SUPPLEMENTARY MATERIALS

Proofs of Theorems

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

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

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

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

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 $g(x)$.

Lemma A.1: 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, $\hat{b}_{jh} - \hat{b}_{jh_0} = o_p (n^{-2/5})$.
Proof: Let $D_1$ be as in assumption 6(ii). Define

$$a_n = n^{-D_1} \left( \frac{\log n}{nh_0} \right)^{1/2},$$

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

$$a_n = \min_{(h_0, h_1) \in G \cap O} Q_n(x, b_0, b_1, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0).$$

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

(A.1) \( \sup_{(h_0, h_1) \in G} |Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0)| < a_n / 2 \),

then $\hat{b}_h = (\hat{b}_{0h_0}, \hat{b}_{1h_0}) \in O$. 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 $\hat{b}_h \in O$.

Step 1: Let $A_n$ be the event

$$\sup_{(h_0, h_1) \in G} |Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0)| < a_n / 2.$$

Then

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

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

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

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

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

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

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

\[
Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) = \frac{h_0}{h} \sum_{i=1}^{n} \rho \left[ Y_i - b_0 - b_1(X_i - x) \right] K' \left( \frac{X_i - x}{\hat{h}} \right) \left( h - h_0 \right) \\
= \frac{h_0}{h^2} \mathcal{P} \left\{ \rho \left[ Y_i - b_0 - b_1(X_i - x) \right] K' \left( \frac{X_i - x}{\hat{h}} \right) \left( h - h_0 \right) \right\} \\
+ \frac{h_0}{h^2} (\mathcal{P}_n - \mathcal{P}) \left\{ \rho \left[ Y_i - b_0 - b_1(X_i - x) \right] K' \left( \frac{X_i - x}{\hat{h}} \right) \left( h - h_0 \right) \right\},
\]

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

\[
\frac{h_0}{h^2} (\mathcal{P}_n - \mathcal{P}) \left\{ \rho \left[ Y_i - b_0 - b_1(X_i - x) \right] K' \left( \frac{X_i - x}{\hat{h}} \right) \left( h - h_0 \right) \right\} = \text{a.s.} \left[ n^{-D_1} \left( \frac{\log n}{nh_0} \right)^{1/2} \right]
\]

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

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

Therefore,

\[
| Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) | < \frac{a_n}{2}
\]

almost surely for all sufficiently large \( n \).
Step 3: It follows from Theorem 2.37 of Pollard (1984) that
\[ Q_n(x,b_0,b_1,h_0) - Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0) = \ell(x)(b_0 - \hat{b}_{0h_0}) + o(1), \]
where \( \ell \) is a non-zero function that does not depend on \( n \). Therefore, \( \hat{b}_h \in \mathcal{O} \) 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

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

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

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

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

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

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

\[ T_{n1}(x) = \mathcal{P} \left\{ \tau - I[Y - g_x - g'_x (X - x) \leq 0] \right\} K_{h_0}(X - x), \]
\( T_{n2}(x) = (P_n - P) \{ \tau - I[Y - g_x - g'_x(X - x) \leq 0] \} K_{h_0}(X - x), \)

\( T_{n3}(x, b_x) = P \{ I[Y - b_{x0} - b_{x1}(X - x) \leq 0] - I[Y - g_x - g'_x(X - x) \leq 0] \} K_{h_0}(X - x), \)

and

\( T_{n4}(x, b_x) = (P_n - P) \{ I[Y - b_{x0} - b_{x1}(X - x) \leq 0] - I[Y - g_x - g'_x(X - x) \leq 0] \} K_{h_0}(X - x). \)

In these definitions, \( b_x \) is an arbitrary, non-stochastic vector. Then

\[
T_n \{ \tau - I[Y - b_{x0} - b_{x1}(X - x) \leq 0] \} K_{h_0}(X - x) = T_{n1}(x) + T_{n2}(x) + T_{n3}(x, b_x) + T_{n4}(x, b_x).
\]

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

**Analysis of \( T_{n1}(x):** Let \( F_\varepsilon \) denote the distribution function of \( \varepsilon \) in (1.1). Then a Taylor series expansion yields

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

\[
= \frac{h_0^3 k_2}{2} f_x(x) f_\varepsilon(0) g''_x + O(h_0^4)
\]

uniformly over \( x \in \mathcal{S} \).

