Inference for generalized partial functional linear regression

Ting Li and Zhongyi Zhu

Department of Statistics, Fudan University

Supplementary Material

The supplementary material contains additional simulation reports, expressions of some linear operators and details of all the proofs. Section S1 gives the expressions of some linear operators that help to simply the proofs. Section S2 includes the proofs of Theorem 1, Theorem 2 and Theorem 3. Section S3 proves the null limit distribution of the proposed test statistic in Theorem 4. In Section S4, we discuss the potential challenges to the theoretical results if the functional covariate is observed with measurement errors. We provide simulation results with measurement errors to the functional process in Section S5.

S1 Linear operators

In this section, we define some linear operators and give the expressions of the linear operators. All these linear operators help to present the proofs in a more concise way. The current generalized partial functional linear model is more comprehensive and more convenient than the generalized functional linear model studied in Shang and Cheng (2015). Such convenience comes at the price of a harder theoretical investigation. Specifically, the modified conditional expectation $\mathbf{G}(X)$ is supposed to be linear in X in Assumption 4. The decay rates of the coefficients of $\mathbf{G}(X)$ are required to be carefully verified. Further, it takes greater effort to bound the term $E\{I(U)Z\int_0^1 X(t)\beta(t)dt\}$ in the proofs via the inner product in (2.5).

To represent $\ell_{n,\lambda}(\theta)$ by the inner product of the parameter θ, two linear operations R and P_λ are defined as follows,

$$\langle R_u, \theta\rangle = z^\top\gamma + \int_0^1 x(t)\beta(t)dt \quad \text{for any } u \in \mathcal{U} \text{ and } \theta \in \mathcal{H} \tag{S1.1}$$

and

$$\langle P_\lambda\theta_1, \theta_2\rangle = \lambda J(\beta_1, \beta_2) \quad \text{for any } \theta_1, \theta_2 \in \mathcal{H}. \tag{S1.2}$$

Owing to the two operators, we can rewrite $\ell_{n,\lambda}(\theta)$ as

$$\ell_{n,\lambda}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i; \langle R_{U_i}, \theta\rangle) - \frac{1}{2}\langle P_\lambda\theta, \theta\rangle. \tag{S1.3}$$

We separate the joint parameter θ from the covariates X and Z in this manner, and provide a convenient approach to obtain the Fréchet derivatives of $\ell_{n,\lambda}$, which are the premise of deriving the Bahadur representation.

Denote $\Delta\theta = (\Delta\gamma, \Delta\beta)$, the Fréchet derivative of $\ell_{n,\lambda}(\theta)$ with respect to θ is

$$S_{n,\lambda}(\theta)\Delta\theta = D\ell_{n,\lambda}(\theta)\Delta\theta = \frac{1}{n}\sum_{i=1}^{n}\dot{\ell}_a(Y_i;\langle R_{U_i},\theta\rangle)\langle R_{U_i},\Delta\theta\rangle - \langle P_\lambda\theta,\Delta\theta\rangle.$$

Notice that $S_{n,\lambda}(\hat{\theta}_{n,\lambda}) = 0$, and $S_{n,\lambda}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\dot{\ell}_a(Y_i;\langle R_{U_i},\theta_0\rangle)R_{U_i} - P_\lambda\theta_0$ is of interest. The second- and third-order Fréchet derivatives of $\ell_{n,\lambda}(\theta)$ can be derived in the same way and we omit here. Meanwhile, define $S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\dot{\ell}_a(Y_i;\langle R_{U_i},\theta\rangle)R_{U_i}$, $S(\theta) = E\{S_n(\theta)\}$ and $S_\lambda(\theta) = E\{S_{n,\lambda}(\theta)\}$.

In order to obtain the expressions of the two linear operators in (S1.1) and (S1.2), we begin with some preparatory work. Let $K(s,t)$ be the reproducing kernel function of $H^m(\mathbb{I})$, and define $K_t(\cdot) = K(t,\cdot) \in H^m(\mathbb{I})$ for any $t \in \mathbb{I}$. Then $\langle K_t, \beta\rangle_1 = \beta(t)$ for any $\beta \in H^m(\mathbb{I})$ by definition. Also, we define an operator W_λ from $H^m(\mathbb{I})$ to $H^m(\mathbb{I})$ satisfying

$$\langle W_\lambda\beta_1,\beta_2\rangle_1 = \lambda J(\beta_1,\beta_2), \quad \text{for any } \beta_1,\beta_2 \in H^m(\mathbb{I}). \tag{S1.4}$$

Simple calculations lead to expressions of the two operators,

$$K_t(\cdot) = \sum_v \frac{\varphi_v(t)}{1+\lambda\rho_v}\varphi_v(\cdot), \quad (W_\lambda\varphi_v)(\cdot) = \frac{\lambda\rho_v}{1+\lambda\rho_v}\varphi_v(\cdot). \tag{S1.5}$$

It follows that $W_\lambda\beta(\cdot) = \sum_v V(\beta,\varphi_v)\frac{\lambda\rho_v}{1+\lambda\rho_v}\varphi_v(\cdot)$. Meanwhile, we define $\tau(x)(\cdot) \in H^m(\mathbb{I})$ satisfying $\langle\tau(x),\beta\rangle_1 = \int_0^1 x(t)\beta(t)dt$ for any L^2 integrable $x = x(t)$ and $\beta \in H^m(\mathbb{I})$. It is easy to have

$$\tau(x)(t) = \sum_{v=1}^{\infty}\frac{x_v}{1+\lambda\rho_v}\varphi_v(t) \quad \text{for } t \in \mathbb{I}, \tag{S1.6}$$

where $x_v = \langle \tau(x), \varphi_v \rangle_1 = \int_0^1 x(t)\varphi_v(t)dt$. With the aforementioned eigenfunctions, the linear operators W_λ and $\tau(x)(\cdot)$, we can have explicit forms of R_u and P_λ defined in (S1.1) and (S1.2).

Let id be the identity operator such that $id\beta = \beta$, and define $A_j = (id - W_\lambda)\tilde{\beta}_j$ for $\tilde{\beta}_j$ defined in Assumption 4. Then for any $\beta \in H^m(\mathbb{I})$, we have $V(\tilde{\beta}_j, \beta) = \langle A_j, \beta \rangle_1$. Let $\boldsymbol{A} = (A_1, \ldots, A_p)^\top$ and $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_p)^\top$. Note that $\boldsymbol{A}$ and $\tilde{\boldsymbol{\beta}}$ are vectors of functional elements, then

$$V(\tilde{\boldsymbol{\beta}}, \beta) = \langle \boldsymbol{A}, \beta \rangle_1. \tag{S1.7}$$

We can derive the expression of $\boldsymbol{A}$ by taking $\beta = K_t$, one can deduce that

$$\boldsymbol{A}(t) = \langle \boldsymbol{A}, K_t \rangle_1 = \sum_v \frac{V(\tilde{\boldsymbol{\beta}}, \varphi_v)}{1 + \lambda\rho_v}\varphi_v(t) \tag{S1.8}$$

and

$$(W_\lambda \boldsymbol{A})(t) = \sum_v \frac{V(\tilde{\boldsymbol{\beta}}, \varphi_v)\lambda\rho_v}{(1 + \lambda\rho_v)^2}\varphi_v(t).$$

Define $\Omega_2 = E_X\{B(X)\mathbf{G}(X)(\mathbf{G}(X) - \int_0^1 X(t)\boldsymbol{A}(t)dt)^\top\}$. We are ready to obtain the expressions of R_u and P_λ defined in (S1.1) and (S1.2).

Proposition 1. *Let $R_u : u \mapsto (H_u, T_u) \in \mathcal{H}$, we have*

$$\begin{aligned} H_u &= (\Omega_1 + \Omega_2)^{-1}(z - \langle \boldsymbol{A}, \tau(x) \rangle_1), \\ T_u &= \tau(x) - \boldsymbol{A}^\top(\Omega_1 + \Omega_2)^{-1}(z - \langle \boldsymbol{A}, \tau(x) \rangle_1). \end{aligned}$$

Furthermore, P_λ can be expressed as $P_\lambda\theta : (\gamma, \beta) \mapsto (H_u^*, T_u^*) \in \mathcal{H}$, then

$$\begin{aligned} H_u^* &= -(\Omega_1 + \Omega_2)^{-1}\langle \boldsymbol{A}, W_\lambda \beta\rangle_1, \\ T_u^* &= W_\lambda\beta + \boldsymbol{A}^\top(\Omega_1 + \Omega_2)^{-1}\langle \boldsymbol{A}, W_\lambda\beta\rangle_1. \end{aligned}$$

Notice that $(\Omega_1+\Omega_2)^{-1}$ is well defined under Assumption 4 and $\lim_{\lambda\to 0}\Omega_2 = 0$ according to (S1.12).

Proof of Proposition 1. Define $R_u = (H_u, T_u)$, for any $\theta = (\gamma, \beta) \in \mathcal{H}$. According to (2.6) and Assumption 2(b), we have

$$\begin{aligned} \langle (H_u, T_u), (\gamma, \beta)\rangle &= E_U\left\{I(U)\left(Z^\top\gamma + \int_0^1 X(t)\beta(t)dt\right)\left(Z^\top H_u + \int_0^1 X(t)T_u(t)dt\right)\right\} \\ &\quad + \lambda J(T_u, \beta). \end{aligned}$$

By definition (S1.1) of R_u, it also holds

$$\langle (H_u, T_u), (\gamma, \beta)\rangle = z^\top\gamma + \int_0^1 x(t)\beta(t)dt = \gamma^\top z + \langle \tau(x), \beta\rangle_1,$$

then (H_u, T_u) are the solutions of equations

$$\begin{cases} E_U\left\{I(U)ZZ^\top\right\}H_u + E_U\left\{I(U)Z\int_0^1 X(t)T_u(t)dt\right\} = z, \\ E_U\left\{I(U)Z\int_0^1 X(t)\beta(t)dtZ^\top\right\}H_u + \langle\beta, T_u\rangle_1 = \langle\tau(x), \beta\rangle_1. \end{cases} \tag{S1.9}$$

Recall that $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \cdots, \tilde{\beta}_p)^\top$, and $\tilde{\beta}_j$s are defined in Assumption 4, we can rewrite

$$\begin{aligned} & E_U\left\{I(U)Z\int_0^1 X(t)T_u(t)dt\right\} = E_X\left\{B(X)\mathbf{G}(X)\int_0^1 X(t)T_u(t)dt\right\} \\ = \;& E_X\left\{B(X)\int_0^1 X(t)\tilde{\boldsymbol{\beta}}(t)dt\int_0^1 X(t)T_u(t)dt\right\} = \langle \boldsymbol{A}, T_u\rangle_1, \end{aligned}$$

where the last equality follows from the definition of $\boldsymbol{A}$ in (S1.8). Similarly, we have $E_U\left\{I(U)\int_0^1 X(t)\beta(t)dtZ^\top\right\} = \langle \boldsymbol{A}^\top, \beta\rangle_1$. Then we can rewrite (S1.9) as

$$\begin{cases} E_U\left\{I(U)ZZ^\top\right\}H_u + \langle \boldsymbol{A}, T_u\rangle_1 = z, \\ \boldsymbol{A}^\top H_u + T_u = \tau(x). \end{cases} \tag{S1.10}$$

