**Supplemental Material to**
# A SIMPLE ADJUSTMENT FOR BANDWIDTH SNOOPING


**By**

**Timothy B. Armstrong and Michal Kolesár**



**December 2014**
**Revised June 2017**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1961R3S**



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.yale.edu/**

# A Simple Adjustment for Bandwidth Snooping: Supplemental Materials

Timothy B. Armstrong* Michal Kolesár†
Yale University Princeton University

June 28, 2017

This supplement is organized as follows. Section S1 contains auxiliary results used in Appendix A of the main text. Section S2 contains auxiliary results on local polynomial regression. Section S3 proves theorems in Appendix B. Section S4 derives critical values for one-sided confidence intervals and gives tables of one- and two-sided critical values. Finally, Section S5 presents the results of a Monte Carlo study.

The following additional notation, which is also used in the appendix in the main text, is used throughout this supplement. For a sample $\{Z_i\}_{i=1}^n$ and a function $f$ on the sample space, $E_n f(Z_i) = \frac{1}{n}\sum_{i=1}^n f(Z_i)$ denotes the sample mean, and $\mathbb{G}_n f(Z_i) = \sqrt{n}(E_n - E)f(Z_i) = \sqrt{n}[E_n f(Z_i) - Ef(Z_i)]$ denotes the empirical process. We use $t \vee t'$ and $t \wedge t'$ to denote element-wise maximum and minimum, respectively. We use $e_k$ to denote the $k$th basis vector in Euclidean space (where the dimension of the space is clear from context).

## S1 Auxiliary Results

This section contains auxiliary results that are used in the proof of Theorem 3.1 in Appendix A of the main text, and in the proofs of the results from Appendix B of the main text given later in this supplement.

*email: timothy.armstrong@yale.edu
†email: mkolesar@princeton.edu

## S1.1 Tail Bounds for Empirical Processes

We state some tail bounds based on an inequality of Talagrand (1996) and other empirical process results. Throughout this section, we consider a class of functions $\mathcal{G}$ on the sample space $\mathbb{R}^{d_Z}$ with an i.i.d. sample of random variables $Z_1, \ldots, Z_n$. We assume throughout that $\mathcal{G}$ has a polynomial covering number in the sense that, for some $B, W$, $N_1(\delta, Q, \mathcal{G}) \leq B\varepsilon^{-W}$ for all finitely discrete probability measures $Q$, where $N_1$ is defined in, e.g., Pollard (1984), p. 25.

**Lemma S1.1.** *Let $\tilde{\mathcal{G}}$ be a subset of $\mathcal{G}$ such that, for some envelope function $G$ and constant $\overline{g}$, $|g(Z_i)| \leq G(Z_i) \leq \overline{g}$ a.s. for all $g \in \tilde{\mathcal{G}}$. Then, for some constant $K$ that depends only on $\mathcal{G}$,*

$$P\left(\sup_{g\in\tilde{\mathcal{G}}} |\mathbb{G}_n g(Z_i)| \geq K\sqrt{E[G(Z_i)^2]} + t\right) \leq K\exp\left(-\frac{1}{K}\frac{t^2}{E[G(Z_i)^2] + \overline{g}\left\{\sqrt{E[G(Z_i)^2]} + t\right\}/\sqrt{n}}\right)$$

*Proof.* We apply a result of Talagrand (1996) as stated in Equation (3) of Massart (2000). The quantity $v$ from that version of the bound is, in our setting, given by $v = E\sup_{g\in\tilde{\mathcal{G}}}\sum_{i=1}^n [g(Z_i) - Eg(Z_i)]^2$ which, as shown in Massart (2000, p. 882), is bounded by (see also Klein and Rio, 2005)

$$n\sup_{g\in\tilde{\mathcal{G}}} E\{[g(Z_i) - Eg(Z_i)]^2\} + 32\overline{g}E\sup_{g\in\tilde{\mathcal{G}}}\sum_{i=1}^n [g(Z_i) - Eg(Z_i)].$$

By Theorem 2.14.1 in van der Vaart and Wellner (1996),

$$E\sup_{g\in\tilde{\mathcal{G}}}\sum_{i=1}^n [g(Z_i) - Eg(Z_i)] \leq \sqrt{n}K_1\sqrt{E[G(Z_i)]^2}, \tag{1}$$

for a constant $K_1$ that depends only on $\mathcal{G}$. Combined with the fact that $E\{[g(Z_i) - Eg(Z_i)]^2\} \leq E[G(Z_i)^2]$, this gives the bound

$$v \leq nE[G(Z_i)^2] + 32\overline{g}K_1\sqrt{n}\sqrt{E[G(Z_i)]^2}.$$

Applying the bound from equation (3) of Massart (2000) with these quantities gives

$$P\left(\sqrt{n}\sup_{g\in\tilde{\mathcal{G}}}\mathbb{G}_n g(Z_i) \geq K_1\sqrt{n}\sqrt{E[G(Z_i)]^2} + r\right)$$

$$\leq P\left(\sqrt{n}\sup_{g\in\tilde{\mathcal{G}}}\mathbb{G}_n g(Z_i)\geq E\sup_{g\in\tilde{\mathcal{G}}}\sum_{i=1}^{n}[g(Z_i)-Eg(Z_i)]+r\right)$$
$$\leq K_2\exp\left(-\frac{1}{K_2}\frac{r^2}{nE[G(Z_i)^2]+32\overline{g}K_1\sqrt{n}\sqrt{E[G(Z_i)]^2}+\overline{g}r}\right),$$

where the first inequality follows from (1). Substituting $r=\sqrt{n}t$ gives

$$P\left(\sup_{g\in\tilde{\mathcal{G}}}\mathbb{G}_n g(Z_i)\geq K_1\sqrt{E[G(Z_i)]^2}+t\right)$$
$$\leq K_2\exp\left(-\frac{1}{K_2}\frac{t^2}{E[G(Z_i)^2]+32\overline{g}K_1\sqrt{E[G(Z_i)]^2}/\sqrt{n}+\overline{g}t/\sqrt{n}}\right),$$

which gives the result after noting that replacing $K_1$ on the left hand side as well as $K_2$ and $32K_1K_2$ on the right hand side with a larger constant $K$ decreases the left hand side and increases the right hand side, and applying a symmetric bound to $\inf_{g\in\tilde{\mathcal{G}}}\mathbb{G}_n g(Z_i)$. $\square$

Lemma S1.1 gives good bounds for $t$ just larger than $\sqrt{E[G(Z_i)]^2}$, so long as $\sqrt{E[G(Z_i)]^2}/\sqrt{n}$ is small relative to $E[G(Z_i)]^2$ (i.e. so long as $E[G(Z_i)]^2 n$ is large). We now state a version of this result that is specialized to this case.

**Lemma S1.2.** *Let $\tilde{\mathcal{G}}$ be a subset of $\mathcal{G}$ such that, for some envelope function $G$ and constant $\overline{g}$, $|g(Z_i)|\leq G(Z_i)\leq\overline{g}$ a.s. for all $g\in\tilde{\mathcal{G}}$. Then, for some constant $K$ that depends only on $\mathcal{G}$,*

$$P\left(\sup_{g\in\tilde{\mathcal{G}}}|\mathbb{G}_n g(Z_i)|\geq\sqrt{V}a\right)\leq K\exp\left(-\frac{a^2}{K}\right)$$

*for all $V\geq E[G(Z_i)^2]$ and $a>0$ with $a+1\leq\sqrt{V}\sqrt{n}/\overline{g}$.*

*Proof.* Substituting $t=rV^{1/2}$ into the bound from Lemma S1.1 gives, letting $K_1$ be the constant $K$ from that lemma,

$$P\left(\sup_{g\in\tilde{\mathcal{G}}}|\mathbb{G}_n g(Z_i)|\geq(K_1+r)V^{1/2}\right)\leq K_1\exp\left(-\frac{1}{K_1}\frac{r^2V}{V+\overline{g}\left\{V^{1/2}+rV^{1/2}\right\}/\sqrt{n}}\right).$$

For $\overline{g}(1+r)\leq\sqrt{n}V^{1/2}$, this is bounded by $K_1\exp\left(-\frac{r^2}{2K_1}\right)$. Setting $a=K_1+r$ and noting that $K_1\exp\left(-\frac{(a-K_1)^2}{2K_1}\right)\leq K_2\exp\left(-\frac{a^2}{K_2}\right)$ for a large enough constant $K_2$ (and that $\overline{g}(1+a)\leq\sqrt{n}V^{1/2}$

implies $\overline{g}(1+a-K_1) \le \sqrt{n}V^{1/2})$ gives the result. $\square$

## S1.2 Tail Bounds for Kernel Estimators

We specialize some of the results of Section S1.1 to our setting. We are interested in functions of the form $g(x,w) = f(w,h,t)k(x/h)$, where $h$ varies over positive real numbers and $t$ varies over some index set $T$.

We assume throughout the section that $k(x)$ is a bounded kernel function with support $[-A,A]$, with $k(x) \le B_k < \infty$ for all $k$. We also assume that $X_i$ is a real valued random variable with with a density $f_X(x)$ with $f_X(x) \le \overline{f}_X < \infty$ all $x$.

**Lemma S1.3.** *Suppose that $\{(x,w) \mapsto f(w,h,t)k(x/h) | 0 \le h \le \overline{h}, t \in T\}$ is contained in some larger class $\mathcal{G}$ with polynomial covering number, and that, for some constant $B_f$, $|f(W_i,h,t)k(X_i/h)| \le B_f$ for all $h \le \overline{h}$ and $t \in T$ with probability one. Then, for some constant $K$ that depends only on $\mathcal{G}$,*

$$P\left(\sup_{0\le h\le \overline{h}, t\in T} |\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \ge aB_f A^{1/2}\overline{f}_X^{1/2}\overline{h}^{1/2}\right) \le K\exp(-\frac{a^2}{K})$$

*for all $a > 0$ with $a+1 \le A^{1/2}\overline{f}_X^{1/2}\overline{h}^{1/2}n^{1/2}$.*

*Proof.* The result follows from Lemma S1.2, since $B_f I(|X_i| \le A\overline{h})$ is an envelope function for $f(W_i,h,t)k(X_i/h)$ as $h$ and $t$ vary over this set. $\square$

**Lemma S1.4.** *Suppose that the conditions of Lemma S1.3 hold, and let $a(h) = 2\sqrt{K\log\log(1/h)}$ where $K$ is the constant from Lemma S1.3. Then, for a constant $\varepsilon > 0$ that depends only on $K$, $A$ and $\overline{f}_X$,*

$$P\left(|\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \ge a(h)h^{1/2}B_f A^{1/2}\overline{f}_X^{1/2} \text{ some } (\log\log n)/(\varepsilon n) \le h \le \overline{h},\ t\in T\right)$$
$$\le K(\log 2)^{-2}\sum_{(2\overline{h})^{-1}\le 2^k\le\infty} k^{-2}.$$

*Proof.* Let $\mathcal{H}^k = (2^{-(k+1)}, 2^{-k})$. Applying Lemma S1.3 to this set, we have

$$P\left(|\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \ge a(h)h^{1/2}B_f A^{1/2}\overline{f}_X^{1/2} \text{ some } h\in\mathcal{H}^k,\ t\in T\right)$$
$$\le P\left(\sup_{0\le h\le 2^k, t\in T} |\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \ge a(2^{-k})2^{-(k+1)/2}B_f A^{1/2}\overline{f}_X^{1/2}\right)$$

$$\leq K \exp\left(-\frac{[a(2^{-k})2^{-1/2}]^2}{K}\right) = K\exp\left(-2\log\log 2^k\right) = K\exp\left(-2\log(k\log 2)\right) = K[k\log 2]^{-2}$$

so long as $2^{-1/2}a(2^{-k})+1 \leq A^{1/2}\overline{f}_X^{-1/2}2^{-k/2}n^{1/2}$, where the first inequality follows since $a(h) \geq a(2^{-k})$ and $h \geq 2^{-(k+1)}$ for $h \in \mathcal{H}^k$.