**Analysis of \( T_{n2}(x):** Define

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

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

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

\[ -E\{ I[\epsilon_i \leq g_x - g_{X_i} + g'_x(X_i - x)] - I(\epsilon_i \leq 0) \} K_{h_0}(X_i - x) \} \).

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 \( Z^0_x(x) \) by \( Z_{nx}(x) \sim Z^0_x(x) \) as \( n \to \infty \). Then a change of variables and integration by parts yields

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

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

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

\[ \equiv n^{1/2} T^{(1)}_{n2a}(x, h_0) + n^{1/2} T^{(2)}_{n2a}(x, h_0). \]

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

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

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

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

\[
n^{1/2}T_{n2b}(x) = \int I(s \leq 0)K\left(\frac{v-x}{h_0}\right)dZ_{nex}(s,v) = -\int Z_{ex}^0(0,x+h_0\xi)K'(\xi)d\xi - \int [Z_{nex}(0,x+h_0\xi) - Z_{ex}^0(0,x+h_0\xi)]K'(\xi)d\xi = n^{1/2}T_{n2b}^{(1)}(x,h_0) + n^{1/2}T_{n2b}^{(2)}(x,h_0).\]

To bound $n^{1/2}T_{n2b}^{(2)}$, let $\hat{n}_0$ denote the number of observations for which $\varepsilon_i \leq 0$. 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 $\varepsilon$ are independent. We have

\[
F_{nex}(0,x) = n^{-1} \sum_{i=1}^{n} I(\varepsilon_i \leq 0)I(X_i \leq x) = \frac{\hat{n}_0}{n} \hat{n}_0^{-1} \sum_{i=1}^{\hat{n}_0} I(X_i \leq x) = \frac{\hat{n}_0}{n} F_{\hat{n}_0,x}(x).
\]

Moreover, because $\varepsilon$ and $X$ are independent,
\[
F_{n\bar{e}}(0, x) - F_{\bar{e}}(0, x) = \frac{\hat{n}_0}{n} F_{\hat{n}_0, x}(x) - F_\bar{e}(0) F_\bar{e}(x)
\]

\[
= F_{n\bar{e}}(0)[F_{\hat{n}_0, x}(x) - F_\bar{e}(x)] + \left[\frac{\hat{n}_0}{n} - F_\bar{e}(0)\right] F_{\hat{n}_0, x}(x)
\]

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

where \(\xi \sim N(0, V_\xi)\), 
\[V_\xi = F_\bar{e}(0)[1 - F_\bar{e}(0)]\],

\[
n^{1/2} \left[\frac{\hat{n}_0}{n} - F_\bar{e}(0)\right] \xrightarrow{d} \xi,
\]

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

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 \(\hat{n}_{n1}(x)\) of \(r_{n1}(x)\) and constants \(C_4\) and \(C_5\) such that

\[
P\left(\sup_{x \in S} |\hat{n}_{n1}(x)| > C_4 n^{-1/2} \log n \right) < n^{-C_5}.
\]

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 \((n / h_0)^{1/2} T_{n2b}^{(l)}(x, h_0)\) in the sense that there are finite constants \(C_6\) and \(C_7\) such that

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

It follows from Theorem 2.37 of Pollard (1984) that
uniformly over $x \in S$ for any $\delta > 0$. Combining (A.10)-(A.12) yields the result that 

$\left(\frac{n}{h_0}\right)^{1/2} T_{n2}(x)$ can be approximated by the mean-zero Gaussian process 

$\left(\frac{n}{h_0}\right)^{1/2} \left[ T_{n2a}^{(1)}(x,h_0) + T_{n2b}^{(1)}(x,h_0) \right]$. 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

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

where $\delta = (x_1 - x_2)/h_0$. Let $W_1()$ 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 $\eta > 0$

\begin{equation}
(A.14) \quad \lim_{n \to \infty} P \left( \sup_{x \in 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.
\end{equation}

