, z_k)$, $\hat{p}_{k,1}(t, z_k)$, $\hat{p}_{k,2}(t, z_k)$, $\hat{\mathbf{M}}_{kk}(t, z_k)$ and $\hat{\mathbf{M}}_{kl}(t, z_k, z_l)$ for all k (see equations (22), (18), (11), (13), (24) and (25)) and then the following steps:

- *Step 1*: Set $r = 0$. Get initial estimates $\hat{\boldsymbol{\mu}}_k^{[0]}$ for all k satisfying the constraints (12). One may use the adjusted version of $\hat{\mathbf{m}}_k$ right after equation (12) for the initial $\hat{\boldsymbol{\mu}}_k^{[0]}$.

- *Step 2*: In the rth iteration, update $\hat{\boldsymbol{\mu}}_k^{[r-1]}$, for $1 \le k \le d$, by the formulas

$$\begin{aligned}\hat{\boldsymbol{\mu}}_k^{[r]}(t, z_k) &= \hat{\mathbf{m}}_k(t, z_k) - \hat{\boldsymbol{\mu}}_0(t) - \sum_{l<k}^{d} \int \hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \hat{\mathbf{M}}_{kl}(t, z_k, z_l) \hat{\boldsymbol{\mu}}_l^{[r]}(t, z_l)\, dz_l \\ &\quad - \sum_{l>k}^{d} \int \hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \hat{\mathbf{M}}_{kl}(t, z_k, z_l) \hat{\boldsymbol{\mu}}_l^{[r-1]}(t, z_l)\, dz_l.\end{aligned}$$

- *Step 3*: Repeat *Step 2* until the convergence criterion is satisfied. The convergence criterion is that, for all $1 \le k \le d$,

$$\frac{\int |\hat{\mu}_{k,0}^{[r]}(t, z_k) - \hat{\mu}_{k,0}^{[r-1]}(t, z_k)|^2\, dz_k\, dt}{\int |\hat{\mu}_{k,0}^{[r-1]}(t, z_k)|^2\, dz_k\, dt} < \epsilon$$

  for a pre-specified small positive real number ϵ.

We note that the formula in *Step 2* gives an updated $\hat{\boldsymbol{\mu}}_k^{[r]}$ that satisfies automatically the constraints (12), provided that the inputs $\hat{\boldsymbol{\mu}}_l^{[r]}$ for $l < k$ and $\hat{\boldsymbol{\mu}}_l^{[r-1]}$ for $l > k$ satisfy the constraints. This follows from the following three facts: (i) $(\hat{p}_k(t, z_k), \hat{p}_{k,1}(t, z_k), \hat{p}_{k,2}(t, z_k))^\top$ is the first row of the matrix $\hat{\mathbf{M}}_{kk}(t, z_k)$; (ii) the first entry of the vector $\int \hat{\mathbf{M}}_{kk}(t, z_k) \left[\hat{\mathbf{m}}_k(t, z_k) - \hat{\boldsymbol{\mu}}_0(t)\right] dz_k$ equals zero; (iii) the constraint (12) on $\hat{\boldsymbol{\mu}}_l$ is equivalent to requiring that the first entry of the vector $\int \hat{\mathbf{M}}_{kl}(t, z_k, z_l) \hat{\boldsymbol{\mu}}_l(t, z_l)\, dz_l$ equals zero. Thus, once the initial estimates satisfy the constraints, no adjustment is necessary in the subsequent steps.

3 Asymptotic Properties

In this section, we first show in Theorem 1 that the smooth backfitting algorithm converges to the solution of the smooth backfitting equation in the L^2 norm. Then, we give in Theorems 2 and 3 the respective uniform convergence rates of the stochastic and deterministic parts of the solution, which result in the oracle properties of the solution. For the theoretical results, we need Assumptions (A1)-(A8), which can be found in the Appendix. Below in the theorems, $\hat{\boldsymbol{\mu}}_k$ are of $(d+2)$-dimension, and $|\cdot|$ denotes the Euclidean norm.

Theorem 1. *Suppose that Assumptions (A1)-(A8) hold. With probability tending to one, there exists a unique solution of the smooth backfitting equation (27). Moreover, there exist constants* $0 < C < \infty$ *and* $0 < \gamma < 1$ *such that, with probability tending to one,*

$$\int \left|\hat{\boldsymbol{\mu}}_k^{[r]}(t, z_k) - \hat{\boldsymbol{\mu}}_k(t, z_k)\right|^2 p_k(t, z_k)\, dz_k\, dt \le C \cdot \Gamma \cdot \gamma^r, \tag{32}$$

where Γ *is the value of the squared* L^2 *norm of the starting estimators* $\hat{\boldsymbol{\mu}}_k^{[0]}$ *given by* $\Gamma = \sum_{k=1}^d \int |\hat{\boldsymbol{\mu}}_k^{[0]}(t, z_k)|^2 p_k(t, z_k)\, dz_k\, dt$.

The left hand side of (32) has integration with respect to t. In fact, (32) also holds pointwise in t: for each t, there exist constants $0 < C(t) < \infty$ and $0 < \gamma(t) < 1$ that depend on t such that, with probability tending to one, $\int \left|\hat{\boldsymbol{\mu}}_k^{[r]}(t, z_k) - \hat{\boldsymbol{\mu}}_k(t, z_k)\right|^2 p_k(t, z_k)\, dz_k \le C(t) \cdot \Gamma(t) \cdot \gamma(t)^r$, where $\Gamma(t) = \sum_{k=1}^d \int |\hat{\boldsymbol{\mu}}_k^{[0]}(t, z_k)|^2 p_k(t, z_k)\, d\mathbf{z}$. In the proof of Theorem 1, we show that $\gamma(t)$ as a function of t is continuous, which implies γ is bounded by an absolute positive constant, γ, which is strictly less than 1. This plays an important role in proving the uniform results in Theorems 2 and 3 below.

Now, we give some relevant approximations of the estimators $\hat{\boldsymbol{\mu}}_k$, $1 \le k \le d$. For this, we decompose $\hat{\boldsymbol{\mu}}_k$ into two parts, one attributed to δ_{ij} and the other to $m(T_{ij}, \mathbf{Z}_{ij})$ in $Y_{ij} = m(T_{ij}, \mathbf{Z}_{ij}) + \delta_{ij}$, where $\delta_{ij} = w_i(T_{ij}) + e_{ij}$ which are identically distributed for all i and j, independent across i, but may be correlated across j within the ith subject. Let $\hat{\boldsymbol{\mu}}_0^A$ and $\hat{\boldsymbol{\mu}}_0^B$ be, respectively, the version of $\hat{\boldsymbol{\mu}}_0$ obtained by substituting δ_{ij} and $m(T_{ij}, \mathbf{Z}_{ij})$ in place of Y_{ij} in (18). Likewise, define $\hat{\mathbf{m}}_k^A$ and $\hat{\mathbf{m}}_k^B$ to be respectively the corresponding version of $\hat{\mathbf{m}}_k$ with δ_{ij} and $m(T_{ij}, \mathbf{Z}_{ij})$ in place of Y_{ij} in (22). Next, we replace $\hat{\boldsymbol{\mu}}_0$ and $\hat{\mathbf{m}}_k$ in (27) by $\hat{\boldsymbol{\mu}}_0^A$ and $\hat{\mathbf{m}}_k^A$ respectively, and denote the new solution as $\hat{\boldsymbol{\mu}}_k^A$. Then $\hat{\boldsymbol{\mu}}_k^B$ can be defined likewise. Thus,

$$\hat{\boldsymbol{\mu}}_k(t, z_k) = \hat{\boldsymbol{\mu}}_k^A(t, z_k) + \hat{\boldsymbol{\mu}}_k^B(t, z_k), \qquad \hat{\mathbf{m}}_k(t, z_k) = \hat{\mathbf{m}}_k^A(t, z_k) + \hat{\mathbf{m}}_k^B(t, z_k). \tag{33}$$

The next theorem is for the approximation of $\hat{\boldsymbol{\mu}}_k^A$. To state the theorem, let $I_0 = [h_0, 1 - h_0]$ and $I_k = [\min \mathbb{Z}_k + h_k, \max \mathbb{Z}_k - h_k]$ where $\mathbb{Z}_k$ is the support of Z_k for $1 \le k \le d$. Let $\mathbb{Z} = \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_d$