Now, $2^{-1/2}a(2^{-k})+1 \leq A^{1/2}\overline{f}_X^{-1/2}2^{-k/2}n^{1/2}$ will hold iff. $[2^{-1/2}a(2^{-k})+1]2^{k/2} \leq A^{1/2}\overline{f}_X^{-1/2}n^{1/2}$. If $2^k \leq \varepsilon n/\log\log n$ for some $\varepsilon > 0$, we will have $a(2^{-k}) \leq 2\sqrt{K\log\log[\varepsilon n/\log\log n]}$, so that $[2^{-1/2}a(2^{-k})+1]2^{k/2} \leq \{2^{-1/2}\cdot 2\sqrt{K\log\log[\varepsilon n/\log\log n]}+1\}\sqrt{\varepsilon n/\log\log n}$. For large enough $n$, this is bounded by $4\sqrt{K\varepsilon n}$, which is less than $A^{1/2}\overline{f}_X^{-1/2}n^{1/2}$ for $\varepsilon$ small enough as required.

Thus, for $\varepsilon$ defined above,

$$\begin{aligned}
&P\left(|\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \geq a(h)h^{1/2}B_f A^{1/2}\overline{f}_X^{-1/2} \text{ some } (\log\log n)/(\varepsilon n) \leq h \leq \overline{h},\, t\in T\right)\\
&\leq \sum_{(2\overline{h})^{-1}\leq 2^k\leq 2\varepsilon n/\log\log n} P\left(\sup_{0\leq h\leq 2^k, t\in T}|\mathbb{G}_n f(W_i,h,t)k(X_i/h)| \geq a(2^{-k})2^{-(k+1)/2}B_f A^{1/2}\overline{f}_X^{-1/2}\right)\\
&\qquad\qquad\qquad\qquad \leq K(\log 2)^{-2}\sum_{(2\overline{h})^{-1}\leq 2^k\leq 2\varepsilon n/\log\log n} k^{-2},
\end{aligned}$$

which gives the result. $\square$

Using these bounds, we obtain the following uniform bound on $\mathbb{G}_n f(W_i,h,t)k(X_i/h)$.

**Lemma S1.5.** *Under the conditions of Lemma S1.4,*

$$\sup_{(\log\log n)/(\varepsilon n)\leq h\leq\overline{h}, t\in T}\frac{|\mathbb{G}_n f(W_i,h,t)k(X_i/h)|}{(\log\log h^{-1})^{1/2}h^{1/2}} = \mathcal{O}_P(1).$$

*Proof.* Given $\varepsilon > 0$, we can apply Lemma S1.4 to find a $\delta > 0$ such that

$$\sup_{(\log\log n)/(\varepsilon n)\leq h\leq\delta, t\in T}\frac{|\mathbb{G}_n f(W_i,h,t)k(X_i/h)|}{(\log\log h^{-1})^{1/2}h^{1/2}} < 2\sqrt{2K}B_f A^{1/2}\overline{f}_X^{-1/2}$$

with probability at least $1-K(\log 2)^{-2}\sum_{(2\delta)^{-1}\leq 2^k\leq\infty}k^{-2} > 1-\varepsilon/2$. For this choice of $\delta$,

$$\sup_{\delta\leq h\leq\overline{h}, t\in T}\frac{|\mathbb{G}_n f(W_i,h,t)k(X_i/h)|}{(\log\log h^{-1})^{1/2}h^{1/2}} = \mathcal{O}_P(1)$$

by Lemma S1.3. Thus, choosing $C$ large enough so that $C \geq 2\sqrt{2K}B_f A^{1/2}\overline{f}_X^{-1/2}$ and

$$\sup_{\delta \leq h \leq \overline{h}, t \in T} \frac{|\mathbb{G}_n f(W_i, h, t) k(X_i/h)|}{(\log\log h^{-1})^{1/2} h^{1/2}} \leq C$$

with probability at least $1 - \varepsilon/2$ asymptotically, we have

$$\sup_{(\log\log n)/(\varepsilon n) \leq h \leq \overline{h}, t \in T} \frac{|\mathbb{G}_n f(W_i, h, t) k(X_i/h)|}{(\log\log h^{-1})^{1/2} h^{1/2}} \leq C$$

with probability at least $1 - \varepsilon$ asymptotically. □

### S1.3 Gaussian Approximation

This section proves Theorem A.2 in Appendix A.4, which gives a Gaussian process approximation for the process $\hat{\mathbb{H}}_n(h)$ defined in that section.

For convenience, we repeat the setup here. We show that $\frac{1}{\sqrt{h}}\mathbb{G}_n \tilde{Y}_i k(X_i/h) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{Y}_i k(X_i/h)$ is approximated by a Gaussian process with the same covariance kernel. We consider a general setup with $\{(\tilde{X}_i, \tilde{Y}_i)\}_{i=1}^n$ i.i.d., with $\tilde{X}_i \geq 0$ a.s. such that $\tilde{X}_i$ has a density $f_{\tilde{X}}(x)$ on $[0, \overline{x}]$ for some $\overline{x} \geq 0$, with $f_{\tilde{X}}(x)$ bounded away from zero and infinity on this set. We assume that $\tilde{Y}_i$ is bounded almost surely, with $E(\tilde{Y}_i|\tilde{X}_i) = 0$ and $var(\tilde{Y}_i|\tilde{X}_i = x) = f_{\tilde{X}}(x)^{-1}$. We assume that the kernel function $k$ has finite support $[0, A]$ and is differentiable on its support with bounded derivative. For ease of notation, we assume in this section that $\int k(u)^2\, du = 1$. The result applies to our setup with $\tilde{Y}_i$ given in (10) in Appendix A in the main text and $\tilde{X}_i$ given by $|X_i|$.

Let

$$\hat{\mathbb{H}}_n(h) = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \tilde{Y}_i k(\tilde{X}_i/h).$$

**Theorem A.2.** Under the conditions above, there exists, for each $n$, a process $\mathbb{H}_n(h)$ such that, conditional on $(\tilde{X}_1, \ldots, \tilde{X}_n)$, $\mathbb{H}_n$ is a Gaussian process with covariance kernel

$$cov\left(\mathbb{H}_n(h), \mathbb{H}_n(h')\right) = \frac{1}{\sqrt{hh'}} \int k(x/h)k(x/h')\, dx$$

and

$$\sup_{\underline{h}_n \le h \le \overline{x}/A} \left|\hat{\mathbb{H}}_n(h) - \mathbb{H}_n(h)\right| = \mathcal{O}_P\left((n\underline{h}_n)^{-1/4}[\log(n\underline{h}_n)]^{1/2}\right)$$

for any sequence $\underline{h}_n$ with $n\underline{h}_n / \log\log \underline{h}_n^{-1} \to \infty$.

We now prove the result. Let $\hat{G}(x) = \frac{1}{n}\sum_{\tilde{X}_i \le x} \tilde{Y}_i$. With this notation, we can write the process $\hat{\mathbb{H}}_n(h)$ as

$$\hat{\mathbb{H}}_n(h) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \tilde{Y}_i k(\tilde{X}_i/h) = \frac{\sqrt{n}}{\sqrt{h}}\int k(x/h)\, d\hat{G}(x).$$

Let $\hat{g}(x) = \frac{1}{n}\sum_{\tilde{X}_i \le x} f_{\tilde{X}}(\tilde{X}_i)^{-1}$. In Lemma S1.6 below, a process $\mathbb{B}_n(t)$ is constructed that is a Brownian motion conditional on $\tilde{X}_1, \ldots, \tilde{X}_n$ such that $\mathbb{B}_n(n\hat{g}(x))$ is, with high probability conditional on $\tilde{X}_1, \ldots, \tilde{X}_n$, close to $n\hat{G}(x)$. By showing that $\hat{g}(x)$ is close to $x$ with high probability and using properties of the fluctuation of the Brownian motion, it is then shown that $\mathbb{B}_n(n\hat{g}(x))$ can be approximated by $\mathbb{B}_n(nx)$, so that $\hat{\mathbb{H}}_n(h)$ is approximated by the corresponding process with $\hat{G}(x)$ replaced by $\mathbb{B}_n(nx)/n$.

Formally, let $\mathbb{B}_n(t)$ be given by the (conditional) Brownian motion in Lemma S1.6 below, and define

$$\mathbb{H}_n(h) = \frac{1}{\sqrt{nh}}\int k(x/h)\, d\mathbb{B}_n(nx).$$

Note that $\mathbb{H}_n(h) = \frac{1}{\sqrt{h}}\int k(x/h)\, d\tilde{\mathbb{B}}_n(x)$ (where $\tilde{\mathbb{B}}_n(x) = \mathbb{B}_n(nx)/\sqrt{n}$ is another Brownian motion conditional on $\tilde{X}_1, \ldots, \tilde{X}_n$), so that, conditional on $(\tilde{X}_1, \ldots, \tilde{X}_n)$, $\mathbb{H}_n$ is a Gaussian process with the desired covariance kernel.

