SUPPLEMENTARY MATERIAL FOR "SCALABLE ESTIMATION AND INFERENCE FOR CENSORED QUANTILE REGRESSION PROCESS"

BY XUMING HE[1], XIAOOU PAN[2,†] KEAN MING TAN[1,*]
AND WEN-XIN ZHOU[2,‡]

[1]Department of Statistics, University of Michigan, xmhe@umich.edu; [*]keanming@umich.edu

[2]Department of Mathematics, University of California, San Diego, [†]xip024@ucsd.edu; [‡]wez243@ucsd.edu

APPENDIX A: OPTIMIZATION ALGORITHMS

A.1. Low-dimensional setting. Solving the smoothed estimating equations (3) and (4) are equivalent to minimizing the convex loss functions $\widehat{L}_k(\cdot)$ $(k=0,1,\ldots,m)$ given in (6) for which standard quasi-Newton methods can be applied. Here we use the Barzilai and Borwein (BB) method [1], which is a well-known technique in optimization that approximates the secant equation by two consecutive points to find a nearly Newton step size. See Section 3 of [6] for an application of the BB method to non-censored smoothed quantile regression.

Note that each $\widehat{L}_k(\cdot)$ for $k \geq 1$ is a shifted version of $\widehat{L}_0(\cdot)$, so that for any $\boldsymbol{\beta}, \boldsymbol{\beta}' \in \mathbb{R}^p$,

$$\nabla \widehat{L}_k(\boldsymbol{\beta}) - \nabla \widehat{L}_k(\boldsymbol{\beta}') = \nabla \widehat{L}_0(\boldsymbol{\beta}) - \nabla \widehat{L}_0(\boldsymbol{\beta}'), \quad k=1,\ldots,m.$$

At the kth quantile level and given the previous estimates $\widehat{\boldsymbol{\beta}}_0, \ldots, \widehat{\boldsymbol{\beta}}_{k-1}$, the gradient of $\widehat{L}_k(\cdot)$ is given by

$$\nabla \widehat{L}_k(\boldsymbol{\beta}) = \nabla \widehat{L}_0(\boldsymbol{\beta}) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{k-1} \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_j)\{H(\tau_{j+1}) - H(\tau_j)\} \boldsymbol{x}_i.$$

With the above preparations, we summarize the BB method for solving $\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \widehat{L}_k(\boldsymbol{\beta})$ in Algorithm 1.

Algorithm 1 Barzilai-Borwein gradient descent method for minimizing $\widehat{L}_k(\cdot)$.

Input: Censored data $\{(y_i, \boldsymbol{x}_i, \Delta_i)\}_{i=1}^n$, current quantile level $\tau_k \in (0,1)$, previous estimates $\{\widehat{\boldsymbol{\beta}}_j\}_{j=0}^{k-1}$, initial values $\widehat{\boldsymbol{\beta}}_k^{(0)} = \widehat{\boldsymbol{\beta}}_{k-1}$, bandwidth h, step size upper bound $\alpha_{\max}$, tolerance level ϵ.

1: Compute $\widehat{\boldsymbol{\beta}}_k^{(1)} \leftarrow \widehat{\boldsymbol{\beta}}_k^{(0)} - \nabla \widehat{L}_k(\widehat{\boldsymbol{\beta}}_k^{(0)})$
2: **for** $t = 1, 2 \ldots$ **do**
3: $\boldsymbol{s}_t \leftarrow \widehat{\boldsymbol{\beta}}_k^{(t)} - \widehat{\boldsymbol{\beta}}_k^{(t-1)}$, $\boldsymbol{g}_t \leftarrow \nabla \widehat{L}_k(\widehat{\boldsymbol{\beta}}_k^{(t)}) - \nabla \widehat{L}_k(\widehat{\boldsymbol{\beta}}_k^{(t-1)}) = \nabla \widehat{L}_0(\widehat{\boldsymbol{\beta}}_k^{(t)}) - \nabla \widehat{L}_0(\widehat{\boldsymbol{\beta}}_k^{(t-1)})$
4: $\alpha_{t,1} \leftarrow ||\boldsymbol{s}_t||_2^2 / \langle \boldsymbol{s}_t, \boldsymbol{g}_t \rangle$, $\alpha_{t,2} \leftarrow \langle \boldsymbol{s}_t, \boldsymbol{g}_t \rangle / ||\boldsymbol{g}_t||_2^2$
5: $\alpha_t \leftarrow \min\{\alpha_{t,1}, \alpha_{t,2}, \alpha_{\max}\}$ if $\langle \boldsymbol{s}_t, \boldsymbol{g}_t \rangle > 0$, and $\alpha_t \leftarrow 1$ otherwise
6: Update $\widehat{\boldsymbol{\beta}}_k^{(t+1)} \leftarrow \widehat{\boldsymbol{\beta}}_k^{(t)} - \alpha_t \nabla \widehat{L}_k(\widehat{\boldsymbol{\beta}}_k^{(t)})$
7: **end for** when $\|\nabla \widehat{L}_k(\widehat{\boldsymbol{\beta}}_k^{(t)})\|_2 \leq \epsilon$

A.2. High-dimensional setting. In the high dimensional regime, we need to solve the following weighted ℓ_1-penalized programs sequentially:

$$\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\beta}}(\tau_k) \in \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \big\{ \widehat{L}_k(\boldsymbol{\beta}) + \|\boldsymbol{\lambda}_k \circ \boldsymbol{\beta}\|_1 \big\}, \quad k=0,\ldots,m, \tag{A.1}$$

where $\circ$ denotes the Hadamard product, and $\boldsymbol{\lambda}_k = (\lambda_{k,1}, \ldots, \lambda_{k,p})^{\mathrm{T}}$ may depend on the previous estimates $\{\widehat{\boldsymbol{\beta}}_j\}_{j=0}^{k-1}$. To this end, we apply the iterative local adaptive majorize-minimize (I-LAMM) algorithm proposed in [5].

To illustrate the basic ideas, consider the general problem of minimizing a nonlinear function $f(\cdot)$ on $\mathbb{R}^p$. Starting at a given point $\boldsymbol{\beta}_0$, the majorize-minimize (MM) algorithm involves two steps: first, construct a majorizing function $g(\cdot \,|\, \boldsymbol{\beta}_0)$, satisfying

$$g(\boldsymbol{\beta}_0 \,|\, \boldsymbol{\beta}_0) = f(\boldsymbol{\beta}_0) \ \text{ and } \ \underbrace{g(\boldsymbol{\beta} \,|\, \boldsymbol{\beta}_0) \geq f(\boldsymbol{\beta}) \ \text{ for any } \boldsymbol{\beta} \in \mathbb{R}^p}_{\text{global majorization property}};$$

secondly, update $\boldsymbol{\beta}_0$ by $\boldsymbol{\beta}_1 := \mathrm{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \, g(\boldsymbol{\beta} \,|\, \boldsymbol{\beta}_0)$ [8]. Noting that

$$f(\boldsymbol{\beta}_1) \overset{\text{majorization}}{\leq} g(\boldsymbol{\beta}_1 \,|\, \boldsymbol{\beta}_0) \overset{\text{minimization}}{\leq} g(\boldsymbol{\beta}_0 \,|\, \boldsymbol{\beta}_0) = f(\boldsymbol{\beta}_0),$$

the objective value of such an algorithm is non-increasing in each step. In fact, the global majorization property is not necessary to ensure non-increasing objective values. Instead, we only need the following local majorization property

$$g(\boldsymbol{\beta}_0 \,|\, \boldsymbol{\beta}_0) = f(\boldsymbol{\beta}_0) \ \text{ and } \ g(\boldsymbol{\beta}_1 \,|\, \boldsymbol{\beta}_0) \geq f(\boldsymbol{\beta}_1).$$

To construct a proper majorizing function for $\widehat{L}_k(\cdot)$ around $\boldsymbol{\beta}_0$, we define an isotropic quadratic function

$$F(\boldsymbol{\beta}; \phi, \boldsymbol{\beta}_0) := \widehat{L}_k(\boldsymbol{\beta}_0) + \langle \nabla \widehat{L}_k(\boldsymbol{\beta}_0), \boldsymbol{\beta} - \boldsymbol{\beta}_0 \rangle + \frac{\phi}{2} \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2$$

for some $\phi > 0$. It is easy to see that $F(\boldsymbol{\beta}_0; \phi, \boldsymbol{\beta}_0) = \widehat{L}_k(\boldsymbol{\beta}_0)$. Using such a surrogate loss function, the weighted ℓ_1-penalized program $\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \{F(\boldsymbol{\beta}; \phi, \boldsymbol{\beta}_0) + \|\boldsymbol{\lambda}_k \circ \boldsymbol{\beta}\|_1\}$ admits a closed-form solution $\boldsymbol{\beta}_1 = S_{\text{soft}}(\boldsymbol{\beta}_0 - \nabla \widehat{L}_k(\boldsymbol{\beta}_0)/\phi, \, \boldsymbol{\lambda}_k/\phi)$, where $S_{\text{soft}}(\boldsymbol{\beta}, \boldsymbol{\lambda}) := (\mathrm{sign}(\beta_j) \max\{|\beta_j| - \lambda_{k,j}, 0\})_{j=1,\ldots,p}$ is the soft-thresholding operator. Moreover, the quadratic coefficient $\phi > 0$ should be sufficiently large so that the local majorization property $F(\boldsymbol{\beta}_1; \phi, \boldsymbol{\beta}_0) \geq \widehat{L}_k(\boldsymbol{\beta}_1)$ is satisfied. Starting with a relatively small value $\phi = \phi_0$, we iteratively increase ϕ by a factor of $\gamma > 1$ and compute

$$\boldsymbol{\beta}_{1,\ell} = S_{\text{soft}}(\boldsymbol{\beta}_0 - \nabla \widehat{L}_k(\boldsymbol{\beta}_0)/\phi_\ell, \, \boldsymbol{\lambda}_k/\phi_\ell) \ \text{ with } \ \phi_\ell = \gamma^\ell \phi_0, \ \ \ell = 0, 1, \ldots.$$

until the local majorization property holds. Repeating this procedure yields a sequence of iterates $\{\boldsymbol{\beta}_t\}_{t=0,1,\ldots}$ until the stopping criterion is met, say $\|\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_t\|_2 \leq \epsilon$. We treat ϕ_0, γ, ϵ as internal optimization parameters; the default choice is $(\phi_0, \gamma, \epsilon) = (0.5, 1.5, 10^{-5})$.

APPENDIX B: PROOFS OF THE MAIN RESULTS IN SECTION 3.2

To begin with, we revisit and define some notations that will be frequently used. For predetermined grid points $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$, we write $\boldsymbol{\beta}_j^* = \boldsymbol{\beta}^*(\tau_j)$ and $\widehat{\boldsymbol{\beta}}_j = \widehat{\boldsymbol{\beta}}(\tau_j)$, $j = 0, \ldots, m$. Since the estimators $\{\widehat{\boldsymbol{\beta}}_j\}_{j=0}^m$ are constructed sequentially, the statistical error of $\widehat{\boldsymbol{\beta}}_j$ at quantile level τ_j depends on the accumulated errors of $\widehat{\boldsymbol{\beta}}_0, \ldots, \widehat{\boldsymbol{\beta}}_{j-1}$.

For every $r > 0$, define the local ellipse $\Theta(r) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_\Sigma \leq r\}$ under the Σ-induced norm. Under Condition 3.2 on the (random) feature vector $\boldsymbol{x} \in \mathbb{R}^p$, for every $\delta \in (0, 1]$ we set

$$\eta_\delta = \inf\big\{\eta > 0 : \mathbb{E}\big\{(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v})^2 \mathbb{1}(|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}| > \eta)\big\} \leq \delta \ \text{ for all } \boldsymbol{v} \in \mathbb{S}^{p-1}\big\}, \tag{B.1}$$

where $\boldsymbol{z} = \Sigma^{-1/2}\boldsymbol{x}$. Since $\mathbb{E}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v})^2 = 1$ for any $\boldsymbol{v} \in \mathbb{S}^{p-1}$, η_δ is well-defined for each δ, and depends implicitly on the underlying distribution of $\boldsymbol{z}$. Throughout the proof, we write

$$\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ip})^{\mathrm{T}} = \Sigma^{-1/2}\boldsymbol{x}_i, \ i = 1, \ldots, n.$$

For a non-negative kernel function $K(\cdot)$ and a bandwidth $h>0$, we write

$$K_h(u)=h^{-1}K(u/h),\ \ \bar{K}_h(u)=\bar{K}(u/h) \quad \text{and} \quad \bar{K}(u)=\int_{-\infty}^{u}K(t)\,\mathrm{d}t.$$

B.1. Technical lemmas. We first collect several technical lemmas that serve as the building blocks for proving the main results.

LEMMA B.1 (Convexity lemma). *For a vector-valued function $Q(\boldsymbol{\beta}):\mathbb{R}^p\to\mathbb{R}^p$ with positive semi-definite Jacobian, define the corresponding divergence function $D=D_Q:\mathbb{R}^p\times\mathbb{R}^p\to[0,\infty)$ as $D(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\langle Q(\boldsymbol{\beta}_1)-Q(\boldsymbol{\beta}_2),\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\rangle$. For any $\boldsymbol{\beta},\boldsymbol{\beta}'\in\mathbb{R}^p$ and $\eta\in[0,1]$, we have*

$$D(\boldsymbol{\beta}'+\eta(\boldsymbol{\beta}-\boldsymbol{\beta}'),\boldsymbol{\beta}')\le\eta D(\boldsymbol{\beta},\boldsymbol{\beta}').$$

Lemma B.2 below provides a Bernstein-type inequality for the ℓ_2-norm of a sum of centered random vectors, which will be frequently used to bound the smoothed estimating functions.

LEMMA B.2. *Assume Condition 3.2 holds, and let $\{\xi_i\}_{i=1}^n$ be independent random variables satisfying $\mathbb{E}(\xi_i^2|\boldsymbol{x}_i)\le\sigma^2$ and $|\xi_i|\le M$ for some $M\ge\sigma>0$. Then, for any $t>0$,*

$$\left\|\frac{1}{n}\sum_{i=1}^n(\xi_i\boldsymbol{z}_i-\mathbb{E}\,\xi_i\boldsymbol{z}_i)\right\|_2\le 2\sigma\sqrt{\frac{p}{n}}+\sigma\sqrt{\frac{2t}{n}}+M\zeta_p\frac{4t}{3n}$$

holds with probability at least $1-e^{-t}$, where $\zeta_p=\max_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{x}\|_{\Sigma^{-1}}$.

Lemma B.3 provides concentration inequalities for some of the stochastic terms in the estimating functions $\widehat{Q}_0(\cdot)$ and $\widehat{Q}_j(\cdot)$ $(j\ge1)$.

LEMMA B.3. *Let $j=0,1,\ldots,m$ and $t>0$.*

(i) *With probability at least $1-e^{-t}$,*

$$||\widehat{Q}_0(\boldsymbol{\beta}_j^*)-\mathbb{E}\widehat{Q}_0(\boldsymbol{\beta}_j^*)||_{\Sigma^{-1}}\le 2\bar{\tau}_0\left(\sqrt{\frac{p}{n}}++\sqrt{\frac{t}{2n}}+\zeta_p\frac{2t}{3n}\right),$$

where $\bar{\tau}_0=\max(\tau_0,1-\tau_0)$.

(ii) *With probability at least $1-e^{-t}$,*

$$\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\boldsymbol{x}_i\right\|_{\Sigma^{-1}}\le 2\left(\sqrt{\frac{p}{n}}+\sqrt{\frac{t}{2n}}+\zeta_p\frac{2t}{3n}\right).$$

(iii) *With probability at least $1-e^{-t}$,*

$$\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_j}^{\tau_{j+1}}\left\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_j))\right\}\mathrm{d}H(u)\cdot\boldsymbol{x}_i\right\|_{\Sigma^{-1}}$$

$$\lesssim m_4^{1/2}w_j\cdot\delta^*\sqrt{\frac{p+t}{nh}}+w_j\cdot\zeta_p^2\delta^*\frac{t}{nh},$$

where $w_j=H(\tau_{j+1})-H(\tau_j)=\log(\frac{1-\tau_j}{1-\tau_{j+1}})$.

The following lemma concerns the first-order property of the smoothed estimating functions $\widehat{Q}_j(\cdot)$ in (3) and (4). Define the corresponding symmetrized Bregman divergence $D:\mathbb{R}^p\times\mathbb{R}^p\to[0,\infty)$ as

(B.2) $$D(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2):=\langle\widehat{Q}_j(\boldsymbol{\beta}_1)-\widehat{Q}_j(\boldsymbol{\beta}_2),\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\rangle=\langle\widehat{Q}_0(\boldsymbol{\beta}_1)-\widehat{Q}_0(\boldsymbol{\beta}_2),\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\rangle.$$

Note that the divergence D is independent of j.

LEMMA B.4 (Restricted strong convexity). *Assume Conditions 3.1–3.3 hold, and let $h,r>0$ satisfy $4\eta_{1/4}r\le h\le 1$ with $\eta_{1/4}$ defined in (B.1). Then, for any $0\le j\le m$ and $t>0$,*

$$\inf_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\frac{D(\boldsymbol{\beta},\boldsymbol{\beta}_j^*)}{\kappa_l\,||\boldsymbol{\beta}-\boldsymbol{\beta}_j^*||_\Sigma^2}\ge\frac{3}{4}\underline{g}-\bar{g}^{1/2}r^{-1}\left(\frac{5}{4}\sqrt{\frac{hp}{n}}+\sqrt{\frac{ht}{8n}}\right)-cr^{-2}\frac{ht}{n}$$

with probability at least $1-e^{-t}$, where $c=1/4+1/48\approx0.27$.

For $j=0,1,\ldots$ and $r>0$, define

(B.3)
$$\varpi_j(r)=\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\left\|\frac{1}{n}\sum_{i=1}^n\big\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\big\}\boldsymbol{x}_i+\mathbf{H}_j(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)\right\|_{\Sigma^{-1}},$$

(B.4)
$$\omega_j(r)=\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\left\|\frac{1}{n}\sum_{i=1}^n\Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i)-\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*-y_i)\big\}\boldsymbol{x}_i-\mathbf{J}_j(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)\right\|_{\Sigma^{-1}},$$

where $\mathbf{J}_j=\mathbf{J}(\tau_j)=\mathbb{E}\{g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}$ and $\mathbf{H}_j=\mathbf{H}(\tau_j)=\mathbb{E}\{f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}$. The following lemma provides upper bounds for the two suprema $\varpi_j(r)$ and $\omega_j(r)$ for any given $r>0$.

LEMMA B.5. *Assume Conditions 3.1–3.3 hold, and write $m_k=\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{u}|^k$ ($k=3,4$). Let $j=0,1,\ldots,m$ and $r>0$.*

(i) *With probability at least $1-e^{-t}$,*

(B.5) $$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\big\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\big\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}}\lesssim(\kappa_u\bar{f}m_4)^{1/2}\sqrt{\frac{p+t}{nh}}\cdot r,$$

provided that the "effective" sample size satisfies $nh\gtrsim\zeta_p^2(p+t)$. Moreover, for any $\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)$,

$$\left\|\mathbb{E}\big\{\bar{K}_h(y-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta})-\bar{K}_h(y-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*)\big\}\boldsymbol{x}+\mathbf{H}_j(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)\right\|_{\Sigma^{-1}}\le l_1\big(0.5m_3r+\kappa_1h\big)\cdot r.$$

(ii) *With probability at least $1-e^{-t}$,*

(B.6) $$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i)-\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*-y_i)\big\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}}\lesssim(\kappa_u\bar{g}m_4)^{1/2}\sqrt{\frac{p+t}{nh}}\cdot r,$$

provided that $nh \gtrsim \zeta_p^2(p+t)$. Moreover, for any $\boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r)$,

$$\left\|\mathbb{E}\Delta_i\big\{\bar{K}_h(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} - y) - \bar{K}_h(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^* - y)\big\}\boldsymbol{x} - \mathbf{J}_j(\boldsymbol{\beta} - \boldsymbol{\beta}_j^*)\right\|_{\Sigma^{-1}} \le l_1\big(0.5m_3 r + \kappa_1 h\big)\cdot r.$$

LEMMA B.6. *Assume Conditions 3.1–3.3 hold. For any $\tau_L \le \tau_l < \tau_u \le \tau_U$,*

$$\sup_{\tau\in[\tau_l,\tau_u]}\left\|\frac{1}{n}\sum_{i=1}^{n}(1-\mathbb{E})\int_{\tau_l}^{\tau}\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{x}_i\right\|_{\Sigma^{-1}}$$
$$\lesssim \frac{\tau_u-\tau_l}{1-\tau_u}\left(\sqrt{\frac{p+\log n+t}{n}} + \zeta_p\frac{\log n + t}{n}\right)$$

holds with probability at least $1-e^{-t}$.

Moreover, Lemma B.7 provides upper bounds on the approximation error (see (B.12) in the proof of Theorem 3.1), which consists of the smoothing and discretization errors. Let $Q_0(\boldsymbol{\beta}) = \mathbb{E}\widehat{Q}_0(\boldsymbol{\beta})$.

LEMMA B.7. *For each $j = 1,\dots,m$,*

$$\left\|Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}\}\right\|_{\Sigma^{-1}} \tag{B.7}$$
$$\le \frac{1}{2}l_1\kappa_2 h^2\big\{1 + H(\tau_j) - H(\tau_0)\big\} + (\overline{f}/\underline{f})\sum_{\ell=0}^{j-1} w_\ell(\tau_{\ell+1}-\tau_\ell),$$

where $w_\ell = H(\tau_{\ell+1}) - H(\tau_\ell)$. In particular, $\|Q_0(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} \le 0.5 l_1\kappa_2 h^2$.

The next lemma establishes the asymptotic uniform equicontinuity of the centered process $\mathbb{G}_n(\cdot)$ in $\ell^\infty([\tau_L,\tau_U])$. This is an equivalent definition to asymptotic tightness, and is an important step towards the weak convergence stated in Theorem 3.3.

LEMMA B.8 (Asymptotic uniform equicontinuity). *Assume the conditions of Theorem 3.3 hold, and let $\{\boldsymbol{a}_n\}_{n\ge1}$ be a normalized sequence such that $\|\boldsymbol{a}_n\|_\Sigma = 1$. Then, for any $x > 0$,*

$$\lim_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{|\tau_1-\tau_2|<\delta}|\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2)| > x\right\} = 0,$$

where $\mathbb{G}_n(\cdot)$ is defined in (20).

B.2. Proof of Theorem 3.1. We first prove a uniform upper bound over the grid of τ-values—$\{\tau_0,\tau_1,\dots,\tau_m\}$. That is, with probability at least $1-C_1n^{-1}$,

$$\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma = \|\widehat{\boldsymbol{\beta}}(\tau_j) - \boldsymbol{\beta}^*(\tau_j)\|_\Sigma \le r_j,\ \ j = 0,1,\dots,m \tag{B.8}$$

for a sequence of radii $\{r_j\}_{j=0,1,\dots,m}$ and some absolute constant $C_1 > 0$. To begin with, define a crude (sub-optimal) convergence radius $r^\diamondsuit = h/(4\eta_{1/4})$ with $\eta_{1/4}$ given in (B.1). Accordingly, define "intermediate" points $\widetilde{\boldsymbol{\beta}}_j = (1-u_j)\boldsymbol{\beta}_j^* + u_j\widehat{\boldsymbol{\beta}}_j$, where

$$u_j = \sup\big\{u\in[0,1] : u(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)\in\Theta(r^\diamondsuit)\big\}\begin{cases} = 1 & \text{if } \widehat{\boldsymbol{\beta}}_j \in \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit),\\ \in(0,1) & \text{if } \widehat{\boldsymbol{\beta}}_j \notin \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit).\end{cases}$$

It is easy to see that $\widetilde{\boldsymbol\beta}_j=\widehat{\boldsymbol\beta}_j$ if $\widehat{\boldsymbol\beta}_j\in\boldsymbol\beta_j^*+\Theta(r^\diamondsuit)$, and $\widetilde{\boldsymbol\beta}_j\in\boldsymbol\beta_j^*+\partial\Theta(r^\diamondsuit)$ if $\widehat{\boldsymbol\beta}_j\notin\boldsymbol\beta_j^*+\Theta(r_j)$. Here $\partial\Theta(r)$ denotes the boundary of $\Theta(r)$. In either case, we have $\widetilde{\boldsymbol\beta}_j\in\boldsymbol\beta_j^*+\Theta(r^\diamondsuit)$.

