Supplementary Document for the Manuscript entitled

"Simultaneous functional quantile regression"

B. HU, X. HU, H. LIU, J. YOU and J. CAO

Department of Statistics and Actuarial Science, Simon Fraser University

Sauder School of Business, University of British Columbia

School of Economics and Finance, Xi'an Jiaotong University

School of Statistics and Management, Shanghai University of Finance

Department of Statistics and Actuarial Science, Simon Fraser University

This supplementary file contains additional simulation results on the proposed method, an introduction to bivariate splines, an introduction to functional principal component analysis (FPCA) and detailed proofs for the key technical conclusions in the main paper together with some preparation.

S1 Simulation Studies

In this section, we use simulation studies to compare the performance of the proposed method and the conventional method in Kato (2012) based on the estimation of $\beta(t, u)$ under three scenarios. In Scenario 1, the true slope function $\beta(t, u)$ is free from u, that is, $\beta(t, u)$ is just a univariate function of t, which is similar to the simulation study in Kato (2012). In Scenario 2, the true slope

function $\beta(t,u)$ is a bivariate function of both t and u, which is more complicated than the slope function in the first scenario and is not considered in Kato (2012). The data generating models are introduced as follows.

S1.1 Data Generating Models

Scenario 1. In this scenario, we suppose that realizations of the $\{X(t), Y\}$ are generated from

$$X(t) = \sum_{i=1}^{10} i^{-1} r_i \phi_i(t) + \mu(t), \quad \mu(t) = \sqrt{3}\phi_1(t),$$
$$Y = \int_0^1 \rho_1(t) X(t) dt + \epsilon, \quad \rho_1(t) = \sum_{j=1}^{10} \tau_j \phi_j(t),$$

where $\phi_j(t) = 2^{1/2}\cos(j\pi t)$, $\tau_1 = 1$, $\tau_j = 4(-1)^{j+1} j^{-2}$ for $j \geq 2$, r_i are i.i.d. Uniform$(-\sqrt{3}, \sqrt{3})$ random variables, and ϵ is an $N(0,1)$ random variables independent of other random variables. Under this setting, the underlying slope function $\beta(t,u)$ is $\beta(t,u) = \rho_1(t) = \sum_{j=1}^{10} \tau_j \phi_j(t)$.

Scenario 2. In this scenario, we suppose that the realizations of the pair $\{X(t), Y\}$ are generated from

$$X(t) = \sum_{i=1}^{10} i^{-1} r_i \phi_i(t) + \mu(t), \quad \mu(t) = 3\sqrt{3}\phi_1(t),$$
$$Y = \int_0^1 \rho_1(t) X(t) dt + \sigma(X)\epsilon, \quad \sigma(X) = \int_0^1 \rho_2(t) X(t) dt,$$
$$\rho_1(t) = \sum_{j=1}^{10} \tau_j \phi_j(t), \quad \rho_2(t) = \phi_1(t),$$

where $\phi_j(t) = 2^{1/2}\cos(j\pi t)$, $\tau_1 = 1$, $\tau_j = 4(-1)^{j+1}j^{-2}$ for $j \geq 2$, r_i are i.i.d. Uniform$(-\sqrt{3}, \sqrt{3})$ random variables, and ϵ is an Gamma$(1, 2)$ random variable independent of other random variables. Under this setting, the underlying slope function $\beta(t, u)$ is $\beta(t, u) = \rho_1(t) + \rho_2(t)Q_\epsilon(u) = \sum_{j=1}^{10} \tau_j\phi_j(t) + \rho_2(t)Q_\epsilon(u)$, where $Q_\epsilon(u)$ is the u-quantile of the random variable ϵ.

Scenario 3. In this scenario, we suppose that the realizations of the pair $\{X(t), Y\}$ are generated from

$$
\begin{aligned}
X(t) &= \sum_{i=1}^{10} i^{-1} r_i \phi_i(t) + \mu(t), \quad \mu(t) = 3\sqrt{3}\phi_1(t), \\
Y &= \int_0^1 \rho_1(t)X(t)dt + \sigma(X)\epsilon, \quad \sigma(X) = \int_0^1 \rho_2(t)X(t)dt, \\
\rho_1(t) &= \sum_{j=1}^{10} \tau_j\phi_j(t), \quad \rho_2(t) = \phi_1(t),
\end{aligned}
$$

where $\phi_j(t) = 2^{1/2}\cos(j\pi t)$, $\tau_1 = 1$, $\tau_j = 4(-1)^{j+1}j^{-2}$ for $j \geq 2$, r_i are i.i.d. Uniform$(-\sqrt{3}, \sqrt{3})$ random variables, and ϵ is an $Gamma(1, 2)$ random variable independent of other random variables. Under this setting, the underlying slope function $\beta(t, u)$ is $\beta(t, u) = \rho_1(t)+\rho_2(t)Q_\epsilon(u) = \sum_{j=1}^{10} \tau_j\phi_j(t)+\rho_2(t)Q_\epsilon(u)$, where $Q_\epsilon(u)$ is the u-quantile of the random variable ϵ. Note that in this scenario, the error term has the asymmetric distribution, which is different from the other scenarios.

S1.2 Summary of Simulation Results

In the simulations of Scenario 1 and 2, the set A consists of 30 quantiles uniformly distributed within the interval $[0.2, 0.8]$. In the simulations of Scenario 3, the set A consists of 17 quantiles uniformly distributed within the interval $[0.1, 0.9]$. The truncation level m for the functional covariate $\{x_i(t)\}_{i=1}^n$ by using FPCA is chosen as $m = 3$. The candidate sets for the tuning parameters $\lambda_{1,n}$ and $\lambda_{2,n}$ are chosen as $\{10^{-2}, 10^{-2.5}, 10^{-3}, 10^{-3.5}, 10^{-4}\}$ and $\{10^{-1}, 10^{-2}, 10^{-3}\}$. We consider two metrics, the mean squared error (MSE) and the maximum absolute error (MAE), to evaluate the performance of the two methods for the estimation of the slope function $\beta(t, u)$ for the ith repetition of simulation, $i = 1, \ldots, 100$,

$$\text{MSE} = \frac{1}{n_A n_T} \sum_{t \in T} \sum_{u_r \in A} \{\hat{\beta}^{(i)}(t, u_r) - \beta(t, u_r)\}^2, \tag{S1.1}$$

$$\text{MAE} = \max_{t \in T, u_r \in A} |\hat{\beta}^{(i)}(t, u_r) - \beta(t, u_r)|. \tag{S1.2}$$

MSE measures the average deviation between the estimator $\hat{\beta}(t, u)$ and the true $\beta(t, u)$, and MAE measures the maximum deviation between them. The simulation results are summarized in Table S1.

Under Scenario 1, the true slope function $\beta(t, u)$ is a univariate function of time t, and it does not change with the quantile u. We regard Scenario 1 as the simple case. Under Scenario 2, the true slope function $\beta(t, u)$ changes with both t and u, and the error ϵ is Normally distributed. Under Scenario 3, the true slope

Table S1: The average of the mean squared error (MSE) defined in (S1.1) and the average of maximum absolute error (MAE) defined in (S1.2) for the estimations of the slope function $\beta(t, u)$ by using the proposed simultaneously functional quantile regression (SFQR) method and the conventional functional quantile regression (FQR) method under three scenarios with the sample size $n = 300, 400, 500$ respectively in 100 simulation repetitions.

	n	SFQR		FQR		SFQR/FQR	
		MSE	MAE	MSE	MAE	MSE	MAE
Scenario 1	300	0.085	0.699	0.099	0.706	0.848	0.991
	400	0.068	0.630	0.080	0.648	0.849	0.97
	500	0.062	0.600	0.067	0.596	0.927	1.007
Scenario 2	300	0.321	1.383	0.505	1.548	0.635	0.894
	400	0.253	1.241	0.381	1.380	0.664	0.900
	500	0.212	1.201	0.309	1.249	0.686	0.962
Scenario 3	300	0.267	1.196	0.647	1.567	0.412	0.763
	400	0.193	1.009	0.459	1.337	0.422	0.755
	500	0.172	0.945	0.379	1.222	0.453	0.773

function $\beta(t, u)$ also changes with both t and u. But different from Scenario 2, the error ϵ has a asymmetric distribution. We regard both Scenario 2 and 3 as the complex cases. Table S1 shows that for the simple case, the proposed method and the conventional method (Kato, 2012) have similar performance on the estimation of $\beta(t, u)$. For the complex cases, the performance of the proposed method is much better than the conventional method (Kato, 2012) . For

instance, when the sample size $n = 300$, our method has reduced the average MSE by 36.5% in comparison with the conventional method (Kato, 2012). The improvement in MAE is less significant than in MSE, but our method still brings down the average MAE by 10.6%. When the sample size n increases to 500, the improvement gap of the proposed method is decreasing in comparison with the conventional method (Kato, 2012) . Nevertheless, the MSE of our estimate is still 31.6% less than the estimate obtained from the conventional method. This empirical result is supported by our intuition. For the simple case, the functional covariate $X(t)$ has the same effect on all the quantiles of the response variable Y, and therefore, incorporating multiple quantiles into the estimation procedure may not have any advantage compared with the conventional method (Kato, 2012). While for the complex cases, the effect of $X(t)$ on different quantiles of Y is different, and combining the strength of multiple quantiles should provide a better estimation.

Fig.S1, Fig.S2 and Fig.S3 present the boxplots of the mean squared error (MSE) and the maximum absolute error (MAE) for the estimation of $\beta(t, u)$ under the setting that the true slope function $\beta(t, u)$ is a univariate function of t and does not change with u for sample size $n = 300, 400$ and 500. Fig.S4, Fig.S5, Fig.S6 Fig.S7, Fig.S8, and Fig.S9 present the boxplots of the mean squared error (MSE) and the maximum absolute error (MAE) for the estimation of $\beta(t, u)$

under the settings that the true slope function $\beta(t, u)$ changes with both u and t based on sample size $n = 300, 400$ and 500.

Based on Fig.S1-Fig.S9, for MSE, the advantage of our method is significant compared with the conventional method. The proposed method in general has a smaller variance and a smaller median in comparison with the estimator derived from the conventional method. In addition, we can also see that for MSE, the advantage of the proposed method is more significant when the true $\beta(t, u)$ changes with both t and u compared with the situations when the true $\beta(t, u)$ only changes with t. But for MAE, the proposed estimator for $\beta(t, u)$ usually has a larger variation than the estimator obtained from the conventional method.

To compare the influence of truncation level m on the estimation of $\beta(t, u)$, we carry out additional simulations based on 100 repetitions under Scenario 2 and 3 with the sample size $n = 500$. The estimations are obtained with different truncation levels, $m = 3$ and $m = 4$. The MSE of the estimations are calculated based on 17 quantiles uniformly distributed on $[0.1, 0.9]$. The results are summarized in Table S2.

Table S2: The average of the mean squared error (MSE) defined in (S1.1) for the estimations of the slope function $\beta(t, u)$ by using the proposed simultaneously functional quantile regression (SFQR) method of two truncation levels $m = 3, 4$ under two scenarios with the sample size $n = 500$ in 100 simulation repetitions.

Truncation Level m	3	4
Scenario 2	0.181	0.259
Scenario 3	0.172	0.285

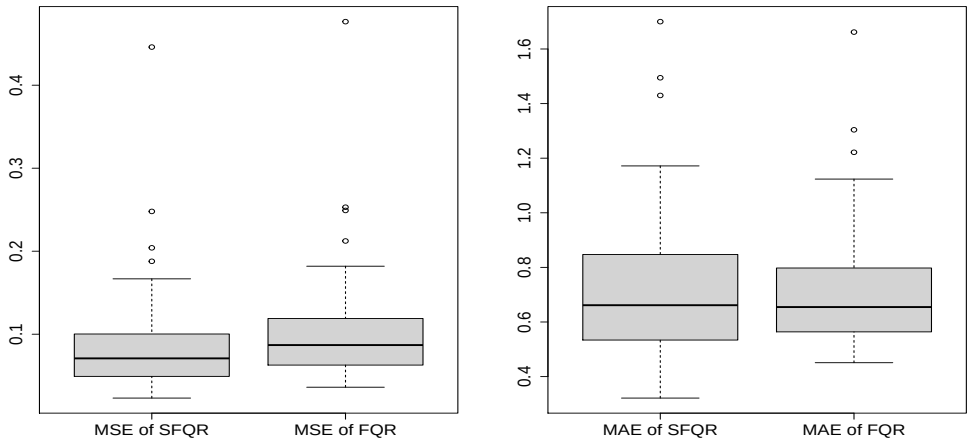

Figure S1: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 1 with sample size 300.

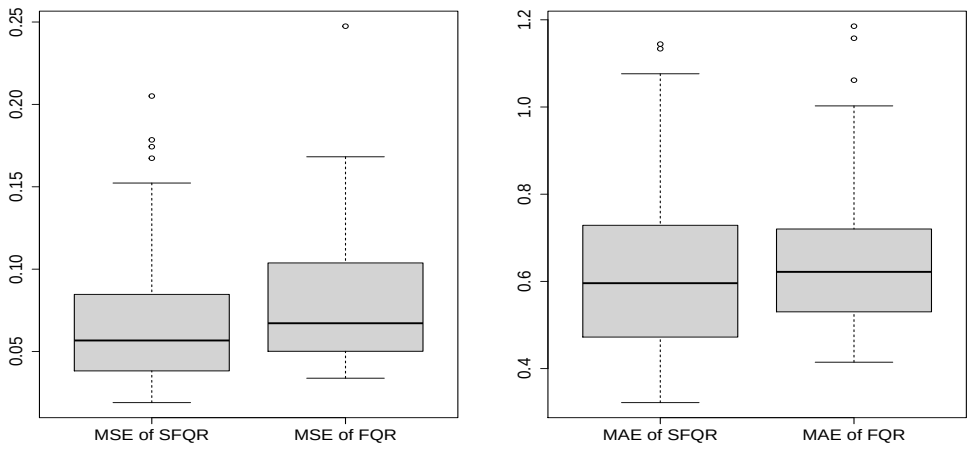

Figure S2: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 1 with sample size 400.