Analysis of $T_{n3}$. We have

$T_{n3}(x, b_x) = \mathcal{P}\{I[Y - b_{x_0} - b_{x_1}(X - x) \leq 0] - I[Y - g_x - g_x'(X - x) \leq 0]\} K_{h_0} (X - x)$

\begin{equation}
(A.15) \quad = \int \{ F_{\epsilon}[b_{x_0} - g_z + b_{x_1}(z - x)] - F_{\epsilon}[g_x - g_z + g_x'(z - x)] \} K_{h_0} (z - x) f_X(z) dz.
\end{equation}

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

\begin{equation}
(A.16a) \quad \sup_{x \in S} |b_{x_0} - g_x| \leq C \left(\frac{\log n}{nh_0}\right)^{1/2}
\end{equation}

and

9
Define

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

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

\[
(A.17) \sup_{x \in \mathcal{S}, \mathbf{b}_x \in \mathcal{V}} | T_{n3}(x, \mathbf{b}_x) - h_0 f_{\varepsilon}(0) f_X(x)(b_0 - g_x) | \leq C_2 \left( \frac{\log n}{n h_0^3} \right)
\]

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

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

and

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

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

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

**Analysis of** \( T_{n4} \). We have

\[
\{ I[Y - b_{x0} - b_{x1}(X - x) \leq 0] - I[Y - g_x - g_x'(X - x) \leq 0] \} K_{h_0}(X - x)
\]

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

\[
- I[b_{x0} - g_X - b_{x1}(X - x) < \varepsilon \leq g_x - g_X + g_x'(X - x)] \} K_{h_0}(X - x).
\]
Let (A.15) and (A.16) hold. Then,

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

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

\[ \sup_{x \in S, h \in \mathcal{V}} |T_{n4}(x, b_x)| \ll h_0^{1/4} \left( \frac{\log n}{n} \right)^{3/4} \]

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

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

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

\[ \mathcal{P}_n \left\{ \tau - I[Y - \hat{b}_{x0} - \hat{b}_{x1}(X - x) \leq 0] \right\} K_{h_0}(X - x) = \frac{h_0^3 K_2}{2} f_x(x) f_x(0) g_x^* \]

\[ + \left( \frac{h_0}{n} \right)^{1/2} \left[ f_x(x) \left( \frac{1 - \tau}{\tau} \right) B_K \right]^{1/2} W_1 \left( \frac{x}{h_0} \right) + h_0 f_x(0) f_x(x) (g_x - g_x) \right] + O \left( h_0^4 \right) \]

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

(A.22) \( = \frac{h_0^3 K_2}{2} f_x(x) f_x(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) \]

\[ + h_0 f_x(0) f_x(x) (g_x - g_x) \right] + o_p \left[ \left( \frac{h_0}{n} \right)^{1/2} \right] \]

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

**Proof of Theorem 3.2:** Define \( \tilde{T}_n(x) = E^* T_{nb}^*(x) \). By definition,
\[ \hat{e}_j = \tilde{e}_j - q_n = \varepsilon_i + g_{X_j} - \hat{g}_{X_j} - q_n. \]

Then conditional on the original data,

\[ \tilde{T}_n(x) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ 1 - \tau^{-1} I(\varepsilon_j \leq \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_{X_i} + \hat{g}_x'(X_i - x) + q_n) \right] K_h(X_i - x). \]

By construction,

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

so

\[ \tilde{T}_n(x) = -\frac{1}{\tau n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ I(\varepsilon_j \leq \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_{X_i} + \hat{g}_x'(X_i - x) + q_n) \right] \]

\[ - I(\varepsilon_j \leq \hat{g}_{X_j} - g_{X_j} + q_n) K_h(X_i - x) + O_p(n^{-1}) \]

\[ = -\frac{1}{\tau n} \sum_{i=1}^{n} P_{n(\varepsilon,Z)} \left[ -I(\varepsilon \leq \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_Z + \hat{g}_x'(X_i - x) + q_n) \right] \]

\[ - I(\varepsilon \leq \hat{g}_Z - g_Z + q_n) K_h(X_i - x) + O_p(n^{-1}), \]