Recall the symmetrized Bregman divergence $D(\boldsymbol\beta_1,\boldsymbol\beta_2)=\langle\widehat Q_j(\boldsymbol\beta_1)-\widehat Q_j(\boldsymbol\beta_2),\boldsymbol\beta_1-\boldsymbol\beta_2\rangle$ defined in (B.2). Applying Lemma B.1 yields that, for each $j=0,1,\ldots,m$,

$$D(\widetilde{\boldsymbol\beta}_j,\boldsymbol\beta_j^*)\le u_j\cdot D(\widehat{\boldsymbol\beta}_j,\boldsymbol\beta_j^*)=u_j\cdot\langle\widehat Q_j(\widehat{\boldsymbol\beta}_j)-\widehat Q_j(\boldsymbol\beta_j^*),\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j^*\rangle.$$

Since $\widehat{\boldsymbol\beta}_j$ solves the estimating equation $\widehat Q_j(\widehat{\boldsymbol\beta}_j)=\mathbf 0$, by the Cauchy–Schwartz inequality we have

$$D(\widetilde{\boldsymbol\beta}_j,\boldsymbol\beta_j^*)\le u_j\cdot\langle-\widehat Q_j(\boldsymbol\beta_j^*),\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j^*\rangle\le\|\widehat Q_j(\boldsymbol\beta_j^*)\|_{\Sigma^{-1}}\cdot\|\widetilde{\boldsymbol\beta}_j-\boldsymbol\beta_j^*\|_\Sigma.$$

For some curvature parameter $\kappa>0$ to be specified, define the event

$$\mathcal F=\bigcap_{j=0}^m\left\{D(\boldsymbol\beta,\boldsymbol\beta_j^*)\ge\kappa\cdot\|\boldsymbol\beta-\boldsymbol\beta_j^*\|_\Sigma^2\ \text{ for all }\ \boldsymbol\beta\in\boldsymbol\beta_j^*+\Theta(r^\diamondsuit)\right\}. \tag{B.9}$$

Conditioning on $\mathcal F$, it follows that

$$\|\widetilde{\boldsymbol\beta}_j-\boldsymbol\beta_j^*\|_\Sigma\le\kappa^{-1}\|\widehat Q_j(\boldsymbol\beta_j^*)\|_{\Sigma^{-1}},\ \ j=0,1,\ldots,m. \tag{B.10}$$

Next we derive upper bounds for $\{\|\widehat Q_j(\boldsymbol\beta_j^*)\|_{\Sigma^{-1}}\}_{j=0,1,\ldots,m}$ sequentially. For each j, we decompose $\widehat Q_j(\boldsymbol\beta_j^*)$ as

$$\widehat Q_j(\boldsymbol\beta_j^*)=\widehat Q_0(\boldsymbol\beta_j^*)-Q_0(\boldsymbol\beta_j^*)-\sum_{\ell=0}^{j-1}w_\ell(\widehat\Delta_\ell+\Delta_\ell)+Q_0(\boldsymbol\beta_j^*)-\sum_{\ell=0}^{j-1}w_\ell\mathbb E\{\bar K_h(y-\boldsymbol x^{\mathrm T}\boldsymbol\beta_\ell^*)\boldsymbol x\}$$

where $Q_0(\boldsymbol\beta)=\mathbb E\widehat Q_0(\boldsymbol\beta)$, $w_\ell=H(\tau_{\ell+1})-H(\tau_\ell)$, and

$$\begin{cases}\widehat\Delta_\ell=\frac1n\sum_{i=1}^n\left\{\bar K_h(y_i-\boldsymbol x_i^{\mathrm T}\widehat{\boldsymbol\beta}_\ell)-\bar K_h(y_i-\boldsymbol x_i^{\mathrm T}\boldsymbol\beta_\ell^*)\right\}\boldsymbol x_i,\\ \Delta_\ell=\frac1n\sum_{i=1}^n(1-\mathbb E)\bar K_h(y_i-\boldsymbol x_i^{\mathrm T}\boldsymbol\beta_\ell^*)\boldsymbol x_i.\end{cases} \tag{B.11}$$

By the triangle inequality,

(B.12)

$$\|\widehat Q_j(\boldsymbol\beta_j^*)\|_{\Sigma^{-1}}\le\underbrace{\|\widehat Q_0(\boldsymbol\beta_j^*)-Q_0(\boldsymbol\beta_j^*)\|_{\Sigma^{-1}}}_{\text{statistical error of the } j\text{th estimating equation}}+\underbrace{\sum_{\ell=0}^{j-1}w_\ell\big(\|\widehat\Delta_\ell\|_{\Sigma^{-1}}+\|\Delta_\ell\|_{\Sigma^{-1}}\big)}_{\text{accumulated error}}$$
$$+\underbrace{\left\|Q_0(\boldsymbol\beta_j^*)-\sum_{\ell=0}^{j-1}w_\ell\mathbb E\{\bar K_h(y-\boldsymbol x^{\mathrm T}\boldsymbol\beta_\ell^*)\boldsymbol x\}\right\|_{\Sigma^{-1}}}_{\text{approximation error}},\ \ j=1,\ldots,m.$$

In particular, $\|\widehat Q_0(\boldsymbol\beta_0^*)\|_{\Sigma^{-1}}\le\|\widehat Q_0(\boldsymbol\beta_0^*)-Q_0(\boldsymbol\beta_0^*)\|_{\Sigma^{-1}}+\|Q_0(\boldsymbol\beta_0^*)\|_{\Sigma^{-1}}$. For the approximation error term in (B.12), by Lemma B.7 we have

$$\left\|Q_0(\boldsymbol\beta_j^*)-\sum_{\ell=0}^{j-1}w_\ell\mathbb E\{\bar K_h(y-\boldsymbol x^{\mathrm T}\boldsymbol\beta_\ell^*)\boldsymbol x\}\right\|_{\Sigma^{-1}} \tag{B.13}$$
$$\le\frac12 l_1\kappa_2h^2+\sum_{\ell=0}^{j-1}w_ja<a+\sum_{\ell=0}^{j-1}w_ja\quad\text{with}\quad a:=\frac12 l_1\kappa_2h^2+\overline f\underline f^{-1}\delta^*.$$

For some $\delta > 0$ to be determined, define the second event

$$\mathcal{G} = \left\{ \max_{0 \le j \le m} \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} \bigvee \max_{0 \le \ell \le m-1} \|\Delta_\ell\|_{\Sigma^{-1}} \le \delta \right\}, \tag{B.14}$$

where the Δ_ℓ's are given in (B.11). Conditioned on $\mathcal{F} \cap \mathcal{G}$, it follows from (B.12)–(B.14) that

$$\|\widehat{Q}_0(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} < \delta + a, \quad \|\widehat{Q}_j(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} < \delta + a + \sum_{\ell=0}^{j-1} w_\ell\big(\delta + a + \|\widehat{\Delta}_\ell\|_{\Sigma^{-1}}\big), \;\; j = 1, \ldots, m.$$

Based on the above general bounds, we iteratively control $\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma$ and $\|\widehat{\Delta}_j\|_{\Sigma^{-1}}$, starting at $j = 0$. By (B.10),

$$\|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\|_\Sigma \le \kappa^{-1}\|\widehat{Q}_0(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} < r_0 := \kappa^{-1}(\delta + a).$$

Provided $r_0 \le r^\diamondsuit$, the intermediate point $\widetilde{\boldsymbol{\beta}}_0$ falls into the interior of the local region $\boldsymbol{\beta}_0^* + \Theta(r^\diamondsuit)$. Via proof by contradiction, we must have $\widehat{\boldsymbol{\beta}}_0 = \widetilde{\boldsymbol{\beta}}_0$ and hence $\widehat{\boldsymbol{\beta}}_0 \in \boldsymbol{\beta}_0^* + \Theta(r_0)$.

Turning to $(\widetilde{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_1)$, it follows from (B.10) with $j = 1$ that $\|\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*\|_\Sigma < \kappa^{-1}\{\delta + a + w_0(\|\widehat{\Delta}_0\|_{\Sigma^{-1}} + \delta + a)\}$. Note that $\widehat{\Delta}_0$ depends on the preceding estimate $\widehat{\boldsymbol{\beta}}_0$. Since, as proved in the last step, $\widehat{\boldsymbol{\beta}}_0 \in \boldsymbol{\beta}_0^* + \Theta(r_0)$ conditioned on $\mathcal{F} \cap \mathcal{G}$, it follows that

$$\|\widehat{\Delta}_0\|_{\Sigma^{-1}} \le \varpi_0(r_0) + \|\mathbf{H}(\tau_0)(\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} \le \varpi_0(r_0) + \overline{f} r_0,$$

where $\varpi_0(\cdot)$ is defined in (B.3). Conditioned further on $\{\varpi_0(r_0) \le \overline{f} r_0\}$, this implies

$$\|\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*\|_\Sigma < r_1 := \kappa^{-1}\big\{\delta + a + w_0(2\overline{f} r_0 + \delta + a)\big\}.$$

As long as $r_1 \le r^\diamondsuit$, $\widetilde{\boldsymbol{\beta}}_1$ lies in the interior of $\boldsymbol{\beta}_1^* + \Theta(r^\diamondsuit)$, which enforces $\widehat{\boldsymbol{\beta}}_1 = \widetilde{\boldsymbol{\beta}}_1$ and hence $\widehat{\boldsymbol{\beta}}_1 \in \boldsymbol{\beta}_1^* + \Theta(r_1)$.

Applying the above argument repeatedly, at the j-th step $(1 \le j \le m)$, we obtain that conditioned on $\{\varpi_{j-1}(r_{j-1}) \le \overline{f} r_{j-1}\}$,

$$\|\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma < \frac{1}{\kappa}\left\{\delta + a + \sum_{\ell=0}^{j-1} w_j\big(\|\widehat{\Delta}_\ell\|_{\Sigma^{-1}} + \delta + a\big)\right\}$$

$$\le \frac{1}{\kappa}\left\{\delta + a + \sum_{\ell=0}^{j-1} w_\ell(2\overline{f} r_\ell + \delta + a)\right\} =: r_j. \tag{B.15}$$

Provided $r_j \le r^\diamondsuit$, by way of contradiction we must have $\widehat{\boldsymbol{\beta}}_j = \widetilde{\boldsymbol{\beta}}_j \in \boldsymbol{\beta}_j^* + \Theta(r_j)$. Equivalently, the above sequence of radii $\{r_j\}_{j=0}^m$ can be recursively defined as

$$r_j = \big(1 + 2\kappa^{-1}\overline{f} w_{j-1}\big) r_{j-1} + \kappa^{-1} w_{j-1}(\delta + a), \;\; j = 1, \ldots, m, \;\; \text{and} \;\; r_0 = \kappa^{-1}(\delta + a),$$

where a is given in (B.13). Taking $C = \kappa^{-1}(2\overline{f} + 1)$, it follows that

$$r_j \le (1 + C w_{j-1}) r_{j-1} \le \cdots \le \prod_{\ell=0}^{j-1}(1 + C w_\ell) \cdot r_0 \le \exp\left(C \sum_{\ell=0}^{j-1} w_\ell\right) \cdot r_0 = \left(\frac{1 - \tau_L}{1 - \tau_j}\right)^C \cdot r_0. \tag{B.16}$$

Thus far we have established the result $\widehat{\boldsymbol{\beta}}_j \in \boldsymbol{\beta}_j^* + \Theta(r_j)$ $(j = 0, 1, \ldots, m)$ as a deterministic claim, but conditioned on the "good" event

$$\mathcal{F} \cap \mathcal{G} \cap \bigcap_{\ell=0}^{m-1} \{\varpi_\ell(r_\ell) \le \overline{f} r_\ell\}$$

with properly chosen κ, δ and $\{r_j\}_{j=0}^m$. By Lemmas B.4 and B.3, we choose $\kappa = (\underline{g}\kappa_l)/2$ and $\delta \asymp \sqrt{(p+\log n)/n} + \zeta_p \log(n)/n$, so that $\mathbb{P}(\mathcal{F}^c) \le (m+1)/n^2$ and $\mathbb{P}(\mathcal{G}^c) \le 2(m+1)/n^2$ as long as $nh \gtrsim p + \log n$. With this choice of δ, and since $a = 0.5 l_1 \kappa_2 h^2 + (\overline{f}/\underline{f})\delta^* \lesssim h^2 + n^{-1/2}$, we obtain from (B.16) that

$$r_j \le \left(\frac{1-\tau_L}{1-\tau_j}\right)^C \cdot r_0 \asymp \left(\frac{1-\tau_L}{1-\tau_j}\right)^C \underline{g}^{-1}\left(\sqrt{\frac{p+\log n}{n}} + \zeta_p \frac{\log n}{n} + h^2\right).$$

Moreover, it follows from Lemma B.5 that with probability at least $1 - m/n^2$,

$$\varpi_\ell(r_\ell) \lesssim \left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r_\ell + h\right)\cdot r_\ell \ \text{ for all } \ \ell = 0,1,\ldots,m-1,$$

provided $nh \gtrsim \zeta_p^2 (p+\log n)^{1/2}$. By the prescribed choice of the bandwidth $h = h_n \asymp \{(p+\log n)/n\}^\gamma$ with $\gamma \in [1/4, 1/2)$, and the sample size requirement $n \gtrsim \zeta_p^{2/(1-\gamma)}(p + \log n)^{(1/2-\gamma)/(1-\gamma)}$, we conclude that the above "good" event occurs with probability at least $1 - C_1 n^{-1}$, and

$$\left(\frac{1-\tau_L}{1-\tau_j}\right)^C \underline{g}^{-1}\sqrt{\frac{p+\log n}{n}} \asymp r_j \le r^\diamondsuit \asymp \left(\frac{p+\log n}{n}\right)^\gamma \ \text{ for all } j = 0,1,\ldots,m.$$

This proves the claim (B.8).

To establish the uniform rate of convergence for $\{\widehat{\boldsymbol{\beta}}(\tau), \tau \in [\tau_L, \tau_U]\}$, define disjoint intervals $\mathcal{I}_j = [\tau_j, \tau_{j+1})$ for $j = 0,1,\ldots,m-1$, and $\mathcal{I}_m = \{\tau_m\}$. For any $\tau \in [\tau_L, \tau_U]$, by the definition of $\widehat{\boldsymbol{\beta}}(\cdot)$ there exits a unique index $j \in \{0,1,\ldots,m\}$ such that $\tau \in \mathcal{I}_j$ and $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}(\tau_j) = \widehat{\boldsymbol{\beta}}_j$. Hence, conditioned on the "good" event that occurs with high probability,

$$\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\|_\Sigma = \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}^*(\tau)\|_\Sigma \le \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}^*_j\|_\Sigma + \|\boldsymbol{\beta}^*(\tau) - \boldsymbol{\beta}^*(\tau_j)\|_\Sigma \le r_j + \underline{f}^{-1}\delta^*.$$

Taking the maximum over j on the right-hand side, and then the supremum over $\tau \in [\tau_L, \tau_U]$ on the left, we obtain

$$\sup_{\tau\in[\tau_L,\tau_U]} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\|_\Sigma \le \max_{0\le j\le m} r_j + \underline{f}^{-1}\delta^* = r_m + \underline{f}^{-1}\delta^*,$$

completing the proof. ☐

B.3. Proof of Theorem 3.2. Similarly to the proof of Theorem 3.1, we divide the proof into two stages. In stage one, we prove a uniform bound over the grid points $\{\tau_0, \ldots, \tau_m\}$; in stage two, we prove the claimed bound (16) which holds uniformly over the interval $[\tau_L, \tau_U]$.

STAGE ONE. As before, we write $\mathbf{J}_j = \mathbf{J}(\tau_j)$ and $\mathbf{H}_j = \mathbf{H}(\tau_j)$ for $j = 0,\ldots,m$, and define the discretized integrated error up to τ_j as

$$\text{(B.17)}\quad \widetilde{e}_0 := \mathbf{J}_0(\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*_0) \ \text{ and } \ \widetilde{e}_j := \underbrace{\mathbf{J}_j(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}^*_j)}_{\text{current step}} + \underbrace{\sum_{\ell=0}^{j-1} w_\ell \mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^*_\ell)}_{\text{preceding steps}}, \quad j = 1,\ldots,m.$$

Let $\{r_j\}_{j=0}^m$ be the sequence of radii from the proof of Theorem 3.1. We will show that

$$\text{(B.18)}\qquad \sup_{j=0,\ldots,m} ||\,\widetilde{e}_j + Q^*_j||_{\Sigma^{-1}} \lesssim \left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r_j + h\right)\cdot r_j,$$

holds with probability at least $1-C_2n^{-1}$ for some absolute constant $C_2>0$, where

(B.19) $$Q_0^* = \widehat{Q}_0(\boldsymbol{\beta}_0^*) = \frac{1}{n}\sum_{i=1}^n \{\Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^* - y_i) - \tau_0\}\boldsymbol{x}_i \quad \text{and}$$

(B.20) $$Q_j^* = \frac{1}{n}\sum_{i=1}^n \left\{\Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i) - \sum_{\ell=0}^{j-1} w_\ell \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \tau_0\right\}\boldsymbol{x}_i \text{ for } j\in[m].$$

We prove the claim (B.18) in a sequential manner, conditioned on some "good" events. Set

$$\mathcal{A} = \bigcap_{j=0}^m \{\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma \le r_j\}, \text{ satisfying } \mathbb{P}(\mathcal{A}) \ge 1 - C_1 n^{-1},$$

For an increasing sequence $0<\lambda_0\le\lambda_1\le\cdots\le\lambda_m$ to be determined, define the event

(B.21) $$\mathcal{E} = \bigcap_{j=0}^{m-1}\{\varpi_j(r_j)\le\lambda_j\cdot r_j,\ \omega_j(r_j)\le\lambda_j\cdot r_j\}\bigcap\{\omega_j(r_m)\le\lambda_m\cdot r_m\},$$

where $\varpi_j(\cdot)$'s and $\omega_j(\cdot)$'s are defined in (B.3) and (B.4), respectively. Recall that $\widehat{Q}_0(\widehat{\boldsymbol{\beta}}_0) = \widehat{Q}_1(\widehat{\boldsymbol{\beta}}_1) = \cdots = \widehat{Q}_m(\widehat{\boldsymbol{\beta}}_m) = \mathbf{0}$. Conditioning on the event $\mathcal{A}\cap\mathcal{E}$, we have

$$\|\widetilde{e}_0 + Q_0^*\|_{\Sigma^{-1}} = \|\widehat{Q}_0(\widehat{\boldsymbol{\beta}}_0) - \widehat{Q}_0(\boldsymbol{\beta}_0^*) - \mathbf{J}_0(\widehat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} \le \lambda_0 r_0,$$

and for $j\in[m]$,

$$\begin{aligned}
\|\widetilde{e}_j + Q_j^*\|_{\Sigma^{-1}} &= \Bigg\|\mathbf{J}_j(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*) - \widehat{Q}_j(\widehat{\boldsymbol{\beta}}_j) + \sum_{\ell=0}^{j-1} w_\ell \mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell^*) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n\left\{\Delta_i\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i) - \sum_{\ell=0}^{j-1} w_\ell\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \tau_0\right\}\boldsymbol{x}_i\Bigg\|_{\Sigma^{-1}} \\
&= \Bigg\|\mathbf{J}_j(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*) - \frac{1}{n}\sum_{i=1}^n \Delta_i\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\}\boldsymbol{x}_i \\
&\qquad + \sum_{\ell=0}^{j-1} w_\ell\mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell^*) + \sum_{\ell=0}^{j-1} w_\ell\cdot\underbrace{\frac{1}{n}\sum_{i=1}^n\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\ell) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\}\boldsymbol{x}_i}_{=\widehat{\Delta}_\ell \text{ in (B.11)}}\Bigg\|_{\Sigma^{-1}} \\
&\le \Bigg\|\frac{1}{n}\sum_{i=1}^n \Delta_i\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\}\boldsymbol{x}_i - \mathbf{J}_j(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)\Bigg\|_{\Sigma^{-1}} \\
&\qquad + \sum_{\ell=0}^{j-1} w_\ell\|\widehat{\Delta}_\ell + \mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell^*)\|_{\Sigma^{-1}} \\
&\le \omega_j(r_j) + \sum_{\ell=0}^{j-1} w_\ell\varpi_\ell(r_\ell) \le \lambda_j r_j + \sum_{\ell=0}^{j-1} w_\ell\lambda_\ell r_\ell.
\end{aligned}$$

Recall from the definition of r_j in (B.15), we have $\sum_{\ell=0}^{j-1} w_\ell r_\ell \le (2\underline{f})^{-1}\kappa r_j$, and hence

$$\|\widetilde{e}_j + Q_j^*\|_{\Sigma^{-1}} \le \{1 + (2\underline{f})^{-1}\kappa\}\cdot\lambda_j r_j,\quad j\in[m].$$

In view of Lemma B.5 with $t = 2\log n$, we set

$$\lambda_j \asymp m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r_j + h,\ \ j = 0, 1, \ldots, m,$$

so that event $\mathcal{E}$ in (B.21) holds with probability at least $1 - 2(m+1)n^{-2}$. This proves (B.18).

STAGE TWO. To generalize (B.18) to (16) on the whole process $\widehat{\boldsymbol{\beta}}(\cdot)$, define disjoint intervals $\mathcal{I}_j = [\tau_j, \tau_{j+1})$ for $j = 0, \ldots, m-1$ and $\mathcal{I}_m = \{\tau_m\}$. For any $\tau \in [\tau_L, \tau_U]$, there exists a unique index $j \in \{0, \ldots, m\}$ such that $\tau \in \mathcal{I}_j$ and $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}_j$. With this notation, the left-hand side of (16) equals

$$\max_{j=0,\ldots,m}\sup_{\tau\in\mathcal{I}_j}\left\|\widehat{\boldsymbol{e}}(\tau) - \frac{1}{n}\sum_{i=1}^n\left\{\tau_L + \int_{\tau_L}^{\tau}\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u) - \Delta_i\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - y_i)\right\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}},$$

where $\widehat{\boldsymbol{e}}(\tau) = \widehat{\boldsymbol{\beta}}_{\mathrm{int}}(\tau) - \boldsymbol{\beta}^*_{\mathrm{int}}(\tau)$. To control the discretization error, conditioning on the event $\mathcal{A}\cap\mathcal{E}$ we have, for each $j = 0, 1, \ldots, m$,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{I}_j}\|\mathbf{J}(\tau)\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\} - \mathbf{J}_j(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}^*_j)\|_{\Sigma^{-1}} \\
&\leq \sup_{\tau\in\mathcal{I}_j}\|\{\mathbf{J}(\tau) - \mathbf{J}_j\}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}^*_j)\|_{\Sigma^{-1}} + \sup_{\tau\in\mathcal{I}_j}\|\mathbf{J}(\tau)\{\boldsymbol{\beta}^*_j - \boldsymbol{\beta}^*(\tau)\}\|_{\Sigma^{-1}} \\
&\leq l_1\underline{f}^{-1}m_3 r_j\delta^* + \bar{g}\underline{f}^{-1}\delta^* \lesssim \delta^*, \qquad \text{(B.22)}
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\tau\in\mathcal{I}_j}\left\|\int_{\tau_L}^{\tau}\mathbf{H}(u)\{\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^*(u)\}\mathrm{d}H(u) - \sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}\mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^*_\ell)\mathrm{d}H(u)\right\|_{\Sigma^{-1}} \qquad \text{(B.23)} \\
&\leq \left\|\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}\big[\mathbf{H}(u)\{\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^*(u)\} - \mathbf{H}_\ell(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^*_\ell)\big]\mathrm{d}H(u)\right\|_{\Sigma^{-1}} \\
&\quad + \sup_{\tau\in\mathcal{I}_j}\left\|\int_{\tau_j}^{\tau}\mathbf{H}(u)\{\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^*(u)\}\mathrm{d}H(u)\right\|_{\Sigma^{-1}} \lesssim \log\left(\frac{1-\tau_0}{1-\tau_{j+1}}\right)\cdot\delta^*.
\end{aligned}$$

Next we control the approximation error for discretizing the linear process $(1/n)\sum_{i=1}^n \boldsymbol{U}_i(\cdot)$. For any interval $\mathcal{I}_j$, write $\boldsymbol{v}(\tau) = \boldsymbol{\beta}^*(\tau) - \boldsymbol{\beta}^*_j$ for $\tau \in \mathcal{I}_j$, satisfying $\|\boldsymbol{v}(\tau)\|_\Sigma \leq \underline{f}^{-1}(\tau_{j+1} - \tau_j)$ by the Lipschitz continuity of $\boldsymbol{\beta}^*(\cdot)$. Moreover, we have

$$\begin{aligned}
&\left\|\mathbb{E}\Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*_j - y_i)\big\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}} \\
&= \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\big[\Delta_i\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*_j - y_i)\}\langle\boldsymbol{u}, \boldsymbol{z}_i\rangle\big] \\
&= \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\int_{-\infty}^{\infty}\left\{\bar{K}\left(\frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - u}{h}\right) - \bar{K}\left(\frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*_j - u}{h}\right)\right\}g(u|\boldsymbol{x})\mathrm{d}u\cdot\langle\boldsymbol{u}, \boldsymbol{z}_i\rangle \\
&= \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} h\cdot\mathbb{E}\int_{-\infty}^{\infty}\{\bar{K}(v + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}(\tau)/h) - \bar{K}(v)\}g(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*_j - hv|\boldsymbol{x})\mathrm{d}v\cdot\langle\boldsymbol{u}, \boldsymbol{z}_i\rangle \\
&\leq \bar{g}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\int_{-\infty}^{\infty}\left\{\int_0^1 K(v + w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}(\tau)/h)\mathrm{d}w\right\}\mathrm{d}v\cdot\langle\boldsymbol{v}(\tau), \boldsymbol{x}_i\rangle\langle\boldsymbol{u}, \boldsymbol{z}_i\rangle
\end{aligned}$$

$$\leq \bar{g} \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}\left(\int_0^1\int_{-\infty}^{\infty} K(v+w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}(\tau)/h)\mathrm{d}v\mathrm{d}w\right)\cdot\langle\boldsymbol{v}(\tau),\boldsymbol{x}_i\rangle\langle\boldsymbol{u},\boldsymbol{z}_i\rangle$$
$$\leq \bar{g} \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}|\langle\boldsymbol{v}(\tau),\boldsymbol{x}_i\rangle\langle\boldsymbol{u},\boldsymbol{z}_i\rangle| \leq \bar{g}\|\boldsymbol{v}(\tau)\|_{\Sigma} \leq \bar{g}\underline{f}^{-1}(\tau_{j+1}-\tau_j).$$

This, combined with Lemma B.5–(ii) with $r=\underline{f}^{-1}\delta^*$ and $t=2\log n$, yields that with probability at least $1-(m+1)n^{-2}$,

$$\text{(B.24)}\qquad \max_{0\leq j\leq m}\sup_{\tau\in\mathcal{I}_j}\left\|\frac{1}{n}\sum_{i=1}^n \Delta_i\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)-y_i)-\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*-y_i)\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}}$$
$$\lesssim \delta^*+\delta^* m_4^{1/2}\sqrt{\frac{p+\log n}{nh}}\lesssim\delta^*$$

as long as $nh\gtrsim\zeta_p^2(p+\log n)$. Similarly, it follows from Lemma B.3–(iii), Lemma B.6 and the union bound that with probability at least $1-2mn^{-2}$,

$$\sup_{\tau\in\mathcal{I}_j}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\left\{\int_{\tau_L}^{\tau}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)-\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\mathrm{d}H(u)\right\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}}$$
$$\leq\left\|\frac{1}{n}\sum_{i=1}^n\sum_{\ell=0}^{j-1}(1-\mathbb{E})\int_{\tau_\ell}^{\tau_{\ell+1}}\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\}\mathrm{d}H(u)\cdot\boldsymbol{z}_i\right\|_2$$
$$+\sup_{\tau\in\mathcal{I}_j}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_j}^{\tau}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{z}_i\right\|_2$$
$$\leq\sum_{\ell=0}^{j-1}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_\ell}^{\tau_{\ell+1}}\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\}\mathrm{d}H(u)\cdot\boldsymbol{z}_i\right\|_2$$
$$+\sup_{\tau\in\mathcal{I}_j}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_j}^{\tau}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{z}_i\right\|_2$$

(B.25)

$$\lesssim\log\left(\frac{1-\tau_0}{1-\tau_j}\right)\left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}}+\zeta_p^2\frac{\log n}{nh}\right)\cdot\delta^*$$
$$+\log\left(\frac{1-\tau_j}{1-\tau_{j+1}}\right)\left(\sqrt{\frac{p+\log n}{n}}+\zeta_p\frac{\log n}{n}\right)\cdot\delta^*$$

for all $j=0,1,\ldots,m$. Turning to the deterministic approximation error, it can be shown that for any $j=0,1,\ldots,m-1$ and $\tau\in\mathcal{I}_j$,

$$\left\|\mathbb{E}\left\{\int_{\tau_L}^{\tau}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)-\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\mathrm{d}H(u)\right\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}}$$
$$\lesssim\log\left(\frac{1-\tau_0}{1-\tau_{j+1}}\right)\cdot\delta^*.$$

Together, (B.18) and (B.22)–(B.25) prove the claimed bounds (14)–(16).

