SUPPLEMENTARY MATERIAL FOR “SCALABLE ESTIMATION AND INFERENCE FOR CENSORED QUANTILE REGRESSION PROCESS”

BY Xuming He, Xiaou Pan, and Kean Ming Tan

1Department of Statistics, University of Michigan, xmehe@umich.edu; *keanming@umich.edu
2Department 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 \( \hat{L}_k(\cdot) \) \((k = 0, 1, \ldots, m)\) 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.

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

\[
\nabla \hat{L}_k(\beta) = \nabla \hat{L}_0(\beta) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{k-1} K_h(y_i - \hat{x}_i^j \hat{\beta}_j) \{H(\tau_{j+1}) - H(\tau_j)\} \hat{x}_i.
\]

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

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

Input: Censored data \( \{(y_i, \hat{x}_i, \Delta_i)\}_{i=1}^{n} \), current quantile level \( \tau_k \in (0, 1) \), previous estimates \( \{\hat{\beta}_j\}_{j=0}^{k-1} \), initial values \( \hat{\beta}_0 = \hat{\beta}_{k-1} \), bandwidth \( h \), step size upper bound \( \alpha_{\max} \), tolerance level \( \epsilon \).

1: Compute \( \hat{\beta}^{(1)}_k = \hat{\beta}_k^{(0)} - \nabla \hat{L}_k(\hat{\beta}_k^{(0)}) \)
2: for \( t = 1, 2, \ldots \) do
3: \( s_t \leftarrow \hat{\beta}_k^{(t)} - \hat{\beta}_k^{(t-1)} \), \( g_t \leftarrow \nabla \hat{L}_k(\hat{\beta}_k^{(t-1)}) - \nabla \hat{L}_k(\hat{\beta}_k^{(t-1)}) = \nabla \hat{L}_0(\hat{\beta}_k^{(t)}) - \nabla \hat{L}_0(\hat{\beta}_k^{(t-1)}) \)
4: \( \alpha_{t,1} \leftarrow ||s_t||_2^2/(s_t, g_t) \), \( \alpha_{t,2} \leftarrow (s_t, g_t)/||g_t||_2 \)
5: \( \alpha_t \leftarrow \min\{\alpha_{t,1}, \alpha_{t,2}, \alpha_{\max}\} \) if \( s_t, g_t > 0 \), and \( \alpha_t \leftarrow 1 \) otherwise
6: Update \( \hat{\beta}_k^{(t+1)} = \hat{\beta}_k^{(t)} - \alpha_t \nabla \hat{L}_k(\hat{\beta}_k^{(t)}) \)
7: end for when \( ||\nabla \hat{L}_k(\hat{\beta}_k^{(t)})||_2 \leq \epsilon \)

A.2. High-dimensional setting. In the high dimensional regime, we need to solve the following weighted \( \ell_1 \)-penalized programs sequentially:

\[
(A.1) \quad \hat{\beta}_k = \arg\min_{\beta \in \mathbb{R}^p} \{\hat{L}_k(\beta) + \|\lambda_k \circ \beta\|_1\}, \quad k = 0, \ldots, m,
\]
where \(\circ\) denotes the Hadamard product, and \(\lambda_k = (\lambda_{k,1}, \ldots, \lambda_{k,p})^T\) may depend on the previous estimates \(\{\hat{\beta}_j\}_{j=0}^{k-1}\). 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 \(\beta_0\), the majorize-minimize (MM) algorithm involves two steps: first, construct a majorizing function \(g(\cdot | \beta_0)\), satisfying
\[
g(\beta | \beta_0) = f(\beta_0) \quad \text{and} \quad g(\beta | \beta_0) \geq f(\beta) \quad \text{for any } \beta \in \mathbb{R}^p;
\]

secondly, update \(\beta_0\) by \(\beta_1 := \arg\min_{\beta \in \mathbb{R}^p} g(\beta | \beta_0)\) [8]. Noting that
\[
f(\beta_1) \leq g(\beta_1 | \beta_0) \leq g(\beta_0 | \beta_0) = f(\beta_0),
\]
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
\[
g(\beta_0 | \beta_0) = f(\beta_0) \quad \text{and} \quad g(\beta_1 | \beta_0) \geq f(\beta_1).
\]

To construct a proper majorizing function for \(\hat{L}_k(\cdot)\) around \(\beta_0\), we define an isotropic quadratic function
\[
F(\beta; \phi, \beta_0) := \hat{L}_k(\beta_0) + \langle \nabla \hat{L}_k(\beta_0), \beta - \beta_0 \rangle + \frac{\phi}{2} \| \beta - \beta_0 \|_2^2
\]
for some \(\phi > 0\). It is easy to see that \(F(\beta_0; \phi, \beta_0) = \hat{L}_k(\beta_0)\). Using such a surrogate loss function, the weighted \(\ell_1\)-penalized program \(\min_{\beta \in \mathbb{R}^p} \{F(\beta; \phi, \beta_0) + \| \lambda_k \circ \beta \|_1\}\) admits a closed-form solution \(\beta_1 = S_{\text{soft}}(\beta_0 - \nabla \hat{L}_k(\beta_0)/\phi, \lambda_k/\phi)\), where \(S_{\text{soft}}(\beta, \lambda) := \{\text{sign}(\beta_j) \max\{\|\beta_j\| - \lambda_{k,j}, 0\}\}_{j=1,...,p}\) is the soft-thresholding operator. Moreover, the quadratic coefficient \(\phi > 0\) should be sufficiently large so that the local majorization property \(F(\beta_1; \phi, \beta_0) \geq \hat{L}_k(\beta_1)\) is satisfied. Starting with a relatively small value \(\phi = \phi_0\), we iteratively increase \(\phi\) by a factor of \(\gamma > 1\) and compute
\[
\beta_{1,\ell} = S_{\text{soft}}(\beta_0 - \nabla \hat{L}_k(\beta_0)/\phi_{\ell}, \lambda_k/\phi_{\ell}) \quad \text{with} \quad \phi_{\ell} = \gamma^\ell \phi_0, \quad \ell = 0, 1, \ldots,
\]
until the local majorization property holds. Repeating this procedure yields a sequence of iterates \(\{\beta_\ell\}_{\ell=0,1,...}\) until the stopping criterion is met, say \(\|\beta_{\ell+1} - \beta_{\ell}\|_2 \leq \epsilon\). We treat \(\phi_0, \gamma, \epsilon\) 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 \(\beta^* = \beta^*(\tau_j)\) and \(\hat{\beta}_j = \hat{\beta}(\tau_j)\), \(j = 0, \ldots, m\). Since the estimators \(\{\hat{\beta}_j\}_{j=0}^{m}\) are constructed sequentially, the statistical error of \(\hat{\beta}_j\) at quantile level \(\tau_j\) depends on the accumulated errors of \(\hat{\beta}_0, \ldots, \hat{\beta}_{j-1}\).

For every \(r > 0\), define the local ellipse \(\Theta(\tau) = \{\delta \in \mathbb{R}^p : \|\delta\|_\Sigma \leq r\}\) under the \(\Sigma\)-induced norm. Under Condition 3.2 on the (random) feature vector \(x \in \mathbb{R}^p\), for every \(\delta \in (0, 1]\) we set
\[
\eta_\delta = \inf\{\eta > 0 : \mathbb{E}\{(z^T\nu)^2 1\{|z^T\nu| > \eta\}\} \leq \delta \quad \text{for all} \quad \nu \in \mathbb{S}^{p-1}\},
\]
where \(z = \Sigma^{-1/2}x\). Since \(\mathbb{E}(z^T\nu)^2 = 1\) for any \(\nu \in \mathbb{S}^{p-1}\), \(\eta_\delta\) is well-defined for each \(\delta\), and depends implicitly on the underlying distribution of \(z\). Throughout the proof, we write
\[
z_i = (z_{i1}, \ldots, z_{ip})^T = \Sigma^{-1/2}x_i, \quad i = 1, \ldots, n.
\]
For a non-negative kernel function $K(\cdot)$ and a bandwidth $h > 0$, we write
\[
K_h(u) = h^{-1}K(u/h), \quad \tilde{K}_h(u) = \tilde{K}(u/h) \quad \text{and} \quad \tilde{K}(u) = \int_{-\infty}^{u} K(t) \, dt.
\]

**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 $Q(\beta) : \mathbb{R}^p \to \mathbb{R}^p$ with positive semi-definite Jacobian, define the corresponding divergence function $D = D_Q : \mathbb{R}^p \times \mathbb{R}^p \to [0, \infty)$ as $D(\beta_1, \beta_2) = \langle Q(\beta_1) - Q(\beta_2), \beta_1 - \beta_2 \rangle$. For any $\beta, \beta' \in \mathbb{R}^p$ and $\eta \in [0, 1]$, we have
\[
D(\beta' + \eta(\beta - \beta'), \beta') \leq \eta D(\beta, \beta').
\]

Lemma B.2 below provides a Bernstein-type inequality for the $\ell_2$-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 $E(\xi_i^2|x_i) \leq \sigma^2$ and $|\xi_i| \leq M$ for some $M \geq \sigma > 0$. Then, for any $t > 0$,
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} (\xi_i z_i - E(\xi_i z_i)) \right\|_2 \leq 2\sigma \sqrt{\frac{p}{n}} + \sigma \sqrt{\frac{2t}{n}} + M \zeta_p \frac{4t}{3n}
\]
holds with probability at least $1 - e^{-t}$, where $\zeta_p = \max_{x \in \mathcal{X}} \|x\|_{\Sigma^{-1}}$.

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

(i) With probability at least $1 - e^{-t},$
\[
\left\| \tilde{Q}_0(\beta_j^*) - E \tilde{Q}_0(\beta_j^*) \right\|_{\Sigma^{-1}} \leq 2\tau_0 \left( \sqrt{\frac{p}{n} + \frac{t}{2n}} + \zeta_p \frac{2t}{3n} \right),
\]
where $\tau_0 = \max(\tau_0, 1 - \tau_0)$.

(ii) With probability at least $1 - e^{-t},$
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \left(1 - E\right) \tilde{K}_h(y_i - x_i^T \beta_j^*) x_i \right\|_{\Sigma^{-1}} \leq 2 \left( \sqrt{\frac{p}{n} + \frac{t}{2n}} + \zeta_p \frac{2t}{3n} \right).
\]

(iii) With probability at least $1 - e^{-t},$
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \left(1 - E\right) \int_{\tau_j}^{\tau_j+1} \{ \tilde{K}_h(y_i - x_i^T \beta^*(u)) - \tilde{K}_h(y_i - x_i^T \beta^*(\tau_j)) \} \, dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \\
\lesssim m_j^{1/2} \bar{w}_j \cdot \delta^* \sqrt{\frac{p + t}{nh}} + \bar{w}_j \cdot \zeta_p^2 \delta^* \frac{t}{nh},
\]
where $w_j = H(\tau_{j+1}) - H(\tau_j) = \log \left( \frac{1 - \tau_j}{1 - \tau_{j+1}} \right)$.
The following lemma concerns the first-order property of the smoothed estimating functions $\hat{Q}_j(\cdot)$ in (3) and (4). Define the corresponding symmetrized Bregman divergence $D : \mathbb{R}^p \times \mathbb{R}^p \to [0, \infty)$ as

\begin{equation}
D(\beta_1, \beta_2) := \langle \hat{Q}_j(\beta_1) - \hat{Q}_j(\beta_2), \beta_1 - \beta_2 \rangle = \langle \check{Q}_0(\beta_1) - \check{Q}_0(\beta_2), \beta_1 - \beta_2 \rangle.
\end{equation}

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 $4\eta_{1/4}^2 r \leq h \leq 1$ with $\eta_{1/4}$ defined in (B.1). Then, for any $0 \leq j \leq m$ and $t > 0$,

\[
\inf_{\beta \in \beta^*_j + \Theta(r)} \frac{D(\beta, \beta^*_j)}{\kappa_1 \|\beta - \beta^*_j\|_\Sigma^2} \geq \frac{3}{4} \frac{g^2}{\beta^2} \geq \frac{5}{4} \sqrt{\frac{ht}{n} + \sqrt{\frac{ht}{8n}}} - cr^{-2} \frac{ht}{n}
\]

with probability at least $1 - e^{-t}$, where $c = 1/4 + 1/48 \approx 0.27$.

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

\begin{equation}
\omega_j(r) := \sup_{\beta \in \beta^*_j + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^n \left\{ \bar{K}_h(y_i - x_i^T \beta) - \bar{K}_h(y_i - x_i^* \beta^*_j) \right\} x_i + H_j(\beta - \beta^*_j) \right\|_{\Sigma^{-1}},
\end{equation}

\begin{equation}
\omega_{j*}(r) := \sup_{\beta \in \beta^*_j + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^n \Delta_i \left\{ \bar{K}_h(x_i^T \beta - y_i) - \bar{K}_h(x_i^* \beta^*_j - y_i) \right\} x_i - J_j(\beta - \beta^*_j) \right\|_{\Sigma^{-1}},
\end{equation}

where $J_j = J(\tau_j) = \mathbb{E}\{g(x^T \beta^*_j | x)x^T\}$ and $H_j = H(\tau_j) = \mathbb{E}\{f_j(x^T \beta^*_j | x)x^T\}$. The following lemma provides upper bounds for the two suprema $\omega_j(r)$ and $\omega_{j*}(r)$ for any given $r > 0$.

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

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

\begin{equation}
\sup_{\beta \in \beta^*_j + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^n (1 - \mathbb{E}) \left\{ \bar{K}_h(y_i - x_i^T \beta) - \bar{K}_h(y_i - x_i^* \beta^*_j) \right\} x_i \right\|_{\Sigma^{-1}} \lesssim (\kappa_m g m_4)^{1/2} \sqrt{\frac{p + t}{nh}} \cdot r,
\end{equation}

provided that the “effective” sample size satisfies $nh \gtrsim \zeta_0^2 (p + t)$. Moreover, for any $\beta \in \beta^*_j + \Theta(r)$,

\[
\left\| \mathbb{E}\{\bar{K}_h(y - x^T \beta) - \bar{K}_h(y - x^* \beta^*_j)\} x + H_j(\beta - \beta^*_j) \right\|_{\Sigma^{-1}} \leq l_1 (0.5 m_3 r + \kappa_1 h) \cdot r.
\]

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

\begin{equation}
\sup_{\beta \in \beta^*_j + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^n (1 - \mathbb{E}) \Delta_i \left\{ \bar{K}_h(x_i^T \beta - y_i) - \bar{K}_h(x_i^* \beta^*_j - y_i) \right\} x_i \right\|_{\Sigma^{-1}} \lesssim (\kappa_m g m_4)^{1/2} \sqrt{\frac{p + t}{nh}} \cdot r,
\end{equation}

\[
\sum_{i=1}^n (1 - \mathbb{E}) \Delta_i \left\{ \bar{K}_h(x_i^T \beta - y_i) - \bar{K}_h(x_i^* \beta^*_j - y_i) \right\} x_i \right\|_{\Sigma^{-1}} \lesssim (\kappa_m g m_4)^{1/2} \sqrt{\frac{p + t}{nh}} \cdot r,
\]
provided that \( n\varphi \gtrsim \zeta \Delta^2 (p + t) \). Moreover, for any \( \beta \in \beta^*_j + \Theta(r) \),

\[
\left\| \mathbb{E} \Delta \{ \widetilde{K}_h(x^T \beta - y) - \widetilde{K}_h(x^T \beta^*_j - y) \} x - \mathbf{J}_j (\beta - \beta^*_j) \right\|_{\Sigma^{-1}} \leq l_1 (0.5m_3 r + \kappa_1 h) \cdot r.
\]

**Lemma B.6.** Assume Conditions 3.1–3.3 hold. For any \( \tau_L \leq \tau_l < \tau_u \leq \tau_U \),

\[
\sup_{\tau \in [\tau_l, \tau_u]} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau_i}^{\tau} \widetilde{K}_h(y_i - x_i^T \beta^*(u)) dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \leq \frac{\tau_u - \tau_l}{1 - \tau_u} \left( \sqrt{\frac{p + \log n + t}{n}} + \zeta_p \frac{\log n + t}{n} \right)
\]

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 \( Q_0(\beta) = \mathbb{E} \hat{Q}_0(\beta) \).

**Lemma B.7.** For each \( j = 1, \ldots, m \),

\[
(B.7) \quad \left\| Q_k(\beta^*_j) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E} \{ \widetilde{K}_h(y - x^T \beta^*_\ell) x \} \right\|_{\Sigma^{-1}} \leq \frac{1}{2} l_1 \kappa_2 h^2 \left\{ 1 + H(\tau_j) - H(\tau_0) \right\} + (\bar{T}/f) \sum_{\ell=0}^{j-1} w_\ell (\tau_{\ell+1} - \tau_\ell),
\]

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

The next lemma establishes the asymptotic uniform equicontinuity of the centered process \( \mathcal{G}_n(\cdot) \) 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 \( \{ \alpha_n \}_{n \geq 1} \) be a normalized sequence such that \( \| \alpha_n \|_{\Sigma} = 1 \). Then, for any \( x > 0 \),

\[
\lim_{\delta \to 0} \lim_{n \to \infty} \limsup \mathbb{P} \left\{ \sup_{|\tau_1 - \tau_2| < \delta} | \mathcal{G}_n(\tau_1) - \mathcal{G}_n(\tau_2) | > x \right\} = 0,
\]

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

**B.2. Proof of Theorem 3.1.** We first prove a uniform upper bound over the grid of \( \tau \)-values—\( \{ \tau_0, \tau_1, \ldots, \tau_m \} \). That is, with probability at least \( 1 - C_1 n^{-1} \),

\[
(B.8) \quad \| \beta^{*}_j - \beta^*_j \|_{\Sigma} = \| \beta(\tau_j) - \beta^*(\tau_j) \|_{\Sigma} \leq r_j, \quad j = 0, 1, \ldots, m
\]

for a sequence of radii \( \{ r_j \}_{j=0,1,\ldots,m} \) and some absolute constant \( C_1 > 0 \). To begin with, define a crude (sub-optimal) convergence radius \( r^\Diamond = h/(4\eta_{1/4}) \) with \( \eta_{1/4} \) given in (B.1).

