SUPPLEMENTARY MATERIALS

Proofs of Theorems

Assumptions 1-6 hold throughout this appendix. We use linear functional notation. For any function f

$$\mathcal{P}f = \int f(x)dP(x); \qquad \mathcal{P}_n f = \int f(x)dP_n(x),$$

where $\mathcal{P}$ and $\mathcal{P}_n$, respectively, are the distribution function and empirical distribution function of (Y, X).

A.1 Replacing the random bandwidth h with the non-stochastic bandwidth h_0.

For any bandwidth s and any $x \in \mathcal{S}$, define

$$Q_n(x, b_0, b_1, s) = n^{-1}\sum_{i=1}^{n} \rho_\tau [Y_i - b_0 - b_1(X_i - x)]K_s(X_i - x).$$

Let d_0 be as in assumption 6 and $h_0 = d_0 n^{-1/5}$. The following lemma shows that replacing the random bandwidth h with the non-stochastic bandwidth h_0 has an asymptotically negligible effect on the local linear estimator of $g(x)$.

<u>Lemma A.1</u>: Define

$$(\hat{b}_{0h}, \hat{b}_{1h}) = \arg \min_{(b_0, b_1)\in\mathcal{G}} Q_n(x, b_0, b_1, h)$$

and

$$(\hat{b}_{0h_0}, \hat{b}_{1h_0}) = \arg \min_{(b_0, b_1)\in\mathcal{G}} Q_n(x, b_0, b_1, h_0).$$

For each $x \in \mathcal{S}$ and $j = 0$ or 1, $\hat{b}_{jh} - \hat{b}_{jh_0} = o_p(n^{-2/5})$

<u>Proof</u>: Let D_1 be as in assumption 6(ii). Define

$$a_n = n^{-D_4}\left(\frac{\log n}{nh_0}\right)^{1/2},$$

where $0 < D_4 < D_1$. Let $\mathcal{O}$ be an open neighborhood of $(\hat{b}_{0h_0}, \hat{b}_{1h_0})$ such that

$$a_n = \min_{(b_0,b_1)\in\mathcal{G}\setminus\mathcal{O}} Q_n(x,b_0,b_1,h_0) - Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0).$$

The proof takes place in three steps. Step 1 shows that if

(A.1) $$\sup_{(b_0,b_1)\in\mathcal{G}} |Q_n(x,b_0,b_1,h) - Q_n(x,b_0,b_1,h_0)| < a_n/2,$$

then $\hat{\boldsymbol{b}}_h \equiv (\hat{b}_{0h}, \hat{b}_{1h}) \in \mathcal{O}$. Step 2 shows that (A.1) holds with probability approaching 1 as $n \to \infty$. Step 3 shows that $\hat{b}_{0h} - \hat{b}_{0h_0} = o_p(n^{-2/5})$ if $\hat{\boldsymbol{b}}_h \in \mathcal{O}$.

Step 1: Let A_n be the event

$$\sup_{(b_0,b_1)\in\mathcal{G}} |Q_n(x,b_0,b_1,h) - Q_n(x,b_0,b_1,h_0)| < a_n/2.$$

Then

(A.2) $$A_n \Rightarrow Q_n(x,\hat{b}_{0h},\hat{b}_{1h},h) > Q_n(x,\hat{b}_{0h},\hat{b}_{1h},h_0) - a_n/2$$

and

(A.3) $$A_n \Rightarrow Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0) > Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h) - a_n/2.$$

But $Q_n(x,\hat{b}_{0h},\hat{b}_{1h},h) \le Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h)$. Therefore, it follows from (A.3) that

(A.4) $$A_n \Rightarrow Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0) > Q_n(x,\hat{b}_{0h},\hat{b}_{1h},h) - a_n/2.$$

Substituting (A.2) into (A.4) yields

(A.5) $$A_n \Rightarrow Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0) > Q_n(x,\hat{b}_{0h},\hat{b}_{1h},h_0) - a_n,$$

Equivalently,

(A.6) $A_n \Rightarrow Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) < a_n$.

Therefore, $A_n \Rightarrow (\hat{b}_{0h}, \hat{b}_{1h}) \in \mathcal{O}$ because $Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) \geq a_n$ if $(\hat{b}_{0h}, \hat{b}_{1h}) \notin \mathcal{O}$.