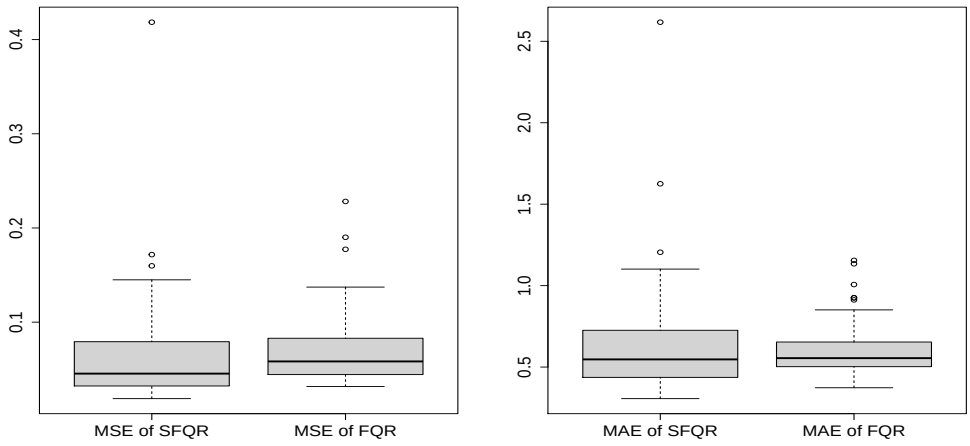

Figure S3: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 1 with sample size 500.

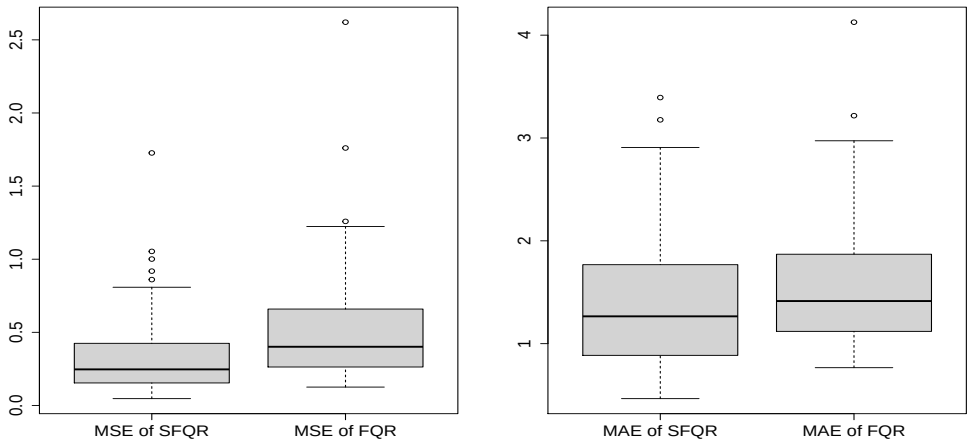

Figure S4: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 2 with sample size 300.

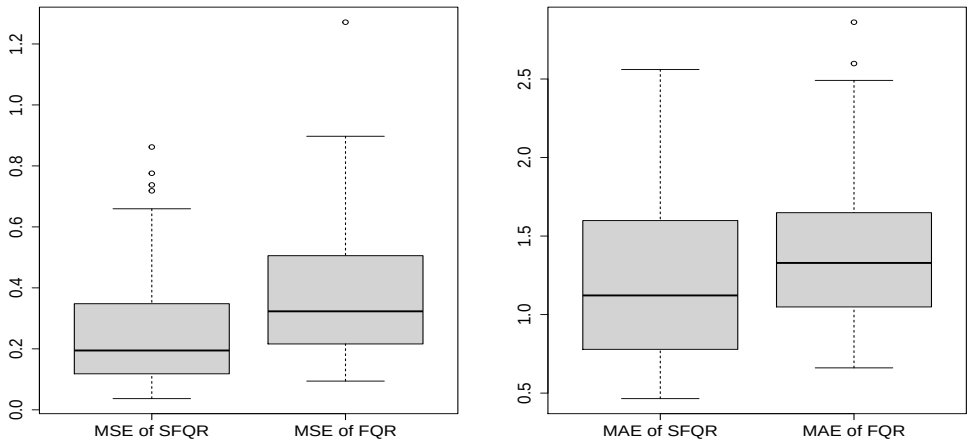

Figure S5: Boxplots of the mean squared error (MSE) defined in (S1.1) and the maximum absolute error (MAE) defined in (S1.2) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 2 with sample size 400.

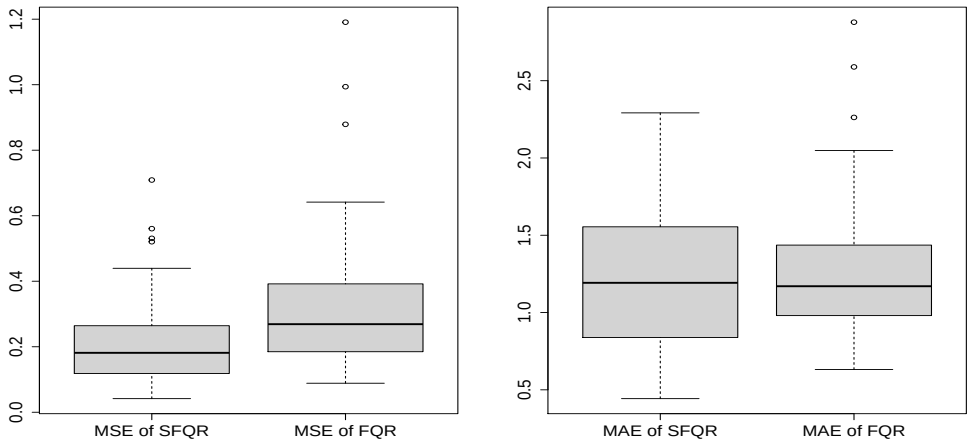

Figure S6: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 2 with sample size 500.

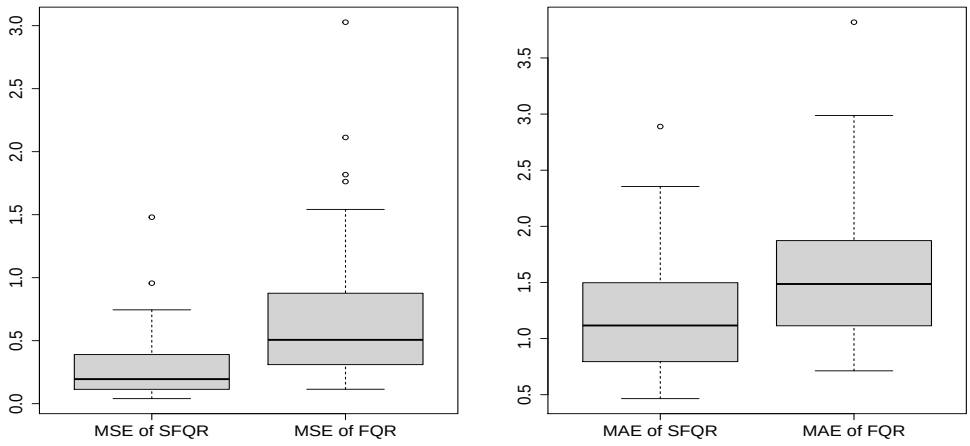

Figure S7: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 3 with sample size 300.

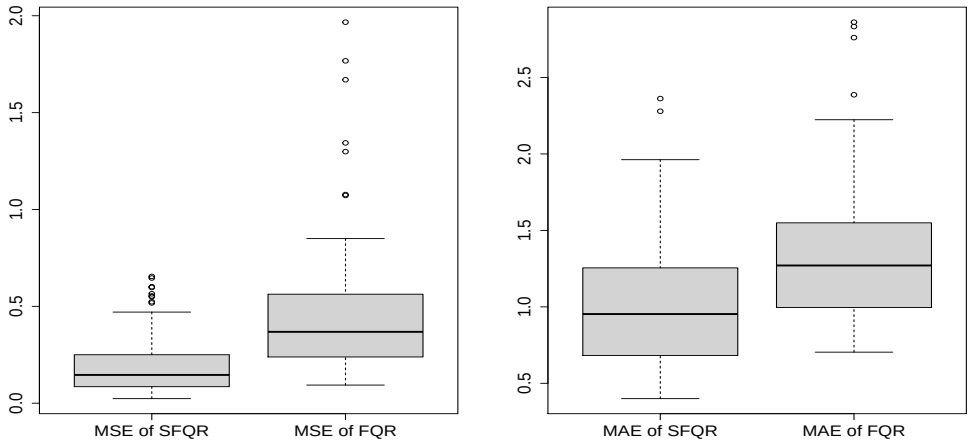

Figure S8: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 3 with sample size 400.

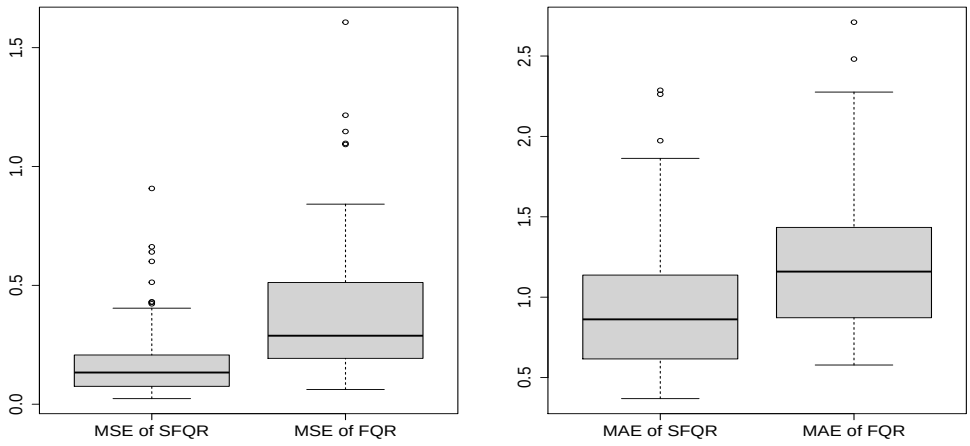

Figure S9: Boxplots of the average of mean squared error (MSE) and the average of maximum absolute error (MAE) for the estimation of $\beta(t, u)$ using the proposed simultaneous functional quantile regression (SFQR) and the conventional functional quantile regression (FQR) under Scenario 3 with sample size 500.

S2 Bernstein Polynomials over Triangulation

A triangulation Δ is a division of a surface (two-dimensional domain) $\boldsymbol{\Omega} \in \mathbb{R}^2$ into a set of triangles $\{\Lambda_1, \ldots, \Lambda_M\} \in \mathbb{R}^2$. That is, $\Delta = \{\Lambda_1, \ldots, \Lambda_M\}$ and $\boldsymbol{\Omega} = \cup_{k=1}^{M} \Lambda_k$, where the interaction of any two triangles in Δ can only be empty or one shared vertex or one shared edge (Wang et al., 2020). In the following, we assume that all the triangles $\{\Lambda_1, \ldots, \Lambda_M\}$ are not degenerated (i.e. each triangle has a nonzero area).

Let $\mathbb{P}_d$ denote the space of polynomials with degree less than or equal to d. Let $C^r(\boldsymbol{\Omega})$ be the space of functions defined on $\boldsymbol{\Omega}$ where the r-th derivatives are continuous with some positive integer r. Let $C^0(\boldsymbol{\Omega})$ be the space of continuous functions defined on $\boldsymbol{\Omega}$. Let Λ be any triangle on $\boldsymbol{\Omega}$, and for any function $s(t, u)$ with domain $\boldsymbol{\Omega}$, we define $s|_\Lambda$ as the restriction of $s(t, u)$ to the triangle Λ. Given a triangulation Δ, we define $S_d^0(\Delta) = \{s \in C^0(\boldsymbol{\Omega}) : s|_\Lambda \in \mathbb{P}_d, \Lambda \in \Delta\}$. Given $0 \leq r \leq d$ and a triangulation Δ, we define $S_d^r(\Delta) = S_d^0(\Delta) \cap C^r(\Delta)$.

Because Bernstein polynomials is a basis for $\mathbb{P}_d$ (Lai and Schumaker, 2007), Bernstein polynomials can be used to represent the bivariate splines, which are defined as follows. For any point $\boldsymbol{z} \in \mathbb{R}^2$ and a given triangle $\Lambda \in \mathbb{R}^2$, there exists a unique barycentric coordinate for the point $\boldsymbol{z}$ relative to the triangle Λ, denoted as (a_1, a_2, a_3). More specifically, denote the vertices of Λ by $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Then a_1, a_2 and a_3 are all functions of $\boldsymbol{v}$ such that $a_1 + a_2 + a_3 = 1$,

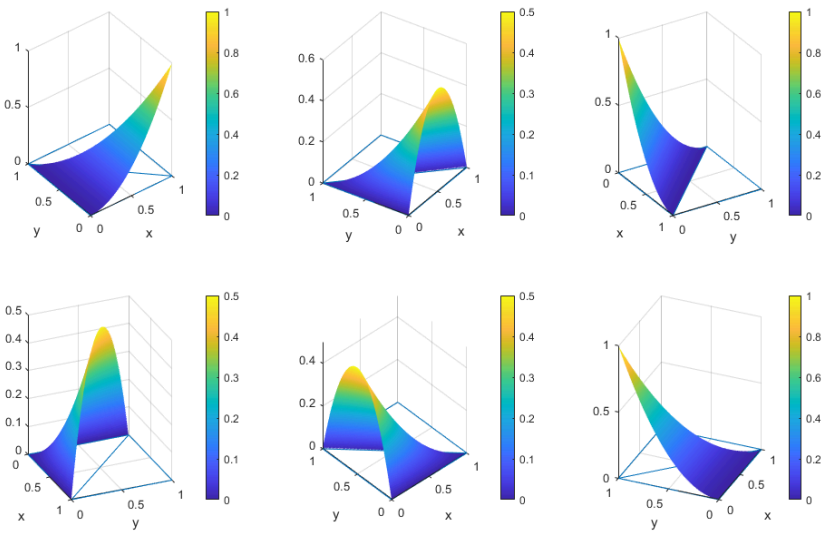

Figure S10: Examples of six Bernstein polynomials of degree $d = 2$ defined in (S2.1) on a single triangle of the triangulation over $[0, 1] \times [0, 1]$.

and $\boldsymbol{z} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3$. With the barycentric coordinate, for any triangle $\Lambda \in \Delta$, Bernstein polynomials of degree d are defined as

$$B^d_{i,j,k} = \frac{d!}{i!j!k!} a_1^i a_2^j a_3^k, \tag{S2.1}$$

for any $i+j+k = d$. There are $(d+2)(d+1)/2$ Bernstein polynomials associated with every triangle $\Lambda \in \Delta$. Figure S10 shows the examples of six Bernstein polynomials defined on a single triangle of the triangulation over $[0, 1] \times [0, 1]$

when $d = 2$. This triangulation consists of only two triangles.

