Appendix C: Proofs of some supplemental results

In this appendix, we prove some technical results which are crucial to proofs of the main theoretical results in Appendix B. Recall that $S(x,z)$ is a $(p+1) \times (p+1)$ diagonal matrix defined as

$$S(x,z) = \text{diag} \{ f_{x \zeta}(x,z)f_\varepsilon(0|x,z), \mu_2 f_{x \zeta}(x,z)f_\varepsilon(0|x,z) \cdot I_p \}$$

with $I_p$ being a $p \times p$ identity matrix, and define

$$W_{h,\lambda}(x,z) = [W_{h,\lambda,0}(x,z), W_{h,\lambda,1}(x,z), \ldots, W_{h,\lambda,p}(x,z)]^T$$

with

$$W_{h,\lambda,0}(x,z) = \frac{1}{n} \sum_{i=1}^n \eta_i(x,z)K_h(X_i - x)\Lambda_\lambda(Z_i,z),$$

$$W_{h,\lambda,j}(x,z) = \frac{1}{n} \sum_{i=1}^n \eta_i(x,z) \left( \frac{X_{ij} - x_j}{h_j} \right) K_h(X_i - x)\Lambda_\lambda(Z_i,z), \quad j = 1, \ldots, p,$$

$\eta_i(x,z) = \tau - I[\varepsilon_i \leq -b_1(x,z)]$, where $b_1(x,z) = Q_\tau(X_i, Z_i) - Q_\tau(x,z) - \sum_{j=1}^p Q_{\tau,j}(x,z) |X_{ij} - x_j|$. Proposition C.1 below gives the Bahadur representation for the local linear quantile estimator uniformly over $x \in \mathcal{D}_x(\varepsilon)$, $z \in \mathcal{D}_z$ and $(h,\lambda) \in \mathcal{K}_n$, where $\mathcal{D}_x(\varepsilon)$ is defined in Assumption 4, $\mathcal{D}_z$ is the support of $Z_i$, and $\mathcal{K}_n$ is defined in the proof of Theorem 3.2. It complements the Bahadur representation results developed in the literature for the setting of purely continuous regressors (c.f., Kong, Linton and Xia, 2010; Guerre and Sabbah, 2012; Su and White, 2012; Kong and Xia, 2017).

**Proposition C.1.** Suppose that Assumptions 1–5 are satisfied, and let

$$\zeta_{h,\lambda}(x,z) = [\zeta_{h,\lambda,0}(x,z), \zeta_{h,\lambda,1}(x,z), \ldots, \zeta_{h,\lambda,p}(x,z)]^T$$

with

$$\zeta_{h,\lambda,0}(x,z) = \tilde{Q}_\tau(x,z; h, \lambda) - Q_\tau(x,z), \quad \zeta_{h,\lambda,j}(x,z) = h_j \tilde{Q}'_{\tau,j}(x,z; h, \lambda) - h_j Q'_{\tau,j}(x,z), \quad j = 1, \ldots, p.$$
Then we have
\[
\hat{\zeta}_{h,\lambda}(x, z) = S^{-1}(x, z)W_{h,\lambda}(x, z) + Op\left(\chi_1^{3/2}(h) + \chi_1(h)\chi_2^{1/2}(h, \lambda)\right)
\]  
(S.1)
uniformly over \(x \in D_x(\epsilon), z \in D_z\) and \((h, \lambda) \in \mathcal{H}_n\), where
\[
\chi_1(h) = (\log n)^{1/2}(nH)^{-1/2}, \quad \chi_2(h, \lambda) = \sum_{j=1}^{p} h_j^2 + \sum_{j=1}^{q} \lambda_j.
\]

The main idea to prove Proposition C.1 is similar to that in the proofs of Theorem 3.1 in Su and White (2012) and Lemma A.1 in Chen et al (2019). We first prove two technical lemmas.

**Lemma C.1.** Suppose that the assumptions in Proposition C.1 are satisfied. Then, we have
\[
\|W_{h,\lambda}(x, z)\| = Op\left(\chi_1(h) + \chi_2(h, \lambda)\right)
\]  
(S.2)
uniformly over \(x \in D_x(\epsilon), z \in D_z\) and \((h, \lambda) \in \mathcal{H}_n\), where \(\chi_1(h)\) and \(\chi_2(h, \lambda)\) are defined in Proposition C.1.

**Proof of Lemma C.1.** For brevity, we only prove that
\[
|W_{h,\lambda,0}(x, z)| = Op\left(\chi_1(h) + \chi_2(h, \lambda)\right)
\]  
(S.3)
uniformly over \(x \in D_x(\epsilon), z \in D_z\) and \((h, \lambda) \in \mathcal{H}_n\), as the proof for \(W_{h,\lambda,j}(x, z), j = 1, \ldots, p\) is exactly the same (with some notational modification). Noting that the support for the discrete covariates only contains a finite number of distinct integers, to prove (S.3), it is sufficient to show that it holds uniformly over \(x \in D_x(\epsilon)\) and \((h, \lambda) \in \mathcal{H}_n\) for each fixed \(z \in D_z\).

By Assumptions 1, 3 and 4, we may prove that
\[
E[|W_{h,\lambda,0}(x, z)|] = E[K_h|X_i - x|\Lambda(z_i, z)]E[|\eta_1(x, z)||X_i, Z_i] = E[K_h|X_i - x|\Lambda(z_i, z)]E[|\eta_1(x, z)||X_i, Z_i]]
\]
\[
= O\left(\sum_{j=1}^{p} h_j^2 + \sum_{j=1}^{q} \lambda_j\right) = O\left(\chi_2(h, \lambda)\right).
\]  
(S.4)

By (S.4), we only need to prove that
\[
|W_{h,\lambda,0}(x, z) - E[W_{h,\lambda,0}(x, z)| = Op\left(\chi_1(h)\right)
\]  
(S.5)
uniformly over \(x \in D_x(\epsilon)\) and \((h, \lambda) \in \mathcal{H}_n\).

We use the argument in the proof of Theorem 3.2 to prove (S.5). Consider covering the bounded set \(D_x(\epsilon)\) by some disjoint sets \(D(k), k = 1, \ldots, K\). Let the center points of \(D(k)\) be \(x(k)\), and let the size of \(D(k)\)
guarantee that
\[
\max_{1 \leq k_1 \leq K} \max_{1 \leq k_2 \leq L_n} \sup_{x \in D(k_1)} \left( nH(k_2) \right)^{1/2} \left[ W_{k_2,0}(x,z) - W_{k_2,0}(x(k_1),z) \right] = O_p \left( (\log n)^{1/2} \right), \tag{S.6}
\]
and
\[
\max_{1 \leq k_1 \leq K} \max_{1 \leq k_2 \leq L_n} \sup_{x \in D(k_1)} \left( nH(k_2) \right)^{1/2} \mathbb{E} \left[ W_{k_2,0}(x,z) - W_{k_2,0}(x(k_1),z) \right] = O \left( (\log n)^{1/2} \right), \tag{S.7}
\]
where \( L_n \) is the number of grid points in \( \mathcal{H}_n \), \( W_{k_2,0}(x,z) = W_h(k_2,\lambda(k_2,0)(x,z), H(k) = \prod_{j=1}^{p} h_j(k) \) with \( h_j(k) \) being the \( j \)-th component of the vector \( h(k) \). The smoothness conditions in Assumptions 1, 3 and 4 are needed to prove \( (S.6) \) and \( (S.7) \).