Let $R_{1,n}(x) = n\hat{G}(x) - \mathbb{B}_n(n\hat{g}(x))$ and $R_{2,n}(x) = \mathbb{B}_n(n\hat{g}(x)) - \mathbb{B}_n(nx)$. Then

$$\hat{\mathbb{H}}_n(h) - \mathbb{H}_n(h) = \frac{1}{\sqrt{nh}}\int k(x/h)\, dR_{1,n}(x) + \frac{1}{\sqrt{nh}}\int k(x/h)\, dR_{2,n}(x).$$

Using the integration by parts formula, we have, for $j = 1, 2$ and $Ah \le \overline{x}$,

$$\frac{1}{\sqrt{nh}}\int k(x/h)\, dR_{j,n}(x) = \frac{R_{j,n}(Ah)k(A)}{\sqrt{nh}} - \frac{1}{\sqrt{nh}}\int_{x=0}^{Ah} R_{j,n}(x)k'(x/h)\frac{1}{h}\, dx$$

The first term is bounded by $\frac{|R_{j,n}(Ah)|k(A)}{\sqrt{nh}}$, and the second term is bounded by

$$\frac{A}{\sqrt{nh}}\left(\sup_{0\leq x\leq Ah}\left|R_{j,n}(x)\right|\right)\left(\sup_{0\leq u\leq A}\left|k'(u)\right|\right)$$

(see Bickel and Rosenblatt, 1973, for a similar derivation). By boundedness of $k'(u)$, it follows that both terms are bounded by a constant times $\frac{1}{\sqrt{nh}}\sup_{0\leq x\leq Ah}\left|R_{j,n}(x)\right|$, so that

$$\sup_{\underline{h}_n\leq h\leq \overline{x}/A}\left|\hat{\mathbb{H}}_n(h)-\mathbb{H}_n(h)\right|\leq K\sup_{\underline{h}_n\leq h\leq \overline{x}/A}\sum_{j=1}^{2}\sup_{0\leq x\leq Ah}\frac{\left|R_{j,n}(x)\right|}{\sqrt{nh}}\leq K\sum_{j=1}^{2}\sup_{0\leq x\leq \overline{x}}\frac{\left|R_{j,n}(x)\right|}{\sqrt{n[(x/A)\vee \underline{h}_n]}}.$$

for some constant $K$. Thus, the result will follow if we can show that $\sup_{0\leq x\leq\overline{x}}\frac{|R_{1,n}(x)|}{\sqrt{n(x\vee\underline{h}_n)}}$ and $\sup_{0\leq x\leq\overline{x}}\frac{|R_{2,n}(x)|}{\sqrt{n(x\vee\underline{h}_n)}}$ converge to zero at the required rate.

We first construct $\mathbb{B}_n(t)$ and show that $\sup_{0\leq x\leq A\overline{x}/A}\frac{|R_{1,n}(x)|}{\sqrt{n(x\vee\underline{h}_n)}}$ converges to zero quickly enough with this construction, using an approximation of Sakhanenko. Denote the the empirical cdf of $\tilde{X}_i$ by $\hat{F}_{\tilde{X}}(x)=\frac{1}{n}\sum_{i=1}^{n}I(\tilde{X}_i\leq x)$, and let $\tilde{X}_{(k)}$ be the $k$th smallest value of $\tilde{X}_i$.

**Lemma S1.6.** *Under the conditions of Theorem A.2, one can construct variables $Z_1,\ldots,Z_n$ such that $Z_i|(\tilde{X}_1,\ldots,\tilde{X}_n)\sim N(0,f_{\tilde{X}}(\tilde{X}_i)^{-1})$ and*

$$P\left(\left|\sum_{\tilde{X}_i\leq x}Z_i-\sum_{\tilde{X}_i\leq x}\tilde{Y}_i\right|>K\log\left[n\hat{F}_{\tilde{X}}(x)+2\right]\ \textit{some}\ 0\leq x\leq\overline{x}\middle|\tilde{X}_1,\ldots,\tilde{X}_n\right)\leq\varepsilon(K)$$

*with probability one, where $\varepsilon(K)$ is a deterministic function with $\varepsilon(K)\to 0$ as $K\to\infty$.*

*Proof.* Using a result of Sakhanenko (1985) as stated in Theorem A of Shao (1995), we can construct $Z_1,\ldots,Z_n$ such that

$$E\exp\left(\lambda A\sup_{0\leq x\leq\tilde{X}_{(k)}}\left|\sum_{\tilde{X}_i\leq x}Z_i-\sum_{\tilde{X}_i\leq x}\tilde{Y}_i\right|\middle|\tilde{X}_1,\ldots,\tilde{X}_n\right)\leq 1+\lambda\sum_{\tilde{X}_i\leq\tilde{X}_{(k)}}f_{\tilde{X}}(\tilde{X}_i)^{-1}$$

where $A$ is a universal constant and $\lambda$ is any constant such that $\lambda E[\exp(\lambda|\tilde{Y}_i|)|\tilde{Y}_i|^3|\tilde{X}_i]\leq E[\tilde{Y}_i^2|\tilde{X}_i]$. Let $\overline{Y}$ be a bound for $\tilde{Y}_i$. Then $\lambda E[\exp(\lambda|\tilde{Y}_i|)|\tilde{Y}_i|^3|\tilde{X}_i]\leq\lambda\exp(\lambda\overline{Y})\overline{Y}E[|\tilde{Y}_i|^2|\tilde{X}_i]$, so the inequality holds for any $\lambda$ with $\lambda\exp(\lambda\overline{Y})\overline{Y}\leq 1$. From now on, we fix $\lambda>0$ so that this inequality holds.

Letting $\underline{f}_{\tilde{X}}$ be a lower bound for $f_{\tilde{X}}(x)$ over $0\leq x\leq\overline{x}$ and applying Markov's inequality, the above bound gives

$$P\left(\lambda A \sup_{0\le x\le \tilde{X}_{(k)}}\left|\sum_{\tilde{X}_i\le x} Z_i - \sum_{\tilde{X}_i\le x}\tilde{Y}_i\right| > t \middle| \tilde{X}_1,\dots,\tilde{X}_n\right)$$
$$\le \exp(-t) E \exp\left(\lambda A \sup_{0\le x\le \tilde{X}_{(k)}}\left|\sum_{\tilde{X}_i\le x} Z_i - \sum_{\tilde{X}_i\le x}\tilde{Y}_i\right| \middle| \tilde{X}_1,\dots,\tilde{X}_n\right) \le \exp(-t)(1+\lambda \underline{f}_{\tilde{X}}^{-1} k).$$

Thus,

$$P\left(\left|\sum_{\tilde{X}_i\le x} Z_i - \sum_{\tilde{X}_i\le x}\tilde{Y}_i\right| > K\log\left[\sum_{i=1}^n I(\tilde{X}_i\le x)+2\right] \text{ some } 0\le x\le \overline{x} \middle| \tilde{X}_1,\dots,\tilde{X}_n\right)$$
$$\le P\left(\sup_{0\le x\le \tilde{X}_{(k)}}\left|\sum_{\tilde{X}_i\le x} Z_i - \sum_{\tilde{X}_i\le x}\tilde{Y}_i\right| > K\log k \text{ some } 2\le k\le n \middle| \tilde{X}_1,\dots,\tilde{X}_n\right)$$
$$\le \sum_{k=2}^n P\left(\lambda A \sup_{0\le x\le \tilde{X}_{(k)}}\left|\sum_{\tilde{X}_i\le x} Z_i - \sum_{\tilde{X}_i\le x}\tilde{Y}_i\right| \ge \lambda A K\log k \middle| \tilde{X}_1,\dots,\tilde{X}_n\right)$$
$$\le \sum_{k=2}^n k^{-\lambda AK}(1+\lambda \underline{f}_{\tilde{X}}^{-1} k) \le \sum_{k=2}^\infty k^{-\lambda AK}(1+\lambda \underline{f}_{\tilde{X}}^{-1} k),$$

which can be made arbitrarily small by making $K$ large. □

Embedding $\sum_{\tilde{X}_i\le x} Z_i$ in a Brownian motion, we can restate the above construction as follows: with probability at least $1-K(\varepsilon)$ conditional on $\tilde{X}_1,\dots,\tilde{X}_n$,

$$\left|n\hat{G}(x) - \mathbb{B}_n(n\hat{g}(x))\right| \le K\log[n\hat{F}_{\tilde{X}}(x)+2] \text{ all } 0\le x\le \overline{x}$$

where $\mathbb{B}_n(t) = \mathbb{B}_n(t;\tilde{X}_1,\dots,\tilde{X}_n)$ is a Brownian motion conditional on $\tilde{X}_1,\dots,\tilde{X}_n$. Let $\overline{f}_{\tilde{X}}$ be an upper bound for the density of $\tilde{X}_i$ on $[0,\overline{x}]$.

**Lemma S1.7.** *Under the conditions of Theorem A.2, for any $\eta>0$,*

$$\hat{F}_{\tilde{X}}(x) \le \overline{f}_{\tilde{X}}\cdot(1+\eta)(x\vee \underline{h}_n)$$

*for all $0\le x\le\overline{x}$ with probability approaching one.*

*Proof.* By Lemma S1.5,

$$\sup_{\underline{h}_n \leq x \leq \overline{x}} \frac{\sqrt{n}|\hat{F}_{\tilde{X}}(x) - F_{\tilde{X}}(x)|}{\sqrt{x \log\log x^{-1}}} = \mathcal{O}_P(1).$$

Thus,

$$\begin{aligned}
\sup_{\underline{h}_n \leq x \leq \overline{x}} \frac{|\hat{F}_{\tilde{X}}(x) - F_{\tilde{X}}(x)|}{x} &= \sup_{\underline{h}_n \leq x \leq \overline{x}} \frac{\sqrt{n}|\hat{F}_{\tilde{X}}(x) - F_{\tilde{X}}(x)|}{\sqrt{x \log\log x^{-1}}} \frac{\sqrt{x \log\log x^{-1}}}{\sqrt{n}x} \\
&= \mathcal{O}_P\left( \sup_{\underline{h}_n \leq x \leq \overline{x}} \frac{\sqrt{\log\log x^{-1}}}{\sqrt{nx}} \right) = \mathcal{O}_P\left( \frac{\sqrt{\log\log \underline{h}_n^{-1}}}{\sqrt{n\underline{h}_n}} \right) = o_P(1)
\end{aligned}$$

where the last step follows since $n\underline{h}_n / \log\log \underline{h}_n^{-1} \to \infty$. Thus, for any $\eta > 0$, we have, with probability approaching one,

$$\hat{F}_{\tilde{X}}(x) \leq \hat{F}_{\tilde{X}}(x \vee \underline{h}_n) \leq F_{\tilde{X}}(x \vee \underline{h}_n) + (\eta \overline{f}_{\tilde{X}})(x \vee \underline{h}_n) \leq \overline{f}_{\tilde{X}} \cdot (1+\eta)(x \vee \underline{h}_n)$$

for all $x$. □

Combining these two lemmas, we have, for large enough $n$,

$$\begin{aligned}
\limsup_n P\left( \left|n\hat{G}(x) - \mathbb{B}_n(n\hat{g}(x))\right| > K \log\left[2n\overline{f}_{\tilde{X}}(x \vee \underline{h}_n) + 2\right] \text{ some } 0 \leq x \leq \overline{x} \right) \\
\leq \varepsilon(K) + \limsup_n P\left( \hat{F}_{\tilde{X}}(x) > \overline{f}_{\tilde{X}} \cdot 2(x \vee \underline{h}_n) \right) \leq \varepsilon(K).
\end{aligned}$$

Since this can be made arbitrarily small by making $K$ large, it follows that

$$\sup_{0 \leq x \leq \overline{x}} \frac{\left|n\hat{G}(x) - \mathbb{B}_n(n\hat{g}(x))\right|}{\sqrt{n(x \vee \underline{h}_n)}} = \mathcal{O}_P\left( \sup_{0 \leq x \leq \overline{x}} \frac{\log\left[2n\overline{f}_{\tilde{X}}(x \vee \underline{h}_n) + 2\right]}{\sqrt{n(x \vee \underline{h}_n)}} \right) = \mathcal{O}_P\left( \frac{\log(n\underline{h}_n)}{\sqrt{n\underline{h}_n}} \right),$$

which gives the required rate for $R_{1,n}(x)$.