Substituting $T_u = \tau(x) - \boldsymbol{A}^\top H_u$ into the first equation of (S1.10), we have

$$\begin{aligned} z &= E_U\left\{I(U)ZZ^\top\right\}H_u + E_U\left\{I(U)Z\int_0^1 X(t)\tau(x)dt\right\} \\ &\quad -E_U\left\{I(U)Z\int_0^1 X(t)\boldsymbol{A}^\top(t)dt\right\}H_u \\ &= E_U\left\{I(U)(Z-\mathbf{G}(X))(Z-\mathbf{G}(X))^\top\right\}H_u + E_U\left\{I(U)Z\int_0^1 X(t)\tau(x)dt\right\} \\ &\quad +E_X\left\{B(X)\mathbf{G}(X)(\mathbf{G}(X) - \int_0^1 X(t)\boldsymbol{A}(t)dt)^\top\right\}H_u \\ &= (\Omega_1+\Omega_2)H_u + \langle \boldsymbol{A}, \tau(x)\rangle_1. \end{aligned}$$

It is easy to see

$$\begin{aligned} H_u &= (\Omega_1+\Omega_2)^{-1}(z - \langle \boldsymbol{A}, \tau(x)\rangle_1), \\ T_u &= \tau(x) - \boldsymbol{A}^\top(\Omega_1+\Omega_2)^{-1}(z - \langle \boldsymbol{A}, \tau(x)\rangle_1). \end{aligned}$$

Similar to the process above, one can get the expression of $P_\lambda\theta$ if we let $z=0$ and replace $\tau(x)$ with $W_\lambda\beta$. □

Lemma 1. *Recall that* $B(X) = E\{I(U)|X\}$, $\mathbf{G}(X) = E\{I(U)Z|X\}/B(X)$

and $\boldsymbol{A}$ *is defined in (S1.8), as* $\lambda \to 0$*, we have*

$$\lim_{\lambda\to 0} E_X\{B(X)(\mathbf{G}(X) - \langle \boldsymbol{A}, \tau(X)\rangle_1)(\mathbf{G}(X) - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} = 0, \quad \text{(S1.11)}$$

$$\lim_{\lambda\to 0} E_X\{B(X)\mathbf{G}(X)(\mathbf{G}(X) - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} = \lim_{\lambda\to 0}\Omega_2 = 0, \quad \text{(S1.12)}$$

$$\lim_{\lambda\to 0} E_U\{I(U)(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} = \Omega_1. \quad \text{(S1.13)}$$

Proof. Since the proofs of (S1.11) and (S1.12) are similar, we only show that (S1.12) holds. For any $j, k \in \{1, 2, \ldots, p\}$, recall that $G_j(X) = \int_0^1 X(t)\tilde{\beta}_j(t)dt = \sum_v V(\tilde{\beta}_j, \varphi_v)X_v$, and $\langle A_j, \tau(X)\rangle_1 = \sum_v \frac{V(\tilde{\beta}_j,\varphi_v)}{1+\lambda\rho_v}X_v$, then

$$\begin{aligned} & E_X\{B(X)G_j(X)(G_k(X) - \langle A_k, \tau(X)\rangle_1)\} \\ = \; & E_X\{B(X)\sum_v V(\tilde{\beta}_j, \varphi_v)X_v \sum_v \frac{V(\tilde{\beta}_k, \varphi_v)\lambda\rho_v}{1+\lambda\rho_v}X_v\}. \end{aligned} \quad \text{(S1.14)}$$

For any $v_1 \neq v_2$, we can derive that

$$\begin{aligned} E_X\{B(X)X_{v_1}X_{v_2}\} &= E_X\{B(X)\int_0^1 X(t)\varphi_{v_1}(t)dt \int_0^1 X(t)\varphi_{v_2}(t)dt\} \\ &= V(\varphi_{v_1}, \varphi_{v_2}) = 0. \end{aligned}$$

Then (S1.14) turns into

$$\sum_v V(\tilde{\beta}_j, \varphi_v)V(\tilde{\beta}_k, \varphi_v)\frac{\lambda\rho_v}{1+\lambda\rho_v} \leq \sum_v V^{1/2}(\tilde{\beta}_j, \tilde{\beta}_j)V^{1/2}(\tilde{\beta}_k, \tilde{\beta}_k)\frac{\lambda\rho_v}{1+\lambda\rho_v}.$$

Under Assumption 4 (b) that $V(\tilde{\beta}_j, \tilde{\beta}_j) < \infty$ and the dominated convergence theorem, we have that the above sum converges to zero as $\lambda \to 0$.

For the proof of (S1.13), simple calculations imply that

$$
\begin{aligned}
& E_U\{I(U)(Z-\langle \boldsymbol{A}, \tau(X)\rangle_1)(Z-\langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} \\
= \; & E_U\{I(U)(Z-\mathbf{G}(X))(Z-\mathbf{G}(X))^\top\} \\
& +2E_U\{I(U)(Z-\mathbf{G}(X))(\mathbf{G}(X)-\langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} \\
& +E_U\{I(U)(\mathbf{G}(X)-\langle \boldsymbol{A}, \tau(X)\rangle_1)(\mathbf{G}(X)-\langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\} \\
= \; & I_1+I_2+I_3,
\end{aligned}
$$

we can easily have $\lim_{\lambda\to 0} I_3 = 0$ according to (S1.11). For I_2, rewrite it as

$$
I_2 = E_X\{E_U\{I(U)(Z-\mathbf{G}(X))|X\}(\mathbf{G}(X)-\langle \boldsymbol{A}, \tau(X)\rangle_1)^\top\}\}.
$$

Recall that $B(X) = E_U\{I(U)|X\}$ and $E_U\{I(U)Z|X\} = \mathbf{G}(X)B(X)$, we have $I_2 = 0$. This completes the proof of (S1.13). □

S2 Proofs of the theoretical results

We need to establish inequalities with respect to the inner product of R_u and its expectation, which are involved in the proofs.

Lemma 2. *Suppose that Assumption 2 and Assumption 3 hold, then for any $u = (x, z)$, $x \in L^2(\mathbb{I}), z \in \mathbb{R}^p$, we get that*

$$
\begin{aligned}
\langle R_u, R_u\rangle \; = \; & (z-\langle \boldsymbol{A}, \tau(x)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(z-\langle \boldsymbol{A}, \tau(x)\rangle_1) \\
& +\langle \tau(x), \tau(x)\rangle_1. \qquad \text{(S2.15)}
\end{aligned}
$$

Meanwhile, as $h \to 0$, there exists a universal constant $C_R > 0$ satisfying $\langle R_u, R_u\rangle \le C_R(1 + \|x\|_{L^2}^2 h^{-(2a+1)})$, and $E_U\{\|R_U\|^2\} \le C_R h^{-1}$.

Proof of Lemma 2. The expression of $\langle R_u, R_u\rangle$ directly follows the definition of R_u. Next we show that the two inequalities hold. Recall that $\tau(x) = \sum_v \frac{X_v}{1+\lambda\rho_v}\varphi_v$ where $X_v = \int_0^1 X(t)\varphi_v(t)dt$. It follows that $\langle \tau(x), \tau(x)\rangle_1 = \sum_v \frac{X_v^2}{1+\lambda\rho_v}$. Under Assumption 2 that $I(U) > C_2^{-1}$, we have

$$\begin{aligned} E_U\{\|R_U\|^2\} &\le C_2 E_U\{I(U)\|R_U\|^2\} \\ &= C_2 E_U\{I(U)(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)\} \\ &\quad + C_2 E_U\left\{I(U)\sum_v \frac{X_v^2}{1+\lambda\rho_v}\right\}. \end{aligned} \tag{S2.16}$$

For the second part of (S2.16), by Assumption 3 that $E[I(U)X_v^2] = V(\varphi_v, \varphi_v) = 1$ and $\rho_v \asymp v^{2k}$, it is easy to derive that

$$\begin{aligned} E_U\left\{I(U)\sum_v \frac{X_v^2}{1+\lambda\rho_v}\right\} &\asymp \sum_v \frac{1}{1+\lambda\rho_v} \le \int_1^\infty \frac{1}{1+\lambda v^{2k}}dv \\ &= h^{-1}\int_1^\infty \frac{1}{1+(hv)^{2k}}d(hv). \end{aligned}$$

Since $\int_1^\infty \frac{1}{1+(hv)^{2k}}d(hv) \le \infty$, it is obvious that there exists a constant C_{R_1}, s.t. $E_U\{I(U)\sum_v \frac{X_v^2}{1+\lambda\rho_v}\} \le C_{R_1}h^{-1}$.

We conclude $E_U\{\|R_U\|^2\} \le C_{R_1}h^{-1}$ by examining the finiteness of the first part in (S2.16). According to (S1.12) and (S1.13), one can verify that

$$\begin{aligned} & E_U\{I(U)(Z-\langle \boldsymbol{A},\tau(X)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(Z-\langle \boldsymbol{A},\tau(X)\rangle_1)\} \\ = \; & \mathrm{tr}(\mathrm{E_U}\{\mathrm{I(U)}(\Omega_1+\Omega_2)^{-1}(\mathrm{Z}-\langle \mathrm{A},\tau(\mathrm{X})\rangle_1)(\mathrm{Z}-\langle \mathrm{A},\tau(\mathrm{X})\rangle_1)^\top\}) \\ = \; & p. \end{aligned}$$

We can use the inequalities $|x_v| \le \|x\|_{L^2}\|\varphi_v\|_{L^2} \le \|x\|_{L^2}C_\varphi v^a$ and the boundness of $(z-\langle \boldsymbol{A},\tau(x)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(z-\langle \boldsymbol{A},\tau(x)\rangle_1)$ to prove $\langle R_u, R_u\rangle \le C_R(1+\|x\|^2_{L^2}h^{-(2a+1)})$. Specifically,

$$\begin{aligned} \langle R_u, R_u\rangle &= (z-\langle \boldsymbol{A},\tau(x)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(z-\langle \boldsymbol{A},\tau(x)\rangle_1)+\sum_v \frac{X_v^2}{1+\lambda\rho_v} \\ &\le (z-\langle \boldsymbol{A},\tau(x)\rangle_1)^\top(\Omega_1+\Omega_2)^{-1}(z-\langle \boldsymbol{A},\tau(x)\rangle_1)+\sum_v \|x\|^2_{L^2}\frac{C_\varphi^2 v^{2a}}{1+\lambda\rho_v} \\ &\le C_{R_2}(1+\|x\|^2_{L^2}h^{-(2a+1)}). \end{aligned}$$

The universal constant can be taken as $C_R = \max(C_{R_1}, C_{R_2})$. □

Denote $T=(Y,Z,X(\cdot))\in\mathcal{T}$, the following lemma proves a vital condition (S2.19) on $H_n(\theta)$ defined as

$$H_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^n[\psi_n(T_i;\theta)R_{U_i}-E_T\{\psi_n(T;\theta)R_U\}], \tag{S2.17}$$

where $\psi_n(T;\theta)$ is a function defined over $\mathcal{T}\times\mathcal{H}$. Define $\mathcal{F}_{p_n}=\{\theta=(\gamma,\beta)\in\mathcal{H}:\gamma^\top\gamma\le 1, \|\beta\|_{L^2}\le 1, J(\beta,\beta)\le p_n\}$, where $p_n\ge 1$. It is worth emphasizing that the proofs of the Bahadur representation count on (S2.19) given in the following lemma.

Lemma 3. *Suppose that Assumptions 2 to 5 hold,* $\psi_n(T_i; 0) = 0$ *a.s., and there exsits a constant* $C_\psi > 0$ *such that the Lipschitz continuity holds,*

$$|\psi_n(T;\theta) - \psi_n(T;\tilde{\theta})| \le C_\psi \|\theta - \tilde{\theta}\|_2 \quad \text{for any } \theta, \tilde{\theta} \in \mathcal{F}_{\mathrm{p_n}}. \tag{S2.18}$$

Then as $n \to \infty$,

$$\sup_{\theta \in \mathcal{F}_{p_n}} \frac{\|H_n(\theta)\|}{p_n^{1/(4m)} \|\theta\|_2^{\zeta} + n^{-1/2}} = O_P((h^{-1} \log\log n)^{1/2}), \tag{S2.19}$$

where $\zeta = 1 - 1/(2m)$.