By the Bonferroni inequality and the Bernstein inequality for independent random variables (e.g., Lemma 2.2.9 in van der Vaart and Wellner, 1996) and noting that the numbers \( K \) and \( L_n \) diverge to infinity at a polynomial rate of \( n \), we may prove that
\[
P \left( \max_{1 \leq k_1 \leq K} \max_{1 \leq k_2 \leq L_n} \left( nH(k_2) \right)^{1/2} \left[ W_{k_2,0}(x(k_1),z) - \mathbb{E} \left[ W_{k_2,0}(x(k_1),z) \right] \right] > C_1 (\log n)^{1/2} \right) \\
\leq \sum_{k_1=1}^{K} \sum_{k_2=1}^{L_n} P \left( \left( nH(k_2) \right)^{1/2} \left[ W_{k_2,0}(x(k_1),z) - \mathbb{E} \left[ W_{k_2,0}(x(k_1),z) \right] \right] > C_1 (\log n)^{1/2} \right) \\
\leq O \left( K \cdot L_n \cdot \exp \left\{ -C_1 \log n \right\} \right) = o(1),
\]
where \( C_1 \) is a positive constant which becomes sufficiently large when \( C_1 \) is large enough. Then, we have
\[
\max_{1 \leq k_1 \leq K} \max_{1 \leq k_2 \leq L_n} \left( nH(k_2) \right)^{1/2} \left[ W_{k_2,0}(x(k_1),z) - \mathbb{E} \left[ W_{k_2,0}(x(k_1),z) \right] \right] = O_p \left( (\log n)^{1/2} \right). \tag{S.8}
\]
With \( (S.6)-(S.8) \), we can prove \( (S.5) \) in a standard way.

Denote
\[
\tilde{\Theta}_{r,h}(x,z) = \left[ Q_r(x,z), h_1 Q'_{r,1}(x,z), \ldots, h_p Q'_{r,p}(x,z) \right]^T, \\
\Delta(\alpha, \beta) \equiv \Delta_{r,h}(\alpha, \beta, x, z) = (nH)^{1/2} \left[ \alpha - Q_{r}(x,z), h_1 \beta_1 - h_1 Q'_{r,1}(x,z), \ldots, h_p \beta_p - h_p Q'_{r,p}(x,z) \right]^T, \\
\bar{\lambda}_{i,h}(x) = \left[ 1, (X_{i,1} - x_1)/h_1, \ldots, (X_{i,p} - x_p)/h_p \right]^T, \\
\eta_{l,h}(x,z,\Delta) = \tau - 1 \left\{ Y_l - \left[ \tilde{\Theta}_{r,h}(x,z) + (nH)^{-1/2} \Delta(\alpha, \beta) \right]^T \bar{\lambda}_{i,h}(x) < 0 \right\},
\]
where \( \Delta \equiv \Delta(\alpha, \beta) \) for further simplification. Let \( W_{h,\lambda}(x,z,\Delta) \) be defined as \( W_{h,\lambda}(x,z) \) but with \( \eta_{l,h}(x,z) \) replaced by \( \eta_{l,h}(x,z,\Delta) \). It is easy to show that when \( \Delta = \Delta(\alpha, \beta) = 0_{p+1}, \eta_{l,h}(x,z,\Delta) \) and \( W_{h,\lambda}(x,z,\Delta) \) are reduced to \( \eta_{l,h}(x,z) \) and \( W_{h,\lambda}(x,z) \), respectively, where \( 0_{p+1} \) is a \((p+1)\)-dimensional null vector.

**Lemma C.2.** Suppose that the assumptions in Proposition C.1 are satisfied. Letting \( \tilde{W}_{h,\lambda}(x,z,\Delta) = \)
With the Bonferroni and Bernstein inequalities, using Assumption 5(ii), we may show that
\[
\|\tilde{W}_{h,\lambda}(x, z, \Delta) - \mathbb{E}[\tilde{W}_{h,\lambda}(x, z, \Delta)]\| = O_P \left( \chi_1^{3/2}(h) + \chi_1(h)\chi_2^{1/2}(h, \lambda) \right)
\]
uniformly over \( x \in D_x(e), z \in D_z, \|\Delta\| \leq \chi(h, \lambda, C_2) \) and \((h, \lambda) \in \mathcal{H}_n\), where
\[
\chi(h, \lambda, C_2) = C_2(nH)^{1/2} (\chi_1(h) + \chi_2(h, \lambda))
\]
with \( C_2 \) being a sufficiently large positive constant.

**Proof of Lemma C.2.** As in the proof of Lemma C.1, we only prove the uniform convergence for \( \tilde{W}_{h,\lambda,0}(x, z, \Delta) \equiv W_{h,\lambda,0}(x, z, \Delta) - W_{h,\lambda,0}(x, z) \), where \( W_{h,\lambda,0}(x, z, \Delta) \) is the first component of the vector \( W_{h,\lambda}(x, z, \Delta) \). Specifically, we aim to show that, for each fixed \( z \in D_z \),
\[
\|\tilde{W}_{h,\lambda,0}(x, z, \Delta) - \mathbb{E}[\tilde{W}_{h,\lambda,0}(x, z, \Delta)]\| = O_P (\tilde{\chi}(h, \lambda))
\]
uniformly over \( x \in D_x(e), \|\Delta\| \leq \chi(h, \lambda, C_2) \) and \((h, \lambda) \in \mathcal{H}_n\), where \( \tilde{\chi}(h, \lambda) = \chi_1^{3/2}(h) + \chi_1(h)\chi_2^{1/2}(h, \lambda) \).

As in the proof of Lemma C.1, for each \( k_1 = 1, \ldots, L_n \), we consider respectively covering the sets \( D_x(e) \) and \( \{ \Delta : \|\Delta\| \leq \chi(k_1, C_2) \} \) by \( \hat{D}_x(k_2) \), \( k_2 = 1, \ldots, \hat{k}_2 \), and \( \hat{j}_{k_3} \), \( k_3 = 1, \ldots, \hat{k}_3 \), where \( \chi(k, C_2) = \chi(h(k), \lambda(k), C_2) \). Let the center points of \( \hat{D}_x(k_2) \) and \( \hat{j}_{k_3} \) be denoted by \( x(k_2) \) and \( \Delta_{k_3} \), respectively. Furthermore, the sizes of \( \hat{D}_x(k_2) \) and \( \hat{j}_{k_3} \) (which are chosen to be sufficiently small) as well as the smoothness conditions in Assumptions 1, 3 and 4 guarantee that
\[
\max_{1 \leq k_1 \leq L_n} \sup_{x \in D_x(e)} \sup_{\|\Delta\| \leq \chi(k_1, C_2)} \tilde{\chi}^{-1}(k_1) \left| \tilde{W}_{k_1,0}(x, z, \Delta) - \mathbb{E}[\tilde{W}_{k_1,0}(x, z, \Delta)] \right| 
\]
\[
\leq \max_{1 \leq k_1 \leq L_n} \max_{1 \leq k_2 \leq \hat{k}_2} \max_{1 \leq k_3 \leq \hat{k}_3} \tilde{W}(k_1, k_2, k_3) + O_P(1),
\]
where \( \tilde{\chi}(k) = \chi(h(k), \lambda(k)) \), \( \tilde{W}(k_1, k_2, k_3) = \tilde{W}(h(k_1), h(k_2), h(k_3), 0)(x, z, \Delta) \) and
\[
\tilde{W}(k_1, k_2, k_3) = \tilde{\chi}^{-1}(k_1) \left| \tilde{W}_{k_1,0}(x(k_2), z, \Delta_{k_3} - \mathbb{E}[\tilde{W}_{k_1,0}(x(k_2), z, \Delta_{k_3})] \right|.
\]