Define the function $LL(x) = \log\log x$ for $\log\log x \geq 1$ and $LL(x) = 1$ otherwise. Given $K$, let $B_n(K)$ be the event that

$$|n\hat{g}(x) - nx| \leq K\sqrt{n(x \vee \underline{h}_n) LL(x/\underline{h}_n)} \text{ all } 0 \leq x \leq \overline{x},$$

and let $C_n(K)$ be the event that

$$\left|\mathbb{B}_n(t') - \mathbb{B}_n(t)\right| \le K\sqrt{(|t'-t| \vee 1)\cdot \log(t \vee t' \vee 2)} \text{ all } 0 \le t, t' < \infty.$$

**Lemma S1.8.** *On the event $B_n(K) \cap C_n(K)$, for large enough $n$,*

$$\begin{aligned}
&\frac{|R_{2,n}(x)|}{\sqrt{n(x \vee \underline{h}_n)}} \le K^{3/2}[n(x \vee \underline{h}_n)]^{-1/4}\{LL(x/\underline{h}_n)\}^{1/4} \cdot \{\log 2 + \log[n(x \vee \underline{h}_n)]\}^{1/2} \\
&\le K^{3/2}(n\underline{h}_n)^{-1/4} \cdot \{\log 2 + \log[n\underline{h}_n]\}^{1/2}
\end{aligned}$$

*for all $0 \le x \le \overline{x}$.*

*Proof.* On this event, for all $0 \le x \le \overline{x}$ and large enough $n$,

$$\begin{aligned}
|R_{2,n}(x)| = |\mathbb{B}_n(n\hat{g}(x)) - \mathbb{B}_n(nx)| &\le \sup_{|t-nx| \le K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)}} |\mathbb{B}_n(t) - \mathbb{B}_n(nx)| \\
&\le \sup_{|t-nx| \le K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)}} K\sqrt{(|t-nx| \vee 1)\cdot \log[t \vee (nx) \vee 2]} \\
&\le K\sqrt{K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)} \cdot \log[2n(x \vee \underline{h}_n)]} \\
&= K^{3/2}n^{1/4}(x \vee \underline{h}_n)^{1/4}\{LL(x/\underline{h}_n)\}^{1/4} \cdot \{\log 2 + \log[n(x \vee \underline{h}_n)]\}^{1/2}.
\end{aligned}$$

□

**Lemma S1.9.** *Under the conditions of Theorem A.2, for any $\varepsilon > 0$, there exists a $K$ such that $P(B_n(K)) \ge 1 - \varepsilon$ for large enough $n$.*

*Proof.* Let $\mathcal{X}^k = (2^k\underline{h}_n, 2^{k+1}\underline{h}_n] \cap [0, \overline{x}]$. We have, for $k \ge 2$,

$$\begin{aligned}
&P\left(|n\hat{g}(x) - nx| > K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)} \text{ some } x \in \mathcal{X}^k\right) \\
&= P\left(|\mathbb{G}_n f(\tilde{X}_i)^{-1} I(\tilde{X}_i \le x)| > K\sqrt{x \cdot LL(x/\underline{h}_n)} \text{ some } x \in \mathcal{X}^k\right) \\
&\le P\left(\sup_{x \in \mathcal{X}^k} |\mathbb{G}_n f(\tilde{X}_i)^{-1} I(\tilde{X}_i \le x)| > K\sqrt{2^k\underline{h}_n \cdot LL(2^k)}\right) \\
&\le C\exp\left(-\frac{K^2 LL(2^k)}{C}\right) \le C\exp\left(-\frac{K^2}{C}\log\log(2^k)\right) = C[k\log 2]^{-\frac{K^2}{C}}
\end{aligned}$$

for some constant $C$ by Lemma S1.3. Thus,

$$P\left(|n\hat{g}(x) - nx| > K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)} \text{ some } 4\underline{h}_n \leq x \leq \overline{x}\right) \leq C\sum_{k=2}^{\infty}[k\log 2]^{-K^2/C}$$

which can be made arbitrarily small by making $K$ large. Note also that

$$\begin{aligned}
&P\left(|n\hat{g}(x) - nx| > K\sqrt{n(x \vee \underline{h}_n)LL(x/\underline{h}_n)} \text{ some } 0 \leq x \leq 4\underline{h}_n\right) \\
&\leq P\left(\sup_{0\leq x\leq 4\underline{h}_n} |\mathbb{G}_n f(\tilde{X}_i)^{-1} I(\tilde{X}_i \leq x)| > K\sqrt{\underline{h}_n}\right),
\end{aligned}$$

which can also be made arbitrarily small by choosing $K$ large by Lemma S1.3. Combining these bounds gives the result. □

**Lemma S1.10.** *Under the conditions of Theorem A.2, for any $\varepsilon > 0$, there exists a $K$ such that with probability one for all $n$, $P(C_n(K)|\tilde{X}_1, \ldots, \tilde{X}_n) \geq 1 - \varepsilon$.*

*Proof.* We have

$$\begin{aligned}
&1 - P(C_n(K)|\tilde{X}_1, \ldots, \tilde{X}_n) \\
&\qquad = P\left(\left|\mathbb{B}_n(t') - \mathbb{B}_n(t)\right| > K\sqrt{(|t - t'| \vee 1) \cdot \log(t \vee t' \vee 2)} \text{ some } 0 \leq t, t' < \infty\right) \\
&\qquad = P\left(|\mathbb{B}_n(t+s) - \mathbb{B}_n(t)| > K\sqrt{(s \vee 1) \cdot \log[(t+s) \vee 2]} \text{ some } 0 \leq s, t < \infty\right) \\
&\qquad \leq \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} P\left(|\mathbb{B}_n(t+s) - \mathbb{B}_n(t)| > K\sqrt{(s \vee 1) \cdot \log[(t+s) \vee 2]} \text{ some } (s,t) \in \mathcal{S}_{k,\ell}\right)
\end{aligned}$$

where $\mathcal{S}_{k,\ell} = \{(s,t)|\ell \leq s \leq \ell + 1, (\ell \vee 1)k \leq t \leq (\ell \vee 1)(k+1)\}$. Note that

$$\begin{aligned}
&P\left(|\mathbb{B}_n(t+s) - \mathbb{B}_n(t)| > K\sqrt{(s \vee 1) \cdot \log[(t+s) \vee 2]} \text{ some } (s,t) \in \mathcal{S}_{k,\ell}\right) \\
&\leq P\left(|\mathbb{B}_n(t+s) - \mathbb{B}_n(t)| > K\sqrt{(\ell \vee 1) \cdot \log\{[(\ell \vee 1)k + \ell] \vee 2\}} \text{ some } (s,t) \in \mathcal{S}_{k,\ell}\right) \\
&= P\left(|\mathbb{B}_n(t+s) - \mathbb{B}_n(t)| > K\sqrt{(\ell \vee 1) \cdot \log\{[(\ell \vee 1)k + \ell] \vee 2\}} \text{ some } (s,t) \in \mathcal{S}_{0,\ell}\right) \\
&\leq P\left(|\mathbb{B}_n(t)| > (K/2)\sqrt{(\ell \vee 1) \cdot \log\{[(\ell \vee 1)k + \ell] \vee 2\}} \text{ some } 0 \leq t \leq (\ell \vee 1) + \ell + 1\right) \\
&\leq 4P\left(|\mathbb{B}_n((\ell \vee 1) + \ell + 1)| > (K/2)\sqrt{(\ell \vee 1) \cdot \log\{[(\ell \vee 1)k + \ell] \vee 2\}}\right)
\end{aligned}$$

$$\le 4 \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\frac{(K/2)^2(\ell \vee 1)\cdot \log\{[(\ell\vee 1)k+\ell]\vee 2\}}{(\ell \vee 1)+\ell+1}\right)$$
$$\le 4 \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(K/2)^2 \log\{[(\ell\vee 1)k+\ell]\vee 2\}}{6}\right) = 4\cdot\frac{1}{\sqrt{2\pi}}\cdot\{[(\ell\vee 1)k+\ell]\vee 2\}^{-K^2/24}.$$

The third line follows since $\mathbb{B}_n(t)$ has the same distribution as $\mathbb{B}_n(t+(\ell\vee 1)k)$. The fourth line follows since, if $|\mathbb{B}_n(t+s)-\mathbb{B}_n(t)| > C$ for some $C$ and $(s,t)\in\mathcal{S}_{0,\ell}$, we must have $|\mathbb{B}_n(t)| > C/2$ for some $0 \le t \le (\ell\vee 1)+\ell+1$. The fifth line follows from the reflection principle for the Brownian motion (see Theorem 2.21 in Mörters and Peres, 2010). The sixth line uses the fact that $P(Z\ge x)\le \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ for $x\ge 1$ and $Z\sim N(0,1)$.

Thus,

$$P\left(\left|\mathbb{B}_n(t')-\mathbb{B}_n(t)\right| > K\sqrt{(|t-t'|\vee 1)\cdot\log(t\vee t'\vee 1)} \text{ some } 0\le t,t'<\infty\right)$$
$$\le \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} 4\cdot\frac{1}{\sqrt{2\pi}}\cdot\{[(\ell\vee 1)k+\ell]\vee 2\}^{-K^2/24}.$$

This can be made arbitrarily small by making $K$ large. □

Theorem A.2 now follows since, for any constant $\varepsilon > 0$, there is a constant $K$ such that $\sup_{\underline{h}_n\le h\le \bar{x}/A}|\hat{\mathbb{H}}_n(h)-\mathbb{H}_n(h)|$ is less than $K\{(\log n\underline{h}_n)(n\underline{h}_n)^{-1/2}+(n\underline{h}_n)^{-1/4}[\log(n\underline{h}_n)]^{1/4}\}$ with probability at least $1-\varepsilon$ asymptotically.

## S1.4 Calculations for Extreme Value Limit

This section provides the calculations for the asymptotic distribution derived in Theorem A.3 in Section A.5 of the appendix.