Accordingly, define “intermediate” points \( \beta_j = (1 - u_j) \beta^*_j + u_j \hat{\beta}_j \), where

\[
u_j = \sup \{ u \in [0, 1] : u (\beta_j - \beta^*_j) \in \Theta(r^\Diamond) \} \left\{ \begin{array}{ll} 1 & \text{if } \hat{\beta}_j \in \beta^*_j + \Theta(r^\Diamond), \\ (0, 1) & \text{if } \hat{\beta}_j \notin \beta^*_j + \Theta(r^\Diamond). \end{array} \right. \]
It is easy to see that $\tilde{\beta}_j = \hat{\beta}_j$ if $\tilde{\beta}_j \in \beta_j^* + \Theta(r^\diamond)$, and $\tilde{\beta}_j \in \beta_j^* + \partial \Theta(r^\diamond)$ if $\tilde{\beta}_j \notin \beta_j^* + \Theta(r_j)$. Here $\partial \Theta(r)$ denotes the boundary of $\Theta(r)$. In either case, we have $\tilde{\beta}_j \in \beta_j^* + \Theta(r^\diamond)$.

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

$$D(\hat{\beta}_j, \beta_j^*) \leq u_j \cdot D(\hat{\beta}_j, \beta_j^*) = u_j \cdot \langle \hat{Q}_j(\hat{\beta}_j) - \hat{Q}_j(\beta_j^*), \hat{\beta}_j - \beta_j^* \rangle.$$ 

Since $\tilde{\beta}_j$ solves the estimating equation $\hat{Q}_j(\tilde{\beta}_j) = 0$, by the Cauchy–Schwartz inequality we have

$$D(\tilde{\beta}_j, \beta_j^*) \leq u_j \cdot \langle \hat{Q}_j(\beta_j^*), \tilde{\beta}_j - \beta_j^* \rangle \leq \|\hat{Q}_j(\beta_j^*)\|_{\Sigma^{-1}} \cdot \|\tilde{\beta}_j - \beta_j^*\|_{\Sigma}.$$ 

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

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

Conditioning on $\mathcal{F}$, it follows that

$$(B.10) \quad \|\tilde{\beta}_j - \beta_j^*\|_{\Sigma} \leq \kappa^{-1} \|\hat{Q}_j(\beta_j^*)\|_{\Sigma^{-1}}, \quad j = 0, 1, \ldots, m.$$ 

Next we derive upper bounds for $\|\hat{Q}_j(\beta_j^*)\|_{\Sigma^{-1}}$ sequentially. For each $j$, we decompose $\hat{Q}_j(\beta_j^*)$ as

$$\hat{Q}_j(\beta_j^*) = \hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_\ell (\hat{\Delta}_\ell + \Delta_\ell) + Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\tilde{K}_h(y - x^T_\ell \beta_j^*)x\}$$ 

where $Q_0(\beta) = \mathbb{E}\hat{Q}_0(\beta)$, $w_\ell = H(\tau_{\ell+1}) - H(\tau_\ell)$, and

$$(B.11) \quad \begin{cases} \hat{\Delta}_\ell = \frac{1}{n} \sum_{i=1}^n \{ \tilde{K}_h(x_i - x^T_\ell \tilde{\beta}_\ell) - \tilde{K}_h(y_i - x^T_\ell \beta_j^*) \} \} x_i, \\ \Delta_\ell = \frac{1}{n} \sum_{i=1}^n (1 - \mathbb{E})\tilde{K}_h(y_i - x^T_\ell \beta_j^*) x_i. \end{cases}$$ 

By the triangle inequality,

$$(B.12) \quad \|\hat{Q}_j(\beta_j^*)\|_{\Sigma^{-1}} \leq \|\hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*)\|_{\Sigma^{-1}} + \sum_{\ell=0}^{j-1} w_\ell \left( \|\hat{\Delta}_\ell\|_{\Sigma^{-1}} + \|\Delta_\ell\|_{\Sigma^{-1}} \right)$$ 

$$+ \left\| Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\tilde{K}_h(y - x^T_\ell \beta_j^*)x\} \right\|_{\Sigma^{-1}}, \quad j = 1, \ldots, m.$$ 

In particular, $\|\hat{Q}_0(\beta_j^*)\|_{\Sigma^{-1}} \leq \|\hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*)\|_{\Sigma^{-1}} + \|Q_0(\beta_j^*)\|_{\Sigma^{-1}}$. For the approximation error term in (B.12), by Lemma B.7 we have

$$(B.13) \quad \left\| Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_\ell \mathbb{E}\{\tilde{K}_h(y - x^T_\ell \beta_j^*)x\} \right\|_{\Sigma^{-1}} \leq \frac{1}{2} l_1 \kappa_2 h^2 + \sum_{\ell=0}^{j-1} w_\ell a < a + \sum_{\ell=0}^{j-1} w_\ell a \quad \text{with } a := \frac{1}{2} l_1 \kappa_2 h^2 + J \ell^{-1} \delta.$$
For some $\delta > 0$ to be determined, define the second event

\[(B.14) \quad \mathcal{G} = \left\{ \max_{0 \leq j \leq m} \| \hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*) \|_{\Sigma^{-1}} \leq \max_{0 \leq \ell \leq m-1} \| \Delta_\ell \|_{\Sigma^{-1}} \leq \delta \right\}, \]

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

\[\| \hat{Q}_0(\beta_j^*) \|_{\Sigma^{-1}} < \delta + a, \quad \| \hat{Q}_j(\beta_j^*) \|_{\Sigma^{-1}} < \delta + a + \sum_{\ell=0}^{j-1} w_\ell (\delta + a + \| \hat{\Delta}_\ell \|_{\Sigma^{-1}}), \quad j = 1, \ldots, m. \]

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

\[\| \hat{\beta}_0 - \beta_0^* \|_{\Sigma} \leq \kappa^{-1} \| \hat{Q}_0(\beta_0^*) \|_{\Sigma^{-1}} < r_0 := \kappa^{-1}(\delta + a). \]

Provided $r_0 \leq r^{\diamond}$, the intermediate point $\hat{\beta}_0$ falls into the interior of the local region $\beta_0^* + \Theta(r^{\diamond})$. Via proof by contradiction, we must have $\hat{\beta}_0 = \beta_0$ and hence $\hat{\beta}_0 \in \beta_0^* + \Theta(r_0)$.

Turning to $(\hat{\beta}_1, \hat{\beta}_1)$, it follows from (B.10) with $j = 1$ that $\| \hat{\beta}_1 - \beta_1^* \|_{\Sigma} < \kappa^{-1}\{\delta + a + w_0(\| \hat{\Delta}_0 \|_{\Sigma^{-1}} + \delta + a)\}$. Note that $\hat{\Delta}_0$ depends on the preceding estimate $\hat{\beta}_0$. Since, as proved in the last step, $\hat{\beta}_0 \in \beta_0^* + \Theta(r_0)$ conditioned on $\mathcal{F} \cap \mathcal{G}$, it follows that

\[\| \hat{\Delta}_0 \|_{\Sigma^{-1}} \leq \varpi_0(r_0) + \| H(\tau_0)(\hat{\beta}_0 - \beta_0^*) \|_{\Sigma^{-1}} \leq \varpi_0(r_0) + \bar{f}_0, \]

where $\varpi_0(\cdot)$ is defined in (B.3). Conditioned further on $\{ \varpi_0(r_0) \leq \bar{f}_0 \}$, this implies

\[\| \hat{\beta}_1 - \beta_1^* \|_{\Sigma} < r_1 := \kappa^{-1}\{\delta + a + w_0(2\bar{f}_0 + \delta + a)\}. \]

As long as $r_1 \leq r^{\diamond}$, $\hat{\beta}_1$ lies in the interior of $\beta_1^* + \Theta(r^{\diamond})$, which enforces $\hat{\beta}_1 = \beta_1$ and hence $\hat{\beta}_1 \in \beta_1^* + \Theta(r_1)$.

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

\[\| \hat{\beta}_j - \beta_j^* \|_{\Sigma} < \frac{1}{\kappa} \left\{ \delta + a + \sum_{\ell=0}^{j-1} w_j(\| \hat{\Delta}_\ell \|_{\Sigma^{-1}} + \delta + a) \right\} \]

\[\leq \frac{1}{\kappa} \left\{ \delta + a + \sum_{\ell=0}^{j-1} w_\ell(2\bar{f}_\ell + \delta + a) \right\} :=: r_j. \]

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

\[r_j = (1 + 2\kappa^{-1}\bar{f}w_{j-1})r_{j-1} + \kappa^{-1}w_{j-1}(\delta + a), \quad j = 1, \ldots, m, \quad \text{and} \quad r_0 = \kappa^{-1}(\delta + a), \]

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

\[(B.16) \quad r_j \leq (1 + Cw_{j-1})r_{j-1} \leq \cdots \leq \prod_{\ell=0}^{j-1} (1 + Cw_\ell) \cdot r_0 \leq \exp \left( C \sum_{\ell=0}^{j-1} w_\ell \right) \cdot r_0 = \left( \frac{1 - \tau_L}{1 - \tau_j} \right)^C \cdot r_0. \]

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

\[\mathcal{F} \cap \mathcal{G} \cap \bigcap_{\ell=0}^{m-1} \{ \varpi_\ell(r_\ell) \leq \bar{f}_r \}. \]
with properly chosen $\kappa, \delta$ and $\{r_j\}_{j=0}^m$. By Lemmas B.4 and B.3, we choose $\kappa = (\sqrt{p} \log n)/2$ and $\delta \asymp \sqrt{(p + \log n)/n} + \zeta_p \log(n)/n$, so that $P(F^c) \leq (m+1)/n^2$ and $P(G^c) \leq 2(m+1)/n^2$ as long as $nh \gtrsim p + \log n$. With this choice of $\delta$, and since $\alpha = 0.5l_1 \kappa_2 h^2 + (J/f) \delta^* \lesssim h^2 + n^{-1/2}$, we obtain from (B.16) that

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

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

$$\omega_\ell(r_\ell) \lesssim \left(m_4^{1/2} \sqrt{\frac{p + \log n}{n} + m_3 r_\ell + h}\right) \cdot r_\ell$$

provided $nh \gtrsim \zeta^2_p (p + \log n)^{1/2}$. 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 $n \gtrsim \zeta^2_p (1/2 - \gamma)/(1 - \gamma)$, we conclude that the above “good” event occurs with probability at least $1 - C_1 n^{-1}$, and

$$\left(1 - \frac{\tau_j}{1 - \tau_j}\right)^C \cdot r_\ell \lesssim r_j \leq r^\diamond \asymp \left(\frac{p + \log n}{n}\right)^\gamma$$

for all $j = 0, 1, \ldots, m$.

This proves the claim (B.8).

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

$$\|\hat{\beta}(\tau) - \beta^*(\tau)\|_\Sigma = \|\hat{\beta}_j - \beta^*(\tau)\|_\Sigma \leq \|\hat{\beta}_j - \beta^*_j\|_\Sigma + \|\beta^*(\tau) - \beta^*_j\|_\Sigma \leq r_j + f^{-1} \delta^*.$$

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

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

completing the proof. \(\square\)

**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 $J_j = J(\tau_j)$ and $H_j = H(\tau_j)$ for $j = 0, \ldots, m$, and define the discretized integrated error up to $\tau_j$ as

(B.17) $\bar{e}_0 := J_0(\beta_0 - \beta_0^*)$ and $\bar{e}_j := J_j(\hat{\beta}_j - \beta_j^*) + \sum_{\ell=0}^{j-1} w_\ell H_\ell(\hat{\beta}_\ell - \beta_\ell^*)$, $j = 1, \ldots, m$.

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

(B.18) $\sup_{j=0, \ldots, m} \|\bar{e}_j + Q_j^\star\|_{\Sigma^{-1}} \lesssim \left(m_4^{1/2} \sqrt{\frac{p + \log n}{nh} + m_3 r_j + h}\right) \cdot r_j$.
holds with probability at least $1 - C_2 n^{-1}$ for some absolute constant $C_2 > 0$, where

(B.19) \[ Q^*_0 = \hat{Q}_0(\beta^*_0) = \frac{1}{n} \sum_{i=1}^{n} \{ \Delta_i \hat{K}_h(x_i^T \beta^*_0 - y_i) - \tau_0 \} x_i \quad \text{and} \quad \]

(B.20) \[ Q^*_j = \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta_i \hat{K}_h(x_i^T \beta^*_j - y_i) - \sum_{\ell=0}^{j-1} w_\ell \hat{K}_h(y_i - x_i^T \beta^*_\ell) - \tau_0 \right\} x_i \text{ for } j \in [m]. \]

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

\[ \mathcal{A} = \bigcap_{j=0}^{m} \left\{ \| \hat{\beta}_j - \beta^*_j \|_{\Sigma} \leq r_j \right\}, \text{ satisfying } P(\mathcal{A}) \geq 1 - C_1 n^{-1}, \]

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

(B.21) \[ \mathcal{E} = \bigcap_{j=0}^{m-1} \left\{ \omega_j(r_j) \leq \lambda_j \cdot r_j, \quad \omega_j(r_j) \leq \lambda_j \cdot r_j \right\} \bigcap \left\{ \omega_j(r_m) \leq \lambda_m \cdot r_m \right\}, \]

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

\[ \| \bar{e}_0 + Q^*_0 \|_{\Sigma^{-1}} = \| \hat{Q}_0(\hat{\beta}_0) - \hat{Q}_0(\beta^*_0) - J_0(\hat{\beta}_0 - \beta^*_0) \|_{\Sigma^{-1}} \leq \lambda_0 r_0, \]

and for $j \in [m],

\[ \| \bar{e}_j + Q^*_j \|_{\Sigma^{-1}} = \left\| J_j(\hat{\beta}_j - \beta^*_j) - \hat{Q}_j(\hat{\beta}_j) + \sum_{\ell=0}^{j-1} w_\ell H_\ell(\hat{\beta}_\ell - \beta^*_\ell) + \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta_i \hat{K}_h(x_i^T \beta^*_j - y_i) - \sum_{\ell=0}^{j-1} w_\ell \hat{K}_h(y_i - x_i^T \beta^*_\ell) - \tau_0 \right\} x_i \right\|_{\Sigma^{-1}} \]

\[ = \left\| J_j(\hat{\beta}_j - \beta^*_j) - \frac{1}{n} \sum_{i=1}^{n} \Delta_i \{ \hat{K}_h(x_i^T \beta^*_j - y_i) - \hat{K}_h(x_i^T \beta^*_j - y_i) \} x_i \right. \]

\[ + \left. \sum_{\ell=0}^{j-1} w_\ell H_\ell(\hat{\beta}_\ell - \beta^*_\ell) + \sum_{\ell=0}^{j-1} w_\ell \cdot \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{K}_h(y_i - x_i^T \beta^*_\ell) - \hat{K}_h(y_i - x_i^T \beta^*_\ell) \right\} x_i \right\|_{\Sigma^{-1}} \]

\[ = \Delta_i \text{ in (B.11)} \]

\[ \leq \left\| \frac{1}{n} \sum_{i=1}^{n} \Delta_i \{ \hat{K}_h(x_i^T \beta^*_j - y_i) - \hat{K}_h(x_i^T \beta^*_j - y_i) \} x_i - J_j(\hat{\beta}_j - \beta^*_j) \right\|_{\Sigma^{-1}} \]

\[ + \sum_{\ell=0}^{j-1} w_\ell \| \hat{\Delta}_\ell + H_\ell(\hat{\beta}_\ell - \beta^*_\ell) \|_{\Sigma^{-1}} \]

\[ \leq \omega_j(r_j) + \sum_{\ell=0}^{j-1} w_\ell \omega(\ell) \leq \lambda_j r_j + \sum_{\ell=0}^{j-1} w_\ell \lambda_\ell \ell. \]