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

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

and let \( P_v \) and \( P_{nv} \), respectively, denote the population and empirical measures of \( v \). Then,
\[
\hat{T}_n(x) = -\frac{1}{\tau n} \sum_{i=1}^{n} \mathcal{P}_v \{-I[v \leq \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \leq q_n)\} K_h(X_i - x) + O_p(n^{-1})
\]

(A.23) 

\[
= -\frac{1}{\tau n} \sum_{i=1}^{n} \mathcal{P}_v \{-I[v \leq \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \leq q_n)\} K_h(X_i - x)
\]

\[
+ \frac{1}{\tau n} \sum_{i=1}^{n} (\mathcal{P}_{n^2} - \mathcal{P}_v) \{-I[v \leq \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \leq q_n)\} K_h(X_i - x) + O_p(n^{-1}).
\]

The summands on the right-hand side of (A.23) are non-zero only if \(|X_i - x| \leq h_n\), \(g(x)\) is continuous, and \(\hat{g}(x) - g(x) \to^p 0\) uniformly over \(x \in S\) by Proposition 2 of Guerre and Sabbah (2012). In addition, the empirical process \(\varphi_n(t) = (\mathcal{P}_{n^2} - \mathcal{P}_v)I(v \leq t)\) is stochastically equicontinuous. Therefore, the second term on the right-hand side of (A.23) is \(O_p(h_0n^{-1/2})\), and

\[
\hat{T}_n(x) = -\frac{1}{\tau n} \sum_{i=1}^{n} \mathcal{P}_v \{-I[v \leq \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n] - I(v \leq q_n)\} K_h(X_i - x) + O_p(h_0n^{-1/2}).
\]

Because \(\varepsilon\) and \(X\) are independent,

(A.24) 

\[
\hat{T}_n(x) = -\frac{1}{\tau n} \sum_{i=1}^{n} \mathcal{P}_v \mathcal{P}_e \{-I[\varepsilon \leq \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n]
\]

\[-I(\varepsilon \leq \hat{g}_Z - g_Z + q_n)\} K_h(X_i - x) + O_p(h_0n^{-1/2}).
\]

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 \(x \in S\). Moreover, in the summand on the right-hand side of (A.24), only terms
for which \(|X_t - x| \leq h_0\) are non-zero. Therefore, arguments like those used to obtain (A.24) show that

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

\[
\equiv \tilde{T}_{na}(x) + \tilde{T}_{nb}(x) + O_{\rho}(h_0 n^{-1/2}),
\]

where

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

and

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

Standard calculations for kernel estimators show that

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

uniformly over \(x \in S\) and

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

uniformly over \(x \in S\).

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

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

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

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

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

uniformly over \( x \in S \). Therefore,

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

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

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

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

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

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

uniformly over \( x \in S \). Combining (A.25) and (A.26) yields
\[
\left( \frac{n}{h_0} \right)^{1/2} \tilde{T}_n = \frac{\kappa_2 \sigma_0^{5/2} f_\varepsilon(0) f_X(x)}{2\tau} g^*(x) - \frac{f_\varepsilon(0) f_X(x) \psi_0(x)}{\tau} W_1 \left( \frac{x}{h_0} \right) + o_p(1)
\]

\[
= \frac{A_1(x) \kappa_2}{2} g^*(x) - A_2(x) W_1 \left( \frac{x}{h_0} \right) + o_p(1)
\]

uniformly over \( x \in S \). 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

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

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

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

whenever

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

Because

$$\sup_{x \in S} \hat{\lambda}(x) \geq \sup_{|x-x^*| \leq \delta} \lambda(x),$$

it suffices to show that

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

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

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

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

$$x^* - \delta = x_0 < x_1 < \ldots < x_{J_n} = x^* + \delta.$$
Let \( x_j - x_{j-1} > 2h_0 \) for each \( j = 1, \ldots, J_n \) and \( J_n \to \infty \) as \( n \to \infty \). This is possible because \( h_0 \to 0 \) and \( \delta \) remains fixed as \( n \to \infty \). Then \( \Delta(x_0), \ldots, \Delta(x_J) \) 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 \( \text{Var}[\Delta(x)] \geq \sigma_{min}^2 > 0 \), Then as \( n \to \infty \),

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

and

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

**REFERENCES**