It remains to control the bias term $\mathbb{E}\boldsymbol{U}_i(\cdot)$. Define the non-smoothed version of $\boldsymbol{U}_i(\cdot)$ as

$$\boldsymbol{V}_i(\tau) = \left[\tau_L + \int_{\tau_L}^{\tau} \mathbb{1}\{y_i > \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\mathrm{d}H(u) - \Delta_i \mathbb{1}\{y_i < \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\}\right]\boldsymbol{x}_i, \;\; \tau \in [\tau_L, \tau_U].$$

By the martingale property, $\mathbb{E}\boldsymbol{V}_i(\tau) = \mathbf{0}$ for every $\tau \in [\tau_L, \tau_U]$. Note that

$$\boldsymbol{U}_i(\tau) - \boldsymbol{V}_i(\tau) = \Bigg(\Delta_i\big[\mathbb{1}\{y_i < \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\} - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - y_i)\big] \\ + \int_{\tau_L}^{\tau}\big[\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)) - \mathbb{1}\{y_i > \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\big]\mathrm{d}H(u)\Bigg)\boldsymbol{x}_i.$$

Following the same calculations that lead to (E.16), we obtain

$$\sup_{\tau\in[\tau_L,\tau_U]} \|\mathbb{E}\boldsymbol{U}_i(\tau)\|_{\Sigma^{-1}} = \sup_{\tau\in[\tau_L,\tau_U]} \|\mathbb{E}\{\boldsymbol{U}_i(\tau) - \boldsymbol{V}_i(\tau)\}\|_{\Sigma^{-1}} \le 0.5 l_1\kappa_2\big\{1 + \log\big(\tfrac{1-\tau_L}{1-\tau_U}\big)\big\}h^2.$$

This completes the proof of the theorem. □

B.4. Proof of Theorem 3.3. Assume without loss of generality that $||\boldsymbol{a}_n||_\Sigma = 1$; otherwise, we simply replace $\boldsymbol{a}_n$ by $\boldsymbol{a}_n/||\boldsymbol{a}_n||_\Sigma$. Following the general result in Theorem 1.5.4 of [15], the claimed weak convergence (22) is a direct consequence of the weak convergence of finite-dimensional marginals and the asymptotic tightness of $\mathbb{G}_n(\cdot)$.

For the former, via the Cramér–Wold device, it is equivalent to show that for any finite set of values $\{\tau_\ell\}_{\ell=1}^L \subseteq [\tau_L, \tau_U]$ and $(\gamma_1, \ldots, \gamma_L)^{\mathrm{T}} \in \mathbb{R}^L$,

$$\sum_{\ell=1}^L \gamma_\ell \mathbb{G}_n(\tau_\ell) \xrightarrow{\mathrm{d}} \sum_{\ell=1}^L \gamma_\ell \mathbb{G}(\tau_\ell), \tag{B.26}$$

with $\mathbb{G}(\cdot)$ defined in (22). For $i = 1, \ldots, n$, define centered variables $W_i = \sum_{\ell=1}^L \gamma_\ell\langle \boldsymbol{a}_n, \boldsymbol{U}_{0i}(\tau_\ell)\rangle$ with $\boldsymbol{U}_{0i}(\tau) := \boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)$, so that $\sum_{\ell=1}^L \gamma_\ell \mathbb{G}_n(\tau_\ell) = n^{-1/2}\sum_{i=1}^n W_i$. Moreover,

$$\begin{aligned}
\mathrm{Var}(W_i) &= \mathrm{Var}\left(\sum_{\ell=1}^L \gamma_\ell\langle \boldsymbol{a}_n, \boldsymbol{U}_{0i}(\tau_\ell)\rangle\right) \\
&= \sum_{\ell_1=1}^L\sum_{\ell_2=1}^L \gamma_{\ell_1}\gamma_{\ell_2}\cdot \mathrm{Cov}\left(\langle \boldsymbol{a}_n, \boldsymbol{U}_{0i}(\tau_{\ell_1}), \langle \boldsymbol{a}_n, \boldsymbol{U}_{0i}(\tau_{\ell_2})\rangle\right) \\
&= \sum_{\ell_1=1}^L\sum_{\ell_2=1}^L \gamma_{\ell_1}\gamma_{\ell_2}\cdot \boldsymbol{a}_n^{\mathrm{T}}\,\mathbb{E}\big\{\boldsymbol{U}_{0i}(\tau_{\ell_1})\boldsymbol{U}_{0i}(\tau_{\ell_2})^{\mathrm{T}}\big\}\boldsymbol{a}_n \\
&= \sum_{\ell_1=1}^L\sum_{\ell_2=1}^L \gamma_{\ell_1}\gamma_{\ell_2}\cdot \boldsymbol{a}_n^{\mathrm{T}}\,\mathbb{E}\big\{\boldsymbol{U}_{i}(\tau_{\ell_1})\boldsymbol{U}_{i}(\tau_{\ell_2})^{\mathrm{T}}\big\}\boldsymbol{a}_n \\
&\quad - \sum_{\ell_1=1}^L\sum_{\ell_2=1}^L \gamma_{\ell_1}\gamma_{\ell_2}\cdot \boldsymbol{a}_n^{\mathrm{T}}\mathbb{E}\big\{\boldsymbol{U}_{i}(\tau_{\ell_1})\big\}\mathbb{E}\big\{\boldsymbol{U}_{i}(\tau_{\ell_2})^{\mathrm{T}}\big\}\boldsymbol{a}_n \\
&\to \sum_{\ell_1=1}^L\sum_{\ell_2=1}^L \gamma_{\ell_1}\gamma_{\ell_2}\cdot H(\ell_1,\ell_2) = \mathrm{Var}\left(\sum_{\ell=1}^L \gamma_\ell\mathbb{G}(\tau_\ell)\right) \;\text{ as }\; n\to\infty,
\end{aligned}$$

where $H(\cdot,\cdot)$ is defined in (21). The finite-dimensional weak convergence (B.26) then follows from the central limit theorem.

Turning to the asymptotic tightness of $\mathbb{G}_n(\cdot)$, an equivalent characterization is the asymptotic uniform equicontinuity in probability; see Theorem 1.5.7 in [15] and the definition above it. That is, for any $x > 0$,

$$\lim_{\delta \to 0} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{|\tau_1-\tau_2|<\delta} \big|\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2)\big| > x \right\} = 0,$$

which is ensured by Lemma B.8.

Finally, the existence of almost surely continuous sample paths of $\mathbb{G}(\cdot)$ follows from Addendum 1.5.8 in [15]. □

APPENDIX C: PROOFS OF THE MAIN RESULTS IN SECTION 3.3

C.1. Technical lemmas. In this section, we provide the technical lemmas needed to establish the validity of the multiplier bootstrap procedure. Recall that e_i's are i.i.d. Rademacher random variables that are independent of the observed data $\mathbb{D}_n = \{y_i, \Delta_i, \boldsymbol{x}_i\}_{i=1}^n$. Similarly to (B.2), we define the symmetrized Bregman divergence $D^\flat : \mathbb{R}^p \times \mathbb{R}^p \to [0,\infty)$ in the bootstrap world as

$$D^\flat(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = \langle \widehat{Q}_j^\flat(\boldsymbol{\beta}_1) - \widehat{Q}_j^\flat(\boldsymbol{\beta}_2), \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\rangle, \tag{C.1}$$

which is also independent of j, where $\widehat{Q}_j^\flat(\cdot)$'s are the randomly perturbed estimating equations defined in (7) and (8).

LEMMA C.1 (Conditional restricted strong convexity). *Assume Conditions 3.1 and 3.3 hold. Let $r = h/(4\eta_{1/4})$ with $\eta_{1/4}$ defined in (B.1), and $t \ge 0$. Suppose the "effective" sample size satisfies $nh \gtrsim \max\{p, \zeta_p t^{1/2}\}$. Then, there exists an event $\mathcal{E}_1(t)$ with $\mathbb{P}\{\mathcal{E}_1(t)\} \ge 1 - (m+3)e^{-t}$ such that, conditioning on $\mathcal{E}_1(t)$,*

$$\inf_{j\in\{0,\dots,m\}} \inf_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \frac{D(\boldsymbol{\beta},\boldsymbol{\beta}_j^*)}{\kappa_l\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_j^*\|_\Sigma^2} \ge \frac{1}{2}\underline{g}, \quad \textit{and}$$

$$\mathbb{P}^*\left\{ \inf_{j\in\{0,\dots,m\}} \inf_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \frac{D^\flat(\boldsymbol{\beta},\boldsymbol{\beta}_j^*)}{\kappa_l\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_j^*\|_\Sigma^2} \ge \frac{1}{2}\underline{g} \right\} \ge 1-(m+1)e^{-t}.$$

LEMMA C.2. *Assume Condition 3.2 holds. Let $\{\xi_i\}_{i=1}^n$ be independent random variables satisfying $|\xi_i| \le M$ for some $M > 0$, and $\{e_i\}_{i=1}^n$ are Rademacher random variables independent of the data $\{\boldsymbol{x}_i,\xi_i\}_{i=1}^n$. Then, there exists an event $\mathcal{E}_2$ depending on $\{\boldsymbol{x}_i,\xi_i\}_{i=1}^n$ such that (i) $\mathbb{P}(\mathcal{E}_2) \ge 1-n^{-2}$, and (ii) conditioned on $\mathcal{E}_2$,*

$$\left\| \frac{1}{n}\sum_{i=1}^n (e_i \cdot \xi_i \boldsymbol{z}_i) \right\|_2 \lesssim \sqrt{\frac{p+\log n}{n}}$$

holds with $\mathbb{P}^*$-probability (over $\{e_i\}_{i=1}^n$) at least $1-n^{-2}$ as long as the sample size satisfies $n \gtrsim \zeta_p^2 \log n$.

The following lemma provides upper bounds for two Rademacher weighted stochastic processes. For $j = 0,1,\dots,m$ and any $r > 0$, define

$$\Gamma_j(r) := \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \left\| \frac{1}{n}\sum_{i=1}^n e_i\big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - \bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\big\}\boldsymbol{z}_i \right\|_2, \tag{C.2}$$

$$\Gamma_j^{\Delta}(r) := \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \left\| \frac{1}{n}\sum_{i=1}^n e_i \Delta_i \big\{ \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\big\}\boldsymbol{z}_i \right\|_2, \tag{C.3}$$

where $\{e_i\}_{i=1}^n$ is a sequence of independent Rademacher random variables that are independent of $\{y_i, \Delta_i, \boldsymbol{x}_i\}_{i=1}^n$, and $\boldsymbol{z}_i = \Sigma^{-1/2}\boldsymbol{x}_i$.

LEMMA C.3. *Assume that Conditions 3.1–3.3 hold, and $K(\cdot)$ in Condition 3.1 is l_K-Lipschitz continuous. Given any $0 < r \le \zeta_p$, there exists an event $\mathcal{E}_3$ with $\mathbb{P}(\mathcal{E}_3) \ge 1 - 3n^{-1}$ such that, with $\mathbb{P}^*$-probability at least $1-(m+1)n^{-2}$ conditioned on $\mathcal{E}_3$,*

$$\sup_{j\in\{0,\ldots,m\}} \Gamma_j(r) \lesssim r\sqrt{\frac{p+\log n}{nh}}\left(m_4^{1/2} + \zeta_p^2\sqrt{\frac{p\log n}{nh}}\right),$$

provided $n \gtrsim \zeta_p^2 \log n$. The same uniform bound also applies to $\Gamma_j^{\Delta}(r)$.

LEMMA C.4. *Assume that Conditions 3.1–3.3 hold, and $K(\cdot)$ in Condition 3.1 is l_K-Lipschitz continuous. Then, there exists an event $\mathcal{E}_4$ with $\mathbb{P}(\mathcal{E}_4) \ge 1-(m+3n+1)n^{-2}$ such that, with $\mathbb{P}^*$-probability at least $1-n^{-2}$ conditional on $\mathcal{E}_4$,*

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} &\|\mathbf{J}_j(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*) - \{\widehat{Q}_j^\flat(\boldsymbol{\beta}) - \widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*)\}\|_{\Sigma^{-1}} \\ &\lesssim \left\{ m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r + h + \zeta_p^2 \frac{(p\log n)^{1/2}(p+\log n)^{1/2}}{nh}\right\}\cdot r\end{aligned}$$

holds uniformly over $j = 0,1,\ldots,m$, where $\mathbf{J}_j = \mathbb{E}\{g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}$.

LEMMA C.5. *Assume Conditions 3.1–3.4 hold, and let $r > 0$. Then, there exists an event $\mathcal{E}_5$ with $\mathbb{P}(\mathcal{E}_5) \ge 1 - mn^{-2}$ such that conditioning on $\mathcal{E}_5$,*

$$\begin{aligned}\sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell\in\boldsymbol{\beta}_\ell^*+\Theta(r)\}} &\left\| \frac{1}{n}\sum_{i=1}^n\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}} \mathrm{d}H(u)\Big[\mathbf{H}_\ell(\boldsymbol{\beta}_\ell-\boldsymbol{\beta}_\ell^*) - W_i\big\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell)\big\}\boldsymbol{x}_i\Big]\right\|_{\Sigma^{-1}} \\ &\lesssim \log\left(\tfrac{1-\tau_0}{1-\tau_j}\right)\cdot\left\{\left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r + h\right)\cdot r + \max_{\ell=0,\ldots,j-1}\Gamma_\ell(r)\right\}\end{aligned}$$

holds for all $j = 1,\ldots,m$, where $\mathbf{H}_\ell = \mathbb{E}\{f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}$ and $\Gamma_\ell(r)$ is defined in (C.2).

The following lemma establishes the asymptotic uniform equicontinuity of the process $\mathbb{G}_n^\flat(\cdot) = n^{-1/2}\sum_{i=1}^n e_i\langle\boldsymbol{a}_n, \boldsymbol{U}_i(\cdot)\rangle$ in $\ell^\infty([\tau_L,\tau_U])$, thus validating the asymptotic tightness of $\mathbb{G}_n^\flat(\cdot)$, where $\boldsymbol{U}_i(\cdot)$ is defined in (15).

LEMMA C.6. *Assume that the conditions of Theorem 3.6 hold. For any $x > 0$ and sequence of vectors $\boldsymbol{a}_n$ satisfying $\|\boldsymbol{a}_n\|_\Sigma = 1$, conditioned on any observed data $\mathbb{D}_n = \{(y_i,\Delta_i,\boldsymbol{x}_i\}_{i=1}^n$, we have*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{|\tau_1-\tau_2|<\delta}\big|\mathbb{G}_n^\flat(\tau_1) - \mathbb{G}_n^\flat(\tau_2)\big| > x\right\} = 0. \tag{C.4}$$

where $\mathbb{G}_n^\flat(\cdot) = n^{-1/2}\sum_{i=1}^n e_i\langle\boldsymbol{a}_n,\boldsymbol{U}_i(\cdot)\rangle$ with $\boldsymbol{U}_i(\tau)$ defined in (15).

C.2. Proof of Theorem 3.4. Similar to the proof of Theorem 3.1, we first prove a uniform bound over the grid points $\{\tau_0, \ldots, \tau_m\}$. Recall the bootstrapped SEE $\widehat{Q}_j^\flat(\boldsymbol{\beta})$ given in (7) and (8), and $\widehat{Q}_j^\flat(\widehat{\boldsymbol{\beta}}_j^\flat) = \mathbf{0}$. Following the localized argument as in the proof of Theorem 3.1, for the same radius parameter $r^\diamondsuit$ therein, define $\widetilde{\boldsymbol{\beta}}_j^\flat = \boldsymbol{\beta}_j^* + \gamma_j(\widehat{\boldsymbol{\beta}}_j^\flat - \boldsymbol{\beta}_j^*)$ with $\gamma_j := \sup\{\gamma \in [0,1] : \gamma(\widehat{\boldsymbol{\beta}}_j^\flat - \boldsymbol{\beta}_j^*) \in \Theta(r^\diamondsuit)\}$, so that $\widetilde{\boldsymbol{\beta}}_j^\flat = \widehat{\boldsymbol{\beta}}_j^\flat$ if $\widehat{\boldsymbol{\beta}}_j^\flat \in \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit)$ and $\widetilde{\boldsymbol{\beta}}_j^\flat \in \boldsymbol{\beta}_j^* + \partial\Theta(r^\diamondsuit)$ if $\widehat{\boldsymbol{\beta}}_j^\flat \notin \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit)$. Consequently,

$$D^\flat(\widetilde{\boldsymbol{\beta}}_j^\flat, \boldsymbol{\beta}_j^*) \le \rho_j \cdot \langle -\widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*), \widetilde{\boldsymbol{\beta}}_j^\flat - \boldsymbol{\beta}_j^* \rangle \le \|\widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} \cdot \|\widetilde{\boldsymbol{\beta}}_j^\flat - \boldsymbol{\beta}_j^*\|_\Sigma,$$

where $D^\flat$ is defined in (C.1).

In addition to $\mathcal{F}$ in (B.9), define

$$\mathcal{F}^\flat = \bigcap_{j=0}^{m} \left\{ D^\flat(\boldsymbol{\beta}, \boldsymbol{\beta}_j^*) \ge \kappa \cdot \|\boldsymbol{\beta} - \boldsymbol{\beta}_j^*\|_\Sigma^2 \ \text{ for all } \ \boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit) \right\}. \tag{C.5}$$

Conditioned on $\mathcal{F}^\flat$, we have for all $j = 0, 1, \ldots, m$ that

$$\|\widetilde{\boldsymbol{\beta}}_j^\flat - \boldsymbol{\beta}_j^*\|_\Sigma \le \kappa^{-1} \|\widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}}.$$

For the bootstrapped estimating function, by the triangle inequality we have

$$\begin{aligned}
&\|\widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} \\
&\le \left\| \frac{1}{n}\sum_{i=1}^n e_i \left\{ \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_j^* - y_i) - \tau_0 - \sum_{\ell=0}^{j-1} w_\ell \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_\ell^*) \right\} \boldsymbol{x}_i \right\|_{\Sigma^{-1}} \\
&\quad + \left\| \frac{1}{n}\sum_{i=1}^n \sum_{\ell=0}^{j-1} e_i w_\ell \{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\widehat{\boldsymbol{\beta}}_\ell^\flat) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_\ell^*)\}\boldsymbol{x}_i \right\|_{\Sigma^{-1}} \\
&\quad + \left\| \widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell\big(\widehat{\Delta}_\ell^\flat + \Delta_\ell\big) + Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}_i\} \right\|_{\Sigma^{-1}}
\end{aligned}$$

(C.6)

$$\begin{aligned}
&\le \underbrace{\left\| \frac{1}{n}\sum_{i=1}^n e_i \left\{ \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_j^* - y_i) - \tau_0 - \sum_{\ell=0}^{j-1} w_\ell \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_\ell^*) \right\} \boldsymbol{x}_i \right.}_{:=\widetilde{Q}_j^\flat(\boldsymbol{\beta}_j^*)}\left.\vphantom{\sum_{i=1}^n}\right\|_{\Sigma^{-1}} \\
&\quad + \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} + \sum_{\ell=0}^{j-1} w_\ell\big(\|\widetilde{\Delta}_\ell^\flat\|_{\Sigma^{-1}} + \|\widehat{\Delta}_\ell^\flat\|_{\Sigma^{-1}} + \|\Delta_\ell\|_{\Sigma^{-1}}\big) \\
&\quad + \left\| Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}_i\} \right\|_{\Sigma^{-1}},
\end{aligned}$$

for $j \ge 1$, where $w_\ell = H(\tau_{\ell+1}) - H(\tau_\ell)$,

$$\widetilde{\Delta}_\ell^\flat = \frac{1}{n}\sum_{i=1}^n e_i\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\widehat{\boldsymbol{\beta}}_\ell^\flat) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_\ell^*)\}\boldsymbol{x}_i,$$

$$\widehat{\Delta}_\ell^\flat = \frac{1}{n}\sum_{i=1}^n \{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\widehat{\boldsymbol{\beta}}_\ell^\flat) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}_\ell^*)\}\boldsymbol{x}_i$$

and Δ_ℓ is defined in (B.11). In particular,

$$\|\widehat{Q}_0^\flat(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} = \left\| \frac{1}{n}\sum_{i=1}^n e_i \left\{ \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^* - y_i) - \tau_0 \right\} \boldsymbol{x}_i \right\|_{\Sigma^{-1}}$$
$$+ \|\widehat{Q}_0(\boldsymbol{\beta}_0^*) - Q_0(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} + \|Q_0(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}}.$$

The rest of the proof is similar to that of Theorem 3.1, and thus we skip some of the technical details. For some $\delta > 0$ to be determined, define the event

(C.7)
$$\mathcal{G}^\flat = \left\{ \max_{0\le j\le m} \|\widetilde{Q}_j^\flat(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} \bigvee \max_{0\le j\le m} \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} \bigvee \max_{0\le \ell\le m-1} \|\Delta_\ell\|_{\Sigma^{-1}} \le \delta \right\}.$$

Conditioned on $\mathcal{F}^\flat \cap \mathcal{G}^\flat$, it follows from Lemma B.7, (C.6) and (C.7) that

$$\|\widehat{Q}_0^\flat(\boldsymbol{\beta}_0^*)\|_{\Sigma^{-1}} < 2\delta + a,$$
$$\widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*)\|_{\Sigma^{-1}} < 2\delta + a + \sum_{\ell=0}^{j-1} w_\ell\big(\delta + a + \|\widetilde{\Delta}_\ell^\flat\|_{\Sigma^{-1}} + \|\widehat{\Delta}_\ell\|_{\Sigma^{-1}}\big), \;\; j = 1, \ldots, m,$$

where a is defined in (B.13). Similarly to (B.16), conditioned on the "good" event

$$\mathcal{F}^\flat \cap \mathcal{G}^\flat \cap \bigcap_{\ell=0}^{m-1} \{\varpi_\ell(r_\ell) \vee \Gamma_\ell(r_\ell) \le \overline{f} r_\ell\}, \tag{C.8}$$

the convergence radii $\{r_j\}_{j=0}^m$ are recursively defined as

(C.9)
$$r_j = \big(1 + 3\kappa^{-1}\overline{f} w_{j-1}\big) r_{j-1} + \kappa^{-1} w_{j-1}(\delta + a), \;\; j = 1, \ldots, m, \;\; \text{and} \;\; r_0 = \kappa^{-1}(2\delta + a).$$

Denoting $C = \kappa^{-1}(3\overline{f} + 1)$, it follows that $r_j \le \big(\frac{1-\tau_L}{1-\tau_j}\big)^C \cdot r_0$.

Next we complement the above deterministic analysis with probabilistic bounds. By Lemmas B.3, B.5 and Lemmas C.1–C.3, we choose $\kappa = (\underline{g}\kappa_l)/2$ and $\delta \asymp \sqrt{(p + \log n)/n} + \zeta_p \log(n)/n$ so that there exists an event $\mathcal{E}$ with $\mathbb{P}(\mathcal{E}) \ge 1 - C_1 n^{-1}$ such that conditioned on $\mathcal{E}$,

$$\mathbb{P}^*(\mathcal{F}^\flat \cap \mathcal{G}^\flat) \ge 1 - C_2 n^{-1},$$
$$\sup_{\ell\in\{0,\ldots,m-1\}} \varpi_\ell(r_\ell) \lesssim \left(m_4^{1/2}\sqrt{\frac{p + \log n}{nh}} + m_3 r_\ell + h \right) \cdot r_\ell$$

and

$$\sup_{\ell\in\{0,\ldots,m-1\}} \Gamma_\ell(r_\ell) \lesssim r_\ell \sqrt{\frac{p + \log n}{nh}} \left(m_4^{1/2} + \zeta_p^2 \sqrt{\frac{p\log n}{nh}} \right) \;\; \text{with } \mathbb{P}^*\text{-probability at least } 1 - n^{-1},$$

provided $nh \gtrsim \zeta_p^2 (p + \log n)^{1/2}$ and $n \gtrsim \zeta_p^2 \log n$. Moreover, the uniform bound (11) holds conditioned on $\mathcal{E}$. Consequently, it follows from (C.9) and (B.13) that

$$r_j \le \left(\frac{1-\tau_L}{1-\tau_j}\right)^C \cdot r_0 \asymp \left(\frac{1-\tau_L}{1-\tau_j}\right)^C \underline{g}^{-1}\left(\sqrt{\frac{p + \log n}{n}} + \zeta_p \frac{\log n}{n} + h^2 \right)$$

holds uniformly over $j \in \{0, \ldots, m\}$.