Step 2: A Taylor series expansion of $Q_n(x, b_0, b_1, h)$ about $\hat{h} = h_0$ yields

$$Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) = \frac{h_0}{\tilde{h}} \frac{1}{nh_0} \sum_{i=1}^{n} \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K' \left(\frac{X_i - x}{\tilde{h}} \right) \left(\frac{h - h_0}{\tilde{h}} \right)$$

$$= \frac{h_0}{\tilde{h}^2} \mathcal{P} \left\{ \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K' \left(\frac{X_i - x}{\tilde{h}} \right) \left(\frac{h - h_0}{h_0} \right) \right\}$$

$$+ \frac{h_0}{\tilde{h}^2} (\mathcal{P}_n - \mathcal{P}) \left\{ \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K' \left(\frac{X_i - x}{\tilde{h}} \right) \left(\frac{h - h_0}{h_0} \right) \right\},$$

where $\tilde{h}$ is between h_0 and h. By assumption 6 and Theorem 2.37 of Pollard (1984)

$$\frac{h_0}{\tilde{h}^2} (\mathcal{P}_n - \mathcal{P}) \left\{ \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K' \left(\frac{X_i - x}{\tilde{h}} \right) \left(\frac{h - h_0}{h_0} \right) \right\} =^{a.s.} O\left[n^{-D_1} \left(\frac{(\log n)^{1+\varepsilon}}{nh_0} \right)^{1/2} \right]$$

for any $\varepsilon > 0$. Standard calculations for kernel estimators combined with assumption 3 yield the result that

$$\frac{h_0}{\tilde{h}^2} \mathcal{P} \left\{ \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K' \left(\frac{X_i - x}{\tilde{h}} \right) \left(\frac{h - h_0}{h_0} \right) \right\} = O\left(n^{-D_1} h_0^2 \right).$$

Therefore,

$$| Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) | < a_n / 2$$

almost surely for all sufficiently large n.

Step 3: It follows from Theorem 2.37 of Pollard (1984) that

$$Q_n(x,b_0,b_1,h_0) - Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0) =^{a.s.} \ell(x)(b_0 - \hat{b}_{0h_0}) + o(1),$$

where ℓ is a non-zero function that does not depend on n. Therefore, $\hat{\boldsymbol{b}}_h \in \mathcal{O}$ implies that $\hat{b}_{0h} - \hat{b}_{0h_0} = o_p(n^{-2/5})$. Q.E.D.

A.2 Proofs of Theorem 3.1, Theorem 3.2, and Corollary 3.3

Proof of Theorem 3.1: The constraint $(b_0,b_1) \in \mathcal{G}$ in Lemma A.1 is non-binding with probability approaching 1 as $n \to \infty$. Therefore, it suffices to consider the local linear estimator of $g(x)$ obtained in Section 2.1 with the non-stochastic bandwidth h_0 in place of h. Denote this estimator by $\hat{g}_x$. Denote the estimator of $g'(x)$ by $\hat{g}'_x$. Let $g''_x = d^2 g(x)/dx^2$. Then

$$\text{(A.7)} \qquad (\hat{g}_x, \hat{g}'_x) = \arg\min_{b_0,b_1} n^{-1} \sum_{i=1}^{n} \rho_\tau [Y_i - b_0 - b_1(X_i - x)] K_{h_0}(X_i - x)$$

$$= \arg\min_{b_0,b_1} \mathcal{P}_n \left\{ \rho_\tau [Y - b_0 - b_1(X - x)] K_{h_0}(X - x) \right\}.$$

For each $x \in \mathcal{S}$, define $\boldsymbol{b}_x = (b_{x0}, b_{x1})'$ to be an arbitrary 2×1 vector. An argument similar to that used to prove Lemma A.2 of Ruppert and Carroll (1980) shows that he first-order conditions for (A.7) are

$$\text{(A.8)} \quad \mathcal{P}_n \left\{ \tau - I[Y - b_{x0} - b_{x1}(X - x) \le 0] \right\} K_{h_0}(X - x) = O_p(h_0 / n).$$

$$\mathcal{P}_n \left\{ \tau - I[Y - b_{x0} - b_{x1}(X - x) \le 0] \right\} (X - x) K_{h_0}(X - x) = O_p(h_0 / n).$$

As is shown below, the asymptotic form of $\hat{g}_x - g_x$ depends only on (A.8). Therefore, only (A.8) is treated in the remainder of the proof. Define

$$T_{n1}(x) = \mathcal{P} \left\{ \tau - I[Y - g_x - g'_x(X - x) \le 0] \right\} K_{h_0}(X - x),$$

$$T_{n2}(x) = (\mathcal{P}_n - \mathcal{P})\left\{\tau - I[Y - g_x - g'_x(X-x) \le 0]\right\} K_{h_0}(X-x) ,$$

$$T_{n3}(x, \boldsymbol{b}_x) = \mathcal{P}\left\{I[Y - b_{x0} - b_{x1}(X-x) \le 0] - I[Y - g_x - g'_x(X-x) \le 0]\right\} K_{h_0}(X-x) ,$$

and

$$T_{n4}(x, \boldsymbol{b}_x) = (\mathcal{P}_n - \mathcal{P})\left\{I[Y - b_{x0} - b_{x1}(X-x) \le 0] - I[Y - g_x - g'_x(X-x) \le 0]\right\} K_{h_0}(X-x) .$$

In these definitions, $\boldsymbol{b}_x$ is an arbitrary, non-stochastic vector. Then

$$\mathcal{P}_n\left\{\tau - I[Y - b_{x0} - b_{x1}(X-x) \le 0]\right\} K_{h_0}(X-x) = T_{n1}(x) + T_{n2}(x) + T_{n3}(x, \boldsymbol{b}_x) + T_{n4}(x, \boldsymbol{b}_x) .$$

We now derive the asymptotic forms of T_{n1}, T_{n2}, T_{n3}, and T_{n4}.