With the Bonferroni and Bernstein inequalities, using Assumption 5(ii), we may show that
\[
P \left( \max_{1 \leq k_1 \leq L_n} \max_{1 \leq k_2 \leq \hat{k}_2} \max_{1 \leq k_3 \leq \hat{k}_3} \tilde{W}(k_1, k_2, k_3) \geq C_3 \right)
\]
\[
\leq \sum_{k_1=1}^{L_n} \sum_{k_2=1}^{\hat{k}_2} \sum_{k_3=1}^{\hat{k}_3} P \left( \left| \tilde{W}_{k_1,0}(x(k_2), z, \Delta_{k_3}) - \mathbb{E}[\tilde{W}_{k_1,0}(x(k_2), z, \Delta_{k_3})] \right| \geq C_3 \tilde{\chi}(k_1) \right)
\]
\[
\leq O \left( L_n \cdot \hat{k}_2 \cdot \hat{k}_3 \cdot \exp \left( -\tilde{C}_3 \log n \right) \right) = o(1),
\]
where \( \tilde{C}_3 \) is a positive constant which would be sufficiently large if \( C_3 > 0 \) is large enough. As a result, we
have
\[
\max_{1 \leq k_1 \leq L_n} \max_{1 \leq k_2 \leq K_2} \max_{1 \leq k_3 \leq K_3} \tilde{W}(k_1, k_2, k_3) = O_p(1). \tag{S.12}
\]

By (S.11) and (S.12), we complete the proof of (S.10).

With the help of Lemmas C.1 and C.2 above, we can provide a detailed proof of Proposition C.1.

**Proof of Proposition C.1.** As in Lemma C.2, we let \( \chi(h, \lambda, C_2) = C_2(nH)^{1/2} (\chi_1(h) + \chi_2(h, \lambda)) \). By Assumptions 1, 3 and 4, and the definition of \( \tilde{W}_{h,\lambda}(x, z, \Delta) \) in Lemma C.2, we may show that
\[
E \left[ \tilde{W}_{h,\lambda}(x, z, \Delta) - W_{h,\lambda}(x, z) \right] = \frac{1}{n} \sum_{i=1}^{n} E \left[ \eta_i h(x_i, z_i, \Delta) \eta_{i,h}(x_i, z_i) \right] = -\frac{(nH)^{-1/2}}{S(x, z)\Delta} + O \left( \| \Delta \| (nH)^{-1/2} + \| \Delta \|^2 (nH)^{-1} \right) \tag{S.13}
\]
uniformly over \( x \in \mathcal{D}_x(e), z \in \mathcal{D}_z, \| \Delta \| \leq \chi(h, \lambda, C_2) \) and \( (h, \lambda) \in \mathcal{H}_n \). Meanwhile, by Assumption 5(ii), if \( \| \Delta \| \leq \chi(h, \lambda, C_2) = C_2(nH)^{1/2} (\chi_1(h) + \chi_2(h, \lambda)) \), we have
\[
\chi_2(h, \lambda)\| \Delta \|(nH)^{-1/2} = O \left( \chi_1(h)\chi_2(h, \lambda) + \chi_2^2(h, \lambda) \right) = o \left( \chi_1(h)\chi_2^2(h, \lambda) \right) \tag{S.14}
\]
and
\[
\| \Delta \|^2(nH)^{-1} = O \left( \chi_1^2(h) + \chi_2^2(h, \lambda) \right) = o \left( \chi_1^2(h) + \chi_1(h)\chi_2^2(h, \lambda) \right). \tag{S.15}
\]
Combining (S.13)–(S.15), we readily have that
\[
E \left[ \tilde{W}_{h,\lambda}(x, z, \Delta) \right] = -\frac{(nH)^{-1/2}}{S(x, z)\Delta} + O \left( \chi_1^2(h) + \chi_1(h)\chi_2^2(h, \lambda) \right), \tag{S.16}
\]
which together with Lemma C.2, leads to
\[
\left\| W_{h,\lambda}(x, z, \Delta) - W_{h,\lambda}(x, z) \right\| = O_p \left( \chi_1^2(h) + \chi_1(h)\chi_2^2(h, \lambda) \right) \tag{S.17}
\]
uniformly over \( x \in \mathcal{D}_x(e), z \in \mathcal{D}_z, \| \Delta \| \leq \chi(h, \lambda, C_2) \) and \( (h, \lambda) \in \mathcal{H}_n \).

Using (S.17) and following the arguments in the proofs of Lemma A2 in Ruppert and Carroll (1980) and Lemma A.5 in Su and White (2012), to complete the proof of Proposition C.1, we only need to show that, with probability approaching one,
\[
\left\| \tilde{\Delta}_{h,\lambda}(x, z) \right\| \leq \chi(h, \lambda, C_2) \tag{S.18}
\]
uniformly over \( x \in \mathcal{D}_x(e), z \in \mathcal{D}_z \) and \( (h, \lambda) \in \mathcal{H}_n \), when \( C_2 > 0 \) is chosen to be sufficiently large, where
\[
\tilde{\Delta}_{h,\lambda}(x, z) = (nH)^{1/2} \left[ \tilde{O}(x, z) - Q(x, z), h_1 \tilde{Q}_t(x, z) - h_1 Q_t(x, z), \ldots, h_p \tilde{Q}_{t,p}(x, z) - h_p Q_{t,p}(x, z) \right]^T.
\]
Let $\mathcal{E}_{n1}$ denote the event that
\[
\inf_{\|\Delta\| = \chi(h, \lambda, C_2)} \left\{ -\Delta^\top W_{h,\lambda}(x, z, \Delta) \right\} < C_2 C_4 (nH)^{1/2} [\chi(h) + \chi_2(h, \lambda)]^2
\]
uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$, and let $\mathcal{E}_{n2}$ denote the event that
\[
\inf_{\|\Delta\| = \chi(h, \lambda, C_2)} \left\{ \Delta^\top S(x, z) \Delta - (nH)^{1/2} \Delta^\top W_{h,\lambda}(x, z) \right\} \geq 2 C_2 C_4 [\chi(h, \lambda, C_2)/C_2]^2
\]
uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$.

By Assumption 3, there exists a positive constant $C$ such that the diagonal elements of $S(x, z)$ are larger than $C$ uniformly over $x \in \mathcal{D}_x(e)$ and $z \in \mathcal{D}_z$. Then, we can prove that
\[
\inf_{\|\Delta\| = \chi(h, \lambda, C_2)} \left\{ \Delta^\top S(x, z) \Delta - (nH)^{1/2} \Delta^\top W_{h,\lambda}(x, z) \right\} \geq 2 C_2 C_4 [\chi(h, \lambda, C_2)/C_2]^2
\]
uniformly over $x \in \mathcal{D}_x(e)$ and $z \in \mathcal{D}_z$. If $\mathcal{E}^c_{n2}$ (the complement of $\mathcal{E}_{n2}$) holds, from (S.19), we must have
\[
\chi(h, \lambda, C_2)(nH)^{1/2} \|W_{h,\lambda}(x, z)\| \geq (C C_2 C_4) [\chi(h, \lambda, C_2)/C_2]^2,
\]
or equivalently
\[
\|W_{h,\lambda}(x, z)\| \geq (C C_2 C_4) [\chi(h) + \chi_2(h, \lambda)]
\]
uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$, which, by Lemma C.1, holds with probability tending to zero (by choosing $C_2 > 0$ to be sufficiently large). The above arguments show that
\[
P(\mathcal{E}_{n2}^c) \to 0 \text{ as } n \to \infty,
\]
when the positive constant $C_2$ is large enough.