Theorem 2. *Under the assumptions of Theorem 1,*

$$\sup_{t \in I_0, z_k \in I_k} \left|\hat{\boldsymbol{\mu}}_k^A(t, z_k) - \hat{\mathbf{m}}_k^A(t, z_k)\right| = o_p((nEN)^{-1/3}).$$

Next, we give a uniform approximation of $\hat{\boldsymbol{\mu}}_k^B$. Define

$$\boldsymbol{\mu}_k(t,z_k) = \left(\mu_k(t,z_k), h_0\partial\mu_k(t,z_k)/\partial t, \mathbf{0}_{k-1}^\top, h_k\partial\mu_k(t,z_k)/\partial z_k, \mathbf{0}_{d-k}^\top\right)^\top.$$

This is in fact what $\hat{\boldsymbol{\mu}}_k(t,z_k)$ aims at. Also, define

$$\beta_k(t,z_k) = \frac{1}{2}\left(\int u^2K(u)du\right)\left[\frac{\partial^2\mu_k(t,z_k)}{\partial t^2}h_0^2 + \frac{\partial^2\mu_k(t,z_k)}{\partial z_k^2}h_k^2\right] - c_k(t), \tag{34}$$

where $c_k(t)$ is chosen so that $\int \beta_k(t,z_k)p_k(z_k|t)\,dz_k = 0$. Let $\boldsymbol{\beta}_k(t,z_k) = \beta_k(t,z_k)\mathbf{1}_1$.

Theorem 3. *Under the assumptions of Theorem 1,*

$$\sup_{t\in I_0, z_k\in I_k}\left|\hat{\boldsymbol{\mu}}_k^B(t,z_k) - \boldsymbol{\mu}_k(t,z_k) - \boldsymbol{\beta}_k(t,z_k)\right| = o_p((nEN)^{-1/3}).$$

The vector of functions $\boldsymbol{\beta}_k(t,z_k)$ is the leading bias of $\hat{\boldsymbol{\mu}}_k(t,z_k)$ as an estimator of $\boldsymbol{\mu}_k(t,z_k)$. The theorem implies that the conditional bias of $\hat{\mu}_{k,0}(t,z_k)$ given $\{(T_{ij},\mathbf{Z}_{ij})\}$ equals $\beta_k(t,z_k) + o_p((nEN)^{-1/3})$. Since all entries of $\boldsymbol{\beta}_k$ except the first one are zero, the theorem also implies that the conditional biases of the corresponding estimators of $h_0\partial\mu_k(t,z_k)/\partial t$ and $h_k\partial\mu_k(t,z_k)/\partial z_k$ are of magnitude $o_p((nEN)^{-1/3})$. This means that the conditional biases of the estimators of the derivatives $\partial\mu_k(t,z_k)/\partial t$ and $\partial\mu_k(t,z_k)/\partial z_k$ are $o_p((nEN)^{-1/6})$. If one applies the local quadratic or even higher-order local polynomial smoothing, then one may derive the leading bias terms for the estimators of the derivatives as well, smoothness of $\mu_k(t,z_k)$ being permitted.

The preceding two theorems gives the following results which demonstrate the joint asymptotic distribution of $\hat{\mu}_{k,0}(t,z_k)$.

Theorem 4. *Suppose that $(nEN)^{1/6}h_k \to a_k$ for $0 \le k \le d$. Let t and z_k be fixed points in $(0,1)$ and $\mathbb{Z}_k^o$, for $1 \le k \le d$. Then, under the assumptions of Theorem 1,*

$$(nEN)^{1/3}\begin{bmatrix}\hat{\mu}_{1,0}(t,z_1) - \mu_1(t,z_1)\\ \vdots \\ \hat{\mu}_{d,0}(t,z_d) - \mu_d(t,z_d)\end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix}\beta_1^*(t,z_1)\\ \vdots\\ \beta_d^*(t,z_d)\end{bmatrix}, \begin{bmatrix} v_1(t,z_1) & \cdots & 0\\ 0 & \ddots & 0\\ 0 & \cdots & v_d(t,z_d)\end{bmatrix}\right),$$

where

$$\beta_k^*(t,z_k) = \frac{1}{2}\left(\int u^2K(u)du\right)\left[\frac{\partial^2\mu_k(t,z_k)}{\partial t^2}a_0^2 + \frac{\partial^2\mu_k(t,z_k)}{\partial z_k^2}a_k^2\right] - c_k^*(t),$$

$c_k^*(t)$ is chosen to satisfy $\int \beta_k^*(t,z_k)p_k(z_k|t)\,dz_k = 0$, and

$$v_k(t,z_k) = \int K^2(u)du\, Var(Y^*|T = t, Z_k(t) = z_k)\, p_k(t,z_k)^{-1}a_0^{-1}a_k^{-1}.$$

We can see from Theorem 4 that each estimator $\hat{\mu}_k(t,z_k)$ achieves the oracle asymptotic bias and variance. This means that it has the same asymptotic distribution of the local linear estimator which is based on the knowledge of all other components $\mu_0(t)$ and $\mu_l(t,z_l)$, $l \neq k$.

4 Bandwidth Choice

Bandwidth choice is an open question for functional data. **?** propose the proper way to perform cross validation by a leave-one-subject-out approach. This has been implemented satisfactorily in Jiang and Wang (2010) to estimate the mean function when there is only one covariate, so we recommend its use to estimate the overall mean function $\mu_0(t)$. As for the bandwidths of the component functions, however, leave-one-subject-out cross validation would be very time consuming, especially for the additive model (4) , when there are multiple bandwidths to be selected for the component function. A simple and appealing approach was proposed in for a varying coefficient model using a plug-in method based on a polynomial approximation of the second derivatives of the component functions. The situation is more complicated for our setting due to the bivariate nature of the component functions and the fact that the time component is involved in all components. However, a clever profile approach alleviates these difficulties and we establish in this section a new plug-in method for the choice of the bandwidths $h_k, k = 0, \ldots, d$ used in estimating the additive components. Numerical performance of this new bandwidth choice method is satisfactory as reflected in the simulation study in Section 5.

For compact output intervals, $\mathcal{T}$ for time t, $\mathcal{L}_k, k = 1, \ldots, d$ for covariates, the asymptotic mean integrated squared error (MISE) of $\hat{\mu}_k(t, z_k), 1 \le k \le d$ defined on those intervals is

$$\begin{aligned}\text{MISE}_k \;=\; & \int_{\mathcal{T}\times\mathcal{L}_k} \left[\frac{1}{2}\int u^2 K(u)du \left(\frac{\partial^2 \mu_k(t, z_k)}{\partial t^2} h_0^2 + \frac{\partial^2 \mu_k(t, z_k)}{\partial z_k^2} h_k^2\right)\right]^2 p_k(t, z_k) dz_k dt \\ & + \frac{\int K^2(u)du}{nENh_0h_k} \int_{\mathcal{T}\times\mathcal{L}_k} \text{Var}(Y^*|T = t, Z_k(t) = z_k) dz_k dt. \qquad (35)\end{aligned}$$

We would like to choose the optimal bandwidths $h_k, k = 0, \ldots, d$ which minimize $\sum_{k=1}^d \text{MISE}_k$. The basic idea is similar to profile methods: for a given h_0, choose $h_k(h_0)$, dependent on h_0, which minimizes MISE_k. Denote such minimum by $\text{MISE}_k(h_0)$. Then choose the optimal h_0 which minimizes $\sum_{k=1}^d \text{MISE}_k(h_0)$. The optimal $h_k, 1 \le k \le d$ are correspondingly obtained.