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

\[ \| \bar{e}_j + Q^*_j \|_{\Sigma^{-1}} \leq \left\{ 1 + (2\bar{T})^{-1} \kappa \right\} \cdot \lambda_j r_j, \quad j \in [m]. \]
In view of Lemma B.5 with \( t = 2 \log n \), we set
\[
\lambda_j \asymp m_j^{1/2} \sqrt{\frac{p + \log n}{nh}} + m_3 r_j + h, \quad j = 0, 1, \ldots, m,
\]
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 \( \hat{\beta}(\cdot) \), 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 \( \hat{\beta}(\tau) = \hat{\beta}_j \). With this notation, the left-hand side of (16) equals
\[
\max_{j=0,\ldots,m} \sup_{\tau \in \mathcal{I}_j} \left\| \hat{e}(\tau) - \frac{1}{n} \sum_{i=1}^n \left\{ \tau_L + \int_{\tau_L}^\tau \tilde{K}_h(y_i - x_i^T \beta^*(u))dH(u) - \Delta_i \tilde{K}_h(x_i^T \beta^*(\tau) - y_i) \right\} x_i \right\|_{\Sigma^{-1}},
\]
where \( \hat{e}(\tau) = \hat{\beta}_{\text{int}}(\tau) - \beta^*_j(\tau) \). To control the discretization error, conditioning on the event \( \mathcal{A} \cap \mathcal{E} \) we have, for each \( j = 0, 1, \ldots, m \),
\[
\sup_{\tau \in \mathcal{I}_j} \| \mathbf{J}(\tau) \{ \hat{\beta}(\tau) - \beta^*(\tau) \} - \mathbf{J}_j(\hat{\beta}_j - \beta^*_j) \|_{\Sigma^{-1}}
\leq \sup_{\tau \in \mathcal{I}_j} \| \mathbf{J}(\tau) - \mathbf{J}_j \| \{ \hat{\beta}_j - \beta^*_j \} \|_{\Sigma^{-1}} + \sup_{\tau \in \mathcal{I}_j} \| \mathbf{J}(\tau) \{ \beta^*_j - \beta^*(\tau) \} \|_{\Sigma^{-1}}
\]
(B.22)
\[
\leq l_1 \frac{1}{f-1} m_3 r_j \delta^* + \overline{g} \frac{1}{f-1} \delta^* \lesssim \delta^*,
\]
and
\[
\sup_{\tau \in \mathcal{I}_j} \left\| \int_{\tau_L}^\tau \mathbf{H}(u) \{ \hat{\beta}(u) - \beta^*(u) \} dH(u) - \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \mathbf{H}_L(\hat{\beta}_\ell - \beta^*_\ell) dH(u) \right\|_{\Sigma^{-1}}
\leq \left\| \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \left[ \mathbf{H}(u) \{ \hat{\beta}(u) - \beta^*(u) \} - \mathbf{H}_L(\hat{\beta}_\ell - \beta^*_\ell) \right] dH(u) \right\|_{\Sigma^{-1}}
\lesssim \log \left( \frac{1 - \tau_0}{1 - \tau_{j+1}} \right) \cdot \delta^*.
\]
Next we control the approximation error for discretizing the linear process \( 1/n \sum_{i=1}^n U_i(\cdot) \). For any interval \( \mathcal{I}_j \), write \( v(\tau) = \beta^*(\tau) - \beta^*_j(\tau) \) for \( \tau \in \mathcal{I}_j \), satisfying \( \| v(\tau) \|_{\Sigma} \leq \frac{1}{f-1}(\tau_{j+1} - \tau_j) \) by the Lipschitz continuity of \( \beta^*(\cdot) \). Moreover, we have
\[
\left\| \mathbb{E} \Delta_i \left\{ \tilde{K}_h(x_i^T \beta^*(\tau) - y_i) - \tilde{K}_h(x_i^T \beta^*_j - y_i) \right\} x_i \right\|_{\Sigma^{-1}}
\leq \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E} \left[ \Delta_i \left\{ \tilde{K}_h(x_i^T \beta^*(\tau) - y_i) - \tilde{K}_h(x_i^T \beta^*_j - y_i) \right\} \langle u, z_i \rangle \right]
\leq \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E} \left[ \int_{-\infty}^{\infty} \left\{ K \left( x_i^T \beta^*(\tau) - u \right) - \tilde{K} \left( x_i^T \beta^*_j - u \right) \right\} g(u|x)du \cdot \langle u, z_i \rangle \right]
\leq \mathcal{C} \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E} \left[ \int_{-\infty}^{\infty} \left\{ \tilde{K}(v + x_i^T v(\tau)/h) - \tilde{K}(v) \right\} g(x_i^T \beta^*_j - hv|x)dv \cdot \langle u, z_i \rangle \right]
\leq \overline{g} \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E} \left[ \int_{-\infty}^{\infty} \left\{ \int_{0}^{1} K(v + w x_i^T v(\tau)/h)dw \right\} dv \cdot \langle v(\tau), x_i \rangle \langle u, z_i \rangle \right]
Together, (B.18) and (B.22)–(B.25) prove the claimed bounds (14)–(16) for all $n h$ as long as $1 \leq \tau_j^{*}$ for any $j = 0, \ldots, m$. Turning to the deterministic approximation error, it can be shown that

$$
\sum_{j=0}^{m} \left( \sqrt{\frac{p + \log n}{n h}} \right) \leq \delta^* \cdot 2 \log \left( \frac{1 - \tau_j}{1 - \tau_{j+1}} \right) \left( \frac{\log n}{n} + \zeta_p \frac{\log n}{\sqrt{n h}} \right) \cdot \delta^*
$$

for all $j = 0, \ldots, m$. Turning to the deterministic approximation error, it can be shown that

$$
\sup_{\tau \in I_j} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - E) \left\{ \int_{\tau}^{\tau+1} \bar{K}_h(y_i - x_i^T \beta^*(u)) dH(u) \right\} \right\|_2 \leq \delta^*
$$

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

$$
\sup_{\tau \in I_j} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - E) \left\{ \int_{\tau}^{\tau+1} \bar{K}_h(y_i - x_i^T \beta^*(u)) dH(u) \right\} \right\|_2 \leq \delta^*
$$

for all $j = 0, \ldots, m$. Turning to the deterministic approximation error, it can be shown that

$$
\sup_{\tau \in I_j} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - E) \left\{ \int_{\tau}^{\tau+1} \bar{K}_h(y_i - x_i^T \beta^*(u)) dH(u) \right\} \right\|_2 \leq \delta^*
$$

for all $j = 0, \ldots, m$. Turning to the deterministic approximation error, it can be shown that

$$
\sup_{\tau \in I_j} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - E) \left\{ \int_{\tau}^{\tau+1} \bar{K}_h(y_i - x_i^T \beta^*(u)) dH(u) \right\} \right\|_2 \leq \delta^*
$$

for all $j = 0, \ldots, m$. Turning to the deterministic approximation error, it can be shown that
It remains to control the bias term $\mathbb{E} U_i(\cdot)$. Define the non-smoothed version of $U_i(\cdot)$ as
\[
V_i(\tau) = \left[ \tau_L + \int_{\tau_L}^\tau 1 \{ y_i > x_i^T \beta^*(u) \} dH(u) - \Delta_i 1 \{ y_i < x_i^T \beta^*(\tau) \} \right] x_i, \quad \tau \in [\tau_L, \tau_U].
\]
By the martingale property, $\mathbb{E} V_i(\tau) = 0$ for every $\tau \in [\tau_L, \tau_U]$. Note that
\[
U_i(\tau) - V_i(\tau) = \left( \Delta_i \left[ 1 \{ y_i < x_i^T \beta^*(\tau) \} - K_h(x_i^T \beta^*(\tau) - y_i) \right] + \int_{\tau_L}^\tau \left[ K_h(y_i - x_i^T \beta^*(u)) - 1 \{ y_i > x_i^T \beta^*(u) \} \right] dH(u) \right) x_i.
\]
Following the same calculations that lead to (E.16), we obtain
\[
\sup_{\tau \in [\tau_L, \tau_U]} \| \mathbb{E} U_i(\tau) \|_{\Sigma^{-1}} = \sup_{\tau \in [\tau_L, \tau_U]} \| \mathbb{E} \{ U_i(\tau) - V_i(\tau) \} \|_{\Sigma^{-1}} \leq 0.5 L_1 \kappa_2 \left\{ 1 + \log \left( \frac{1}{L_1 \kappa_2} \right) \right\} h^2.
\]
This completes the proof of the theorem.

**B.4. Proof of Theorem 3.3.** Assume without loss of generality that $\|\alpha_n\|_\Sigma = 1$; otherwise, we simply replace $\alpha_n$ by $\alpha_n/\|\alpha_n\|_\Sigma$. 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 $G_{n}(\cdot)$.

For the former, via the Cramér–Wold device, it is equivalent to show that for any finite set of values $\{ \tau_i \}_{i=1}^L \subseteq [\tau_L, \tau_U]$ and $\gamma_1, \ldots, \gamma_L \in \mathbb{R}$,
\begin{equation}
\sum_{\ell=1}^L \gamma_\ell G_n(\tau_\ell) \xrightarrow{d} \sum_{\ell=1}^L \gamma_\ell G(\tau_\ell),
\end{equation}
with $G(\cdot)$ defined in (22). For $i = 1, \ldots, n$, define centered variables $W_i = \sum_{\ell=1}^L \gamma_\ell \langle \alpha_n, U_{0i}(\tau_\ell) \rangle$ with $U_{0i}(\tau) := U_i(\tau) - \mathbb{E} U_i(\tau)$, so that $\sum_{\ell=1}^L \gamma_\ell G_n(\tau_\ell) = n^{-1/2} \sum_{i=1}^n W_i$. Moreover,
\[
\text{Var}(W_i) = \text{Var} \left( \sum_{\ell=1}^L \gamma_\ell \langle \alpha_n, U_{0i}(\tau_\ell) \rangle \right)
\]
\[
= \sum_{\ell_1=1}^L \sum_{\ell_2=1}^L \gamma_{\ell_1} \gamma_{\ell_2} \cdot \text{Cov} \left( \langle \alpha_n, U_{0i}(\tau_{\ell_1}) \rangle, \langle \alpha_n, U_{0i}(\tau_{\ell_2}) \rangle \right)
\]
\[
= \sum_{\ell_1=1}^L \sum_{\ell_2=1}^L \gamma_{\ell_1} \gamma_{\ell_2} \cdot \alpha_n^T \mathbb{E} \left\{ U_{0i}(\tau_{\ell_1}) U_{0i}(\tau_{\ell_2})^T \right\} \alpha_n
\]
\[
= \sum_{\ell_1=1}^L \sum_{\ell_2=1}^L \gamma_{\ell_1} \gamma_{\ell_2} \cdot \alpha_n^T \mathbb{E} \left\{ U_i(\tau_{\ell_1}) U_i(\tau_{\ell_2})^T \right\} \alpha_n
\]
\[
- \sum_{\ell_1=1}^L \sum_{\ell_2=1}^L \gamma_{\ell_1} \gamma_{\ell_2} \cdot \alpha_n^T \mathbb{E} \left\{ U_i(\tau_{\ell_1}) \right\} \mathbb{E} \left\{ U_i(\tau_{\ell_2})^T \right\} \alpha_n
\]
\[
\to \sum_{\ell_1=1}^L \sum_{\ell_2=1}^L \gamma_{\ell_1} \gamma_{\ell_2} \cdot H(\ell_1, \ell_2) = \text{Var} \left( \sum_{\ell=1}^L \gamma_\ell G(\tau_\ell) \right) \text{ as } n \to \infty,
\]
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 $x > 0$,

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

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].

\section*{Appendix C: Proofs of the Main Results in Section 3.3}

\subsection*{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 $\mathbb{D}_n = \{y_i, \Delta_i, x_i\}_{i=1}^n$. Similarly to (B.2), we define the symmetrized Bregman divergence $D^\beta : \mathbb{R}^p \times \mathbb{R}^p \to [0, \infty)$ in the bootstrap world as

\begin{align*}
(D^\beta(\beta_1, \beta_2) = \langle \hat{Q}_j^\beta(\beta_1) - \hat{Q}_j^\beta(\beta_2), \beta_1 - \beta_2 \rangle, \tag{C.1}
\end{align*}

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

\begin{lemma}[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 $t \geq 0$. Suppose the “effective” sample size satisfies $nh \gtrsim \max\{p, \zeta \}$ and $\zeta \leq M$ for some $M > 0$, and $\{e_i\}_{i=1}^n$ are Rademacher random variables independent of the data $\{x_i, \xi_i\}_{i=1}^n$. Then, there exists an event $\mathcal{E}_1(t)$ with $\mathbb{P}(\mathcal{E}_1(t)) \geq 1 - (m + 1)e^{-t}$, such that, conditioning on $\mathcal{E}_1(t)$,

\begin{align*}
\inf_{j \in \{0, \ldots, m\}} \inf_{\beta \in \beta_j^* + \Theta(r)} \frac{D(\beta, \beta_j^*)}{\kappa_i \|\beta - \beta_j^*\|_2^2} \geq \frac{1}{2}g, \quad \text{and} \quad \mathbb{P}^* \left\{ \inf_{j \in \{0, \ldots, m\}} \inf_{\beta \in \beta_j^* + \Theta(r)} \frac{D^\beta(\beta, \beta_j^*)}{\kappa_i \|\beta - \beta_j^*\|_2^2} \geq \frac{1}{2}g \right\} \geq 1 - (m + 1)e^{-t}.
\end{align*}

\end{lemma}

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

\begin{align*}
\left\| \frac{1}{n} \sum_{i=1}^n (e_i \cdot \xi_i \mathbf{z}_i) \right\|_2 \lesssim \sqrt{\frac{p + \log n}{n}}
\end{align*}

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

\begin{align*}
\Gamma_j(r) := \sup_{\beta \in \beta_j^* + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^n e_i \left\{ \hat{K}_h(y_i - x_i^\top \beta) - \hat{K}_h(y_i - x_i^\top \beta_j^*) \right\} \mathbf{z}_i \right\|_2, \tag{C.2}
\end{align*}
(C.3) \[ \Gamma_j^\Delta(r) := \sup_{\beta \in \mathcal{B}_j + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^{n} e_i \Delta_i \left\{ \hat{K}_h(x_i^3 \beta - y_i) - \hat{K}_h(x_i^3 \beta_j - y_i) \right\} x_i \right\|_2, \]

where \( \{e_i\}_{i=1}^{n} \) is a sequence of independent Rademacher random variables that are independent of \( \{y_i, \Delta_i, x_i\}_{i=1}^{n} \), and \( x_i = \Sigma^{-1/2} x_i \).

**Lemma C.3.** Assume that Conditions 3.1–3.3 hold, and \( K(\cdot) \) in Condition 3.1 is \( l_K \)-Lipschitz continuous. Then, there exists an event \( E_3 \) with \( \mathbb{P}(E_3) \geq 1 - 3n^{-1} \) such that, with \( \mathbb{P}^\ast \)-probability at least 1 – \((m+1)n^{-2} \) conditioned on \( E_3 \),

\[ \sup_{j \in \{0, \ldots, m\}} \Gamma_j(r) \leq r \sqrt{\frac{p + \log n}{nh}} \left( \frac{m^{1/2} + \zeta^2}{p \log n} \right), \]

provided \( n \geq \zeta_p^2 \log n \). 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 \( E_4 \) with \( \mathbb{P}(E_4) \geq 1 - (m + 3n + 1)n^{-2} \) such that, with \( \mathbb{P}^\ast \)-probability at least 1 – \( n^{-2} \) conditional on \( E_4 \),

\[ \sup_{\beta \in \mathcal{B}_j + \Theta(r)} \left\| J_j(\beta - \beta_j) - \{ \hat{Q}^\beta_j(\beta) - \hat{Q}^\beta_j(\beta_j) \} \right\|_{\Sigma^{-1}} \]

\[ \lesssim \left\{ \frac{m^{1/2}}{p + \log n} + m_3 r + h + \zeta_p^2 \left( \frac{p \log n}{n} \right)^{1/2} \left( \frac{\log n}{n} \right)^{1/2} \right\} \cdot r \]

holds uniformly over \( j = 0, 1, \ldots, m \), where \( J_j = \mathbb{E}\{g(x^\top \beta_j^\ast|x)x^\top x\} \).

**Lemma C.5.** Assume Conditions 3.1–3.4 hold, and let \( r > 0 \). Then, there exists an event \( E_5 \) with \( \mathbb{P}(E_5) \geq 1 - mn^{-2} \) such that conditioning on \( E_5 \),

\[ \sup_{\mathcal{G}_n^b(\cdot)} \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{l=0}^{L-1} \int_{T_l} dH(n) \left[ H_{\ell}^\beta(\beta_l - \beta_l^\ast) - W_0 \{ \hat{K}_h(y_i - x_i^\top x_l \beta_l) - \hat{K}_h(y_i - x_i^\top x_l \beta_l^\ast) \} x_i \right] \right\|_{\Sigma^{-1}} \]

\[ \lesssim \log \left( \frac{1 - r_0}{1 - r_j} \right) \cdot \left\{ m_4^{1/2} \left( \frac{p + \log n}{n} + m_3 r + h \right) \cdot r + \max_{\ell=0, \ldots, m-1} \Gamma_{\ell}(r) \right\} \]

holds for all \( j = 1, \ldots, m \), where \( H_{\ell} = \mathbb{E}\{f(g(x^\top \beta_l^\ast|x)|x)x^\top x\} \) and \( \Gamma_{\ell}(r) \) is defined in (C.2).

The following lemma establishes the asymptotic uniform equicontinuity of the process \( \mathcal{G}_n^b(\cdot) = n^{-1/2} \sum_{i=1}^{n} e_i(a_n, U_i(\cdot)) \) in \( \ell_\infty([\tau_L, \tau_U]) \), thus validating the asymptotic tightness of \( \mathcal{G}_n^b(\cdot) \), where \( U_i(\cdot) \) is defined in (15).