S3 Functional Principal Component Analysis (FPCA)

Assume that we observe independent and identically distributed data pairs $\{y_i, x_i(t)\}_{i=1}^n$ as the realizations of $\{Y, X(t)\}$. Now we start to describe the standard FPCA (Ramsay and Silverman, 2005; Ramsay et al., 2009). Let $G(s,t) = \text{cov}\,(X(s), X(t))$ be the covariance function of $X(t)$. We have the spectral decomposition on the covariance function:

$$G(s,t) = \sum_{k=1}^{\infty} \kappa_k \phi_k(s)\phi_k(t),$$

where $\kappa_1 \geq \kappa_2 \geq \cdots \geq 0$. Suppose the realizations $\{x_i(t)\}_{i=1}^n$ are all observed on the same set $T \subset \mathscr{T}$. We first estimate the mean curve $\mu(t)$ of $X(t)$ evaluated at each $t \in T$ by

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t).$$

Given $\hat{\mu}(t)$, we estimate $G(s,t)$ by

$$\hat{G}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \{x_i(s) - \hat{\mu}(s)\}\{x_i(t) - \hat{\mu}(t)\},$$

for any $s, t \in T$. Similarly, we have the spectral decomposition of $\hat{G}(s,t)$ as

$$\hat{G}(s,t) = \sum_{k=1}^{\infty} \hat{\kappa}_k \hat{\phi}_k(s)\hat{\phi}_k(t),$$

where $\{\hat{\phi}_k(t)\}_{k=1}^{\infty}$ are estimators of the eigenfunctions, and $\hat{\kappa}_1 \geq \hat{\kappa}_2 \geq \ldots \geq 0$ are the estimator of corresponding eigenvalues.

Given $\hat{\mu}(t)$ and $\{\hat{\phi}_k(t)\}_{k=1}^m$, for a truncation level m, we estimate the corresponding functional principal component scores $\{\hat{\boldsymbol{\xi}}_i = (\xi_{i,1}, \ldots, \xi_{i,m})^{\mathrm{T}}\}_{i=1}^n$ by

$$\hat{\xi}_{i,k} = \int \{x_i(t) - \hat{\mu}(t)\} \, \hat{\phi}_k(t) dt, \tag{S3.1}$$

for $k = 1, \ldots, m$.

S4 Proofs for the Theorems

We use $\|\cdot\|_{l^2}$ to denote the l^2-norm of a vector. We assume that A1-A5 given in the main paper are satisfied.

Lemma 1. *De Boor et al. (1978) Let l and d be integers with $0 \leq l \leq d$. For a given set of knots T on $[a, b]$. a B-spline basis of degree d is defined based on the knots T, denoted as a vector $\boldsymbol{b}_0^{\mathrm{T}}(x)$. Then for any function $g(x) \in \boldsymbol{W}_q^{l+1}([a, b])$ and for any $1 \leq q \leq \infty$, any $0 \leq r \leq l$, there exists a coefficient vector $\boldsymbol{\gamma}_0$ such that*

$$\|D^r(\boldsymbol{b}_0^{\mathrm{T}}(x)\boldsymbol{\gamma}_0 - g(x))\|_{L^q([a,b])} \leq Ch^{l+1-r} \left\|D^{l+1}g\right\|_{L^q([a,b])}.$$

where C is a constant and h is the largest length of a knot interval in T.

Lemma 2. *Lai and Schumaker (2007) Let $\{B(t, u)_j\}_{j=1}^{J}$ be the Bernstein polynomials basis defined on the spline space $S_d^r(\Delta)$ over a π-quasi-uniform triangulation Δ. Then there exist positive constants c_1 and c_2, which only depend on d, r and shape parameter π, such that*

$$c_1|\Delta|^2 \|\boldsymbol{\gamma}\|_{l^2}^2 \leq \|\boldsymbol{B}(t, u)^{\mathrm{T}}\boldsymbol{\gamma}\|_2^2 \leq c_2|\Delta|^2 \|\boldsymbol{\gamma}\|_{l^2}^2.$$

Given any domain $\boldsymbol{\Omega} \in \mathbb{R}^2$, for any $1 \leq q \leq \infty$, we define the $q-$norm of

any function $f \in \mathbf{\Omega}$ by

$$\|f\|_{q,\mathbf{\Omega}} = \begin{cases} \left(\int_{\mathbf{\Omega}} |f(u)|^q du\right)^{1/q}, & 1 \le q < \infty \\ \text{ess sup}_{u\in\mathbf{\Omega}}|f(u)|, & q = \infty, \end{cases}$$

Lemma 3. *Lai and Schumaker (2007) Suppose Δ is a quasi-uniform triangulation over a polygonal domain $\mathbf{\Omega}$, and $f(x,y) \in W_q^{d+1}(\mathbf{\Omega})$, where $W_q^{d+1}(\mathbf{\Omega})$ is a Sobolev space defined on $\mathbf{\Omega}$. Suppose $d \ge 3r+2$. Then, for any integers a_1, a_2, $0 \le a_1 + a_2 \le d$, there exists $f^*(x,y) \in S_d^r(\Delta)$ such that*

$$\left\|\nabla_x^{a_1}\nabla_y^{a_2}(f^* - f)\right\|_{q,\mathbf{\Omega}} \le C|\Delta|^{d+1-a_1-a_2}|g|_{d+1,q,\mathbf{\Omega}},$$

for some constant C, which only depends on d, r and π, the shape parameter of the triangulation Δ.

Define $L_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) = \frac{1}{n_A}\sum_{r=1}^{n_A} E_n\left\{\rho_{u_r}\left(v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta})\right)\right\} + \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta}$, where $v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) = Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta}$. Suppose $(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}})$ is a solution of

$$\min_{\boldsymbol{\gamma}_0, \boldsymbol{\theta}} L_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}), \tag{S4.1}$$

and define $\hat{\boldsymbol{\gamma}} = \boldsymbol{Q}\hat{\boldsymbol{\theta}}$ and $\hat{\beta}(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}$.

For convenience, we first define several notations. Let $\beta_k(u) = \int \beta(t,u)\phi_k(t)dt$ for $k = 1, \ldots, m$. For $i = 1, \ldots, n$ and $k = 1, \ldots, m$, let $\eta_{i,k} = \kappa_k^{-1/2}\xi_{i,k}$, $\hat{\eta}_{i,k} = \kappa_k^{-1/2}\hat{\xi}_{i,k}$, $d_k(u) = \kappa_k^{1/2}\beta_k(u)$, $\hat{d}_k(u) = \kappa_k^{1/2}\sum_{j=1}^{J}\hat{\gamma}_j\hat{p}_{k,j}(u)$. where $\hat{\gamma}_j$ is the jth entry of $\hat{\boldsymbol{\gamma}}$. For $k = 0$, let $\hat{\eta}_{i,0} = \eta_{i,0} = 1$ and $d_0(u) = c_0(u)$ as well as

$\hat{d}_0(u) = \boldsymbol{b}_0^{\mathrm{T}}(u)\hat{\boldsymbol{\gamma}}_0$.

Then for any $i = 1, \ldots, n$, define

$$\boldsymbol{d}^m(u) = (d_1(u), \ldots, d_m(u))^{\mathrm{T}},$$

$$\boldsymbol{d}^{m+1}(u) = (d_0(u), d_1(u), \ldots, d_m(u))^{\mathrm{T}},$$

$$\hat{\boldsymbol{d}}^m(u) = (\hat{d}_1(u), \ldots, \hat{d}_m(u))^{\mathrm{T}},$$

$$\hat{\boldsymbol{d}}^{m+1}(u) = (\hat{d}_0(u), \hat{d}_1(u), \ldots, \hat{d}_m(u))^{\mathrm{T}},$$

$$\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,m})^{\mathrm{T}},$$

$$\hat{\boldsymbol{\eta}}_i^m = (\hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,m})^{\mathrm{T}},$$

$$\hat{\boldsymbol{\eta}}_i^{m+1} = (\hat{\eta}_{i,0}, \hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,m})^{\mathrm{T}} = (1, \hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,m})^{\mathrm{T}}.$$

For a sequence $\{a_i\}_{i=1}^n$, define $E_n\{a_i\} = \frac{1}{n}\sum_{i=1}^n a_i$.

Lemma 4. *(Kato, 2012) Let* $\mathbb{S}^m = \{\boldsymbol{h} \in \mathbb{R}^{m+1} | \, \|\boldsymbol{h}\|_{l^2} = 1\}$. *Then,*

$$-E_n[\{u - \mathbb{1}(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot (\boldsymbol{d}^{m+1}(u) + M\sqrt{m/n}\boldsymbol{h}))\}(\hat{\boldsymbol{\eta}}_i^{m+1} \cdot \boldsymbol{h})]$$
$$\ge c_1 M(1 - o_p(1))\sqrt{m/n} - O_p(\sqrt{m/n}) - M^2 o_p(\sqrt{m/n}),$$

where c_1 *and* M *are some positive constants, as well as the stochastic orders are evaluated uniformly for any* $u \in (0, 1)$ *and any* $\boldsymbol{h} \in \mathbb{S}^m$

Lemma 5. *Let* $(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}})$ *be a solution of* (S4.1), $\lambda_{1,n} \asymp n_A^{-1} n_T^{-1} m^{-1/2} n |\Delta|^{d+1}$,

and $\lambda_{2,n} = o(\lambda_{1,n} n_A n_T |\Delta|^4)$. *The estimator* $\hat{\beta}(t,u)$ *can be expressed as*

$$\begin{aligned}\hat{\beta}(t,u) &= \sum_{k=1}^{m} \kappa_k^{-1/2} \hat{d}_k(u) \hat{\phi}_k(t) + \sum_{k=m+1}^{\infty} \kappa_k^{-1/2} \hat{d}_k(u) \hat{\phi}_j(t) \\ &= \sum_{k=1}^{m} \kappa_k^{-1/2} \hat{d}_k(u) \hat{\phi}_k(t) + \hat{\beta}_{re}(t,u),\end{aligned}$$

where $\kappa_k^{-1/2} \hat{d}_k(u) = \int \hat{\beta}(t,u) \hat{\phi}_k(t) dt$ *and* $\hat{\beta}_{re}(t,u) = \sum_{k=m+1}^{\infty} \kappa_k^{-1/2} \hat{d}_k(u) \hat{\phi}_j(t)$.

Then, we have

$$\left\| \boldsymbol{B}_{A,T}^{\mathrm{T}} \boldsymbol{Q} \hat{\boldsymbol{\theta}} \right\|_{l^2} \leq \lambda_{1,n}^{-1} (1 - o(1)) O_p(m n^{-\frac{1}{2}} \vee m |\Delta|^{d+1}),$$

and

$$\hat{\beta}_{re}(t,u) \approx O_p(m^{1/2} n^{-1/2}).$$

Proof: Under the condition A5 given in the main paper, and by Lemma 1 and Lemma 3, there exist $\boldsymbol{\gamma}_0^*$ and $\boldsymbol{\theta}^*$ such that

$$\sup_{(t,u) \in \mathscr{T} \times \mathscr{A}} |\beta(t,u) - \boldsymbol{B}^{\mathrm{T}}(t,u) \boldsymbol{Q} \boldsymbol{\theta}^*| \leq C_1 |\Delta|^{d+1},$$

$$\sup_{u \in \mathscr{A}} |c(u) - \boldsymbol{b}_0^{\mathrm{T}}(u) \boldsymbol{\gamma}_0^*| \leq C_2 |\Delta|^{d+1},$$

for some constant C_1 and C_2. Since $(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}})$ is the minimizer of (S4.1), then

$$L_n(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}) \leq L_n(\hat{\boldsymbol{\gamma}}_0, \boldsymbol{\theta}^*). \tag{S4.2}$$

For any $u_r \in A$, let

$$(\tilde{d}_0(u_r), \tilde{\boldsymbol{d}}^m(u_r)) = \text{argmin}_{f_0, f_1, \ldots, f_m} E_n \left\{ \rho_{u_r} \left(Y_i - f_0 - \hat{\boldsymbol{\eta}}_i^m \cdot \boldsymbol{f}^m \right) \right\},$$

where $\boldsymbol{f}^m = (f_1, \ldots, f_m)^{\mathrm{T}}$. From (Kato, 2012), we know that

$$\sup_{u \in \mathscr{A}} \left\{ (d_0(u) - c_0(u))^2 + \left\| \tilde{\boldsymbol{d}}^m(u) - \boldsymbol{d}^m(u) \right\|_{l^2}^2 \right\} = O_p(mn^{-1}).$$

Therefore,

$$\begin{aligned}
&\left| n_A^{-1} \sum_{r=1}^{n_A} E_n \left\{ \rho_{u_r} \left(Y_i - \tilde{d}_0(u_r) - \hat{\boldsymbol{\eta}}_i^m \cdot \tilde{\boldsymbol{d}}^m \right) \right\} - n_A^{-1} \sum_{r=1}^{n_A} E_n \left\{ \rho_{u_r} \left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r) \boldsymbol{Q} \boldsymbol{\theta}^*) \right) \right\} \right| \\
&\quad \leq \{n_A n\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \left\{ \left| \tilde{d}_0(u_r) + \hat{\boldsymbol{\eta}}_i^m \cdot \tilde{\boldsymbol{d}}^m(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r) Q \boldsymbol{\theta}^* \right| \right. \\
&\qquad \left. + |2u_r - 1| \left| \tilde{d}_0(u_r) + \hat{\boldsymbol{\eta}}_i^m \cdot \tilde{\boldsymbol{d}}^m(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r) Q \boldsymbol{\theta}^* \right| \right\}, \\
&\quad \leq 2 n_A^{-1} n^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \left\{ \left| \hat{\boldsymbol{\eta}}_i^m \cdot \tilde{\boldsymbol{d}}^m(u_r) - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r) Q \boldsymbol{\theta}^* \right| + \left| \tilde{d}_0(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \boldsymbol{\gamma}_0^* \right| \right\}, \\
&\quad \leq 2 n_A^{-1} n^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \left\{ \|\hat{\boldsymbol{\eta}}_i^m\|_{l^2} \left\| \tilde{\boldsymbol{d}}^m(u_r) - \boldsymbol{\kappa} \hat{\boldsymbol{P}}(u_r) Q \boldsymbol{\theta}^* \right\|_{l^2} + \left| \tilde{d}_0(u_r) - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \boldsymbol{\gamma}_0^* \right| \right\}, \\
&\quad \leq O_p(m^{1/2}) \times O_p(m^{1/2} n^{-1/2} \vee |\Delta|^{d+1}) = O_p(mn^{-1/2} \vee m^{1/2} |\Delta|^{d+1}),
\end{aligned} \tag{S4.3}$$

where $\boldsymbol{\kappa}$ is a m by m diagonal matrix with the diagonal entries being $(\kappa_1^{1/2}, \ldots, \kappa_m^{1/2})$.