To calculate MISE_k, we need to estimate

$$D_k = \int_{\mathcal{T}\times\mathcal{L}_k} \left(\frac{\partial^2 \mu_k(t, z_k)}{\partial t^2}\right)^2 p_k(t, z_k) dz_k dt, \qquad E_k = \int_{\mathcal{T}\times\mathcal{L}_k} \left(\frac{\partial^2 \mu_k(t, z_k)}{\partial z_k^2}\right)^2 p_k(t, z_k) dz_k dt,$$

$$F_k = \int_{\mathcal{T}\times\mathcal{L}_k} \left(\frac{\partial^2 \mu_k(t, z_k)}{\partial t^2} \frac{\partial^2 \mu_k(t, z_k)}{\partial z_k^2}\right) p_k(t, z_k) dz_k dt, \qquad G_k(t, z_k) = \text{Var}(Y^*|T = t, Z_k(t) = z_k).$$

The second derivatives in D_k, E_k and F_k can be estimated by fitting a cubic polynomial on T and $\mathbf{Z}$. We first subtract the overall mean function obtained by fitting a quadratic polynomial on T and we have $\tilde{Y}_{ij} = Y_{ij} - (\hat{\alpha}_0 + \hat{\alpha}_1 T_{ij} + \hat{\alpha}_2 T_{ij}^2)$ where

$$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2) = \underset{\alpha_0,\alpha_1,\alpha_2}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(Y_{ij} - (\alpha_0 + \alpha_1 T_{ij} + \alpha_2 T_{ij}^2)\right)^2 .$$

Then D_k, E_k and F_k are estimated by

$$\hat{D}_k = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(2\hat{\beta}_2 + 6\hat{\beta}_3 T_{ij} + 2\hat{\beta}_{k,5} Z_{ijk}\right)^2 ,$$

$$\hat{E}_k = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(2\hat{\beta}_{k,2} + 6\hat{\beta}_{k,3} Z_{ijk} + 2\hat{\beta}_{k,6} T_{ij}\right)^2 ,$$

$$\hat{F}_k = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\hat{\beta}_2 + 6\hat{\beta}_3 T_{ij} + 2\hat{\beta}_{k,5} Z_{ijk}\right)\left(2\hat{\beta}_{k,2} + 6\hat{\beta}_{k,3} Z_{ijk} + 2\hat{\beta}_{k,6} T_{ij}\right),$$

where

$$(\{\hat{\beta}_l\}_{0\le l\le 3}, \{\hat{\beta}_{k,m}\}_{1\le k\le d, 1\le m\le 6}) = \underset{\beta_l,\beta_{k,m}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\tilde{Y}_{ij} - \sum_{l=0}^{3} \beta_l T_{ij}^l - \sum_{k=1}^{d} \Big[\beta_{k,1} Z_{ijk} + \beta_{k,2} Z_{ijk}^2 \right.$$
$$\left. + \beta_{k,3} Z_{ijk}^3 + \beta_{k,4} T_{ij} Z_{ijk} + \beta_{k,5} T_{ij}^2 Z_{ijk} + \beta_{k,6} T_{ij} Z_{ijk}^2\Big]\right)^2 .$$

Denote the residual, the quantity in the parenthesis above with the β's replaced by its estimates, as R_{ij}. We estimate G_k by

$$\hat{G}_k(t, z_k) = \hat{\gamma}_0 + \hat{\gamma}_1 T_{ij} + \hat{\gamma}_2 Z_{ijk},$$

where

$$(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2) = \underset{\gamma_0,\gamma_1,\gamma_2}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(R_{ij}^2 - \gamma_0 - \gamma_1 T_{ij} - \gamma_2 Z_{ijk}\right)^2 .$$

With $\hat{D}_k$, $\hat{E}_k$, $\hat{F}_k$ and $\hat{G}_k(t, z_k)$ plugged in (35), we obtain an approximation of MISE_k, $1 \le k \le d$ and the optimal bandwidths are correspondingly obtained from these approximates.

5 Simulation

Let $d = 2$, $t \in [0, 1]$, and

$$m(t, \mathbf{z}) = 1 + (t - 0.5)^2 + t z_1^2 + (1 - t) z_2^2.$$

We generate 500 datasets with $n = 100$ subjects in each dataset. N_i's are i.i.d. random variables from a discrete uniform distribution between 2 and 10. For the ith subject, T_{ij}, $j = 1, \ldots, N_i$ are i.i.d from uniform distribution on $[0, 1]$. Two longitudinal covariates are generated by

$$Z_{ij1} = 2T_{ij} - 0.5 + a_{i1} + b_{i1}T_{ij} + c_{ij1},$$

$$Z_{ij2} = 0.5 + 4(T_{ij} - 0.5)^3 + a_{i2} + b_{i2}T_{ij} + c_{ij2},$$

where $(a_{i1}, b_{i1}, a_{i2}, b_{i2})^\top$ are i.i.d. samples from a multivariate normal distribution with zero mean and covariance

$$\begin{bmatrix} 0.09 & -0.03 & 0.045 & -0.03 \\ -0.03 & 0.04 & -0.03 & 0.02 \\ 0.045 & -0.03 & 0.09 & -0.03 \\ -0.03 & 0.02 & -0.03 & 0.04 \end{bmatrix},$$

and c_{ij1} and c_{ij2} are i.i.d. from $N(0, 0.4^2)$ and both independent of $(\alpha_{i1}, \beta_{i1}, \alpha_{i2}, \beta_{i2})^\top$.

The resulting overall mean function is

$$\begin{aligned} \mu(t) &= 1 + (t - 0.5)^2 + t(2t - 0.5)^2 + (1 - t)[0.5 + 4(t - 0.5)^3]^2 \\ &+ 0.04t^2 - 0.06t + 0.25, \end{aligned}$$

and the component functions after taking into account the identification condition (7) are:

$$\begin{aligned} \mu_1(t, z_1) &= t\left[z_1^2 - 0.04t^2 + 0.06t - 0.25 - (2t - 0.5)^2\right], \\ \mu_2(t, z_2) &= (1 - t)\left[z_2^2 - 0.04t^2 - 0.06t - 0.25 - (0.5 + 4(t - 0.5)^3)^2\right]. \end{aligned}$$

The stochastic components w_{ij} are generated from $w_{ij} = w_i(T_{ij}) = \sum_{q=1}^4 A_{iq}\phi_q(T_{ij})$, where A_{iq} are independent random variables from $N(0, \lambda_q)$ and $\lambda_q = 1/(q + 1)^2$ for $q = 1, \ldots, 4$, and

$$\begin{aligned} \phi_1(t) = \sqrt{2}\cos(2\pi t), \quad \phi_2(t) = \sqrt{2}\sin(2\pi t), \\ \phi_3(t) = \sqrt{2}\cos(4\pi t), \quad \phi_4(t) = \sqrt{2}\sin(4\pi t). \end{aligned}$$

The observed data are $Y_{ij} = m(T_{ij}, \mathbf{Z}_{ij}) + w_{ij} + e_{ij}$, where e_{ij} are i.i.d. random noises from $N(0, 0.05^2)$.

For this simulation, a Gaussian kernel is used. The threshold ϵ for the convergence of the smooth backfitting algorithm is 10^{-4}. The summary statistics of the 500 triplets of the optimal bandwidths (h_0, h_1, h_2) using the plug-in method are shown in Table **??**. Individual histograms are in Figure **??**.

In addition to the plug-in method introduced in Section 4, we also use grid search to find the optimal bandwidths. Denote $|\mathcal{A}|$ as the range of an interval $\mathcal{A}$. Define $\text{AMISE} = \sum_{k=1}^{d} \text{AMISE}_k$ where

$$\text{AMISE}_k = |\mathcal{T}|^{-1}|\mathcal{L}_k|^{-1} \int_{\mathcal{T}\times\mathcal{L}_k} \left[\hat{\mu}_{k,0}(t, z_k) - \mu_k(t, z_k)\right]^2 dz_k dt.$$

By grid search, the optimal bandwidths which minimizes AMISE are $(h_0, h_1, h_2) = (0.1653, 0.6065, 0.7408)$ with $\text{AMISE}_1 = 0.0836$, $\text{AMISE}_2 = 0.0338$ and the total $\text{AMISE} = 0.1174$. They are fairly close to the results from the plug-in method which has the corresponding $\text{AMISE} = 0.2220$ with $\text{AMISE}_1 = 0.1846$ and $\text{AMISE}_2 = 0.0373$. This suggests that the new plug-in method works well. The estimated $\mu_1(t, z_1)$ and $\mu_2(t, z_2)$ by both grid search and the plug-in method are compared with the true functions in Figure **??**.