**Lemma C.6.** Assume the conditions of Theorem 3.6 hold. For any \( x > 0 \) and sequence of vectors \( a_n \) satisfying \( \|a_n\|_{\Sigma} = 1 \), conditioned on any observed data \( \mathcal{D}_n = \{y_i, \Delta_i, x_i\}_{i=1}^{n} \), we have

\[ \lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{|\tau_1 - \tau_2| < \delta} \left\| \mathcal{G}_n^b(\tau_1) - \mathcal{G}_n^b(\tau_2) \right\| > x = 0. \]

where \( \mathcal{G}_n^b(\cdot) = n^{-1/2} \sum_{i=1}^{n} e_i(a_n, U_i(\cdot)) \) with \( U_i(\tau) \) 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 \( \hat{Q}_j^*(\beta) \) given in (7) and (8), and \( \tilde{Q}_j^*(\beta^*) = 0 \). Following the localized argument as in the proof of Theorem 3.1, for the same radius parameter \( r^\odot \) therein, define \( \beta_j^* = \beta_j^* + \gamma_j (\beta_j - \beta_j^*) \) with 
\[
\gamma_j := \sup \{ \gamma \in [0, 1] : \gamma (\beta_j^* - \beta_j^*) \in \Theta(r^\odot) \}
\]
so that if \( \beta_j^* \in \beta_j^* + \Theta(r^\odot) \) and \( \beta_j^* \in \beta_j^* + \Theta(r^\odot) \) if \( \beta_j^* \notin \beta_j^* + \Theta(r^\odot) \). Consequently,
\[
D^\circ(\beta_j^*, \beta_j^*) \leq \rho_j \cdot \langle -\tilde{Q}_j^*(\beta_j^*), \tilde{Q}_j - \beta_j^* \rangle \leq \|\tilde{Q}_j^*(\beta_j^*)\|_{\Sigma^{-1}} \cdot \|\beta_j^* - \beta_j^*\|_{\Sigma},
\]
where \( D^\circ \) is defined in (C.1).

In addition to \( F \) in (B.9), define
\[
F^\circ = \bigcap_{j=0}^m \left \{ D^\circ(\beta, \beta_j^*) \geq \kappa \cdot \|\beta - \beta_j^*\|_{\Sigma}^2 \text{ for all } \beta \in \beta_j^* + \Theta(r^\odot) \right \}.
\]
Conditioned on \( F^\circ \), we have for all \( j = 0, 1, \ldots, m \) that \( ||\hat{\beta}_j^* - \beta_j^*||_{\Sigma} \leq \kappa^{-1} \|\tilde{Q}_j^*(\beta_j^*)\|_{\Sigma^{-1}} \).

For the bootstrapped estimating function, by the triangle inequality we have
\[
\|\tilde{Q}_j^*(\beta_j^*)\|_{\Sigma^{-1}} \leq \left\| \frac{1}{n} \sum_{i=1}^n \left\{ \Delta_i \hat{K}_h(x_i^T \beta_j^* - y_i) - \tau_0 - \sum_{\ell=0}^{j-1} w_{\ell} \hat{K}_h(y_i - x_i^T \beta_j^*) \right\} x_i \right\|_{\Sigma^{-1}}
\]
\[+ \left\| \frac{1}{n} \sum_{i=1}^n \sum_{\ell=0}^{j-1} e_i w_{\ell} \left\{ \hat{K}_h(y_i - x_i^T \beta_j^*) - \hat{K}_h(y_i - x_i^T \beta_j^*) \right\} x_i \right\|_{\Sigma^{-1}}
\]
\[+ \left\| \tilde{Q}_0(\beta_j^*) - Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_{\ell} (\hat{\Delta}_j^\circ + \Delta_\ell) + Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_{\ell} E \{ \hat{K}_h(y - x^T \beta_j^*) x_i \} \right\|_{\Sigma^{-1}}
\]
\[= : Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_{\ell} E \{ \hat{K}_h(y - x^T \beta_j^*) x_i \} \right\|_{\Sigma^{-1}},
\]
for \( j \geq 1 \), where \( w_{\ell} = H(\tau_{\ell+1}) - H(\tau_\ell) \),
\[
\hat{\Delta}_\ell = \frac{1}{n} \sum_{i=1}^n e_i \left\{ \hat{K}_h(y_i - x_i^T \beta_j^*) - \hat{K}_h(y_i - x_i^T \beta_j^*) \right\} x_i,
\]
\[
\hat{\Delta}_\ell = \frac{1}{n} \sum_{i=1}^n \left\{ \hat{K}_h(y_i - x_i^T \beta_j^*) - \hat{K}_h(y_i - x_i^T \beta_j^*) \right\} x_i.
\]
and $\Delta_\ell$ is defined in (B.11). In particular,

$$
\|\tilde{Q}_0^\ell(\beta_0^\ast)\|_{\Sigma^{-1}} = \left\| \frac{1}{n} \sum_{i=1}^{n} e_i \left\{ \Delta_1 \tilde{K}_h(x_i^T \beta_0^\ast - y_i) - \tau_0 \right\} x_i \right\|_{\Sigma^{-1}} + \|\tilde{Q}_0(\beta_0^\ast) - Q_0(\beta_0^\ast)\|_{\Sigma^{-1}} + \|Q_0(\beta_0^\ast)\|_{\Sigma^{-1}} .
$$

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)

$$
\mathcal{G}_\delta = \left\{ \max_{0 \leq j \leq m} \|\tilde{Q}_j^\ell(\beta_j^\ast)\|_{\Sigma^{-1}} \sqrt{\max_{0 \leq j \leq m} \|\tilde{Q}_0(\beta_j^\ast) - Q_0(\beta_j^\ast)\|_{\Sigma^{-1}}} \max_{0 \leq p \leq m-1} \|\Delta_p\|_{\Sigma^{-1}} \leq \delta \right\} .
$$

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

$$
\|\tilde{Q}_0(\beta_0^\ast)\|_{\Sigma^{-1}} < 2\delta + a ,
$$

$$
\|\tilde{Q}_j^\ell(\beta_j^\ast)\|_{\Sigma^{-1}} < 2\delta + a + \sum_{\ell=0}^{j-1} w_\ell (\delta + a + \|\tilde{Q}_\ell\|_{\Sigma^{-1}} + \|\Delta_\ell\|_{\Sigma^{-1}}) , \quad j = 1, \ldots, m ,
$$

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

(C.8)

$$
\mathcal{F}^p \cap \mathcal{G}_\delta \cap \bigcap_{\ell=0}^{m-1} \{ \mathcal{P}(r_\ell) \vee \mathcal{P}(\Gamma_\ell) \leq \tilde{F} r_\ell \} ,
$$

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

(C.9)

$$
r_j = (1 + 3\kappa^{-1} \tilde{F} w_{j-1}) r_{j-1} + \kappa^{-1} w_{j-1} (\delta + a) , \quad j = 1, \ldots, m , \quad \text{and} \quad r_0 = \kappa^{-1} (2\delta + a).
$$

Denoting $C = \kappa^{-1} (3\tilde{F} + 1)$, it follows that $r_j \leq \left( \frac{1 - \tau_j}{1 - r_j} \right)^{C} r_0$.

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

$$
\mathbb{P}(\mathcal{F}^p \cap \mathcal{G}_\delta) \geq 1 - C_2 n^{-1} ,
$$

$$
\sup_{\ell \in \{0, \ldots, m-1\}} \mathcal{P}(r_\ell) \lesssim \left( \frac{1}{n} \sqrt{\frac{p + \log n}{nh}} + m_3 r_\ell + h \right) \cdot r_\ell
$$

and

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

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

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

holds uniformly over $j \in \{0, \ldots, m\}$. 
Recall that \( m_3 \) and \( m_4 \) are dimension-free moment parameters. Given the bandwidth \( h = h_n \sim \{(p + \log n)/n\}^\gamma \) with \( \gamma \in [1/4, 1/2] \), and under the sample size requirement \( n \gtrsim \zeta_0^2/(1-\gamma)(p + \log n)/(1/2-\gamma)/(p \log n)^{1/(2-2\gamma)} \), 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

\[
\left( 1 - \frac{\tau}{1 - \tau} \right)^C g^{-1} \sqrt{\frac{p + \log n}{n}} \geq \tau_j \leq \left( \frac{p + \log n}{n} \right)^\gamma \text{ for all } j = 0, 1, \ldots, m.
\]

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. \( \square \)

### 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.

**STEP 1.** (Uniform bound over \( \{\tau_0, \ldots, \tau_m\} \)) For simplicity, we write \( J_j = J(\tau_j) \) and \( H_j = H(\tau_j) \) for \( j = 0, \ldots, m \), and define the accumulated bootstrap errors as

\[
e_0^\circ := J_0(\beta_0^\circ - \beta_0) \quad \text{and} \quad e_j^\circ := J_j(\beta_j^\circ - \beta_j) + \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} H_\ell(\beta_\ell^\circ - \beta_\ell) \, dH(u), \quad j = 1, \ldots, m.
\]

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

(C.10)

\[
\sup_{j=0,\ldots,m} \|e_j^\circ + Q_j^\circ\|_{\Sigma^{-1}} \lesssim m_4^{1/2} \frac{p + \log n}{nh^{1/2}} + h \sqrt{\frac{p + \log n}{n}} + C_p \frac{p + \log n}{n^{3/2}h}
\]

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

\[
Q_0^\circ = \frac{1}{n} \sum_{i=1}^{n} e_i \{ \Delta_i \hat{K}_h(x_i^T \beta_0^\circ - y_i) - \tau_0 \} x_i,
\]

\[
Q_j^\circ = \frac{1}{n} \sum_{i=1}^{n} e_i \left\{ \Delta_i \hat{K}_h(x_i^T \beta_j^\circ - y_i) - \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \hat{K}_h(y_i - x_i^T \beta_\ell^\circ) \, dH(u) - \tau_0 \right\} x_i
\]

for \( j = 1, \ldots, m \).

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

\[
\max_{0 \leq j \leq m} \|\hat{\beta}_j - \beta_j\|_{\Sigma} \lesssim \sqrt{\frac{p + \log n}{n}}, \quad \text{and}
\]

(C.11)

\[
\mathbb{P}^* \left( \max_{0 \leq j \leq m} \|\hat{\beta}_j - \beta_j\|_{\Sigma} \lesssim \sqrt{\frac{p + \log n}{n}} \right) \geq 1 - C_5 n^{-1}.
\]

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

\[
\|e_0^\circ + Q_0^\circ\|_{\Sigma^{-1}} \leq \|e_0 + Q_0\|_{\Sigma^{-1}} + \left\| J_0(\hat{\beta}_0^\circ - \beta_0) + \frac{1}{n} \sum_{i=1}^{n} W_i(\Delta_i \hat{K}_h(x_i^T \beta_0^\circ - y_i) - \tau_0) x_i \right\|_{\Sigma^{-1}} =: I_0 + I_0.
\]
and for \( j \geq 1, \)
\[
\| \hat{e}_{\text{int}}(\tau_j) + Q_j^* \|_{\Sigma^{-1}} \\
\leq \| \hat{e}_{\text{int}}(\tau_j) + Q_j^* \|_{\Sigma^{-1}} + \left\| \mathbf{J}_j(\hat{\beta}_j^* - \beta_j^*) + \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \mathbf{H}_\ell(\hat{\beta}_\ell^* - \beta_\ell^*) dH(u) \right\|_{\Sigma^{-1}} \\
+ \frac{1}{n} \sum_{i=1}^{n} W_i \left\{ \Delta_i \tilde{K}_h(x_i^T \hat{\beta}_j^* - y_i) - \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \tilde{K}_h(y_i - x_i^T \hat{\beta}_\ell^*) dH(u) - \tau_0 \right\} x_i \right\|_{\Sigma^{-1}} \\
=: I_j + II_j,
\]
where \( \hat{e}_{\text{int}}(\tau_j) \) 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 \( \mathcal{E}_2, \)
\[
(C.12) \quad \max_{0 \leq j \leq m} I_j = \| \hat{e}_{\text{int}}(\tau_j) + Q_j^* \|_{\Sigma^{-1}} \lesssim m_4^{1/2} \frac{p + \log n}{nh^{1/2}} + h \sqrt{\frac{p + \log n}{n}}.
\]
It remains to bound \( II_j \) for \( j = 0, 1, \ldots, m. \) Recall that \( \hat{Q}_j^*(\hat{\beta}_j^*) = 0, \) we have
\[
II_0 = \| J_0(h_0^0 - \beta_0^0) - \{ \hat{Q}_0^*(\hat{\beta}_0^0) - \hat{Q}_0(\beta_0^0) \} \|_{\Sigma^{-1}}
\]
and for \( j = 1, \ldots, m, \)
\[
II_j \leq \| J_j(\hat{\beta}_j^* - \beta_j^*) - \{ \hat{Q}_j^*(\hat{\beta}_j^*) - \hat{Q}_j^*(\beta_j^*) \} \|_{\Sigma^{-1}} \\
+ \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \mathbf{H}_\ell(\hat{\beta}_\ell^* - \beta_\ell^*) - W_i \left\{ \tilde{K}_h(y_i - x_i^T \hat{\beta}_\ell^*) + \tilde{K}_h(y_i - x_i^T \hat{\beta}_\ell) x_i \right\} \right\|_{\Sigma^{-1}}.
\]
Putting together the pieces, and taking \( r \asymp \sqrt{(p + \log n)/n} \), we conclude that conditioning on \( \mathcal{E}_1 \cap \mathcal{E}_2, \)
\[
(C.13) \quad \| J_j(\beta - \beta_j^*) - \{ \hat{Q}_j^*(\beta) - \hat{Q}_j(\beta) \} \|_{\Sigma^{-1}} \\
+ \sup_{\hat{\beta}_j^* \in \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \mathbf{H}_\ell(\beta - \beta_\ell^*) - \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=0}^{j-1} \int_{\tau_{\ell}}^{\tau_{\ell+1}} \mathbf{H}_\ell(\beta - \beta_\ell^*) - W_i \left\{ \tilde{K}_h(y_i - x_i^T \beta_\ell^*) + \tilde{K}_h(y_i - x_i^T \hat{\beta}_\ell) x_i \right\} \right\|_{\Sigma^{-1}}
\]
holds with \( \mathbb{P}^* \)-probability at least \( 1 - C_3 n^{-1} \).

Let \( \mathcal{E}_3 \sim \mathcal{E}_5 \) 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 \( j = 0, 1, \ldots, m, \)
\[
(C.14) \quad II_j \lesssim \left\{ m_4^{1/2} \frac{p + \log n}{nh^{1/2}} + h \sqrt{\frac{p + \log n}{n}} \\
+ \zeta_2 \frac{(p + \log n) \log n}{n^{3/2} h} + \max_{0 \leq \ell \leq j-1} \Gamma_\ell(r) \right\} \cdot \left\{ \log \left( \frac{1}{r} \right) \right\} \vee 1
\]
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 \( \log \left( \frac{1}{r} \right) \leq \log \left( \frac{1 - \tau_0}{1 - \tau_2} \right) \) 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 \( \mathcal{E}_3, \)
\[
(C.15) \quad \max_{0 \leq \ell \leq m} \Gamma_\ell(r) \lesssim m_4^{1/2} \frac{p + \log n}{nh^{1/2}} + \zeta_2 \frac{(p + \log n) \log n}{n^{3/2} h}
\]
holds with \( \mathbb{P}^* \)-probability at least \( 1 - C\delta n^{-1} \) 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 \( \|a_n\|_\Sigma = 1 \) throughout the proof; otherwise, we first rescale the vectors \( a_n \) so that the same arguments apply.

Conditioned on the observed data \( \mathbb{D}_n = \{(y_i, \Delta_i, x_i)\}_{i=1}^n \), we have \( \mathbb{E}^* \langle a_n, U^p_\tau(\tau) \rangle = 0 \) 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 \( G^p_n(\cdot) \) 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 \( G^p_n(\cdot) \) given \( \mathbb{D}_n \) converges to \( (\cdot, \cdot) \) defined in (21), which is the limit of the (unconditional) covariance function of \( G_n(\cdot) \). To this end, for any \( s, t \in [\tau_L, \tau_U] \), note that

\[
\text{Cov}^* \left( G^p_n(s), G^p_n(t) \right) = \mathbb{E}^* \left\{ \mathbb{G}^p_n(s) \mathbb{G}^p_n(t) \right\} \\
= \frac{1}{n} \mathbb{E}^* \left\{ \sum_{i=1}^n \langle a_n, e_i U_i(s) \rangle \right\} \cdot \left\{ \sum_{i=1}^n \langle a_n, e_i U_i(t) \rangle \right\} \\
= \frac{1}{n} \sum_{i=1}^n a_n^T U_i(s) U_i(t)^T a_n \xrightarrow{a.a.} H(s, t),
\]

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

\[
\psi_j(r, q) = \sup_{\beta \in \beta_j^* + \Theta(r) \cap \Lambda(q)} \left\| \frac{1}{n} \sum_{i=1}^n (1 - \mathbb{E}) \left\{ \tilde{K}_h(y_i - x_i^T \beta) - \bar{K}_h(y_i - x_i^T \beta_j^*) \right\} x_i \right\|_\infty,
\]

where \( \Lambda(q) := \{ u \in \mathbb{R}^p : \|u\|_1 \leq q \|u\|_\Sigma \} \) 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 \( 1 - e^{-t} \),

\[
\|\tilde{Q}_0(\beta_j^*) - Q_0(\beta_j^*)\|_\infty \leq \tau_0 \left\{ \sigma \sqrt{\frac{2t + 2 \log(2p)}{n}} + \frac{t + \log(2p)}{3n} \right\},
\]

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

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

\[
\left\| \frac{1}{n} \sum_{i=1}^n (1 - \mathbb{E}) \tilde{K}_h(y_i - x_i^T \beta_j^*) x_i \right\|_\infty \leq \sigma \sqrt{\frac{2t + 2 \log(2p)}{n}} + \frac{t + \log(2p)}{3n}.
\]
PROOF. It suffices to prove part (i) since the second inequality can be obtained from the same argument. Fix \( j \), we have

\[
\|\hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*)\|_\infty = \max_{1 \leq k \leq p} \left| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \left\{ \Delta_i \hat{K}_h(x_i^T \beta_j^* - y_i) - \tau_0 \right\} x_{ik} \right|
\]

where \( \xi_{ij} = \Delta_i \hat{K}_h(x_i^T \beta_j^* - y_i) - \tau_0 \) is such that \( |\xi_{ij}| \leq \tau_0 = \max(\tau_0, 1 - \tau_0) \) and \( \mathbb{E}(\xi_{ij} x_{ik})^2 \leq \tau_0^2 \sigma_{kk} \). Applying Bernstein’s inequality yields that, with probability at least \( 1 - 2e^{-z} \),

\[
\left| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \xi_{ij} x_{ik} \right| \leq \tau_0 \left( \frac{\sqrt{2z}}{n} + \frac{z}{3n} \right)
\]

for any \( 1 \leq k \leq p \).