Recall that m_3 and m_4 are dimension-free moment parameters. Given the bandwidth $h = h_n \asymp \{(p+\log n)/n\}^\gamma$ with $\gamma \in [1/4, 1/2)$, and under the sample size requirement $n \gtrsim \zeta_p^{2/(1-\gamma)}(p+\log n)^{(1/2-\gamma)/(1-\gamma)}(p\log n)^{1/(2-2\gamma)}$, we conclude that conditioned on $\mathcal{E}$, the "good" event (C.8) occurs with $\mathbb{P}^*$-probability at least $1 - C_3 n^{-1}$, and

$$\left(\frac{1-\tau_L}{1-\tau_j}\right)^C \underline{g}^{-1}\sqrt{\frac{p+\log n}{n}} \asymp r_j \le r^\diamondsuit \asymp \left(\frac{p+\log n}{n}\right)^\gamma \text{ for all } j = 0, 1, \ldots, m.$$

This proves the uniform bound over $\tau \in \{\tau_0, \tau_1, \ldots, \tau_m\}$, which can naturally be extended to $\tau \in [\tau_L, \tau_U]$ following the last paragraph in the proof of Theorem 3.1. ☐

C.3. Proof of Theorem 3.5. We divide the proof into two steps as in the proof of Theorem 3.2. Recall that $W_i = e_i + 1$, where e_i's are independent Rademacher variables.
<u>STEP 1.</u> (Uniform bound over $\{\tau_0, \ldots, \tau_m\}$) For simplicity, we write $\mathbf{J}_j = \mathbf{J}(\tau_j)$ and $\mathbf{H}_j = \mathbf{H}(\tau_j)$ for $j = 0, \ldots, m$, and define the accumulated bootstrap errors as

$$\widetilde{e}^\flat_{\rm int}(\tau_0) := \mathbf{J}_0(\widehat{\boldsymbol\beta}^\flat_0 - \widehat{\boldsymbol\beta}_0) \text{ and}$$

$$\widetilde{e}^\flat_{\rm int}(\tau_j) := \mathbf{J}_j(\widehat{\boldsymbol\beta}^\flat_j - \widehat{\boldsymbol\beta}_j) + \sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}} \mathbf{H}_\ell(\widehat{\boldsymbol\beta}^\flat_\ell - \widehat{\boldsymbol\beta}_\ell)\mathrm{d}H(u), \;\; j = 1, \ldots, m.$$

We claim that there exits an event $\mathcal{F}$ on which (14)–(16) hold such that $\mathbb{P}(\mathcal{F}) \ge 1 - C_4 n^{-1}$, and

(C.10)

$$\sup_{j=0,\ldots,m} ||\,\widetilde{e}^\flat_{\rm int}(\tau_j) + Q_j^{*\flat}||_{\Sigma^{-1}} \lesssim m_4^{1/2}\frac{p+\log n}{nh^{1/2}} + h\sqrt{\frac{p+\log n}{n}} + \zeta_p^2\frac{(p+\log n)(p\log n)^{1/2}}{n^{3/2}h}$$

with $\mathbb{P}^*$-probability at least $1 - C_5 n^{-1}$ conditioned on $\mathcal{F}$, where

$$Q_0^{*\flat} = \frac{1}{n}\sum_{i=1}^n e_i\{\Delta_i \bar K_h(\boldsymbol{x}_i^{\rm T}\boldsymbol\beta_0^* - y_i) - \tau_0\}\boldsymbol{x}_i,$$

$$Q_j^{*\flat} = \frac{1}{n}\sum_{i=1}^n e_i\left\{\Delta_i \bar K_h(\boldsymbol{x}_i^{\rm T}\boldsymbol\beta_j^* - y_i) - \sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}} \bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol\beta_\ell^*)\,\mathrm{d}H(u) - \tau_0\right\}\boldsymbol{x}_i$$

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

From Theorem 3.4 and its proof, we see that there exists an event $\mathcal{E}_1$ with $\mathbb{P}(\mathcal{E}_1) \ge 1 - C_1 n^{-1}$ such that conditioned on $\mathcal{E}_1$,

$$\max_{0\le j\le m} \|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_j^*\|_\Sigma \lesssim \sqrt{\frac{p+\log n}{n}}, \text{ and}$$

$$\text{(C.11)} \qquad \mathbb{P}^*\left(\max_{0\le j\le m} \|\widehat{\boldsymbol\beta}^\flat_j - \boldsymbol\beta_j^*\|_\Sigma \lesssim \sqrt{\frac{p+\log n}{n}}\right) \ge 1 - C_3 n^{-1}.$$

We then prove the claim (C.10). For $j = 0$, by the triangle inequality we have

$$\begin{aligned}
&||\,\widetilde{e}^\flat_{\rm int}(\tau_0) + Q_0^{*\flat}||_{\Sigma^{-1}} \\
&\le ||\,\widetilde{e}_{\rm int}(\tau_0) + Q_0^*||_{\Sigma^{-1}} + \left\|\mathbf{J}_0(\widehat{\boldsymbol\beta}^\flat_0 - \boldsymbol\beta_0^*) + \frac{1}{n}\sum_{i=1}^n W_i\{\Delta_i \bar K_h(\boldsymbol{x}_i^{\rm T}\boldsymbol\beta_0^* - y_i) - \tau_0\}\boldsymbol{x}_i\right\|_{\Sigma^{-1}} \\
&=: \mathrm{I}_0 + \mathrm{II}_0,
\end{aligned}$$

and for $j \geq 1$,

$$
\begin{aligned}
&||\,\widetilde{e}^{\flat}_{\rm int}(\tau_j) + Q^{*\flat}_j||_{\Sigma^{-1}} \\
&\leq ||\,\widetilde{e}_{\rm int}(\tau_j) + Q^{*}_j||_{\Sigma^{-1}} + \Bigg\| \mathbf{J}_j(\widehat{\boldsymbol\beta}^{\flat}_j - \boldsymbol\beta^*_j) + \sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}} \mathbf{H}_\ell(\widehat{\boldsymbol\beta}^{\flat}_\ell - \boldsymbol\beta^*_\ell){\rm d}H(u) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n W_i\bigg\{\Delta_i \bar K_h(\boldsymbol{x}_i^{\rm T}\boldsymbol\beta^*_j - y_i) - \sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}} \bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol\beta^*_\ell){\rm d}H(u) - \tau_0\bigg\}\boldsymbol{x}_i\Bigg\|_{\Sigma^{-1}} \\
&=: {\rm I}_j + {\rm II}_j,
\end{aligned}
$$

where $\widetilde{e}_{\rm int}(\tau_j)$ and Q^*_j are defined in (B.17) and (B.19)–(B.20). Let $\mathcal{E}_2$ be the event that (14)–(16) hold. Then $\mathbb{P}(\mathcal{E}_2) \geq 1 - C_2 n^{-1}$ for some constant C_2, and conditioned on $\mathcal{E}_2$,

$$
\max_{0\leq j\leq m} {\rm I}_j = ||\,\widetilde{e}_{\rm int}(\tau_j) + Q^*_j||_{\Sigma^{-1}} \lesssim m_4^{1/2}\frac{p+\log n}{nh^{1/2}} + h\sqrt{\frac{p+\log n}{n}}. \tag{C.12}
$$

It remains to bound ${\rm II}_j$ for $j = 0, 1, \ldots, m$. Recall that $\widehat{Q}^{\flat}_j(\widehat{\boldsymbol\beta}^{\flat}_j) = \mathbf{0}$, we have

$$
{\rm II}_0 = \|\,\mathbf{J}_0(\widehat{\boldsymbol\beta}^{\flat}_0 - \boldsymbol\beta^*_0) - \{\widehat{Q}^{\flat}_0(\widehat{\boldsymbol\beta}^{\flat}_0) - \widehat{Q}^{\flat}_0(\boldsymbol\beta^*_0)\}\,\|_{\Sigma^{-1}}
$$

and for $j = 1, \ldots, m$,

$$
\begin{aligned}
{\rm II}_j &\leq \|\,\mathbf{J}_j(\widehat{\boldsymbol\beta}^{\flat}_j - \boldsymbol\beta^*_j) - \{\widehat{Q}^{\flat}_j(\widehat{\boldsymbol\beta}^{\flat}_j) - \widehat{Q}^{\flat}_j(\boldsymbol\beta^*_j)\}\,\|_{\Sigma^{-1}} \\
&\quad + \Bigg\|\frac{1}{n}\sum_{i=1}^n\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}{\rm d}H(u)\Big[\mathbf{H}_\ell(\widehat{\boldsymbol\beta}_\ell - \boldsymbol\beta^*_\ell) - W_i\big\{\bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol\beta^*_\ell) + \bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\widehat{\boldsymbol\beta}_\ell)\boldsymbol{x}_i\big\}\Big]\Bigg\|_{\Sigma^{-1}}.
\end{aligned}
$$

Putting together the pieces, and taking $r \asymp \sqrt{(p+\log n)/n}$, we conclude that conditioning on $\mathcal{E}_1 \cap \mathcal{E}_2$,

(C.13)

$$
\begin{aligned}
{\rm II}_j &\leq \sup_{\boldsymbol\beta\in\boldsymbol\beta^*_j+\Theta(r)} \|\,\mathbf{J}_j(\boldsymbol\beta - \boldsymbol\beta^*_j) - \{\widehat{Q}^{\flat}_j(\boldsymbol\beta) - \widehat{Q}^{\flat}_j(\boldsymbol\beta^*_j)\}\,\|_{\Sigma^{-1}} \\
&+ \sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol\beta_\ell\in\boldsymbol\beta^*_\ell+\Theta(r)\}} \Bigg\|\frac{1}{n}\sum_{i=1}^n\sum_{\ell=0}^{j-1}\int_{\tau_\ell}^{\tau_{\ell+1}}{\rm d}H(u)\Big[\mathbf{H}_\ell(\boldsymbol\beta_\ell - \boldsymbol\beta^*_\ell) - W_i\big\{\bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol\beta^*_\ell) + \bar K_h(y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol\beta_\ell)\boldsymbol{x}_i\big\}\Big]\Bigg\|_{\Sigma^{-1}}
\end{aligned}
$$

holds with $\mathbb{P}^*$-probability at least $1 - C_3 n^{-1}$.

Let $\mathcal{E}_3$–$\mathcal{E}_5$ be the events from Lemmas C.3–C.5, so that $\mathbb{P}(\mathcal{E}_3 \cap \mathcal{E}_4 \cap \mathcal{E}_5) \geq 1 - C_4 n^{-1}$. Applying Lemmas C.4 and C.5 to (C.13) yields that for any $j = 0, 1, \ldots, m$,

$$
\begin{aligned}
{\rm II}_j \lesssim \bigg\{ & m_4^{1/2}\frac{p+\log n}{nh^{1/2}} + h\sqrt{\frac{p+\log n}{n}} \\
& + \zeta_p^2\frac{(p+\log n)(p\log n)^{1/2}}{n^{3/2}h} + \max_{0\leq\ell\leq j-1}\Gamma_\ell(r)\bigg\}\cdot\Big\{\log\Big(\tfrac{1-\tau_0}{1-\tau_j}\Big)\vee 1\Big\}
\end{aligned} \tag{C.14}
$$

holds with $\mathbb{P}^*$-probability at least $1 - n^{-2}$ conditioned on $\mathcal{E}_4 \cap \mathcal{E}_5$, where $\Gamma_\ell(r)$ is defined in (C.2). Note that $\log\big(\frac{1-\tau_0}{1-\tau_j}\big) \leq \log\big(\frac{1-\tau_0}{1-\tau_m}\big)$ is bounded by a constant. For $\Gamma_\ell(r)$, it follows from Lemma C.3 with $r \asymp \sqrt{(p+\log n)/n}$ that conditioned on $\mathcal{E}_3$,

$$
\max_{0\leq\ell\leq m}\Gamma_\ell(r) \lesssim m_4^{1/2}\frac{p+\log n}{nh^{1/2}} + \zeta_p^2\frac{(p+\log n)(p\log n)^{1/2}}{n^{3/2}h} \tag{C.15}
$$

holds with $\mathbb{P}^*$-probability at least $1 - C_5 n^{-1}$ provided $n \gtrsim \zeta_p^2 \log n$.

Finally, define the event $\mathcal{F} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4 \cap \mathcal{E}_5$, satisfying $\mathbb{P}(\mathcal{F}) \geq 1 - Cn^{-1}$ for some constant $C > 0$ independent of (n, p). Combining (C.12), (C.14) and (C.15) proves (C.10), as claimed.

STEP 2. The arguments from Stage Two in the proof of Theorem 3.2 can be similarly applied to bridge the gap between discrete and continuous uniform bounds. Thus the details are omitted. □

C.4. Proof of Theorem 3.6. Without loss of generality, we assume $||\boldsymbol{a}_n||_\Sigma = 1$ throughout the proof; otherwise, we first rescale the vectors $\boldsymbol{a}_n$ so that the same arguments apply.

Conditioned on the observed data $\mathbb{D}_n = \{(y_i, \Delta_i, \boldsymbol{x}_i)\}_{i=1}^n$, we have $\mathbb{E}^* \langle \boldsymbol{a}_n, \boldsymbol{U}_i^\flat(\tau)\rangle = 0$ for any $\tau \in [\tau_L, \tau_U]$. By the asymptotic (conditional) tightness established in Lemma C.6 and the central limit theorem, the limiting distribution of $\mathbb{G}_n^\flat(\cdot)$ given $\mathbb{D}_n$ is a zero-mean Gaussian process. Following the arguments in Appendix 1 of [10], it suffices to show that the conditional covariance function of $\mathbb{G}_n^\flat(\cdot)$ given $\mathbb{D}_n$ converges to $H(\cdot,\cdot)$ defined in (21), which is the limit of the (unconditional) covariance function of $\mathbb{G}_n(\cdot)$. To this end, for any $s, t \in [\tau_L, \tau_U]$, note that

$$\begin{aligned}
\mathrm{Cov}^*\left(\mathbb{G}_n^\flat(s), \mathbb{G}_n^\flat(t)\right) &= \mathbb{E}^*\{\mathbb{G}_n^\flat(s)\mathbb{G}_n^\flat(t)\} \\
&= \frac{1}{n}\mathbb{E}^*\left\{\sum_{i=1}^n \langle \boldsymbol{a}_n, e_i \boldsymbol{U}_i(s)\rangle\right\} \cdot \left\{\sum_{i=1}^n \langle \boldsymbol{a}_n, e_i \boldsymbol{U}_i(t)\rangle\right\} \\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{a}_n^{\mathrm{T}} \boldsymbol{U}_i(s) \boldsymbol{U}_i(t)^{\mathrm{T}} \boldsymbol{a}_n \xrightarrow{\text{a.s.}} H(s,t),
\end{aligned}$$

where the almost sure convergence follows from the strong law of large numbers. This completes the proof. □

APPENDIX D: PROOF OF THEOREM 4.1

For $j = 0, 1, \ldots,$ and $r, q > 0$, define

(D.1)

$$\psi_j(r,q) = \sup_{\boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r) \cap \Lambda(q)} \left\| \frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\}\boldsymbol{x}_i \right\|_\infty,$$

where $\Lambda(q) := \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\|_1 \leq q\|\boldsymbol{u}\|_\Sigma\}$ is a cone-like set.

D.1. Technical lemmas.

LEMMA D.1. *Let $j = 0, 1, \ldots, m$, and $t > 0$.*

(i) *With probability at least $1 - e^{-t}$,*

$$\|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty \leq \bar{\tau}_0 \left\{\sigma\sqrt{\frac{2t + 2\log(2p)}{n}} + \frac{t + \log(2p)}{3n}\right\},$$

where $\bar{\tau}_0 = \max(\tau_0, 1-\tau_0)$ and $\sigma^2 = \max_{1 \leq k \leq p} \sigma_{kk} \leq 1$.

(ii) *With probability at least $1 - e^{-t}$,*

$$\left\|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\boldsymbol{x}_i\right\|_\infty \leq \sigma\sqrt{\frac{2t + 2\log(2p)}{n}} + \frac{t + \log(2p)}{3n}.$$

PROOF. It suffices to prove part (i) since the second inequality can be obtained from the same argument. Fix j, we have

$$\|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty = \max_{1\le k\le p}\left|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\underbrace{\left\{\Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i) - \tau_0\right\}}_{=:\,\xi_{ij}} x_{ik}\right|,$$

where $\xi_{ij} = \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i) - \tau_0$ is such that $|\xi_{ij}| \le \bar{\tau}_0 = \max(\tau_0, 1-\tau_0)$ and $\mathbb{E}(\xi_{ij}x_{ik})^2 \le \bar{\tau}_0^2\sigma_{kk}$. Applying Bernstein's inequality yields that, with probability at least $1-2e^{-z}$,

$$\left|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\xi_{ij}x_{ik}\right| \le \bar{\tau}_0\left(\sigma_{kk}^{1/2}\sqrt{\frac{2z}{n}} + \frac{z}{3n}\right) \quad \text{for any } 1\le k\le p.$$

The claimed bound then follows by taking $z = t + \log(2p)$ and the union bound. □

LEMMA D.2. *For any $t>0$, we have that with probability at least $1-e^{-t}$,*

$$\psi_j(r,q) \lesssim \left(\frac{q}{h}\sqrt{\frac{\log p}{n}} + \overline{f}^{1/2}\sqrt{\frac{t+\log p}{nh}}\right)\cdot r + \frac{t+\log p}{n}.$$

PROOF. For any j fixed, and $k = 1,\ldots,p$, define after a change of variable $\boldsymbol{v} = \boldsymbol{\beta} - \boldsymbol{\beta}_j^*$ that

$$\psi_{j,k}(r,q) = \sup_{\boldsymbol{v}\in\Theta(r)\cap\Lambda(q)}\left|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\underbrace{\left\{\bar{K}_h(\varepsilon_{ij} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}) - \bar{K}_h(\varepsilon_{ij})\right\}x_{ik}}_{=:\,g_{\boldsymbol{v}}(y_i,\boldsymbol{x}_i)}\right|,$$

where $\varepsilon_{ij} = y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*$. Then $\psi_j(r,q) \le \max_{1\le k\le p}\psi_{j,k}(r,q)$. Note that $\sup_{\boldsymbol{v}}|g_{\boldsymbol{v}}(y_i,\boldsymbol{x}_i)| \le |x_{ik}| \le 1$. Let σ be any positive number such that $\sigma^2 \ge \sup_{\boldsymbol{v}\in\Theta(r)\cap\Lambda(q)}\mathbb{E}g_{\boldsymbol{v}}^2(y_i,\boldsymbol{x}_i)$. By Theorem 7.3 in [3]—an improved version of Talagrand's inequality, we obtain that for any $z>0$,

$$\psi_{j,k}(r,q) \le \mathbb{E}\psi_{j,k}(r,q) + \sqrt{\left\{\sigma^2 + 2\mathbb{E}\psi_{j,k}(r,q)\right\}\frac{2z}{n}} + \frac{z}{3n} \tag{D.2}$$

holds with probability at least $1-e^{-z}$. For the second moment $\mathbb{E}g_{\boldsymbol{v}}^2(y_i,\boldsymbol{x}_i)$, by a change of variable and Minkowski's integral inequality we derive that

$$\begin{aligned}
\mathbb{E}g_{\boldsymbol{v}}^2(y_i,\boldsymbol{x}_i) &= \mathbb{E}\left[x_{ik}^2\int_{-\infty}^{\infty}\left\{\bar{K}_h(u-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}) - \bar{K}_h(u)\right\}^2 f_y(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* + u|\boldsymbol{x}_i)\mathrm{d}u\right]\\
&= h\mathbb{E}\left[x_{ik}^2\int_{-\infty}^{\infty}\left\{\bar{K}(v-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h) - \bar{K}(v)\right\}^2 f_y(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* + vh|\boldsymbol{x}_i)\mathrm{d}v\right]\\
&\le \overline{f}h^{-1}\mathbb{E}\left[x_{ik}^2(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2\int_{-\infty}^{\infty}\left\{\int_0^1 K(v - w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h)\mathrm{d}w\right\}^2\mathrm{d}v\right]\\
&\le \overline{f}h^{-1}\mathbb{E}\left(x_{ik}^2(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2\left[\int_0^1\left\{\int_{-\infty}^{\infty}K^2(v - w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h)\mathrm{d}v\right\}^{1/2}\mathrm{d}w\right]^2\right)\\
&\le \kappa_u\overline{f}h^{-1}\mathbb{E}(x_{ik}\cdot\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2 \le \kappa_u\overline{f}h^{-1}r^2, \quad \text{valid for any } \boldsymbol{v}\in\Theta(r).
\end{aligned}$$

It remains to bound $\mathbb{E}\psi_{j,k}(r,q)$ in the concentration inequality (D.2). Note that $g_{\boldsymbol{v}}(y_i,\boldsymbol{x}_i)$ is (κ_u/h)-Lipschitz continuous in $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}$, i.e., for any $\boldsymbol{v},\boldsymbol{v}'$, $|g_{\boldsymbol{v}}(y_i,\boldsymbol{x}_i) - g_{\boldsymbol{v}'}(y_i,\boldsymbol{x}_i)| \le$

$(\kappa_u/h)|\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}'|$. Hence, it follows from Rademacher symmetrization and Talagrand's contraction principle that

$$\mathbb{E}\psi_{j,k}(r,q) \le 2\mathbb{E}\left\{ \sup_{\boldsymbol{v}\in\Theta(r)\cap\Lambda(q)} \left| \frac{1}{n}\sum_{i=1}^n e_i g_{\boldsymbol{v}}(y_i, \boldsymbol{x}_i) \right| \right\}$$

$$\le 4\kappa_u \mathbb{E}\left\{ \sup_{\boldsymbol{v}\in\Theta(r)\cap\Lambda(q)} \left| \frac{1}{nh}\sum_{i=1}^n e_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v} \right| \right\} \le 4\kappa_u \frac{qr}{nh}\cdot \mathbb{E}\left\| \sum_{i=1}^n e_i \boldsymbol{x}_i \right\|_\infty,$$

where $e_1, \ldots, e_n$ are independent Rademacher variables. By Hoeffding's moment inequality,

$$\mathbb{E}_e \left\| \sum_{i=1}^n e_i \boldsymbol{x}_i \right\|_\infty \le \max_{1\le k\le p} \left(\sum_{i=1}^n x_{ik}^2 \right)^{1/2} \sqrt{2\log(2p)},$$

where $\mathbb{E}_e$ denotes the expectation over $\{e_i\}_{i=1}^n$. Plugging this into the previous bound yields

$$\mathbb{E}\psi_{j,k}(r,q) \le 4\kappa_u \frac{qr}{h}\sqrt{\frac{2\log(2p)}{n}}.$$

Finally, the claimed result follows by taking $z = t + \log p$ in (D.2) and the union bound. □

The following result extends the restricted strong property in Lemma B.4 to high dimensions. It follows from Proposition 4.2 in [13] with slight modifications.

LEMMA D.3. *Assume Conditions 3.1–3.3 hold, and let $h, r > 0$ satisfy $4\eta_{1/4} r \le h \le \underline{g}/(2l_1)$ with $\eta_{1/4}$ defined in (B.1). Then, for any $0 \le j \le m$ and $t > 0$,*

$$\mathbb{P}\left\{ D(\boldsymbol{\beta}, \boldsymbol{\beta}_j^*) \ge \frac{1}{2}\underline{g}\kappa_l \cdot ||\boldsymbol{\beta} - \boldsymbol{\beta}_j^*||_\Sigma^2 \text{ for all } \boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r)\cap\Lambda(q) \right\} \ge 1 - e^{-t}$$

provided that $n \gtrsim h(q/r)^2(t \vee \log p)$.