In addition, by definition of $(\tilde{d}_0(u_r), \tilde{\boldsymbol{d}}^m(u_r))$,

$$\sum_{r=1}^{n_A} E_n \left\{ \rho_{u_r} \left(Y_i - \tilde{d}_0(u_r) - \hat{\boldsymbol{\eta}}_i^m \cdot \tilde{\boldsymbol{d}}^m(u_r) \right) \right\} \leq \sum_{r=1}^{n_A} E_n \left\{ \rho_{u_r} \left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r) \hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r) \boldsymbol{Q} \hat{\boldsymbol{\theta}}) \right) \right\}. \tag{S4.4}$$

Since

$$n_A^{-1}\sum_{r=1}^{n_A} E_n\left\{\rho_{u_r}\left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\right)\right\}$$
$$< n_A^{-1}\sum_{r=1}^{n_A} E_n\left\{\rho_{u_r}\left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta}^*)\right)\right\},$$

then by (S4.3) and (S4.4),

$$\Bigg| n_A^{-1}\sum_{r=1}^{n_A} E_n\left\{\rho_{u_r}\left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\right)\right\}$$
$$- n_A^{-1}\sum_{r=1}^{n_A} E_n\left\{\rho_{u_r}\left(Y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0^* - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta}^*)\right)\right\}\Bigg| \leq O_p(mn^{-1/2} \vee m^{1/2}|\Delta|^{d+1}). \tag{S4.5}$$

By Lemma 2 and Wang et al. (2020),

$$\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{\theta}} \leq C|\Delta|^{-2}\left\|\hat{\boldsymbol{\theta}}\right\|_{l^2}^2 = C|\Delta|^{-4}\left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2.$$

On the other hand,

$$\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}} \approx n_A n_T\left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2.$$

Since $\lambda_{2,n} = o(\lambda_{1,n}n_A n_T|\Delta|^4)$, when n_A and n_T are large enough,

$$\frac{\lambda_{1,n}^{-1}\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}} = o(1). \tag{S4.6}$$

Combine (S4.2), (S4.5) and (S4.6), we have

$$\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}} \leq O_p(\lambda_{1,n}^{-1}mn^{-1/2} \vee \lambda_{1,n}^{-1}m^{1/2}|\Delta|^{d+1}) + \boldsymbol{\theta}^{*\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta}^*,$$
$$= O_p(\lambda_{1,n}^{-1}mn^{-1/2} \vee \lambda_{1,n}^{-1}m^{1/2}|\Delta|^{d+1}) + O(n_A n_T),$$

which implies $\left\|\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}\right\|_{l^2} = O_p(\lambda_{1,n}^{-1/2}m^{1/2}n^{-1/4} \vee \lambda_{1,n}^{-1/2}m^{1/4}|\Delta|^{(d+1)/2} \vee n_A^{1/2}n_T^{1/2})$.

By Lemma 3, there exists a vector $\hat{\boldsymbol{\theta}}_m$ such that $\max_{t\in T,u\in A}|\hat{\beta}_m^*(t,u) - \sum_{k=1}^m\left(\int\hat{\beta}(t,u)\hat{\phi}_k(t)dt\right)\hat{\phi}_k(t)| \leq C|\Delta|^{d+1}$, where $\hat{\beta}_m^*(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}_m$. Since $(\hat{\boldsymbol{\gamma}}_0,\hat{\boldsymbol{\theta}})$ is a minimizer of (S4.1), then $L_n(\hat{\boldsymbol{\gamma}}_0,\hat{\boldsymbol{\theta}}) \leq L_n(\hat{\boldsymbol{\gamma}}_0,\hat{\boldsymbol{\theta}}_m)$. That is,

$$
\begin{aligned}
&n^{-1}n_A^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}\right) + \lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}} \\
&+\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{\theta}} \leq n^{-1}n_A^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right) \\
&+\lambda_{1,n}\hat{\boldsymbol{\theta}}_m^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}_m + \lambda_{2,n}\hat{\boldsymbol{\theta}}_m^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{\theta}}_m. \qquad \text{(S4.7)}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&n^{-1}n_A^{-1}\Bigg|\sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}\right) \\
&\qquad - \sum_{r=1}^{n_A}\sum_{i=1}^{n}\rho_{u_r}\left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right)\Bigg|, \\
&\qquad = n^{-1}n_A^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\Bigg\{\left|\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right| \\
&\qquad + (2u_r-1)\left(\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right)\Bigg\}, \\
&\qquad \leq n^{-1}n_A^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\left\{2\left|\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right|\right\}, \\
&\qquad \leq n^{-1}n_A^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\left\{2\left\|\hat{\boldsymbol{\xi}}_i\right\|_{l^2}\left\|\hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{P}}(u_r)Q\hat{\boldsymbol{\theta}}_m\right\|_{l^2}\right\}, \\
&\qquad = O_p(m^{1/2}|\Delta|^{d+1}), \qquad \text{(S4.8)}
\end{aligned}
$$

provided that $E_n\left\{\left\|\hat{\boldsymbol{\xi}}_i\right\|_{l^2}\right\} = O_p(m^{1/2})$ (Kato, 2012). By Lemma 2, (S4.7) and (S4.8), we have

$$
\begin{aligned}
& n_A^{-1}n_T^{-1}\left\{\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_m^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}_m\right\} \\
& \leq O_p(\lambda_{1,n}^{-1}n_A^{-1}n_T^{-1}m^{1/2}|\Delta|^{d+1}) + O_p(\lambda_{1,n}^{-1}\lambda_{2,n}n_A^{-1}n_T^{-1}|\Delta|^{-2}\left\|\hat{\boldsymbol{\theta}}_m\right\|_{l^2}^2), \\
& = O_p(\lambda_{1,n}^{-1}n_A^{-1}n_T^{-1}m^{1/2}|\Delta|^{d+1} \vee \lambda_{1,n}^{-1}\lambda_{2,n}n_A^{-1}n_T^{-1}|\Delta|^{-4}) = O_p(m/n). \quad \text{(S4.9)}
\end{aligned}
$$

Then, by (S4.9)

$$
\begin{aligned}
& \left\|\hat{\beta}(t,u) - \sum_{k=1}^{m}\left(\int \hat{\beta}(t,u)\hat{\phi}_k(t)dt\right)\hat{\phi}_k(t)\right\|_{L^2(\boldsymbol{\Omega})}^2, \\
= & \left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2 - \left\|\sum_{k=1}^{m}\left(\int \hat{\beta}(t,u)\hat{\phi}_k(t)dt\right)\hat{\phi}_k(t)\right\|_{L^2(\boldsymbol{\Omega})}^2 \\
= & \left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2 - \left\|\sum_{k=1}^{m}\kappa_k^{-1/2}\hat{d}_k(u)\hat{\phi}_k(t) - \hat{\beta}_m^*(t,u) + \hat{\beta}_m^*(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2, \\
\leq & \left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2 - \left\|\hat{\beta}_m^*(t,u)\right\|_{L^2(\boldsymbol{\Omega})}^2 + O(|\Delta|^{d+1}), \\
\approx & n_A^{-1}n_T^{-1}\left\{\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_m^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}_m\right\} + O(|\Delta|^{d+1}), \\
= & O_p(m/n).
\end{aligned}
$$

Therefore, we conclude that $\hat{\beta}_{re}(t,u) \approx O_p(m^{1/2}n^{-1/2})$.

Lemma 6. *Let* $(\hat{\gamma}_0, \hat{\boldsymbol{\theta}})$ *be a solution of* (S4.1). *Then,*

$$\left\| n_A^{-1} n^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \left\{ u_r - \mathbb{1}(Y_i < \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 + \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}}) \right\} \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r)\boldsymbol{Q} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right) \right\|_{l^2}$$
$$\leq \frac{c\{m + (n_A \wedge C|\Delta|^{-1})\}}{n_A n} \max_{i \in 1,\ldots,n} \left\| \hat{\boldsymbol{\xi}}_i \right\|_{l^2} \max_{u_r \in A} \sigma_{\max}\{\hat{\boldsymbol{P}}(u_r)\} \left\| \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\|_{l^2}$$
$$+ \left\| \left(2\lambda_{1,n} \hat{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{B}_{A,T} \boldsymbol{B}_{A,T}^{\mathrm{T}} \boldsymbol{Q} + 2\lambda_{2,n} \hat{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{D} \boldsymbol{Q} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right) \right\|_{l^2}, \tag{S4.10}$$

where $\sigma_{\max}(x)$ *is the largest singular value of a matrix* x.

Proof: The check function $\rho_u(x)$ is

$$\rho_u(x) = |x| + (2u - 1)x,$$

then we can rewrite (S4.1) as

$$\min_{\boldsymbol{\gamma}_0, \boldsymbol{\theta}} \frac{1}{n_A} \sum_{r=1}^{n_A} E_n\{|v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta})| + (2u_r - 1)v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta})\} + \lambda_{1,n} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{B}_{A,T} \boldsymbol{B}_{A,T}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{\theta}$$
$$+ \lambda_{2,n} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{D} \boldsymbol{Q} \boldsymbol{\theta}. \tag{S4.11}$$

To further remove the absolute value sign, the minimization problem (S4.11) is equivalent to the following constrained optimization problem

$$\min_{\boldsymbol{\varphi}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}} g(\boldsymbol{\varphi}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}) = \frac{1}{n_A} \sum_{r=1}^{n_A} E_n\{\varphi_{i,r} + (2u_r - 1)v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta})\} + \lambda_{1,n} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{B}_{A,T} \boldsymbol{B}_{A,T}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{\theta}$$
$$+ \lambda_{2,n} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{D} \boldsymbol{Q} \boldsymbol{\theta} \tag{S4.12}$$

such that

$$-\varphi_{i,r} + v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) \leq 0,$$

$$-\varphi_{i,r} - v_{i,r}(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) \leq 0.$$

Suppose ($\hat{\boldsymbol{\varphi}}$, $\hat{\boldsymbol{\gamma}}_0$,$\hat{\boldsymbol{\theta}}$) is a solution of (S4.12). Define three sets

$$I_1 = \{(i,r) : v_{i,r}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}) > 0\},$$

$$I_2 = \{(i,r) : v_{i,r}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}) < 0\},$$

$$I_3 = \{(i,r) : v_{i,r}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}) = 0\}.$$

The gradient of objective function g is

$$\nabla g = \frac{1}{n}\begin{bmatrix} \mathbf{1} \\ \frac{1}{n_A}\sum_{r=1}^{n_A}\sum_{i=1}^{n}(2u_r - 1)\nabla v_{i,r} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ 2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q} + 2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} \end{bmatrix}$$

The active constraints are given by

$$\begin{bmatrix} -\boldsymbol{e}_{i,r} \\ \nabla v_{i,r} \end{bmatrix} \quad \forall (i,r) \in I_1,$$

$$\begin{bmatrix} -\boldsymbol{e}_{i,r} \\ -\nabla v_{i,r} \end{bmatrix} \quad \forall (i,r) \in I_2,$$

and

$$\begin{bmatrix} -\boldsymbol{e}_{i,r} \\ \nabla v_{i,r} \end{bmatrix}, \quad \begin{bmatrix} -\boldsymbol{e}_{i,r} \\ -\nabla v_{i,r} \end{bmatrix} \quad \forall (i,r) \in I_3,$$

where $\boldsymbol{e}_i$ is a $n_A n$-vector with a one in the position corresponding to the pair (i,r) and zeros elsewhere.

By checking Kuhn-Tucker condition, we have the following necessary con-

dition for the optimality (El-Attar et al., 1979),

$$\nabla g + \frac{1}{n}\sum_{(i,r)\in I_1}\delta_{i,r}\begin{bmatrix}-\boldsymbol{e}_{i,r}\\ \nabla v_{i,r}\end{bmatrix} + \frac{1}{n}\sum_{(i,r)\in I_2}\delta_{i,r}\begin{bmatrix}-\boldsymbol{e}_{i,r}\\ -\nabla v_{i,r}\end{bmatrix} + \frac{1}{n}\sum_{(i,r)\in I_3}\left\{\delta_{i,r}\begin{bmatrix}-\boldsymbol{e}_{i,r}\\ \nabla v_{i,r}\end{bmatrix} + \nu_{i,r}\begin{bmatrix}-\boldsymbol{e}_{i,r}\\ -\nabla v_{i,r}\end{bmatrix}\right\} = \begin{bmatrix}0\\ 0\end{bmatrix}, \tag{S4.13}$$

where $\delta_{i,r} \geq 0$ and $\nu_{i,r} \geq 0$. Since the sets I_1, I_2 and I_3 are disjoint, then the top nn_A equations of (S4.13) imply that

$$\delta_{i,r} = 1 \quad \forall (i,r) \in I_1, \tag{S4.14}$$

$$\delta_{i,r} = 1 \quad \forall (i,r) \in I_2, \tag{S4.15}$$

$$\delta_{i,r} + \nu_{i,r} = 1 \quad \forall (i,r) \in I_3. \tag{S4.16}$$

Now we consider the rest of equations in (S4.13). By (S4.14), (S4.15) and (S4.16), we can derive that

$$\frac{1}{n_A n}\sum_{(i,r)\in I_1\cup I_2}\left\{\text{sign}(v_{i,r}) + 2u_r - 1\right\}\nabla v_{i,r} + \frac{1}{n_A n}\sum_{(i,r)\in I_3}\left\{\delta_{i,r} - \nu_{i,r} + 2u_r - 1\right\}\nabla v_{i,r} + \begin{bmatrix}\mathbf{0}\\ 2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q} + 2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\end{bmatrix} = \mathbf{0}, \tag{S4.17}$$

where $\nabla v_{i,r}$ is defined as

$$\nabla v_{i,r} = \begin{bmatrix}-\boldsymbol{b}_0^{\mathrm{T}}(u_r)\\ -\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\end{bmatrix}.$$

The equations (S4.17) implies that

$$\frac{1}{n_A n}\sum_{(i,r)\in I_1\cup I_2\cup I_3}\{u_r-\mathbb{1}(Y_i<\boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0+\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\}\boldsymbol{b}_0^{\mathrm{T}}(u_r)$$
$$=\frac{1}{n_A n}\sum_{(i,r)\in I_3}\{1/2-(\delta_{i,r}-\nu_{i,r})/2-\mathbb{1}(Y_i<\boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0+\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\}\boldsymbol{b}_0^{\mathrm{T}}(u_r), \tag{S4.18}$$

and

$$\frac{1}{n_A n}\sum_{(i,r)\in I_1\cup I_2\cup I_3}\{1-u_r-\mathbb{1}(Y_i<\boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0+\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\}\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}$$
$$=\frac{1}{n_A n}\sum_{(i,r)\in I_3}\{1/2-(\delta_{i,r}-\nu_{i,r})/2-\mathbb{1}(Y_i<\boldsymbol{b}_0^{\mathrm{T}}(u_r)\hat{\boldsymbol{\gamma}}_0+\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\hat{\boldsymbol{\theta}})\}\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}$$
$$-2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}-2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}. \tag{S4.19}$$

Let n_B denote the number of B-spline basis functions, and n_{knots} denote the number of its interior knots. The relationship bewteen n_B, n_{knots} and d is

$$n_B=d+n_{knots}.$$

Since we only require the chosen B-spline basis has the same approximation power as the bivariate splines used for the approximation of $\beta(t,u)$, we can choose the interior knots uniformly distributed within the range of u with $n_{knots}=C|\Delta|^{-1}$ for some constant C.