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

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

\[
\psi_j(r, q) \leq \left( \frac{q}{h} \sqrt{\log p} \sqrt{\frac{2\sqrt{t + \log p}}{n}} \right) \cdot \frac{t + \log p}{n}.
\]

**PROOF.** For any \( j \) fixed, and \( k = 1, \ldots, p \), define after a change of variable \( v = \beta - \beta_j^* \) that

\[
\psi_{j,k}(r, q) = \sup_{v \in \Theta(r) \cap A(q)} \left| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \left\{ \hat{K}_h(\varepsilon_{ij} - x_i^T v) - \hat{K}_h(\varepsilon_{ij}) \right\} x_{ik} \right|
\]

where \( \varepsilon_{ij} = y_i - x_i^T \beta_j^* \). Then \( \psi_j(r, q) \leq \max_{1 \leq k \leq p} \psi_{j,k}(r, q) \). Note that \( \sup_v |g_v(y_i, x_i)| \leq |x_{ik}| \leq 1 \). Let \( \sigma \) be any positive number such that \( \sigma^2 \geq \sup_{v \in \Theta(r) \cap A(q)} \mathbb{E}g_v^2(y_i, x_i) \). By Theorem 7.3 in [3]—an improved version of Talagrand’s inequality, we obtain that for any \( z > 0 \),

(D.2) \[
\psi_{j,k}(r, q) \leq \mathbb{E}\psi_{j,k}(r, q) + \sqrt{\left\{ \sigma^2 + 2\mathbb{E}g_v^2(y_i, x_i) \right\} \frac{2z}{n} + \frac{z}{3n}}
\]

holds with probability at least \( 1 - e^{-z} \). For the second moment \( \mathbb{E}g_v^2(y_i, x_i) \), by a change of variable and Minkowski’s integral inequality we derive that

\[
\mathbb{E}g_v^2(y_i, x_i) = \mathbb{E} \left[ x_{ik}^2 \int_{-\infty}^{\infty} \left\{ \hat{K}_h(u - x_i^T v) - \hat{K}_h(u) \right\}^2 f_y(x_i^T \beta_j^* + u|x_i)du \right]
\]

\[
= h \mathbb{E} \left[ x_{ik}^2 \int_{-\infty}^{\infty} \left\{ \hat{K}(v - x_i^T v/h) - \hat{K}(v) \right\}^2 f_y(x_i^T \beta_j^* + v/h|x_i)dv \right]
\]

\[
\leq \frac{h}{n} \mathbb{E} \left[ x_{ik}^2 (x_i^T v)^2 \int_{-\infty}^{\infty} \left\{ \int_0^1 K(v - w x_i^T v/h)dw \right\}^2 dw \right]
\]

\[
\leq \frac{h}{n} \mathbb{E} \left[ (x_i^T v)^2 \left\{ \int_0^1 \left\{ \int_{-\infty}^{\infty} K^2(v - w x_i^T v/h)dw \right\}^{1/2} dw \right\}^2 \right]
\]

\[
\leq \kappa u h^{-1} \mathbb{E}(x_{ik} \cdot x_i^T v)^2 \leq \kappa u_j h^{-1} r^2, \text{ valid for any } v \in \Theta(r).
\]

It remains to bound \( \mathbb{E}\psi_{j,k}(r, q) \) in the concentration inequality (D.2). Note that \( g_v(y_i, x_i) \) is \((\kappa_u/h)\)-Lipschitz continuous in \( x_i^T v \), i.e., for any \( v, v' \),

\[
|g_v(y_i, x_i) - g_{v'}(y_i, x_i)| \leq \frac{\kappa_u}{h} \|v - v'\|.
\]
(\kappa_u/h)|x_i^\tau v - x_i^\nu v'|$. Hence, it follows from Rademacher symmetrization and Talagrand’s contraction principle that
\[
E\psi_{j,k}(r, q) \leq 2E\left\{ \sup_{v \in \Theta(r) \cap \Lambda(q)} \left| \frac{1}{n} \sum_{i=1}^n e_i g_v(y_i, x_i) \right| \right\}
\leq 4\kappa_u E\left\{ \sup_{v \in \Theta(r) \cap \Lambda(q)} \left| \frac{1}{nh} \sum_{i=1}^n e_i \bar{x}_i^r v \right| \right\} \leq 4\kappa_u \frac{qr}{nh} \cdot E \left\| \sum_{i=1}^n e_i x_i \right\|_\infty,
\]
where \(e_1, \ldots, e_n\) are independent Rademacher variables. By Hoeffding’s moment inequality,
\[
E_e \left\| \sum_{i=1}^n e_i x_i \right\|_\infty \leq \max_{1 \leq k \leq p} \left( \sum_{i=1}^n x_{i,k}^2 \right)^{1/2} \sqrt{2 \log(2p)},
\]
where \(E_e\) denotes the expectation over \(\{e_i\}_{i=1}^n\). Plugging this into the previous bound yields
\[
E\psi_{j,k}(r, q) \leq 4\kappa_u \frac{qr}{h} \sqrt{\frac{2 \log(2p)}{n}}.
\]
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 \(4\eta_1/4r \leq h \leq g/(2\eta_1)\) with \(\eta_1\) defined in (B.1). Then, for any \(0 \leq j \leq m\) and \(t > 0\),
\[
P\left\{ D(\beta, \beta_j^*) \geq \frac{1}{2} q \kappa_1 \cdot \|\beta - \beta_j^*\|_2^2 \text{ for all } \beta \in \beta_j^* + \Theta(r) \cap \Lambda(q) \right\} \geq 1 - e^{-t}
\]
provided that \(n \geq h(q/r)^2(t \vee \log p)\).

**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 \(r\)-levels, \(\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U\). Again, we start with constructing intermediate points \(\bar{\beta}_j = (1 - u_j)\beta_j^*_j + u_j \beta_j\) that satisfy \(\bar{\beta}_j \in \beta_j^* + \Theta(r^\Diamond)\), where \(r^\Diamond = h/(4\eta_1/4)\). For each \(\bar{\beta}_j\), by the first-order optimality condition, there exists some subgradient \(\bar{g}_j \in \partial \|\bar{\beta}_j\|_1\) such that \(\bar{Q}_j(\bar{\beta}_j) + \lambda_j \cdot \bar{g}_j = 0\) and \(\langle \bar{g}_j, \bar{\beta}_j \rangle = \|\bar{\beta}_j\|_1\). Consequently, for each \(j = 0, 1, \ldots, m\),
\[
D(\bar{\beta}_j, \beta_j^*) \leq u_j D(\bar{\beta}_j, \beta_j^*) = u_j \langle -\lambda_j \bar{g}_j - \bar{Q}_j(\beta_j^*), \bar{\beta}_j - \beta_j^* \rangle
\]
\begin{equation}
\leq \lambda_j \left( \|\bar{\beta}_j - \beta_j^*\|_1 - \| (\bar{\beta}_j - \beta_j^*) S_j \|_1 \right) + \langle -\bar{Q}_j(\beta_j^*), \bar{\beta}_j - \beta_j^* \rangle,
\end{equation}
where \(S_j = \text{supp}(\beta_j^*)\). Denote the cardinality of \(S_j\) by \(s_j\), satisfying \(s_j \leq s\) for all \(j\). Consider the decomposition
\[
\bar{Q}_j(\beta_j^*) = \bar{Q}_0(\beta_j^*) - Q_0(\beta_j^*) - \sum_{\ell=1}^{j-1} w_{\ell} (\Delta_{\ell} + \Delta_{\ell})
\]
\begin{align*}
&+ Q_0(\beta_j^*) - \sum_{\ell=0}^{j-1} w_{\ell} E\{ \tilde{K}_h(y - x^\tau \beta_j^* x) \} + \sum_{\ell=0}^{j-1} w_{\ell} B_\ell(\hat{\beta}_\ell),
\end{align*}
where $\Delta_{\ell} (\ell = 0, 1, \ldots, m - 1)$ are given in (B.11),

$$
\tilde{\Delta}_{\ell} = \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \{ \tilde{K}_h(y_i - x_i^T \tilde{\beta}_{\ell}) - \tilde{K}_h(y_i - x_i^T \beta^*) \} x_i,
$$

and $B_\ell(\beta) = \mathbb{E}\{ \tilde{K}_h(y - x^T \beta) \} - \mathbb{E}\{ \tilde{K}_h(y - x^T \beta^*) \} x$. 

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

$$(D.4) \quad \|\langle \tilde{Q}_j(\beta^*_j), \tilde{\beta}_j - \beta^*_j \rangle\|^{\Sigma}$$

$$
\leq \left\{ \|Q_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} + \sum_{\ell=0}^{j-1} \|Q_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} \right\} \|\tilde{\beta}_j - \beta^*_j\|_{\Sigma}$$

$$
= \left\{ \|Q_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} + \sum_{\ell=0}^{j-1} \|Q_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} \right\} \|\tilde{\beta}_j - \beta^*_j\|_{\Sigma}$$

and $\langle \tilde{Q}_0(\beta^*_0), \tilde{\beta}_0 - \beta^*_0 \rangle \leq \|\tilde{Q}_0(\beta^*_0) - Q_0(\beta^*_0)\|_{\Sigma} \|\beta - \beta^*\|_{\Sigma} + a \|\tilde{\beta}_0 - \beta^*_0\|_{\Sigma}$, where

$$(D.5) \quad a = 0.5 l_1 \kappa_2 h^2 + \tilde{f}^{-1} \delta^* \quad \text{and} \quad W_j = \sum_{\ell=0}^{j-1} w_\ell = \int_{r_\ell}^{r_{j+1}} dH(u) = \log \left( \frac{1-r_\ell}{1-r_j} \right).$$

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

$$
\mathcal{G} = \left\{ \|\tilde{Q}_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} \leq \frac{\lambda_0}{2} \right\}
$$

$$
\bigcap_{j=1}^{m} \left\{ \|\tilde{Q}_0(\beta^*_j) - Q_0(\beta^*_j)\|_{\Sigma} + \sum_{\ell=0}^{j-1} w_\ell \|\Delta_{\ell}\|_{\Sigma} \leq \frac{\lambda_j}{3}, \quad j = 1, \ldots, m \right\} \quad \text{and}
$$

$$
\mathcal{F} = \bigcap_{j=0}^{m} \left\{ D(\beta, \beta^*_j) \geq \kappa \cdot \|\beta - \beta^*_j\|^2_{\Sigma} \quad \text{for all} \quad \beta \in \beta^*_j + \Theta(\delta^*) \cap \Lambda(q_j) \right\}.
$$

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

$$(D.6) \quad 0 \leq D(\tilde{\beta}_0, \beta^*_0) < \frac{\lambda_0}{2} (3 \|\tilde{\beta}_0 - \beta^*_0\|_{\Sigma} 1 - \|\beta^*_0\|_{\Sigma} 1 + a \|\tilde{\beta}_0 - \beta^*\|_{\Sigma},$$

thus implying the cone-like constraint $\|\langle \tilde{\beta}_0 - \beta^*_0, \beta^*_0 \rangle \|_{\Sigma} 1 \leq 3 \|\beta^*_0\|_{\Sigma} 1 + (2a/\lambda_0) \|\tilde{\beta}_0 - \beta^*\|_{\Sigma}$. Taking $q_0 = 4(s_0/\gamma)^{1/2} + 2a/\lambda_0$, we see that $\tilde{\beta}_0$ falls into the cone-like set $\beta^*_0 + \Lambda(q_0)$, and so does $\tilde{\beta}_0$. Hence, $D(\tilde{\beta}_0, \beta^*_0) \geq \kappa \cdot \|\tilde{\beta}_0 - \beta^*_0\|^2_{\Sigma}$. Combining this with (D.6) yields, after some algebra, that

$$(D.7) \quad \|\tilde{\beta}_0 - \beta^*_0\|_{\Sigma} < r_0 := \kappa^{-1} \left\{ 1.5(s/\gamma)^{1/2} \lambda_0 + a \right\}.$$
Provided $r_0 \leq r^\diamond$, $\tilde{\beta}_0$ lies in the interior of the local region $\beta^*_0 + \Theta(r^\diamond)$. As before, we argue by contradiction that $\tilde{\beta}_0$ coincides with $\tilde{\beta}_0$, thus implying $\tilde{\beta}_0 \in \beta^*_0 + \Theta(r_0) \cap \Lambda(q_0)$.

For $(\tilde{\beta}_1, \beta_1)$, from (D.3) and (D.4) it follows that

$$0 \leq D(\tilde{\beta}_1, \beta_1)$$

$$< \lambda_1 \left( \| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1 - \| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1 \right)$$

$$+ \left( \lambda_1/3 + w_0 \| \Delta_0 \|_\infty \right) \| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1 + \{ a + w_0 (\mathcal{F} r_0 + a) \} \| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1,$$

which in turn implies $\tilde{\beta}_1 \in \beta^*_1 + \Lambda(q_1)$ with $q_1 := 6(s_1/\gamma)^{1/2} + 3\{ (1 + w_0) a + \mathcal{F} w_0 r_0 \}/\lambda_1$. On the event $\mathcal{F}, D(\tilde{\beta}_1, \beta_1) \geq \kappa \cdot \| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1$. Combining the upper and lower bounds yields

$$\| (\tilde{\beta}_1 - \beta^*_1)_{S_1} \|_1 < \kappa^{-1} \left\{ \frac{5}{3} (s/\gamma)^{1/2} \lambda_1 + (1 + w_0) a + w_0 \mathcal{F} r_0 \right\}.$$  \hspace{1cm} (D.8)

Provided $r_1 \leq r^\diamond$, we reach the conclusion that $\tilde{\beta}_1 \in \beta^*_1 + \Theta(r_1) \cap \Lambda(q_1)$.

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

$$0 \leq D(\tilde{\beta}_j, \beta^*_j)$$

$$< \lambda_j \left( \| (\tilde{\beta}_j - \beta^*_j)_{S_1} \|_1 - \| (\tilde{\beta}_j - \beta^*_j)_{S_1} \|_1 \right)$$

$$+ \left( \frac{\lambda_j}{3} + \sum_{\ell=0}^{j-1} w_\ell \| \Delta_\ell \|_\infty \right) \| (\tilde{\beta}_j - \beta^*_j)_{S_1} \|_1 + \left\{ a + \sum_{\ell=0}^{j-1} w_\ell (\mathcal{F} r_\ell + a) \right\} \| (\tilde{\beta}_j - \beta^*_j)_{S_1} \|_1,$$

Conditioning on the event $\{ \sum_{\ell=0}^{j-1} w_\ell \psi(r_\ell, q_\ell) \leq \lambda_j/3 \}$, we obtain the cone-like constraint $\tilde{\beta}_j \in \beta^*_j + \Lambda(q_j)$ with $q_j := 6(s_j/\gamma)^{1/2} + 3\{ (1 + W_j) a + \mathcal{F} \sum_{\ell=0}^{j-1} w_\ell r_\ell \}/\lambda_j$, thus implying

$$\| (\tilde{\beta}_j - \beta^*_j)_{S_1} \|_1 < \kappa^{-1} \left\{ \frac{5}{3} (s/\gamma)^{1/2} \lambda_j + (1 + W_j) a + \mathcal{F} \sum_{\ell=0}^{j-1} w_\ell r_\ell \right\}. \hspace{1cm} (D.9)$$

In view of (D.7)–(D.9), we write

$$A_j = \frac{1}{\kappa} \left\{ \frac{5}{3} (s/\gamma)^{1/2} \lambda_j + (1 + W_j) a \right\}, \hspace{0.5cm} c_j = \kappa^{-1} \mathcal{F} w_j, \hspace{0.5cm} j = 0, 1, \ldots,$$

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

$$r_j \leq \prod_{\ell=0}^{j-1} (1 + c_\ell) \cdot A_j \leq e^{\sum_{\ell=0}^{j-1} c_\ell} \cdot A_j = C_j \cdot A_j \hspace{0.5cm} \text{with} \hspace{0.5cm} C_j := \left( \frac{1 - \tau_j}{1 - \tau_j} \right)^{J/\kappa}.$$
In particular, we write \( C_0 = 1 \). As long as \( \max_{0 \leq j \leq m} r_j \leq r^\oplus \), we have established the result \( \hat{\beta}_j \in \beta_j^* + \Theta(r_j) \cap \Lambda(q_j) \) (\( j = 0, 1, \ldots, m \)) as a deterministic claim, conditioned on the event

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

with properly chosen \( \kappa > 0 \) and regularization parameters \( \lambda_0, \lambda_1, \ldots, \lambda_m \), where \( q_0 = 4(s_0/\gamma)^{1/2} + 2a/\lambda_0 \) and \( q_j = 6(s_j/\gamma)^{1/2} + 3\{(1 + W_j)a + \mathcal{F} \sum_{\ell=0}^{j-1} w_\ell r_\ell\}/\lambda_j \) for \( j \geq 1 \) 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 \( 1 - 2(m + 1)p^{-2} \),

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

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

\[
h^2 \lesssim (s/\lambda_0)^{1/2} \sqrt{\frac{\log p}{n}}.
\]