Analysis of $T_{n1}(x)$: Let F_ε denote the distribution function of ε in (1.1). Then a Taylor series expansion yields

$$T_{n1}(x) = \tau - \int F_\varepsilon[g(z) - g_x - g'_x(z-x)] K_{h_0}(z-x) f_X(z) dz$$

$$= \frac{h_0^3 \kappa_2}{2} f_X(x) f_\varepsilon(0) g''_x + O\left(h_0^4\right) \tag{A.9}$$

uniformly over $x \in \mathcal{S}$.

Analysis of $T_{n2}(x)$: Define

$$T_{n2a}(x) = \frac{\tau}{n}\sum_{i=1}^{n}\left\{K_{h_0}(X_i - x) - E[K_{h_0}(X_i - x)]\right\} ,$$

$$T_{n2b}(x) = n^{-1}\sum_{i=1}^{n}\{I(\varepsilon_i \le 0) K_{h_0}(X_i - x) - E[I(\varepsilon_i \le 0) K_{h_0}(X_i - x)]\} ,$$

and

$$T_{n2c}(x) = \frac{\tau}{n}\sum_{i=1}^{n}\Big(\{I[\varepsilon_i \le g_x - g_{X_i} + g'_x(X_i - x)] - I(\varepsilon_i \le 0)\}K_{h_0}(X_i - x)$$

$$-E\{I[\varepsilon_i \le g_x - g_{X_i} + g'_x(X_i - x)] - I(\varepsilon_i \le 0)\}K_{h_0}(X_i - x)\Big).$$

Then

$$T_{n2} = T_{n2a} + T_{n2b} + T_{n2c}.$$

Let F_X and F_{nX} and, respectively, denote the distribution and empirical distributions functions of X. Define the stochastic process $Z_{nx}(x) = n^{1/2}[F_{nX}(x) - F_X(x)]$. Define the limit process $Z_x^0(x)$ by $Z_{nx}(x) \rightsquigarrow Z_x^0(x)$ as $n \to \infty$. Then a change of variables and integration by parts yields

$$n^{1/2}T_{n2a}(x) = \tau\int K\left(\frac{v-x}{h_0}\right)dZ_{nx}(v)$$

$$= -\tau\int Z_{nx}(x + h_0\xi)K'(\xi)d\xi$$

$$= -\tau\int Z_x^0(x + h_0\xi)K'(\xi)d\xi - \tau\int[Z_{nx}(x + h_0\xi) - Z_x^0(x + h_0\xi)]K'(\xi)d\xi$$

$$\equiv n^{1/2}T_{n2a}^{(1)}(x, h_0) + n^{1/2}T_{n2a}^{(2)}(x, h_0).$$

It follows from Theorem 3 of Komlós, Major, and Tusnády (1975) that there are processes $\tilde{Z}_{nx}$ and $\tilde{Z}_x^0$ having the same distributions as Z_{nx} and Z_x^0 such that

$$P\left[\sup_{x\in\mathcal{S}} |\tilde{Z}_{nx}(x) - \tilde{Z}_x^0(x)| > C_1 n^{-1/2}\log n\right] < n^{-C_2},$$

where C_1 and C_2 are constants. Therefore, $(n/h_0)^{1/2}T_{n2a}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2}T_{n2a}^{(1)}(x, h_0)$ in the sense that

(A.10) $P\left[\sup_{x\in\mathcal{S}} |(n/h_0)^{1/2}T_{n2a}(x)-(n/h_0)^{1/2}T_{n2a}^{(1)}(x,h_0)|>C_3(nh_0)^{-1/2}\log n\right]<n^{-C_2}$.