D.2. Proof of the theorem. Following the argument as in the proof of Theorem 3.1, it suffices to derive a uniform bound on the grid of τ-levels, $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$. Again, we start with constructing intermediate points $\{\widetilde{\boldsymbol{\beta}}_j = (1-u_j)\boldsymbol{\beta}_j^* + u_j\widehat{\boldsymbol{\beta}}_j\}_{j=0,1,\ldots,m}$ that satisfy $\widetilde{\boldsymbol{\beta}}_j \in \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit)$, where $r^\diamondsuit = h/(4\eta_{1/4})$. For each $\widehat{\boldsymbol{\beta}}_j$, by the first-order optimality condition, there exists some subgradient $\widehat{\boldsymbol{g}}_j \in \partial\|\widehat{\boldsymbol{\beta}}_j\|_1$ such that $\widehat{Q}_j(\widehat{\boldsymbol{\beta}}_j) + \lambda_j \cdot \widehat{\boldsymbol{g}}_j = \mathbf{0}$ and $\langle \widehat{\boldsymbol{g}}_j, \widehat{\boldsymbol{\beta}}_j\rangle = \|\widehat{\boldsymbol{\beta}}_j\|_1$. Consequently, for each $j = 0, 1, \ldots, m$,

$$D(\widetilde{\boldsymbol{\beta}}_j, \boldsymbol{\beta}_j^*) \le u_j D(\widehat{\boldsymbol{\beta}}_j, \boldsymbol{\beta}_j^*) = u_j\langle -\lambda_j\widehat{\boldsymbol{g}}_j - \widehat{Q}_j(\boldsymbol{\beta}_j^*), \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\rangle$$

$$\le \lambda_j\big(\|(\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)_{\mathcal{S}_j}\|_1 - \|(\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*)_{\mathcal{S}_j^c}\|_1\big) + \langle -\widehat{Q}_j(\boldsymbol{\beta}_j^*), \widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\rangle, \tag{D.3}$$

where $\mathcal{S}_j = \mathrm{supp}(\boldsymbol{\beta}_j^*)$. Denote the cardinality of $\mathcal{S}_j$ by s_j, satisfying $s_j \le s$ for all j. Consider the decomposition

$$\widehat{Q}_j(\boldsymbol{\beta}_j^*) = \widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=1}^{j-1} w_\ell(\widetilde{\Delta}_\ell + \Delta_\ell)$$

$$+ Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}\} + \sum_{\ell=0}^{j-1} w_\ell B_\ell(\widehat{\boldsymbol{\beta}}_\ell),$$

where Δ_ℓ ($\ell = 0, 1, \ldots, m-1$) are given in (B.11),

$$\widetilde{\Delta}_\ell = \frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\ell) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\big\}\boldsymbol{x}_i,$$

$$\text{and} \;\; B_\ell(\boldsymbol{\beta}) = \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{x}\} - \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}\}.$$

Recall from the proof of Lemma B.5 that $\|B_\ell(\boldsymbol{\beta})\|_{\Sigma^{-1}} \le \overline{f}\|\boldsymbol{\beta} - \boldsymbol{\beta}_\ell^*\|_\Sigma$. Then, by Hölder's inequality,

$$\begin{aligned}
\text{(D.4)} \qquad & |\langle \widehat{Q}_j(\boldsymbol{\beta}_j^*), \widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\rangle| \\
&\le \left\{ \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty + \sum_{\ell=0}^{j-1} w_\ell\big(\|\widetilde{\Delta}_\ell\|_\infty + \|\Delta_\ell\|_\infty\big)\right\}\|\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_1 \\
&\quad + \underbrace{\left\| Q_0(\boldsymbol{\beta}_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\boldsymbol{x}\}\right\|_{\Sigma^{-1}}}_{<(1+W_j)a \text{ by (B.13)}} \|\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma \\
&\quad + \overline{f}\sum_{\ell=0}^{j-1} w_\ell \|\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell^*\|_\Sigma \cdot \|\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j^*\|_\Sigma, \;\; j = 1, \ldots, m,
\end{aligned}$$

and $|\langle \widehat{Q}_0(\boldsymbol{\beta}_0^*), \widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\rangle| \le \|\widehat{Q}_0(\boldsymbol{\beta}_0^*) - Q_0(\boldsymbol{\beta}_0^*)\|_\infty \|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\|_1 + a\|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\|_\Sigma$, where

$$\text{(D.5)} \qquad a = 0.5 l_1 \kappa_2 h^2 + \overline{f}\underline{f}^{-1}\delta^* \;\; \text{and} \;\; W_j = \sum_{\ell=0}^{j-1} w_\ell = \int_{\tau_L}^{\tau_j} \mathrm{d}H(u) = \log\big(\tfrac{1-\tau_L}{1-\tau_j}\big).$$

For some positive sequence $\{q_j\}_{j=0,1,\ldots,m}$ and curvature parameter $\kappa > 0$ to be determined, define the "good" events

$$\begin{aligned}
\mathcal{G} &= \left\{ \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty \le \frac{\lambda_0}{2}\right\} \\
&\quad \bigcap \left\{ \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty + \sum_{\ell=0}^{j-1} w_\ell\|\Delta_\ell\|_\infty \le \frac{\lambda_j}{3}, \, j = 1, \ldots, m\right\} \text{ and} \\
\mathcal{F} &= \bigcap_{j=0}^m \left\{ D(\boldsymbol{\beta}, \boldsymbol{\beta}_j^*) \ge \kappa \cdot \|\boldsymbol{\beta} - \boldsymbol{\beta}_j^*\|_\Sigma^2 \;\text{ for all }\; \boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r^\diamondsuit) \cap \Lambda(q_j)\right\}.
\end{aligned}$$

Conditioned on $\mathcal{F} \cap \mathcal{G}$, it follows from (D.3) that

$$\text{(D.6)} \qquad 0 \le D(\widetilde{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0^*) < \frac{\lambda_0}{2}\big(3\|(\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)_{\mathcal{S}_0}\|_1 - \|(\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)_{\mathcal{S}_0^{\mathrm{c}}}\|_1\big) + a\|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*\|_\Sigma,$$

thus implying the cone-like constraint $\|(\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)_{\mathcal{S}_0^{\mathrm{c}}}\|_1 \le 3\|(\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*)_{\mathcal{S}_0}\|_1 + (2a/\lambda_0)\|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*\|_\Sigma$. Taking $q_0 = 4(s_0/\underline{\gamma})^{1/2} + 2a/\lambda_0$, we see that $\widetilde{\boldsymbol{\beta}}_0$ falls into the cone-like set $\boldsymbol{\beta}_0^* + \Lambda(q_0)$, and so does $\widehat{\boldsymbol{\beta}}_0$. Hence, $D(\widetilde{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0^*) \ge \kappa \cdot \|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\|_\Sigma^2$. Combining this with (D.6) yields, after some algebra, that

$$\text{(D.7)} \qquad \|\widetilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0^*\|_\Sigma < r_0 := \kappa^{-1}\big\{1.5(s/\underline{\gamma})^{1/2}\lambda_0 + a\big\}.$$

Provided $r_0 \le r^\diamondsuit$, $\widetilde{\boldsymbol\beta}_0$ lies in the interior of the local region $\boldsymbol\beta_0^* + \Theta(r^\diamondsuit)$. As before, we argue by contradiction that $\widehat{\boldsymbol\beta}_0$ coincides with $\widetilde{\boldsymbol\beta}_0$, thus implying $\widehat{\boldsymbol\beta}_0 \in \boldsymbol\beta_0^* + \Theta(r_0) \cap \Lambda(q_0)$.

For $(\widetilde{\boldsymbol\beta}_1, \widehat{\boldsymbol\beta}_1)$, from (D.3) and (D.4) it follows that

$$\begin{aligned} 0 &\le D(\widetilde{\boldsymbol\beta}_1, \boldsymbol\beta_1^*) \\ &< \lambda_1\big(\|(\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*)_{\mathcal S_1}\|_1 - \|(\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*)_{\mathcal S_1^c}\|_1\big) \\ &\quad + \big(\lambda_1/3 + w_0\|\widetilde\Delta_0\|_\infty\big)\|\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*\|_1 + (a + w_0 a + \overline f w_0 \|\widehat{\boldsymbol\beta}_0 - \boldsymbol\beta_0^*\|_\Sigma)\|\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*\|_\Sigma. \end{aligned}$$

We have already shown that $\widehat{\boldsymbol\beta}_0 \in \boldsymbol\beta_0^* + \Theta(r_0) \cap \Lambda(q_0)$ conditioning on $\mathcal F \cap \mathcal G$. Then $\|\widetilde\Delta_0\|_\infty \le \psi_0(r_0, q_0)$, where $\psi_j(\cdot,\cdot)$ is defined in (D.1). Conditioned further on $\{w_0\psi_0(r_0,q_0) \le \lambda_1/3\}$, we have

$$\begin{aligned} 0 &\le D(\widetilde{\boldsymbol\beta}_1, \boldsymbol\beta_1^*) \\ &< \frac{\lambda_1}{3}\big(5\|(\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*)_{\mathcal S_1}\|_1 - \|(\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*)_{\mathcal S_1^c}\|_1\big) + \{a + w_0(\overline f r_0 + a)\}\|\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*\|_\Sigma, \end{aligned}$$

which in turn implies $\widetilde{\boldsymbol\beta}_1 \in \boldsymbol\beta_1^* + \Lambda(q_1)$ with $q_1 := 6(s_1/\underline\gamma)^{1/2} + 3\{(1+w_0)a + \overline f w_0 r_0\}/\lambda_1$. On the event $\mathcal F$, $D(\widetilde{\boldsymbol\beta}_1, \boldsymbol\beta_1^*) \ge \kappa \cdot \|\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*\|_\Sigma^2$. Combining the upper and lower bounds yields

$$\|\widetilde{\boldsymbol\beta}_1 - \boldsymbol\beta_1^*\|_\Sigma < r_1 := \kappa^{-1}\big\{\tfrac53 (s/\underline\gamma)^{1/2}\lambda_1 + (1+w_0)a + w_0\overline f r_0\big\}. \tag{D.8}$$

Provided $r_1 \le r^\diamondsuit$, we reach the conclusion that $\widehat{\boldsymbol\beta}_1 \in \boldsymbol\beta_1^* + \Theta(r_1) \cap \Lambda(q_1)$.

We now recurse this argument, in particular controlling the error terms $\|\widehat\Delta_\ell\|_\infty$ sequentially, so that at the j-th step $(1 \le j \le m)$, $\widetilde{\boldsymbol\beta}_j$ satisfies

$$\begin{aligned} 0 &\le D(\widetilde{\boldsymbol\beta}_j, \boldsymbol\beta_j^*) \\ &< \lambda_j\big(\|(\widetilde{\boldsymbol\beta}_j - \boldsymbol\beta_j^*)_{\mathcal S_j}\|_1 - \|(\widetilde{\boldsymbol\beta}_j - \boldsymbol\beta_j^*)_{\mathcal S_j^c}\|_1\big) \\ &\quad + \left(\frac{\lambda_j}{3} + \sum_{\ell=0}^{j-1} w_\ell\|\widetilde\Delta_\ell\|_\infty\right)\|\widetilde{\boldsymbol\beta}_j - \boldsymbol\beta_j^*\|_1 + \left\{a + \sum_{\ell=0}^{j-1} w_\ell(\overline f r_\ell + a)\right\}\|\widetilde{\boldsymbol\beta}_j - \boldsymbol\beta_j^*\|_\Sigma. \end{aligned}$$

Conditioning on the event $\{\sum_{\ell=0}^{j-1} w_\ell\psi_\ell(r_\ell, q_\ell) \le \lambda_j/3\}$, we obtain the cone-like constraint $\widetilde{\boldsymbol\beta}_j \in \boldsymbol\beta_j^* + \Lambda(q_j)$ with $q_j := 6(s_j/\underline\gamma)^{1/2} + 3\{(1+W_j)a + \overline f \sum_{\ell=0}^{j-1} w_\ell r_\ell\}/\lambda_j$, thus implying

$$\|\widetilde{\boldsymbol\beta}_j - \boldsymbol\beta_j^*\|_\Sigma < r_j := \kappa^{-1}\left\{\tfrac53 (s/\underline\gamma)^{1/2}\lambda_j + (1+W_j)a + \overline f \sum_{\ell=0}^{j-1} w_\ell r_\ell\right\}. \tag{D.9}$$

In view of (D.7)–(D.9), we write

$$A_j = \frac{1}{\kappa}\big\{\tfrac53 (s/\underline\gamma)^{1/2}\lambda_j + (1+W_j)a\big\}, \;\; c_j = \kappa^{-1}\overline f w_j, \;\; j = 0, 1, \ldots,$$

where $W_0 = 0$, so that the sequence of radii $\{r_j\}_{j\ge0}$ satisfies $r_j \le A_j + \sum_{\ell=0}^{j-1} c_\ell r_\ell$. Since $A_0 \le A_1 \le A_2 \le \cdots$, we have

$$r_j \le \prod_{\ell=0}^{j-1}(1 + c_\ell)\cdot A_j \le e^{\sum_{\ell=0}^{j-1} c_\ell}\cdot A_j = C_j \cdot A_j \;\text{ with }\; C_j := \left(\frac{1-\tau_L}{1-\tau_j}\right)^{\overline f/\kappa}.$$

In particular, we write $C_0 = 1$. As long as $\max_{0\le j\le m} r_j \le r^{\diamondsuit}$, we have established the result $\widehat{\boldsymbol{\beta}}_j \in \boldsymbol{\beta}_j^* + \Theta(r_j) \cap \Lambda(q_j)$ $(j = 0, 1, \ldots, m)$ as a deterministic claim, conditioned on the event

$$\mathcal{F} \cap \mathcal{G} \cap \bigcap_{j=1}^{m} \left\{ \sum_{\ell=0}^{j-1} w_\ell \psi_\ell(r_\ell, q_\ell) \le \frac{\lambda_j}{3} \right\}$$

with properly chosen $\kappa > 0$ and regularization parameters $\lambda_0, \lambda_1, \ldots, \lambda_m$, where $q_0 = 4(s_0/\underline{\gamma})^{1/2} + 2a/\lambda_0$ and $q_j = 6(s_j/\underline{\gamma})^{1/2} + 3\{(1+W_j)a + \overline{f}\sum_{\ell=0}^{j-1} w_\ell r_\ell\}/\lambda_j$ for $j \ge 1$ with W_j given in (D.5).

Next we choose $\{\lambda_j\}_{j=0,1,\ldots,m}$ in a sequential manner so that the above good event occurs with high probability. Applying the two inequalities in Lemma D.1, both with $t = 2\log p$, implies that with probability at least $1 - 2(m+1)p^{-2}$,

$$\|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty + \sum_{\ell=0}^{j-1} w_\ell \|\Delta_\ell\|_\infty \lesssim (1+W_j)\sqrt{\frac{\log p}{n}} \quad \text{for all } j = 0, 1, \ldots, m,$$

where $W_0 \equiv 0$. Throughout, assume the following upper bound constraint on the magnitude of h:

$$h^2 \lesssim (s/\lambda_l)^{1/2}\sqrt{\frac{\log p}{n}}.$$

Starting at $j = 0$, set $\lambda_0 \asymp \sqrt{\log(p)/n}$ so that $q_0 \lesssim (s/\underline{\gamma})^{1/2}$ and $r_0 = \kappa^{-1}\{1.5(s/\underline{\gamma})^{1/2}\lambda_0 + a\} \lesssim \kappa^{-1}(s/\underline{\gamma})^{1/2}\lambda_0$. With this choice of (λ_0, r_0, q_0), it follows from Lemma D.2 with $t = 2\log p$ that, with probability at least $1 - p^{-2}$,

$$\psi_0(r_0, q_0) \lesssim \frac{s\lambda_0}{\kappa\underline{\gamma}h}\sqrt{\frac{\log p}{n}} + \frac{\log p}{n}.$$

Recall that $a \asymp h^2 + \delta^* \lesssim h^2 + n^{-1/2}$. We then choose $\lambda_1 \asymp (1+W_1)\sqrt{\log(p)/n}$ so that $\lambda_1 \ge 3\max\left\{w_0\psi_0(r_0, q_0), \|\widehat{Q}_0(\boldsymbol{\beta}_1^*) - Q_0(\boldsymbol{\beta}_1^*)\|_\infty + w_0\|\Delta_\ell\|_\infty\right\}$ as long as $h \gtrsim (\kappa\underline{\gamma})^{-1}s\sqrt{\log(p)/n}$. Furthermore, it follows that

$$q_1 \lesssim C_1(s/\underline{\gamma})^{1/2} \quad \text{and} \quad r_1 \lesssim C_1(s/\underline{\gamma})^{1/2}\lambda_1/\kappa.$$

For a general $j \ge 2$, assume we already have $\lambda_\ell \asymp (1+W_\ell)\sqrt{\log(p)/n}$, $q_\ell \lesssim C_\ell(s/\underline{\gamma})^{1/2}$ and $r_\ell \lesssim C_\ell(s/\underline{\gamma})^{1/2}\lambda_\ell/\kappa$ for $\ell = 1, \ldots, j-1$. And with probability at least $1 - jp^{-2}$,

$$\psi_\ell(r_\ell, q_\ell) \lesssim \frac{q_\ell}{h}\sqrt{\frac{\log p}{n}}r_\ell + \frac{\log p}{n}, \quad \ell = 0, 1, \ldots, j-1.$$

The accumulated error can thus be bounded by

$$\sum_{\ell=0}^{j-1} w_\ell \psi_\ell(r_\ell, q_\ell) \lesssim (1+W_{j-1})\sum_{\ell=0}^{j-1} w_\ell C_\ell^2 \frac{s\log p}{\kappa\underline{\gamma}nh} + W_j\frac{\log p}{n}.$$

Provided that $h \gtrsim (\kappa\underline{\gamma})^{-1}s\sqrt{\log(p)/n}$,

$$(1+W_j)\sqrt{\frac{\log p}{n}} \asymp \lambda_j \ge 3\max\left\{\sum_{\ell=0}^{j-1} w_\ell\psi_\ell(r_\ell, q_\ell), \|\widehat{Q}_0(\boldsymbol{\beta}_j^*) - Q_0(\boldsymbol{\beta}_j^*)\|_\infty + \sum_{\ell=0}^{j-1} w_\ell\|\Delta_\ell\|_\infty\right\}.$$

and therefore the event that involves λ_j is certified.

With the above choice of $\{r_j\}_{j=0,1,\ldots,m}$ and the lower bound constraint on the magnitude of the bandwidth—$h \gtrsim (\kappa\underline{\gamma})^{-1}s\sqrt{\log(p)/n}$, we have

$$C_j\kappa^{-1}(1+W_j)(s/\underline{\gamma})^{1/2}\sqrt{\frac{\log p}{n}} \asymp r_j \le r^{\diamondsuit} \asymp h \text{ for all } j=0,1,\ldots,m.$$

Finally, by the restricted strong convexity lemma—Lemma D.3 with $r = h/(4\eta_{1/4})$, $q \asymp (s/\underline{\gamma})^{1/2}$ and $t = 2\log p$—we take $\kappa = (\underline{g}\kappa_l)/2$ so that event $\mathcal{F}$ happens with probability at least $1-(m+1)p^{-2}$ provided that the "effective" sample size satisfies $nh \gtrsim s\log p$. □

APPENDIX E: PROOF OF TECHNICAL LEMMAS

This section contains the proofs of all technical lemmas from Sections B and C.

E.1. Proof of Lemma B.1. Fix $\boldsymbol{\beta}, \boldsymbol{\beta}' \in \mathbb{R}^p$, and define the function $f(\eta) = \langle Q(\boldsymbol{\beta}_\eta) - Q(\boldsymbol{\beta}'), \boldsymbol{\beta}-\boldsymbol{\beta}'\rangle$ for $\eta \in [0,1]$. Since $Q(\cdot)$ is differentiable with a positive semi-definite Jacobian, we have $f'(\eta) = \langle \nabla Q(\boldsymbol{\beta}_\eta)(\boldsymbol{\beta}-\boldsymbol{\beta}'), \boldsymbol{\beta}-\boldsymbol{\beta}'\rangle \ge 0$, and hence $f(\cdot)$ is non-decreasing. Consequently, for any $\eta \in [0,1]$,

$$\begin{aligned} D(\boldsymbol{\beta}_\eta, \boldsymbol{\beta}^*) &= \langle Q(\boldsymbol{\beta}_\eta) - Q(\boldsymbol{\beta}'), \boldsymbol{\beta}_\eta - \boldsymbol{\beta}'\rangle = \eta\langle Q(\boldsymbol{\beta}_\eta) - Q(\boldsymbol{\beta}'), \boldsymbol{\beta}-\boldsymbol{\beta}'\rangle \\ &= \eta f(\eta) \le \eta f(1) = \eta\langle Q(\boldsymbol{\beta}) - Q(\boldsymbol{\beta}'), \boldsymbol{\beta}-\boldsymbol{\beta}'\rangle = \eta D(\boldsymbol{\beta}, \boldsymbol{\beta}'), \end{aligned}$$

as claimed. □

E.2. Proof of Lemma B.2. First, by the variational representation of $\|\cdot\|_2$,

$$\Delta := \left\|\frac{1}{n}\sum_{i=1}^n(\xi_i\boldsymbol{z}_i - \mathbb{E}\xi_i\boldsymbol{z}_i)\right\|_2 = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})f_{\boldsymbol{u}}(\xi_i,\boldsymbol{z}_i),$$

where $f_{\boldsymbol{u}}(\xi_i,\boldsymbol{z}_i) := \langle \boldsymbol{u}, \xi_i\boldsymbol{z}_i\rangle$ satisfies $|f_{\boldsymbol{u}}(\xi_i,\boldsymbol{z}_i)| \le M\zeta_p$ and $\mathbb{E}\{f^2_{\boldsymbol{u}}(\xi_i,\boldsymbol{z}_i)\} = \mathbb{E}\{\xi_i^2\langle\boldsymbol{u},\boldsymbol{z}_i\rangle^2\} \le \sigma^2$. Applying a refined Talagrand's inequality (see, e.g., Theorem 7.3 in [3]) yields that with probability at least $1-e^{-t}$,

$$\Delta \le 2\mathbb{E}\Delta + \sigma\sqrt{\frac{2t}{n}} + \frac{4M\zeta_p t}{3n}.$$

It remains to bound $\mathbb{E}\Delta$. By the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}\Delta &\le \left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n(\xi_i\boldsymbol{z}_i - \mathbb{E}\xi_i\boldsymbol{z}_i)\right\|_2^2\right)^{1/2} \le \frac{1}{n}\left(\sum_{i=1}^n\mathbb{E}\|\xi_i\boldsymbol{z}_i - \mathbb{E}\xi_i\boldsymbol{z}_i\|_2^2\right)^{1/2} \\ &\le \frac{1}{n}\left\{\sum_{i=1}^n\mathbb{E}\big(\xi_i^2\|\boldsymbol{z}_i\|_2^2\big)\right\}^{1/2} \le \frac{\sigma}{n^{1/2}}\big(\mathbb{E}\|\boldsymbol{z}_i\|_2^2\big)^{1/2} = \sigma\sqrt{\frac{p}{n}}. \end{aligned}$$

Combining the above two displays gives

$$\left\|\frac{1}{n}\sum_{i=1}^n(\xi_i\boldsymbol{z}_i - \mathbb{E}\xi_i\boldsymbol{z}_i)\right\|_2 \le 2\sigma\sqrt{\frac{p}{n}} + \sigma\sqrt{\frac{2t}{n}} + M\zeta_p\frac{4t}{3n}$$

holds with probability at least $1-e^{-t}$. □

E.3. Proof of Lemma B.3. Proof of (i). For each $j = 0, 1, \dots, m$, define random variables $\xi_{ij} = \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i) - \tau_0$, so that the centered process can be written as $\Sigma^{-1/2}\{\widehat{Q}_0(\boldsymbol{\beta}_j^*) - \mathbb{E}\widehat{Q}_0(\boldsymbol{\beta}_j^*)\} = (1/n)\sum_{i=1}^n(\xi_{ij}\boldsymbol{z}_i - \mathbb{E}\xi_{ij}\boldsymbol{z}_i)$. Since $\Delta_i \in \{0,1\}$ and $0 \le \bar{K}(\cdot) \le 1$, we have $|\xi_{ij}| \le \bar{\tau}_0 = \max(\tau_0, 1-\tau_0)$. In particular, for $j = 0$, it is shown in the proof of Lemma C.2 in [6] that $\mathbb{E}(\xi_{i0}^2|\boldsymbol{x}_i) \le \tau_0(1-\tau_0) + (1+\tau_0)l_1\kappa_2 h^2$. For general $j \ge 1$, we can simply use the crude second moment bound $\mathbb{E}(\xi_{ij}^2|\boldsymbol{x}_i) \le \bar{\tau}_0^2$. The claimed bound of (i) then follows directly from Lemma B.2.

Proof of (ii). The bound follows trivially from Lemma B.2 and the facts that $|\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)| \le 1$ and $\mathbb{E}\{\bar{K}_h^2(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)|\boldsymbol{x}_i\} \le 1$.

Proof of (iii). The proof is based on a similar argument used in the proof of Lemma B.2. Fix j, set $\boldsymbol{v} = \boldsymbol{\beta}_{j+1}^* - \boldsymbol{\beta}_j^*$ so that $\|\boldsymbol{v}\|_\Sigma \le \underline{f}^{-1}\delta^*$. By the monotonicity of $u \mapsto \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)$ and $\bar{K}_h(\cdot)$, we have

$$\left|\int_{\tau_j}^{\tau_{j+1}}\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u) - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_j) - y_i)\}\mathrm{d}H(u)\right|$$

$$\le w_j\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_{j+1}^* - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\} \le \kappa_u w_j h^{-1}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v} \le \kappa_u \underline{f}^{-1} w_j h^{-1}\zeta_p\delta^*, \tag{E.1}$$

implying the boundedness, where the last step follows from Condition 3.4 and (10). To control the (conditional) second moment, note that

$$\begin{aligned}
&\mathbb{E}\left[\left\{\int_{\tau_j}^{\tau_{j+1}}\left\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u) - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\right\}\mathrm{d}H(u)\right\}^2\middle|\boldsymbol{x}_i\right] \\
&\le w_j^2\int_{-\infty}^{\infty}\left\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_{j+1}^* - u) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - u)\right\}^2 f_y(u|\boldsymbol{x})\mathrm{d}u \\
&= w_j^2 h\int_{-\infty}^{\infty}\left\{\bar{K}(v + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h) - \bar{K}(v)\right\}^2 f_y(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j - hv|\boldsymbol{x})\mathrm{d}v \\
&\le \bar{f} w_j^2 h^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2\int_{-\infty}^{\infty}\left\{\int_0^1 K(v + w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h)\mathrm{d}w\right\}^2\mathrm{d}v \\
&\overset{(*)}{\le} \bar{f} w_j^2 h^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2\left(\int_0^1\left\{\int_{-\infty}^{\infty}K^2(v + w\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v}/h)\mathrm{d}v\right\}^{1/2}\mathrm{d}w\right)^2 \\
&\le \kappa_u \bar{f} w_j^2 h^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^2,
\end{aligned}$$

where Minkowski's integral inequality is applied in step $(*)$. Turning to the unconditional second moment, we have for any unit vector $\boldsymbol{u}$ that

$$\mathbb{E}\left\{\int_{\tau_j}^{\tau_{j+1}}\left\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u) - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_j) - y_i)\right\}\mathrm{d}H(u)\langle\boldsymbol{u}, \boldsymbol{z}_i\rangle\right\}^2$$

$$\le \kappa_u \bar{f} w_j^2 h^{-1}\left\{\mathbb{E}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{v})^4\right\}^{1/2}\left\{\mathbb{E}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u})^4\right\}^{1/2} \le \kappa_u \bar{f}\underline{f}^{-2} m_4 w_j^2 h^{-1}\delta^{*2}, \tag{E.2}$$

where m_4 is given in (9). Combining (E.1) and (E.2) with Talagrand's inequality as in Lemma B.2 proves the claimed bound. □

E.4. Proof of Lemma B.4. Throughout the proof, for any fixed $j = 0, 1, \dots, m$, we write $\boldsymbol{\beta}^* = \boldsymbol{\beta}_j^*$, $\widehat{Q}(\cdot) = \widehat{Q}_j(\cdot)$ and $Q(\cdot) = \mathbb{E}\widehat{Q}(\cdot)$ for simplicity. Recall the smoothed estimating functions defined in (3), (4), and the induced metric (symmetrized Bregman divergence)

$$D(\boldsymbol{\beta},\boldsymbol{\beta}^*) = \frac{1}{n}\sum_{i=1}^n \Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i)-\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-y_i)\big\}\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*), \tag{E.3}$$

where $\bar{K}_h(\cdot)=\bar{K}(\cdot/h)$. Given $h,r>0$, define the events $\mathcal{E}_i=\{|\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-y_i|\le h/2\}\cap\{|\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)|\le\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma\cdot h/(2r)\}$ for $i=1,\ldots,n$. For any $\boldsymbol{\beta}\in\boldsymbol{\beta}^*+\Theta(r)$, it is easy to see that $|y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}|\le h$ on $\mathcal{E}_i$, hence implying

$$D(\boldsymbol{\beta},\boldsymbol{\beta}^*)\ge\frac{\kappa_l}{nh}\sum_{i=1}^n\Delta_i\{\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\}^2\mathbb{1}_{\mathcal{E}_i}. \tag{E.4}$$

It then suffices to bound the right-hand side of (E.4) from below uniformly over $\boldsymbol{\beta}\in\boldsymbol{\beta}^*+\Theta(r)$.