6 Application

The methodology in Section 2 is implemented to the CD4 data from the Multicenter AIDS Cohort Study or MACS (S.Chmiel et al. (1987)). CD4 cells are important to immune system and often decline when a patient is HIV infected. Thus the amount of CD4 cells is an important index to monitor AIDS progression.

For this data, there are 2376 observations from 369 men infected with HIV. The time origin of each subject is the time of seroconversion. Several factors may be related to the number of CD4 cells, such as age at seroconversion and the level of depression, which we adopt as two covariates and denote respectively by Z_1 and Z_2. Here age, Z_1, is time-invariant while depression scores, Z_2, are recoded over time (in years) and is thus longitudinal. Depression is measured by the CESD scale: higher CESD score indicates more severe depression.

The bandwidths obtained by the plug-in method are $(h_0, h_1, h_2) = (0.95, 7.3144, 17.1198)$. Figure **??** shows that on the average the amount of CD4 cells started to decline two years before seroconversion and then decline sharply around seroconvergence time until one year later when the decline continues at a slower rate. For the dynamic component function involving age, the left panel of Figure **??** reveals a strong time dynamic feature of the effect of age, suggesting a lack of fit for the simpler model (3) adopted in the literature. Here we adopt a reversed time scale in Figure **??** for better graphical effect. After seroconversion (time 0), young people's CD4 counts decrease slower than middle-aged men. One year after seroconversion, CD4 counts of older people decline faster than younger and middle-age people. The right panel of Figure **??** suggests that CD 4 counts decline at a faster pace for subjects that are more depressed.

7 Discussion and Conclusion

The model proposed in the paper not only enjoys the flexibility of modeling but also captures the time-dynamic features of longitudinal data. The estimates for component functions in terms of local linear regression and smooth backfitting algorithm can achieve the oracle bias and variance. The optimal rate of convergence is $(nEN)^{1/3}$ which is the same as that of a two dimensional local linear smoother for independent response data. A practical and simple plug-in method for bandwidth choice is provided and its satisfactory performance is reflected in the simulation.

This paper mainly focuses on fitting the mean function. In fact, functional principal component analysis (FPCA) can be conducted afterwards to estimate eigenvalues and eigenfunctions. Due to multiple covariates in additive models, mean-adjusted FPCA (mFPCA) proposed by Jiang and Wang (2010) is a feasible approach.

Model (4) is more general than Model (2) and thus can be used for model checking. The CD4 count data demonstrate this and suggest that existing longitudinal additive model in the literature, i.e. model (2), does not fit this data. However, if the estimates of some component functions reveal little dependency on time, Model (4) can be simplified to model (2) such that each of those components is a one dimensional function depending on a single covariate. A worthy future direction is to derive inference procedure for such a model checking approach.

8 Appendix

8.1 Assumptions and Basic Lemma

Below we collect the assumptions we use to prove the theorems.

(A1) The density function $p(t, \mathbf{z})$ of $(T, \mathbf{Z}(T))$ is supported on $[0, 1] \times \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_d$, continuous and bounded away from zero and infinity.

(A2) The kernel K is symmetric and Lipschitz continuous.

(A3) The component functions $\mu_0(t)$ and $\mu_k(t, z_k)$ for $1 \le k \le d$ are twice partially continuously differentiable on $[0, 1]$ and $[0, 1] \times \mathbb{Z}_k$, respectively.

(A4) N_i, $1 \le i \le n$, are i.i.d. copies of N, which may depend on n.

(A5) Ee^N is bounded.

(A6) The bandwidths h_k for $0 \le k \le d$ are asymptotic to $(nEN)^{-1/6}$.

(A7) The conditional second moment $E[Y^2 \,|T = t, \mathbf{Z}(t) = \mathbf{z}]$ is continuous in t and $\mathbf{z}$ on $[0, 1] \times \mathbb{Z}$.

(A8) $E|\delta_{11}|^k < \infty$ for $k > 3$.

(A1), (A2) and (A3) are typical assumptions in the smoothing literature. (A4) is used by Yao et al. (2005) and Jiang and Wang (2010). (A5) is useful to control covariance within subject in the basic lemma below. (A6) assumes that the bandwidths have the optimal asymptotic rate for two-dimensional smoothing, which is also assumed in Jiang and Wang (2010). (A7) and (A8) are both useful in the lemma, to control the main deviation and the tail respectively.

The following lemma is useful for the proof of the theorems. It is for the uniform convergence of the kernel weighted average

$$W_n(t, z_k) = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk})\delta_{ij}.$$

Typically, this kind of results can be derived for independent data by applying an exponential inequality. A little complication arises in our setting since N_i are random variables and δ_{ij} are dependent within subjects. Recall that the conditional mean $E(\delta_{ij} \,|\, T_{ij}, \mathbf{Z}_{ij})$ equals zero. Thus $EW_n(t, z_k) = 0$.

Lemma 1. *Assume that (A1)-(A8) hold. Then,*

$$\sup_{t\in[0,1],z_k\in\mathbb{Z}_k} |\, W_n(t, z_k)\,| = O_p((nEN)^{-1/3} \sqrt{\log(nEN)}).$$

Proof. We sketch the proof. Write $K_{ijk} = K_{h_0,h_k}(t, z_k; T_{ij}, Z_{ijk})$ and let $a_n = nEN$. By Markov inequality and conditioning arguments, it follows that, for a small positive number α,

$$\begin{aligned} P\left[|\delta_{ij}| \le a_n^{1/3-\alpha} \text{ for all } i, j\right] &\ge E\left[1 - \sum_{i=1}^{n} \sum_{j=1}^{N_i} P(|\delta_{ij}| > a_n^{1/3-\alpha})\right] \\ &= 1 - a_n P(|\delta_{11}| > a_n^{1/3-\alpha}) \\ &\ge 1 - (\text{const})a_n^{1+k\alpha-k/3}, \end{aligned}$$

where k is the number in Assumption (A8). Thus, if we take $0 < \alpha < 1/3 - 1/k$, then with probability tending to one, $|\delta_{ij}|$ are bounded by $a_n^{1/3-\alpha}$ for all i, j. With (A8), we can also prove, by conditioning arguments again, that $Ea_n^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} K_{ijk}\delta_{ij} I(|\delta_{ij}| > a_n^{1/3-\alpha}) = o(a_n^{-1/3})$ uniformly for $t, z_k \in [0,1] \times \mathbb{Z}_k$. These considerations and Lipschitz continuity of the kernel K give the lemma if we prove

$$\sup_{t \in [0,1], z_k \in \mathbb{Z}_k} P\left[a_n^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \xi_{ij}(t, z_k) > C a_n^{-1/3} \sqrt{\log a_n}\right] \le a_n^{-\rho(C)}, \tag{36}$$

for some function ρ such that $\rho(C) \to \infty$ as $C \to \infty$, where $\xi_{ij}(t, z_k) = K_{ijk}\delta_{ij} I(|\delta_{ij}| \le a_n^{1/3-\alpha}) - EK_{ijk}\delta_{ij} I(|\delta_{ij}| \le a_n^{1/3-\alpha})$. This can be shown by applying Markov inequality again as the left hand side of (36) is bounded by

$$\begin{aligned} & a_n^{-C}\left[1 + \frac{1}{2} a_n^{-4/3} (\log a_n) E\left[e^{c_1 N a_n^{-\alpha} (\log a_n)^{1/2}} \Big(\sum_{j=1}^N \xi_{1j}(t, z_k)\Big)^2\right]\right]^n \\ \le\ & a_n^{-C}\left[1 + \frac{1}{2} a_n^{-4/3} (\log a_n) E\, N^2 e^{c_1 N a_n^{-\alpha} (\log a_n)^{1/2}} E\xi_{11}(t, z_k)^2\right]^n \\ \le\ & a_n^{c_2 - C}, \end{aligned} \tag{37}$$

where c_i for $i = 1, 2$ are some absolute constants. The second inequality above follows from (i) $(1 + x)^n \le e^{nx}$ for all $x > -1$; (ii) $a_n^{-1/3} \sup_{t \in [0,1], z_k \in \mathbb{Z}_k} E\xi_{11}(t, z_k)^2$ is bounded; (iii) $E\, N^2 e^{c_1 N a_n^{-\alpha} (\log a_n)^{1/2}}$ is bounded. Note that (ii) holds by (A7) Also, (iii) follows from (A5). □