The proof of Lemma 3 is similar to the proof of Lemma 3.4 of Shang and Cheng (2015) by using Lemma 2 and modern empirical process theory, so we omit here

With the preparations above, we can prove Theorem 1 and Theorem 2.

Proof of Theorem 1 . The proof of Theorem 1 follows from the proof of Proposition 3.5 of Shang and Cheng (2015) by using Lemmas 2–3, Assumptions 1–6, the conditions in Theorem 1 and the Cauchy's inequality. □

Proof of Theorem 2. The proof of Theorem 2 follows directly from the proof of Theorem 3.6 of Shang and Cheng (2015) and is omitted here. □

Proof of Theorem 3 . The proof of the joint distribution depends on the Cramér-Wald device. Denote $\theta_0^* = \theta_0 - P_\lambda \theta_0 = (\gamma_0^*, \beta_0^*)$. For any $\tilde{z} \in \mathbb{R}^p$, and $u^* = (\tilde{z}, \tilde{x}_0)$, we will derive the distribution of

$$\begin{aligned} & \{\tilde{z}^\top(\hat{\gamma}_{n,\lambda} - \gamma_0^*) + \int_0^1 \tilde{x}_0(t)\hat{\beta}_{n,\lambda}(t)dt - \int_0^1 \tilde{x}_0(t)\beta_0^*(t)dt\} \\ = & \langle R_{u^*}, \hat{\theta}_{n,\lambda} - \theta_0^* \rangle \end{aligned} \tag{S2.20}$$

where $\tilde{x}_0 = x_0 \cdot \sigma_{x_0}^{-1}$. Then we will show that the bias converges to zero, which can be found in Lemma 4.

Recall that $S_{n,\lambda}(\theta_0) = \frac{1}{n}\sum_{i=1}^n \epsilon_i R_{U_i} - P_\lambda \theta_0$. For the distribution of (S2.20), under the condition $\|R_{u^*}\| = \mathrm{O}(1)$ and by Theorem 2, we have

$$|\langle R_{u^*}, \hat{\theta}_{n,\lambda} - \theta_0 - S_{n,\lambda}(\theta_0)\rangle| \leq \|R_{u^{**}}\| \|\hat{\theta}_{n,\lambda} - \theta_0 - S_{n,\lambda}(\theta_0)\| \leq O_p(a_n).$$

Then we will derive the asymptotic distribution of $\langle R_{u^*}, S_{n,\lambda}(\theta_0)\rangle$.

Direct calculations lead to

$$\langle R_{u^*}, S_{n,\lambda}(\theta_0)\rangle = \frac{1}{n}\sum_{i=1}^n \epsilon_i(\tilde{z}^\top H_{U_i} + \langle \tau(x_0), T_{U_i}\rangle_1) - \langle P_\lambda \theta_0, R_{u^*}\rangle,$$

where H_{U_i}, T_{U_i} are defined in Proposition S1.1, then

$$\begin{aligned} M_i & \triangleq \tilde{z}^\top H_{U_i} + \langle \tau(\tilde{x}_0), T_{U_i}\rangle_1 \qquad \text{(S2.21)} \\ & = (\tilde{z} - \langle \boldsymbol{A}, \tau(\tilde{x}_0)\rangle_1)^\top (\Omega_1 + \Omega_2)^{-1}(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1) + \langle \tau(\tilde{x}_0), \tau(X)\rangle_1. \end{aligned}$$

It follows from Assumption 2 and $E(\epsilon^2|U) = I(U)$ that

$$\begin{aligned} s_n^2 &= nE\{\epsilon^2|\tilde{z}^\top H_{U_i} + \langle \tau(\tilde{x}_0), T_{U_i}\rangle_1|^2\} \\ &= nE\{I(U)\langle \tau(\tilde{x}_0), \tau(X)\rangle_1^2\} \\ &\quad +2n(\tilde{z} - \langle \boldsymbol{A}, \tau(\tilde{x}_0)\rangle_1)^\top E\{I(U)\langle \tau(\tilde{x}_0), \tau(X)\rangle_1 (Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)\} \\ &\quad +n(\tilde{z} - \langle \boldsymbol{A}, \tau(\tilde{x}_0)\rangle_1)^\top E\{I(U) H_U H_U^\top\}(\tilde{z} - \langle \boldsymbol{A}, \tau(\tilde{x}_0)\rangle_1). \quad \text{(S2.22)} \end{aligned}$$

Recall that $B(X) = E\{I(U)|X\}$, it is easy to verify that

$$\begin{aligned} &E\{I(U)\langle \tau(\tilde{x}_0), \tau(X)\rangle_1^2\} = V(\tau(\tilde{x}_0), \tau(\tilde{x}_0)) \\ = &\sum_{v=1}^\infty \frac{x_{0v}^2}{(1+\lambda\rho_v)^2} \cdot {\sigma_{x_0}^2}^{-1} = 1. \quad \text{(S2.23)} \end{aligned}$$

Meanwhile, Lemma 1 implies that as $\lambda \to 0$,

$$E\{I(U) H_U H_U^\top\} \to \Omega_1^{-1}. \quad \text{(S2.24)}$$

Thus, it can be derived from Lemma 4, (S2.23) and (S2.24),

$$s_n^2 = n\{1 + \tilde{z}^\top \Omega_1^{-1} \tilde{z}\} = n(\tilde{z}^\top, 1)^\top \Psi(\tilde{z}^\top, 1) \asymp n, \quad \text{(S2.25)}$$

where Ψ is defined in Theorem 3.

Recall that M_i are defined in (S2.21). By Lemma 4 and $\|\tau(X)\|_1 \leq C_R h^{-\frac{(2a+1)}{2}} \cdot \|X_i\|_{L^2}$ from the proof of Lemma 2, we can obtain

$$M_i \leq \tilde{z}^\top \Omega^{-1}(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1) + C_R h^{-\frac{(2a+1)}{2}} \cdot \|X_i\|_{L^2} \cdot \|\tau(\tilde{x}_0)\|_1.$$

Denote c^* as the largest element of the matrix $\Omega^{-1}\tilde{z}\tilde{z}^\top\Omega^{-1}$, then c^* is finite due to the definiteness of Ω_1. Cauchy's inequality indicates that

$$\begin{aligned} M_i^2 &\leq 2(Z_i - \langle \boldsymbol{A}, \tau(X_i)\rangle_1)^\top \Omega^{-1}\tilde{z}\tilde{z}^\top\Omega^{-1}(Z_i - \langle \boldsymbol{A}, \tau(X_i)\rangle_1) \\ &\quad +2C_R^2 h^{-(2a+1)} \cdot \|X_i\|_{L^2}^2 \cdot \|\tau(\tilde{x}_0)\|_1^2 \\ &\leq 2c^*(Z - \langle \boldsymbol{A}, \tau(X_i)\rangle_1)^\top (Z - \langle \boldsymbol{A}, \tau(X_i)\rangle_1) \\ &\quad +2C_R^2 h^{-(2a+1)} \cdot \|X_i\|_{L^2}^2 \cdot \|\tau(\tilde{x}_0)\|_1^2. \end{aligned}$$

Next we will check the Lindeberg's condition. Since $\log(h^{-1}) = O(\log n)$ holds, we can choose a large constant $\tilde{C} > 0$ such that $h^{-(2a+1)}n^{-\tilde{C}} = o(1)$. Then, for any $\varepsilon > 0$, one can obtain

$$\frac{n}{s_n^2} E\{\epsilon_i^2 M_i^2 I(\epsilon_i^2 M_i^2 \geq \varepsilon^2 s_n^2)\} \lesssim E\{\epsilon_i^4 M_i^4\}^{1/2} P(\epsilon_i^2 M_i^2 \geq \epsilon^2 s_n^2)^{1/2}. \quad \text{(S2.26)}$$

Recall that $E(\epsilon_i^4|U) < \infty$, it is easy to check that

$$E\{\epsilon_i^4 M_i^4\} = E\{E(\epsilon_i^4|U)M_i^4\} \lesssim E\{M_i^4\} = O(h^{-2(2a+1)}). \qquad \text{(S2.27)}$$

Meanwhile, one can deduce that

$$\begin{aligned} & P(\epsilon_i^2 M_i^2 \geq \epsilon^2 s_n^2) \\ \leq\ & P(s^*|\epsilon_i| \geq \tilde{C}\log n) + P(s^*|(Z - \langle \boldsymbol{A}, \tau(X_i)\rangle_1)^\top (Z - \langle \boldsymbol{A}, \tau(X_i)\rangle_1)| \geq \tilde{C}\log n) \\ & +P\left(s^*\|X\|_{L^2} \geq s^*\sqrt{\frac{h^{2a+1}}{C_R}\left(\frac{s^{*3}\varepsilon^2 n}{(\tilde{C}\log n)^2\|\tau(\tilde{x}_0)\|_1^2} - \tilde{C}\log n\right)}\right). \end{aligned}$$

Owing to the conditions $E\{\exp(s^*|\epsilon|)\} < \infty$, (3.2) and (3.5), we have

$$\begin{aligned} & P(\epsilon_i^2 M_i^2 \geq \epsilon^2 s_n^2) \\ \leq\ & 2n^{-\tilde{C}} + \exp\left(-s^*\sqrt{\frac{h^{2a+1}}{C_R}\left(\frac{s^{*3}\varepsilon^2 n}{(\tilde{C}\log n)^2} - \tilde{C}\log n\right)}\right). \end{aligned} \quad \text{(S2.28)}$$

Substituting (S2.27) and (S2.28) into (S2.26), one can verify that

$$\begin{aligned} & \frac{n}{s_n^2} E\{\epsilon_i^2 M_i^2 I(\epsilon_i^2 M_i^2 \geq \varepsilon^2 s_n^2)\} \\ \lesssim\ & O(h^{-(2a+1)})\left[2n^{-\tilde{C}} + \exp\left(-s^*\sqrt{\frac{h^{2a+1}}{C_R}\left(\frac{s^{*3}\varepsilon^2 n}{(\tilde{C}\log n)^2\|\tau(\tilde{x}_0)\|_1^2} - \tilde{C}\log n\right)}\right)\right]^{1/2}. \end{aligned}$$

Then the Lindeberg's condition holds under the condition $nh^{2a+1} \gg (\log n)^4$ and suitable choice of $\tilde{C}$, which implies $s_n^{-1}\sum_i \epsilon_i M_i \xrightarrow{d} N(0,1)$. □

Lemma 4. *Suppose that there exists $b \in ((2a+1)/2k, a/k+1]$, such that for $j = 1, \cdots, p$, $\tilde{\beta}_j$ satisfies (3.10). If $n^{1/2}\lambda^{\frac{1+b-a/k}{2}} = o(1)$ and $h = o(1)$, then for any $x_0 \in L^2(\mathbb{I})$, and $\tilde{x}_0 = x_0 \cdot \sigma_{x_0}^{-1}$, we have*

$$\langle \boldsymbol{A}, \tau(\tilde{x}_0)\rangle_1 = o(1), \quad \text{(S2.29)}$$

$$E\{I(U)\langle \tau(\tilde{x}_0), \tau(X)\rangle_1 (Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)\} = o(1). \quad \text{(S2.30)}$$

Recall that $\theta_0^* = \theta_0 - P_\lambda \theta_0 = (\gamma_0^*, \beta_0^*)$, then

$$\begin{pmatrix} \sqrt{n}(\gamma_0 - \gamma_0^*) \\ \sqrt{n}\{\int_0^1 \tilde{x}_0(t)(\beta_0(t) - W_\lambda\beta_0(t) - \beta_0^*(t))dt\} \end{pmatrix} \to 0. \quad \text{(S2.31)}$$