We want to prove
\[
P(\mathcal{E}_{n1}) \to 0 \text{ as } n \to \infty.
\]
By (S.20), it is sufficient to show
\[
P(\mathcal{E}_{n1} \cap \mathcal{E}_{n2}) \to 0 \text{ as } n \to \infty.
\]
Note that when the events $\mathcal{E}_{n1}$ and $\mathcal{E}_{n2}$ hold jointly, we must have that
\[
\sup_{\|\Delta\| = \chi(h, \lambda, C_2)} \left\{ (nH)^{1/2} \left[ \Delta^\top W_{h,\lambda}(x, z, \Delta) - \Delta^\top W_{h,\lambda}(x, z) \right] + \Delta^\top S(x, z) \Delta \right\} \geq C_2 C_4 [\chi(h, \lambda, C_2)/C_2]^2
\]
uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$. Hence, by (S.21), $\mathcal{E}_{n1}$ and $\mathcal{E}_{n2}$ hold jointly with probability approaching zero by choosing $C_2 > 0$ to be sufficiently large. This completes the proofs of (S.21) and (S.22).

Next, we consider a more general case of $\|\Delta\| \geq \chi(h, \lambda, C_2)$. It is easy to check that $-\Delta_1^T W_{h,\lambda}(x, z, \varpi \Delta)$ is non-decreasing for $\varpi \geq 1$ when it is treated as a function of $\varpi$ (with the components $h, \lambda, x$ and $\Delta$ fixed). For $\|\Delta\| \geq \chi(h, \lambda, C_2)$, we define

$$\varpi_s = \frac{\|\Delta\|}{\chi(h, \lambda, C_2)}, \quad \Delta_s = \varpi_s^{-1} \Delta.$$ 

It is straightforward to show that $\|\Delta_s\| = \chi(h, \lambda, C_2)$ and

$$-\Delta_1^T W_{h,\lambda}(x, z, \Delta) = -\Delta_1^T W_{h,\lambda}(x, z, \varpi_s \Delta_s) \geq -\Delta_1^T W_{h,\lambda}(x, z, \Delta_s).$$

Then, using the Cauchy-Schwarz inequality, we have

$$\|W_{h,\lambda}(x, z, \Delta)\| \geq \frac{-\Delta_1^T W_{h,\lambda}(x, z, \Delta_s)}{\chi(h, \lambda, C_2)},$$

which together with (S.21), indicates that the event that

$$\inf_{\|\Delta\| \geq \chi(h, \lambda, C_2)} \| (nH)^{1/2} W_{h,\lambda}(x, z, \Delta) \| < \chi(h, \lambda, C_4) \equiv C_4 (nH)^{1/2} [\chi_1(h) + \chi_2(h, \lambda)]$$

uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$, holds with probability approaching zero by letting $C_4 > 0$ be sufficiently large. Therefore, letting $\mathcal{E}_{n3}$ denote the event that

$$\|W_{h,\lambda}(x, z, \hat{\Delta}_{h,\lambda}(x, z))\| < C_4 [\chi_1(h) + \chi_2(h, \lambda)]$$

uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$, and $\mathcal{E}_{n4}$ denote the event that

$$\|\hat{\Delta}_{h,\lambda}(x, z)\| \geq \chi(h, \lambda, C_2)$$

uniformly over $x \in \mathcal{D}_x(e)$, $z \in \mathcal{D}_z$ and $(h, \lambda) \in \mathcal{H}_n$, we have $P(\mathcal{E}_{n3} \cap \mathcal{E}_{n4}) \to 0$ as $n \to \infty$. On the other hand, from Lemma A.5 in (Su and White 2012), we readily have $P(\mathcal{E}_{n3}^c) \to 0$ as $n \to \infty$. Then, we can prove that

$$P(\mathcal{E}_{n4}) \leq P(\mathcal{E}_{n3} \cap \mathcal{E}_{n4}) + P(\mathcal{E}_{n3}^c) \to 0 \text{ as } n \to \infty.$$ 

This leads to (S.18), completing the proof of Proposition C.1. \hfill \blacksquare

**PROPOSITION C.2.** Suppose that the assumptions in Theorem 2.1 are satisfied and $m \asymp n$. Let

$$CV_{22}(h, \lambda) = \frac{1}{n - m} \sum_{j=m+1}^n M(X_j) \int_0^{\mathcal{E}_{n1,\lambda}(X_j, Z_j)} (I(e_j \leq w) - I(e_j \leq 0)) \, dw, \quad (S.23)$$
Then we have uniformly over $(X_j, Z_j) \in \mathcal{P}_{g, \lambda}(\mu, \nu, \gamma)$, where

$$W_{\lambda, \mu, \nu, \gamma}(X_j, Z_j) = \frac{1}{m} \sum_{i=1}^{m} \eta_i(X_j, Z_j) \mathcal{K}_h(X_j - X_i) \mathcal{A}_\lambda(Z_i, Z_j).$$

Then

$$CV_{22}(h, \lambda) = CV_{22}(h, \lambda) + O_P \left( \chi_1(h) \chi_2(h, \lambda) \right)$$

(S.25)

uniformly over $(h, \lambda) \in \mathcal{H}_m$, where $\chi_3(h, \lambda) = \sqrt{\log n/n} \left[ \chi_1(h) + \chi_2(h, \lambda) \right]$.

**Proof of Proposition C.2.** The main idea of the proof is similar to that used in the proof of Lemma B.4 in *Chen et al.* (2019). By Proposition C.1 and Lemma C.1 and noting that $m \asymp n$, we may show that

$$\left| \zeta_{\lambda, \mu, \nu, \gamma}(X_j, Z_j) \right| = O_P \left( \chi_1(h) + \chi_2(h, \lambda) \right)$$

(S.26)

uniformly over $(h, \lambda) \in \mathcal{H}_m$ and $m + 1 \leq j \leq n$ with $X_j \in \mathcal{D}_x(\varepsilon)$. Letting

$$\Lambda_j(h, \lambda) = M(X_j) \int_{0}^{\zeta_{\lambda, \mu, \nu, \gamma}(X_j, Z_j)} \left( I[e_j \leq w] - I[e_j \leq 0] \right) dw,$$

and $\mathcal{F}_{j-1}$ be a $\sigma$-field generated by $(X_j, Z_j)$ and $(Y_i, X_i, Z_i)$, $i \leq j - 1$, by (S.26) and Assumptions 3 and 4, we may show that

$$E \left[ \Lambda_j(h, \lambda) | \mathcal{F}_{j-1} \right] = M(X_j) \int_{0}^{\zeta_{\lambda, \mu, \nu, \gamma}(X_j, Z_j)} \left( F_e(w|X_j, Z_j) - F_e(0|X_j, Z_j) \right) dw$$