8.2 Proof of Theorem 1

We first define explicitly $\mathcal{H}^t$ and $\mathcal{H}^t(\hat{\mathbf{M}})$ mentioned in Projection Interpretation in Section 2.2.2. For $\mathbf{P}(t, \mathbf{z})$ in (31), define

$$\begin{aligned} \mathcal{H}^t \ =\ & \{\mathbf{f}(\mathbf{z}) = (f(\mathbf{z}), f_1(\mathbf{z}), f_{1,2}(z_1), \ldots, f_{d,2}(z_d))^\top : f(\mathbf{z}) = \sum_{k=1}^d f_k(z_k), \\ & f_1(\mathbf{z}) = \sum_{k=1}^d f_{k,1}(z_k), \|\mathbf{f}\|_{*(t)}^2 = \int \mathbf{f}(\mathbf{z})^\top \mathbf{P}(t, \mathbf{z}) \mathbf{f}(\mathbf{z}) d\mathbf{z} < \infty\}, \end{aligned} \tag{38}$$

and the inner product on $\mathcal{H}^t$ is defined as, for $\mathbf{f}, \mathbf{g} \in \mathcal{H}^t$,

$$\langle \mathbf{f}, \mathbf{g} \rangle_{*(t)} = \int \mathbf{f}(\mathbf{z})^\top \mathbf{P}(t, \mathbf{z}) \mathbf{g}(\mathbf{z}) d\mathbf{z}.$$

For $\hat{\mathbf{M}}(t,\mathbf{z})$ in (24), define

$$\mathcal{H}^t(\hat{\mathbf{M}}) = \{\mathbf{f}(\mathbf{z}) = (f(\mathbf{z}), f_1(\mathbf{z}), f_{1,2}(z_1), \ldots, f_{d,2}(z_d))^\top : f(\mathbf{z}) = \sum_{k=1}^{d} f_k(z_k),$$
$$f_1(\mathbf{z}) = \sum_{k=1}^{d} f_{k,1}(z_k), \|\mathbf{f}\|^2_{\hat{\mathbf{M}}(t)} = \int \mathbf{f}(\mathbf{z})^\top \hat{\mathbf{M}}(t,\mathbf{z})\mathbf{f}(\mathbf{z})d\mathbf{z} < \infty\}. \quad (39)$$

and the inner product on $\mathcal{H}^t(\hat{\mathbf{M}})$ is defined as, for $\mathbf{f}, \mathbf{g} \in \mathcal{H}^t(\hat{\mathbf{M}})$,

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\hat{\mathbf{M}}(t)} = \int \mathbf{f}(\mathbf{z})^\top \hat{\mathbf{M}}(t,\mathbf{z})\mathbf{g}(\mathbf{z})d\mathbf{z}.$$

Recall from Section 2.2.2 the $(d+2)$-dimensional vectors of the function estimators $\hat{\mathbf{m}}$, $\hat{\boldsymbol{\mu}}_k$ for $0 \le k \le d$ and $\hat{\mathbf{m}}_k$ for $1 \le k \le d$ are respectively defined in (28), (18), (20) and (22). As we discussed there, taking the constraints (12) into consideration, $\sum_{k=1}^d \hat{\boldsymbol{\mu}}_k(t,\cdot)$ can be regarded as the projection of $(\hat{\mathbf{m}}(t,\cdot) - \hat{\boldsymbol{\mu}}_0(t))$ onto a space of additive functions $\mathcal{H}^{0t}(\hat{\mathbf{M}})$ defined as

$$\mathcal{H}^{0t}(\hat{\mathbf{M}}) = \{\mathbf{f} \in \mathcal{H}^t(\hat{\mathbf{M}}) : \int (1, \mathbf{0}_{d+1}^\top)\hat{\mathbf{M}}(t,\mathbf{z})\mathbf{f}_k(\mathbf{z})\, d\mathbf{z} = 0.\}$$

where $\mathbf{f}_k = (f_k, f_{k,1}, \mathbf{0}_{k-1}^\top, f_{k,2}, \mathbf{0}_{d-k}^\top)^\top$.

To treat the individual components $\hat{\boldsymbol{\mu}}_k$, we consider the following closed subspaces of $\mathcal{H}^{0t}(\hat{\mathbf{M}})$: for $1 \le k \le d$

$$\mathcal{H}_k^{0t}(\hat{\mathbf{M}}) = \{\mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}}) : \mathbf{f}(\mathbf{z}) = (f(z_k), f_1(z_k), \mathbf{0}_{k-1}^\top, f_2(z_k), \mathbf{0}_{d-k}^\top)^\top\}.$$

Thus, $\mathcal{H}^{0t}(\hat{\mathbf{M}}) = \mathcal{H}_1^{0t}(\hat{\mathbf{M}}) + \cdots + \mathcal{H}_d^{0t}(\hat{\mathbf{M}})$. Define $\mathcal{H}^{0t}$ and $\mathcal{H}_k^{0t}$, likewise, from the space $\mathcal{H}^t$ equipped with the norm $\|\cdot\|_{*(t)}$. Let $\hat{\Pi}_k^t$ denote the projection operator onto $\mathcal{H}_k^{0t}(\hat{\mathbf{M}})$. Likewise, define Π_k^t for $\mathcal{H}_k^{0t}$. It can be verified that $\hat{\Pi}_k^t(\hat{\mathbf{m}}(t,\cdot)) = \hat{\Pi}_k^t(\hat{\mathbf{m}}(t,\cdot) - \hat{\boldsymbol{\mu}}_0(t)) = \hat{\mathbf{m}}_k(t,\cdot) - \hat{\boldsymbol{\mu}}_0(t)$. Put $\hat{\Phi}^t = (I - \hat{\Pi}_d^t) \times \cdots \times (I - \hat{\Pi}_1^t)$ and $\Phi^t = (I - \Pi_d^t) \times \cdots \times (I - \Pi_1^t)$.

As in the proof of Theorem 1 in Mammen et al. (1999), we can verify that there exists a function γ such that $0 < \gamma(t) < 1$ and

$$\sup\{\|\Phi^t\mathbf{f}\|^2_{*(t)} : \mathbf{f} \in \mathcal{H}^{0t} \text{ and } \|\mathbf{f}\|^2_{*(t)} \le 1\} \le \gamma(t)$$

for all $t \in [0,1]$. To prove the theorem, we only show:

(i) γ is bounded on $[0,1]$;

(ii) $\sup_{t\in I_0} \sup\{\|(\hat{\Phi}^t - \Phi^t)\mathbf{f}\|^2_{*(t)} : \mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}}) \text{ and } \|\mathbf{f}\|^2_{*(t)} \le 1\} \xrightarrow{p} 0$;

(iii) $\sup_{t\in[0,1]} \sup\{\|(\hat{\Phi}^t - \Phi^t)\mathbf{f}\|^2_{*(t)} : \mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}}) \text{ and } \|\mathbf{f}\|^2_{*(t)} \leq 1\}$ is bounded in probability.

The other details of the proof can be done using standard theory of local polynomial smoothing and projection as in Mammen et al. (1999).

To prove (i), we first note that

$$\gamma(t) = 1 - \prod_{k=1}^{d-1} \sin^2(a_k(t))$$

where $a_k(t)$ is the minimal angle between the two subspaces $\mathcal{H}_k^{0t}$ and $S_k^{0t} \equiv \mathcal{H}_{k+1}^{0t} + \cdots + \mathcal{H}_d^{0t}$, that is

$$\cos(a_k(t)) = \sup_{(\mathbf{f},\mathbf{g})\in L(t)} \frac{\langle \mathbf{f}, \mathbf{g}\rangle_{*(t)}}{\|\mathbf{f}\|_{*(t)} \cdot \|\mathbf{g}\|_{*(t)}},$$

where $L(t) = \{(\mathbf{f}, \mathbf{g}) : \mathbf{f} \in \mathcal{H}_k^{0t} \cap (\mathcal{H}_k^{0t} \cap S_k^{0t})^{\perp},\ \mathbf{g} \in S_k^{0t} \cap (\mathcal{H}_k^{0t} \cap S_k^{0t})^{\perp}\}$.