Now consider T_{n2b}. Let $F_{\varepsilon x}$ denote the distribution function of (ε, X) and $F_{n\varepsilon x}$ denote the empirical distribution function. Define $Z_{n\varepsilon x}(\varepsilon,x)=n^{1/2}[F_{n\varepsilon x}(\varepsilon,x)-F_{\varepsilon x}(\varepsilon,x)]$, and let $Z_{\varepsilon x}^0(\varepsilon,x)$ denote the limiting Gaussian process of $Z_{n\varepsilon x}(\varepsilon,x)$. Integration by parts and a change of variables yields

$$
\begin{aligned}
n^{1/2}T_{n2b}(x) &= \int I(s\le 0)K\left(\frac{v-x}{h_0}\right)dZ_{n\varepsilon x}(s,v) \\
&= -\int Z_{\varepsilon x}^0(0,x+h_0\xi)K'(\xi)d\xi-\int [Z_{n\varepsilon x}(0,x+h_0\xi)-Z_{\varepsilon x}^0(0,x+h_0\xi)]K'(\xi)d\xi \\
&\equiv n^{1/2}T_{n2b}^{(1)}(x,h_0)+n^{1/2}T_{n2b}^{(2)}(x,h_0).
\end{aligned}
$$

To bound $n^{1/2}T_{n2b}^{(2)}$, let $\hat{n}_0$ denote the number of observations for which $\varepsilon_i\le 0$. Assume without loss of generality that these are the first $\hat{n}_0$ observations. The corresponding X_i 's are a random sample of X , because X and ε are independent. We have

$$
\begin{aligned}
F_{n\varepsilon x}(0,x) &= n^{-1}\sum_{i=1}^{n} I(\varepsilon_i\le 0)I(X_i\le x) \\
&= \frac{\hat{n}_0}{n}\hat{n}_0^{-1}\sum_{i=1}^{\hat{n}_0} I(X_i\le x) \\
&\equiv \frac{\hat{n}_0}{n}F_{\hat{n}_0 x}(x).
\end{aligned}
$$

Moreover, because ε and X are independent,

$$F_{n\varepsilon x}(0,x) - F_{\varepsilon x}(0,x) = \frac{\hat{n}_0}{n} F_{\hat{n}_0 x}(x) - F_\varepsilon(0) F_x(x)$$

$$= F_{n\varepsilon}(0)[F_{\hat{n}_0 x}(x) - F_x(x)] + \left[\frac{\hat{n}_0}{n} - F_\varepsilon(0)\right] F_{\hat{n}_0 x}(x)$$

$$= F_\varepsilon(0)\hat{n}_0^{-1/2} Z_x^0(x) + n^{-1/2}\xi F_{\hat{n}_0}(x) + F_\varepsilon(0)\hat{n}_0^{-1/2} r_{n1}(x) + F_{\hat{n}_0 x}(0) n^{-1/2} r_{n2},$$

where $\xi \sim N(0, V_\xi)$, $V_\xi = F_\varepsilon(0)[1 - F_\varepsilon(0)]$,

$$n^{1/2}\left[\frac{\hat{n}_0}{n} - F_\varepsilon(0)\right] \to^d \xi,$$

$$r_{n1}(x) = \hat{n}_0^{1/2}[F_{\hat{n}_0 x}(x) - F_x(x)] - Z_x^0(x),$$

and

$$r_{n2} = n^{1/2}\left[\frac{\hat{n}_0}{n} - F_U(0)\right] - \xi.$$

By Theorem 3 of Komlós, Major, and Tusnády (1975), there are a version $\tilde{r}_{n1}(x)$ of $r_{n1}(x)$ and constants C_4 and C_5 such that

$$P\left(\sup_{x \in \mathcal{S}} |\tilde{r}_{n1}(x)| > C_4 n^{-1/2} \log n\right) < n^{-C_5}.$$

A similar result applies to $r_{n2}(x)$. Therefore, $(n/h_0)^{1/2} T_{n2b}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2} T_{n2b}^{(1)}(x, h_0)$ in the sense that there are finite constants C_6 and C_7 such that

$$\text{(A.11)} \quad P\left[\sup_{x \in \mathcal{S}} |(n/h_0)^{1/2} T_{n2b}(x) - (n/h_0)^{1/2} T_{n2b}^{(1)}(x, h_0)| > C_6 (nh_0)^{-1/2} \log n\right] < n^{-C_7}.$$

It follows from Theorem 2.37 of Pollard (1984) that

(A.12) $\left(\frac{n}{h_0}\right)^{1/2} T_{n2c}(x) =^{a.s.} o[h_0^{3/2}(\log n)^{1/2+\delta}]$

uniformly over $x \in \mathcal{S}$ for any $\delta > 0$. Combining (A.10)-(A.12) yields the result that $(n/h_0)^{1/2} T_{n2}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2}[T_{n2a}^{(1)}(x,h_0) + T_{n2b}^{(1)}(x,h_0)]$. The sample paths of this process are uniformly continuous in h_0 (Dudley 1967). A straightforward but lengthy calculation shows that the covariance function of this process converges to

(A.13) $C(x_1,x_2) = \left(\frac{1-\tau}{\tau}\right) f_X(x_1) \int K(\zeta) K(\zeta+\delta) d\zeta$,

where $\delta = (x_1 - x_2)/h_0$. Let $W_1(\cdot)$ denote the mean-zero Gaussian process whose covariance function is $C(x_1,x_2)/C(x_1,x_1)$. Then it follows from Theorem 5.8 of Boucheron, Lugosi, and Massart (2013) and criterion B of Loève (1978, p. 268) that for any $\eta > 0$

(A.14) $\lim_{n\to\infty} P\left\{ \sup_{x\in\mathcal{S}} \left| \left(\frac{n}{h_0}\right)^{1/2} T_{n2}(x) - \left[f_X(x) \left(\frac{1-\tau}{\tau}\right) B_K \right]^{1/2} W_1\left(\frac{x}{h_0}\right) \right| > \eta \right\} = 0$.