As described in the proof of Theorem A.3, we use Theorem 12.3.5 of Leadbetter et al. (1983) applied to the process $\mathbb{X}(t)=\mathbb{H}(e^t)$, which is stationary, with, in the case where $k(A)\ne 0$, $\alpha = 1$ and $C=\frac{Ak(A)^2}{\int k(u)^2\,du}$ and, in the case where $k(A)=0$, $\alpha=2$ and $C=\frac{\int\left[k'(u)u+\frac{1}{2}k(u)\,du\right]^2du}{2\int k(u)^2\,du}$. In the notation of that theorem, we have

$$r(t)=cov\left(\mathbb{X}(s),\mathbb{X}(s+t)\right)=\frac{e^{\frac{1}{2}t}\int k(ue^t)k(u)\,du}{\int k(u)^2\,du}.$$

Since $r(t)$ is bounded by a constant times $e^{\frac{1}{2}t}\cdot e^{-t}$, the condition $r(t)\log t \overset{t\to\infty}{\to} 0$ holds, so it remains to verify that $r(t) = 1 - C|t|^\alpha + o(|t|^\alpha)$ with $\alpha$ and $C$ given above.

Since $k(ue^t)k(u)$ has a continuous derivative with respect to $t$ on its support, which for $t \geq 0$ is $[-Ae^{-t}, Ae^{-t}]$, it follows by Leibniz's rule and symmetry of $k$ that, for $t \geq 0$ $\frac{d}{dt}\int k(ue^t)k(u)\,du = -2Ae^{-t}k(A)k(Ae^{-t}) + \int k'(ue^t)k(u)ue^t\,du$ for $t \geq 0$. Thus, for $t \geq 0$,

$$\begin{aligned}\frac{d}{dt_+}r(t) &= \frac{e^{\frac{1}{2}t}\frac{d}{dt_+}\int k(ue^t)k(u)\,du + \frac{1}{2}e^{\frac{1}{2}t}\int k(ue^t)k(u)\,du}{\int k(u)^2\,du} \\ &= \frac{e^{\frac{1}{2}t}\left[-2Ae^{-t}k(A)k(Ae^{-t}) + \int k'(ue^t)k(u)ue^t\,du\right] + \frac{1}{2}e^{\frac{1}{2}t}\int k(ue^t)k(u)\,du}{\int k(u)^2\,du}.\end{aligned}$$

Thus,

$$\left.\frac{d}{dt_+}r(t)\right|_{t=0} = \frac{-2Ak(A)^2 + \int k'(u)k(u)u\,du + \frac{1}{2}\int k(u)^2\,du}{\int k(u)^2\,du} = \frac{-Ak(A)^2}{\int k(u)^2\,du}$$

where the last step follows by noting that, applying integration by parts with $k(u)u$ playing the part of $u$ and $k'(u)du$ playing the part of $dv$,

$$\begin{aligned}\int k(u)k'(u)u\,du &= \left[k(u)^2u\right]_{-A}^{A} - \int k(u)[k(u) + k'(u)u]\,du \\ &= 2k(A)^2A - \int k(u)^2\,du - \int k(u)k'(u)u\,du\end{aligned}$$

so that $\int k(u)k'(u)u\,du = k(A)^2A - \frac{1}{2}\int k(u)^2\,du$. For the case where $k(A) \neq 0$, it follows from this and a symmetric argument for $t \leq 0$ that $r(t) = 1 - C|t| - o(|t|)$ for $C = \frac{Ak(A)^2}{\int k(u)^2\,du}$ as required.

For the case where $k(A) = 0$, applying Leibniz's rule as above shows that $r(t)$ is differentiable with,

$$r'(t) = e^{\frac{1}{2}t}\frac{\int k'(ue^t)k(u)ue^t\,du + \frac{1}{2}\int k(ue^t)k(u)\,du}{\int k(u)^2\,du}.$$

Thus, $r'(0) = 0$ (using the integration by parts identity above) and $r(t)$ is twice differentiable with

$$r''(t) = e^{\frac{1}{2}t}\frac{\frac{d}{dt}\int k'(ue^t)k(u)ue^t\,du + \frac{1}{2}\left(\frac{d}{dt}\int k(ue^t)k(u)\,du + \int k'(ue^t)k(u)ue^t\,du + \int k(ue^t)k(u)\,du\right)}{\int k(u)^2\,du}.$$

We have

$$\frac{d}{dt}\int k'(ue^t)k(u)ue^t\,du = \frac{d}{dt}\int k'(v)k(ve^{-t})ve^{-t}\,dv$$

$$= \int k'(v)k'(ve^{-t})(-ve^{-t})ve^{-t}\,dv - \int k'(v)k(ve^{-t})ve^{-t}\,dv$$

and $\frac{d}{dt}\int k(ue^t)k(u)\,du = \int k'(ue^t)k(u)ue^t\,du$, so this gives

$$r''(t) = e^{\frac{1}{2}t}\frac{-\int k'(v)k'(ue^{-t})u^2e^{-2t}\,du - \frac{1}{2}\int k'(ue^t)k(u)ue^t\,du}{\int k(u)^2\,du} + \frac{1}{2}e^{\frac{1}{2}t}\frac{\int k'(ue^t)k(u)ue^t\,du + \frac{1}{2}\int k(ue^t)k(u)\,du}{\int k(u)^2\,du}.$$

Thus,

$$r''(0) = \frac{-\int [k'(u)u]^2\,du + \frac{1}{4}\int k(u)^2\,du}{\int k(u)^2\,du}.$$

Since, by the integration by parts argument above, $\frac{1}{4}\int k(u)^2\,du = \frac{1}{2}\int k(u)^2\,du - \frac{1}{4}\int k(u)^2\,du = -\int k(u)k'(u)u\,du - \frac{1}{4}\int k(u)^2\,du$, this is equal to

$$\frac{-\int [k'(u)u]^2\,du - \int k(u)k'(u)u\,du - \frac{1}{4}\int k(u)^2\,du}{\int k(u)^2\,du} = -\frac{\int \left[k'(u)u + \frac{1}{2}k(u)\right]^2\,du}{\int k(u)^2\,du}$$

which gives the required expansion with $C$ given by one half of the negative of the above display and $\alpha = 2$.

### S1.5 Delta Method

We state some results that allow us to obtain influence function representations with the necessary uniform rate for differentiable functions of estimators. These results amount to applying the delta method to our setting and keeping track of the uniform rates.

Let $\hat{\beta}(h)$ be an estimator of a parameter $\beta(h) \in \mathbb{R}^{d_\beta}$ with influence function representation

$$\sqrt{nh}(\hat{\beta}(h) - \beta(h)) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\psi_\beta(W_i,h)k(X_i/h) + R_{1,n}(h)$$

for some function $\psi_\beta$ and a kernel function $k$, where $\psi_\beta(W_i,h)k(X_i/h)$ has mean zero and $\sup_{\underline{h}_n\le h\le\overline{h}}|R_{1,n}(h)| = o_P(1/\sqrt{\log\log \underline{h}_n^{-1}})$. Let $g$ be a function from $\mathbb{R}^{d_\beta}$ to $\mathbb{R}^{d_\theta}$ and consider the parameter $\theta(h) = g(\beta(h))$ and the estimator $\hat{\theta}(h) = g(\hat{\beta}(h))$.

Let $\hat{V}_\beta(h)$ be an estimate of $V_\beta(h) = \frac{1}{h}E\psi_\beta(W_i,h)\psi_\beta(W_i,h)'k(X_i/h)^2$, the (pointwise in $h$)

asymptotic variance of $\hat{\beta}(h)$. A natural estimator of the asymptotic variance $V_\theta(h)$ of $\hat{\theta}$ is

$$\hat{V}_\theta(h) = D_g(\hat{\beta}(h))'\hat{V}_\beta(h)D_g(\hat{\beta}(h))'.$$

**Lemma S1.11.** *Suppose that $\beta(h)$ is bounded uniformly over $h \leq \overline{h}_n$ where $\overline{h}_n = \mathcal{O}(1)$ and*

*(i) For large enough $n$, $g$ is differentiable on an open set containing the range of $\beta(h)$ over $h \leq \overline{h}_n$, with Lipschitz continuous derivative $D_g$.*

*(ii) $\psi_\beta$ and $k$ are bounded, $k$ has finite support, and the class of functions $(w, x) \mapsto \psi_\beta(w, h)k(x/h)$ has polynomial uniform covering number.*

*(iii) $|X_i|$ has a bounded density on $[0, \overline{h}_n]$ for large enough $n$.*

*Then, if $n\underline{h}_n/(\log\log n)^3 \to \infty$,*

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \left| \sqrt{nh}(\hat{\theta}(h) - \theta(h)) - \frac{1}{\sqrt{nh}} \sum_{i=1}^n D_g(\beta(h))\psi_\beta(W_i, h)k(X_i/h) \right| = o_P\left(1\Big/\sqrt{\log\log \underline{h}_n^{-1}}\right).$$

*If, in addition, $\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \|\hat{V}_\beta(h) - V_\beta(h)\| \xrightarrow{p} 0$, then, for some constant $K$ and some $R_{n,2}(h)$ with $\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}}|R_{n,2}(h)| = \mathcal{O}_P(1)$,*

$$\left\|\hat{V}_\theta(h) - V_\theta(h)\right\| \leq K \left\|\hat{V}_\beta(h) - V_\beta(h)\right\| + R_{n,2}(h)$$

*for all $\underline{h}_n \leq h \leq \overline{h}_n$ with probability approaching one.*

*Proof.* By a first order Taylor expansion, we have, for some $\beta^*(h)$ with $\|\beta^*(h) - \beta(h)\| \leq \|\hat{\beta}(h) - \beta(h)\|$,

$$\begin{aligned}
&\sqrt{nh}(\hat{\theta}(h) - \theta(h)) = \sqrt{nh}(g(\hat{\beta}(h)) - g(\beta(h))) = \sqrt{nh}D_g(\beta^*(h))(\hat{\beta}(h) - \beta(h)) \\
&= D_g(\beta^*(h))\frac{1}{\sqrt{nh}}\sum_{i=1}^n \psi_\beta(W_i, h)k(X_i/h) + D_g(\beta^*(h))R_{1,n}(h) \\
&= \frac{1}{\sqrt{nh}}\sum_{i=1}^n D_g(\beta(h))\psi_\beta(W_i, h)k(X_i/h) + [D_g(\beta^*(h)) - D_g(\beta(h))]\frac{1}{\sqrt{nh}}\sum_{i=1}^n \psi_\beta(W_i, h)k(X_i/h) \\
&+ D_g(\beta^*(h))R_{1,n}(h)
\end{aligned}$$

Applying Lemma A.2, $\hat{\beta}(h) - \beta(h)$ is $\mathcal{O}_P(\sqrt{\log\log h^{-1}}/\sqrt{nh})$ uniformly over $\underline{h}_n \le h \le \overline{h}_n$ and $\frac{1}{\sqrt{nh}}\sum_{i=1}^n \psi_\beta(W_i, h)k(X_i/h)$ is $\mathcal{O}_P(\sqrt{\log\log h^{-1}})$ uniformly over $\underline{h}_n \le h \le \overline{h}_n$. so that, by the Lipschitz condition on $D_g$, the second term is $\mathcal{O}_P(\log\log h^{-1}/\sqrt{nh})$ uniformly over $\underline{h}_n \le h \le \overline{h}_n$, which is $o_P(1/\sqrt{\log\log \underline{h}_n^{-1}})$ uniformly over $\underline{h}_n \le h \le \overline{h}_n$ since $\sqrt{n\underline{h}_n}/(\log\log \underline{h}_n)^{3/2} \to \infty$. The last term is $o_P(1/\sqrt{\log\log h_n^{-1}})$ uniformly over $\underline{h}_n \le h \le \overline{h}_n$ by the conditions on $R_{1,n}(h)$, the uniform consistency of $\hat{\beta}(h)$ and the Lipschitz condition on $D_g$.