It is sufficient to prove that $\cos(a_k(t))$ is continuous in t, which implies (i). In fact, we first note that, if $\mathbf{f} \in \mathcal{H}_k^{0t}$ for some t, then $\mathbf{f} \in \mathcal{H}_k^{0t'}$ for all other t'. This follows from the fact that the density ratio $p(t', \mathbf{z})/p(t, \mathbf{z})$ is bounded in t and $\mathbf{z}$. Also, we note that $\mathcal{H}_k^{0t} \cap S_k^{0t} = \{\mathbf{0}_{d+2}\}$. Thus, the set $L(t)$ actually does not depend on t. Let us call it now L. Let t_0 be an arbitrary point in $[0, 1]$. We get

$$\begin{aligned}
&|\cos(a_k(t)) - \cos(a_k(t_0))| \\
&\leq \sup_{(\mathbf{f},\mathbf{g})\in L} \left| \int \mathbf{f}(\mathbf{z})^\top \mathbf{D}\mathbf{g}(\mathbf{z}) \left(\frac{p(\mathbf{z}|t)}{\|\mathbf{f}\|_{*(t)}\|\mathbf{g}\|_{*(t)}} - \frac{p(\mathbf{z}|t_0)}{\|\mathbf{f}\|_{*(t_0)}\|\mathbf{g}\|_{*(t_0)}} \right) d\mathbf{z} \right| \\
&= \sup_{(\mathbf{f},\mathbf{g})\in L} \left| \int \frac{\mathbf{f}(\mathbf{z})^\top \mathbf{D}\mathbf{g}(\mathbf{z})}{\|\mathbf{f}\|_{*(t_0)}\|\mathbf{g}\|_{*(t_0)}} p(\mathbf{z}|t_0) \left[\left(\frac{\|\mathbf{f}\|_{*(t_0)}\|\mathbf{g}\|_{*(t_0)}}{\|\mathbf{f}\|_{*(t)}\|\mathbf{g}\|_{*(t)}} - 1 \right) \left(\frac{p(\mathbf{z}|t)}{p(\mathbf{z}|t_0)} - 1 \right) \right.\right. \\
&\qquad \left.\left. + \left(\frac{\|\mathbf{f}\|_{*(t_0)}\|\mathbf{g}\|_{*(t_0)}}{\|\mathbf{f}\|_{*(t)}\|\mathbf{g}\|_{*(t)}} - 1 \right) + \left(\frac{p(\mathbf{z}|t)}{p(\mathbf{z}|t_0)} - 1 \right) \right] d\mathbf{z} \right|, \qquad (40)
\end{aligned}$$

where $\mathbf{D}$ is defined in (31). Since $p(\mathbf{z}|t)/p(\mathbf{z}|t_0) \to 1$ as $t \to t_0$ uniformly for $\mathbf{z}$, it follows that

$$\sup_{(\mathbf{f},\mathbf{g})\in L} \left| \frac{\|\mathbf{f}\|_{*(t_0)}\|\mathbf{g}\|_{*(t_0)}}{\|\mathbf{f}\|_{*(t)}\|\mathbf{g}\|_{*(t)}} - 1 \right| \to 0 \text{ as } t \to t_0,$$

and consequently the right hand side of (40) tends to zero as $t \to t_0$. Therefore, $\cos(a_k(t))$ is continuous in t and thus (i) holds.

To prove (ii) and (iii), it suffices to show that for each $1 \leq k \leq d$

$$\sup_{t\in I_0} \sup\{\|(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}\|^2_{*(t)} : \mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}}) \text{ and } \|\mathbf{f}\|^2_{*(t)} \leq 1\} = o_p(1), \qquad (41)$$

$$\sup_{t\in[0,1]} \sup\{\|(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}\|^2_{*(t)} : \mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}}) \text{ and } \|\mathbf{f}\|^2_{*(t)} \leq 1\} = O_p(1). \qquad (42)$$

We first note that, for $\mathbf{f} \in \mathcal{H}^{0t}(\hat{\mathbf{M}})$,

$$\hat{\Pi}_k^t(\mathbf{f}) = \mathbf{f}_k + \sum_{l \neq k} \int \hat{\mathbf{M}}_{kk}(t, \cdot)^{-1} \hat{\mathbf{M}}_{kl}(t, \cdot, z_l) \mathbf{f}_l(z_l)\, dz_l.$$

On the left hand side $\mathbf{f} = (\sum_{l=1}^d f_l, \sum_{l=1}^d f_{l,1}, f_{1,2}, \ldots, f_{d,2})^\top$ is $(d+2)$−dimensional while on the right hand side $\mathbf{f}_l = (f_l, f_{l,1}, f_{l,2})^\top$ for $1 \leq l \leq d$ are three-dimensional. $\hat{\Pi}_k^t(\mathbf{f})$ is considered to be the three-dimensional representation of the corresponding $(d+2)$−dimensional vector of functions in $\mathcal{H}_k^{0t}(\hat{\mathbf{M}})$. For the projection Π_k^t we have the representation

$$\Pi_k^t(\mathbf{f}) = \mathbf{f}_k + \sum_{l \neq k} \int \mathbf{J} \cdot \mathbf{f}_l(z_l) \frac{p_{kl}(\cdot, z_l|t)}{p_k(\cdot|t)}\, dz_l,$$

where $\mathbf{J}$ is a 3×3 diagonal matrix with entries $(1, 1, 0)$. This implies $\mathbf{J} \cdot \mathbf{f}_l(z_l) = (f_l(z_l), f_{l,1}(z_l), 0)^\top$. Using Minkowski and Cauchy-Schwarz inequalities, we can prove for all $\mathbf{f} = \mathbf{f}_1 + \cdots + \mathbf{f}_d \in \mathcal{H}^{0t}(\hat{\mathbf{M}})$ that

$$\|(\hat{\Pi}_k^t - \Pi_k^t)\mathbf{f}\|_{*(t)} \leq \sum_{l \neq k} \|\mathbf{f}_l\|_{*(t)} \left[\int \|\mathbf{Q}_{kl}(t, z_k, z_l)\|_{HS}^2\, p_k(z_k|t) p_l(z_l|t)\, dz_k\, dz_l \right]^{1/2},$$

where $\|\mathbf{A}\|_{HS}$ for a matrix $\mathbf{A}$ denotes the Hilbert-Schmidt norm of $\mathbf{A}$ and

$$\mathbf{Q}_{kl}(t, z_k, z_l) = \frac{\hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \hat{\mathbf{M}}_{kl}(t, z_k, z_l)}{p_l(z_l|t)} - \mathbf{J} \frac{p_{kl}(z_k, z_l|t)}{p_k(z_k|t) p_l(z_l|t)}. \tag{43}$$

By the standard theory of kernel smoothing (standard method of local polynomial smoothing?) and a three-dimensional extension of Lemma 1 with δ_{ij} replaced by 1 , one can verify

$$\sup_{t \in I_0, z_k \in I_k, z_l \in I_l} |\mathbf{Q}_{kl}(t, z_k, z_l)| = o_p(1), \qquad \sup_{t \in [0,1], z_k \in \mathbb{Z}_k, z_l \in \mathbb{Z}_l} |\mathbf{Q}_{kl}(t, z_k, z_l)| = O_p(1). \tag{44}$$

This establishes (41) and (42).