Analysis of T_{n3}. We have

$$T_{n3}(x, \boldsymbol{b}_x) = \mathcal{P}\{I[Y - b_{x0} - b_{x1}(X-x) \le 0] - I[Y - g_x - g_x'(X-x) \le 0]\} K_{h_0}(X-x)$$

(A.15) $\qquad = \int \{F_\varepsilon[b_{x0} - g_z + b_{x1}(z-x)] - F_\varepsilon[g_x - g_z + g_x'(z-x)]\} K_{h_0}(z-x) f_X(z) dz.$

Suppose there is a constant $C_1 < \infty$ such that

(A.16a) $\qquad \sup_{x\in\mathcal{S}} |b_{x0} - g_x| \le C\left(\frac{\log n}{nh_0}\right)^{1/2}$

and

(A.16b) $$\sup_{x\in\mathcal{S}} | b_{x1} - g'_x | \le C_2 \left(\frac{\log n}{n h_0^3} \right)^{1/2}.$$

Define

$$\mathcal{V} = \{ \boldsymbol{b}_x : \text{(A.16a) and (A.16b) hold for all } x \in \mathcal{S} \}.$$

Then the change of variables $\xi = (z-x)/h_0$ and Taylor series expansions about zero of the F_ε terms in the integral on the right-hand side of (A.15) yield

(A.17) $$\sup_{x\in\mathcal{S};\, \boldsymbol{b}_x\in\mathcal{V}} | T_{n3}(x, \boldsymbol{b}_x) - h_0 f_\varepsilon(0) f_X(x)(b_0 - g_x) | \le C_2 \left(\frac{\log n}{n} \right)$$

for some constant $C_2 < \infty$ and all sufficiently large n. It follows from Proposition 2 of Guerre and Sabbah (2012) that

(A.18) $$\sup_{x\in\mathcal{S}} | \hat{g}_x - g_x | = O_p\left[\left(\frac{\log n}{n h_0} \right)^{1/2} \right]$$

and

(A.19) $$\sup_{x\in\mathcal{S}} | \hat{g}'_x - g'_x | = O_p\left[\left(\frac{\log n}{n h_0^3} \right)^{1/2} \right].$$

Let $\hat{\boldsymbol{b}}_x = (\hat{g}_x, \hat{g}'_x)$. Then (A.17)-(A.19) imply that

(A.20) $$\sup_{x\in\mathcal{S}} | T_{n3}(x, \hat{\boldsymbol{b}}_x) - h_0 f_\varepsilon(0) f_X(x)(\hat{b}_0 - g_x) | = O_p\left[\left(\frac{\log n}{n} \right) \right].$$

Analysis of T_{n4}. We have

$$\left\{ I[Y - b_{x0} - b_{x1}(X-x) \le 0] - I[Y - g_x - g'_x(X-x) \le 0] \right\} K_{h_0}(X-x)$$

$$= \{ I[g_x - g_X + g'_x(X-x) < \varepsilon \le b_{x0} - g_X - b_{x1}(X-x)]$$

$$- I[b_{x0} - g_X - b_{x1}(X-x) < \varepsilon \le g_x - g_X + g'_x(X-x)] \} K_{h_0}(X-x).$$

Let (A.15) and (A.16) hold. Then,

$$\mathcal{P}\left(\left\{I[Y-b_{x0}-b_{x1}(X-x)\le 0]-I[Y-g_x-g_x'(X-x)\le 0]\right\}^2 K_{h_0}(X-x)^2\right)\le C\left(\frac{h_0\log n}{n}\right)$$

for some $C<\infty$ and all sufficiently large n. It follows from Theorem (2.37) of Pollard (1984) that

$$\sup_{x\in\mathcal{S},\,\boldsymbol{b}_x\in\mathcal{V}} |T_{n4}(x,\boldsymbol{b}_x)| \ll h_0^{1/4}\left(\frac{\log n}{n}\right)^{3/4}$$

almost surely. Therefore, it follows from (A.18) and (A.19) that

(A.21) $$T_{n4}(x,\hat{\boldsymbol{b}}_x)=O_p\left[h_0^{1/4}\left(\frac{\log n}{n}\right)^{3/4}\right].$$