Define $\boldsymbol{Z}_i(u)=\left[\boldsymbol{b}_0^{\mathrm{T}}(u),\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u)\boldsymbol{Q}\right]$, and $\hat{\boldsymbol{\Gamma}}=\begin{bmatrix}\hat{\boldsymbol{\gamma}}_0\\ \hat{\boldsymbol{\theta}}\end{bmatrix}$. Recall that I_3 is

defined as

$$I_3 = \{(i,r) : v_{i,r}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}) = 0\}.$$

Let

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1(u_1) \\ \boldsymbol{Z}_2(u_1) \\ \cdots \\ \boldsymbol{Z}_n(u_1) \\ \boldsymbol{Z}_1(u_2) \\ \cdots \\ \boldsymbol{Z}_n(u_r) \\ \cdots \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_0^{\mathrm{T}}(u_1), \hat{\boldsymbol{\xi}}_1^{\mathrm{T}} \hat{\boldsymbol{P}}(u_1)\boldsymbol{Q} \\ \boldsymbol{b}_0^{\mathrm{T}}(u_1), \hat{\boldsymbol{\xi}}_2^{\mathrm{T}} \hat{\boldsymbol{P}}(u_1)\boldsymbol{Q} \\ \cdots, \cdots \\ \boldsymbol{b}_0^{\mathrm{T}}(u_2), \hat{\boldsymbol{\xi}}_1^{\mathrm{T}} \hat{\boldsymbol{P}}(u_2)\boldsymbol{Q} \\ \cdots, \cdots \\ \boldsymbol{b}_0^{\mathrm{T}}(u_r), \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r)\boldsymbol{Q} \\ \cdots, \cdots \end{bmatrix} = [\boldsymbol{M}_1, \boldsymbol{M}_2],$$

where $\boldsymbol{M}_1 = \begin{bmatrix} \boldsymbol{b}_0^{\mathrm{T}}(u_1) \\ \boldsymbol{b}_0^{\mathrm{T}}(u_1) \\ \cdots \\ \boldsymbol{b}_0^{\mathrm{T}}(u_2) \\ \cdots \\ \boldsymbol{b}_0^{\mathrm{T}}(u_r) \\ \cdots \end{bmatrix}$ and $\boldsymbol{M}_2 = \begin{bmatrix} \hat{\boldsymbol{\xi}}_1^{\mathrm{T}} \hat{\boldsymbol{P}}(u_1)\boldsymbol{Q} \\ \hat{\boldsymbol{\xi}}_2^{\mathrm{T}} \hat{\boldsymbol{P}}(u_1)\boldsymbol{Q} \\ \cdots \\ \hat{\boldsymbol{\xi}}_1^{\mathrm{T}} \hat{\boldsymbol{P}}(u_2)\boldsymbol{Q} \\ \cdots \\ \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r)\boldsymbol{Q} \\ \cdots \end{bmatrix}$. The matrix $\boldsymbol{M}_1$ is a $n_A n$ by n_{knots} matrix. It is easy to see that rank$(\boldsymbol{M}_1) \leq (n_A \wedge C|\Delta|^{-1})$ and rank$(\boldsymbol{M}_2) \leq m$. Therefore, rank$(\boldsymbol{Z}) \leq m + (n_A \wedge C|\Delta|^{-1})$. Then by Sard's theorem, almost

surely, Card$(I_3) \leq m + (n_A \wedge C|\Delta|^{-1})$. Combined with (S4.18) and (S4.19), we have almost surely,

$$
\begin{aligned}
&E_n\left[\frac{1}{n_A}\sum_{r=1}^{n_A}\left\{\left(u_r - \mathbb{1}\left\{Y_i - \boldsymbol{Z}_i(u_r)\hat{\boldsymbol{\Gamma}} \leq 0\right\}\right)\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\right\}\right]\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\\
=&\frac{c}{n_A n}\sum_{(i,r)\in I_3}\left(u_r - \mathbb{1}\left\{Y_i - \boldsymbol{Z_i}(u_r)\hat{\boldsymbol{\Gamma}} \leq 0\right\}\right)\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right),\\
\leq&\frac{c\{m + (n_A \wedge C|\Delta|^{-1})\}}{n_A n}\max_{i,r}\left|\left(u_r - \mathbb{1}\left\{Y_i - \boldsymbol{Z}_i(u_r)\hat{\boldsymbol{\Gamma}} \leq 0\right\}\right)\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\right|\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\|_{l^2}\\
&+\left\|\left(2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q} + 2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right\|_{l^2},\\
\leq&\frac{c\{m + (n_A \wedge C|\Delta|^{-1})\}}{n_A n}\max_{i\in 1,\ldots,n}\left\|\hat{\boldsymbol{\xi}}_i\right\|_{l^2}\max_{u_r\in A}\sigma_{\max}\{\hat{\boldsymbol{P}}(u_r)\}\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\|_{l^2}\\
&+\left\|\left(2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q} + 2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right\|_{l^2},
\end{aligned}
$$

where $\sigma_{\max}(x)$ is the largest singular value of a matrix x.

S4.1 Proof of Theorem 1

Denote the minimizer of (S4.1) by $(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}})$. Define $\hat{\beta}(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\hat{\boldsymbol{\theta}}$,

$$A_0 = \{r \in (1,\ldots,n_A) : \left\|\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r)\right\|_{l^2} \geq M\sqrt{m/n},\ \text{for some constant } M > 0\},$$

and

$$A_1 = \{r \in (1,\ldots,n_A) : \left\|\hat{\beta}_{u_r}(t) - \beta_{u_r}(t)\right\|_{L^2} \geq M\kappa_m^{-1/2}m^{1/2}n^{-1/2},\ \text{for some constant } M > 0\}.$$

Then by Lemma 4, we have for some $M > 0$

$$
\begin{aligned}
&- n_A^{-1} \sum_{r=1}^{n_A} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \left(\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right) \right] \\
= &- n_A^{-1} \sum_{r \in (A \setminus A_0)} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \left(\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right) \right] \\
&- n_A^{-1} \sum_{r \in A_0} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \left(\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right) \right], \\
\ge &- n_A^{-1} \sum_{r \in (A \setminus A_0)} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \left(\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right) \right] \\
&+ n_A^{-1} \left\{ M(1 - o_p(1)) m^{1/2} n^{-1/2} - O_p\left(m^{1/2} n^{-1/2} \right) \right\} \sum_{r \in A_0} \left\| \hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right\|_{l^2}.
\end{aligned}
\tag{S4.20}
$$

For the first term in (S4.20), almost surely

$$
\begin{aligned}
&\left| -n_A^{-1} \sum_{r \in (A \setminus A_0)} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \left(\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right) \right] \right| \\
&\le n_A^{-1} \sum_{r \in (A \setminus A_0)} E_n \left[\left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \left\| \hat{\boldsymbol{\eta}}_i^{m+1} \right\| \left\| \hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right\| \right] \\
&= n_A^{-1} n^{-1} \sum_{(i,r) \in (A \setminus A_0) \cap I_3} \left\{ u - \mathbb{1}\left(Y_i \le \hat{\boldsymbol{\eta}}_i^{m+1} \cdot \hat{\boldsymbol{d}}^{m+1}(u_r) \right) \right\} \left\| \hat{\boldsymbol{\eta}}_i^{m+1} \right\| \left\| \hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r) \right\| \\
&= o_p\left(n_A^{-1} n^{-1} m |\Delta|^{-1} \right) \times O\left(m^{1/2} n^{-1/2} \right) = o_p\left(n_A^{-1} n^{-3/2} |\Delta|^{-1} m^{3/2} \right).
\end{aligned}
\tag{S4.21}
$$

On the other hand, by standard matrix theory,

$$\begin{aligned}\max_{u_r\in A}\sigma_{\max}\{\hat{\boldsymbol{P}}(u_r)\} &\le \max_{u_r\in A}\left\|\hat{\boldsymbol{P}}(u_r)\right\|_F,\\ &\le \left\{cm\max_{u_r\in A}\sum_{b_j(t,u_r)\neq 0}\left(\int_{\Delta_j} b_j(t,u_r)\hat{\phi}_k(t)dt\right)^2\right\}^{1/2},\\ &= cm^{1/2}|\Delta|^{1/2}.\end{aligned}$$

Then, for the first term of (S4.10), we have

$$\begin{aligned}&\max_{i\in 1,\ldots,n}\left\|\hat{\boldsymbol{\xi}}_i\right\|_{l^2}\max_{u_r\in A}\sigma_{\max}\{\hat{\boldsymbol{P}}(u_r)\}\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\|_{l^2}\\ \le& o_p\left((\log n)^{-1}n^{1/2}m^{-1/2}\right)cm^{1/2}|\Delta|^{-1/2}\left\|\hat{\beta}-\beta^*\right\|_{L^2(\boldsymbol{\Omega})},\\ =& o\left((\log n)^{-1}n^{1/2}|\Delta|^{-1/2}\right)\left\|\hat{\beta}(t,u)-\sum_{k=1}^m\beta_k(u)\hat{\phi}_k(t)+\sum_{k=1}^m\beta_k(u)\hat{\phi}_k(t)-\beta^*(t,u))\right\|_{L^2(\boldsymbol{\Omega})},\\ \lesssim& o\left((\log n)^{-1}n^{1/2}|\Delta|^{-1/2}\right)\left\{n_A^{-1/2}\left(\sum_{r=1}^{n_A}\kappa_m^{-1}\left\|\hat{\boldsymbol{d}}^m(u_r)-\boldsymbol{d}^m(u_r)\right\|_{l^2}^2\right)^{1/2}+O(|\Delta|^{d+1})\right\},\end{aligned}\tag{S4.22}$$

provided that $\max_{1\le i\le n}\left\|\hat{\boldsymbol{\xi}}_i\right\|_{l^2}=o((\log n)^{-1}n^{1/2}m^{-1/2})$ (Kato, 2012). By Lemma 2 and Wang et al. (2020),

$$\left\|\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\right\|_{l^2}\le C|\Delta|^{-2}\left\|\hat{\boldsymbol{\theta}}\right\|_{l^2}=C|\Delta|^{-3}\left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}.$$

By Lemma 5,

$$\left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}\approx n_A^{-1/2}n_T^{-1/2}\left\|\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}\right\|_{l^2}\le O_p(m^{3/4}n^{-3/4}|\Delta|^{(d+1)/2}\vee m^{1/2}n^{-1/2})+C.$$

In addition, $\left\|\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\right\| \leq \Gamma_{max}(\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}})\left\|\hat{\boldsymbol{\theta}}\right\|_{l^2} = C|\Delta|n_A n_T\left\|\hat{\beta}(t,u)\right\|_{L^2(\boldsymbol{\Omega})}$.

Then we conclude that

$$\begin{aligned}
&\left\|\left(2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}+2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right)\right\|_{l^2}\\
\leq&\left\|2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{\theta}}\right\|_{l^2}+\left\|2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{\theta}}\right\|_{l^2}\\
&+\left\|\left(2\lambda_{1,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}+2\lambda_{2,n}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\right)\boldsymbol{\theta}^*\right\|_{l^2}\\
=&O_p(m^{-1/2}n|\Delta|^{d+1})(1+o(1))
\end{aligned}\tag{S4.23}$$

Given $|\Delta| = o\left(m^{(1+\alpha)/(2d+2)}n^{-3/(2d+2)}\right)$, from (S4.22), if

$$n_A^{-1/2}\left(\sum_{r=1}^{n_A}\kappa_m^{-1}\left\|\hat{\boldsymbol{d}}^m(u_r)-\boldsymbol{d}^m(u_r)\right\|_{l^2}^2\right)^{1/2} \leq O_p(|\Delta|^{d+1}) = o_p\left(m^{(1+\alpha)/2}n^{-1/2}\right),$$

then by Lemma 5, we can conclude $\left\|\hat{\beta}(t,u)-\beta(t,u)\right\|_{L^2(\boldsymbol{\Omega})} \approx O_p\left(\kappa_m^{-1/2}m^{1/2}n^{-1/2}\right)$.