8.3 Proof of Theorem 2

Let $\Delta_{1n} = (nEN)^{-1/3}$. Arguing as in the proof of Theorem 2 in Mammen et al. (1999), the theorem follows if we prove

$$\sup_{t \in I_0, z_k \in I_k} \left| \int \hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \hat{\mathbf{M}}_{kl}(t, z_k, z_l) \hat{\mathbf{m}}_l^A(t, z_l)\, dz_l \right| = o_p(\Delta_{1n}), \tag{45}$$

$$\sup_{t \in I_0} \int \left| \int \hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \hat{\mathbf{M}}_{kl}(t, z_k, z_l) \hat{\mathbf{m}}_l^A(t, z_l)\, dz_l \right|^2 p(z_k|t)\, dz_k = o_p(\Delta_{1n}^2), \tag{46}$$

where $\hat{\mathbf{m}}_l^A = (\hat{m}_{l,0}^A, \hat{m}_{l,1}^A, \hat{m}_{l,2}^A)^\top$ on the left hand sides denote the three-dimensional versions of their original definitions given in (33).The integral on the left hand side of (45) equals

$$\mathbf{r}_{n1}(t, z_k) + \mathbf{r}_{n2}(t, z_k) \equiv \int \mathbf{Q}_{kl}(t, z_k, z_l)\hat{\mathbf{m}}_l^A(t, z_l)p_l(z_l|t)\, dz_l + \int \mathbf{J} \cdot \hat{\mathbf{m}}_l^A(t, z_l)\frac{p_{kl}(z_k, z_l|t)}{p_k(z_k|t)}\, dz_l,$$

where $\mathbf{Q}_{kl}(t, z_k, z_l)$ is defined in (43) and $\mathbf{J}$ is a 3×3 diagonal matrix with entries $(1, 1, 0)$.

Let $\Delta_{2n} = (nEN)^{-1/3}\sqrt{\log(nEN)}$ and $\Delta_{3n} = (nEN)^{-5/12}\sqrt{\log(nEN)}$. By Lemma 1, it follows that $\sup_{t,\in[0,1],z_l\in\mathbb{Z}_l} |\hat{\mathbf{m}}_l^A(t, z_l)| = O_p(\Delta_{2n})$. This together with (44) gives

$$\sup_{t\in I_0, z_k\in I_k} |\mathbf{r}_{n1}(t, z_k)| = o_p(\Delta_{1n}), \quad \sup_{t\in I_0} \int |\mathbf{r}_{n1}(t, z_k)|^2 p(z_k|t)\, dz_k = o_p(\Delta_{1n}^2). \qquad (47)$$

To treat $\mathbf{r}_{n2}$, define $\mathbf{r}_{n2,a}(t, z_k)$ for $a = 1, 2$ by

$$\mathbf{r}_{n2,a}(t, z_k) = \mathcal{N}_s^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_{h_0}(t; T_{ij})\delta_{ij} \int \mathbf{R}_a(t, z_l)\mathbf{v}_{ijl}(t, z_l)K_{h_l}(z_l,\ Z_{ijl})\frac{p_{kl}(z_k, z_l|t)}{p_k(z_k|t)}\, dz_l,$$

where $\mathbf{R}_1(t, z_l) = \mathbf{J}(\hat{\mathbf{M}}_{ll}^{-1}(t, z_l) - \mathbf{J}_*^{-1} p_l(z_l|t)^{-1})$, $\mathbf{R}_2(t, z_l) = \mathbf{J}\mathbf{J}_*^{-1} p_l(z_l|t)^{-1}$ and $\mathbf{J}_*$ is a 3×3 diagonal matrix with entries $1, \int u^2 K(u)\, du, \int u^2 K(u)\, du$. Note that $\mathbf{r}_{n2} = \mathbf{r}_{n2,1} + \mathbf{r}_{n2,2}$. Arguing as in the derivation of (47), we can prove (47) with $\mathbf{r}_{n1}$ replaced by $\mathbf{r}_{n2,1}$. By change of variables, we have

$$\begin{aligned}&\int \mathbf{v}_{ijl}(t, z_l)K_{h_l}(z_l,\ Z_{ijl})\frac{p_{kl}(z_k, z_l|t)}{p_k(z_k|t)}\, dz_l \\ &\qquad = \int (1, (T_{ij} - t)/h_0, u)^\top \left[\frac{p_{kl}(z_k, Z_{ijl} + uh_l|t)}{p_k(z_k|t) \int K_{h_l}(x - Z_{ijl} - uh_l)\, dx}\right] K(u)\, du,\end{aligned}$$

which is a vector of random variables that depend only on $(T_{ij}, \mathbf{Z}_{ij})$ and are bounded by a constant uniformly for (i, j). Applying a univariate version of Lemma 1, we get $\sup_{t\in[0,1],z_k\in\mathbb{Z}_k} |\mathbf{r}_{n2,2}(t, z_k)| = O_p(\Delta_{3n})$. This implies that (47) holds with $\mathbf{r}_{n1}$ replaced by $\mathbf{r}_{n2,2}$ as well, which completes the proofs of (45) and (46).

8.4 Proof of Theorem 3

Let $\hat{\mathbf{m}}^B(t, \mathbf{z})$ be a version of the full-dimensional local linear estimator $\hat{\mathbf{m}}(t, \mathbf{z})$ obtained by substituting $m(T_{ij}, \mathbf{Z}_{ij})$ for Y_{ij}. In other words, $\hat{\mathbf{m}}^B(t, \mathbf{z})$ is the solution to equation (28) with Y_{ij} replaced by $m(T_{ij}, \mathbf{Z}_{ij})$. Recall that $\hat{\boldsymbol{\mu}}_k^B(t, \cdot)$ are obtained by projecting $\hat{\mathbf{m}}^B(t, \cdot) - \hat{\boldsymbol{\mu}}_0^B(t)$ onto $\mathcal{H}^{0t}(\hat{\mathbf{M}})$ and also that $\mathbf{m}(t, \mathbf{z}) = (m(t, \mathbf{z}), m_1(t, \mathbf{z}), m_{1,2}(t, \mathbf{z}), \ldots,$

$m_{d,2}(t,\mathbf{z}))^\top$, where $m_1(t,\mathbf{z}) = h_0\partial m(t,\mathbf{z})/\partial t$ and $m_{k,2}(t,\mathbf{z}) = h_k\partial m(t,\mathbf{z})/\partial z_k$. Define

$$\begin{aligned}\tilde{\boldsymbol{\mu}}^{B,1}(t,\mathbf{z}) &= \hat{\mathbf{M}}(t,\mathbf{z})^{-1}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\frac{1}{2}\mathbf{v}_{ij}(t,\mathbf{z})K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\left[\left(\frac{T_{ij}-t}{h_0}\right)^2\left(\frac{\partial^2\mu_0(t)}{\partial t^2}\right)h_0^2\right.\\ &\qquad\left.+\sum_{l=1}^{d}\left(\frac{T_{ij}-t}{h_0}\right)^2\left(\frac{\partial^2\mu_l(t,z_l)}{\partial t^2}\right)h_0^2+\sum_{l=1}^{d}\left(\frac{Z_{ijl}-z_l}{h_l}\right)^2\left(\frac{\partial^2\mu_l(t,z_l)}{\partial z_l^2}\right)h_l^2\right]\\ \tilde{\boldsymbol{\mu}}^{B,2}(t,\mathbf{z}) &= \hat{\mathbf{m}}^B(t,\mathbf{z})-\mathbf{m}(t,\mathbf{z})-\tilde{\boldsymbol{\mu}}^{B,1}(t,\mathbf{z})\end{aligned}$$

where $\hat{\mathbf{M}}(t,\mathbf{z})$ is defined in (24). By the Taylor expansion of $m(T_{ij},\mathbf{Z}_{ij}) = \mu_0(T_{ij}) + \sum_{l=1}^d \mu_l(T_{ij},\mathbf{Z}_{ij})$ at $(t,\mathbf{z})$, it follows that

$$\tilde{\boldsymbol{\mu}}^{B,2}(t,\mathbf{z}) = \hat{\mathbf{M}}(t,\mathbf{z})^{-1}\frac{1}{\mathcal{N}_s}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\mathbf{v}_{ij}(t,\mathbf{z})K_{\mathbf{h}}(t,\mathbf{z};T_{ij},\mathbf{Z}_{ij})\eta_{nij}(t,\mathbf{z}), \tag{48}$$

where the random variable $\eta_{nij}(t,\mathbf{z})$ satisfy $\sup_{i,j}\sup_{t\in[0,1],\mathbf{z}\in\mathbb{Z}}|\eta_{nij}(t,\mathbf{z})| \le \varepsilon_n\Delta_{1n}$ for some real sequence ε_n converging to zero. Let $\hat{\boldsymbol{\mu}}_k^{B,j}(t,\cdot)$ for $j = 1,2$ be the solution of the smooth backfitting equation (27) with $\hat{\mathbf{m}}_k(t,\cdot) - \hat{\boldsymbol{\mu}}_0(t)$ replaced by $\hat{\Pi}_k^t(\tilde{\boldsymbol{\mu}}^{B,j}(t,\cdot))$. Arguing as in the proof of Theorem 2 and using (48), we can show

$$\sup_{t\in I_0, z_k\in I_k}|\hat{\boldsymbol{\mu}}_k^{B,2}(t,z_k)| = o_p(\Delta_{1n}).$$