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

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

Recall that \( a \propto h^2 + \delta^\star \lesssim h^2 + n^{-1/2} \). We then choose \( \lambda_1 \propto (1 + W_1) \sqrt{\log(p)/n} \) so that \( \lambda_1 \geq 3 \max\{w_0 \psi_0(r_0, q_0), \|\hat{Q}_0(\beta_1^*) - Q_0(\beta_1^*)\|_\infty + w_0 \|\Delta_1\|_\infty\} \) as long as \( h \gtrsim (\kappa \gamma)^{-1}s\sqrt{\log(p)/n} \). Furthermore, it follows that

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

For a general \( j \geq 2 \), assume we already have \( \lambda_\ell \propto (1 + W_\ell) \sqrt{\log(p)/n} \), \( q_\ell \lesssim C_\ell(s/\gamma)^{1/2} \) and \( r_\ell \lesssim C_\ell(s/\gamma)^{1/2} \lambda_\ell/\kappa \) 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

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

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

\[
(1 + W_j) \sqrt{\frac{\log p}{n}} \lesssim \lambda_j \gtrsim 3 \max\left\{ \sum_{\ell=0}^{j-1} w_\ell \psi_\ell(r_\ell, q_\ell), \|\hat{Q}_0(\beta_j^*) - Q_0(\beta_j^*)\|_\infty + \sum_{\ell=0}^{j-1} w_\ell \|\Delta_\ell\|_\infty \right\}
\]

and therefore the event that involves \( \lambda_j \) is certified.
With the above choice of \( \{r_j\}_{j=0,1,...,m} \) and the lower bound constraint on the magnitude of the bandwidth—\( h \gtrsim (\kappa^2) \frac{1}{s} \sqrt{\log(p)/n} \), we have

\[
C_j \kappa^{-1}(1 + W_j)(s/\kappa)^{1/2} \sqrt{\log p \over n} \preceq r_j \preceq r^\diamond \gtrsim h \quad \text{for all } j = 0, 1, \ldots, m.
\]

Finally, by the restricted strong convexity lemma—Lemma D.3 with \( r = h/(4n_{1/4}) \), \( q \propto (s/\kappa)^{1/2} \) and \( t = 2 \log p \) we take \( \kappa = (g_\kappa) / 2 \) so that event \( \mathcal{F} \) happens with probability at least \( 1 - (m + 1)p^{-2} \) provided that the “effective” sample size satisfies \( nh \gtrsim s \log p \). \hfill \square

**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 \( \beta, \beta' \in \mathbb{R}^p \), and define the function \( f(\eta) = \langle Q(\beta_\eta) - Q(\beta'), \beta - \beta' \rangle \) for \( \eta \in [0, 1] \). Since \( Q(\cdot) \) is differentiable with a positive semi-definite Jacobian, we have \( f'(\eta) = \langle \nabla Q(\beta_\eta)(\beta - \beta'), \beta - \beta' \rangle \geq 0 \), and hence \( f(\cdot) \) is non-decreasing. Consequently, for any \( \eta \in [0, 1] \),

\[
D(\beta,\beta^*) = \langle Q(\beta) - Q(\beta'), \beta - \beta' \rangle = \eta(Q(\beta) - Q(\beta'), \beta - \beta') = \eta f(\eta) \leq \eta f(1) = \eta(Q(\beta) - Q(\beta'), \beta - \beta') = \eta D(\beta,\beta'),
\]

as claimed. \hfill \square

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

\[
\Delta := \left\| {1 \over n} \sum\limits_{i=1}^{n} (z_i \xi_i - \mathbb{E} \xi_i z_i) \right\|_2 = \sup_{u \in \mathbb{R}^p} {1 \over n} \sum_{i=1}^{n} (1 - \mathbb{E} f_u(\xi_i, z_i)),
\]

where \( f_u(\xi_i, z_i) := \langle u, \xi_i z_i \rangle \) satisfies \( |f_u(\xi_i, z_i)| \leq M \zeta p \) and \( \mathbb{E} \{ f_u^2(\xi_i, z_i) \} = \mathbb{E} \{ \xi_i^2 \langle u, z_i \rangle^2 \} \leq \sigma^2 \). Applying a refined Talagrand’s inequality (see, e.g., Theorem 7.3 in [3]) yields that with probability at least \( 1 - e^{-t} \),

\[
\Delta \leq 2 \mathbb{E} \Delta + \sigma \sqrt{{2t \over n} + {4M \zeta p t \over 3n}}.
\]

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

\[
\mathbb{E} \Delta \leq \left( \mathbb{E} \left\| {1 \over n} \sum_{i=1}^{n} (z_i \xi_i - \mathbb{E} \xi_i z_i) \right\|_2^2 \right)^{1/2} \leq \left( \sum_{i=1}^{n} \mathbb{E} \| \xi_i z_i - \mathbb{E} \xi_i z_i \|_2^2 \right)^{1/2} \leq \sigma \sqrt{{p \over n}}.
\]

Combining the above two displays gives

\[
\left\| {1 \over n} \sum_{i=1}^{n} (z_i \xi_i - \mathbb{E} \xi_i z_i) \right\|_2 \leq 2 \sigma \sqrt{{p \over n}} + \sigma \sqrt{{2t \over n} + M \zeta p^4 t} \over 3n
\]

holds with probability at least \( 1 - e^{-t} \). \hfill \square
E.3. Proof of Lemma B.3. Proof of (i). For each \( j = 0, 1, \ldots, m \), define random variables \( \xi_{ij} = \Delta_i \tilde{K}_h(x_i^j \beta_j^* - y_i - \tau_0) \), so that the centered process can be written as \( \Sigma^{-1/2} \{ \tilde{Q}_0(\beta_j^*) - \mathbb{E}\tilde{Q}_0(\beta_j^*) \} = (1/n) \sum_{i=1}^n (\xi_{ij} z_i - \mathbb{E}\xi_{ij} z_i) \). Since \( \Delta_i \in \{0, 1\} \) and \( 0 \leq \tilde{K}(\cdot) \leq 1 \), we have \( |\xi_{ij}| \leq \tau_0 = \max(\tau_0, 1 - \tau_0) \). In particular, for \( j = 0 \), it is shown in the proof of Lemma C.2 in [6] that \( \mathbb{E}(\xi_{ij}^2 | x_i) \leq \tau_0(1 - \tau_0) + (1 + \tau_0)h^2 \). For general \( j \geq 1 \), we can simply use the crude second moment bound \( \mathbb{E}(\xi_{ij}^2 | x_i) \leq \tau_0^2 \). 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 \( |\tilde{K}_h(x_i^j \beta_j^* - y_i) - \tilde{K}_h(x_i^j \beta_j^* - y_i)| \leq 1 \) and \( \mathbb{E}\tilde{K}_h^2(x_i^j \beta_j^* - y_i | x_i) \leq 1 \).

Proof of (iii). The proof is based on a similar argument used in the proof of Lemma B.2. Fix \( j \), set \( v = \beta_{j+1}^* - \beta_j^* \) so that \( \|v\| \leq f^{-1} \delta^* \). By the monotonicity of \( u \mapsto x_i^j \beta_j^*(u) \) and \( \tilde{K}_h(\cdot) \), we have

\[
\left| \int_{\tau_j}^{\tau_{j+1}} \{ \tilde{K}_h(x_i^j \beta_j^*(u) - y_i) - \tilde{K}_h(x_i^j \beta_j^*(\tau_j) - y_i) \} \, dH(u) \right| \leq w_j \{ \tilde{K}_h(x_i^j \beta_{j+1}^* - y_i) - \tilde{K}_h(x_i^j \beta_j^* - y_i) \} \leq \kappa_u w_j h^{-1} x_i^j \|v\| \leq \kappa_u f^{-1} w_j h^{-1} \delta^*,
\]

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

\[
\mathbb{E}\left[ \left( \int_{\tau_j}^{\tau_{j+1}} \{ \tilde{K}_h(x_i^j \beta_j^*(u) - y_i) - \tilde{K}_h(x_i^j \beta_j^*(\tau_j) - y_i) \} \, dH(u) \right)^2 \right] \leq w_j^2 h \int_{-\infty}^{\infty} \{ \tilde{K}(v + x_i^j v/h) - \tilde{K}(v) \}^2 f_y(u | x) \, du \leq \tilde{f} w_j^2 h^{-1} (x_i^j v)^2 \int_{-\infty}^{\infty} \left( \int_0^1 K(v + w x_i^j v/h) \, dw \right)^2 \, dv \leq \tilde{f} w_j^2 h^{-1} (x_i^j v)^2 \left( \int_0^1 \left( \int_{-\infty}^{\infty} K^2(v + w x_i^j v/h) \, dv \right) \frac{1}{2} \, dw \right)^2 \leq \kappa_u \tilde{f} w_j^2 h^{-1} (x_i^j v)^2,
\]

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

\[
\mathbb{E}\left\{ \int_{\tau_j}^{\tau_{j+1}} \{ \tilde{K}_h(x_i^j \beta_j^*(u) - y_i) - \tilde{K}_h(x_i^j \beta_j^*(\tau_j) - y_i) \} \, dH(u) \right\} \leq \kappa_u \tilde{f} w_j^2 h^{-1} \left\{ \mathbb{E}(x_i^j v)^4 \right\}^{1/2} \left\{ \mathbb{E}(z_i^j u)^4 \right\}^{1/2} \leq \kappa_u \tilde{f} \int_{-2}^{-2} m_4 w_j^2 h^{-1} \delta^2,
\]

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 \( j = 0, 1, \ldots, m \), we write \( \beta^* = \beta_j^* \), \( \tilde{Q}(\cdot) = \tilde{Q}_j(\cdot) \) and \( Q(\cdot) = \mathbb{E}\tilde{Q}(\cdot) \) for simplicity. Recall the smoothed estimating
functions defined in (3), (4), and the induced metric (symmetrized Bregman divergence)

\[ D(\beta, \beta^*) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \{ \bar{K}_h(x_i^T \beta - y_i) - \bar{K}_h(x_i^T \beta^* - y_i) \} x_i^T (\beta - \beta^*) , \]

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

\[ D(\beta, \beta^*) \geq \frac{\kappa_l}{nh} \sum_{i=1}^{n} \Delta_i \{ x_i^T (\beta - \beta^*) \}^2 1_{E_i} . \]

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

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

\[ u^2 1(|u| \leq R/2) \leq \varphi_R(u) \leq u^2 1(|u| \leq R) . \]

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

\[ D(\beta, \beta^*) \geq \frac{\kappa_l}{\| \beta - \beta^* \|_\Sigma^2} \geq D_0(\delta) := \frac{1}{n\eta h} \sum_{i=1}^{n} \omega_i \cdot \varphi_{h/(2r)}(z_i^T \delta) , \]

where \( \omega_i := 1 \{ |x_i^T \beta^* - y_i| \leq h/2, \Delta_i = 1 \} \).

We first bound the expectation \( \mathbb{E}\{ D_0(\delta) \} \), and then control the concentration of \( D_0(\delta) \) around \( \mathbb{E}\{ D_0(\delta) \} \). When \( 0 < h \leq 1 \), Condition 3.3 ensures that

\[ g \cdot h \leq \mathbb{E}(\omega_i | x_i) = \int_{x_i^T \beta^* - h/2}^{x_i^T \beta^* + h/2} g(u | x_i) du \leq \mathbb{E} \] almost surely.

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

\[ \mathbb{E}\{ \omega_i \cdot \varphi_{h/(2r)}(z_i^T \delta) \} \geq gh \cdot \mathbb{E}\varphi_{h/(2r)}(z_i^T \delta) \geq gh \cdot \mathbb{E}\{ (z_i^T \delta)^2 1 (|z_i^T \delta| \leq h/(4r)) \} \]

\[ = gh \cdot \{ 1 - \mathbb{E}(z_i^T \delta)^2 1 (|z_i^T \delta| > h/(4r)) \} , \]

which further implies

\[ \inf_{\delta \in S^{p-1}} \mathbb{E}\{ D_0(\delta) \} \geq g \cdot \left[ 1 - \sup_{u \in S^{p-1}} \mathbb{E}\{ (z_i^T \delta)^2 1 (|z_i^T \delta| > h/(4r)) \} \right] . \]

By the definition of \( \eta \) in (B.1), as long as \( 0 < r \leq h/(4n_1/4) \),

\[ \inf_{\delta \in S^{p-1}} \mathbb{E}\{ D_0(\delta) \} \geq \frac{3}{4} g . \]

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

\[ \Lambda = \sup_{\delta \in S^{p-1}} \{ -D_0(\delta) + \mathbb{E} D_0(\delta) \} . \]

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

\[ 0 \leq \left( \omega_i/h \right) \varphi_{h/(2r)}(z_i^T \delta) \leq \omega_i \min\{ (4r)^{-2} h, (4r)^{-1} |z_i^T \delta| \} . \]
Combining this with (E.7) yields
\[ E \{ (\omega_i/h)^2 \varphi_{h/(2r)} (z_i^T \delta) \} \leq (4r)^{-2} E \{ (\omega_i | x_i) (z_i^T \delta)^2 \} \leq (4r)^{-2} gh. \]
With the above preparations, we apply a refined Talagrand’s inequality—Theorem 7.3 in [3]—to obtain that, for any \( t > 0 \),
\[ \Lambda \leq E \Lambda + (E \Lambda)^{1/2} \sqrt{\frac{ht}{4r^2n}} + \tilde{g}^{1/2} \sqrt{\frac{ht}{8r^2n}} + \frac{h}{(4r)^2} \frac{t}{3n} \]
(E.10)
with probability at least \( 1 - e^{-t} \). It remains to bound \( E \Lambda \). To this end, we define
\[ \mathcal{E}(\delta; z_i, y_i) = \frac{\omega_i}{h} \varphi_{h/(2r)} (z_i^T \delta) = \frac{1}{h} \varphi_{\omega_i h/(2r)} (\omega_i z_i^T \delta), \quad \delta \in \mathbb{S}^{p-1}. \]
where the second equality follows from the property that \( \varphi_{cR}(cu) = c^2 \varphi_R(u) \) for any \( c \geq 0 \). By the Lipschitz continuity of \( \varphi_R(\cdot) \), \( \mathcal{E}(\delta; z_i, y_i) \) is \( (2r)^{-1} \)-Lipschitz continuous in \( \omega_i z_i^T \delta \), and \( \mathcal{E}(\delta; z_i, y_i) = 0 \) for any \( \delta \) such that \( \omega_i z_i^T \delta = 0 \). Furthermore, define the subset \( T \subseteq \mathbb{R}^n \) as
\[ T = \{ t = (t_1, \ldots, t_n)^T : t_i = \omega_i z_i^T \delta, i = 1, \ldots, n, \delta \in \mathbb{S}^{p-1} \}, \]
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 \( |\phi(t) - \phi(s)| \leq |t - s| \) for all \( t, s \in \mathbb{R} \). Let \( \epsilon_1, \ldots, \epsilon_n \) be independent Rademacher random variables, and denote by \( \mathbb{E} \) the expectation taken only with respect to \( \epsilon_i \)'s. Then, via a standard symmetrization and contraction argument (see, e.g. Lemma 6.3 and Theorem 4.12 in [9]), we have
\[ \mathbb{E} \epsilon \Lambda \leq 2 \mathbb{E}_\epsilon \left\{ \sup_{\delta \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \mathcal{E}(\delta; z_i, y_i) \right\} = \frac{1}{r} \mathbb{E}_\epsilon \left\{ \sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \phi_i(t_i) \right\} \leq \frac{1}{r} \mathbb{E}_\epsilon \left\{ \sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \omega_i z_i^T \delta \right\} \leq \frac{1}{r} \mathbb{E}_\epsilon \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \omega_i z_i \right\|_2. \]
Taking the expectation over \( \{(z_i, y_i)\}_{i=1}^{n} \) on both sides yields \( E \Lambda \leq \tilde{g}^{1/2} \sqrt{hp/(r^2n)} \). Substituting this into (E.10), we obtain
(E.11)
\[ \Lambda \leq \tilde{g}^{1/2} r^{-1} \left( \frac{5}{4} \sqrt{\frac{hp}{r^2n}} + \sqrt{\frac{ht}{8r^2n}} + (1/4 + 1/48) r^{-2} \frac{ht}{n} \right) \]
with probability at least \( 1 - e^{-t} \).
Finally, combining (E.6), (E.8), (E.9) and (E.11) completes the proof. \( \square \)

**E.5. Proof of Lemma B.5.** To begin with, define the centered process
\[ S(\beta) = \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) K_h(y_i - x_i^T \beta) z_i, \quad \beta \in \mathbb{R}^p. \]
After a change of variable \( v = \Sigma^{1/2}(\beta - \beta_j^*) \), we have
\[ \sup_{\beta \in \beta_j^* + \Theta(r)} \| S(\beta) - S(\beta_j^*) \|_2 = \sup_{v \in \mathbb{B}^p(r)} \| S(\beta_j^* + \Sigma^{-1/2} v) - S(\beta_j^*) \|_2. \]