Proof. First we show that (S2.29) holds. By the definition of $\boldsymbol{A}$ in (S1.7), for any $j = 1, \cdots, p$, we have

$$\langle A_j, \tau(\tilde{x}_0)\rangle_1 = \sigma_{x_0}^{-1} V(\tilde{\beta}_j, \tau(x_0)).$$

Recall that $\tau(x_0) = \sum_v \frac{x_{0v}}{1+\lambda\rho_v}\varphi_v$, it is easy to see that $V(\tilde{\beta}_j, \tau(x_0)) = \sum_v V(\tilde{\beta}_j, \varphi_v)\frac{x_{0v}}{1+\lambda\rho_v}$. By the Cauchy's inequality, $x_{0v} \le \|x_0\|_{L^2}\|\varphi_v\|_{L^2}$ and $\|\varphi_v\|_{L^2} \le C_{\varphi_v} v^a$ in Assumption 3, we have

$$\begin{aligned}
|V(\tilde{\beta}_j, \tau(x_0))|^2 &\le \sum_v V(\tilde{\beta}_j, \varphi_v)^2 \|x_0\|_{L^2}^2 v^{2a}(1+\rho_v)^{b-a/k}\frac{1}{(1+\lambda\rho_v)^2(1+\rho_v)^{b-a/k}} \\
&\lesssim \sum_v V(\tilde{\beta}_j, \varphi_v)^2 \|x_0\|_{L^2}^2 {\rho_v}^{a/k}(1+\rho_v)^{b-a/k}\frac{1}{(1+\lambda\rho_v)^2(1+\rho_v)^{b-a/k}} \\
&= O\left(\frac{1}{(1+\rho_v)^{b-a/k}}\right) = O(1),
\end{aligned}$$

where the last equality follows from $x_0 \in L^2(\mathbb{I})$, condition (3.10), $\rho_v \asymp v^{2k}$ and $2k(b - a/k) > 1$. As $\sigma_{x_0}^{-1} = o(1)$, we can directly have

$$\langle A_j, \tau(\tilde{x}_0)\rangle_1 = \sigma_{x_0}^{-1} V(\tilde{\beta}_j, \tau(x_0)) = o(1).$$

Next we show that (S2.30) holds. Since

$$\begin{aligned}
&E\{I(U)\langle\tau(\tilde{x}_0), \tau(X)\rangle_1(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)\} \\
=\ &\sigma_{x_0}^{-1} E\{I(U)\langle\tau(x_0), \tau(X)\rangle_1(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)\},
\end{aligned}$$

it is sufficient to show that for any $j = 1, \cdots, p$,

$$E\{I(U)\langle\tau(x_0), \tau(X)\rangle_1(Z_j - \langle A_j, \tau(X)\rangle_1)\} = O(1).$$

Under Assumption 4 and $E(I(U)Z_j|X) = B(X)\int_0^1 \tilde{X}(t)\beta_j(t)dt$, we have

$$\begin{aligned} & E\{I(U)\langle\tau(x_0),\tau(X)\rangle_1(Z_j - \langle A_j,\tau(X)\rangle_1)\} \\ = \; & V(\tilde{\beta}_j - A_j,\tau(x_0)) \leq V(\tilde{\beta}_j,\tau(x_0) = O(1). \end{aligned}$$

Then $E\{I(U)\langle\tau(\tilde{x}_0),\tau(X)\rangle_1(Z - \langle \boldsymbol{A},\tau(X)\rangle_1)\} = o(1)$ follows immediately from $\sigma_{x_0}^{-1} = o(1)$.

In the end, we show that the bias converges to zero. Rewrite

$$\begin{aligned} & \begin{pmatrix} \sqrt{n}(\gamma_0 - \gamma_0^*) \\ \sqrt{n}\{\int_0^1 \tilde{x}_0(t)\beta_0(t)dt - \int_0^1 \tilde{x}_0(t)W_\lambda\beta_0(t)dt - \int_0^1 \tilde{x}_0(t)\beta_0^*(t)dt\} \end{pmatrix} \\ = \; & \sqrt{n}\begin{pmatrix} (\Omega_1+\Omega_2)^{-1}\langle \boldsymbol{A}, W_\lambda\beta_0\rangle \\ -\int_0^1 \tilde{x}_0(t)\boldsymbol{A}^\top(t)dt(\Omega_1+\Omega_2)^{-1}\langle \boldsymbol{A}, W_\lambda\beta_0\rangle \end{pmatrix}. \end{aligned} \tag{S2.32}$$

From (S2.29), we can directly have that $\int_0^1 \tilde{x}_0(t)\boldsymbol{A}^\top(t)dt = \langle \boldsymbol{A}^\top, \tau(\tilde{x}_0)\rangle = o(1)$. We only need to show that $\|\langle \boldsymbol{A}, W_\lambda\beta_0\rangle\|_{l^2} = o(n^{-1/2})$ because $(\Omega_1+\Omega_2)$ is positive definite. Recall $W_\lambda\beta_0 = \sum_v \frac{V(\beta_0,\varphi_v)}{1+\lambda\rho_v}\lambda\rho_v\varphi_v$, for any $j = 1,\cdots,p$,

$$\langle A_j, W_\lambda\beta_0\rangle = V(\tilde{\beta}_j, W_\lambda\beta) = \sum_v V(\beta_0,\varphi_v)V(\tilde{\beta},\varphi_v)\frac{\lambda\rho_v}{1+\lambda\rho_v}.$$

Note that β_0 admits $\sum_v V(\beta_0,\varphi_v)^2\rho_v < \infty$, then

$$\begin{aligned} \langle A_j, W_\lambda\beta_0\rangle^2 \; & \leq \; \sum_v V(\beta_0,\varphi_v)^2\frac{\lambda\rho_v}{1+\lambda\rho_v}\sum_v V(\tilde{\beta},\varphi_v)^2\frac{\lambda\rho_v}{1+\lambda\rho_v} \\ & \lesssim \; \lambda\sum_v V(\beta_0,\varphi_v)^2\rho_v^{b-a/k}\frac{\lambda\rho_v^{1-b+a/k}}{1+\lambda\rho_v} \\ & \lesssim \; \lambda^{1+b-a/k}, \end{aligned}$$

where the last inequality follows from (3.10). Therefore, $n^{1/2}\lambda^{\frac{1+b-a/k}{2}} = o(1)$ implies $\|\langle \boldsymbol{A}, W_\lambda \beta_0\rangle\|_{l^2} = o(n^{-1/2})$. $\square$

S3 Proofs of the limit distributions

Proof of Theorem 4. Let $\theta_0 = (\gamma_0, \beta_0) = 0$ be the true parameter under H_0, and $\hat{\theta}^0 = (\hat{\gamma}^0, \hat{\beta}^0)$ be the maximizer over $\mathcal{H}$. In analogy to Shang and Cheng (2015), we have

$$T_P = n^{-1}\|\sum_{i=1}^{n} \epsilon_i R_{U_i}\|^2 + n\|W_\lambda \beta_0\|_1^2 + n^{1/2}\|S_{n,\lambda}(\theta_0)\| \cdot o_p(1) + o_p(h^{-1/2}). \tag{S3.33}$$

The null limit distribution depends on the term $n^{-1}\|\sum_{i=1}^n \epsilon_i R_{U_i}\|^2$, and we can rewrite it as

$$\|\sum_{i=1}^{n} \epsilon_i R_{U_i}\|^2 = \sum_{i=1}^{n} \epsilon_i^2 \langle R_{U_i}, R_{U_i}\rangle + 2\sum_{1\le i<j\le n} \epsilon_i\epsilon_j \langle R_{U_i}, R_{U_j}\rangle.$$

Denote $W_{ij} = 2\epsilon_i\epsilon_j\langle R_{U_i}, R_{U_j}\rangle$ and define $W(n) = \sum_{i<j} W_{ij}$. It is easy to verify that for $i<j$,

$$E\{W_{ij}|\epsilon_i, U_i\} = 2\langle R_{U_i}, E\{\epsilon_i\epsilon_j R_{U_j}|\epsilon_i, U_i\}\rangle = 2\epsilon_i\langle R_{U_i}, E\{\epsilon_j R_{U_j}|\epsilon_i, U_i\}\rangle = 0.$$

Hence, $W(n)$ is clean in the sense of de Jong (1987).

Define $\sigma(n)^2 = E\{W(n)^2\}$ and

$$\begin{aligned} G_I &= \sum_{i<j} E\{W_{ij}^4\}, \\ G_{II} &= \sum_{i<j<k} (E\{W_{ij}^2 W_{ik}^2\} + E\{W_{ji}^2 W_{jk}^2\} + E\{W_{ki}^2 W_{kj}^2\}), \\ G_{IV} &= \sum_{i<j<k<l} (E\{W_{ij} W_{ik} W_{lj} W_{lk}\} + E\{W_{ij} W_{il} W_{kj} W_{kl}\} + E\{W_{ik} W_{il} W_{jk} W_{jl}\}). \end{aligned}$$

According to Proposition 3.2 of de Jong (1987), we can derive the limit distribution of $W(n)$ if G_I, G_{II}, G_{IV} are of lower orders than $\sigma(n)^4$.

It is easy to see that

$$E\{W_{ij}^4\} = 2^4 E\{\epsilon_i^4 \epsilon_j^4 |\langle R_{U_i}, R_{U_j}\rangle|^4\} \leq 16 E\{\epsilon^4 \|R_U\|^4\}^2 \leq 16 M_4^2 E\{|\langle R_U, R_U\rangle|^2\}^2.$$

Recall that $E\{I(U) X_v^2\} = V(\varphi_v, \varphi_v) = 1$ where $X_v = \int_0^1 X(t)\varphi_v(t)dt$, then $E\{X_v^4\} \leq E\{X_v^2\}^2 \leq C_2^2 E\{I(U) X_v^2\}^2 = C_2^2$. From (S2.15) and (S1.13), we can directly have

$$\begin{aligned} & E\{|\langle R_U, R_U\rangle|^2\} \\ \leq\ & 2E\{[(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top \Omega_1^{-1} (Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)]^2 + |\langle \tau(X), \tau(X)\rangle_1|^2\}. \end{aligned}$$

We will deal with the two terms respectively. For the first term, by (3.5) in Assumption 3.6 and the positive definiteness of Ω_1, we can see that

$$E\{[(Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)^\top \Omega_1^{-1} (Z - \langle \boldsymbol{A}, \tau(X)\rangle_1)]^2 < \infty. \qquad \text{(S3.34)}$$

For the second term, direct calculations give us

$$\begin{aligned} 2E\{|\langle \tau(X), \tau(X)\rangle_1|^2\} &= 2E\{|\sum_v \frac{X_v^2}{1+\lambda\rho_v}|^2\} \\ &\le 2E\{|\sum_v \frac{X_v^4}{1+\lambda\rho_v}|\sum_v \frac{1}{1+\lambda\rho_v}\} = O(h^{-2}). \end{aligned}$$

Thus, $E\{|\langle R_U, R_U\rangle|^2\} = O(h^{-2})$, $E\{W_{ij}^4\} = O(h^{-4})$ and $G_I = O(n^2h^{-4})$.