If not, then

$$\begin{aligned}
\text{(S4.22)} &= o_p\left((\log n)^{-1}n^{1/2}|\Delta|^{-1/2}\right)\left(n_A^{-1/2}\sum_{r=1}^{n_A}\kappa_m^{-1/2}\left\|\hat{\boldsymbol{d}}^m(u_r)-\boldsymbol{d}^m(u_r)\right\|_{l^2}^2+O(|\Delta|^{d+1})\right)^{1/2},\\
&= o_p\left((\log n)^{-1}n_A^{-1/2}n^{1/2}|\Delta|^{-1/2}\kappa_m^{-1/2}\right)\sum_{r\in A_0}\left\|\hat{\boldsymbol{d}}^m(u_r)-\boldsymbol{d}^m(u_r)\right\|_{l^2}(1+o(1)).
\end{aligned}\tag{S4.24}$$

By Lemma 6, (S4.23) and (S4.24) , almost surely

$$\begin{aligned}
&-n_A^{-1}\sum_{r=1}^{n_A}E_n\left[\left\{u-\mathbb{1}\left(Y_i\leq\hat{\boldsymbol{\eta}}_i^{m+1}\cdot\hat{\boldsymbol{d}}^{m+1}(u_r)\right)\right\}\hat{\boldsymbol{\eta}}_i^{m+1}\cdot\left(\hat{\boldsymbol{d}}^{m+1}(u_r)-\boldsymbol{d}^{m+1}(u_r)\right)\right]\\
&\leq o_p\left(n_A^{-3/2}|\Delta|^{-3/2}n^{-1/2}\kappa_m^{-1/2}\right)\sum_{r\in A_0}\left\|\hat{\boldsymbol{d}}^m(u_r)-\boldsymbol{d}^m(u_r)\right\|_{l^2}+O_p(m^{-1/2}n|\Delta|^{d+1})(1+o(1))
\end{aligned}\tag{S4.25}$$

Then by Lemma 4, (S4.20), (S4.21), (S4.25) and $n_A^{-1}|\Delta|^{-1}m^{(\alpha-1)/3} = o(1)$, we have almost surely

$$\left\{M(1-o_p(1))m^{1/2}n^{-1/2} - O_p\left(m^{1/2}n^{-1/2}\right)\right\}\sum_{r\in A_0}\left\|\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r)\right\|_{l^2}$$
$$\leq o_p\left(n_A^{-3/2}|\Delta|^{-3/2}n^{-1/2}\kappa_m^{-1/2}\right)\sum_{r\in A_0}\left\|\hat{\boldsymbol{d}}^{m}(u_r) - \boldsymbol{d}^{m}(u_r)\right\|_{l^2} + O_p(m^{-1/2}n|\Delta|^{d+1})(1+o(1)), \tag{S4.26}$$

which implies that

$$\sum_{r\in A_0}\left\|\hat{\boldsymbol{d}}^{m+1}(u_r) - \boldsymbol{d}^{m+1}(u_r)\right\|_{l^2} = O_p(n_A n^{1/2}|\Delta|^{d+1}).$$

Based on the definition of A_0 and A_1, we have $A_1 \subset A_0$ and $|A_0| = o_p(m^{-1-\alpha}n^{-1/2}n_A)$.

Then by Lemma 4,

$$\left\|\hat{\beta}(t,u) - \sum_{k=1}^{m}\beta_k(u)\phi_k(t)\right\|_{L^2(\boldsymbol{\Omega})} \lesssim n_A^{-1/2}\left(\sum_{r=1}^{n_A}\kappa_m^{-1}\left\|\hat{\boldsymbol{d}}^{m}(u_r) - \boldsymbol{d}^{m}(u_r)\right\|_{l^2}^2\right)^{1/2} = O_p(\kappa_m^{-1/2}m^{1/2}n^{-1/2}).$$

By the condition A5 given in the main paper,

$$\left\|\beta(t,u) - \sum_{k=1}^{m}\beta_k(u)\phi_k(t)\right\|_{L^2(\boldsymbol{\Omega})} = O(m^{-(2\zeta+1)/2}).$$

Therefore, we conclude that

$$\left\|\hat{\beta}(t,u) - \beta(t,u)\right\|_{L^2(\boldsymbol{\Omega})} \approx O_p\left(\kappa_m^{-1/2}m^{1/2}n^{-1/2} \vee m^{-2\zeta+1}\right).$$

S4.2 Proof of Theorem 2

Assume that p_0 is known and let $m = p_0$. Under the condition A5 given in the main paper, and by Lemma 1 and Lemma 3, there exist $\boldsymbol{\gamma}_0^*$ and $\boldsymbol{\theta}^*$ such that

$$\sup_{(t,u)\in\mathscr{T}\times\mathscr{A}} |\beta(t,u) - \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\boldsymbol{\theta}^*| \leq C_1|\Delta|^{d+1},$$

$$\sup_{u\in\mathscr{A}} |c(u) - \boldsymbol{b}_0^{\mathrm{T}}(u)\boldsymbol{\gamma}_0^*| \leq C_2|\Delta|^{d+1},$$

for some constant C_1 and C_2. Let $\boldsymbol{\Gamma}^* = (\boldsymbol{\gamma}_0^{*\mathrm{T}}, \boldsymbol{\theta}^{*\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\delta} = \sqrt{n}(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}_0^*, \boldsymbol{\theta} - \boldsymbol{\theta}^*)^{\mathrm{T}}$. Let $\boldsymbol{P}(u_r)$ denote the p_0 by J matrix with the (k,j)-entry being $\int_{(t,u_r)\in\Delta_j} \phi_k(t)b_j(t,u)dt$, where $j = 1,\ldots,J$ is the index for the bivariate splines basis. Then define

$$\psi_u(x) = u - \mathbb{1}(x \leq 0), \quad \hat{\omega}_{i,r} = y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*, \quad \text{and} \quad \omega_{i,r} = y_i - \boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*,$$

where

$$\boldsymbol{Z}_i^*(u_r) = \left[\boldsymbol{b}_0^{\mathrm{T}}(u_r), \boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{Q}\right].$$

We further define $L_n^{0,1}(\boldsymbol{\Gamma},\boldsymbol{\delta})$ and $L_n^{0,2}(\boldsymbol{\Gamma},\boldsymbol{\delta})$ as follow,

$$\begin{aligned} L_n^{0,1}(\boldsymbol{\Gamma},\boldsymbol{\delta}) &= \{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\frac{\boldsymbol{\delta}}{\sqrt{n}}\left\{\mathbb{1}(y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0) - u_r\right\} \\ &= -\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\frac{\boldsymbol{\delta}}{\sqrt{n}}\psi_{u_r}(\hat{\omega}_{i,r}), \end{aligned}$$

and

$$L_n^{0,2}(\boldsymbol{\Gamma}) = (nn_A)^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} \{\mathbb{1}\left(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq t\right) - \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0)\} \, dt$$
$$= n_A^{-1} \sum_{r=1}^{n_A} G_r(\boldsymbol{\Gamma}, \boldsymbol{\delta}),$$

where

$$G_r(\boldsymbol{\Gamma}, \boldsymbol{\delta}) = n^{-1} \sum_{i=1}^{n} \int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} \{\mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq t) - \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0)\} \, dt.$$

Lemma 7. *Under the conditions of Theorem 1,*

$$L_n^{0,2}(\boldsymbol{\Gamma}, \boldsymbol{\delta}) = \frac{1}{2n} \boldsymbol{\delta}^{\mathrm{T}} \boldsymbol{\Sigma}_1 \boldsymbol{\delta} + o_p(1),$$

where

$$\boldsymbol{\Sigma}_1 = n_A^{-1} \sum_{r=1}^{n_A} E\left[f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*) \boldsymbol{Z}_i^{\mathrm{T}}(u_r) \boldsymbol{Z}_i(u_r)\right].$$

Proof: By Taylor expansion,

$$\begin{aligned}
E\left[G_r(\boldsymbol{\Gamma}, \boldsymbol{\delta})\right] &= E\left[E\left[G_r(\boldsymbol{\Gamma}, \boldsymbol{\delta}) \mid \hat{\boldsymbol{\xi}}_i, i = 1, \ldots, n\right]\right] \\
&= E\left[n^{-1} \sum_{i=1}^{n} \int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} \{F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* + t) - F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\} \, dt\right] \\
&= n^{-1} \sum_{i=1}^{n} E\left[\int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} \{f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)t + R_{i,r}(t)\} \, dt\right] \\
&= n^{-1} \sum_{i=1}^{n} E\left[\frac{f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)}{2} t^2 \Big|_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} + R_{i,r}(t)\right] \\
&= n^{-1} \sum_{i=1}^{n} E\left[\frac{1}{2n} f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*) \boldsymbol{\delta}^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}}(u_r) \boldsymbol{Z}_i(u_r) \boldsymbol{\delta}\right] + \sum_{i=1}^{n} E\left[R_{i,r}(t)\right] \\
&= \frac{1}{2n} \boldsymbol{\delta}^{\mathrm{T}} \left\{n^{-1} \sum_{i=1}^{n} E\left[f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*) \boldsymbol{Z}_i^{\mathrm{T}}(u_r) \boldsymbol{Z}_i(u_r)\right]\right\} \boldsymbol{\delta} + \sum_{i=1}^{n} E\left[R_{i,r}(t)\right],
\end{aligned}$$

where $\{R_{i,r}(t)\}$ are the remainder terms of Taylor expansion. Regarding these remainder terms $\{R_{i,r}(t)\}$, we have

$$\sum_{i=1}^{n} E\left[R_{i,r}(t)\right] \leq \text{constant} \cdot n^{-3/2} \sum_{i=1}^{n} E\left[(\boldsymbol{Z}_i(u_r)\boldsymbol{\delta})^3\right] \leq O(n^{-1/2}) E\left[\|\boldsymbol{Z}_i(u_r)\|_{l^2}^3\right] \|\boldsymbol{\delta}\|_{l^2}^3 = o(1).$$

Then, we can conclude that

$$E\left[L_n^{0,2}(\boldsymbol{\Gamma})\right] = \frac{1}{2n}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\Sigma}_1\boldsymbol{\delta} + o(1).$$

Next, we need to prove $E\left[L_n^{0,2}(\boldsymbol{\Gamma})^2\right] = o(1)$.

$$\begin{aligned}
E\left[\left(L_n^{0,2}(\boldsymbol{\Gamma})\right)^2\right] =& E\left[E\left[\left(L_n^{0,2}(\boldsymbol{\Gamma})\right)^2 \mid \hat{\boldsymbol{\xi}}_i, i=1,\ldots,n\right]\right] \\
\leq& E\Bigg[E\Bigg[(n_A n)^{-1} \sum_{r=1}^{n_A}\sum_{i=1}^{n}\Bigg\{\int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}} \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq t) \\
&- \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0)dt\Bigg\}^2 \mid \hat{\boldsymbol{\xi}}_i, i=1,\ldots,n\Bigg]\Bigg] \\
\leq& E\Bigg[(n_A n)^{-1} \sum_{r=1}^{n_A}\sum_{i=1}^{n}\int_0^{|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}|} |\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}| E\Bigg[\Big\{\mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq t) \\
&- \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0)\Big\}^2 dt \mid \hat{\boldsymbol{\xi}}_i, i=1,\ldots,n\Bigg]\Bigg] \\
=& E\Bigg[(n_A n)^{-1} \sum_{r=1}^{n_A}\sum_{i=1}^{n} |\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}| \int_0^{|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}|} E\Bigg[\Big|\mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq t) \\
&- \mathbb{1}(Y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \leq 0)\Big| dt \mid \hat{\boldsymbol{\xi}}_i, i=1,\ldots,n\Bigg]\Bigg].
\end{aligned}$$

Thus,

$$
\begin{aligned}
E\left[\left(L_n^{0,2}(\boldsymbol{\Gamma})\right)^2\right] \leq & E\Bigg[(n_A n)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}| \\
& \int_0^{|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}|} E\big[\mathbb{1}\big(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* < Y_i < \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* + t\big)\big]dt \mid \hat{\boldsymbol{\xi}}_i, i=1,\ldots,n\Bigg]\Bigg] \\
= & E\Bigg[(n_A n)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}| \\
& \int_0^{|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}|} \left\{F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* + t) - F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\right\}dt\Bigg] \\
= & (n_A n)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n} E\left[\frac{|\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}|}{2n} f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{Z}_i^{\mathrm{T}}(u_r)\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}\right] \\
& + (n_A n)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n} E\left[R_{i,r}(t)\right] \\
= & o(1).
\end{aligned}
$$

Define

$$
\boldsymbol{V}_0 = -(n_A n)^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i^*(u_r)\psi_{u_r}(\omega_{i,r}),
$$

and let $\tilde{\boldsymbol{Z}}_i$ and $\boldsymbol{\psi}$ denote $(\boldsymbol{Z}_i^{\mathrm{T}}(u_1),\ldots,\boldsymbol{Z}_i^{\mathrm{T}}(u_{n_A}))$ and $\left(\psi_{u_1},\ldots,\psi_{u_{n_A}}\right)$ respectively.

Lemma 8. *Under the conditions of Theorem 1 and A6 given in the main paper,*

$$
\sqrt{n}\boldsymbol{V}_0 \to N(\boldsymbol{0},\boldsymbol{U}_2),
$$

in distribution, where

$$
\boldsymbol{U}_2 = n_A^{-2}E\left[\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{U}_1\tilde{\boldsymbol{Z}}_i\right].
$$

Proof: First notice that

$$
\begin{aligned}
&\text{cov}\left(\psi_{u_r}(\omega_{i,r}), \psi_{u_r}(\omega_{j,r})\right) \\
&= \begin{cases} 0, & i \neq j \\ u_{r'} - u_r u_{r'} + O\left(\left|\int \sum_{k=1}^{p_0} \xi_{i,k}\phi_k(t)\left(\beta(t,u_r) - \beta^*(t,u_r)\right)dt\right|\right), & \text{otherwise} \end{cases} \\
&= \begin{cases} 0, & i \neq j \\ u_r \wedge u_{r'} - u_r u_{r'} + O_p(|\Delta|^{d+1}), & \text{otherwise} \end{cases}
\end{aligned}
$$

Since $E\left[\psi_{u_r}(\omega_{i,r}) \mid \boldsymbol{\xi}_i\right] = u_r - F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*) = O_p(|\Delta|^{d+1})$, then $E\left[\boldsymbol{V}_0 \mid \boldsymbol{\xi}_i\right] = O_p(|\Delta|^{d+1})$. On the other hand, $Var\left(\boldsymbol{V}_0 \mid \boldsymbol{\xi}_i\right) = \text{var}\left(n^{-1}\sum_{i=1}^n n_A^{-1}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{\psi}\right) = (n_A n)^{-2}\sum_{i=1}^n \tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{U}_1\tilde{\boldsymbol{Z}}_i$, where $\boldsymbol{U}_1$ is a matrix with its (r,r')-entry being $u_r \wedge u_{r'} - u_r u_{r'}$. Then the co-variance matrix of $\boldsymbol{V}_0$ is given by

$$
\begin{aligned}
\text{var}(\boldsymbol{V}_0) =& \text{var}\left(E\left(\boldsymbol{V}_0 \mid \boldsymbol{\xi}_i\right)\right) + E\left(\text{var}\left(\boldsymbol{V}_0 \mid \boldsymbol{\xi}_i\right)\right) \\
=& O_p(|\Delta|^{d+1}) + n^{-1}\left[n_A^{-2}E\left[\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{U}_1\tilde{\boldsymbol{Z}}_i\right]\right] \\
=& O_p(|\Delta|^{d+1}) + n^{-1}\boldsymbol{U}_2.
\end{aligned}
$$

Regarding $\boldsymbol{U}_2$, we have

$$
\begin{aligned}
\boldsymbol{U}_2 \leq& \frac{2}{n_A^2}E\left[\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_i\right] \leq \frac{2}{n_A^2}n_A \max_r E\left\|\boldsymbol{Z}_i(u_r)\right\|_2^2 \\
=& \frac{2}{n_A}\max_r\left\{\left\|\boldsymbol{b}_0(u_r)\right\|_2^2 + E\left\|\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{Q}\right\|_2^2\right\} \\
=& O\left(n_A^{-1}|\Delta|^{-1}\right).
\end{aligned}
$$

Based on $n_A n|\Delta|^{d+2} = o(1)$, we just prove that

$$\sqrt{n}\boldsymbol{V}_0 \to N(\mathbf{0}, \boldsymbol{U}_2)$$

in distribution.