For the second claim, note that

$$\begin{aligned}
\hat{V}_\theta - V_\theta &= D_g(\hat{\beta}(h))\hat{V}_\beta(h)D_g(\hat{\beta}(h))' - D_g(\beta(h))V_\beta(h)D_g(\beta(h))' \\
&= [D_g(\hat{\beta}(h)) - D_g(\beta(h))]\hat{V}_\beta(h)D_g(\hat{\beta}(h))' + D_g(\beta(h))[\hat{V}_\beta(h) - V_\beta(h)]D_g(\hat{\beta}(h))' \\
&\quad + D_g(\beta(h))V_\beta(h)[D_g(\hat{\beta}(h)) - D_g(\beta(h))]'.
\end{aligned}$$

The first and last terms converge at a $\sqrt{\log\log h^{-1}}/\sqrt{nh}$ rate uniformly over $\underline{h}_n \le h \le \overline{h}_n$ by Lemma A.2 and the Lipschitz continuity on $D_g$. The second term is bounded by a constant times $\|\hat{V}_\beta(h) - \hat{V}_\beta(h)\|$ uniformly over $\underline{h}_n \le h \le \overline{h}$ with probability approaching one by the uniform consistency of $\hat{\beta}(h)$ and the Lipschitz continuity of $D_g$. $\square$

### S1.6 Sufficient Conditions Based on Non-normalized Influence Function

In some cases, it will be easier to verify the conditions for an influence function approximation to $\sqrt{nh}(\hat{\theta}(h) - \theta(h))$ rather than the normalized version $\sqrt{nh}(\hat{\theta}(h) - \theta(h))/\hat{\sigma}(h)$. The following lemma is useful in these cases.

**Lemma S1.12.** *Suppose that the following conditions hold for some $\tilde{\psi}(W_i, h)$.*

1. *$E\tilde{\psi}(W_i, h)k(X_i/h) = 0$ and $k$ is bounded and symmetric with finite support $[-A, A]$.*

2. *$|X_i|$ has a density $f_{|X|}$ with $f_{|X|}(0) > 0$, $\tilde{\psi}(W_i, h)k(X_i/h)$ is bounded uniformly over $h \le \underline{h}_n$ and, for some deterministic function $\ell(h)$ with $\ell(h)\log\log h^{-1} \to 0$ as $h \to 0$, the following expressions are bounded by $\ell(t)$: $|f_{|X|}(t) - f_{|X|}(0)|$, $|E[\tilde{\psi}(W_i, 0)||X_i| = t] - E[\tilde{\psi}(W_i, 0)||X_i| = 0]|$, $|var[\tilde{\psi}(W_i, 0)||X_i| = t] - var[\tilde{\psi}(W_i, 0)||X_i| = 0]|$ and $|(\tilde{\psi}(W_i, t) - \tilde{\psi}(W_i, 0))k(X_i/h)|$.*

*Let $\sigma^2(h) = \frac{1}{h}var(\tilde{\psi}(W_i, h)k(X_i/h))$ for $h > 0$ and let $\sigma^2(0) = var[\tilde{\psi}(W_i, 0)||X_i| = 0]f_{|X|}(0) \cdot \int_{u=0}^\infty k(u)^2\,du$. Let $\psi(W_i, h) = \tilde{\psi}(W_i, h)/\sigma(h)$ so that $\frac{1}{h}var[\psi(W_i, h)k(X_i/h)] = 1$. Suppose that*

*$var\,[\tilde\psi(W_i,0)||X_i|=0]>0$. Then the above assumptions hold with $\tilde\psi$ replaced by $\psi$ for $h$ small enough and with $\ell(t)$ possibly redefined.*

*Proof.* First, note that the only condition we need to verify is the one that involves $|[\psi(W_i,h)-\psi(W_i,0)]k(X_i/h)|$, since the remaining conditions are only changed by multiplication by a constant when $\tilde\psi$ is replaced by $\psi$. Note that

$$\sigma^2(h)-\frac{1}{h}var(\tilde\psi(W_i,0)k(X_i/h))=\frac{1}{h}var(\tilde\psi(W_i,h)k(X_i/h))-\frac{1}{h}var(\tilde\psi(W_i,0)k(X_i/h))=$$
$$\frac{1}{h}var\{[\tilde\psi(W_i,h)-\tilde\psi(W_i,0)]k(X_i/h)\}+2\frac{1}{h}cov\{[\tilde\psi(W_i,h)-\tilde\psi(W_i,0)]k(X_i/h),\tilde\psi(W_i,0)k(X_i/h)\}.$$

Since $|(\tilde\psi(W_i,h)-\tilde\psi(W_i,0))k(X_i/h)|\le\ell(h)I(|X_i|\le Ah)$, $\tilde\psi(W_i,h)k(X_i/h)$ and $\tilde\psi(W_i,0)k(X_i/h)$ are bounded, the last two terms are bounded by a constant times $\ell(h)\frac{1}{h}EI(|X_i|\le Ah)$, which is bounded by a constant times $\ell(h)$ by the assumption on the density of $|X_i|$.

Thus, let us consider

$$\frac{1}{h}var(\tilde\psi(W_i,0)k(X_i/h))$$
$$=\frac{1}{h}\int_{x=0}^{\infty}var\,[\tilde\psi(W_i,0)||X_i|=x]\,k(x/h)^2f_{|X|}(x)\,dx+\frac{1}{h}var\left\{E\,[\tilde\psi(W_i,0)||X_i|]\,k(X_i/h)\right\}.$$

Arguing as in the proof of Lemma A.6 (using the fact that $E\tilde\psi(W_i,h)k(X_i/h)=0$ and taking limits), it can be seen that $E\,[\tilde\psi(W_i,0)||X_i|=0]=0$ under these conditions. Thus, the last term is bounded by $\ell(Ah)^2\frac{1}{h}Ek(X_i/h)^2$. The first term is equal to $var(\tilde\psi(W_i,0)\mid|X_i|=0)f_{|X|}(0)\cdot\int_{u=0}^{\infty}k(u)^2\,du$ plus a term that is bounded by a constant times $\ell(Ah)$.

It follows that, letting $\sigma^2(0)=var\,[\tilde\psi(W_i,0)||X_i|=0]\,f_{|X|}(0)\int_{u=0}^{\infty}k(u)^2\,du$ as defined above, we have, for some constant $K$, $|\sigma^2(h)-\sigma^2(0)|\le K\ell(Ah)$. Thus,

$$|[\psi(W_i,h)-\psi(W_i,0)]k(X_i/h)|$$
$$\le\frac{1}{\sigma(0)}\,|[\tilde\psi(W_i,h)-\tilde\psi(W_i,0)]k(X_i/h)|+|\tilde\psi(W_i,h)k(X_i/h)|\cdot\left|\frac{1}{\sigma(h)}-\frac{1}{\sigma(0)}\right|.$$

The first term is bounded by a constant times $\ell(h)$ by assumption. The last term is bounded by a constant times $|\sigma^2(h)-\sigma^2(0)|$, which is bounded by a constant times $\ell(Ah)$ as shown above. □

## S2 Local polynomial estimators: regression discontinuity/estimation at the boundary

This section gives primitive conditions for smooth functions of estimates based on local polynomial estimates at the boundary, or at a discontinuity in the regression function. The results are used in Section S3 below to verify the conditions of Theorem 3.1 for the applications in Section 4 in the main text. Throughout this section, we consider a setup with $\{(X_i, Y_i')'\}_{i=1}^n$ i.i.d. with $X_i$ a real valued random variable and $Y_i$ taking values in $\mathbb{R}^{d_Y}$. We consider smooth functions of the left and right hand limits of the regression function at a point, which we normalize to be zero.

Let $(\hat{\beta}_{u,j,1}(h), \hat{\beta}_{u,j,2}(h)/h, \dots, \hat{\beta}_{u,j,r+1}(h)/h^r)$ be the coefficients of an $r$th order local polynomial estimate of $E[Y_{i,j}|X_i = 0_+]$ based on the subsample with $X_i \geq 0$ with a kernel function $k^*$. Similarly, let $(\hat{\beta}_{\ell,j,1}(h), \hat{\beta}_{\ell,j,2}(h)/h, \dots, \hat{\beta}_{\ell,j,r+1}(h)/h^r)$ be the coefficients of an $r$th order local polynomial estimate of $E[Y_{i,j}|X_i = 0_-]$ based on the subsample with $X_i < 0$, where the polynomial is taken in $|X_i|$ rather than $X_i$ (this amounts to multiplying even elements of $\beta_{\ell,j}$ by $-1$). The scaling by powers of $h$ is used to handle the different rates of convergence of the different coefficients. Let $p(x) = (1, x, x^2, \dots, x^r)'$, and define $\hat{\beta}_{u,j} = (\hat{\beta}_{u,j,1}(h), \hat{\beta}_{u,j,2}(h), \dots, \hat{\beta}_{u,j,r+1}(h))$ and $\hat{\beta}_{\ell,j} = (\hat{\beta}_{\ell,j,1}(h), \hat{\beta}_{\ell,j,2}(h), \dots, \hat{\beta}_{\ell,j,r+1}(h))$. Let $p(x) = (1, x, x^2, \dots, x^r)'$. Then $\hat{\beta}_{u,j}$ minimizes

$$\sum_{i=1}^n (Y_{i,j} - p(|X_i/h|)'\beta_{u,j})^2 I(X_i \geq 0) k^*(X_i/h)$$

and $\hat{\beta}_{\ell,j}$ minimizes

$$\sum_{i=1}^n (Y_{i,j} - p(|X_i/h|)'\beta_{u,j})^2 I(X_i < 0) k^*(X_i/h).$$

Define

$$\begin{aligned}
\Gamma_u(h) &= \tfrac{1}{h} E p(|X_i/h|) p(|X_i/h|)' k^*(X_i/h) I(X_i \geq 0), \\
\Gamma_\ell(h) &= \tfrac{1}{h} E p(|X_i/h|) p(|X_i/h|)' k^*(X_i/h) I(X_i < 0), \\
\hat{\Gamma}_u(h) &= \tfrac{1}{nh} \textstyle\sum_{i=1}^n p(|X_i/h|) p(|X_i/h|)' k^*(X_i/h) I(X_i \geq 0) \qquad \text{and} \\
\hat{\Gamma}_\ell(h) &= \tfrac{1}{nh} \textstyle\sum_{i=1}^n p(|X_i/h|) p(|X_i/h|)' k^*(X_i/h) I(X_i < 0).
\end{aligned}$$

Let $\mu_{k^*,\ell} = \int_0^\infty u^\ell k^*(u)\, du$, and let $M$ be the matrix with $i,j$th element given by $\mu_{k^*,i+j-2}$.