Meanwhile, since $E\{W_{ij}^2 W_{it}^2\} \le E\{W_{ij}^4\}$ holds for $i < j < t$, we have $G_{II} = O(n^3h^{-4})$. Finally, we will derive the bound rate of G_{IV}. For $i < j < t < l$, denote $\tilde{Z}_i = Z_i - \langle \boldsymbol{A}, \tau(X_i)\rangle_1$. It can be shown that

$$\begin{aligned} & E\{W_{ij}W_{it}W_{lj}W_{lt}\} \\ = \; & 2^4 E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \langle R_{U_i}, R_{U_j}\rangle\langle R_{U_i}, R_{U_t}\rangle\langle R_{U_l}, R_{U_j}\rangle\langle R_{U_l}, R_{U_t}\rangle\} \\ = \; & 16E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \tilde{Z}_i^\top\Omega_1^{-1}\tilde{Z}_j \cdot \tilde{Z}_i^\top\Omega_1^{-1}\tilde{Z}_t \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_j \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_t\} \\ & +(16)(4)E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \langle\tau(X_i), \tau(X_j)\rangle_1 \cdot \tilde{Z}_i^\top\Omega_1^{-1}\tilde{Z}_t \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_j \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_t\} \\ & +(16)(4)E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \langle\tau(X_i), \tau(X_j)\rangle_1\langle\tau(X_i), \tau(X_t)\rangle_1 \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_j \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_t\} \\ & +(16)(2)E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \langle\tau(X_i), \tau(X_j)\rangle_1\langle\tau(X_l), \tau(X_t)\rangle_1 \cdot \tilde{Z}_i^\top\Omega_1^{-1}\tilde{Z}_t \cdot \tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_j\} \\ & +(16)(4)E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \langle\tau(X_i), \tau(X_j)\rangle_1\langle\tau(X_i), \tau(X_t)\rangle_1\langle\tau(X_l), \tau(X_j)\rangle_1\tilde{Z}_l^\top\Omega_1^{-1}\tilde{Z}_t\} \\ & +16E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_t^2 \cdot \langle\tau(X_i), \tau(X_j)\rangle_1\langle\tau(X_i), \tau(X_t)\rangle_1\langle\tau(X_l), \tau(X_j)\rangle_1\langle\tau(X_l), \tau(X_t)\rangle_1\} \\ = \; & S_1 + S_2 + S_3 + S_4 + S_5 + S_6. \end{aligned}$$

Note that in (S1.13), $E\{I(U)\tilde{Z}_i\tilde{Z}_i^\top\} \to \Omega_1$ as $\lambda \to 0$, then

$$E\{I(U)\tilde{Z}^\top\Omega_1^{-1}\tilde{Z}\} = \mathrm{tr}\{\Omega_1^{-1}\mathrm{E}(\mathrm{I}(\mathrm{U})\tilde{\mathrm{Z}}\tilde{\mathrm{Z}}^\top)\} = \mathrm{p}. \qquad \text{(S3.35)}$$

It follows directly that $S_1 = E\{\epsilon^2 \tilde{Z}^\top \Omega_1^{-1} \tilde{Z}\}^4 = E\{I(U) Z^\top \Omega_1^{-1} \tilde{Z}\}^4 = p^4$.

For $i = 1, \cdots, n$ and $v \geq 1$, define $X_v^i = \int_0^1 X_i(t)\varphi_v(t)dt$. One can verify $E\{\epsilon_i^2 \tilde{Z}_i X_v^i\} = E\{I(U)\tilde{Z}_i X_v^i\} = 0$. Recall the definition of $\mathbf{G}(X)$ and $\boldsymbol{A}$ in Assumption 4 and (S1.8), we have

$$
\begin{aligned}
& E\{I(U)\tilde{Z}_i X_v^i\} = E\{E\{I(U)(Z_i - \langle \boldsymbol{A}, \tau(X_i)\rangle)X_v^i | X\}\} \\
= \; & E\{B(X)(\mathbf{G}(X) - \langle \boldsymbol{A}, \tau(X)\rangle)X_v\} = V(\tilde{\beta} - \boldsymbol{A}, \varphi_v) \\
= \; & \sum_v \frac{V(\tilde{\beta}, \varphi_v)}{1 + \lambda\rho_v} \lambda\rho_v \to 0.
\end{aligned}
$$

Condition (3.10) implies that $\sum_v V(\tilde{\beta}, \varphi_v) < \infty$, then the last limit holds as $\lambda \to 0$ by applying the dominated convergence theorem. On the other hand, one can deduce that

$$
\begin{aligned}
S_2 &= E\{\epsilon_i^2 \epsilon_j^2 \tilde{Z}_i^\top \Omega_1^{-1} \tilde{Z}_j \langle \tau(X_i), \tau(X_j)\rangle_1\} \cdot E\{\epsilon^2 \tilde{Z}^\top \Omega_1^{-1} \tilde{Z}\}^2 \\
&= E\left\{\epsilon_i^2 \epsilon_j^2 \tilde{Z}_i^\top \Omega_1^{-1} \tilde{Z}_j \sum_v \frac{X_v^i X_v^j}{1 + \lambda\rho_v}\right\} \cdot p^2 \\
&= \sum_v \frac{E\{\epsilon_i^2 \tilde{Z}^\top X_v^i\} \Omega_1^{-1} E\{\epsilon_j^2 \tilde{Z}_j X_v^j\}}{1 + \lambda\rho_v} \cdot p^2 = 0.
\end{aligned}
$$

Similar to the calculations of S_2, it is easy to find that $S_3 = S_4 = S_5 = 0$.

For S_6, we have

$$\begin{aligned} S_6 &= \sum_{v_1,v_2,v_3,v_4} \frac{E\{\epsilon_i^2 X_{v_1}^i X_{v_2}^i\}E\{\epsilon_j^2 X_{v_1}^j X_{v_3}^j\}E\{\epsilon_l^2 X_{v_3}^l X_{v_4}^l\}E\{\epsilon_t^2 X_{v_2}^t X_{v_4}^t\}}{(1+\lambda\rho_{v_1})(1+\lambda\rho_{v_2})(1+\lambda\rho_{v_3})(1+\lambda\rho_{v_4})} \\ &= \sum_{v_1,v_2,v_3,v_4} \frac{\delta_{v_1,v_2}\delta_{v_1,v_3}\delta_{v_3,v_4}\delta_{v_2,v_4}}{(1+\lambda\rho_{v_1})(1+\lambda\rho_{v_2})(1+\lambda\rho_{v_3})(1+\lambda\rho_{v_4})} \\ &= \sum_v \frac{1}{(1+\lambda\rho_v^4)} = O(h^{-1}). \end{aligned}$$

The summation of S_1 to S_6 leads to $G_{IV} = O(n^4 h^{-1})$.

Now we set out to calculate the order of $\sigma(n)^2$. Specifically,

$$\begin{aligned} \sigma(n)^2 &= E\{W(n)^2\} = 4\sum_{i<j} E\{\epsilon_i^2\epsilon_j^2 |\langle R_{U_i}, R_{U_j}\rangle|^2\} \\ &= 4C_n^2 E\{\epsilon_i^2\epsilon_j^2 (\tilde{Z}_i^\top \Omega_1^{-1}\tilde{Z}_j + \langle \tau(X_i), \tau(X_j)\rangle_1)^2\} \\ &= 4C_n^2 E\{\epsilon_i^2\epsilon_j^2 (\tilde{Z}_i^\top \Omega_1^{-1}\tilde{Z}_j \tilde{Z}_j^\top \Omega_1^{-1}\tilde{Z}_i + \langle \tau(X_i), \tau(X_j)\rangle_1^2)\}. \end{aligned}$$

Notice that $E\{\epsilon_i^2\epsilon_j^2(Z_i^\top \boldsymbol{A} Z_j Z_j^\top \boldsymbol{A} Z_i\} = p$ and

$$\begin{aligned} E\{\epsilon_i^2\epsilon_j^2\langle \tau(X_i), \tau(X_j)\rangle_1^2\} &= E\{\epsilon_i^2\epsilon_j^2 \sum_v \frac{{X_v^i}^2 {X_v^j}^2}{(1+\lambda\rho_v})^2 \\ = \sum_v \frac{E\{I(U_i){X_v^i}^2\}E\{I(U_j){X_v^j}^2\}}{(1+\lambda\rho_v)^2} &= \sum_v \frac{1}{(1+\lambda\rho_v)^2}. \end{aligned}$$

Recall that $\sigma_l^2 = h\sum_v (1+\lambda\rho_v)^{-l}$, then

$$\sigma(n)^2 = 2n(n-1)(p + \sum_v \frac{1}{(1+\lambda\rho_v)^2}) \asymp 2n^2(p + h^{-1}\sigma_2^2).$$

It is obvious that G_I, G_{II}, G_{IV} are of lower orders than $\sigma(n)^4 \asymp 4n^4 h^{-2}\sigma_2^2$.

Then by Proposition 3.2 of de Jong (1987), as $n \to \infty$, we have

$$\frac{W(n)}{\sqrt{2n^2 h^{-1}\sigma_2^2}} \xrightarrow{d} N(0,1).$$

Since

$$E\{|\sum_{i=1}^{n}[\epsilon_i^2\|R_{U_i}\|^2 - E\{\epsilon_i^2\|R_{U_i}^2\|\}]|^2\} \leq nE\{\epsilon^4|\langle R_U, R_U\rangle|^2\} = O(nh^{-2}),$$

then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\|R_{U_i}\|^2 &= E\{\epsilon_i^2\|R_{U_i}\|^2\} + O_p(n^{-1/2}h^{-1})\\ &= p + h^{-1}\sigma_1^2 + O_p(n^{-1/2}h^{-1}).\end{aligned}$$

Thus,

$$\begin{aligned}n\|S_{n,\lambda}(\theta_0)\|^2 &= \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\|R_{U_i}\|^2 + \frac{1}{n}W(n) + O_p(h^{-1/2}+n\lambda)\\ &= O_p(h^{-1}+n^{-1/2}h^{-1}+h^{-1/2}+n\lambda) = O_p(h^{-1}).\end{aligned}$$

Therefore, it follows by (S3.33) that

$$\begin{aligned}T_P &= \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\|R_{U_i}\|^2 + \frac{1}{n}W(n) + n\|W_\lambda\beta_0\|_1 + o_p(h^{-1/2})\\ &= p + h^{-1}\sigma_1^2 + \frac{1}{n}W(n) + n\|W_\lambda\beta_0\|_1 + o_p(h^{-1/2}).\end{aligned}$$

This leads to the conclusion that as $n \to \infty$,

$$\begin{aligned}&\frac{T_P - (h^{-1}\sigma_1^2 + n\|W_\lambda\beta_0\|_1 + p)}{\sqrt{2(\sigma_2^2h^{-1}+p)}}\\ =\ &(2u_n + 2p\sigma^2)^{-1/2}(\sigma^2\cdot T_P - (u_n + p\sigma^2 + n\sigma^2\|W_\lambda\beta_0\|_1)) \xrightarrow{d} N(0,1).\end{aligned}$$

Besides, it can be shown that $n\|W_\lambda\beta_0\|_1 = o(n\lambda) = o(u_n)$. Therefore σ^2T_P is asymptotically $N(u_n + p\sigma^2, 2u_n + 2p\sigma^2)$. This completes the proof.

□

S4 Impacts of measurement errors

The theoretical results are based on the underlying assumption that the functional covariate $X(t)$ is observed completely. However, $X(t)$ is usually observed intermittently and with errors in practice. Here we discuss potential challenges to achieving similar theoretical results if we plug in an empirical version of $X(t)$.

We observe that

$$W_{ij} = X_i(t_{ij}) + e_{ij}, \quad j = 1, \ldots, m_i,$$

where e_{ij} are independent zero-mean errors independent of X_i, with $Var(e_{ij}) = \sigma_e^2$. We smooth each curve to obtain an estimate $\hat{X}_i(t) = \hat{\theta}_0(t)$ of X_i by a local linear regression,

$$(\hat{\theta}_0, \hat{\theta}_1) = \arg\min_{(\theta_0,\theta_1)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{W_{ij} - \theta_0 - \theta_1(t_{ij} - t)\}^2 K\{(t_{ij} - t)/h_w\},$$

where $K(\cdot)$ is a kernel function and h_w is the bandwidth for the smoothing step. If dense measurements are made on each curve, we can effectively eliminate effects from measurement errors and pretend that we know the true curve. We can use $\hat{X}_i(t)$ to perform estimation and hypothesis testings. The following conditions used in Kong et al. (2016) ensure that $\|\hat{X}_i(t) - X_i(t)\|_{L^2} = o_p(n^{-1/2})$. Denote that $\tilde{m} = \inf_{i=1,\ldots,n} m_i$.