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

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

$$\nabla \Psi(v) = \frac{1}{n} \sum_{i=1}^{n} \{ \phi_i, v z_i z_i^T - \mathbb{E}(\phi_i z z^T) \},$$

where $\phi_i, v = K_h(z_i^T v - \varepsilon_i)$, $\phi_i = K_h(z_i^T v - \varepsilon)$ and $\varepsilon_i = y_i - x_i^T \beta_j^*$. It is easy to see that $0 \leq \phi_i, v \leq \kappa_u/h$ with $\kappa_u = \sup_{u \in \mathbb{R}} K(u)$. For any $g, h \in \mathbb{S}^{p-1}$ and $|\lambda| \leq \min\{nh/(\kappa_u \zeta^2), n/\bar{f}\}$, by independence and the elementary inequality $e^u \leq 1 + u^2 e^{|u|/2}$, we obtain that

$$\mathbb{E}\exp\{ \lambda g^T \nabla \Psi(v) h \}$$

$$\leq 1 + \frac{\lambda^2}{2n^2} e^{\frac{\lambda |\lambda|}{n} \|g z^T h\|_2} \mathbb{E}\left\{ \phi_v z^T g z^T h - \mathbb{E}(\phi_v z^T g z^T h) \right\}^2 e^{\frac{\kappa_u |\lambda|}{nh} |z^T g z^T h|}$$

$$\leq 1 + \frac{\lambda^2}{2n^2} \mathbb{E}\left\{ \phi_v z^T g z^T h - \mathbb{E}(\phi_v z^T g z^T h) \right\}^2 e^{\kappa_u \zeta^2 |\lambda|/(nh)}$$

(E.12) \leq \left\{ 1 + \frac{(e \lambda)^2}{2n^2} \mathbb{E}(\phi_v z^T g z^T h)^2 \right\}^n,$$
With $v = \Sigma^{1/2}(\beta - \beta^*)$, note that
\[ \mathbb{E}\{K_h(y - \langle x, \beta^*_j \rangle + t(\beta - \beta^*_j))|x\} \]
\[ = \frac{1}{h} \int_{-\infty}^{\infty} \frac{u - tz^v}{h} f_y(x^\top \beta^*_j + u|x)du = \int_{-\infty}^{\infty} K(v)f_y(x^\top \beta^*_j + tz^v + hv|x)dv. \]

By the Lipschitz continuity of $f_y(\cdot|x)$, we have
\[ \left\| \mathbb{E}\{K_h(y - x^\top \beta) - K_h(y - x^\top \beta^*_j)\} z + \mathbb{E}\{f_y(x^\top \beta^*_j|x)zz^\top v\} \right\|_2 \]
\[ = \left\| \mathbb{E}\int_0^1 \int_{-\infty}^{\infty} K(v)f_y(x^\top \beta^*_j + tz^v + hv|x) - f_y(x^\top \beta^*_j|x)zz^\top dvdt \cdot v \right\|_2 \]
\[ \leq l_1 \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}\int_0^1 \int_{-\infty}^{\infty} K(v)(t|z^\top v| + hv|x|)dvdt \cdot |z^\top uz^\top v| \leq l_1 (0.5m_3v + \kappa_1 h)r, \]
as claimed. \hfill \Box

**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 $L := [(\tau_u - \tau_l)/(2\epsilon)] + 1$ subintervals, centered at the points $\tau^k$ for $k \in [L]$, and each of length at most $2\epsilon$. For any $\tau \in [\tau_l, \tau_u]$, there exists some $k$ such that $|\tau - \tau^k| \leq \epsilon$, and hence
\[ \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau}^{\tau^k} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \]
\[ \leq \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau}^{\tau^k} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \]
\[ + \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau^k}^{\tau} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \]
\[ \leq \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau}^{\tau^k} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot z_i \right\|_2 + 2\zeta_\rho |H(\tau) - H(\tau^k)|. \]

For any given $k \in [L]$, applying Lemma B.2 yields that with probability at least $1 - e^{-v}$,
\[ \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau}^{\tau^k} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot z_i \right\|_2 \leq |H(\tau^k) - H(\tau_l)| \left( \sqrt{\frac{p + v}{n}} + \zeta_\rho \frac{v}{n} \right). \]

Recall that $H(u) = -\log(1 - u)$, $u \in (0, 1)$, we have $|H(u) - H(v)| \leq |u - v|/(1 - u \vee v)$. 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
\[ \sup_{\tau \in [\tau_l, \tau_u]} \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{\tau}^{\tau^k} \tilde{K}_h(y_i - x_i^\top \beta^*(u))dH(u) \cdot x_i \right\|_{\Sigma^{-1}} \]
\[ \leq \frac{\tau_u - \tau_l}{1 - \tau_u} \left( \sqrt{\frac{p + \log n + t}{n}} + \zeta_\rho \frac{\log n + t}{n} \right) \]
holds with probability at least $1 - e^{-t}$. This proves the claimed result. \hfill \Box
E.7. Proof of Lemma B.7. For $Q_0(\beta) = \mathbb{E}\{\Delta \tilde{K}_h(x^T \beta - y) - \tau_0\} x$, it follows from integration by parts and change of variables that

\[
\mathbb{E}\{\Delta \tilde{K}_h(x^T \beta^* - y) | x\} = \int_{-\infty}^{\infty} \tilde{K}\left(\frac{x^T \beta^* - t}{h}\right) dG(t | x) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x^T \beta^* - t}{h}\right) G(t | x) dt
\]

\[
= \int_{-\infty}^{\infty} K(u) G(x^T \beta^* + hu | x) du
\]

(E.15)

\[
= G(x^T \beta^* | x) + \int_{-\infty}^{\infty} K(u) \int_{x^T \beta^*}^{x^T \beta^* + hu} \{g(t | x) - g(x^T \beta^* | x)\} dt du.
\]

On the other hand, using the martingale property gives

\[
\mathbb{E}\left[\sum_{\ell=0}^{T_{\tau_j}} \mathbb{1}\{y \geq x^T \beta^*_\ell(u)\} dH(u) | x\right] = \mathbb{E}\{N(x^T \beta^*_\ell | x)\} = \mathbb{P}(y \leq x^T \beta^*_\ell; \Delta = 1 | x) = G(x^T \beta^*_\ell | x).
\]

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

(E.16) \[
\left\| \mathbb{E}\left[\Delta \tilde{K}_h(x^T \beta^*_\ell - y) - \int_{0}^{T_{\tau_j}} \mathbb{1}\{y \geq x^T \beta^*_\ell(u)\} dH(u)\right] x \right\|_{\Sigma^{-1}} \leq \frac{1}{2} l_1 \kappa_2 h^2.
\]

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

\[
\mathbb{E}\{\tilde{K}_h(y - x^T \beta^*_\ell) | x\} = 1 - F_y(x^T \beta^*_\ell | x) - \int_{-\infty}^{\infty} K(u) \int_{x^T \beta^*_\ell}^{x^T \beta^*_\ell + hu} \{f_y(t | x) - f_y(x^T \beta^*_\ell | x)\} dt du.
\]

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

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

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

\[
0 \leq \mathbb{E}\left(\sum_{\ell=0}^{j-1} \int_{T_{\tau_\ell}}^{T_{\tau_{\ell+1}}} \mathbb{1}\{y \geq x^T \beta^*_\ell - 1(y \geq x^T \beta^*(u))\} dH(u) | x\right)
\]

\[
\leq \mathbb{E}\left(\sum_{\ell=0}^{j-1} \int_{T_{\tau_\ell}}^{T_{\tau_{\ell+1}}} \mathbb{1}\{y \geq x^T \beta^*_\ell - 1(y \geq x^T \beta^*_\ell)\} dH(u) | x\right)
\]

\[
= \sum_{\ell=0}^{j-1} \int_{T_{\tau_\ell}}^{T_{\tau_{\ell+1}}} \{F_y(x^T \beta^*_{\ell+1} | x) - F_y(x^T \beta^*_\ell | x)\} dH(u)
\]

\[
\leq \tilde{f} \sum_{\ell=0}^{j-1} w_\ell : x^T (\beta^*_{\ell+1} - \beta^*_\ell).
\]

Recall from the last paragraph of Section 3.1 that $\|\beta^*_{\ell+1} - \beta^*_\ell\|_\Sigma \leq \tilde{f}^{-1} |\tau_{\ell+1} - \tau_\ell|$ for $\ell = 0, 1, \ldots, m - 1$. Under the condition of no censoring below $\tau_0 = \tau_L$, we have $\mathbb{E}\int_{0}^{\tau_0} \mathbb{1}\{y \geq \tilde{f} \sum_{\ell=0}^{j-1} w_\ell : x^T (\beta^*_{\ell+1} - \beta^*_\ell)\} dH(u)$
moments of $V$ so that Lemma A.1 in the Appendix of [7] gives

\[ V \text{ define the centered random variables} \]

Then, for any $\| \cdot \|$ leads to

\[ ||\cdot|| \]

\[ (E.18) \]

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,

\[ \mathbb{E}(a_n, U_i(\tau_1) - U_i(\tau_2))^2 \]

Moreover, for any $q > 2$,

\[ |\mathbb{E}(a_n, U_i(\tau_1) - U_i(\tau_2))| \]

Putting together the pieces, we obtain that for any $\tau_1 < \tau_2$ satisfying $\tau_2 - \tau_1 \geq (\zeta_p/(m_4h)^1/2|\mathbb{E}(a_n, U_i(\tau_1) - U_i(\tau_2))|^2)\zeta_p^{q/2-1} (\tau_2 - \tau_1)^2$.

(E.18) \[ ||G_n(\tau_1) - G_n(\tau_2)||_\psi \lesssim (m_4h)^1/2|\tau_2 - \tau_1|, \]

where $|| \cdot ||_\psi$ denotes the $\psi$-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 $\mathcal{P}(\tau_L, \tau_U, \epsilon, d) \lesssim h^{1/2-\epsilon}$. Let $\bar{\eta} = 2\sqrt{\zeta_p/(m_4n)}$, so that $\lim_{n \to \infty} \bar{\eta} \to 0$, and (E.18) holds for all $\tau_1, \tau_2$ satisfying $d(\tau_1, \tau_2) \geq \bar{\eta}/2$. Applying Lemma A.1 in the Appendix of [7] gives

\[ \sup_{|\tau_1 - \tau_2| < \delta} |G_n(\tau_1) - G_n(\tau_2)| = \sup_{d(\tau_1, \tau_2) < h^{-1/2}\delta} |G_n(\tau_1) - G_n(\tau_2)| \]

(E.19) \[ \leq S_1 + \sup_{d(s, t) < \bar{\eta}/2 + \bar{\eta}/\bar{\eta}} |G_n(s) - G_n(t)|, \]
where the set \( \tilde{\mathcal{T}} \) contains at most \( \mathcal{P}([\tau_L, \tau_U], \tilde{\eta}, d) \lesssim \eta^{1/2} \tilde{\eta}^{-1} \lesssim \sqrt{nh/\zeta_p} \) points, and \( S_1 \) is a random variable satisfying
\[
(\text{E.20})
\]
\[
\mathbb{P}(|S_1| > x) \lesssim \left\{ \int_{\tilde{\eta}/2}^{\eta} \psi^{-1}(\mathcal{P}([\tau_L, \tau_U], \epsilon, d))d\epsilon + (h^{-1/2} \delta + 2\tilde{\eta}) \psi^{-1}(\mathcal{P}^2([\tau_L, \tau_U], \eta, d)) \right\}^{2L} x^{-2L}
\]
for any \( \eta \geq \tilde{\eta} \) and \( x > 0 \). Note that
\[
\int_{\tilde{\eta}/2}^{\eta} \psi^{-1}(\mathcal{P}([\tau_L, \tau_U], \epsilon, d))d\epsilon \lesssim h^{(4L)-1} \int_{\tilde{\eta}/2}^{\eta} \epsilon^{-2(2L)-1}d\epsilon = h^{(4L)-1} \cdot \frac{\eta^{1-(2L)-1} - (\tilde{\eta}/2)^{1-(2L)-1}}{1 - (2L)-1},
\]
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
\[
\mathbb{P}(|S_1| > x) \lesssim \left\{ h^{(4L)-1} \cdot \frac{\eta^{1-(2L)-1} - (\tilde{\eta}/2)^{1-(2L)-1}}{1 - (2L)-1} + (h^{-1/2} \delta + 2\tilde{\eta}) \cdot h^{(2L)-1} \eta^{-L-1} \right\}^{2L} x^{-2L}
\]
\[
= \left\{ 2h^{1/8} - 2h^{1/4}(\tilde{\eta}/2)^{1/2} + h^{1/4} \delta + 2\tilde{\eta} \right\}^2 x^{-2}.
\]
Since \( \tilde{\eta} \to 0 \) and \( h \to 0 \) as \( n \to \infty \), we have for any \( x > 0 \) that
\[
(\text{E.21}) \quad \lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(|S_1| > x) = 0.
\]

It remains to deal with the supremum on the right-hand side of (E.19). For any fixed \( t \in \tilde{\mathcal{T}} \), and \( s \in [\tau_L, \tau_U] \) satisfying \( |s-t| < h^{1/2} \tilde{\eta} = 2 \sqrt{\zeta_p h/(m_q n)} \),
\[
\sup_{|s-t| < h^{1/2} \tilde{\eta}} \left| G^\ast_n (s) - G^\ast_n (t) \right| \leq \sup_{|s-t| < h^{1/2} \tilde{\eta}} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{t}^{s} \tilde{K}_h(y_i - x_i^T \beta^\ast(u))dH(u) \cdot a_i^T x_i \right. \\
+ \left. \sup_{|s-t| < h^{1/2} \tilde{\eta}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \Delta_i \tilde{K}_h(x_i^T \beta^\ast(t) - y_i) - \Delta_i \tilde{K}_h(x_i^T \beta^\ast(s) - y_i) \right) a_i^T x_i \right\}
\]
\[
(\text{E.22}) \quad := I + II.
\]

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
\[
I \leq \sup_{|s-t| < h^{1/2} \tilde{\eta}} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E} \int_{t}^{s} \tilde{K}_h(y_i - x_i^T \beta^\ast(u))dH(u) \cdot a_i^T x_i \right\}
\]
\[
+ \sup_{|s-t| < h^{1/2} \tilde{\eta}} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - \mathbb{E}) \int_{t}^{s} \tilde{K}_h(y_i - x_i^T \beta^\ast(u))dH(u) \cdot a_i^T x_i \right\}
\]
\[
\lesssim n^{-1/2} h^{1/2} \tilde{\eta} \sum_{i=1}^{n} \mathbb{E} |a_i^T x_i| + n^{1/2} h^{1/2} \tilde{\eta} \left( \sqrt{\frac{p + \log n + v}{n}} + \zeta_p \frac{\log n + v}{n} \right)
\]
\[
(\text{E.23}) \quad \lesssim (\zeta_p h)^{1/2} \left( \sqrt{\frac{p + \log n + v}{n}} + \zeta_p \frac{\log n + v}{n} \right)
\]
with probability at least $1 - e^{-v}$ for any $v > 0$. Turning to the second term, by the Lipschitz continuity of $\|\beta^*(\cdot)\|_\Sigma$ as in (10), $|s - t| < h^{1/2}\eta$ indicates $\|\beta^*(s) - \beta^*(t)\|_\Sigma \leq \int_{-1}^{1} h^{1/2}\eta = 2\int_{-1}^{1} \sqrt{\zeta}h/(m_4n)$. Denote $r := 2\int_{-1}^{1} \sqrt{\zeta}h/(m_4n)$, we have

$$\begin{align*}
\Pi &\leq \sup_{|s - t| < h^{1/2}\eta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\{ \Delta_i \tilde{K}_h(x_i^\top \beta^*(t) - y_i) - \Delta_i \tilde{K}_h(x_i^\top \beta^*(s) - y_i) \} a_i^\top x_i \right| \\
&+ \sup_{\beta \in \beta^*(t) + \Theta(r)} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - \mathbb{E})\{ \Delta_i \tilde{K}_h(x_i^\top \beta^*(t) - y_i) - \Delta_i \tilde{K}_h(x_i^\top \beta - y_i) \} a_i^\top x_i \right|
\end{align*}$$

(E.24)

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

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 $n \gtrsim c_p^{5/2}(p + \log n)$, it follows from (E.22)–(E.24) with $v = 2\log n$ that, for any $t \in \tilde{T}$,

$$\sup_{s:|s - t| < h^{1/2}\eta} \left| G_n(s) - G_n(t) \right| \lesssim c_p^{1/2} \left( \frac{p + \log n}{n} \right)^{1/5}$$

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

$$\sup_{d(s,t) < \eta, t \in \tilde{T}} \left| G_n(s) - G_n(t) \right| = \sup_{|s - t| < h^{1/2}\eta, t \in \tilde{T}} \left| G_n(s) - G_n(t) \right| \lesssim c_p^{1/2} \left( \frac{p + \log n}{n} \right)^{1/5}$$

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

(E.25) \quad \limsup_{n \to \infty} \mathbb{P}\left\{ \sup_{d(s,t) < \eta, t \in \tilde{T}} \left| G_n(s) - G_n(t) \right| > x \right\} = 0, \ \text{valid for any } x > 0.