For $R>0$, define the function $\varphi_R(u)=u^2\mathbb{1}(|u|\le R/2)+\{u\,\mathrm{sign}(u)-R\}^2\mathbb{1}(R/2<|u|\le R)$, which is R-Lipschitz continuous and satisfies the following properties: $\varphi_{cR}(cu)=c^2\varphi_R(u)$ for any $c\ge0$, $\varphi_0(u)=0$, and

$$u^2\mathbb{1}(|u|\le R/2)\le\varphi_R(u)\le u^2\mathbb{1}(|u|\le R). \tag{E.5}$$

For $\boldsymbol{\beta}\in\boldsymbol{\beta}^*+\Theta(r)$, consider the change of variable $\boldsymbol{\delta}=\Sigma^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)/\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma$. Together, (E.4) and (E.5) imply

$$\frac{D(\boldsymbol{\beta},\boldsymbol{\beta}^*)}{\kappa_l\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma^2}\ge D_0(\boldsymbol{\delta}):=\frac{1}{nh}\sum_{i=1}^n\omega_i\cdot\varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}), \tag{E.6}$$

where $\omega_i:=\mathbb{1}(|\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-y_i|\le h/2,\Delta_i=1)$.

We first bound the expectation $\mathbb{E}\{D_0(\boldsymbol{\delta})\}$, and then control the concentration of $D_0(\boldsymbol{\delta})$ around $\mathbb{E}\{D_0(\boldsymbol{\delta})\}$. When $0<h\le1$, Condition 3.3 ensures that

$$\underline{g}\cdot h\le\mathbb{E}(\omega_i|\boldsymbol{x}_i)=\int_{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-h/2}^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*+h/2}g(u|\boldsymbol{x}_i)\,\mathrm{d}u\le\bar{g}\cdot h\ \text{ almost surely.} \tag{E.7}$$

It then follows from (E.5) and (E.7) that

$$\begin{aligned}\mathbb{E}\big\{\omega_i\cdot\varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})\big\}&\ge\underline{g}h\cdot\mathbb{E}\varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})\ge\underline{g}h\cdot\mathbb{E}\big\{(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})^2\mathbb{1}\big(|\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}|\le h/(4r)\big)\big\}\\&=\underline{g}h\cdot\big\{1-\mathbb{E}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})^2\mathbb{1}\big(|\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}|>h/(4r)\big)\big\},\end{aligned}$$

which further implies

$$\inf_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\mathbb{E}\{D_0(\boldsymbol{\delta})\}\ge\underline{g}\cdot\left[1-\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\big\{(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})^2\mathbb{1}\big(|\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}|>h/(4r)\big)\big\}\right].$$

By the definition of η_ξ in (B.1), as long as $0<r\le h/(4\eta_{1/4})$,

$$\inf_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\mathbb{E}\{D_0(\boldsymbol{\delta})\}\ge\frac{3}{4}\underline{g}. \tag{E.8}$$

Turning to the random process $\{D_0(\boldsymbol{\delta})-\mathbb{E}D_0(\boldsymbol{\delta}):\boldsymbol{\delta}\in\mathbb{S}^{p-1}\}$, it suffices to bound

$$\Lambda=\sup_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\big\{-D_0(\boldsymbol{\delta})+\mathbb{E}D_0(\boldsymbol{\delta})\big\}. \tag{E.9}$$

By the fact that $0\le\varphi_R(u)\le\min\{(R/2)^2,(R/2)|u|\}$ for all $u\in\mathbb{R}$, we have

$$0\le(\omega_i/h)\varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})\le\omega_i\min\big\{(4r)^{-2}h,(4r)^{-1}|\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}|\big\}.$$

Combining this with (E.7) yields

$$\mathbb{E}\{(\omega_i/h)^2\varphi^2_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})\} \le (4r)^{-2}\mathbb{E}\{\mathbb{E}(\omega_i|\boldsymbol{x}_i)(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta})^2\} \le (4r)^{-2}\bar{g}h.$$

With the above preparations, we apply a refined Talagrand's inequality—Theorem 7.3 in [3]—to obtain that, for any $t > 0$,

$$\Lambda \le \mathbb{E}\Lambda + (\mathbb{E}\Lambda)^{1/2}\sqrt{\frac{ht}{4r^2n}} + \bar{g}^{1/2}\sqrt{\frac{ht}{8r^2n}} + \frac{h}{(4r)^2}\frac{t}{3n}$$

$$\le \frac{5}{4}\mathbb{E}\Lambda + \bar{g}^{1/2}\sqrt{\frac{ht}{8r^2n}} + (1/4 + 1/48)\frac{ht}{r^2n} \tag{E.10}$$

with probability at least $1 - e^{-t}$. It remains to bound $\mathbb{E}\Lambda$. To this end, we define

$$\mathcal{E}(\boldsymbol{\delta}; \boldsymbol{z}_i, y_i) = \frac{\omega_i}{h}\varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}) = \frac{1}{h}\varphi_{\omega_i h/(2r)}(\omega_i\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}), \quad \boldsymbol{\delta} \in \mathbb{S}^{p-1}.$$

where the second equality follows from the property that $\varphi_{cR}(cu) = c^2\varphi_R(u)$ for any $c \ge 0$. By the Lipschitz continuity of $\varphi_R(\cdot)$, $\mathcal{E}(\boldsymbol{\delta}; \boldsymbol{z}_i, y_i)$ is $(2r)^{-1}$-Lipschitz continuous in $\omega_i\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}$, and $\mathcal{E}(\boldsymbol{\delta}; \boldsymbol{z}_i, y_i) = 0$ for any $\boldsymbol{\delta}$ such that $\omega_i\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta} = 0$. Furthermore, define the subset $T \subseteq \mathbb{R}^n$ as

$$T = \{\boldsymbol{t} = (t_1, \ldots, t_n)^{\mathrm{T}} : t_i = \omega_i\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}, i = 1, \ldots, n, \boldsymbol{\delta} \in \mathbb{S}^{p-1}\},$$

and contractions $\phi_i : \mathbb{R} \to \mathbb{R}$ as $\phi_i(t) = (2r/h) \cdot \varphi_{\omega_i h/(2r)}(t)$. In fact, the Lipschitz continuity of $\varphi_R(\cdot)$ implies $|\phi(t) - \phi(s)| \le |t - s|$ for all $t, s \in \mathbb{R}$. Let $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher random variables, and denote by $\mathbb{E}_\epsilon$ the expectation taken only with respect to ϵ_i's. Then, via a standard symmetrization and contraction argument (see, e.g. Lemma 6.3 and Theorem 4.12 in [9]), we have

$$\mathbb{E}_\epsilon\Lambda \le 2\mathbb{E}_\epsilon\left\{\sup_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\frac{1}{n}\sum_{i=1}^n\epsilon_i\mathcal{E}(\boldsymbol{\delta}; \boldsymbol{z}_i, y_i)\right\} = \frac{1}{r}\mathbb{E}_\epsilon\left\{\sup_{\boldsymbol{t}\in T}\frac{1}{n}\sum_{i=1}^n\epsilon_i\phi_i(t_i)\right\}$$

$$\le \frac{1}{r}\mathbb{E}_\epsilon\left(\sup_{\boldsymbol{t}\in T}\frac{1}{n}\sum_{i=1}^n\epsilon_i t_i\right) = \frac{1}{r}\mathbb{E}_\epsilon\left\{\sup_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\frac{1}{n}\sum_{i=1}^n\epsilon_i\cdot\omega_i\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\delta}\right\} \le \frac{1}{r}\mathbb{E}_\epsilon\left\|\frac{1}{n}\sum_{i=1}^n\epsilon_i\omega_i\boldsymbol{z}_i\right\|_2.$$

Taking the expectation over $\{(\boldsymbol{z}_i, y_i)\}_{i=1}^n$ on both sides yields $\mathbb{E}\Lambda \le \bar{g}^{1/2}\sqrt{hp/(r^2n)}$. Substituting this into (E.10), we obtain

$$\Lambda \le \bar{g}^{1/2}r^{-1}\left(\frac{5}{4}\sqrt{\frac{hp}{r^2n}} + \sqrt{\frac{ht}{8r^2n}}\right) + (1/4 + 1/48)r^{-2}\frac{ht}{n} \tag{E.11}$$

with probability at least $1 - e^{-t}$.

Finally, combining (E.6), (E.8), (E.9) and (E.11) completes the proof. □

E.5. Proof of Lemma B.5. To begin with, define the centered process

$$S(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n(1 - \mathbb{E})\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{z}_i, \quad \boldsymbol{\beta} \in \mathbb{R}^p.$$

After a change of variable $\boldsymbol{v} = \Sigma^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_j^*)$, we have

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}||S(\boldsymbol{\beta}) - S(\boldsymbol{\beta}_j^*)||_2 = \sup_{\boldsymbol{v}\in\mathbb{B}^p(r)}\|\underbrace{S(\boldsymbol{\beta}_j^* + \Sigma^{-1/2}\boldsymbol{v}) - S(\boldsymbol{\beta}_j^*)}_{=:\,\Psi(\boldsymbol{v})}\|_2.$$

Since the empirical process $\Psi(\boldsymbol{v})$ is continuous with respect to $\boldsymbol{v}$, we will apply the concentration bound from Theorem A.3 in [12] to control the supremum $\sup_{\boldsymbol{v}\in\mathbb{B}^p(r)}\|\Psi(\boldsymbol{v})\|_2$.

First, note that the function $\Psi(\cdot):\mathbb{R}^p\to\mathbb{R}^p$ satisfies $\Psi(\mathbf{0})=\mathbf{0}$, $\mathbb{E}\{\Psi(\boldsymbol{v})\}=\mathbf{0}$

$$\nabla\Psi(\boldsymbol{v})=\frac{1}{n}\sum_{i=1}^n\big\{\phi_{i,\boldsymbol{v}}\boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}-\mathbb{E}\big(\phi_{\boldsymbol{v}}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\big)\big\},$$

where $\phi_{i,\boldsymbol{v}}=K_h(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}-\varepsilon_i)$, $\phi_{\boldsymbol{v}}=K_h(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}-\varepsilon)$ and $\varepsilon_i=y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*$. It is easy to see that $0\le\phi_{i,\boldsymbol{v}}\le\kappa_u/h$ with $\kappa_u=\sup_{u\in\mathbb{R}}K(u)$. For any $\boldsymbol{g},\boldsymbol{h}\in\mathbb{S}^{p-1}$ and $|\lambda|\le\min\{nh/(\kappa_u\zeta_p^2),n/\overline{g}\}$, by independence and the elementary inequality $e^u\le 1+u+u^2e^{|u|}/2$, we obtain that

$$
\begin{aligned}
&\mathbb{E}\exp\big\{\lambda\boldsymbol{g}^{\mathrm{T}}\nabla\Psi(\boldsymbol{v})\boldsymbol{h}\big\}\\
&\le\left[1+\frac{\lambda^2}{2n^2}e^{\frac{\overline{g}|\lambda|}{n}\mathbb{E}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}|}\mathbb{E}\big\{\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}-\mathbb{E}(\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h})\big\}^2e^{\frac{\kappa_u|\lambda|}{nh}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}|}\right]^n\\
&\overset{\text{(i)}}{\le}\left[1+\frac{\lambda^2}{2n^2}e^{\overline{g}|\lambda|/n}\mathbb{E}\big\{\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}-\mathbb{E}(\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h})\big\}^2e^{\frac{\kappa_u|\lambda|}{nh}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}|}\right]^n\\
&\le\left[1+\frac{e\lambda^2}{2n^2}\mathbb{E}\big\{\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}-\mathbb{E}(\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h})\big\}^2e^{\kappa_u\zeta_p^2|\lambda|/(nh)}\right]^n\\
&\le\left\{1+\frac{(e\lambda)^2}{2n^2}\mathbb{E}\big(\phi_{\boldsymbol{v}}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}\big)^2\right\}^n, \qquad \text{(E.12)}
\end{aligned}
$$

where inequality (i) follows from the bound $\mathbb{E}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{g}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{h}|\le 1$. For $\phi_{\boldsymbol{v}}=K_h(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}-\varepsilon)$, under Condition 3.3, its conditional second moment can be bounded by

$$
\begin{aligned}
\mathbb{E}\big(\phi_{\boldsymbol{v}}^2|\boldsymbol{x}\big)&=\frac{1}{h^2}\int_{-\infty}^{\infty}K^2\bigg(\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*-t}{h}\bigg)f_y(t|\boldsymbol{x})\mathrm{d}t\\
&=\frac{1}{h}\int_{-\infty}^{\infty}K^2(u)f_y(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*+hu|\boldsymbol{x})\mathrm{d}u\le\frac{\kappa_u\overline{f}}{h}. \qquad \text{(E.13)}
\end{aligned}
$$

Substituting this into (E.12) yields

$$\mathbb{E}\exp\big\{\lambda\boldsymbol{g}^{\mathrm{T}}\nabla\Psi(\boldsymbol{v})\boldsymbol{h}\big\}\le\big\{1+\kappa_u\overline{f}e^2m_4\lambda^2/(2n^2h)\big\}^n\le\exp\big\{\kappa_u\overline{f}e^2m_4\lambda^2/(2nh)\big\}.$$

This verifies condition (A.4) in [12]. Therefore, applying Theorem A.3 therein, we obtain that with probability at least $1-e^{-t}$,

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)}\left\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\big\{\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})-\bar{K}_h(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*)\big\}\boldsymbol{z}_i\right\|_2\\
&=\sup_{\boldsymbol{v}\in\mathbb{B}^p(r)}\|\Psi(\boldsymbol{v})\|_2\lesssim(\kappa_u\overline{f}m_4)^{1/2}\sqrt{\frac{p+t}{nh}}\cdot r \qquad \text{(E.14)}
\end{aligned}
$$

as long as $nh\gtrsim\zeta_p^2(p+t)^{1/2}$. This proves (B.5), and (B.6) can be obtained from the same argument.

Turning to the mean difference approximation, applying the mean value theorem for vector-valued functions implies

$$
\begin{aligned}
&\mathbb{E}\{\bar{K}_h(y-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta})-\bar{K}_h(y-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*)\}\boldsymbol{z}\\
&=-\int_0^1\mathbb{E}\big\{K_h(y-\langle\boldsymbol{x},\boldsymbol{\beta}_j^*+t(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)\rangle)\boldsymbol{z}\boldsymbol{x}^{\mathrm{T}}\big\}\mathrm{d}t\cdot(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*).
\end{aligned}
$$

With $\boldsymbol{v}=\Sigma^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)$, note that

$$\begin{aligned}
&\mathbb{E}\big\{K_h(y-\langle \boldsymbol{x},\boldsymbol{\beta}_j^*+t(\boldsymbol{\beta}-\boldsymbol{\beta}_j^*)\rangle)|\boldsymbol{x}\big\}\\
&=\frac{1}{h}\int_{-\infty}^{\infty}K\left(\frac{u-t\boldsymbol{z}^{\mathrm T}\boldsymbol{v}}{h}\right)f_y(\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*+u|\boldsymbol{x})\mathrm{d}u=\int_{-\infty}^{\infty}K(v)f_y(\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*+t\boldsymbol{z}^{\mathrm T}\boldsymbol{v}+hv|\boldsymbol{x})\mathrm{d}v.
\end{aligned}$$

By the Lipschitz continuity of $f_y(\cdot|\boldsymbol{x})$, we have

$$\begin{aligned}
&\Big\|\mathbb{E}\{\bar K_h(y-\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta})-\bar K_h(y-\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*)\}\boldsymbol{z}+\mathbb{E}\big\{f_y(\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*|\boldsymbol{x})\boldsymbol{z}\boldsymbol{z}^{\mathrm T}\big\}\boldsymbol{v}\Big\|_2\\
&=\Big\|\mathbb{E}\int_0^1\int_{-\infty}^{\infty}K(v)\big\{f_y(\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*+t\boldsymbol{z}^{\mathrm T}\boldsymbol{v}+hv|\boldsymbol{x})-f_y(\boldsymbol{x}^{\mathrm T}\boldsymbol{\beta}_j^*|\boldsymbol{x})\big\}\boldsymbol{z}\boldsymbol{z}^{\mathrm T}\mathrm{d}v\mathrm{d}t\cdot\boldsymbol{v}\Big\|_2\\
&\le l_1\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\int_0^1\int_{-\infty}^{\infty}K(v)\big(t|\boldsymbol{z}^{\mathrm T}\boldsymbol{v}|+h|v|\big)\mathrm{d}v\mathrm{d}t\cdot|\boldsymbol{z}^{\mathrm T}\boldsymbol{u}\boldsymbol{z}^{\mathrm T}\boldsymbol{v}|\le l_1\big(0.5m_3r+\kappa_1h\big)r,
\end{aligned}$$

as claimed. □

E.6. Proof of Lemma B.6. For any $\epsilon\in(0,\tau_u-\tau_l)$, we divide the interval $[\tau_l,\tau_u]$ into $L:=\lceil(\tau_u-\tau_l)/(2\epsilon)\rceil+1$ subintervals, centered at the points τ^k for $k\in[L]$, and each of length at most 2ϵ. For any $\tau\in[\tau_l,\tau_u]$, there exists some k such that $|\tau-\tau^k|\le\epsilon$, and hence

$$\begin{aligned}
&\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_l}^{\tau}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{x}_i\bigg\|_{\Sigma^{-1}}\\
&\le\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_l}^{\tau^k}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{x}_i\bigg\|_{\Sigma^{-1}}\\
&\qquad+\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau^k}^{\tau}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{x}_i\bigg\|_{\Sigma^{-1}}\\
&\le\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_l}^{\tau^k}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{z}_i\bigg\|_2+2\zeta_p|H(\tau)-H(\tau^k)|.
\end{aligned}$$

For any given $k\in[L]$, applying Lemma B.2 yields that with probability at least $1-e^{-v}$,

$$\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_l}^{\tau^k}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{z}_i\bigg\|_2\lesssim|H(\tau^k)-H(\tau_l)|\left(\sqrt{\frac{p+v}{n}}+\zeta_p\frac{v}{n}\right).$$

Recall that $H(u)=-\log(1-u)$, $u\in(0,1)$, we have $|H(u)-H(v)|\le|u-v|/(1-u\vee v)$. Finally, taking $\epsilon=(\tau_u-\tau_l)/(2n)$, $v=\log L+t$ $(t>0)$, and the union bound over $k=1,\ldots,L$, we conclude that

$$\begin{aligned}
&\sup_{\tau\in[\tau_l,\tau_u]}\bigg\|\frac{1}{n}\sum_{i=1}^n(1-\mathbb{E})\int_{\tau_j}^{\tau}\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{x}_i\bigg\|_{\Sigma^{-1}}\\
&\qquad\qquad\lesssim\frac{\tau_u-\tau_l}{1-\tau_u}\left(\sqrt{\frac{p+\log n+t}{n}}+\zeta_p\frac{\log n+t}{n}\right)
\end{aligned}$$

holds with probability at least $1-e^{-t}$. This proves the claimed result. □

E.7. Proof of Lemma B.7. For $Q_0(\boldsymbol{\beta}) = \mathbb{E}\{\Delta \bar{K}_h(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} - y) - \tau_0\}\boldsymbol{x}$, it follows from integration by parts and change of variables that

$$\mathbb{E}\{\Delta \bar{K}_h(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^* - y)|\boldsymbol{x}\} = \int_{-\infty}^{\infty} \bar{K}\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^* - t}{h}\right) \mathrm{d}G(t|\boldsymbol{x}) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^* - t}{h}\right) G(t|\boldsymbol{x})\mathrm{d}t$$

$$= \int_{-\infty}^{\infty} K(u) G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^* + hu|\boldsymbol{x})\mathrm{d}u$$

$$= G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*|\boldsymbol{x}) + \int_{-\infty}^{\infty} K(u) \int_{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*}^{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^* + hu} \{g(t|\boldsymbol{x}) - g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*|\boldsymbol{x})\}\mathrm{d}t\,\mathrm{d}u. \tag{E.15}$$

On the other hand, using the martingale property gives

$$\mathbb{E}\left[\int_0^{\tau_j} \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\mathrm{d}H(u)\middle|\boldsymbol{x}\right] = \mathbb{E}\{N(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*)|\boldsymbol{x}\} = \mathbb{P}(y \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*, \Delta = 1|\boldsymbol{x}) = G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^*|\boldsymbol{x}).$$

Together, the last two displays and the Lipschitz continuity of $g(\cdot|\boldsymbol{x})$ imply

$$\left\|\mathbb{E}\left[\Delta \bar{K}_h(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_j^* - y) - \int_0^{\tau_j} \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\mathrm{d}H(u)\right]\boldsymbol{x}\right\|_{\Sigma^{-1}} \leq \frac{1}{2} l_1 \kappa_2 h^2. \tag{E.16}$$

Next, for $\ell = 0, 1, \ldots, j-1$,

$$\mathbb{E}\big\{\bar{K}_h(y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)|\boldsymbol{x}\big\} = 1 - F_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*|\boldsymbol{x}) - \int_{-\infty}^{\infty} K(v) \int_{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*}^{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^* + hv} \{f_y(t|\boldsymbol{x}) - f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*|\boldsymbol{x})\}\,\mathrm{d}t\mathrm{d}v.$$

This, combined with the Lipschitz continuity of $f(\cdot|\boldsymbol{x})$, implies

$$\left\|\mathbb{E}\left[\sum_{\ell=0}^{j-1} w_\ell \big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \mathbb{1}(y_i \geq \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*)\big\}\boldsymbol{x}_i\right]\right\|_{\Sigma^{-1}} \leq \frac{1}{2} l_1 \kappa_2 h^2 \sum_{\ell=0}^{j-1} w_\ell. \tag{E.17}$$

It remains to compare $\int_0^{\tau_j} \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\mathrm{d}H(u)$ and $\sum_{\ell=0}^{j-1} w_\ell \mathbb{1}(y_i \geq \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) + \tau_0$. By the global linear conditional quantile model assumption, the function $u \mapsto \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}$ is non-increasing in $u \in [\tau_L, \tau_U]$. Consequently,

$$\begin{aligned}
0 &\leq \mathbb{E}\left(\sum_{\ell=0}^{j-1} \int_{\tau_\ell}^{\tau_{\ell+1}} \big[\mathbb{1}(y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\big]\,\mathrm{d}H(u)\middle|\boldsymbol{x}\right) \\
&\leq \mathbb{E}\left(\sum_{\ell=0}^{j-1} \int_{\tau_\ell}^{\tau_{\ell+1}} \big\{\mathbb{1}(y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \mathbb{1}(y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_{\ell+1}^*)\big\}\,\mathrm{d}H(u)\middle|\boldsymbol{x}\right) \\
&= \sum_{\ell=0}^{j-1} \int_{\tau_\ell}^{\tau_{\ell+1}} \big\{F_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_{\ell+1}^*|\boldsymbol{x}) - F_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*|\boldsymbol{x})\big\}\,\mathrm{d}H(u) \\
&\leq \bar{f} \sum_{\ell=0}^{j-1} w_\ell \cdot \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{\beta}_{\ell+1}^* - \boldsymbol{\beta}_\ell^*).
\end{aligned}$$

Recall from the last paragraph of Section 3.1 that $\|\boldsymbol{\beta}_{\ell+1}^* - \boldsymbol{\beta}_\ell^*\|_\Sigma \leq \underline{f}^{-1}|\tau_{\ell+1} - \tau_\ell|$ for $\ell = 0, 1, \ldots, m-1$. Under the condition of no censoring below $\tau_0 = \tau_L$, we have $\mathbb{E}\int_0^{\tau_0} \mathbb{1}\{y \geq$

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\mathrm{d}H(u) = \tau_0$. Putting together the pieces, we conclude that

$$\left\|\mathbb{E}\left[\int_0^{\tau_j} \mathbb{1}\{y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\,\mathrm{d}H(u) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{1}(y \geq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \tau_0\right]\boldsymbol{x}\right\|_{\Sigma^{-1}}$$
$$\leq (\overline{f}/\underline{f}) \sum_{\ell=0}^{j-1} w_\ell(\tau_{\ell+1} - \tau_\ell).$$

Combining this with (E.16) and (E.17) proves (B.7). □

E.8. Proof of Lemma B.8. Fix $\delta > 0$, for any $\tau_1, \tau_2 \in [\tau_L, \tau_U]$ satisfying $|\tau_1 - \tau_2| < \delta$, define the centered random variables $V_i := \langle \boldsymbol{a}_n, \boldsymbol{U}_{0i}(\tau_1) - \boldsymbol{U}_{0i}(\tau_2)\rangle$ for $i = 1, \ldots, n$, where $\boldsymbol{U}_{0i}(\tau) = \boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)$. Assume without loss of generality that $\tau_2 \geq \tau_1$. For some constant $L \geq 1$ to be determined, applying Rosenthal's inequality to $S := \sum_{i=1}^n V_i$ yields

$$\left(\mathbb{E}S^{2L}\right)^{1/(2L)} \leq C_L\left\{\left(\sum_{i=1}^n \mathbb{E}V_i^2\right)^{1/2} + \left(\sum_{i=1}^n \mathbb{E}V_i^{2L}\right)^{1/(2L)}\right\},$$

where $C_L > 0$ is a constant depending only on L. We then bound the second and higher-order moments of V_i's. By Minkowski's integral inequality as in the proof of Lemma B.3,

$$\mathbb{E}\langle \boldsymbol{a}_n, \boldsymbol{U}_i(\tau_1) - \boldsymbol{U}_i(\tau_2)\rangle^2$$
$$= \mathbb{E}\left\{-\int_{\tau_1}^{\tau_2} \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u))\mathrm{d}H(u) + \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_2) - y_i) - \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_1) - y_i)\right\}^2 \langle \boldsymbol{a}_n, \boldsymbol{x}_i\rangle^2$$
$$\lesssim (\overline{f} \vee \overline{g})\left\{(1-\tau_2)^{-2} + \underline{f}^{-2} m_4 h^{-1}\right\}(\tau_2 - \tau_1)^2.$$

Moreover, for any $q > 2$,

$$|\mathbb{E}\langle \boldsymbol{a}_n, \boldsymbol{U}_i(\tau_1) - \boldsymbol{U}_i(\tau_2)\rangle^q|$$
$$\lesssim \zeta_p^{q/2-1}\mathbb{E}\langle \boldsymbol{a}_n, \boldsymbol{U}_i(\tau_1) - \boldsymbol{U}_i(\tau_2)\rangle^2 \lesssim (\overline{f} \vee \overline{g})\left\{(1-\tau_2)^{-2} + \underline{f}^{-2} m_4 h^{-1}\right\}\zeta_p^{q/2-1}(\tau_2 - \tau_1)^2.$$

Putting together the pieces, we obtain that for any $\tau_1 < \tau_2$ satisfying $\tau_2 - \tau_1 \geq (\zeta_p/m_4)^{1/2}(h/n)^{1/2}$,

$$\left(\mathbb{E}S^{2L}\right)^{1/(2L)} \lesssim n^{1/2}(m_4/h)^{1/2}|\tau_2 - \tau_1| + n^{1/(2L)}\zeta_p^{(L-1)/(2L)}\left\{(m_4/h)^{1/2}|\tau_2 - \tau_1|\right\}^{1/L}$$
$$\lesssim n^{1/2}(m_4/h)^{1/2}|\tau_2 - \tau_1|.$$

Note that $\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2) = n^{-1/2}S$. Hence, taking $\psi(x) = x^{2L}$ in the above inequality leads to

$$||\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2)||_\psi \lesssim (m_4/h)^{1/2}|\tau_2 - \tau_1|, \tag{E.18}$$

where $||\cdot||_\psi$ denotes the ψ-Orlicz norm; see Section 2.2 in [15].