Define

$$\begin{aligned} L_n^0(\boldsymbol{\Gamma}) &= \{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \rho_{u_r} \left(y_i - \boldsymbol{b}_0^{\mathrm{T}}(u_r)\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\xi}}_i^{\mathrm{T}} \hat{\boldsymbol{P}}(u_r)\boldsymbol{Q}\boldsymbol{\theta} \right) \\ &= \{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \rho_{u_r} \left(y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma} \right), \end{aligned} \tag{S4.27}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_0, \boldsymbol{\theta})^{\mathrm{T}}$, and $P_\lambda(\boldsymbol{\Gamma}) = \lambda_{1,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\theta} + \lambda_{2,n}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\boldsymbol{\theta}$.

Recall that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^* + \boldsymbol{\delta}/\sqrt{n}$. Then minimizing

$$L_n(\boldsymbol{\Gamma}) = L_n^0(\boldsymbol{\Gamma}) + P_\lambda(\boldsymbol{\Gamma})$$

is equivalent to minimizing

$$\begin{aligned} LL_n(\boldsymbol{\delta}) = \{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \left\{ \rho_{u_r}\left(y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n} - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \right) - \rho_{u_r}\left(y_i - \boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^* \right) \right\} \\ + P_\lambda(\boldsymbol{\delta}/\sqrt{n} + \boldsymbol{\Gamma}^*). \end{aligned} \tag{S4.28}$$

Applying the Knight's identity (Knight, 1998),

$$\rho_u(x-y) - \rho_u(x) = y\left\{\mathbb{1}(x \le 0) - u\right\} + \int_0^y \left\{\mathbb{1}(x \le t) - \mathbb{1}(x \le 0)\right\} dt,$$

on the first term of (S4.28), then we have

$$
\begin{aligned}
L_n^0(\boldsymbol{\Gamma}) =& \{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\Big\{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}\,\{\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\le 0)-u_r\}\\
&+\int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}}\{\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\le t)-\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\le 0)\}\,dt\Big\}\\
=& \{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}\,\{\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\le 0)-u_r\}\\
&+\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\int_0^{\boldsymbol{Z}_i(u_r)\boldsymbol{\delta}/\sqrt{n}}\{\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\le t)-\mathbb{1}(y_i-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\le 0)\}\,dt\\
=& L_n^{0,1}(\boldsymbol{\Gamma})+L_n^{0,2}(\boldsymbol{\Gamma}).
\end{aligned}
$$

From Lemma 7 and Lemma 8, under the conditions $\lambda_{1,n}\asymp n_A^{-1}n_T^{-1}m^{-1/2}n|\Delta|^{d+1}$, and $\lambda_{2,n}=o(\lambda_{1,n}n_An_T|\Delta|^4)$, we have

$$
\begin{aligned}
LL_n(\boldsymbol{\delta}) =& -\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\frac{1}{\sqrt{n}}\boldsymbol{\delta}\psi_{u_r}(\hat{\omega}_{i,r})\\
&+\frac{1}{2n}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\Sigma}_1\boldsymbol{\delta}+E\left[R(t)\right]+P_\lambda(\frac{1}{\sqrt{n}}\boldsymbol{\delta}+\boldsymbol{\Gamma}^*)\\
=& -\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\frac{1}{\sqrt{n}}\boldsymbol{\delta}\psi_{u_r}(\hat{\omega}_{i,r})\\
&+\frac{1}{2n}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\Sigma}_1\boldsymbol{\delta}+\boldsymbol{\delta}^{\mathrm{T}}\left[\lambda_{1,n}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\end{bmatrix}+\lambda_{2,n}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\end{bmatrix}\right]\boldsymbol{\delta}\\
&+\frac{2}{\sqrt{n}}\boldsymbol{\delta}^{\mathrm{T}}\left[\lambda_{1,n}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\end{bmatrix}+\lambda_{2,n}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\end{bmatrix}\right]\boldsymbol{\Gamma}^*+o(1)\\
=& \boldsymbol{V}\boldsymbol{\delta}(1+o_p(1))+\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\Sigma}_2\boldsymbol{\delta}+o(1),
\end{aligned}
$$

with

$$\boldsymbol{V} = -\left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\boldsymbol{Z}_i(u_r)\psi_{u_r}(\hat{\omega}_{i,r}),$$

$$\boldsymbol{\Sigma}_2 = \frac{1}{2n}\boldsymbol{\Sigma}_1 + \lambda_{1,n}\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}_{A,T}\boldsymbol{B}_{A,T}^{\mathrm{T}}\boldsymbol{Q}\end{bmatrix} + \lambda_{2,n}\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\end{bmatrix}.$$

Then by the convexity lemma (Pollard, 1991) and quadratic approximation lemma (Fan and Gijbels, 1996), we have $\hat{\boldsymbol{\delta}} = -\frac{1}{2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{V}^{\mathrm{T}} + o_p(1)$. Let $\tilde{\boldsymbol{B}}(t,u) = (\boldsymbol{0}_{1\times n_B}, \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q})^{\mathrm{T}}$, and $\beta^*(t,u) = \boldsymbol{B}^{\mathrm{T}}(t,u)\boldsymbol{Q}\boldsymbol{\theta}^*$. Then, we have $\hat{\beta}(t,u) - \beta^*(t,u) = \tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\hat{\boldsymbol{\delta}}/\sqrt{n}$. Therefore,

$$\begin{aligned}\hat{\beta}(t,u) - \beta(t,u) &= \tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\hat{\boldsymbol{\delta}}/\sqrt{n} + \beta^*(t,u) - \beta(t,u)\\ &= -\frac{1}{2n}\tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}_2^{-1}\sqrt{n}\boldsymbol{V} + \beta^*(t,u) - \beta(t,u) + o_p(1).\end{aligned}$$

Now, we need to prove the asymptotic normality of $\boldsymbol{V}$. We first decompose $\boldsymbol{V}_0 - \sqrt{n}\boldsymbol{V}$ into three terms and then calculate the order for each of them.

$$\begin{aligned}&\left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i(u_r)\psi_{u_r}(\hat{\omega}_{i,r}) - \left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i^*(u_r)\psi_{u_r}(\omega_{i,r})\\ =&\left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\left\{\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r)\right\}\left\{\psi_{u_r}(\hat{\omega}_{i,r}) - \psi_{u_r}(\omega_{i,r})\right\}\\ &+\left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i^*(u_r)\left\{\psi_{u_r}(\hat{\omega}_{i,r}) - \psi_{u_r}(\omega_{i,r})\right\}\\ &+\left\{nn_A\right\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\left\{\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r)\right\}\psi_{u_r}(\omega_{i,r})\\ =&\boldsymbol{V}_1 + \boldsymbol{V}_2 + \boldsymbol{V}_3.\end{aligned}$$

First notice that

$$
\begin{aligned}
&E\left[\psi_{u_r}(\hat{\omega}_{i,r})\middle|\boldsymbol{Z}_i(u_r),\boldsymbol{Z}_i^*(u_r)\right]\\
=&u_r-F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\\
=&u_r-F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)+F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)-F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\\
=&o_p(1)+f_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)\left(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\right)+O\left(\left(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*-\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*\right)^2\right).
\end{aligned}
$$

Then for the first term $\boldsymbol{V}_1=\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\{\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r)\}\{\psi_{u_r}(\hat{\omega}_{i,r})-\psi_{u_r}(\omega_{i,r})\}$,

$$
\begin{aligned}
E|\boldsymbol{V}_1|\leq&E\left[\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r)|\,|\psi_{u_r}(\hat{\omega}_{i,r})-\psi_{u_r}(\omega_{i,r})|\right]\\
=&E\left[\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r)|\,E\left[|\psi_{u_r}(\hat{\omega}_{i,r})-\psi_{u_r}(\omega_{i,r})|\mid\boldsymbol{Z}_i,\boldsymbol{Z}_i^*\right]\right]\\
=&E\left[\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)-F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)|\,|\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r)|\right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E|\boldsymbol{V}_1|\leq&E\left[\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|F_i(\boldsymbol{Z}_i(u_r)\boldsymbol{\Gamma}^*)-F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)|^2\right]^{1/2}\\
&\times E\left[\{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}|\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r)|^2\right]^{1/2}\\
\leq&O\left(\max_r E\left[(\boldsymbol{Z}_i(u_r)-\boldsymbol{Z}_i^*(u_r))^2\right]^{1/2}\right)\\
=&O\left(\max_r E\left[\left\{\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)-\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r)\right\}^2\right]^{1/2}\right)\\
=&O\left(\max_r E\left[\left\{\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)-\boldsymbol{\xi}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)+\boldsymbol{\xi}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r)-\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r)\right\}^2\right]^{1/2}\right)
\end{aligned}
$$

According to (Hall and Hosseini-Nasab, 2006) and (Hall and Hosseini-Nasab,

2009), we have

$$\left\|\hat{\phi}_k - \phi_k\right\|_{L^2(\boldsymbol{\Omega})} \leq \text{constant} \cdot s_k^{-1} n^{-1/2},$$

and for any $c > 0$,

$$E\left\|\hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i\right\|^c \leq \text{constant} \cdot s_{p_0}^{-c} n^{-c/2},$$

where $s_k = \min_{r \leq k}(\kappa_r - \kappa_{r+1})$. Therefore, we can conclude that

$$E|\boldsymbol{V}_1| = O(n^{-1/2}|\Delta|).$$

Similarly,

$$\begin{aligned} E\left[\boldsymbol{V}_1^2\right] \leq & E\left[\{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \{\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r)\}^2 \{\psi_{u_r}(\hat{\omega}_{i,r}) - \psi_{u_r}(\omega_{i,r})\}^2\right] \\ \leq & \max_r E\left[(\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r))^4\right]^{1/2} \max_r E\left[(\psi_{u_r}(\hat{\omega}_{i,r}) - \psi_{u_r}(\omega_{i,r}))^4\right]^{1/2} \\ = & \text{constant} \cdot n^{-1}|\Delta|^2. \end{aligned}$$

Then we have

$$\boldsymbol{V}_1 = O_p(n^{-1/2}|\Delta|).$$

For the second term $\boldsymbol{V}_2 = \{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \boldsymbol{Z}_i^*(u_r) \{\psi_{u_r}(\hat{\omega}_{i,r}) - \psi_{u_r}(\omega_{i,r})\}$, we first define

$$R_r(\boldsymbol{t}_r) = \sum_{i=1}^{n} \boldsymbol{Z}_i^*(u_r) \{\psi_{u_r}(\omega_{i,r} - \boldsymbol{\xi}_i^{\mathrm{T}} \boldsymbol{t}_r) - \psi_{u_r}(\omega_{i,r})\}$$

for any vector such that $\|\boldsymbol{t}_r\| \leq C_t$ with some constant C_t. According to (Li et al., 2022), we have $\sup \|R_r(\boldsymbol{t}_r) - E[R_r(\boldsymbol{t}_r)]\| = O_p\left(n^{1/2} \log(n) \|\boldsymbol{t}_r\|^{1/2}\right)$,

and therefore,

$$E\left[R_r(\boldsymbol{t}_r)\right] = \sum_{i=1}^{n} E\left[\boldsymbol{Z}_i^*(u_r)\left\{F_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*) - F_i\left(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^* - [\boldsymbol{b}_0(u_r)^{\mathrm{T}}, \boldsymbol{\xi}_i^{\mathrm{T}}]\,\boldsymbol{t}_r\right)\right\}\right]$$
$$= -\,nE\big[\boldsymbol{Z}_i^*(u_r) f_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)\left([\boldsymbol{b}_0(u_r)^{\mathrm{T}}, \boldsymbol{\xi}_i^{\mathrm{T}}]\,\boldsymbol{t}_r\right)\big] + O\left(nE\left[\left([\boldsymbol{b}_0(u_r)^{\mathrm{T}}, \boldsymbol{\xi}_i^{\mathrm{T}}]\,\boldsymbol{t}_r\right)^2 \boldsymbol{Z}_i^*(u_r)\right]\right). \tag{S4.29}$$

Define $\boldsymbol{\Sigma}_{c,r} = E\big[f_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)\boldsymbol{b}_0\boldsymbol{b}_0^{\mathrm{T}}\big]$ and $\boldsymbol{\Sigma}_{3,r} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}(u_r)E\big[f_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\big]$.