(S.27)

uniformly over $(h, \lambda) \in \mathcal{H}_m$. Furthermore, by the uniform Bahadur representation (S.1) in Proposition C.1, we have

$$\zeta_{\lambda, \mu, \nu, \gamma}(X_j, Z_j) = \zeta_{\lambda, \mu, \nu, \gamma}(X_j, Z_j) + O_P \left( \chi_1(h) \chi_2(h, \lambda) \right)$$

(S.28)

uniformly over $(h, \lambda) \in \mathcal{H}_m$ and $m + 1 \leq j \leq n$ with $X_j \in \mathcal{D}_x(\varepsilon)$, which together with (S.26) and (S.27), leads to

$$\frac{1}{n - m} \sum_{j=m+1}^{n} E \left[ \Lambda_j(h, \lambda) | \mathcal{F}_{j-1} \right] = CV_{22}(h, \lambda) + O_P \left( \chi_1(h) \chi_2(h, \lambda) \right)$$

(S.29)

uniformly over $(h, \lambda) \in \mathcal{H}_m$, as

$$\chi_1(h) \chi_2(h, \lambda) = \chi_1^{5/2}(h) \chi_2^{1/2}(h, \lambda) \leq \chi_1^{5/2}(h) + \chi_2^{3/2}(h, \lambda).$$
and
\[ \chi_1^{3/2}(h) \chi_2(h, \lambda) \leq \chi_1^{5/2}(h) + \chi_1(h) \chi_2^{3/2}(h, \lambda). \]

By (S.29), we only need to prove
\[ \frac{1}{n - m} \sum_{j=m+1}^{n} \{ \Lambda_j(h, \lambda) - \mathbb{E} [ \Lambda_j(h, \lambda) | \mathcal{F}_{j-1} ] \} = O_p(\chi_3(h, \lambda)) \]
uniformly over \((h, \lambda) \in \mathcal{H}_m \). Note that for each \((h, \lambda) \in \mathcal{H}_m \),
\[ \{ \Lambda_j(h, \lambda) - \mathbb{E} [ \Lambda_j(h, \lambda) | \mathcal{F}_{j-1} ] , \mathcal{F}_j : j = m + 1, \ldots, n \} \]
is a sequence of martingale differences. Let \( \mathcal{E}_{n5} \) be the event that
\[ | \zeta_{M_1, h, \lambda}(X_j, Z_j) | \leq C_5 [ \chi_1(h) + \chi_2(h, \lambda) ] \]
uniformly over \((h, \lambda) \in \mathcal{H}_m \) and \( m + 1 \leq j \leq n \) with \( X_j \in \mathcal{D}_x(e) \). By (S.26), it is easy to prove that \( P(\mathcal{E}_{n5}) \to 0 \) by choosing \( C_5 > 0 \) to be sufficiently large. Hence, in order to prove (S.30), it is sufficient to show that
\[ P \left( \max_{1 \leq k \leq L_m} \left| \sum_{j=m+1}^{n} \{ \Lambda_{jk} - \mathbb{E} [ \Lambda_{jk} | \mathcal{F}_{j-1} ] \} \right| > C_6, \mathcal{E}_{n5} \right) \to 0, \]
where \( \Lambda_{jk} = \Lambda_j(h(k), \lambda(k)) \) and \( \chi_3(k) = \chi_3(h(k), \lambda(k)) \). By the Bonferroni inequality and the exponential inequality for martingale differences (e.g., Theorem 1.2A in de la Peña, 1999), we may show that
\[ \sum_{k=1}^{L_m} P \left( \left| \sum_{j=m+1}^{n} \{ \Lambda_{jk} - \mathbb{E} [ \Lambda_{jk} | \mathcal{F}_{j-1} ] \} \right| > C_6 \chi_3(k), \mathcal{E}_{n5} \right) \leq L_m \exp \left( -\overline{C}_6 \log n \right) \to 0, \]
where \( \overline{C}_6 \) is a positive constant which becomes sufficiently large when \( C_6 \) is large enough. This proves (S.31), completing the proof of Proposition C.2.

**Proposition C.3.** Suppose that the assumptions in Theorem 2.1 are satisfied and \( m \asymp n \). Letting \( \text{CV}_{22,B,1}(h, \lambda) \) and \( \text{CV}_{22,B,2}(h, \lambda) \) be defined in (B.7) in the proof of Theorem 2.1, then we have
\[ \text{CV}_{22,B,1}(h, \lambda) = o_p \left( \chi_2^2(h, \lambda) + \frac{1}{mH} \right), \]
(S.33)
and
\[
CV_{22,2}(h, \lambda) = \frac{1}{2(n-m)} \sum_{j=m+1}^{n} b_m^2(X_j, Z_j; h, \lambda) f_e(0|X_j, Z_j) M(X_j) + o_P \left( \frac{\chi_2^2(h, \lambda)}{mH} \right)
\]  \hspace{1cm} (S.34)

uniformly over \((h, \lambda) \in \mathcal{H}_m\).

**Proof of Proposition C.3.** For notational simplicity, we let
\[
\eta_i(X_j, Z_j; h, \lambda) = [\eta_i(X_j, Z_i) - \bar{\eta}_i] K_h(X_i - X_j) \Lambda \lambda(Z_i, Z_j), \quad f_M(X_j, Z_j) = \frac{f_e(0|X_j, Z_j) M(X_j)}{f^2(X_j, Z_j)}.
\]

In the sequel, “uniformly over \(j\) and \((h, \lambda)\)” means that “uniformly over \(j = m + 1, \ldots, n\) with \(X_j \in \mathcal{D}_x(\epsilon)\) and \((h, \lambda) \in \mathcal{H}_m\).” We first prove that
\[
\frac{H}{m} \sum_{i=1}^{m} \eta_i^2(X_j, Z_j; h, \lambda) = O_P \left( \chi_2^2(h, \lambda) + \chi_2^{1/2}(h, \lambda) \chi_1(h) \right)
\]  \hspace{1cm} (S.35)

uniformly over \(j\) and \((h, \lambda)\). Following the standard arguments in (S.13), we may show that
\[
\frac{H}{m} \sum_{i=1}^{m} E \left[ \eta_i^2(X_j, Z_j; h, \lambda) | X_j, Z_j \right] = O_P \left( \chi_2(h, \lambda) \right)
\]  \hspace{1cm} (S.36)

uniformly over \(j\) and \((h, \lambda)\). On the other hand, for given \(X_j, Z_j\) and each \((h, \lambda) \in \mathcal{H}_m\),
\[
\{ \eta_i^2(X_j, Z_j; h, \lambda) : i = 1, \ldots, m \}
\]
is a sequence of i.i.d. random variables. Using the techniques in the proof of Lemma C.1, we can prove that
\[
\frac{H}{m} \sum_{i=1}^{m} \left\{ \eta_i^2(X_j, Z_j; h, \lambda) - E \left[ \eta_i^2(X_j, Z_j; h, \lambda) | X_j, Z_j \right] \right\} = O_P \left( \chi_2^{1/2}(h, \lambda) \chi_1(h) \right).
\]  \hspace{1cm} (S.37)

Then, (S.35) can be proved using (S.36) and (S.37). By (S.35), Assumption 5(ii) and the Law of Large Numbers, we can complete the proof of (S.33).