(A-1). For any $C > 0$, there exists an $\epsilon > 0$ such that $\sup_{t\in\mathbb{I}}\{E|X(s)|^C\} < \infty$, and $\sup_{s,t\in\mathbb{I}}\{E[(|s-t|^{-\epsilon}|X(s)-X(t)|)^C]\} < \infty$.

(A-2). X is twice continuously differentiable on $\mathbb{I}$ with probability 1, and $\int E(X^{(2)}(t))^4 dt < \infty$, where $X^{(2)}(t)$ denotes the second derivative of $X(t)$.

(A-3). The observation points $\{t_{ij}, j = 1, \ldots, m_i\}$ are deterministic and ordered increasingly for $i = 1, \ldots, n$. There exist densities g_i uniformly smooth over i, satisfying $\int_0^1 g_i(t)dt = 1$ and $0 < c_1 < \inf_i\{\inf_{t\in\mathbb{I}} g_i(t)\} < \sup_i\{\sup_{t\in\mathbb{I}} g_i(t)\} < c_2 < \infty$. The t_{ij}s are generated according to $t_{ij} = G_i^{-1}\{j/(m_i+1)\}$, where G_i^{-1} is the inverse of $G_{ij} = \int_{-\infty}^t g_i(s)ds$. The kernel density function is smooth and compactly supported.

(A-4). $\sup_i \sup\{t_{i(j+1)} - t_{ij}, j = 1, \ldots, m_i\} = O(\tilde{m}^{-1})$, $h_w \sim \tilde{m}^{-1/5}$, $\tilde{m}n^{-5/4} \to \infty$.

Such a "smooth first, then perform estimation" procedure was widely adopted in the literature (Li et al., 2010; Zhang and Chen, 2007; Wong et al., 2019). From the simulation results below, it can be seen that the smoothing procedure is quite useful especially when the variance of e_{ij} is small and the curves are densely observed.

The penalized estimator using $\hat{X}_i(t)$ instead of $X_i(t)$ is obtained by

$\hat{\tilde{\theta}}_{n,\lambda} = (\hat{\tilde{\gamma}}_{n,\lambda}, \hat{\tilde{\beta}}_{n,\lambda}) = \arg\sup_{\theta\in\mathcal{H}} \tilde{\ell}_{n,\lambda}(\theta)$, where

$$\tilde{\ell}_{n,\lambda}(\theta) = \left\{ \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i; Z_i^\top \gamma + \int_0^1 \hat{X}_i(t)\beta(t)dt) - (\lambda/2)J(\beta,\beta) \right\}. \quad \text{(S4.36)}$$

The Fréchet derivative of $\tilde{\ell}_{n,\lambda}(\theta)$ with respect to θ is

$$\tilde{S}_{n,\lambda}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_a(Y_i; \langle R_{\hat{U}_i}, \theta\rangle) R_{\hat{U}_i} - P_\lambda \theta,$$

where $\hat{U}_i = (\hat{X}_i, Z)$. Also, define $\tilde{S}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_a(Y_i; \langle R_{\hat{U}_i}, \theta\rangle) R_{\hat{U}_i}$, $\tilde{S}(\theta) = E\{\tilde{S}_n(\theta)\}$, $\tilde{S}_\lambda(\theta) = E\{\tilde{S}_{n,\lambda}(\theta)\}$ and $\tilde{\epsilon}_i = \dot{\ell}_a(Y_i; Z_i^\top \gamma_0 + \int_0^1 \hat{X}_i(t)\beta_0(t)dt)$. Let $\tilde{H}_n(\theta)$ be the term when using $\hat{X}_i(t)$ in $H_n(\theta)$ defined in (S2.17).

By examining the proofs of the theoretical results, roughly, it is required to quantify the asymptotic orders of several important types of expressions. Denote $\tilde{S}_\lambda(\tilde{\theta}_\lambda) = 0, S_\lambda(\theta_\lambda) = 0$, the expressions are as follows,

$$\eta_1 = \tilde{S}_\lambda(\theta_0),\ \eta_2 = \theta + D\tilde{S}_\lambda(\theta_0)\theta,\ \eta_3 = E\{\langle R_{\hat{U}_i}, \theta\rangle^2 \|R_{\hat{U}_i}\| - \langle R_{U_i}, \theta\rangle^2 \|R_{U_i}\|\},$$

$$\eta_4 = \int_0^1\int_0^1 sE\{|\langle R_{\hat{U}_i}, \theta_1-\theta_2\rangle| \cdot |\langle R_{\hat{U}_i}, \theta_2 + s(\theta_1-\theta_2)\rangle| \cdot \|R_{\hat{U}_i}\|\}dsds',$$
$$-\int_0^1\int_0^1 sE\{|\langle R_{U_i}, \theta_1-\theta_2\rangle| \cdot |\langle R_{U_i}, \theta_2 + s(\theta_1-\theta_2)\rangle| \cdot \|R_{U_i}\|\}dsds',$$

$$\eta_5 = |[D\tilde{S}_\lambda(\tilde{\theta}_\lambda) - \tilde{S}_\lambda(\theta_0)]\theta\theta' - [DS_\lambda(\theta_\lambda) - S_\lambda(\theta_0)]\theta\theta'|,$$

$$\eta_6 = E\{\tilde{\epsilon}_i \|R_{\hat{U}_i}\| - \epsilon_i \|R_{U_i}\|\},\ \eta_7 = \sup_{\theta\in\mathcal{F}_{p_n}} \|\tilde{H}_n(\theta) - H_n(\theta)\|,$$

$$\eta_8 = E\{(|\dot{\ell}_a(Y_i; \langle R_{\hat{U}_i}, \tilde{\theta}_\lambda\rangle) - \dot{\ell}_a(Y_i; \langle R_{\hat{U}_i}, \theta_0\rangle)|)\|R_{\hat{U}_i}\|\},$$

$$\eta_9 = \sup_{\theta'=1} E\{|\langle R_{\hat{U}_i}, \theta\rangle|^2 \cdot |\langle R_{\hat{U}_i}, \theta'\rangle|^2\} - \sup_{\theta'=1} E\{|\langle R_{U_i}, \theta\rangle|^2 \cdot |\langle R_{U_i}, \theta'\rangle|^2\}.$$

Denote $\mathbb{B}(\varepsilon) = \{\theta \in \mathcal{H} : \|\theta\| \le \varepsilon\}$. The following lemmas provide the conditions under which the theoretical results still hold when we plug in an empirical version of $X(t)$.

Lemma 5. *If* $\|\eta_1\| = O(h^k)$*, and for any* $\theta, \theta_1, \theta_2 \in \mathbb{B}(2(J(\beta_0,\beta_0)+1)^{1/2}h^k)$*, the following conditions hold,* $\|\eta_2\| \leq h^{-1/2}\|\theta\|^2$*,* $\eta_3 \leq h^{-1/2}\|\theta\|^2$*,* $\eta_4 \leq 1/2\|\theta_1 - \theta_2\|$*,* $\eta_5 = o(1)$*,* $\eta_6 = o((nh)^{-1})$*,* $\eta_7 = o(p_n^{1/(4m)}(h^{-1}\log\log n)^{1/2})$ *and* $\eta_8 = o((nh)^{-1})$*. Meanwhile, for any* $\theta \in \mathbb{B}(C(nh)^{-1/2})$*, where* $C > 0$ *is a constant,* $|\eta_9| = o(\|\theta\|)$*. Then we have*

$$\|\hat{\tilde{\theta}}_{n,\lambda} - \theta_0\| = O_p((nh)^{-1/2} + h^k).$$

Lemma 6. *Suppose the conditions in Lemma 5 are satisfied. Recall that* a_n *is defined in Theorem 2. Additionally, for* $\theta = \hat{\tilde{\theta}}_{n,\lambda} - \theta_0$*, the following conditions hold,* $\|\tilde{S}_{n,\lambda}(\theta+\theta_0) - \tilde{S}_{n,\lambda}(\theta_0) - E\{\tilde{S}_{n,\lambda}(\theta+\theta_0) - \tilde{S}_{n,\lambda}(\theta_0)\}\| \leq O(n^{-1/2}h^{-\frac{4ma+6m-1}{4m}}r_n(\log n)^2(\log\log n)^{1/2})$*,* $\|E\{D\tilde{S}_{n,\lambda}(\theta_0)\theta - \theta\}\| = o_p(a_n)$*, and* $\|\int_0^1\int_0^1 sE\{D\tilde{S}_{n,\lambda}(\theta_0 + ss'\theta)\theta\theta ds ds'\}\| \leq O(h^{-1/2}r_n^2)$*. Then we have*

$$\|\hat{\tilde{\theta}}_{n,\lambda} - \theta_0 - \tilde{S}_{n,\lambda}(\theta_0)\| = O_p(a_n).$$

Lemma 7. *Suppose the conditions in Lemma 6 hold. Denote* $\hat{u}^* = (\tilde{z}, \hat{\tilde{x}}_0)$ *for any* $\tilde{z} \in \mathbb{R}^p$*. If* $\langle R_{\hat{u}^*} - R_{u^*}, \hat{\tilde{\theta}}_{n,\lambda} - \hat{\theta}_{n,\lambda}\rangle = o_p(n^{-1/2})$*,* $\langle R_{\hat{u}^*} - R_{u^*}, \hat{\theta}_{n,\lambda} - \theta_0\rangle = o_p(n^{-1/2})$*, and* $\langle R_{\hat{u}^*} - R_{u^*}, \hat{\tilde{\theta}}_{n,\lambda} - \hat{\theta}_{n,\lambda}\rangle = o_p(n^{-1/2})$*, then the joint independence result in Theorem 3 can still be achieved if we use an empirical version of* $\tilde{x}_0(t)$*.*

Lemma 8. *For the penalized likelihood ratio test statistic*

$$\tilde{T}_P = -2n\{\tilde{\ell}_{n,\lambda}(\theta_0) - \tilde{\ell}_{n,\lambda}(\hat{\tilde{\theta}}_{n,\lambda})\}. \tag{S4.37}$$

If the conditions in Lemma 6 are satisfied, further

$$\sup_{\|\theta-\theta_0\|\leq C((nh)^{-1/2}+h^k)} n|\tilde{\ell}_{n,\lambda}(\theta) - \ell_{n,\lambda}(\theta)| = o_p(u_n + p\sigma^2),$$

where σ^2 and u_n are defined in Theorem 4, then $\sigma^2\tilde{T}_P$ is also asymptotically $N(u_n + p\sigma^2, 2u_n + 2p\sigma^2)$.

Let $\hat{C}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \hat{B}(\hat{X}_i)\hat{X}_i(s)\hat{X}_i(t)$ be an estimate of C, where $\hat{B}(\hat{X}_i) = -\frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_a(y_i; z_i^\top\hat{\tilde{\gamma}}_{n,\lambda} + \int_0^1 \hat{x}_i(t)\hat{\tilde{\beta}}_{n,\lambda}(t)dt)$. Then we can obtain an estimate of $V(\beta_1,\beta_2)$ such that $\hat{V}(\beta_1,\beta_2) = \int_0^1 \hat{C}(s,t)\beta_1(s)\beta_2(t)dsdt$. Denote $(\hat{\rho}_v, \hat{\varphi}_v)$ as the eigen-pairs driven by $\hat{C}$. The last step is to show that the limit distribution also holds if we use $\hat{\sigma}_l$ instead of σ_l in practice. The key step is to show $|\hat{\sigma}_l^2 - \sigma_l^2| = o_p(1)$.