The rest of the proof is based on a packing argument, and is inspired by the proof of Lemma A.3 in [4]. Define the metric $d(\cdot,\cdot)$ as $d(s,t) = h^{-1/2}|s - t|$ for $s, t \in [\tau_L, \tau_U]$. Then, for any $\epsilon > 0$, the packing number $\mathcal{P}([\tau_L, \tau_U], \epsilon, d) \lesssim h^{1/2}\epsilon^{-1}$. Let $\bar{\eta} = 2\sqrt{\zeta_p/(m_4 n)}$, so that $\lim_{n\to\infty} \bar{\eta} \to 0$, and (E.18) holds for all τ_1, τ_2 satisfying $d(\tau_1, \tau_2) \geq \bar{\eta}/2$. Applying Lemma A.1 in the Appendix of [7] gives

$$\sup_{|\tau_1 - \tau_2| < \delta} \left|\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2)\right| = \sup_{d(\tau_1,\tau_2) < h^{-1/2}\delta} \left|\mathbb{G}_n(\tau_1) - \mathbb{G}_n(\tau_2)\right|$$
$$\leq S_1 + \sup_{d(s,t) < \bar{\eta},\, t \in \widetilde{\mathcal{T}}} \left|\mathbb{G}_n(s) - \mathbb{G}_n(t)\right|, \tag{E.19}$$

where the set $\widetilde{\mathcal{T}}$ contains at most $\mathcal{P}([\tau_L,\tau_U],\bar\eta,d)\lesssim h^{1/2}\bar\eta^{-1}\lesssim\sqrt{nh/\zeta_p}$ points, and S_1 is a random variable satisfying

(E.20)
$$\mathbb{P}(|S_1|>x)\lesssim\left\{\int_{\bar\eta/2}^{\eta}\psi^{-1}(\mathcal{P}([\tau_L,\tau_U],\epsilon,d))\mathrm{d}\epsilon+(h^{-1/2}\delta+2\bar\eta)\psi^{-1}(\mathcal{P}^2([\tau_L,\tau_U],\eta,d))\right\}^{2L}x^{-2L}$$

for any $\eta\geq\bar\eta$ and $x>0$. Note that

$$\int_{\bar\eta/2}^{\eta}\psi^{-1}(\mathcal{P}([\tau_L,\tau_U],\epsilon,d))\mathrm{d}\epsilon\lesssim h^{(4L)^{-1}}\int_{\bar\eta/2}^{\eta}\epsilon^{-(2L)^{-1}}\mathrm{d}\epsilon=h^{(4L)^{-1}}\cdot\frac{\eta^{1-(2L)^{-1}}-(\bar\eta/2)^{1-(2L)^{-1}}}{1-(2L)^{-1}},$$

and $\psi^{-1}(\mathcal{P}^2([\tau_L,\tau_U],\eta,d))\lesssim h^{(2L)^{-1}}\eta^{-L^{-1}}$. Substituting these into (E.20) with $\eta=h^{-1/4}$ and $L=1$ implies

$$\begin{aligned}\mathbb{P}(|S_1|>x)&\lesssim\left\{h^{(4L)^{-1}}\cdot\frac{\eta^{1-(2L)^{-1}}-(\bar\eta/2)^{1-(2L)^{-1}}}{1-(2L)^{-1}}+(h^{-1/2}\delta+2\bar\eta)\cdot h^{(2L)^{-1}}\eta^{-L^{-1}}\right\}^{2L}\cdot x^{-2L}\\&=\left\{2h^{1/8}-2h^{1/4}(\bar\eta/2)^{1/2}+h^{1/4}\delta+2\bar\eta h^{3/4}\right\}^2\cdot x^{-2}.\end{aligned}$$

Since $\bar\eta\to0$ and $h\to0$ as $n\to\infty$, we have for any $x>0$ that

$$\lim_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}(|S_1|>x)=0. \tag{E.21}$$

It remains to deal with the supremum on the right-hand side of (E.19). For any fixed $t\in\widetilde{\mathcal{T}}$, and $s\in[\tau_L,\tau_U]$ satisfying $|s-t|<h^{1/2}\bar\eta=2\sqrt{\zeta_p h/(m_4 n)}$,

$$\begin{aligned}&\sup_{|s-t|<h^{1/2}\bar\eta}\left|\mathbb{G}_n(s)-\mathbb{G}_n(t)\right|\\&\leq\sup_{|s-t|<h^{1/2}\bar\eta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_t^s\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{a}_n^{\mathrm T}\boldsymbol{x}_i\right|\\&\quad+\sup_{|s-t|<h^{1/2}\bar\eta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\Delta_i\bar K_h(\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(t)-y_i)-\Delta_i\bar K_h(\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(s)-y_i)\right\}\boldsymbol{a}_n^{\mathrm T}\boldsymbol{x}_i\right|\\&:=\mathrm{I}+\mathrm{II}.\end{aligned}\tag{E.22}$$

We bound the two terms on the right-hand side respectively. For the first one, applying the triangle inequality and Lemma B.6 to the centered random quantity yields

$$\begin{aligned}\mathrm{I}&\leq\sup_{|s-t|<h^{1/2}\bar\eta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbb{E}\int_t^s\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{a}_n^{\mathrm T}\boldsymbol{x}_i\right|\\&\quad+\sup_{|s-t|<h^{1/2}\bar\eta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\mathbb{E})\int_t^s\bar K_h(y_i-\boldsymbol{x}_i^{\mathrm T}\boldsymbol{\beta}^*(u))\mathrm{d}H(u)\cdot\boldsymbol{a}_n^{\mathrm T}\boldsymbol{x}_i\right|\\&\lesssim n^{-1/2}h^{1/2}\bar\eta\sum_{i=1}^n\mathbb{E}|\boldsymbol{a}_n^{\mathrm T}\boldsymbol{x}_i|+n^{1/2}h^{1/2}\bar\eta\left(\sqrt{\frac{p+\log n+v}{n}}+\zeta_p\frac{\log n+v}{n}\right)\\&\lesssim(\zeta_p h)^{1/2}\left(\sqrt{\frac{p+\log n+v}{n}}+\zeta_p\frac{\log n+v}{n}\right)\end{aligned}\tag{E.23}$$

with probability at least $1-e^{-v}$ for any $v>0$. Turning to the second term, by the Lipschitz continuity of $\|\boldsymbol{\beta}^*(\cdot)\|_\Sigma$ as in (10), $|s-t|<h^{1/2}\bar{\eta}$ indicates $\|\boldsymbol{\beta}^*(s)-\boldsymbol{\beta}^*(t)\|_\Sigma \le \underline{f}^{-1}h^{1/2}\bar{\eta} = 2\underline{f}^{-1}\sqrt{\zeta_p h/(m_4 n)}$. Denote $r := 2\underline{f}^{-1}\sqrt{\zeta_p h/(m_4 n)}$, we have

$$
\begin{aligned}
\mathrm{II} \le & \sup_{|s-t|<h^{1/2}\bar{\eta}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}\big\{ \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(t)-y_i) - \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(s)-y_i)\big\} \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{x}_i \right| \\
& + \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}^*(t)+\Theta(r)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (1-\mathbb{E})\big\{ \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(t)-y_i) - \Delta_i \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i)\big\} \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{x}_i \right|
\end{aligned}
$$

(E.24)

$$
\lesssim (\zeta_p h)^{1/2} + \zeta_p^{1/2}\sqrt{\frac{p+\log n}{n}}
$$

with probability at least $1-e^{-v}$ for any $v>0$, where Lemma B.5–(ii) is applied in the last step. With the bandwidth $h \asymp \{(p+\log n)/n\}^{2/5}$ and under the sample size requirement $n \gtrsim \zeta_p^{5/2}(p+\log n)$, it follows from (E.22)–(E.24) with $v=2\log n$ that, for any $t\in\widetilde{\mathcal{T}}$,

$$
\sup_{s:|s-t|<h^{1/2}\bar{\eta}} \big|\mathbb{G}_n(s)-\mathbb{G}_n(t)\big| \lesssim \zeta_p^{1/2}\left(\frac{p+\log n}{n}\right)^{1/5}
$$

holds with probability at least $1-n^{-2}$. Since $|\widetilde{\mathcal{T}}| \lesssim \sqrt{nh/\zeta_p}$, taking the union bound over $t\in\widetilde{\mathcal{T}}$ renders

$$
\sup_{d(s,t)<\bar{\eta},\, t\in\widetilde{\mathcal{T}}} \big|\mathbb{G}_n(s)-\mathbb{G}_n(t)\big| = \sup_{|s-t|<h^{1/2}\bar{\eta},\, t\in\widetilde{\mathcal{T}}} \big|\mathbb{G}_n(s)-\mathbb{G}_n(t)\big| \lesssim \zeta_p^{1/2}\left(\frac{p+\log n}{n}\right)^{1/5}
$$

with probability at least $1-\widetilde{c}/n$ for some constant $\widetilde{c}>0$. Provided $\zeta_p^{5/2}(p+\log n)=o(n)$, this implies

$$
\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{d(s,t)<\bar{\eta},\, t\in\widetilde{\mathcal{T}}} \big|\mathbb{G}_n(s)-\mathbb{G}_n(t)\big| > x \right\} = 0, \quad \text{valid for any } x>0. \tag{E.25}
$$

Finally, putting together (E.19), (E.21) and (E.25) completes the proof. □

E.9. Proof of Lemma C.1. For simplicity, we omit the subscript j as in the proof of Lemma B.4. Using arguments similar to those that lead (E.4) and (E.6), we obtain

$$
\begin{aligned}
D^\flat(\boldsymbol{\beta},\boldsymbol{\beta}^*) &= \frac{1}{n}\sum_{i=1}^n \Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-y_i)\big\}(1+e_i)\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \\
&\ge \kappa_l \cdot \|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma^2 \cdot \underbrace{\frac{1}{nh}\sum_{i=1}^n (1+e_i)\omega_i \cdot \varphi_{h/(2r)}(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v})}_{=:D_0^\flat(\boldsymbol{v})},
\end{aligned}
$$

where $\boldsymbol{v}=\Sigma^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)/\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma \in \mathbb{S}^{p-1}$, $\omega_i=\Delta_i\mathbb{1}(|\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*-y_i|\le h/2)\in\{0,1\}$, and $\varphi(\cdot)$ is as in (E.5). Recall the definition of $D_0(\boldsymbol{v})$ in (E.6), we have

$$
\inf_{\boldsymbol{\beta}\in\boldsymbol{\beta}^*+\Theta(r)} \frac{D^\flat(\boldsymbol{\beta},\boldsymbol{\beta}^*)}{\kappa_l\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_\Sigma^2} \ge \inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}} D_0(\boldsymbol{v}) - \sup_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\{D_0(\boldsymbol{v})-D_0^\flat(\boldsymbol{v})\}. \tag{E.26}
$$

A lower bound for $\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}} D_0(\boldsymbol{v})$ can be derived from Lemma B.4. Let $\mathcal{E}_{\rm rsc}(t)$ be the event that the bounds in Lemma B.4 with $r = h/(4\eta_{1/4})$ hold uniformly over $j = 0,1,\ldots,m$, so that $\mathbb{P}\{\mathcal{E}_{\rm rsc}(t)\} \geq 1-(m+1)e^{-t}$. It suffices to control the bootstrap error

$$\Gamma_n := \sup_{\boldsymbol{v}\in\mathbb{S}^{p-1}} \{D_0(\boldsymbol{v}) - D_0^\flat(\boldsymbol{v})\} = \sup_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}} \frac{1}{nh}\sum_{i=1}^n e_i\omega_i \cdot \varphi_{h/(2r)}(\boldsymbol{z}_i^{\rm T}\boldsymbol{v}).$$

Since $\varphi_R(u) \leq (R/2)^2$, $\omega_i \in \{0,1\}$ and $e_i \in \{-1,1\}$, we have $\mathbb{E}^*\{e_i\omega_i\cdot\varphi_{h/(2r)}(\boldsymbol{z}_i^{\rm T}\boldsymbol{v})\}^2 \leq (h/4r)^4\omega_i$ and $|e_i\omega_i\cdot\varphi_{h/(2r)}(\boldsymbol{z}_i^{\rm T}\boldsymbol{v})| \leq (h/4r)^2$. Then, by Theorem 7.3 in [3],

$$\Gamma_n \leq 2\mathbb{E}^*(\Gamma_n) + \frac{h}{(4r)^2}\left\{\left(\frac{1}{n}\sum_{i=1}^n\omega_i\right)^{1/2}\sqrt{\frac{2t}{n}} + \frac{4t}{3n}\right\}$$

holds with probability at least $1-e^{-t}$. Furthermore, by the Lipschitz continuity of $u \to \varphi_R(u)$ and Ledoux-Talagrand contraction principle,

$$\begin{aligned}\mathbb{E}^*(\Gamma_n) &\leq \frac{1}{2r}\mathbb{E}^*\left(\sup_{\boldsymbol{\delta}\in\mathbb{S}^{p-1}}\frac{1}{n}\sum_{i=1}^n e_i\omega_i\cdot\boldsymbol{z}_i^{\rm T}\boldsymbol{v}\right)\\ &= \frac{1}{2r}\mathbb{E}^*\left\|\frac{1}{n}\sum_{i=1}^n e_i\omega_i\boldsymbol{z}_i\right\|_2 \leq \frac{1}{2rn^{1/2}}\left(\frac{1}{n}\sum_{i=1}^n\omega_i\|\boldsymbol{z}_i\|_2^2\right)^{1/2}.\end{aligned}$$

Next we deal with the data-dependent quantities $(1/n)\sum_{i=1}^n\omega_i$ and $(1/n)\sum_{i=1}^n\omega_i\|\boldsymbol{z}_i\|_2^2$. Note that $\mathbb{E}(\omega_i|\boldsymbol{x}_i) \leq \bar{g}h$ and thus $\mathbb{E}(\omega_i\|\boldsymbol{z}_i\|_2^2) \leq \bar{g}ph$. Moreover, $\omega_i\|\boldsymbol{z}_i\|_2^2 \leq \zeta_p^2$ (almost surely) and $\mathbb{E}(\omega_i^2\|\boldsymbol{z}_i\|_2^4) = \mathbb{E}(\omega_i\|\boldsymbol{z}_i\|_2^4) \leq \bar{g}\zeta_p^2 ph$. By Bernstein's inequality, together

$$\text{(E.27)}\quad \frac{1}{n}\sum_{i=1}^n\omega_i \leq \mathbb{E}\omega_i + \sqrt{\mathbb{E}\omega_i\cdot\frac{2t}{n}} + \frac{t}{3n} \leq \left(\sqrt{\mathbb{E}\omega_i} + \sqrt{\frac{t}{2n}}\right)^2 \leq \left(\sqrt{\bar{g}h} + \sqrt{\frac{t}{2n}}\right)^2$$

and

$$\text{(E.28)}\quad \frac{1}{n}\sum_{i=1}^n\omega_i\|\boldsymbol{z}_i\|_2^2 \leq \bar{g}ph + (\bar{g}ph)^{1/2}\zeta_p\sqrt{\frac{2t}{n}} + \zeta_p^2\frac{t}{3n} \leq \frac{3}{2}\bar{g}ph + \zeta_p^2\frac{4t}{3n}$$

hold with probability at least $1-2e^{-t}$. Let $\mathcal{E}_{\rm loc}(t)$ be the event that (E.27) and (E.28) hold. Putting the above four bounds together, we conclude that conditioned on $\mathcal{E}_{\rm loc}(t)$,

$$\text{(E.29)}\quad \begin{aligned}\sup_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\{D_0(\boldsymbol{v}) - D_0^\flat(\boldsymbol{v})\} &\leq \left(\frac{3}{2}\bar{g}ph + \zeta_p^2\frac{4t}{3n}\right)^{1/2}\frac{1}{2rn^{1/2}}\\ &\quad + \frac{h}{(4r)^2}\left\{\left(\sqrt{\bar{g}h} + \sqrt{\frac{t}{2n}}\right)\sqrt{\frac{2t}{n}} + \frac{4t}{3n}\right\}\end{aligned}$$

holds with $\mathbb{P}^*$-probability greater than $1-e^{-t}$.

Define the radius $r = h/(4\eta_{1/4})$ and event $\mathcal{E}_1(t) = \mathcal{E}_{\rm rsc}(t)\cap\mathcal{E}_{\rm loc}(t)$. Together, Lemma B.4, (E.26) and (E.29) imply that, with $\mathbb{P}^*$-probability at least $1-e^{-t}$ conditioned on $\mathcal{E}_1(t)$,

$$D^\flat(\boldsymbol{\beta},\boldsymbol{\beta}_j^*) \geq \frac{1}{2}\underline{g}\kappa_l\|\boldsymbol{\beta}-\boldsymbol{\beta}_j^*\|_\Sigma^2$$

holds uniformly over $\boldsymbol{\beta}\in\boldsymbol{\beta}_j^* + \Theta(r)$ as long as $nh \gtrsim \max(p, \zeta_p t^{1/2})$. Taking the union bound over $j = 0,1,\ldots,m$ concludes the proof. □

E.10. Proof of Lemma C.2. We proceed the proof via a covering argument. Let $\mathcal{N}$ be an $(1/2)$-net of the unit sphere $\mathbb{S}^{p-1}$ with cardinality $|\mathcal{N}| \le 5^p$ such that

$$\left\| \frac{1}{n}\sum_{i=1}^n (e_i \cdot \xi_i \boldsymbol{z}_i) \right\|_2 \le 2 \max_{\boldsymbol{u}\in\mathcal{N}_\epsilon} \frac{1}{n}\sum_{i=1}^n e_i \cdot \xi_i \langle \boldsymbol{u}, \boldsymbol{z}_i\rangle .$$

Since e_i's are independent Rademacher random variables and $|\xi_i| \le M$, conditional on the data $\{\boldsymbol{x}_i, \xi_i\}_{i=1}^n$, it follows from Hoeffding's inequality (see, e.g., Theorem 2.8 in [2]) that, with $\mathbb{P}^*$-probability (over $\{e_i\}_{i=1}^n$) at least $1 - e^{-u}$,

$$\frac{1}{n}\sum_{i=1}^n e_i \cdot \xi_i \langle \boldsymbol{u}, \boldsymbol{z}_i\rangle \le M \lambda_{\max}^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}\right)\sqrt{\frac{2u}{n}}, \tag{E.30}$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a symmetric matrix A. For the normalized empirical design matrix $(1/n)\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}$ satisfying $\mathbb{E}(\boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}) = \mathbf{I}_p$ and $\|\boldsymbol{z}_i\|_2 \le \zeta_p$ (almost surely), it follows from Theorem 5.41 in [16] that with probability at least $1 - n^{-2}$,

$$\lambda_{\max}^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}\right) \le 1 + C\zeta_p\sqrt{\frac{\log(2pn)}{n}}. \tag{E.31}$$

We denote by $\mathcal{E}_2$ the event that (E.31) holds.

Finally, taking a union bound over $\boldsymbol{u} \in \mathcal{N}$ and setting $u = 2(p + \log n)$ in (E.30), we conclude that with $\mathbb{P}^*$-probability at least $1 - n^{-2}$ conditioned on $\mathcal{E}_2$,

$$\left\| \frac{1}{n}\sum_{i=1}^n (e_i \cdot \xi_i \boldsymbol{z}_i) \right\|_2 \lesssim \left(1 + \zeta_p\sqrt{\frac{\log n}{n}}\right)\sqrt{\frac{p+\log n}{n}} \lesssim \sqrt{\frac{p+\log n}{n}},$$

as long as the sample size satisfies $n \gtrsim \zeta_p^2 \log n$. This proves the claimed bound. □

E.11. Proof of Lemma C.3. To avoid unnecessary repetitions, we only provide details for bounding $\Gamma_j(r)$. After a change of variable $\boldsymbol{v} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_j^*) \in \mathbb{B}^p(r)$ for $\boldsymbol{\beta} \in \boldsymbol{\beta}_j^* + \Theta(r)$, we write

(E.32)

$$\Gamma_j^{\Delta}(r) = \sup_{\boldsymbol{v}\in\mathbb{B}^p(r)} \|G_j(\boldsymbol{v})\|_2 := \sup_{\boldsymbol{v}\in\mathbb{B}^p(r)} \left\| \frac{1}{n}\sum_{i=1}^n e_i \cdot \{\bar{K}_h(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v} - \varepsilon_{ij}) - \bar{K}_h(-\varepsilon_{ij})\}\boldsymbol{z}_i \right\|_2,$$

where $\varepsilon_{ij} = y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^*$. The process $\{G_j(\cdot)\}$ satisfies $G_j(\mathbf{0}) = \mathbf{0}$, $\mathbb{E}^*\{G_j(\boldsymbol{v})\} = \mathbf{0}$ and $\nabla G_j(\boldsymbol{v}) = (1/n)\sum_{i=1}^n e_i \phi_{ij,\boldsymbol{v}} \boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}$, where $\phi_{ij,\boldsymbol{v}} = K_h(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v} - \varepsilon_{ij})$ and $K_h(x) = K(x/h)/h$. For any $\lambda \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{w} \in \mathbb{S}^{p-1}$, we have

$$\mathbb{E}^* \exp\{\lambda n^{1/2} \boldsymbol{u}^{\mathrm{T}} \nabla G_j(\boldsymbol{v}) \boldsymbol{w}\} = \prod_{i=1}^n \mathbb{E}^* \exp\{\lambda n^{-1/2} e_i \phi_{ij,\boldsymbol{v}} \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u} \cdot \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{w}\}$$

$$\le \prod_{i=1}^n \exp\left\{\frac{\lambda^2}{2n}\phi_{ij,\boldsymbol{v}}^2 (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u} \cdot \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{w})^2\right\} = \exp\left\{\frac{\lambda^2}{2n}\sum_{i=1}^n \phi_{ij,\boldsymbol{v}}^2 (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u} \cdot \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{w})^2\right\}.$$

Note that $\phi_{ij,\boldsymbol{v}} \le \kappa_u/h$, and by Hölder's inequality,

$$\frac{1}{n}\sum_{i=1}^n \phi_{ij,\boldsymbol{v}}^2 (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u} \cdot \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{w})^2 \le \frac{\kappa_u}{h} \cdot \left\{\frac{1}{n}\sum_{i=1}^n \phi_{ij,\boldsymbol{v}} (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u})^4\right\}^{1/2} \left\{\frac{1}{n}\sum_{i=1}^n \phi_{ij,\boldsymbol{v}} (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{w})^4\right\}^{1/2}.$$

Given $r > 0$, define the supremum

$$\Psi(r) = \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} \psi_{\boldsymbol{u},\boldsymbol{v}}(r) := \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} \frac{1}{n}\sum_{i=1}^n K_h(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v} - \varepsilon_{ij})(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u})^4.$$

Under this notation,

$$\sup_{\boldsymbol{v}\in\mathbb{B}^p(r)} \log \mathbb{E}^* \exp\{\lambda n^{1/2}\boldsymbol{u}^{\mathrm{T}}\nabla G_j(\boldsymbol{v})\boldsymbol{w}\} \le \frac{\kappa_u}{2h}\Psi(r)\lambda^2.$$

It then follows from a conditional version of Theorem A.3 in [12] that

$$\sup_{\boldsymbol{v}\in\mathbb{B}^p(r)} \|G_j(\boldsymbol{v})\|_2 \lesssim \kappa_u^{1/2}\Psi(r)^{1/2}\cdot r\sqrt{\frac{p+\log n}{nh}} \tag{E.33}$$

holds with $\mathbb{P}^*$-probability at least $1-n^{-2}$.