Then from (S4.29), we have

$$R_r^{\mathrm{T}}(\boldsymbol{t}_r) = -n\left[\boldsymbol{\Sigma}_{c,r}, \boldsymbol{\Sigma}_{3,r}\right]\boldsymbol{t}_r + O\left(n\left\|\boldsymbol{t}_r\right\|_2^2\right) + O_p\left(n^{1/2}\log(n)\left\|\boldsymbol{t}_r\right\|_2^{1/2}\right).$$

By (Li et al., 2022), we know that there exists a random matrix $\boldsymbol{C}_\xi$ such that

$$\hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i = n^{-1/2}\boldsymbol{C}_\xi\boldsymbol{\xi}_i + O_p(n^{-1}).$$

The dimension of $\boldsymbol{C}_\xi = (c_{k,k'})$ is p_0 by p_0 where $c_{k,k'} = 0$ if $k = k'$ and $c_{k,k'} = n^{-1/2}(\kappa_k - \kappa_{k'})^{-1}\sum_{i=1}^{n}\xi_{ik}\xi_{ik'}$ if $k \neq k'$. Then,

$$\begin{aligned}
\hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r) - \boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r) &= \hat{\boldsymbol{\xi}}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r) - \boldsymbol{\xi}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r) + \boldsymbol{\xi}_i^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r) - \boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{P}(u_r)\\
&= n^{-1/2}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{C}_\xi^{\mathrm{T}}\hat{\boldsymbol{P}}(u_r) + O_p\left(n^{-1}|\Delta|\right) + O_p\left(n^{-3/4}|\Delta|\right)\\
&= n^{-1/2}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{C}_\xi^{\mathrm{T}}\left(\boldsymbol{P}(u_r) + O_p\left(n^{-3/4}|\Delta|\right)\right) + O_p\left(n^{-1}|\Delta|\right) + O_p\left(n^{-3/4}|\Delta|\right)\\
&= n^{-1/2}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{C}_\xi^{\mathrm{T}}\boldsymbol{P}(u_r) + o_p(1).
\end{aligned}$$

Choose $\boldsymbol{t}_r$ as $\boldsymbol{t}_r^{\mathrm{T}} = \left[\boldsymbol{0}, n^{-1/2}\boldsymbol{\Gamma}^{*\tau}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}(u_r)\boldsymbol{C}_\xi\right]$, and define $\boldsymbol{\Sigma}_3 = n_A^{-1}\sum_{r=1}^{n_A}\boldsymbol{\Sigma}_{3,r}\boldsymbol{C}_\xi^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{Q}$.

Then,

$$\begin{aligned}\boldsymbol{V}_2 &= \{nn_A\}^{-1}\sum_{r=1}^{n_A}\sum_{i=1}^{n}\boldsymbol{Z}_i^*(u_r)\left\{\psi_{u_r}(\omega_{i,r}-[\boldsymbol{b}_0(u_r)^{\mathrm{T}},\boldsymbol{\xi}_i^{\mathrm{T}}]\,\boldsymbol{t}_r)-\psi_{u_r}(\omega_{i,r})\right\}\\ &= -\,n^{-1/2}n_A^{-1}\sum_{r=1}^{n_A}\boldsymbol{\Sigma}_{3,r}\boldsymbol{C}_\xi^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{\Gamma}^*\boldsymbol{Q}\left(1+o_p(1)\right)\\ &= -\,n^{-1/2}\boldsymbol{\Sigma}_3\boldsymbol{\Gamma}^*\left(1+o_p(1)\right),\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_3\boldsymbol{\Gamma}^* =& n_A^{-1}\sum_{r=1}^{n_A}\boldsymbol{\Sigma}_{3,r}\boldsymbol{C}_\xi^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{Q}\boldsymbol{\Gamma}^*\\ =& n_A^{-1}\sum_{r=1}^{n_A}f_i(\boldsymbol{Z}_i^*(u_r)\boldsymbol{\Gamma}^*)\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}(u_r)E\left[\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\right]\boldsymbol{C}_\xi^{\mathrm{T}}\boldsymbol{P}(u_r)\boldsymbol{Q}\boldsymbol{\Gamma}^*.\end{aligned}$$

We know that $E\left[\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\right]$ is a diagonal matrix with the main diagonal entries being the variances of functional principal component scores. For $\boldsymbol{C}_\xi$, its main diagonal are zeros and its off diagonal entries are of order $O_p(1)$ because under the conditions A2 and A3 given in the main paper, we have

$$(\kappa_k-\kappa_{k'})^{-2}E\left[(\xi_{ik}\xi_{ik'})^2\right]\le(\kappa_k-\kappa_{k'})^{-2}\left(E\left[\xi_{ik}^4\right]E\left[\xi_{ik'}^4\right]\right)^{1/2}=O(1).$$

On the other hand, recall that for each r, $\boldsymbol{P}(u_r)$ is a p_0 by J matrix with the (k,j)-entry being $\int_{(t,u_r)\in\Delta_j}\phi_k(t)b_j(t,u)dt$, where $j=1,\ldots,J$ is the index for the bivariate splines basis and $J=O(|\Delta|^{-1})$. By the choice of $\boldsymbol{\Gamma}^*$, we know that

$$\boldsymbol{P}(u_r)\boldsymbol{Q}\boldsymbol{\Gamma}^*=\begin{bmatrix}\beta_1(u_r)\\ \ldots\\ \beta_{p_0}(u_r)\end{bmatrix},$$

where $\beta_k(u) = \int \beta(t,u)\phi_k(t)dt$, $k = 1, \ldots, p_0$. Given that p_0 is finite, we can conclude that

$$\boldsymbol{V}_2 = O_p(n^{-1/2}|\Delta|).$$

For the third term $\boldsymbol{V}_3 = \{nn_A\}^{-1} \sum_{r=1}^{n_A} \sum_{i=1}^{n} \{\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r)\} \psi_{u_r}(\omega_{i,r})$, since

$$E\left[\psi_{u_r}(\omega_{i,r}) \mid \boldsymbol{\xi}_i\right] = O_p(|\Delta|^{d+1}),$$

$$E\left[\psi_{u_r}^2(\omega_{i,r}) \mid \boldsymbol{\xi}_i\right] = u_r - u_r^2 + O_p(|\Delta|^{d+1}),$$

then we have

$$E\left[\boldsymbol{V}_3\right] = O_p(|\Delta|^{d+1}),$$

and

$$E\left[\boldsymbol{V}_3^2\right] \leq n_A^{-1} \sum_{r=1}^{n_A} u_r(1-u_r) E\left[(\boldsymbol{Z}_i(u_r) - \boldsymbol{Z}_i^*(u_r))^2\right] = O_p(n^{-1}|\Delta|^2),$$

which implies that

$$\boldsymbol{V}_3 = O_p(n^{-1/2}|\Delta|).$$

Based on the above calculation and Lemma 8, under the conditions of Theorem 1, by Slutsky's theorem we have

$$\sqrt{n}\boldsymbol{V} \to N(\boldsymbol{0}, \boldsymbol{U}_2/n) \tag{S4.30}$$

in distribution. Finally we prove the asymptotic normality of $\hat{\beta}(t,u) - \beta(t,u)$.

Recall that $\hat{\beta}(t,u) - \beta(t,u)$ admits the following decomposition,

$$\hat{\beta}(t,u) - \beta(t,u) = -\frac{1}{2n}\tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}_2^{-1}\sqrt{n}\boldsymbol{V} + \beta^*(t,u) - \beta(t,u) + o_p(1).$$

Under the conditions of Theorem 1, the bias of $\hat{\beta}(t,u) - \beta(t,u)$ is asymptotically negligible. Then by (S4.30) and Slutsky's theorem, we can conclude that

$$\frac{1}{\sqrt{\sigma_\beta(t,u)}}\left(\hat{\beta}(t,u) - \beta(t,u)\right) \to N(0,1)$$

in distribution, where $\sigma_\beta(t,u) = \tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(t,u)$.

S4.3 Proof of Theorem 3

Define $\vartheta(t,u) = \sigma_\beta(t,u)^{-1/2}\tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}_2^{-1}\boldsymbol{U}_2^{1/2}n^{-1/2}\sum_{i=1}^n \tilde{\boldsymbol{Z}}_i$, where $\tilde{\boldsymbol{Z}}_1,\ldots,\tilde{\boldsymbol{Z}}_n \overset{iid}{\sim} N(0,\boldsymbol{I})$ and the dimension of $\tilde{\boldsymbol{Z}}_i$ is same as the dimension of $\boldsymbol{Z}_i$, $i = 1,\ldots,n$. Note that $\vartheta(t,u)$ is a Gaussian random field with $E\left[\vartheta(t,u)\right] = 0$ and $Var\left[\vartheta(t,u)\right] = 1$ for any (t,u), and its covariance function is given by

$$Cov\left[\vartheta(t,u),\vartheta(t',u')\right] = \sigma_\beta^{-1/2}(t,u)\sigma_\beta^{-1/2}(t',u')\tilde{\boldsymbol{B}}^{\mathrm{T}}(t,u)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(t',u').$$

For part (1), similar to the proofs for Theorem 3 in (Ma, 2016), Theorem 3 in (Guan et al., 2021), and Theorem 5 in (Belloni et al., 2019), by the strong approximation theorem (Csörgo and Révész, 2014), we can prove that

$$\sup_{t,u}\left|\sigma_\beta^{-1/2}(t,u)\left\{\hat{\beta}(t,u) - \beta(t,u)\right\} - \vartheta(t,u)\right| = o_p(1). \tag{S4.31}$$

For part (2), based on the triangulation, we first partition $\boldsymbol{\Omega}$,the domain of $\beta(t,u)$, into M triangles with vertices $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_{J_M}$. Then we can construct

the simultaneous confidence regions (SCRs) for the estimator $\hat{\beta}(t,u)$ over a subset of Ω, $\Omega_s = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{J_M})$.

For any $\boldsymbol{v}_j$ and $\boldsymbol{v}_{j'} \in \Omega_s$, we notice that

$$|Cov\left(\vartheta(\boldsymbol{v}_j), \vartheta(\boldsymbol{v}_{j'})\right)| = \begin{cases} 0, & \Delta_j \neq \Delta_{j'} \\ 1, & j = j' \\ \sigma_\beta^{-1/2}(\boldsymbol{v}_j)\sigma_\beta^{-1/2}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'}), & j \neq j', \Delta_j = \Delta_{j'}. \end{cases}$$

By definition of $\sigma_\beta(t,u)$, for any $v_j \in \Omega_s$, we have $\sigma_\beta(\boldsymbol{v}_j) = \mathrm{tr}\left\{\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(\boldsymbol{v}_j)\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right\}$, and therefore, $\lambda_{min}\left(\boldsymbol{\Sigma}\right)\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_j)\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right) \leq \sigma_\beta(\boldsymbol{v}_j) \leq \lambda_{max}\left(\boldsymbol{\Sigma}\right)\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_j)\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right)$, where $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ represent the minimum and maximum eigenvalues of the matrix. Then,

$$\begin{aligned}
&\sigma_\beta^{-1/2}(\boldsymbol{v}_j)\sigma_\beta^{-1/2}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'}) \\
=&\sigma_\beta^{-1/2}(\boldsymbol{v}_j)\sigma_\beta^{-1/2}(\boldsymbol{v}_{j'})\mathrm{tr}\left(\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\right) \\
=&\sigma_\beta^{-1/2}(\boldsymbol{v}_j)\sigma_\beta^{-1/2}(\boldsymbol{v}_{j'})\mathrm{tr}\left(\boldsymbol{\Sigma}\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right) \\
\leq&\sigma_\beta^{-1/2}(\boldsymbol{v}_j)\sigma_\beta^{-1/2}(\boldsymbol{v}_{j'})\lambda_{max}\left\{\boldsymbol{\Sigma}\right\}\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right) \\
\leq&\lambda_{min}\left\{\boldsymbol{\Sigma}\right\}^{-1}\lambda_{max}\left\{\boldsymbol{\Sigma}\right\} \\
&\times\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_j)\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right)^{-1/2}\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_{j'})\right)^{-1/2}\mathrm{tr}\left(\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\right) \\
=&\lambda_{min}\left\{\boldsymbol{\Sigma}\right\}^{-1}\lambda_{max}\left\{\boldsymbol{\Sigma}\right\}\frac{\tilde{\boldsymbol{B}}^{\mathrm{T}}(\boldsymbol{v}_j)\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})}{\|\tilde{\boldsymbol{B}}(\boldsymbol{v}_j)\|\|\tilde{\boldsymbol{B}}(\boldsymbol{v}_{j'})\|} \\
\leq&\lambda_{min}\left\{\boldsymbol{\Sigma}\right\}^{-1}\lambda_{max}\left\{\boldsymbol{\Sigma}\right\}.
\end{aligned}$$

Since this upper bound does not depend on the location of $\boldsymbol{v}_j$ and $\boldsymbol{v}_{j'}$, then there must exist constants c_1 and c_2 such that

$$\lambda_{min}\left\{\boldsymbol{\Sigma}\right\}^{-1}\lambda_{max}\left\{\boldsymbol{\Sigma}\right\} \leq c_1 c_2^{J_M} \leq c_1 c_2^{j-j'},$$

for any $1 \leq j, j' \leq J_M$. Now combined with Lemma 1 in (Ma and Yang, 2011), we can conclude that for any $a \in (0,1)$,

$$\lim_{n\to\infty} P\left\{\sup_j |\vartheta(\boldsymbol{v}_j)| \leq Q_\beta(a)\right\} = 1 - a,$$

where $Q_\beta(a) = (2\log J_M)^{1/2} - (2\log J_M)^{-1/2}\left\{\log(-0.5\log(1-a)) + 0.5\left[\log\right.\right.$ $\left.\left.(\log J_M) + \log 4\pi\right]\right\}$.

References

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics 213*(1), 4–29.

Csörgo, M. and P. Révész (2014). *Strong Approximations in Probability and Statistics*. Academic press.

De Boor, C., C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor (1978). *A Practical Guide to Splines*, Volume 27. New York: Springer.

El-Attar, R., M. Vidyasagar, and S. Dutta (1979). An algorithm for l_1-norm minimization with application to nonlinear l_1-approximation. *SIAM Journal on Numerical Analysis 16*(1), 70–86.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. New York: CRC Press.

Guan, X., H. Liu, J. You, and Y. Zhou (2021). Estimation and inference for dynamic single-index varying-coefficient models. *Statistica Sinica*. DOI:10.5705/ss.202019.0467.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 109–126.

Hall, P. and M. Hosseini-Nasab (2009). Theory for high-order bounds in functional principal components analysis. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 146, pp. 225. Cambridge University Press.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics 40*(6), 3108–3136.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *The Annals of statistics 26*(2), 755–770.

Lai, M.-J. and L. L. Schumaker (2007). *Spline Functions on Triangulations*. Cambridge: Cambridge University Press.

Li, M., K. Wang, A. Maity, and A.-M. Staicu (2022). Inference in functional linear quantile regression. *Journal of Multivariate Analysis 190*, 104985.

Ma, S. (2016). Estimation and inference in functional single-index models. *Annals of the Institute of Statistical Mathematics 68*(1), 181–208.

Ma, S. and L. Yang (2011). A jump-detecting procedure based on spline estimation. *Journal of Nonparametric Statistics 23*(1), 67–81.

REFERENCES

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Ramsay, J. O., G. Hooker, and S. Graves (2009). *Functional Data Analysis with R and MATLAB*. New York: Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (Second ed.). New York: Springer.

Wang, Y., G. Wang, L. Wang, and R. T. Ogden (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics 76*, 427–437.