Now combine (A.9), (A.14), (A.20), and (A.21) to obtain

$$\mathcal{P}_n\left\{\tau-I[Y-\hat{b}_{x0}-\hat{b}_{x1}(X-x)\le 0]\right\}K_{h_0}(X-x)=\frac{h_0^3\kappa_2}{2}f_X(x)f_\varepsilon(0)g_x''$$

$$+\left(\frac{h_0}{n}\right)^{1/2}\left[f_X(x)\left(\frac{1-\tau}{\tau}\right)B_K\right]^{1/2}W_1\left(\frac{x}{h_0}\right)+h_0 f_\varepsilon(0)f_X(x)(\hat{g}_x-g_x)|+O\left(h_0^4\right)$$

$$+O_p\left[\left(\frac{\log n}{n}\right)\right]+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]+O_p\left[h_0^{1/4}\left(\frac{\log n}{n}\right)^{3/4}\right]$$

(A.22) $$=\frac{h_0^3\kappa_2}{2}f_X(x)f_\varepsilon(0)g_x''+\left(\frac{h_0}{n}\right)^{1/2}\left[f_X(x)\left(\frac{1-\tau}{\tau}\right)B_K\right]^{1/2}W_1\left(\frac{x}{h_0}\right)$$

$$+h_0 f_\varepsilon(0)f_X(x)(\hat{g}_x-g_x)|+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x\in\mathcal{S}$. The theorem follows from combining (A.8) and (A.22). Q.E.D.

<u>Proof of Theorem 3.2</u>: Define $\tilde{T}_n(x)=E^*T_{nb}^*(x)$. By definition,

$$\hat{\varepsilon}_j = \tilde{\varepsilon}_j - q_n = \varepsilon_i + g_{X_j} - \hat{g}_{X_j} - q_n .$$

Then conditional on the original data,

$$\tilde{T}_n(x) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}[1-\tau^{-1}I(\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_X + \hat{g}'_x(X_i - x) + q_n)]K_{h_0}(X_i - x) .$$

By construction,

$$n^{-1}\sum_{j=1}^{n}[1-\tau^{-1}I(\hat{\varepsilon}_j \le 0)] = O_p(n^{-1}) ,$$

so

$$\tilde{T}_n(x) = -\frac{1}{\tau n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{I[\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n]$$
$$- I(\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1})$$
$$= -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{n(\varepsilon,Z)}\{-I[\varepsilon \le \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n]$$
$$- I(\varepsilon \le \hat{g}_Z - g_Z + q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1}),$$

where $\mathcal{P}_{n(\varepsilon,Z)}$ is the empirical measure of (ε, X) and this notation is used instead of $\mathcal{P}_{n(\varepsilon,X)}$ to avoid confusion with the data $\{X_j : j = 1,...,n\}$. Define

$$v = \tilde{\varepsilon} - (\hat{g}_Z - g_Z) ,$$

and let $\mathcal{P}_v$ and $\mathcal{P}_{nv}$, respectively, denote the population and empirical measures of v. Then,

$$\tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{nv}\{-I[v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1})$$

$$\text{(A.23)} \qquad = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{v}\{-I[v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \le q_n)\}K_{h_0}(X_i - x)$$

$$+\frac{1}{\tau n}\sum_{i=1}^{n}(\mathcal{P}_{nv} - \mathcal{P}_{v})\{-I[v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1}).$$

The summands on the right-hand side of (A.23) are non-zero only if $|X_i - x| \le h_n$, $g(x)$ is continuous, and $\hat{g}(x) - g(x) \to^p 0$ uniformly over $x \in \mathcal{S}$ by Proposition 2 of Guerre and Sabbah (2012). In addition, the empirical process $\varphi_n(t) = (\mathcal{P}_{nv} - \mathcal{P}_v)I(v \le t)$ is stochastically equicontinuous. Therefore, the second term on the right-hand side of (A.23) is $O_p(h_0 n^{-1/2})$, and

$$\tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{v}\{-I[v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2}).$$

Because ε and X are independent,

$$\text{(A.24)} \qquad \tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_Z\mathcal{P}_\varepsilon\{-I[\varepsilon \le \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n]$$

$$- I(\varepsilon \le \hat{g}_Z - g_Z + q_n)\}K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2}).$$