Following similar procedures in Kong et al. (2016), we can have $\int\int(\hat{C}(s,t) - C(s,t))^2 dsdt = O_p(n^{-1})$ if conditions (A-1)-(A-4) hold. Then in analogy to the arguments of Shang and Cheng (2015), we can have $|\hat{\sigma}_l^2 - \sigma_l^2| = o_p(1)$.

In general, the proofs of the theoretical developments rely heavily on the inner products defined in (2.5) and (2.6), which involve the fully observed trajectory. Apart from figuring out the errors to the eigen-system, we not only need to explore the impacts of measurement errors on the inner products, but also need to clarify the effects on several expressions in relation to $X(t)$ in complex forms. It requires greater effort to verify the conditions in Lemmas 5-8. These issues need to be addressed in future research.

S5 Simulation results with measurement errors

In this section, we conduct additional simulations to explore the impacts of measurement errors of the functional variable on the performance of the proposed test. *Example 1* explores the impact of the variance of measurement errors on the performance of the proposed test. *Example 2* investigates the effect of the sparsity of the observation points.

Example 1. We compute the sizes and powers of the proposed test when testing $H_0 : \beta = 0$ and $\gamma = 0$ and $H_0 : \beta = 0$ under the PFLM setting and the PFLGM setting with sample size $n \in \{100, 500\}$. We run 1000 replicates for each case. Data are simulated in the same way as that in *Case 1* in the main text except that the functional predictor $X_i(t)$ are not fully observed. We assume the actual observation X_{ij} is the realization of $X_i(t)$ at 200 evenly spaced points $\{T_{ij}, j = 1, \cdots, 200\}$ with i.i.d. error $e_{ij} \sim N(0, \sigma_e^2)$, and $\sigma_e \in \{0.5, 1, 1.5\}$.

Table S1 and Table S2 provide the results when testing $H_0 : \beta = 0$ and $\gamma = 0$ under the PFLM setting and the PFLGRM setting. Sizes and powers when testing $H_0 : \beta = 0$ are summarized in Tables S3-S4. Recall that T_P denotes the proposed test, T_S, T_W, T_L and T_F denote the score test, Wald test, modified likelihood ratio test and F test in Kong, Staicu, and Maity (2016), and T_W^* denotes the test method of Su, Di, and Hsu (2017). It can be seen that if the errors are small, the sizes and powers behave similar to the sizes and powers when $X(t)$s are fully observed. Meanwhile, we also plot changes of sizes and powers with σ_e ranging from 0.5 to 4 when testing $H_0 : \beta = 0$ under the PFLM setting in Figure S5. Under the alternative hypothesis, we set $\xi = 0.1$ and $B = 1$. The proposed method still outperforms the competing methods in all scenarios.

Example 2. The data settings are similar to that in *Example 1*, except that $X_i(t)$ are observed with fewer observation points. We set the number of points to be $\tilde{m} \in \{30, 50, 100\}$. The variance of the measurement errors is fixed at $\sigma_e = 1$. The results are summarized in Tables S5 - S8. We can see that all the methods lose power as the sparsity level becomes higher. However, when observation points are sufficiently dense, the results are similar to knowing the entire trajectory of each X_i.

Table S1: Sizes and powers in the PFLM setting when testing $H_0 : \beta = 0$ and $\gamma = 0$ with measurement errors.

n	σ_e	(γ_1, γ_2)	$B = 0$	$\xi = 0.1$ $B = 0.1$	$\xi = 0.1$ $B = 0.5$	$\xi = 0.1$ $B = 1$	$\xi = 0.5$ $B = 0.1$	$\xi = 0.5$ $B = 0.5$	$\xi = 0.5$ $B = 1$
100	0.5	(0.0,0.0)	5.4	8.0	20.0	63.1	9.0	54.8	98.5
		(0.1,0.1)	21.2	21.7	35.4	71.7	21.0	63.6	98.9
		(0.2,0.2)	64.9	62.1	74.7	90.7	64.8	87.5	99.5
		(0.3,0.3)	94.0	94.8	95.9	98.8	93.5	98.1	100
	1.0	(0.0,0.0)	5.5	8.2	18.4	60.7	7.7	53.7	97.6
		(0.1,0.1)	20.2	21.2	34.7	69.1	20.6	63.1	98.7
		(0.2,0.2)	63.5	63.8	72.6	90.1	63.5	87.6	99.4
		(0.3,0.3)	94.1	92.7	96.1	98.5	94.0	98.2	100
	1.5	(0.0,0.0)	5.5	7.2	18.4	58.0	7.5	52.8	98.8
		(0.1,0.1)	20.3	20.7	33.1	70.4	21.1	62.9	99.7
		(0.2,0.2)	63.9	61.5	70.7	89.6	62.9	85.7	99.2
		(0.3,0.3)	94.5	92.9	95.2	98.4	94.3	97.8	100
500	0.5	(0.0,0.0)	5.2	10.6	71.7	100	17.1	100	100
		(0.1,0.1)	72.2	75.9	96.5	100	79.8	100	100
		(0.2,0.2)	99.9	100	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100
	1.0	(0.0,0.0)	5.3	10.1	71.7	100	15.2	99.6	100
		(0.1,0.1)	74.0	74.8	96.7	100	78.7	100	100
		(0.2,0.2)	100	99.9	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100
	1.5	(0.0,0.0)	5.1	9.7	73.6	100	14.1	99.9	100
		(0.1,0.1)	73.2	73.7	97.2	100	78.6	100	100
		(0.2,0.2)	100	100	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100

Table S2: Sizes and powers in the PFLGRM setting when testing $H_0 : \beta = 0$ and $\gamma = 0$ with measurement errors.

n	σ_e			$\xi = 0.1$			$\xi = 0.5$		
		(γ_1, γ_2)	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	0.5	(0.0,0.0)	5.6	5.2	6.5	13.5	5.2	10.1	43.4
		(0.1,0.1)	7.4	7.6	8.3	18.6	7.7	14.3	45.4
		(0.2,0.2)	14.9	14.9	19.3	26.2	13.5	24.5	57.2
		(0.3,0.3)	28.9	29.6	33.6	42.3	32.0	42.7	64.8
	1.0	(0.0,0.0)	5.5	5.4	5.2	12.8	5.5	10.4	41.6
		(0.1,0.1)	7.3	7.2	8.6	17.2	6.9	13.1	40.9
		(0.2,0.2)	14.3	13.1	17.9	23.8	11.8	24.3	55.2
		(0.3,0.3)	29.8	28.7	32.1	41.2	29.7	40.3	63.3
	1.5	(0.0,0.0)	5.6	5.2	6.1	11.9	6.0	9.0	35.4
		(0.1,0.1)	6.6	6.4	8.0	16.3	6.6	12.9	39.7
		(0.2,0.2)	13.5	12.2	15.6	23.8	11.8	20.1	53.0
		(0.3,0.3)	26.5	27.7	30.0	38.3	27.8	39.1	60.5
500	0.5	(0.0,0.0)	5.4	5.4	19.2	65.6	6.3	57.6	99.7
		(0.1,0.1)	19.8	21.2	38.2	79.0	21.1	72.2	100
		(0.2,0.2)	67.7	66.2	78.7	95.8	71.5	93.1	100
		(0.3,0.3)	97.5	97.8	98.5	99.7	97.4	99.5	100
	1.0	(0.0,0.0)	5.2	5.1	18.9	66.8	5.0	57.8	99.4
		(0.1,0.1)	20.5	20.0	38.4	78.0	20.5	68.0	99.7
		(0.2,0.2)	67.3	62.4	76.4	93.6	69.2	91.9	100
		(0.3,0.3)	96.6	96.4	97.7	99.2	95.7	99.2	100
	1.5	(0.0,0.0)	4.9	5.0	18.3	66.7	5.5	55.3	99.7
		(0.1,0.1)	15.3	18.0	35.9	76.0	16.3	65.8	99.4
		(0.2,0.2)	63.7	63.3	76.4	92.5	62.8	92.1	100
		(0.3,0.3)	96.3	95.2	97.5	99.1	96.5	99.5	100

Table S3: Sizes and powers in the PFLM setting when testing $H_0 : \beta = 0$ with measurement errors.

n	σ_e			$\xi = 0.1$			$\xi = 0.5$		
			$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	0.5	T_P	5.1	20.2	46.1	89.5	20.2	80.7	99.6
		T_S	5.5	5.8	18.5	59.2	7.1	52.4	98.7
		T_W	5.7	6.0	19.0	59.7	7.3	53.2	98.8
		T_L	5.8	6.1	19.0	59.9	7.4	53.5	98.9
		T_F	5.3	5.6	17.8	58.6	6.9	51.7	98.6
		T_W^*	5.4	6.7	18.1	56.1	6.5	47.7	98.0
	1.0	T_P	5.4	18.6	45.5	87.4	19.4	80.4	99.9
		T_S	5.3	5.2	18.2	59.4	7.1	50.5	98.6
		T_W	5.5	5.5	19.5	60.7	7.4	51.0	98.6
		T_L	5.7	5.7	19.5	61.1	7.5	51.5	98.6
		T_F	5.2	5.1	17.3	58.3	6.7	49.5	98.4
		T_W^*	5.6	5.8	16.2	55.6	6.9	45.7	97.0
	1.5	T_P	5.2	16.0	43.0	86.8	18.3	81.5	99.8
		T_S	5.2	6.3	16.9	60.5	5.9	50.2	98.2
		T_W	5.8	7.3	17.9	61.2	6.1	50.4	98.3
		T_L	5.7	7.4	18.2	61.2	6.2	50.5	98.3
		T_F	5.4	5.8	16.2	59.8	5.7	49.2	98.1
		T_W^*	5.4	6.8	16.5	55.5	5.4	45.4	97.1
500	0.5	T_P	5.6	22.5	92.3	100	35.1	100	100
		T_S	5.8	7.5	72.4	100	13.8	100	100
		T_W	5.5	7.5	72.8	100	14.0	100	100
		T_L	5.8	7.5	72.8	100	13.4	100	100
		T_F	5.7	7.3	72.3	100	13.8	100	100
		T_W^*	5.2	7.5	64.6	100	11.6	100	100
	1.0	T_P	5.3	20.5	92.1	100	32.5	100	100
		T_S	5.6	7.1	71.8	100	11.8	99.8	100
		T_W	5.7	7.2	71.9	100	12.1	99.8	100
		T_L	5.6	7.2	71.9	100	12.2	99.8	100
		T_F	5.4	7.1	71.7	100	11.8	99.8	100
		T_W^*	5.3	7.5	63.3	100	11.0	99.5	100
	1.5	T_P	5.5	19.5	92.2	100	32.4	100	100
		T_S	5.5	6.3	69.3	100	12.0	100	100
		T_W	5.6	6.4	69.3	100	12.2	100	100
		T_L	5.6	6.4	69.3	100	12.4	100	100
		T_F	5.4	6.1	68.9	100	11.7	100	100
		T_W^*	5.2	6.0	62.3	100	10.5	100	100