Let $\hat{\alpha}_u(h) = (\hat{\beta}_{1,1,u}(h), \dots \hat{\beta}_{1,d_Y,u}(h))'$ and $\hat{\alpha}_\ell(h) = (\hat{\beta}_{1,1,\ell}(h), \dots \hat{\beta}_{1,d_Y,\ell}(h))'$, and similarly for $\alpha_u(h)$ and $\alpha_\ell(h)$ (i.e. $\alpha_u$ and $\alpha_\ell$ contain the constant terms in the local polynomial regressions for each $j$). Let $\hat{\alpha}(h) = (\hat{\alpha}_u(h)', \hat{\alpha}_\ell(h)')$ and $\alpha(h) = (\alpha_u(h)', \alpha_\ell(h)')$. We are interested in $\theta(h) = g(\alpha(h))$ for a differentiable function $g$ from $\mathbb{R}^{2d_Y}$ to $\mathbb{R}$, and an estimator $\hat{\theta}(h) = \hat{g}(\alpha(h))$. We consider standard errors defined by the delta method applied to the robust covariance matrix formula obtained by treating the local linear regressions as a system of $2d_Y$ weighted least squares regressions. Let $\nu_u(h) = e_1'\Gamma_u(h)^{-1}$ and let $\nu_\ell(h) = e_1'\Gamma_\ell(h)^{-1}$. Let $\hat{\nu}_u(h) = e_1'\hat{\Gamma}_u(h)^{-1}$ and let $\nu_\ell(h) = e_1'\hat{\Gamma}_\ell(h)^{-1}$. Let $\psi_\alpha(X_i, Y_i, h)$ be the $(2d_Y) \times 1$ random vector with $j$th element given by

$$\psi_{\alpha,j}(X_i, Y_i, h) = \begin{cases} \nu_u(h) p(|X_i/h|)[Y_{i,j} - p(|X_i/h|)'\beta_{u,j}(h)] I(X_i \geq 0) & \text{if } j = 1, \dots, d_Y, \\ \nu_\ell(h) p(|X_i/h|)[Y_{i,j-d_Y} - p(|X_i/h|)'\beta_{\ell,j-d_Y}(h)] I(X_i < 0) & \text{if } j = d_Y + 1, \dots, 2d_Y. \end{cases}$$

Let $\hat{\psi}_\alpha(X_i, Y_i, h)$ be defined analogously,

$$\hat{\psi}_{\alpha,j}(X_i, Y_i, h) = \begin{cases} \hat{\nu}_u(h) p(|X_i/h|)[Y_{i,j} - p(|X_i/h|)'\hat{\beta}_{u,j}(h)] I(X_i \geq 0) & \text{if } j = 1, \dots, d_Y, \\ \hat{\nu}_\ell(h) p(|X_i/h|)[Y_{i,j-d_Y} - p(|X_i/h|)'\hat{\beta}_{\ell,j-d_Y}(h)] I(X_i < 0) & \text{if } j = d_Y + 1, \dots, 2d_Y. \end{cases}$$

Let

$$V_\alpha(h) = \frac{1}{h} E\psi_\alpha(X_i, Y_i, h)\psi_\alpha(X_i, Y_i, h)' k^*(X_i/h)^2$$

and let

$$\hat{V}_\alpha(h) = \frac{1}{h} E_n\hat{\psi}_\alpha(X_i, Y_i, h)\hat{\psi}_\alpha(X_i, Y_i, h)' k^*(X_i/h)^2.$$

Let $\hat{\sigma}(h) = D_g(\hat{\alpha}(h))\hat{V}_\alpha(h)D_g(\hat{\alpha}(h))'$, and $\sigma(h) = D_g(\alpha(h))V_\alpha(h)D_g(\alpha(h))'$, where $D_g$ is the derivative of $g$.

We make the following assumption throughout this section. In the following assumption, $\ell(t)$ is an arbitrary nondecreasing function satisfying $\lim_{t\downarrow 0} \ell(t) \log\log t^{-1} = 0$.

**Assumption S2.1.** *(i) $X_i$ has a density $f_X(x)$ with $|f_X(x) - f_{X,-}| \leq \ell(x)$ for $x < 0$ and $|f_X(x) - f_{X,+}| \leq \ell(x)$ for some $f_{X,+} > 0$ and $f_{X,-} > 0$.*

*(ii)* $Y_i$ *is bounded and, for some matrices* $\Sigma_-$ *and* $\Sigma_+$ *and vectors* $\tilde{\mu}_-$ *and* $\tilde{\mu}_+$, $\tilde{\Sigma}(x) = var(Y_i|X_i = x)$ *and* $\tilde{\mu}(x) = E(Y_i|X_i = x)$ *satisfy* $\|\tilde{\Sigma}(x) - \Sigma_+\| \leq \ell(x)$ *and* $\|\tilde{\mu}(x) - \tilde{\mu}_+\| \leq \ell(x)$ *for* $x > 0$ *and* $\|\tilde{\Sigma}(x) - \Sigma_-\| \leq \ell(x)$ *and* $\|\tilde{\mu}(x) - \tilde{\mu}_-\| \leq \ell(x)$ *for* $x < 0$.

*(iii)* $k^*$ *is symmetric with finite support* $[-A, A]$*, is bounded with a bounded, uniformly continuous first derivative on* $(0, A)$*, and satisfies* $\int k(u)\,du \neq 0$*, and the matrix* $M$ *is invertible.*

*(iv)* $D_g$ *is bounded and is Lipschitz continuous on an open set containing the range of* $\alpha(h)$ *over* $\overline{h}_n$ *for* $n$ *large enough.*

*(v)* $D_{g,u}(\alpha(0))\tilde{\Sigma}_+ D_{g,u}(\alpha(0)) > 0$ *or* $D_{g,\ell}(\alpha(0))\tilde{\Sigma}_- D_{g,u}(\ell) > 0$.

*(vi)* $\overline{h}_n = \mathcal{O}(1)$ *and* $n\underline{h}_n/(\log\log n)^3 \to \infty$.

**Theorem S2.1.** *Under Assumption S2.1, Assumptions 3.1 and Assumption 3.2 hold with* $k(u) = e_1' M^{-1} p(|u|) k^*(u)$ *and* $\psi$ *defined below so long as* $n\underline{h}_n/(\log\log \underline{h}_n^{-1})^3 \to \infty$ *and* $\overline{h}_n$ *is small enough for large* $n$.

Throughout, we assume that $\overline{h}_n$ is small enough so that $\|\Gamma_u(h)^{-1}\|$ and $\|\Gamma_\ell(h)^{-1}\|$ are bounded uniformly over $h \leq \overline{h}_n$ for large enough $n$ (this will hold for small enough $\overline{h}_n$ by Lemma S2.4 below).

**Lemma S2.1.** *Suppose that Assumption S2.1 holds. Then*

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \left\|\hat{\Gamma}_u(h) - \Gamma_u(h)\right\| = \mathcal{O}_P(1),$$

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \left\|\hat{\Gamma}_u(h)^{-1} - \Gamma_u(h)^{-1}\right\| = \mathcal{O}_P(1),$$

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{nh}{\log\log h^{-1}} \Bigg\|\hat{\beta}_{u,j}(h) - \beta_{u,j}(h) - \frac{1}{h} E_n \Gamma_u(h)^{-1} p(X_i/h) k^*(X_i/h)[Y_i - p(X_i/h)'\beta(h)] I(X_i \geq 0)\Bigg\| = \mathcal{O}_P(1),$$

*and*

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \left\|\hat{\beta}_{u,j}(h) - \beta_{u,j}(h)\right\| = \mathcal{O}_P(1)$$

*for each* $j$*. The same holds with* $I(X_i \geq 0)$ *replaced by* $I(X_i < 0)$*,* $\Gamma_u$ *replaced by* $\Gamma_\ell$*,* $\hat{\Gamma}_u$ *replaced by* $\hat{\Gamma}_\ell$*, etc.*

*Proof.* The first display follows from Lemma A.2. For the second display, note that $\hat{\Gamma}(h)^{-1} - \Gamma(h)^{-1} = -\hat{\Gamma}(h)^{-1}(\hat{\Gamma}(h) - \Gamma(h))\Gamma(h)^{-1}$, so $\|\hat{\Gamma}(h)^{-1} - \Gamma(h)^{-1}\| \leq \|\hat{\Gamma}(h)^{-1}\|\|\hat{\Gamma}(h) - \Gamma(h)\|\|\Gamma(h)^{-1}\|$. $\|\Gamma(h)^{-1}\|$ is bounded by assumption and $\|\hat{\Gamma}(h)^{-1}\|$ is $\mathcal{O}_P(1)$ uniformly over $\underline{h}_n \leq h \leq \overline{h}_n$ by this and the first display in the lemma. For the third display, note that

$$\hat{\beta}_{u,j}(h) - \beta_{u,j}(h) = \hat{\Gamma}_u(h)^{-1}\frac{1}{h}E_n p(X_i/h)k^*(X_i/h)[Y_i - p(X_i/h)'\beta(h)]I(X_i \geq 0).$$

Thus, letting $\mathcal{B} = -\frac{1}{h}E_n\Gamma_u(h)^{-1}p(X_i/h)k^*(X_i/h)[Y_i - p(X_i/h)'\beta(h)]I(X_i \geq 0)$,

$$\begin{aligned}\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{nh}{\log\log h^{-1}} \|\hat{\beta}_{u,j}(h) - \beta_{u,j}(h)\mathcal{B}\| \\ \leq \sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \left\|\hat{\Gamma}_u(h)^{-1} - \Gamma_u(h)^{-1}\right\| \\ \cdot \sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \left\|\frac{1}{h}E_n p(X_i/h)k^*(X_i/h)[Y_i - p(X_i/h)'\beta(h)]I(X_i \geq 0)\right\|.\end{aligned}$$

The first term is $\mathcal{O}_P(1)$ by the second display in the lemma. The second term is $\mathcal{O}_P(1)$ by Lemma A.2. The last display in the lemma follows from the third display and Lemma A.2. □

Applying the above lemma, we obtain the following.