Next, we turn to the proof of (S.34). Note that
\[
\frac{1}{2m^2(n-m)} \sum_{j=m+1}^{n} f_M(X_j, Z_j) \sum_{i_1=1}^{m} \sum_{i_2=1, i_2 \neq i_1}^{m} E \left[ \eta_{i_1}(X_j, Z_j; h, \lambda) | X_j, Z_j \right] \cdot E \left[ \eta_{i_2}(X_j, Z_j; h, \lambda) | X_j, Z_j \right]
\]
\[
= \frac{m(m-1)}{2m^2(n-m)} \sum_{j=m+1}^{n} f_M(X_j, Z_j) \left\{ E \left[ \eta_1(X_j, Z_j; h, \lambda) | X_j, Z_j \right] \right\}^2
\]
\[
= \frac{1}{2(n-m)} \sum_{j=m+1}^{n} b^2(X_j, Z_j; h, \lambda) f_e(0|X_j, Z_j) M(X_j) + o_P \left( \frac{\chi_2^2(h, \lambda)}{m} + \chi_2^2(h, \lambda) \right)
\]
10
\[ = \frac{1}{2(n - m)} \sum_{j=m+1}^{n} b^2(X_j, Z_j; h, \lambda) f_e(0|X_j, Z_j) M(X_j) + o_P \left( \chi_2^2(h, \lambda) \right). \]  
(S.38)

Let

\[ \eta_{i, h, \lambda}^c(X_j, Z_j) = \eta_i(X_j, Z_j; h, \lambda) - E \left[ \eta_i(X_j, Z_j; h, \lambda) | X_j, Z_j \right] \]

and

\[ R_i(X_j, Z_j; h, \lambda) = \eta_{i, h, \lambda}^c(X_j, Z_j) \sum_{k=1}^{i-1} \eta_{k, h, \lambda}^c(X_j, Z_j) \]

for \( i = 2, \ldots, m \). Write

\[ \sum_{i=2}^{m} R_i(X_j, Z_j; h, \lambda) = \sum_{i=2}^{m_*} R_i(X_j, Z_j; h, \lambda) + \sum_{i=m_*+1}^{m} R_i(X_j, Z_j; h, \lambda), \]

where \( m_* = [m^{2/3}H^{1/3}] \). Let \( \mathcal{F}_{i-1}(X_j, Z_j) \) be a sigma-field generated by \( (X_j, Z_j), (X_k, Z_k) \) with \( k \leq i \) and \( \epsilon_k \) with \( k \leq i - 1 \). Note that \( \{ R_i(X_j, Z_j; h, \lambda), \mathcal{F}_i(X_j, Z_j) : i = 2, \ldots, m \} \) is a sequence of martingale differences with mean zero. As in the proof of Proposition C.2, using the Bonferroni inequality and the exponential inequality for martingale differences (e.g., de la Peña, 1999), we can prove that

\[ \frac{1}{m^2(n - m)} \sum_{j=m+1}^{n} \sum_{i=2}^{m_*} R_i(X_j, Z_j; h, \lambda) = o_P \left( \frac{\chi_2^2(h, \lambda) \log^{1/2} n}{m H} \right) = o_P \left( \frac{1}{m H} \right) \]  
(S.39)

uniformly over \( (h, \lambda) \in \mathcal{H}_m \). On the other hand, we can similarly show that

\[ \frac{1}{m^2(n - m)} \sum_{j=m+1}^{n} \sum_{i=m_*+1}^{m} R_i(X_j, Z_j; h, \lambda) = o_P \left( \frac{\chi_2^2(h, \lambda) \log n}{m H^{1/2}} \right) = o_P \left( \frac{1}{m H} \right) \]  
(S.40)

uniformly over \( (h, \lambda) \in \mathcal{H}_m \). By (S.39) and (S.40), we have

\[ \frac{1}{m^2(n - m)} \sum_{j=m+1}^{n} \sum_{i=1}^{m} \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \eta_{i_1, h, \lambda}^c(X_j, Z_j) \eta_{i_2, h, \lambda}^c(X_j, Z_j) = o_P \left( \frac{1}{m H} \right) \]  
(S.41)

uniformly over \( (h, \lambda) \in \mathcal{H}_m \). Analogously, we can also prove that

\[ \frac{1}{m^2(n - m)} \sum_{j=m+1}^{n} \sum_{i=1}^{m} E \left[ \eta_{i, h, \lambda}^c(X_j, Z_j) | X_j, Z_j \right] \sum_{i_2=1}^{m} \eta_{i_2, h, \lambda}^c(X_j, Z_j) = o_P \left( \frac{1}{m H} \right) \]  
(S.42)

uniformly over \( j \) and \( (h, \lambda) \). By (S.38), (S.41) and (S.42), we prove (S.34).

**Proposition C.4.** Suppose that the assumptions in Theorem 2.1 are satisfied and \( m \asymp n \). Letting
CV\textsuperscript{2, V, 1}(h, \lambda) \text{ and } CV\textsuperscript{2, V, 2}(h, \lambda) \text{ be defined in (B.9) in the proof of Theorem 2.1, then we have}

\[
CV\textsuperscript{2, V, 1}(h, \lambda) = \frac{1}{2(n - m)} \sum_{j=m+1}^{n} \sigma_{n}(X_j, Z_j \mid h, \lambda) f_x(0 \mid X_j, Z_j) M(X_j) + o_P \left( \frac{1}{mH} \right),
\]

(S.43)

and

\[
CV\textsuperscript{2, V, 2}(h, \lambda) = o_P \left( \frac{1}{mH} \right)
\]

(S.44)

uniformly over \((h, \lambda) \in \mathcal{H}_m\).

**Proof of Proposition C.4.** For notational simplicity, throughout this proof, “uniformly over \((h, \lambda)\)” means that “uniformly over \(j = m + 1, \ldots, n\) with \(X_j \in \mathcal{D}_x(\varepsilon)\) and \((h, \lambda) \in \mathcal{H}_m\).” By some standard calculations, we may show that

\[
\frac{H}{m} \sum_{i=1}^{m} E \left[ \tilde{\eta}_i^2 K^2_h(X_i - X_j) \Lambda^2_{\lambda}(Z_i, Z_j) \right] = \tau (1 - \tau) \nu^2_0 f_{x \perp}(X_j, Z_j) + O(\chi_2(h, \lambda))
\]

(S.45)

uniformly over \(j \text{ and } (h, \lambda)\). On the other hand, with the arguments in the proof of Lemma C.1, we can prove that

\[
\frac{H}{m} \sum_{i=1}^{m} \{ \tilde{\eta}_i^2 K_h^2(X_i - X_j) \Lambda^2_{\lambda}(Z_i, Z_j) - E \left[ \tilde{\eta}_i^2 K_h^2(X_i - X_j) \Lambda^2_{\lambda}(Z_i, Z_j) \mid X_j, Z_j \right] \} = O_P(\chi_1(h))
\]

(S.46)

uniformly over \(j \text{ and } (h, \lambda)\). Combining (S.45) and (S.46), we can complete the proof of (S.43).