Define

$$\hat{A}_1(x, X_i, Z) = \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n$$

and

$$\hat{A}_2(Z) = \hat{g}_Z - g_Z + q_n .$$

We have $|\hat{g}(x) - g(x)| = O_p[(nh_0)^{-1/2}(\log n)^{1/2}]$ and $|\hat{g}'(x) - g'(x)| = O_p[(nh_0^3)^{-1/2}(\log n)^{1/2}]$ uniformly over $x \in \mathcal{S}$. Moreover, in the summand on the right-hand side of (A.24), only terms for which $|X_i - x| \le h_0$ are non-zero. Therefore, arguments like those used to obtain (A.24) show that

$$\tilde{T}_n(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}\mathcal{P}_Z[\hat{A}_1(x, X_i, Z) - \hat{A}_2(Z)]K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2})$$

$$\equiv \tilde{T}_{na}(x) + \tilde{T}_{nb}(x) + O_p(h_0 n^{-1/2}),$$

where

$$\tilde{T}_{na}(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[g_x - g_{X_i} + g'_x(X_i - x)]K_{h_0}(X_i - x)$$

and

$$\tilde{T}_{nb}(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[(\hat{g}_x - g_x) - (\hat{g}_{X_i} - g_{X_i}) + (\hat{g}'_x - g'_x)(X_i - x)]K_{h_0}(X_i - x)\,.$$

Standard calculations for kernel estimators show that

$$\tilde{T}_{na}(x) = -\frac{\kappa_2 h_0^3 f_\varepsilon(0) g''(x) f_X(x)}{2\tau} + O(h_0^4)\,,$$

uniformly over $x \in \mathcal{S}$ and

(A.25) $$\left(\frac{n}{h_0}\right)^{1/2}\tilde{T}_{na} = -\frac{\kappa_2 d_0^{5/2} f_\varepsilon(0) g''(x) f_X(x)}{2\tau} + O(h_0)$$

uniformly over $x \in \mathcal{S}$.

Now consider $\tilde{T}_{nb}(x)$. It follows from Theorem 3.1 that

$$(\hat{g}_x - g_x) - (\hat{g}_{X_i} - g_{X_i}) = \frac{d_0^{5/2}\kappa_2}{2}(nh_0)^{-1/2}[g''(x) - g''(X_i)] + (nh_0)^{-1/2}\psi_0(x)W_1\left(\frac{x}{h_0}\right)$$

$$-(nh_0)^{-1/2}\psi_0(X_i)W_1\left(\frac{X_i}{h_0}\right) + o_p[(nh_0)^{-1/2}].$$

Combining this result with assumption 4 yields

$$-\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[(\hat{g}_x - g_x)-(\hat{g}_{X_i}-g_{X_i})]K_{h_0}(X_i - x)$$

$$= -\frac{f_\varepsilon(0)}{\tau n}(nh_0)^{-1/2}\psi_0(x)W_1\left(\frac{x}{h_0}\right)\sum_{i=1}^{n}K_{h_0}(X_i - x)$$

$$+\frac{f_\varepsilon(0)}{\tau n}(nh_0)^{-1/2}\sum_{i=1}^{n}\psi_0(X_i)W_1\left(\frac{X_i}{h_0}\right)K_{h_0}(X_i - x)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

$$= -\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x \in \mathcal{S}$. Therefore,

$$\tilde{T}_{nb}(x) = -\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)$$

$$-\frac{f_\varepsilon(0)}{\tau n}(\hat{g}'_x - g'_x)\sum_{i=1}^{n}(X_i - x)K_{h_0}(X_i - x)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x \in \mathcal{S}$. In addition, $|\hat{g}'(x)-g'(x)| = O_p[(nh_0^3)^{-1/2}(\log n)^{1/2}]$ uniformly over $x \in \mathcal{S}$, and

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - x)K_{h_0}(X_i - x) \ll h_0^{3/2}\left(\frac{\log n}{n}\right)^{1/2} = o\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

almost surely uniformly over $x \in \mathcal{S}$ by Theorem 2.37 of Pollard (1984). Therefore,

$$\text{(A.26)}\quad \tilde{T}_{nb}(x) = -\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x \in \mathcal{S}$. Combining (A.25) and (A.26) yields

$$\left(\frac{n}{h_0}\right)^{1/2}\tilde{T}_n = \frac{\kappa_2 d_0^{5/2} f_\varepsilon(0) f_X(x)}{2\tau} g''(x) - \frac{f_\varepsilon(0) f_X(x)\psi_0(x)}{\tau} W_1\left(\frac{x}{h_0}\right) + o_p(1)$$

$$= \frac{A_1(x)\kappa_2}{2} g''(x) - A_2(x) W_1\left(\frac{x}{h_0}\right) + o_p(1)$$

uniformly over $x \in \mathcal{S}$. This proves part (i) of the theorem. Part (ii) is an immediate consequence of part (i), uniform consistency of $\hat{f}_X(x)$, and consistency of $\hat{f}_\varepsilon(0)$. Q.E.D.