Thus it suffices to show

$$\sup_{t\in I_0, z_k\in I_k}|\hat{\boldsymbol{\mu}}_k^{B,1}(t,z_k) - \boldsymbol{\beta}_k(t,z_k)| = o_p(\Delta_{1n}), \tag{49}$$

where $\boldsymbol{\beta}_k(t,z_k) = \beta_k(t,z_k)\mathbf{1}_1$ and $\beta_k(t,z_k)$ and $\mathbf{1}_1$ are respectively defined in (34) and (19). We first note that, for a $(d+2)$-dimensional vector of d-variate functions $\mathbf{f}(\mathbf{z}) = (f(\mathbf{z}), f_1(\mathbf{z}), f_{1,2}(\mathbf{z}),\dots,f_{d,2}(\mathbf{z}))^\top$, the three-dimensional representation of the projection of $\mathbf{f}$ onto $\mathcal{H}_k^{0t}(\hat{\mathbf{M}})$ is given by

$$\hat{\Pi}_k^t(\mathbf{f})(z_k) = \hat{\mathbf{M}}_{kk}(t,z_k)^{-1}\int \mathbf{J}_k^\top\hat{\mathbf{M}}(t,\mathbf{z})\mathbf{f}(\mathbf{z})\,d\mathbf{z}_{-k} - \hat{c}_k(t)\mathbf{1}_*,$$

where $\mathbf{1}_* = (1,0,0)^\top$ and $\mathbf{J}_k$ is defined in (19). We can determine $\hat{c}_k(t)$ such that the first entry of $\int \hat{\mathbf{M}}_{kk}(t,z_k)\hat{\Pi}_k^t(\mathbf{f})(z_k)\,dz_k$ equals zero. In fact, $\hat{c}_k(t) = \hat{p}_T(t)^{-1}\int \mathbf{1}_1^\top\hat{\mathbf{M}}(t,\mathbf{z})\mathbf{f}(\mathbf{z})\,d\mathbf{z}$ for all $1\le k\le d$. To get the uniform approximation (49), we use the uniqueness of $\hat{\boldsymbol{\mu}}_k^{B,1}(t,\cdot)$ and the fact $\hat{\Pi}_k^t(\tilde{\boldsymbol{\mu}}^{B,1}(t,\cdot) - \sum_{k=1}^d\hat{\boldsymbol{\mu}}_k^{B,1}(t,\cdot)) = \mathbf{0}_3$ for $1\le k\le d$. We have

$$\begin{aligned}&\hat{\mathbf{M}}_{kk}(t,z_k)^{-1}\int \mathbf{J}_k^\top\hat{\mathbf{M}}(t,\mathbf{z})\tilde{\boldsymbol{\mu}}^{B,1}(t,\mathbf{z})\,d\mathbf{z}_{-k}\\ &\quad= \mathbf{1}_*\frac{1}{2}\left(\int u^2K(u)\,du\right)\left[h_0^2\frac{\partial^2\mu_0(t)}{\partial t^2}+h_0^2\frac{\partial^2\mu_k(t,z_k)}{\partial t^2}+h_k^2\frac{\partial^2\mu_k(t,z_k)}{\partial z_k^2}\right.\\ &\qquad\left.+\sum_{l\neq k}\int\frac{p_{kl}(t,z_k,z_l)}{p_k(t,z_k)}\left(h_0^2\frac{\partial^2\mu_k(t,z_l)}{\partial t^2}+h_l^2\frac{\partial^2\mu_l(t,z_l)}{\partial z_l^2}\right)dz_l\right]+o_p(\Delta_{1n})\end{aligned}$$

uniformly for $t \in I_0$ and $z_k \in I_k$. Note that $\mathbf{J}$ is a 3×3 diagonal matrix with entries $1, 1, 0$. Then, we also obtain

$$\begin{aligned}
&\hat{\mathbf{M}}_{kk}(t, z_k)^{-1} \int \mathbf{J}_k^\top \hat{\mathbf{M}}(t, \mathbf{z}) \sum_{l=1}^{d} \hat{\boldsymbol{\mu}}_l^{B,1}(t, z_l)\, d\mathbf{z}_{-k} \\
&\quad = \hat{\boldsymbol{\mu}}_k^{B,1}(t, z_k)[1 + o_p(1)] + \sum_{l \neq k} \int \frac{p_{kl}(t, z_k, z_l)}{p_k(t, z_k)} \mathbf{J} \hat{\boldsymbol{\mu}}_l^{B,1}(t, z_l)\, dz_l [1 + o_p(1)]
\end{aligned}$$

uniformly for $t \in I_0$ and $z_k \in I_k$, where $\hat{\boldsymbol{\mu}}_l^{B,1}$ for $1 \leq l \leq d$ on the right hand side of the equation denote the three-dimensional representations of $\hat{\boldsymbol{\mu}}_l^{B,1}$ on the left hand side. For the correction term $\hat{c}_k(t)$, we have $\hat{c}_k(t) = c(t) + o_p(\Delta_{1n})$ uniformly for $t \in I_0$, where

$$c(t) = \frac{1}{2}\left(\int u^2 K(u)\, du\right)\left[h_0^2 \frac{\partial^2 \mu_0(t)}{\partial t^2} + \sum_{l=1}^{d} \int p_l(z_l|t)\left(h_0^2 \frac{\partial^2 \mu_l(t, z_k)}{\partial t^2} + h_l^2 \frac{\partial^2 \mu_l(t, z_k)}{\partial z_l^2}\right) dz_l\right].$$

Comparing the two expressions of $\hat{\Pi}_k^t(\tilde{\boldsymbol{\mu}}^{B,1}(t, \cdot))$ and $\hat{\Pi}_k^t(\sum_{l=1}^d \hat{\boldsymbol{\mu}}_l^{B,1}(t, \cdot))$, we conclude (49).

8.5 Proof of Theorem 4

The theorem follows from Theorems 2 and 3 and derivation of the joint asymptotic distribution of $\hat{m}_k(t, z_k)$.

Reference

Berhane, K. and Tibshirani, R. J. (1998). Generalized additive models for longitudinal data. *The Canadian Journal of Statistics*, 26:517–535.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17:453–555.

Carroll, R. J., Maity, A., Mammen, E., and Yu, K. (2009). Nonparametric additive regression for repeatedly measured data. *Biometrika*, 96:383–398.

Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26:943–971.

Jiang, C.-R. and Wang, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, 38:1194–1226.

Lee, Y. K., Mammen, E., and Park, B. U. Projection-type estimation for varying coefficient regression models. *Bernoulli, forthcoming*.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Serial B*, 61:381–400.

Linton, O. B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, 84:469–473.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82:93–100.

Mammen, E., Linton, O. B., and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27:1443–1490.

Opsomer, J. D. (2000). Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73:166–179.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regre. *The Annals of Statistics*, 25:186–211.

S.Chmiel, J., Detels, R., Kaslow, R. A., van Raden, M., Kingsley, L. A., and Brookmeyer, R. (1987). Factors associated with prevalent human immunodeficiency virus (hiv) infection in the multicenter aids cohort study. *American Journal of Epidemiology*, 126:568–577.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, 13:689–705.

Wang, J. and Yang, L. (2009). Efficient and fast spline-backfitted kernel smoothing of additive models. *Annals of the Institute of Statistical Mathematics*, 61:663–690.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics*, 35:2474–2503.

Xue, L., Qu, A., and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association*, 105:15181530.

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

You, J. and Zhou, H. (2007). Two-stage efficient estimation of longitudinal nonparametric additive models. *Statistics & Probability Letters*, 77:1666–1675.

Yu, K., Park, B. U., and Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics*, 36:228–260.