As in the proof of Proposition S.4, we let \(f_M(X_j, Z_j) = f_e(0 \mid X_j, Z_j) M(X_j) / f^2(X_j, Z_j)\). The proof of (S.44) is similar to the proof of (S.41). Define

\[
\tilde{R}_i(h, \lambda) = \tilde{\eta}_i \sum_{j=m+1}^{n} f_M(X_j, Z_j) K_h(X_i - X_j) \Lambda_{\lambda}(Z_i, Z_j) \sum_{k=1}^{i-1} \tilde{\eta}_k K_h(X_k - X_j) \Lambda_{\lambda}(Z_k, Z_j).
\]

and let \(\widetilde{\mathcal{F}}_{i-1}\) be a \(\sigma\)-field generated by \(\mathcal{M}_2\) (the validation set), \((X_k, Z_k)\) with \(k \leq i\) and \(e_k\) with \(k \leq i - 1\). For each \((h, \lambda) \in \mathcal{H}_m\), it is easy to verify that \(\{\tilde{R}_i(h, \lambda), \tilde{\mathcal{F}}_i : i = 2, \ldots, m\}\) is a sequence of martingale differences with mean zero. To prove (S.44), it is sufficient to show that

\[
\frac{1}{m^2(n - m)} \sum_{i=2}^{m} \tilde{R}_i(h, \lambda) = O_P \left( \frac{1}{mH} \right)
\]

(S.47)

uniformly over \((h, \lambda) \in \mathcal{H}_m\). Write

\[
\sum_{i=2}^{m} \tilde{R}_i(h, \lambda) = \sum_{i=2}^{m_0} \tilde{R}_i(h, \lambda) + \sum_{i=m_0 + 1}^{m} \tilde{R}_i(h, \lambda),
\]

12
where $m_o = [m / \log^2 n]$. Using the Bonferroni inequality and the exponential inequality for martingale differences (e.g., de la Peña, 1999) as well as the fact that

\[
\frac{1}{(n - m)} \sum_{j=m+1}^{n} f_M(X_j, Z_j) K_h(X_i - X_j) \Lambda(z_i, z_j) = f_M(X_i, Z_i) + O_P \left( \left( \frac{\log n}{n^H} \right)^{1/2} \right)
\]

uniformly over $i = 1, \ldots, m$ and $(h, \lambda) \in \mathcal{H}_m$, we can prove that

\[
\frac{1}{m^2(n - m)} \sum_{i=2}^{m_o} \tilde{R}_i(h, \lambda) = O_P \left( \frac{m_o \log n}{m^2 H} \right) = o_P \left( \frac{1}{mH} \right)
\]

and

\[
\frac{1}{m^2(n - m)} \sum_{i=m_o+1}^{m} \tilde{R}_i(h, \lambda) = O_P \left( \frac{\log n}{mH^{1/2}} \right) = o_P \left( \frac{1}{mH} \right)
\]

uniformly over $(h, \lambda) \in \mathcal{H}_m$. By (S.49) and (S.50), we prove (S.47), completing the proof of (S.44).

**Proposition C.5.** Suppose that the assumptions in Theorem 2.1 are satisfied and $m \asymp n$. Letting $CV^*_{22, BV}(h, \lambda)$ be defined in (B.6) in the proof of Theorem 2.1, then we have

\[
CV^*_{22, BV}(h, \lambda) = o_P \left( \chi^2_2(h, \lambda) + \frac{1}{mH} \right)
\]

uniformly over $(h, \lambda) \in \mathcal{H}_m$.

**Proof of Proposition C.5.** Note that

\[
CV^*_{22, BV}(h, \lambda) = \frac{1}{n - m} \sum_{j=m+1}^{n} \mathbb{E} \left[ B_{M_1, h, \lambda}(X_j, Z_j) | X_j, Z_j \right] V_{M_1, h, \lambda}(X_j, Z_j) f_e(0|X_j, Z_j) M(X_j) + \frac{1}{n - m} \sum_{j=m+1}^{n} B_{M_1, h, \lambda}(X_j, Z_j) V_{M_1, h, \lambda}(X_j, Z_j) f_e(0|X_j, Z_j) M(X_j) = CV^*_{22, BV, 1}(h, \lambda) + CV^*_{22, BV, 2}(h, \lambda),
\]

where $B_{M_1, h, \lambda}(X_j, Z_j) = B_{M_1, h, \lambda}(X_j, Z_j) - \mathbb{E} \left[ B_{M_1, h, \lambda}(X_j, Z_j) | X_j, Z_j \right]$. As in the proofs of the previous two propositions, in the sequel, “uniformly over $j$ and $(h, \lambda)$” means that “uniformly over $j = m + 1, \ldots, n$ with $X_j \in \mathcal{D}_c(\epsilon)$ and $(h, \lambda) \in \mathcal{H}_m$”.

Following the arguments in the proof of Proposition C.3, we may show that

\[
|B_{M_1, h, \lambda}(X_j, Z_j)| = O_P \left( \chi^{1/2}_2(h, \lambda) \chi_1(h) \right)
\]

uniformly over $j$ and $(h, \lambda)$. On the other hand, following the arguments in the proof of Lemma C.1, we have

\[
|V_{M_1, h, \lambda}(X_j, Z_j)| = O_P \left( \chi_1(h) \right)
\]
uniformly over \( j \) and \( (h, \lambda) \). Combining (S.53) and (S.54), we have

\[
CV^*_{22,B,V,2}(h, \lambda) = O_p \left( \frac{\chi_2^2(h, \lambda)\chi_1^2(h)}{\frac{1}{mH}} \right)
\]

(S.55)

uniformly over \( (h, \lambda) \in H_m \). Similarly, we can show that

\[
CV^*_{22,B,V,1}(h, \lambda) = O_p \left( \chi_2(h, \lambda)\eta^{-1/2} + \chi_2(h, \lambda)\chi_1^2(h) \right) = O_p \left( \frac{\chi_2^2(h, \lambda) + \frac{1}{mH}}{\frac{1}{mH}} \right)
\]

(S.56)

uniformly over \( (h, \lambda) \in H_m \). By (S.52), (S.55) and (S.56), we can complete the proof of (S.51). 

\[\blacksquare\]

**Appendix D: Weak convergence of quantile regression estimation uniformly over \( \tau \)**

In this appendix, we discuss extension of the weak convergence result developed in Qu and Yoon (2015) from the case of continuous regressors to the case of mixed continuous and discrete regressors. As suggested by Qu and Yoon (2015) and discussed in Section 6, we may combine the local linear smoothing method introduced in Section 2 with the linear interpolation technique to obtain quantile regression estimates for all \( \tau \in \mathcal{J} = [\underline{\tau}, \overline{\tau}] \), where \( 0 < \underline{\tau} < \overline{\tau} < 1 \). Specifically, consider a set of grid points (with equal distance) arranged in the increasing order: \( \{\tau_0, \tau_1, \ldots, \tau_n\} \) with \( \tau_0 = \underline{\tau} \) and \( \tau_n = \overline{\tau} \), where \( n \) is a positive integer which may be divergent to infinity as \( n \) increases. For each given \( \tau_k \), \( 0 \leq k \leq n \), we minimize the loss function in (2.2) with respect to \( \alpha \), and obtain the local linear estimate of the \( \tau_k \)-quantile regression function \( \hat{Q}_{\tau_k}(x_0, z_0) \), denoted by \( \hat{Q}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k}) \), where we make the dependence of \( h_{\tau_k} \) and \( \lambda_{\tau_k} \) on \( \tau_k \) explicit. Then, we apply the technique of linear interpolation between \( \hat{Q}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k}) \), i.e.,

\[
\hat{Q}^o_{\tau}(x_0, z_0) = \frac{\tau_{k+1} - \tau}{\tau_{k+1} - \tau_k} \cdot \hat{Q}_{\tau_k}(x_0, z_0; h_{\tau_k}, \lambda_{\tau_k}) + \frac{\tau - \tau_k}{\tau_{k+1} - \tau_k} \cdot \hat{Q}_{\tau_{k+1}}(x_0, z_0; h_{\tau_{k+1}}, \lambda_{\tau_{k+1}})
\]