Proof of Corollary 3.3: A Taylor series expansion of $\hat{\pi}(x,\alpha)$ about $\hat{\lambda}(x) = \lambda(x)$ yields

$$\hat{\pi}(x,\alpha) = \Phi[z_{1-\alpha/2} - \lambda(x) - \Delta(x)] - \Phi[-z_{1-\alpha/2} - \lambda(x) - \Delta(x)] + r_n(x)[\hat{\lambda}(x) - \lambda(x)],$$

where

$$r_n(x) = [\{\phi[z_{1-\alpha/2} - \tilde{\lambda}_1(x) + \Delta(x)] - \phi[-z_{1-\alpha/2} + \tilde{\lambda}_2(x) - \Delta(x)]\}$$

and $\tilde{\lambda}_1(x)$ and $\tilde{\lambda}_2(x)$ are between $\hat{\lambda}(x)$ and $\lambda(x)$. The corollary now follows from Theorem 3.2(ii) and boundedness of $r_n(x)$. Q.E.D.

A.3 Proofs of Theorems 3.4 and 3.5

Proof of Theorem 3.4: Part (i) follows from Corollary 3.3. The process $\Delta(\cdot)$ is a non-stochastic multiple of W_1 and has uniformly continuous sample paths (Dudley 1967). Parts (ii) and (iii) of the theorem follow from arguments identical to those used to prove results (4.12) and (4.13) of HH. Q.E.D.

Proof of Theorem 3.5: It suffices to show that asymptotically, $\hat{\lambda}_{\max} \geq \max_{x\in\mathcal{S}} \lambda(x)$ and $\hat{\lambda}_{\min} \leq \min_{x\in\mathcal{S}} \lambda(x)$. We prove that $\hat{\lambda}_{\max} \geq \max_{x\in\mathcal{S}} \lambda(x)$ asymptotically. The proof for $\hat{\lambda}_{\min}$ is similar.

To show that $\hat{\lambda}_{\max} \ge \max_{x \in \mathcal{S}} \lambda(x)$ asymptotically, observe that $\lambda(x)$ is a continuous function on the compact interval $\mathcal{S}$. Therefore, there is a point $x^* \in \mathcal{S}$ such that $\max_{x \in \mathcal{S}} \lambda(x) = \lambda(x^*)$. Assume that x^* is unique. The proof for a unique x^* holds with minor modifications if x^* is not unique. Given any $\varepsilon > 0$, choose $\delta > 0$ so that

$$| \lambda(x) - \lambda(x^*) | < \varepsilon$$

whenever

$$| x - x^* | \le \delta .$$

Because

$$\sup_{x \in \mathcal{S}} \hat{\lambda}(x) \ge \sup_{|x - x^*| \le \delta} \hat{\lambda}(x) ,$$

it suffices to show that

$$\sup_{|x - x^*| \le \delta} \hat{\lambda}(x) \ge \lambda(x^*) .$$

By Theorem 3.2(ii)

$$\hat{\lambda}(x) = \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \Delta(x) .$$

If $x \in [x^* - \delta, x^* + \delta]$, then

$$\Delta(x) > \varepsilon \Rightarrow \hat{\lambda}(x) = \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \Delta(x) > \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \varepsilon > \frac{\beta(x^*)}{\sigma_{\hat{g}}(x^*)} = \lambda(x^*) .$$

Therefore, $\Delta(x) > \varepsilon$ for some $x \in [x^* - \delta, x^* + \delta]$ implies that $\hat{\lambda}_{\max} \ge \lambda(x^*)$. To prove that $\Delta(x) > \varepsilon$ for some $x \in [x^* - \delta, x^* + \delta]$, let $x_0, ..., x_{J_n}$ be a set of points such that

$$x^* - \delta = x_0 < x_1 < ... < x_{J_n} = x^* + \delta .$$

Let $x_j - x_{j-1} > 2h_0$ for each $j = 1, ..., J_n$ and $J_n \to \infty$ as $n \to \infty$. This is possible because $h_0 \to 0$ and δ remains fixed as $n \to \infty$. Then $\Delta(x_0), ..., \Delta(x_{J_n})$ are independent random variables that are normally distributed with means of 0 and variances that are bounded away from 0 as $n \to \infty$. Let $Var[\Delta(x)] \geq \sigma_{\min}^2 > 0$, Then as $n \to \infty$,

$$P\left\{\bigcap_{j=0}^{J_n} [\Delta(x_j) \leq \varepsilon]\right\} \leq [\Phi(\varepsilon / \sigma_{\min})]^{J_n+1} \to 0,$$

and

$P[\hat{\lambda}_{\max} \geq \lambda(x^*)] \to 1$. Q.E.D.

REFERENCES

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, U.K.: Oxford University Press.

Dudley, R.M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis* **1**, 290-330.

Guerre, E. and Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* **28**, 87-129.

Komlós, J., Major,P., and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s and the sample DF. I. *Zeitschrift fur Warscheinlichkeitstheorie und Verwand Gebiete* **32**, 111-131.

Loève, M. (1978). *Probability Theory II*. New York: Springer-Verlag.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* **75**, 828-838.