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

\[\Box\]

**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

$$\begin{align*}
D^h(\beta, \beta^*) &= \frac{1}{n} \sum_{i=1}^{n} \Delta_i \{ \tilde{K}_h(x_i^\top \beta - y_i) - \tilde{K}_h(x_i^\top \beta^* - y_i) \} (1 + e_i)x_i^\top \beta - \beta^* \\
&\geq \kappa_{l} \cdot \|\beta - \beta^*\|_\Sigma^2 \cdot \frac{1}{nh} \sum_{i=1}^{n} (1 + e_i)\omega_i \cdot \varphi_{h/(2r)}(z_i^\top v),
\end{align*}$$

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

(E.26) \quad \inf_{\beta \in \beta^* + \Theta(r)} \kappa_l \|\beta - \beta^*\|_\Sigma^2 \geq \inf_{v \in \mathbb{S}_p^{-1}} D_0(v) - \sup_{v \in \mathbb{S}_p^{-1}} \{ D_0(v) - D_0(v) \}. 
A lower bound for $\inf_{v \in S^{p-1}} D_0(v)$ can be derived from Lemma B.4. Let $\mathcal{E}_{rsc}(t)$ 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 $\mathbb{P}\{\mathcal{E}_{rsc}(t)\} \geq 1 - (m + 1)e^{-t}$. It suffices to control the bootstrap error

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

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

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

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

$$
\mathbb{E}^*(\Gamma_n) \leq \frac{1}{2r} \mathbb{E}^* \left( \sup_{\delta \in S^{p-1}} \frac{1}{n} \sum_{i=1}^n e_i \omega_i \cdot z_i^T v \right) = \frac{1}{2r} \mathbb{E}^* \left( \frac{1}{n} \sum_{i=1}^n e_i \omega_i z_i \right) \leq \frac{1}{2r} \left( \frac{1}{n} \sum_{i=1}^n \omega_i \|z_i\|^2 \right)^{1/2}.
$$

Next we deal with the data-dependent quantities $(1/n) \sum_{i=1}^n \omega_i$ and $(1/n) \sum_{i=1}^n \omega_i \|z_i\|^2$. Note that $\mathbb{E}(\omega_i | x_i) \leq \overline{g} h$ and thus $\mathbb{E}(\omega_i \|z_i\|^2) \leq \overline{g} \rho h$. Moreover, $\omega_i \|z_i\|^2 \leq \zeta_p^2$ (almost surely) and $\mathbb{E}(\omega_i^2 \|z_i\|^4) = \mathbb{E}(\omega_i \|z_i\|^2) \leq \zeta_p^4 \rho h$. By Bernstein’s inequality, together

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

and

(E.28) $\frac{1}{n} \sum_{i=1}^n \omega_i \|z_i\|^2 \leq \overline{g} \rho h + (\overline{g} \rho h)^{1/2} \zeta_p \left( \frac{2t}{n} + \zeta_p^2 \frac{t}{3n} \right) \leq \frac{3}{2} \overline{g} \rho h + \zeta_p^2 \frac{4t}{3n}$

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

(E.29) $\sup_{v \in S^{p-1}} \{D_0(v) - D_0^b(v)\} \leq \left( \frac{3}{2} \overline{g} \rho h + \zeta_p^2 \frac{4t}{3n} \right)^{1/2} \frac{1}{2r n^{1/2}}$

$$
+ \frac{h}{(4r)^2} \left\{ \left( \sqrt{\overline{g} h + \frac{t}{2n}} \right) \sqrt{\frac{2t}{n}} + \frac{4t}{3n} \right\}^{1/2}
$$

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

Define the radius $r = h/(4\eta_1/4)$ and event $\mathcal{E}_1(t) = \mathcal{E}_{rsc}(t) \cap \mathcal{E}_{loc}(t)$. Together, Lemma B.4, (E.26) and (E.29) imply that, with $\mathbb{P}^*$-probability at least $1 - e^{-t}$ conditioned on $\mathcal{E}_1(t)$,

$$
D^b(\beta, \beta^*_j) \geq \frac{1}{2} \zeta_p \left\| \beta - \beta^*_j \right\|_2^2
$$

holds uniformly over $\beta \in \mathbb{B}_j + \Theta(r)$ as long as $nh \geq \max(p, \zeta_p t^{1/2})$. Taking the union bound over $j = 0, 1, \ldots, m$ concludes the proof. \qed
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 $|\mathcal{N}| \leq 5^p$ such that

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} (e_i \cdot \xi_i z_i) \right\|_2 \leq 2 \max_{u \in \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} e_i \cdot \xi_i \langle u, z_i \rangle.
$$

Since $e_i$'s are independent Rademacher random variables and $|\xi_i| \leq M$, conditional on the data $\{x_i, \xi_i\}_{i=1}^{n}$, 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 $1 - e^{-u}$,

$$
\frac{1}{n} \sum_{i=1}^{n} e_i \cdot \xi_i \langle u, z_i \rangle \leq M \lambda_{\max}^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T \right) \sqrt{\frac{2u}{n}},
$$

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

$$
\lambda_{\max}^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T \right) \leq 1 + C \zeta_p \sqrt{\frac{\log(2p)}{n}}.
$$

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

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

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

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

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 $v = \Sigma^{1/2}(\beta - \beta_j^*) \in \mathbb{B}^p(r)$ for $\beta \in \beta_j^* + \Theta(r)$, we write

(E.32)

$$
\Gamma_j^2(r) = \sup_{v \in \mathbb{B}^p(r)} \|G_j(v)\|_2 := \sup_{v \in \mathbb{B}^p(r)} \left\| \frac{1}{n} \sum_{i=1}^{n} e_i \cdot \{ \bar{K}_h(z_i^T v - \varepsilon_{ij}) - \bar{K}_h(-\varepsilon_{ij}) \} z_i \right\|_2,
$$

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

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

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

Note that $\phi_{ij,v} \leq \kappa_u/h$, and by Hölder’s inequality,

$$
\frac{1}{n} \sum_{i=1}^{n} \phi_{ij,v}^2 (z_i^T u \cdot z_i^T w)^2 \leq \frac{\kappa_u}{h} \cdot \left\{ \frac{1}{n} \sum_{i=1}^{n} \phi_{ij,v}^2 (z_i^T u)^4 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \phi_{ij,v}^2 (z_i^T w)^4 \right\}^{1/2}.
$$
Given $r > 0$, define the supremum

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

Under this notation,

$$
\sup_{v \in \mathbb{S}^{p-1}} \log \mathbb{E}^* \exp \left\{ \lambda n^{1/2} u^T \nabla G_j(v) w \right\} \leq \frac{K_u}{2h} \Psi(r)^2.
$$

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

$$
\text{(E.33)} \quad \sup_{v \in \mathbb{S}^{p-1}} \|G_j(v)\|_2 \leq \kappa_1^{1/2} \Psi(r)^{1/2} \cdot \frac{\sqrt{p + \log n}}{nh}.
$$

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

It remains to control the data-dependent quantity $\Psi(r) = \sup_{u, v \in \mathbb{S}^{p-1}} \psi_{u, v}(r)$. For any $\epsilon_1, \epsilon_2 \in (0, 1)$, let $\{u_1, \ldots, u_{d_1}\}$ and $\{v_1, \ldots, v_{d_2}\}$ be the $\epsilon_1$- and $\epsilon_2$-nets of $\mathbb{S}^{p-1}$ with cardinalities $d_1 \leq (1 + 2/\epsilon_1)^p$ and $d_2 \leq (1 + 2/\epsilon_2)^p$. Given $u, v \in \mathbb{S}^{p-1}$, there exist some $1 \leq \ell \leq d_1$ and $1 \leq k \leq d_2$ such that $\|u - u_\ell\|_2 \leq \epsilon_1$ and $\|v - v_k\|_2 \leq \epsilon_2$. Consider the decomposition

$$
\text{(E.34)} \quad \psi_{u, v}(r) = \psi_{u, v}(r) - \psi_{u, v_k}(r) + \psi_{u, v_k}(r).
$$

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

$$
\text{(E.35)} \quad |\psi_{u, v}(r) - \psi_{u, v_k}(r)| \leq \frac{L K^p}{nh^2} \sum_{i=1}^{n} |z_i^T (v - v_k)(z_i^T u)|^4 \leq \frac{L K^p}{nh^2} \sum_{i=1}^{n} |z_i^T u|^4 \cdot r \epsilon_2.
$$

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

$$
\psi_{u, v_k}(r)^{1/4} = \left\{ \frac{1}{n} \sum_{i=1}^{n} K_h(r z_i^T v_k - \varepsilon_{ij})(z_i^T u)^4 \right\}^{1/4} \leq \left\{ \frac{1}{n} \sum_{i=1}^{n} K_h(r z_i^T v - \varepsilon_{ij})(z_i^T u)^4 \right\}^{1/4} + \left\{ \frac{1}{n} \sum_{i=1}^{n} K_h(r z_i^T v_k - \varepsilon_{ij})(z_i^T u - u_\ell)^4 \right\}^{1/4}
$$

$$
\leq \psi_{u, v_k}(r) + \epsilon_1 \cdot \sup_{u \in \mathbb{S}^{p-1}} \psi_{u, v_k}(r)^{1/4}.
$$

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

$$
\text{(E.36)} \quad \sup_{u \in \mathbb{S}^{p-1}} \psi_{u, v_k}(r) \leq (1 - \epsilon_1)^{-4} \max_{1 \leq \ell \leq d_1} \psi_{u, v_k}(r).
$$

The problem then boils down to controlling the maximum $\max_{(\ell, k) \in [d_1] \times [d_2]} \psi_{u, v_k}(r)$. Note that $\mathbb{E} K_h^2(r z_i^T v_k - \varepsilon_{ij})(z_i^T u)^6 \leq \mathbb{E} K_h(r z_i^T v_k - \varepsilon_{ij})(z_i^T u)^4 \leq \kappa_u \zeta_p^2 / h$. By Bernstein's inequality, we have that with probability at least $1 - e^{-u}$,

$$
\psi_{u, v_k}(r) \leq \mathbb{E} \psi_{u, v_k}(r) + (\mathbb{E} K_h^2)^{1/2} \zeta_p^2 h^{1/2} \sqrt{\frac{2u}{nh} + \kappa_u \zeta_p^2 \frac{u}{3nh}}.
$$
\[ \leq \mathcal{I} m_4 + (\mathcal{F} \kappa_m m_4)^{1/2} \zeta_p \sqrt{\frac{2u}{nh} + \kappa_u \zeta_p^4 n} \]

\[ \leq \frac{3}{2} \mathcal{I} m_4 + \kappa_u \zeta_p^4 \frac{4u}{3nh}. \]

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 \( 1 - n^{-1} \),

\[ \Psi(r) \lesssim m_4 + \zeta_p^4 (p \log n)^{1/2} \frac{n}{nh} \]

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

\[ \Lambda_j^\beta = \hat{Q}_j^\beta - \hat{Q}_j^\beta_0 - \mathbf{J}_j (\beta - \beta^*_j), \quad \beta \in \mathbb{R}^p. \]

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

\[ \Lambda_j^\beta = \frac{1}{n} \sum_{i=1}^n \Delta_i \{ \hat{K}_h(x_i^\beta - y_i) - \hat{K}_h(x_i^{\beta^*_j} - y_i) \} x_i - \mathbf{J}_j (\beta - \beta^*_j). \]

Consequently, by the triangle inequality,

\[ \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta \|_{\Sigma^{-1}} \leq \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta - \Lambda_j^\beta \|_{\Sigma^{-1}} + \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta \|_{\Sigma^{-1}}. \]

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 \( \mathcal{E}_3 \),

\[ \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta - \Lambda_j^\beta \|_2 \lesssim \left( m_4^{1/2} + \zeta_p^2 \sqrt{\frac{p \log n}{nh}} \right) \cdot r \sqrt{\frac{p + \log n}{nh}}, \]

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 \( \mathcal{E}_3' \),

\[ \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta \|_{\Sigma^{-1}} \lesssim \left( m_4^{1/2} \sqrt{\frac{p + \log n}{nh}} + m_3r + h \right) \cdot r, \]

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 \( \mathcal{E}_4 \),

\[ \sup_{\beta \in \beta_j + \Theta(r)} \| \Lambda_j^\beta \|_2 \lesssim \left( m_4^{1/2} \sqrt{\frac{p + \log n}{nh}} + m_3r + h + \zeta_p \left( \frac{p \log n}{nh} \right)^{1/2} \left( \frac{p + \log n}{nh} \right)^{1/2} \right) \cdot r, \]

completing the proof. \( \square \)
E.13. Proof of Lemma C.5. For \( j = 1, \ldots, m \), define the random processes 
\[
R_j^2(\beta_0, \beta_1, \ldots, \beta_{j-1}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=0}^{\tau_{i+1}} dH(u) \left[ W_i \left\{ \tilde{K}_h(y_i - x_i^T \beta_\ell^*) - \tilde{K}_h(y_i - x_i^T \beta_\ell) \right\} z_i - \Sigma^{-1/2} \mathbf{H}_\ell(\beta_\ell - \beta_\ell^*) \right] 
\]
and
\[
R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=0}^{\tau_{i+1}} dH(u) \left\{ \tilde{K}_h(y_i - x_i^T \beta_\ell^*) - \tilde{K}_h(y_i - x_i^T \beta_\ell) \right\} z_i - \Sigma^{-1/2} \mathbf{H}_\ell(\beta_\ell - \beta_\ell^*) ,
\]
and note that \( \mathbb{E}^* R_j^2(\beta_0, \beta_1, \ldots, \beta_{j-1}) = R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) \). By the triangle inequality,
\[
\sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \| R_j^2(\beta_0, \beta_1, \ldots, \beta_{j-1}) \|_2 \leq \sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \| R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) \|_2 + \sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \| R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) - R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) \|_2 .
\]
Let \( \mathcal{E}_5 \) be the event that (B.5) holds for all \( \ell = 0, \ldots, m-1 \) with \( t = 2 \log n \). Then, \( \mathbb{P}(\mathcal{E}_5) \geq 1 - m n^{-2} \), and conditional on \( \mathcal{E}_5 \),
\[
\sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \| R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) \|_2 \leq \log \left( \frac{1 - n^{-\gamma}}{1 - \gamma} \right) \cdot \left( m_4^{1/2} \sqrt{\frac{p + \log n}{n h}} + m_3 r + h \right) \cdot r
\]
holds for all \( j = 0, 1, \ldots, m \) provided that \( n h \gtrsim \gamma^2 (p + t) \). For the second term on the right-hand side of (E.40), note that
\[
\sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \| R_j^2(\beta_0, \beta_1, \ldots, \beta_{j-1}) - R_j(\beta_0, \beta_1, \ldots, \beta_{j-1}) \|_2 \leq \sum_{\ell=0}^{\tau_i} \sup_{\beta \in \beta_i + \Theta(r)} \left\| \frac{1}{n} \sum_{i=1}^{n} c_i \left\{ \tilde{K}_h(y_i - x_i^T \beta_\ell^*) - \tilde{K}_h(y_i - x_i^T \beta_\ell) \right\} z_i \right\|_2 \cdot \log \left( \frac{1 - n^{-\gamma}}{1 - \gamma} \right)
\]
(E.42)
\[
= \sum_{\ell=0}^{\tau_i} \Gamma_\ell(r) \cdot \log \left( \frac{1 - n^{-\gamma}}{1 - \gamma} \right) \leq \max_{\ell=0, \ldots, j-1} \Gamma_\ell(r) \cdot \log \left( \frac{1 - n^{-\gamma}}{1 - \gamma} \right),
\]
where \( \{ \Gamma_\ell(r) \}_{\ell=0}^{j-1} \) are defined in (C.2).
Combining (E.40)–(E.42) proves the claimed result. \( \square \)

E.14. Proof of Lemma C.6. We first show an unconditional version
\[
\lim_{\delta \to 0} \lim_{n \to \infty} \max_{|\tau_1 - \tau_2| < \delta} \sup_{|\tau_2 - \tau_1| < \delta} \mathbb{P} \left\{ \sup_{|\tau_2 - \tau_1| < \delta} \left| \mathbb{G}^b_n(\tau_1) - \mathbb{G}^b_n(\tau_2) \right| > x \right\} = 0
\]
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
\[
\left| \mathbb{G}^b_n(\tau_1) - \mathbb{G}^b_n(\tau_2) \right|_{\psi} \lesssim (m_4/h)^{1/2} |\tau_2 - \tau_1|,
\]
for any \( \tau_1 < \tau_2 \) satisfying \( \tau_2 - \tau_1 \geq (\zeta_p / m_k)\sqrt{(h/n)} \), 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 \( \{ D_n \} \), if (C.4) does not hold, then there is a sequence \( \{ \delta_m : m \in \mathbb{N}^+ \} \) such that \( \lim_{m \to \infty} \delta_m = 0 \), and a subsequence of natural numbers \( \{ n_k : k \in \mathbb{N}^+ \} \subset \mathbb{N}^+ \) such that

\[
\lim_{m \to \infty} \lim_{k \to \infty} \mathbb{P}\left\{ \sup_{|\tau_1 - \tau_2| < \delta_m} \left| G_{n_k}^b(\tau_1) - G_{n_k}^b(\tau_2) \right| > x \right\} > 0 \text{ over } \mathcal{A},
\]

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,

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

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

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

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

\[\square\]

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 \( x_i \) \( (i = 1, \ldots, n) \) from \( \mathcal{N}(0, \Sigma) = (s_{jk})_{1 \leq j, k \leq 5} \), where \( s_{jk} = 0.5|j - k| \). The remaining settings are the same as in Section 5.1. The results on the \( \ell_2 \)-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 \( \ell_2 \)-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 \( \tau \) 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 \( \lambda_k = \{1 + \log\left(\frac{1}{1 - \tau_k}\right)\} \lambda_0 \) for \( k = 1, \ldots, m \) and choosing \( \lambda_0 \) 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 \( \lambda \) values, i.e. \( \lambda_k = \lambda_0 + c \cdot k \) for \( k = 0, 1, \ldots, m \), and used cross-validation to choose \( \lambda_0 \) and \( c \) by a two-dimensional grid search. Since this tuning scheme is not adopted, the selection measures are presented for reference only.
**SUPPLEMENT**

(a) $\ell_2$-error under model (31)  
(b) Estimated quantile effects under model (31)  
(c) Runtime under model (31)  
(d) $\ell_2$-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**


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.


FIG F.3. Plots of $\ell_2$-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 $p = n/100$.

FIG F.4. Box plots of false discovery rate (FDR) for $\ell_1$-penalized CQR (CQR-Lasso) and cross-validated $\ell_1$-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.