(S.57)

for \( \tau_k \leq \tau \leq \tau_{k+1} \). To ensure monotonicity of the estimated quantile regression curves (with respect to \( \tau \)), we may further apply the re-arrangement technique to \( \hat{Q}^o_{\tau}(x_0, z_0) \) defined in (S.57), see, for example, Chernozhukov, Fernández-Val and Galichon (2010). Let \( e_1(\tau) = Y - Q_\tau(X_i, Z_i) \) and \( f_{e,\tau}(\cdot|x_0, z_0) \) be the conditional density function of \( e_1(\tau) \) given \( X_i = x_0 \) and \( Z_i = z_0 \). To generalize Qu and Yoon (2015)’s result to our setting, we need the following smoothness conditions on \( Q_\tau(x_0, z_0) \) and \( f_{e,\tau}(\cdot|x_0, z_0) \) (with respect to \( \tau \)) as well as some mild restriction on the smoothing parameters \( h_{\tau_k} \) and \( \lambda_{\tau_k} \).

**Assumption 6.** (i) The conditional density of \( e_1(\tau) = 0 \) for given \( X_i = x_0 \) and \( Z_i = z_0 \), \( f_{e,\tau}(\cdot|x_0, z_0) \) is Lipschitz continuous and bounded away from zero and infinity uniformly over \( \mathcal{J} \).

(ii) The quantile regression function \( Q_\tau(x_0, z_0) \), its first-order derivative (with respect to \( \tau \)) and \( Q_{\tau_{i,j}}''(x_0, z_0) \) are finite and Lipschitz continuous over \( \mathcal{J} \).
Assumption 7. (i) There exist two functions of $\tau$: $a(\tau)$ and $d(\tau)$ such that $h_{\tau} = a(\tau)n^{-1/(p+4)}1_p$ and $\lambda_{\tau} = d(\tau)n^{-2/(p+4)}1_q$, where $1_r$ is an r-dimensional vector with each component being one. In addition, $a(\tau)$ and $d(\tau)$ are positive and Lipschitz continuous over $T$.

(ii) The number of grid points $\tau_n$ satisfies that $\tau_n/n^{1/(p+4)} \to \infty$.

Combining the above two assumptions and Assumptions 1–4 in Appendix A, we can easily verify Assumptions 1–5 in Qu and Yoon (2015). Following the proof of Theorem 1 in their paper or the proof of Proposition C.1, we may prove the following Bahadur representation uniformly over $\tau \in T$:

$$\sqrt{n}H_\tau \left[ \hat{Q}_\tau(x_0, z_0; h_{\tau}, \lambda_{\tau}) - Q_\tau(x_0, z_0) - b(x_0, z_0; h_{\tau}, \lambda_{\tau}) \right] = s^{-1}(x_0, z_0)W_{n, \tau}(x_0, z_0; h_{\tau}, \lambda_{\tau}) + o_P(1), \quad (S.58)$$

where $x_0$ is an interior point of $D_x$, $H_\tau = \prod_{j=1}^p h_{j, \tau}$ (with $h_{j, \tau}$ being the j-th component of $h_\tau$), $b(x_0, z_0; h_{\tau}, \lambda_{\tau})$ is defined as in Section 3, $s_\tau(x_0, z_0) = f_{xz}(x_0, z_0)f_{x, \tau}(0|x_0, z_0)$, and

$$W_{n, \tau}(x_0, z_0; h_{\tau}, \lambda_{\tau}) = (nH_\tau)^{-1/2} \sum_{i=1}^n \tilde{e}_i, i, K \frac{X_i - x_0}{h_{\tau}} \Lambda_{\lambda_{\tau}}(Z_{i, 0}), \quad \tilde{e}_i, i, \tau = \tau - I(e_i(\tau) \leq 0).$$

Note that the involvement of the discrete kernel in quantile regression estimation only affects the asymptotic bias term. Using (S.58) and following the proof of Theorem 2 in Qu and Yoon (2015), we can prove the following weak convergence result.

Proposition D.1. Suppose that Assumptions 1–4, 6 and 7 are satisfied. Then,

$$\sqrt{n}H_\tau f_{x, \tau}^{1/2}(x_0, z_0)f_{x, \tau}(0|x_0, z_0) \left[ \hat{Q}_\tau(x_0, z_0; h_{\tau}, \lambda_{\tau}) - Q_\tau(x_0, z_0) - b(x_0, z_0; h_{\tau}, \lambda_{\tau}) \right] \Rightarrow G(\tau), \quad (S.59)$$

where $G(\tau)$ is a continuous Gaussian process over $\tau \in T$ with mean zero and covariance function defined by

$$E \left[ G(\tau_1)G(\tau_2) \right] = (\tau_1 \land \tau_2 - \tau_1 \tau_2) [k(\tau_1)k(\tau_2)]^{-p/2} \int K \left( \frac{w}{k(\tau_1)} \right) K \left( \frac{w}{k(\tau_2)} \right) dw, \quad k(\tau) = a(\tau)/a(0.5).$$

A natural question is whether the above weak convergence result is still valid if we replace $h_{\tau}$ and $\lambda_{\tau}$ by the RCV selected bandwidths $\hat{h}_{\tau}$ and $\hat{\lambda}_{\tau}$. Note that Proposition D.1 holds when the theoretically optimal bandwidths $h_{\tau}^0$ and $\lambda_{\tau}^0$ are used. Following the arguments in the proof of Theorem 2.1 in Li and Li (2010), we may extend Proposition D.1 to the more general scenario, using the data-dependent RCV selected smoothing parameters as long as

$$\hat{h}_{j, \tau} - h_{j, \tau}^0 = o_P(n^{-1/(p+4)}), \quad j = 1, \cdots, p \quad (S.60)$$

and

$$\hat{\lambda}_{k, \tau} - \lambda_{k, \tau}^0 = o_P(n^{-2/(p+4)}), \quad k = 1, \cdots, q, \quad (S.61)$$

uniformly over $\tau \in T$, where $\hat{h}_{j, \tau}$ and $h_{j, \tau}^0$ are the j-th components of $\hat{h}_{\tau}$ and $h_{\tau}^0$, respectively, and similarly
$\hat{\lambda}_{k,\tau}$ and $\lambda^0_{k,\tau}$ are the k-th components of $\hat{\lambda}_\tau$ and $\lambda^0_\tau$, respectively. However, it is quite challenging to prove (S.60) and (S.61) using the techniques developed in the present paper. For example, we would have to prove the asymptotic expansion of the CV function in Theorem 2.1 uniformly over both $(h, \lambda)$ and $\tau$. We leave this challenging open question to a future research topic.

Appendix E: Point-wise confidence intervals in the empirical study

We plot point-wise 95% confidence intervals for the 10%, 25%, 75% and 90% conditional quantile functions. The confidence intervals in Figure S.1 are constructed via the standard (resampling with replacement) bootstrap method with undersmoothed bandwidth vectors.

Figure S.1: The 95% point-wise confidence intervals of 10%, 25%, 75% and 90% conditional quantiles

References