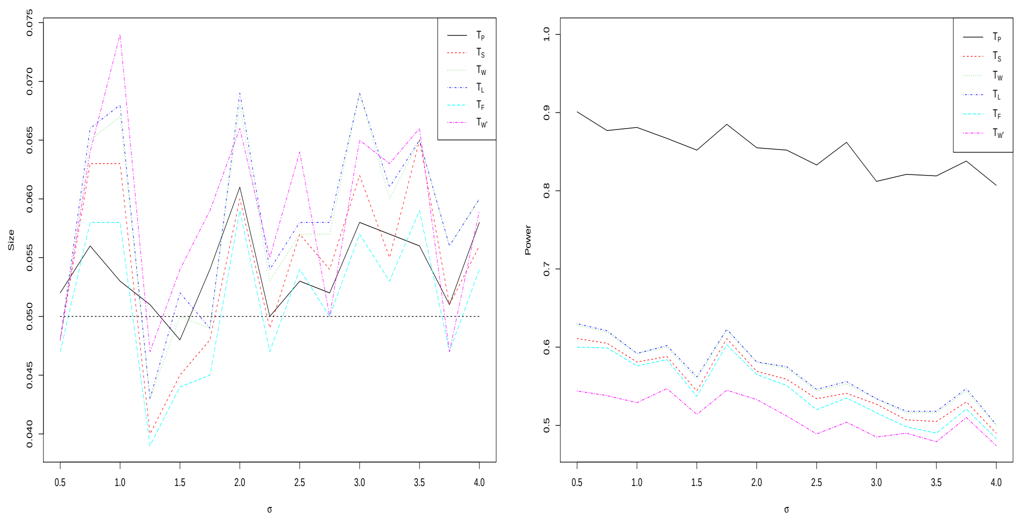

Figure S1: Changes of sizes and powers with σ_e under $H_0 : \beta = 0$ in the PFLM setting

Table S4: Sizes and powers in the PFLGRM setting when testing $H_0 : \beta = 0$ with measurement errors.

n	σ_e		$\xi = 0.1$			$\xi = 0.5$		
(γ_1, γ_2)		$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	0.5	5.5	5.2	7.1	20.4	6.3	17.5	56.0
	1.0	5.4	5.5	6.9	20.0	5.7	15.3	52.5
	1.5	5.3	5.1	6.4	17.3	6.4	15.4	49.4
500	0.5	5.4	6.6	25.7	74.3	8.0	70.0	99.9
	1.0	5.1	5.9	25.9	75.6	7.4	68.7	99.8
	1.5	5.6	5.7	24.4	73.4	7.6	68.3	99.8

Table S5: Sizes and powers in the PFLM setting when testing $H_0 : \beta = 0$ and $\gamma = 0$ with different number of observation points.

n	$\tilde{m}$			$\xi = 0.1$			$\xi = 0.5$		
		(γ_1, γ_2)	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	30	(0.0,0.0)	5.8	7.0	17.4	55.6	7.2	46.9	95.9
		(0.1,0.1)	19.2	19.3	31.9	65.8	18.8	62.5	97.6
		(0.2,0.2)	62.4	62.7	69.8	88.1	63.9	85.0	98.8
		(0.3,0.3)	93.5	94.0	95.1	98.3	93.3	98.3	100
	50	(0.0,0.0)	5.6	7.4	17.6	57.7	7.8	50.8	97.8
		(0.1,0.1)	20.4	18.3	33.1	67.6	20.4	63.5	98.1
		(0.2,0.2)	64.0	64.0	70.1	89.2	64.0	86.1	99.6
		(0.3,0.3)	93.6	93.0	95.7	98.4	94.7	98.0	100
	100	(0.0,0.0)	5.7	7.9	19.1	60.6	7.5	52.6	97.7
		(0.1,0.1)	20.1	19.9	34.5	69.6	21.1	65.0	98.2
		(0.2,0.2)	63.7	63.5	71.4	90.0	64.0	86.6	99.7
		(0.3,0.3)	94.9	95.5	96.8	98.8	95.0	98.2	100
500	30	(0.0,0.0)	5.6	7.2	62.1	99.6	11.3	99.0	100
		(0.1,0.1)	72.3	72.8	94.4	100	75.3	99.7	100
		(0.2,0.2)	100	100	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100
	50	(0.0,0.0)	5.3	9.0	67.8	99.8	11.8	99.3	100
		(0.1,0.1)	71.6	73.7	95.5	100	76.4	100	100
		(0.2,0.2)	100	99.8	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100
	100	(0.0,0.0)	5.0	10.0	68.3	100	14.8	99.4	100
		(0.1,0.1)	73.8	74.2	95.7	100	74.9	100	100
		(0.2,0.2)	100	100	100	100	100	100	100
		(0.3,0.3)	100	100	100	100	100	100	100

Table S6: Sizes and powers in the PFLGRM setting when testing $H_0 : \beta = 0$ and $\gamma = 0$ with different number of observation points.

n	$\tilde{m}$			$\xi = 0.1$			$\xi = 0.5$		
		(γ_1, γ_2)	$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	30	(0.0,0.0)	5.6	5.4	6.4	10.7	5.7	10.0	37.5
		(0.1,0.1)	6.6	6.2	7.5	15.9	5.6	12.2	39.1
		(0.2,0.2)	11.9	11.6	16.3	19.9	11.5	22.1	51.5
		(0.3,0.3)	28.1	29.9	29.4	40.0	29.6	36.6	61.6
	50	(0.0,0.0)	5.1	5.7	6.0	12.8	5.9	11.2	39.5
		(0.1,0.1)	7.2	7.8	8.9	16.7	6.6	12.9	40.7
		(0.2,0.2)	14.1	12.1	17.7	22.2	13.9	24.3	54.0
		(0.3,0.3)	27.2	31.1	31.6	39.7	27.9	38.6	63.2
	100	(0.0,0.0)	5.4	5.5	6.3	13.9	5.6	11.3	40.8
		(0.1,0.1)	7.1	8.5	8.1	17.5	7.2	14.4	41.3
		(0.2,0.2)	14.0	17.6	18.4	25.7	14.6	24.9	54.5
		(0.3,0.3)	31.2	32.8	33.5	42.6	29.6	40.1	64.1
500	30	(0.0,0.0)	5.3	5.6	15.0	62.0	6.1	54.6	98.8
		(0.1,0.1)	19.4	17.2	35.8	74.8	20.9	65.0	99.6
		(0.2,0.2)	66.4	63.8	73.7	92.7	66.9	91.2	99.9
		(0.3,0.3)	96.0	96.5	97.6	99.6	96.1	99.1	100
	50	(0.0,0.0)	5.3	6.3	17.6	67.5	5.0	58.1	99.1
		(0.1,0.1)	20.1	19.7	38.4	75.8	20.0	67.5	99.4
		(0.2,0.2)	67.1	63.4	77.5	93.5	69.1	91.0	99.9
		(0.3,0.3)	96.0	96.7	98.1	99.5	96.2	99.7	100
	100	(0.0,0.0)	5.6	6.3	19.0	66.2	6.0	58.9	99.4
		(0.1,0.1)	20.8	19.9	38.5	77.9	20.8	67.3	99.8
		(0.2,0.2)	67.1	63.4	78.3	93.5	69.6	92.5	100
		(0.3,0.3)	96.9	97.4	98.8	99.7	97.2	99.7	100

Table S7: Sizes and powers in the PFLM setting when testing $H_0 : \beta = 0$ with different number of observation points.

n	$\tilde{m}$			$\xi = 0.1$			$\xi = 0.5$		
			$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	30	T_P	5.3	16.8	41.3	84.0	17.4	78.7	99.6
		T_S	5.1	5.4	17.1	56.4	6.4	49.1	97.6
		T_W	5.5	5.6	17.9	57.2	6.7	50.0	97.2
		T_L	5.8	5.6	18.0	57.4	6.8	50.3	97.6
		T_F	5.2	5.3	17.0	55.7	6.4	48.9	98.1
		T_W^*	5.5	5.5	15.2	53.5	6.1	44.5	96.4
	50	T_P	5.1	16.1	42.6	84.7	19.0	79.2	99.8
		T_S	5.2	5.6	17.8	58.6	6.7	50.2	98.3
		T_W	5.6	5.8	19.1	59.7	7.1	50.3	98.4
		T_L	5.7	5.7	19.0	60.0	7.2	50.5	98.5
		T_F	5.3	5.8	17.3	58.9	6.8	49.2	98.0
		T_W^*	5.6	5.5	16.0	53.6	6.7	45.1	96.6
	100	T_P	5.2	17.2	44.2	86.3	19.7	80.3	99.8
		T_S	5.4	5.5	18.4	58.9	7.2	50.0	98.8
		T_W	5.6	5.7	19.3	60.0	7.5	51.2	98.9
		T_L	5.4	5.5	19.6	61.2	7.1	51.1	98.9
		T_F	5.1	5.3	17.7	58.5	6.6	49.2	98.5
		T_W^*	5.4	5.3	16.3	55.0	6.9	46.0	97.5
500	30	T_P	5.5	18.9	90.6	100	30.6	99.9	100
		T_S	5.3	6.7	68.7	100	12.2	99.6	100
		T_W	5.4	6.9	68.8	100	12.2	99.6	100
		T_L	5.4	6.9	68.8	100	12.2	99.6	100
		T_F	5.3	6.6	68.5	100	12.1	99.5	100
		T_W^*	5.2	6.8	62.7	99.9	10.9	99.5	100
	50	T_P	5.3	20.5	92.1	100	32.5	100	100
		T_S	5.6	7.1	71.8	100	11.8	99.8	100
		T_W	5.7	7.2	71.9	100	12.1	99.8	100
		T_L	5.6	7.2	71.9	100	12.2	99.8	100
		T_F	5.4	7.1	71.7	100	11.8	99.8	100
		T_W^*	5.3	7.5	63.3	100	11.0	99.5	100
	50	T_P	5.2	19.6	91.1	100	31.8	100	100
		T_S	5.4	7.0	69.7	100	12.0	99.9	100
		T_W	5.5	7.0	69.9	100	12.0	99.9	100
		T_L	5.5	7.0	70.0	100	12.1	99.9	100
		T_F	5.3	6.9	69.3	100	11.8	99.9	100
		T_W^*	5.2	6.8	63.2	100	11.0	99.6	100
	100	T_P	5.3	20.7	92.1	100	33.5	100	100
		T_S	5.4	7.2	71.3	100	12.7	99.9	100
		T_W	5.4	7.3	71.5	100	12.7	99.9	100
		T_L	5.4	7.3	71.5	100	12.7	99.9	100
		T_F	5.2	7.1	71.1	100	12.5	99.9	100
		T_W^*	5.3	7.3	63.4	100	11.2	99.8	100

Table S8: Sizes and powers in the PFLGRM setting when testing $H_0 : \beta = 0$ with different number of observation points.

n	$\tilde{m}$		$\xi = 0.1$			$\xi = 0.5$		
(γ_1, γ_2)		$B = 0$	$B = 0.1$	$B = 0.5$	$B = 1$	$B = 0.1$	$B = 0.5$	$B = 1$
100	30	5.7	5.4	6.8	18.9	6.0	13.1	50.0
	50	5.3	5.5	6.9	19.1	6.4	14.5	51.7
	100	5.5	5.7	6.4	20.3	5.9	15.4	52.7
500	30	5.2	6.0	24.1	72.5	7.3	66.5	99.5
	50	5.6	5.7	25.4	72.5	7.4	67.9	99.9
	100	5.3	6.2	27.2	75.5	8.0	67.9	99.9

Bibliography

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields 75*(2), 261–277.

Kong, D., A.-M. Staicu, and A. Maity (2016). Classical testing in functional linear models. *Journal of Nonparametric Statistics 28*(4), 813–838.

Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*(1), 147–159.

Li, Y., N. Wang, and R. J. Carroll (2010). Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association 105*(490), 621–633.

Shang, Z. and G. Cheng (2015). Nonparametric inference in generalized functional linear models. *The Annals of Statistics 43*(4), 1742–1773.

Su, Y.-R., C.-Z. Di, and L. Hsu (2017). Hypothesis testing in functional linear models. *Biometrics 73*(2), 551–561.

Wong, R. K., Y. Li, and Z. Zhu (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association 114*(525), 406–418.

Zhang, J.-T. and J. Chen (2007). Statistical inferences for functional data. *The Annals of Statistics 35*(3), 1052–1079.