It remains to control the data-dependent quantity $\Psi(r) = \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} \psi_{\boldsymbol{u},\boldsymbol{v}}(r)$. For any $\epsilon_1,\epsilon_2\in(0,1)$, let $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{d_1}\}$ and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{d_2}\}$ be the ϵ_1- and ϵ_2-nets of $\mathbb{S}^{p-1}$ with cardinalities $d_1 \le (1+2/\epsilon_1)^p$ and $d_2 \le (1+2/\epsilon_2)^p$. Given $\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}$, there exist some $1\le \ell\le d_1$ and $1\le k\le d_2$ such that $\|\boldsymbol{u}-\boldsymbol{u}_\ell\|_2\le\epsilon_1$ and $\|\boldsymbol{v}-\boldsymbol{v}_k\|_2\le\epsilon_2$. Consider the decomposition

$$\psi_{\boldsymbol{u},\boldsymbol{v}}(r) = \psi_{\boldsymbol{u},\boldsymbol{v}}(r) - \psi_{\boldsymbol{u},\boldsymbol{v}_k}(r) + \psi_{\boldsymbol{u},\boldsymbol{v}_k}(r). \tag{E.34}$$

For $\psi_{\boldsymbol{u},\boldsymbol{v}}(r) - \psi_{\boldsymbol{u},\boldsymbol{v}_k}(r)$, the Lipschitz continuity of $K(\cdot)$ ensures that

(E.35)

$$|\psi_{\boldsymbol{u},\boldsymbol{v}}(r) - \psi_{\boldsymbol{u},\boldsymbol{v}_k}(r)| \le \frac{l_K r}{nh^2}\sum_{i=1}^n |\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{v}-\boldsymbol{v}_k)|(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u})^4 \le \frac{l_K}{h^2}\zeta_p^3\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}\right)\cdot r\epsilon_2.$$

For $\psi_{\boldsymbol{u},\boldsymbol{v}_k}(r)$, using the triangle inequality gives

$$\begin{aligned}
\psi_{\boldsymbol{u},\boldsymbol{v}_k}(r)^{1/4} &= \left\{\frac{1}{n}\sum_{i=1}^n K_h(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}_k - \varepsilon_{ij})\langle \boldsymbol{z}_i, \boldsymbol{u}_\ell + \boldsymbol{u} - \boldsymbol{u}_\ell\rangle^4\right\}^{1/4} \\
&\le \left\{\frac{1}{n}\sum_{i=1}^n K_h(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}_k - \varepsilon_{ij})(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u}_\ell)^4\right\}^{1/4} \\
&\quad + \left\{\frac{1}{n}\sum_{i=1}^n K_h(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}_k - \varepsilon_{ij})\langle \boldsymbol{z}_i, \boldsymbol{u} - \boldsymbol{u}_\ell\rangle^4\right\}^{1/4} \\
&\le \psi_{\boldsymbol{u}_\ell,\boldsymbol{v}_k}(r) + \epsilon_1\cdot \sup\nolimits_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\psi_{\boldsymbol{u},\boldsymbol{v}_k}(r)^{1/4}.
\end{aligned}$$

Taking the supremum over $\boldsymbol{u}\in\mathbb{S}^{p-1}$ and then the maximum over ℓ yields

$$\sup\nolimits_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\psi_{\boldsymbol{u},\boldsymbol{v}_k}(r) \le (1-\epsilon_1)^{-4}\max\nolimits_{1\le\ell\le d_1}\psi_{\boldsymbol{u}_\ell,\boldsymbol{v}_k}(r). \tag{E.36}$$

The problem then boils down to controlling the maximum $\max_{(\ell,k)\in[d_1]\times[d_2]}\psi_{\boldsymbol{u}_\ell,\boldsymbol{v}_k}(r)$. Note that $\mathbb{E}K_h^2(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}_k - \varepsilon_{ij})(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u}_\ell)^8 \le \overline{f}\kappa_u m_4\zeta_p^4/h$ and $K_h(r\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{v}_k - \varepsilon_{ij})(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u}_\ell)^4 \le \kappa_u\zeta_p^4/h$. By Bernstein's inequality, we have that with probability at least $1-e^{-u}$,

$$\psi_{\boldsymbol{u}_\ell,\boldsymbol{v}_k}(r) \le \mathbb{E}\psi_{\boldsymbol{u}_\ell,\boldsymbol{v}_k}(r) + (\overline{f}\kappa_u m_4)^{1/2}\zeta_p^2\sqrt{\frac{2u}{nh}} + \kappa_u\zeta_p^4\frac{u}{3nh}$$

$$\leq \overline{f} m_4 + (\overline{f}\kappa_u m_4)^{1/2}\zeta_p^2\sqrt{\frac{2u}{nh}} + \kappa_u\zeta_p^4\frac{u}{3nh}$$
$$\leq \frac{3}{2}\overline{f} m_4 + \kappa_u\zeta_p^4\frac{4u}{3nh}.$$

Taking $\epsilon_1 = 1 - 2^{-1/4}$, $\epsilon_2 = n^{-2}$ and $u = p\log(1+2/\epsilon_1)(1+2/\epsilon_2) + \log(n)$ in the above bounds, we conclude from (E.31) and (E.34)–(E.36) that with probability at least $1 - n^{-1}$,

$$\Psi(r) \lesssim m_4 + \zeta_p^4\frac{p\log n}{nh} + \zeta_p^3\frac{r}{(nh)^2}$$

as long as $n \gtrsim \zeta_p^2 \log n$. Substituting this bound into (E.33) completes the proof for a particular $j \in \{0, \ldots, m\}$. The claimed uniform bound follows from a union bound over the grid points. □

E.12. Proof of Lemma C.4. For $j = 0, \ldots, m$, define the random process

$$\Lambda_j^\flat(\boldsymbol{\beta}) = \widehat{Q}_j^\flat(\boldsymbol{\beta}) - \widehat{Q}_j^\flat(\boldsymbol{\beta}_j^*) - \mathbf{J}_j(\boldsymbol{\beta} - \boldsymbol{\beta}_j^*), \;\; \boldsymbol{\beta} \in \mathbb{R}^p.$$

The goal is to bound the local fluctuation $\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j^\flat(\boldsymbol{\beta})\|_2$ for a prespecified $r > 0$. Since $\mathbb{E}(W_i) = 1$, we have $\mathbb{E}^*\{\widehat{Q}_j^\flat(\boldsymbol{\beta})\} = \widehat{Q}_j(\boldsymbol{\beta})$, and $\mathbb{E}^*\{\Lambda_j^\flat(\boldsymbol{\beta})\} = \Lambda_j(\boldsymbol{\beta})$, where

$$\Lambda_j(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \Delta_i\big\{\bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - y_i) - \bar{K}_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j^* - y_i)\big\}\boldsymbol{x}_i - \mathbf{J}_j(\boldsymbol{\beta} - \boldsymbol{\beta}_j^*).$$

Consequently, by the triangle inequality,

(E.37)
$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j^\flat(\boldsymbol{\beta})\|_{\Sigma^{-1}} \leq \underbrace{\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j^\flat(\boldsymbol{\beta}) - \Lambda_j(\boldsymbol{\beta})\|_{\Sigma^{-1}}}_{=\Gamma_j^\Delta(r) \text{ by (C.3)}} + \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j(\boldsymbol{\beta})\|_{\Sigma^{-1}}.$$

For the first term on the right-hand side of (E.37), Lemma C.3 guarantees the existence of an event $\mathcal{E}_3$ with $\mathbb{P}(\mathcal{E}_3) \geq 1 - 3n^{-1}$ such that, conditioned on $\mathcal{E}_3$,

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j^\flat(\boldsymbol{\beta}) - \Lambda_j(\boldsymbol{\beta})\|_2 \lesssim \left(m_4^{1/2} + \zeta_p^2\sqrt{\frac{p\log n}{nh}}\right)\cdot r\sqrt{\frac{p+\log n}{nh}}, \tag{E.38}$$

holds with $\mathbb{P}^*$-probability at least $1 - n^{-2}$, provided $n \gtrsim \zeta_p^2 \log n$. Moreover, let $\mathcal{E}_3'$ be the event that (B.6) holds for all $j = 0, \ldots, m$ with $t = 2\log n$, so that $\mathbb{P}(\mathcal{E}_3') \geq 1 - (m+1)n^{-2}$. Conditioning on $\mathcal{E}_3'$,

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j(\boldsymbol{\beta})\|_{\Sigma^{-1}} \lesssim \left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r + h\right)\cdot r, \tag{E.39}$$

provided that $nh \gtrsim \zeta_p^2(p+t)$.

Finally, taking $\mathcal{E}_4 = \mathcal{E}_3 \cap \mathcal{E}_3'$ so that $\mathbb{P}(\mathcal{E}_4) \geq 1 - (m+3n+1)n^{-2}$. Together, (E.37)–(E.39) yield that with $\mathbb{P}^*$-probability at least $1 - n^{-2}$ conditioned on $\mathcal{E}_4$,

$$\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}_j^*+\Theta(r)} \|\Lambda_j^\flat(\boldsymbol{\beta})\|_2 \lesssim \left\{m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r + h + \zeta_p^2\frac{(p\log n)^{1/2}(p+\log n)^{1/2}}{nh}\right\}\cdot r,$$

completing the proof. □

E.13. Proof of Lemma C.5. For $j = 1, \ldots, m$, define the random processes

$$R_j^\flat(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})$$
$$= \frac{1}{n}\sum_{i=1}^n \sum_{\ell=0}^{j-1} \int_{\tau_\ell}^{\tau_{\ell+1}} \mathrm{d}H(u)\Big[W_i\big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell)\big\}\boldsymbol{z}_i - \Sigma^{-1/2}\mathbf{H}_\ell(\boldsymbol{\beta}_\ell - \boldsymbol{\beta}_\ell^*)\Big]$$

and

$$R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})$$
$$= \frac{1}{n}\sum_{i=1}^n \sum_{\ell=0}^{j-1} \int_{\tau_\ell}^{\tau_{\ell+1}} \mathrm{d}H(u)\Big[\big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell)\big\}\boldsymbol{z}_i - \Sigma^{-1/2}\mathbf{H}_\ell(\boldsymbol{\beta}_\ell - \boldsymbol{\beta}_\ell^*)\Big],$$

and note that $\mathbb{E}^* R_j^\flat(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1}) = R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})$. By the triangle inequality,

(E.40)
$$\sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)\}} ||R_j^\flat(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})||_2 \le \sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)\}} ||R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})||_2$$
$$+ \sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)\}} ||R_j^\flat(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1}) - R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})||_2.$$

Let $\mathcal{E}_5$ be the event that (B.5) holds for all $\ell = 0, \ldots, m-1$ with $t = 2\log n$. Then, $\mathbb{P}(\mathcal{E}_5) \ge 1 - mn^{-2}$, and conditional on $\mathcal{E}_5$,

(E.41)
$$\sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)\}} ||R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})||_2 \lesssim \log\left(\tfrac{1-\tau_0}{1-\tau_j}\right) \cdot \left(m_4^{1/2}\sqrt{\frac{p+\log n}{nh}} + m_3 r + h\right) \cdot r$$

holds for all $j = 0, 1, \ldots, m$ provided that $nh \gtrsim \zeta_p^2(p+t)$. For the second term on the right-hand side of (E.40), note that

$$\sup_{\cap_{\ell=0}^{j-1}\{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)\}} ||R_j^\flat(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1}) - R_j(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{j-1})||_2$$
$$\le \sum_{\ell=0}^{j-1} \sup_{\boldsymbol{\beta}_\ell \in \boldsymbol{\beta}_\ell^* + \Theta(r)} \left\| \frac{1}{n}\sum_{i=1}^n e_i\big\{\bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell^*) - \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\ell)\big\}\boldsymbol{z}_i \right\|_2 \cdot \log\left(\tfrac{1-\tau_\ell}{1-\tau_{\ell+1}}\right)$$

(E.42)
$$= \sum_{\ell=0}^{j-1} \Gamma_\ell(r) \cdot \log\left(\tfrac{1-\tau_\ell}{1-\tau_{\ell+1}}\right) \le \max_{\ell=0,\ldots,j-1} \Gamma_\ell(r) \cdot \log\left(\tfrac{1-\tau_0}{1-\tau_j}\right),$$

where $\{\Gamma_\ell(r)\}_{\ell=0}^{j-1}$ are defined in (C.2).

Combining (E.40)–(E.42) proves the claimed result. □

E.14. Proof of Lemma C.6. We first show an unconditional version

(E.43)
$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{|\tau_1-\tau_2|<\delta} \big|\mathbb{G}_n^\flat(\tau_1) - \mathbb{G}_n^\flat(\tau_2)\big| > x \right\} = 0$$

via the proving techniques of Lemma B.8. By the law of total expectation and following the arguments that lead to (E.18), it can be verified that

$$\|\mathbb{G}_n^\flat(\tau_1) - \mathbb{G}_n^\flat(\tau_2)\|_\psi \lesssim (m_4/h)^{1/2}|\tau_2 - \tau_1|,$$

for any $\tau_1 < \tau_2$ satisfying $\tau_2 - \tau_1 \geq (\zeta_p/m_4)^{1/2}(h/n)^{1/2}$, where $\psi(x) = x^2$ so that $||\cdot||_\psi$ coincides with the L_2-norm. The rest follows from the same packing argument as in the proof of Lemma B.8.

Then, we prove the claimed result by contradiction. Conditioned on a sequence of observed data $\{\mathbb{D}_n\}$, if (C.4) does not hold, then there is a sequence $\{\delta_m : m \subset \mathbb{N}^+\}$ such that $\lim_{m\to\infty} \delta_m = 0$, and a subsequence of natural numbers $\{n_k : k \subset \mathbb{N}^+\} \subset \mathbb{N}^+$ such that

$$\lim_{m\to\infty}\lim_{k\to\infty} \underbrace{\mathbb{P}^*\left\{\sup_{|\tau_1-\tau_2|<\delta_m} \left|\mathbb{G}^\flat_{n_k}(\tau_1) - \mathbb{G}^\flat_{n_k}(\tau_2)\right| > x\right\}}_{:=\chi_{m,k}} > 0 \text{ over } \mathcal{A},$$

where $\mathcal{A}$ is an event over the data space with $\mathbb{P}(\mathcal{A}) = p_0 > 0$. This further means there are sufficiently large integers M and K such that $\chi_{m,k} \geq c_0 > 0$ over $\mathcal{A}$ for any $m \geq M$ and $k \geq K$. On the other hand, by the law of total probability,

$$\mathbb{P}\left\{\sup_{|\tau_1-\tau_2|<\delta_m} \left|\mathbb{G}^\flat_{n_k}(\tau_1) - \mathbb{G}^\flat_{n_k}(\tau_2)\right| > x\right\} \geq \mathbb{E}\big[\chi_{m,k}\cdot \mathbb{1}\{\mathcal{A}\}\big] \geq c_0 p_0,$$

for any $m \geq M$ and $k \geq K$. So far, we have identified subsequences $\{\delta_m\}$ and $\{n_k\}$, such that $\lim_{m\to\infty}\delta_m = 0$, $\lim_{k\to\infty} n_k = \infty$, and

$$\lim_{m\to\infty}\lim_{k\to\infty} \mathbb{P}\left\{\sup_{|\tau_1-\tau_2|<\delta_m} \left|\mathbb{G}^\flat_{n_k}(\tau_1) - \mathbb{G}^\flat_{n_k}(\tau_2)\right| > x\right\} \geq c_0 p_0 > 0,$$

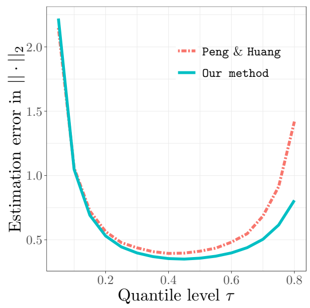
which contradicts with (E.43), thus completes the proof of (C.4). □

APPENDIX F: ADDITIONAL SIMULATION STUDIES

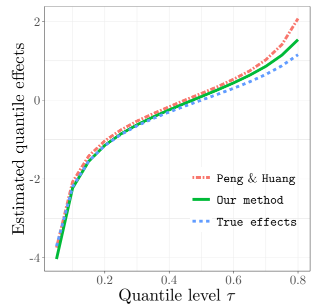
F.1. More comparisons in low dimensions. We first compare the SEE method with its non-smoothed counterpart [11] on small-scale datasets. Let $n \in \{200, 400\}$ and $p = 5$. We generate covariates $\boldsymbol{x}_i$ $(i = 1, \ldots, n)$ from $\mathcal{N}(\mathbf{0}, \Sigma = (\sigma_{jk})_{1\leq j,k\leq 5})$, where $\sigma_{jk} = 0.5^{|j-k|}$. The remaining settings are the same as in Section 5.1. The results on the ℓ_2-error, estimated quantile effects and running time are presented in Figures F.1 and F.2. Taking also into account Figure 1 in the main text, we see that the SEE method exhibits desirable finite-sample performance especially at higher quantile levels. The computational advantage becomes more prominent for large-scale data.

When both the sample size and dimension increase, Figure 3 in Section 5.1 presents the ℓ_2-errors and runtimes, both as functions of n, at $\tau = 0.7$. Since the censoring rates vary from 30% to 45% in these settings, the estimation at $\tau = 0.7$ is prone to some “boundary" issues related to the non-identifiability of upper quantiles. To have a more complete picture of the performance, Figure F.3 below shows the simulation results at quantile levels $\tau \in \{0.3, 0.5, 0.7\}$. The benefit of smoothing becomes more evident as τ increases.

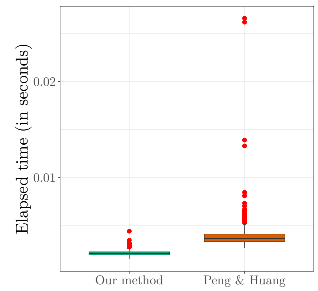
F.2. Reports of selection consistency. For the speed comparison between SCQR-Lasso and CQR-Lasso [17] in Section 5.2, we applied a simpler tuning scheme (to both methods) by setting $\lambda_k = \{1 + \log(\frac{1-\tau_L}{1-\tau_k})\}\lambda_0$ for $k = 1, \ldots, m$ and choosing λ_0 via cross-validation. The variable selection performance in terms of TPR and FDR is shown in Figure F.4, which complements Figure 5. As remarked in Section 5.2, [17] considered equally-spaced λ values, i.e. $\lambda_k = \lambda_0 + c\cdot k$ for $k = 0, 1, \ldots, m$, and used cross-validation to choose λ_0 and c by a two-dimensional grid search. Since this tuning scheme is not adopted, the selection measures are presented for reference only.



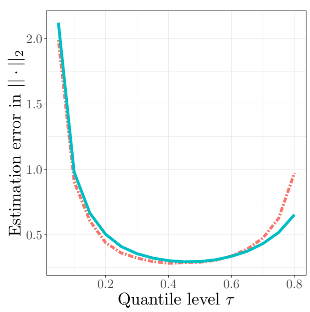
(a) ℓ_2-error under model (31)



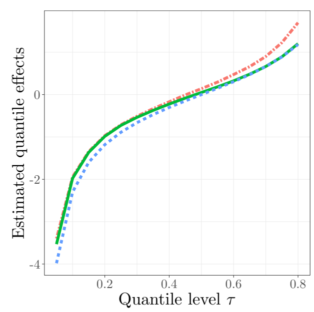
(b) Estimated quantile effects under model (31)



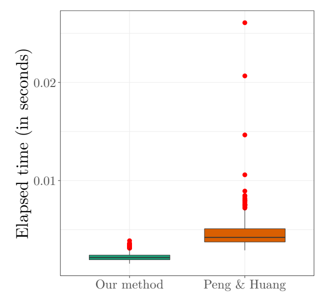
(c) Runtime under model (31)



(d) ℓ_2-error under model (32)



(e) Estimated quantile effects under model (32)



(f) Runtime under model (32)

FIG F.1. *Numerical comparisons between the smoothed approach and [11]'s method when* $(n, p) = (200, 5)$. *See Figure 1 in the main text for more details.*

REFERENCES

[1] BARZILAI, J. and BORWEIN, J. M. (1988). Two-point step size gradient methods. *IMA J. Numer. Anal.* **8** 141–148.

[2] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford.

[3] BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications* **56** 213–247. Birkhäuser, Basel.

[4] CHAO, S.-K., VOLGUSHEV, S. and CHENG, G. (2017). Quantile processes for semi and nonparametric regression. *Electron. J. Stat.* **11** 3272–3331.

[5] FAN, J., LIU, H., SUN, Q. and ZHANG, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Ann. Statist.* **46** 814–841.

[6] HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2021). Smoothed quantile regression with large-scale inference. *J. Econom.*, to appear. https://doi.org/10.1016/j.jeconom.2021.07.010.

[7] KLEY, T., VOLGUSHEV, S., DETTE, H. and HALLIN, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli* **22** 1770–1807.

[8] LANGE, K., HUNTER, D. R. and YANG, I. (2000). Optimization transfer using surrogate objective functions. *J. Comput. Graph. Stat.* **9** 1–20.

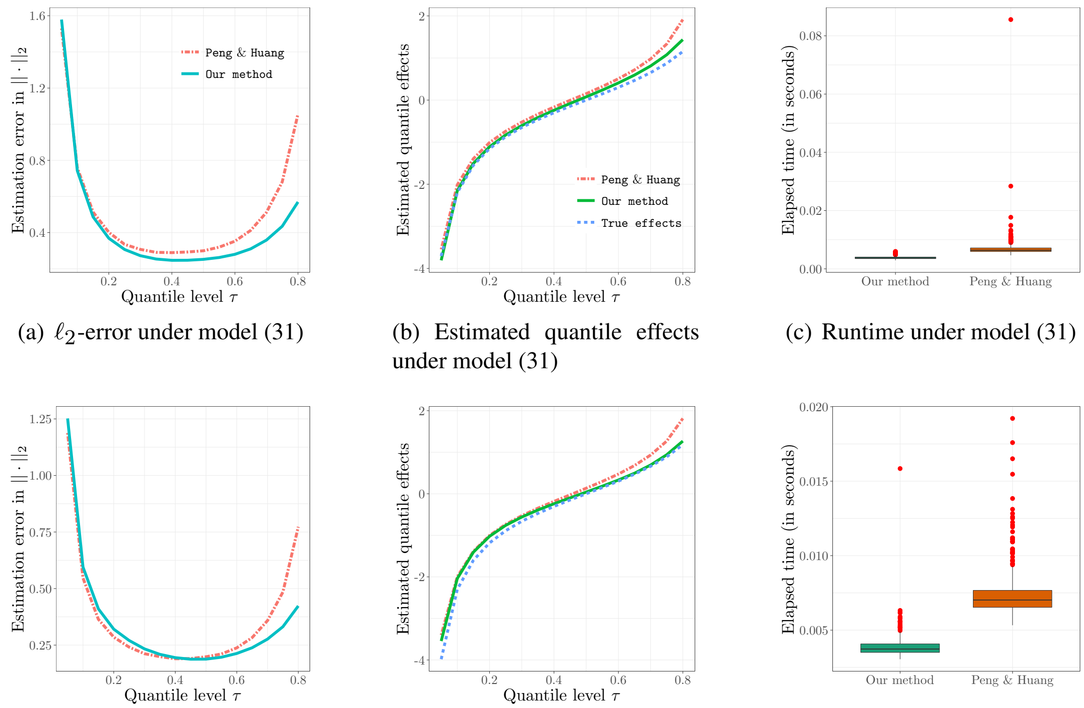
[9] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag, Berlin.

[10] LIN, D. Y., WEI, L. J. and YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80** 557–572.

[11] PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **103** 637–649.

[12] SPOKOINY, V. (2013). Bernstein–von Mises theorem for growing parameter dimension. *arXiv preprint arXiv:1302.3430*.

[13] TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. Roy. Statist. Soc. Ser. B* **84** 205–233.



(a) ℓ_2-error under model (31)

(b) Estimated quantile effects under model (31)

(c) Runtime under model (31)

(d) ℓ_2-error under model (32)

(e) Estimated quantile effects under model (32)

(f) Runtime under model (32)

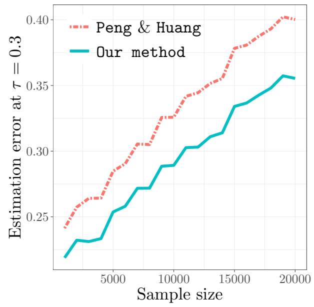
FIG F.2. *Numerical comparisons between the smoothed approach and [11]'s method when* $(n, p) = (400, 5)$. *See Figure 1 in the main text for more details.*

[14] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. Ser. B* **58** 267–288.

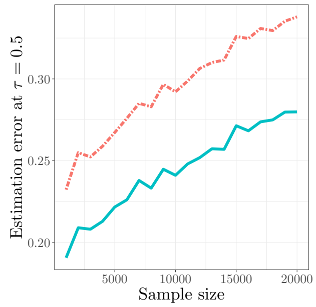
[15] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

[16] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* (Y. Eldar and G. Kutyniok, eds.) 210–268. Cambridge Univ. Press, Cambridge.

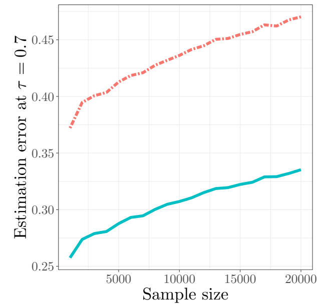
[17] ZHENG, Q., PENG, L. and HE, X. (2018). High dimensional censored quantile regression. *Ann. Statist.* **46** 308–343.



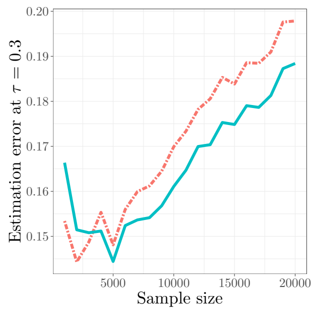
(a) ℓ_2-error at $\tau = 0.3$ under model (31)



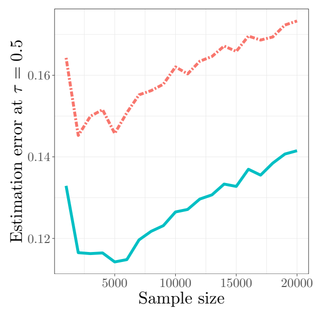
(b) ℓ_2-error at $\tau = 0.5$ under model (31)



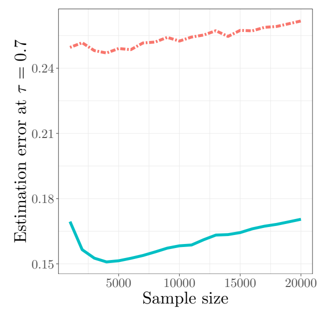
(c) ℓ_2-error at $\tau = 0.7$ under model (31)



(d) ℓ_2-error at $\tau = 0.3$ under model (32)

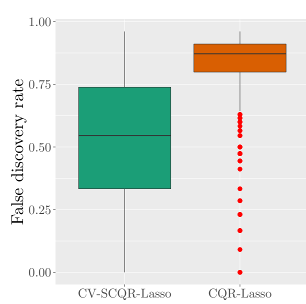


(e) ℓ_2-error at $\tau = 0.5$ under model (32)

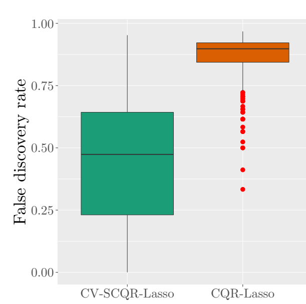


(f) ℓ_2-error at $\tau = 0.7$ under model (32)

FIG F.3. *Plots of ℓ_2-errors versus sample size at $\tau \in \{0.3, 0.5, 0.7\}$ for [11]'s method and the smoothed approach under models (31)–(32) with increasing (n, p) subject to $p = n/100$.*



(a) FDR under homoscedastic model



(b) FDR under heteroscedastic model

FIG F.4. *Box plots of false discovery rate (FDR) for ℓ_1-penalized CQR (CQR-Lasso) and cross-validated ℓ_1-penalized smoothed CQR (CV-SCQR-Lasso). For the homoscedastic model, the average true positive rate (TPR) is 1.00 for both methods; for the heteroscedastic model, the average TPR is 0.9872 for CV-SCQR-Lasso, and 1.00 for CQR-Lasso. Other settings are the same as Figure 5.*