**Lemma S2.2.** *Under Assumption S2.1,*

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{nh}{\log\log h^{-1}} \left\|\hat{\alpha}(h) - \alpha(h) - \frac{1}{nh}\sum_{i=1}^{n}\psi_\alpha(X_i, Y_i, h)k^*(X_i/h)\right\| = \mathcal{O}_P(1)$$

*and*

$$\sup_{\underline{h}_n \leq h \leq \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \|\hat{V}_\alpha(h) - V_\alpha(h)\| = \mathcal{O}_P(1).$$

*Proof.* The first claim follows by Lemma S2.1. The second claim follows by using the fact that $\hat{V}_\alpha(h)$ is a Lipschitz continuous function of the $\hat{\beta}$ and $\hat{\nu}$ terms and terms that can be handled with Lemma A.2. □

**Lemma S2.3.** *Suppose that Assumption S2.1 holds. Then*

$$\sup_{\underline{h}_n \le h \le \overline{h}_n} \sqrt{nh} \left\| \hat{\theta}(h) - \theta(h) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} D_g(\alpha(h)) \psi_\alpha(X_i, Y_i, h) k^*(X_i/h) \right\| = o_P\left(1/\sqrt{\log\log \underline{h}_n^{-1}}\right)$$

*and*

$$\sup_{\underline{h}_n \le h \le \overline{h}_n} \frac{\sqrt{nh}}{\sqrt{\log\log h^{-1}}} \|\hat{\sigma}(h) - \sigma(h)\| = \mathcal{O}_P(1).$$

*Proof.* By Lemma S2.2,

$$\sup_{\underline{h}_n \le h \le \overline{h}_n} \left\| \sqrt{nh}\,(\hat{\alpha}(h) - \alpha(h)) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \psi_\alpha(X_i, Y_i, h) k^*(X_i/h) \right\|$$
$$= \mathcal{O}_P\left( \sup_{\underline{h} \le h \le \overline{h}_n} (\log\log h^{-1})/\sqrt{nh} \right) = \mathcal{O}_P\left( (\log\log \underline{h}_n^{-1})/\sqrt{n\underline{h}_n} \right) = o_P\left(1/\sqrt{\log\log \underline{h}_n^{-1}}\right)$$

since $(\log\log \underline{h}_n^{-1})^{3/2}/\sqrt{n\underline{h}_n} \to 0$. Thus, the result follows by Lemma S1.11. □

Let $m_j(x,h) = p(x/h)'\beta_{u,j}(h)$ for $x \ge 0$ and $m_j(x,h) = p(x/h)'\beta_{\ell,j-d_Y}(h)$ for $x < 0$. Let $D_{g,u}(\alpha)$ be the row vector with the first $d_Y$ elements of $D_g(\alpha)$, and let $D_{g,\ell}(\alpha)$ be the row vector with the remaining $d_Y$ elements. With this notation, we have

$$D_g(\alpha(h))\psi_\alpha(X_i, Y_i, h)$$
$$= \left\{ I(X_i \ge 0)\nu_u(h) p(|X_i/h|) D_{g,u}(\alpha(h)) + I(X_i < 0)\nu_\ell(h) p(|X_i/h|) D_{g,\ell}(\alpha(h)) \right\} [Y_i - m(X_i, h)].$$

Let $\gamma_{u,j}(h) = \frac{1}{h} E Y_{i,j} p(|X_i/h|) k^*(X_i/h) I(X_i \ge 0)$ and $\gamma_{u,\ell}(h) = \frac{1}{h} E Y_{i,j} p(|X_i/h|) k^*(X_i/h) I(X_i < 0)$. Let $\gamma_{u,j}(0)$ be the $(r+1) \times 1$ vector with $q$th element given by $f_{X,+}\tilde{\mu}_{+,j}\mu_{k^*,q}$. Let $\gamma_{\ell,j}(0)$ be the $(r+1) \times 1$ vector with $q$th element given by $f_{X,-}\tilde{\mu}_{-,j}\mu_{k^*,q}$. Let $\alpha(0) = (\tilde{\mu}_+', \tilde{\mu}_-')'$ (it will be shown below that $\lim_{h\to 0} \alpha(h) = \alpha(0)$).

We now verify the conditions of the main result with $k(u) = e_1' M^{-1} p(|u|) k^*(u)$ and

$$\psi(W_i, h) = \frac{D_g(\alpha(h))\psi_\alpha(X_i, Y_i, h)}{e_1' M^{-1} p(|X_i/h|)\sigma(h)}$$

for $h > 0$ and

$$\psi(W_i, 0) = \frac{1}{\sigma(0)} \left[ D_{g,u}(\alpha(0)) f_{X,+}^{-1}(Y_i - \mu_+) I(X_i \geq 0) + D_{g,\ell}(\alpha(0)) f_{X,-}^{-1}(Y_i - \mu_-) I(X_i < 0) \right]$$

where $\sigma^2(0) = \lim_{h \to 0} \sigma^2(h)$ (this choice of $\psi(W_i, 0)$ will be justified by the calculations below).

**Lemma S2.4.** *Under Assumption S2.1, for some constant K,*

$$\begin{aligned} \|\Gamma_u(h) - f_{X,+} M\| &\leq K\ell(Ah), \\ \|\Gamma_\ell(h) - f_{X,-} M\| &\leq K\ell(Ah), \\ \|\gamma_u(h) - \gamma_u(0)\| &\leq K\ell(Ah), \\ \textit{and} \qquad \|\gamma_\ell(h) - \gamma_\ell(0)\| &\leq K\ell(Ah). \end{aligned}$$

*Proof.* We have

$$\begin{aligned} \gamma_{u,j}(h) &= \frac{1}{h} E Y_{i,j} p(|X_i/h|) k^*(X_i/h) I(X_i \geq 0) = \frac{1}{h} \int_{x=0}^{\infty} \tilde{\mu}_j(x) p(x/h) k^*(x/h) f_X(x)\, dx \\ &= \int_{x=0}^{\infty} \tilde{\mu}_j(uh) p(u) k^*(u) f_X(uh)\, dx. \end{aligned}$$

Thus, by boundedness of $k^*$, the quantity $\|\gamma_{u,j}(h) - \gamma_{u,j}(0)\|$ is bounded by a constant times $\sup_{0 \leq x \leq Ah} |\tilde{\mu}_j(x) f_X(x) - \tilde{\mu}_{+,j} f_{X,+}|$, which is bounded by a constant times $\ell(Ah)$ by assumption. Similarly,

$$\begin{aligned} \Gamma_{u,j,m}(h) &= \frac{1}{h} E (X_i/h)^{j+m-2} k^*(X_i/h) I(X_i \geq 0) = \frac{1}{h} \int_{x=0}^{\infty} (x/h)^{j+m-2} k^*(x/h) f_X(x)\, dx \\ &= \int_{x=0}^{\infty} u^{j+m-2} k^*(u) f_X(uh)\, du, \end{aligned}$$

so $|\Gamma_{u,j,m}(h) - f_{X,+} M_{j,m}|$ is bounded by a constant times $\sup_{0 \leq x \leq Ah} |f_X(x) - f_{X,+}| \leq \ell(Ah)$. The proof for $\Gamma_\ell$ and $\gamma_\ell$ is similar. □

Note that $\beta_{u,j}(h) = \Gamma_u(h)^{-1} \gamma_{u,j}(h) \to \tilde{\mu}_{+,j} M^{-1}(1, \mu_{k^*,1}, \ldots, \mu_{k^*,r})' = \tilde{\mu}_{+,j}(1, 0, \ldots, 0)'$ as $h \to 0$, where the last equality follows since $M^{-1}(1, \mu_{k^*,1}, \ldots, \mu_{k^*,r})'$ is the first column of $M^{-1}M = I_{r+1}$ (the second through $r$th elements of $\beta_{u,j}$ are given by the corresponding coefficients of the local polynomial scaled by powers of $h$, so this is a result of the fact that the coefficients of the local polynomial do not increase too quickly as $h \to 0$). By these calculations and Lemma S2.4, we obtain the following.

**Lemma S2.5.** *Under Assumption S2.1, for some constant $K$ and $h$ small enough,*

$$\begin{aligned} &\left|\beta_{u,j}(h) - \tilde{\mu}_{+,j}(1,0,\ldots,0)'\right| \leq K\ell(Ah), \\ \textit{and} \qquad &\left|\beta_{\ell,j}(h) - \tilde{\mu}_{-,j}(1,0,\ldots,0)'\right| \leq K\ell(Ah). \end{aligned}$$

*Proof.* The result is immediate from Lemma S2.4, the fact that $\|\Gamma_u(h)^{-1}\|$ and $\|\Gamma_\ell(h)^{-1}\|$ are bounded uniformly over small enough $h$ (which follows from Lemma S2.4 and invertibility of $M$) and fact that the function that takes $\Gamma$ and $\gamma$ to $\Gamma^{-1}\gamma$ is Lipschitz over $\Gamma$ and $\gamma$ with $\Gamma^{-1}$ and $\gamma$ bounded. □

Note that, since $\alpha(h)$ is made up of the first component of each of the $\beta_{u,j}(h)$ and $\beta_{\ell,j}(h)$ vectors, the above lemma also implies that $|\alpha(h) - \alpha(0)| \leq K\ell(Ah)$ for $\alpha(0)$ defined above. For convenience, let us also define $\beta_{u,j}(0)$ and $\beta_{\ell,j}(0)$ to be the limits of $\beta_{u,j}(h)$ and $\beta_{\ell,j}(h)$ derived above.

**Lemma S2.6.** *Under Assumption S2.1, for some constant $K$ and $h$ small enough,*

$$\|\nu_u(h) - e_1'M^{-1}f_{X,+}^{-1}\| \leq K\ell(Ah) \qquad \textit{and} \qquad \|\nu_\ell(h) - e_1'M^{-1}f_{X,-}^{-1}\| \leq K\ell(Ah).$$

*Proof.* The result is follows immediately from Lemma S2.4 and the the fact that $\|\Gamma_u(h)^{-1}\|$ and $\|\Gamma_\ell(h)^{-1}\|$ are bounded over small enough $h$. □

**Lemma S2.7.** *Under Assumption S2.1, for some constant $K$ and $h$ small enough,*

$$|[\sigma(h)\psi(W_i,h) - \sigma(0)\psi(W_i,0)]\,k(X_i/h)| \leq K\ell(Ah).$$

*Proof.* We have

$$\begin{aligned} &[\sigma(h)\psi(W_i,h) - \sigma(0)\psi(W_i,0)]\,k(X_i/h) = D_g(\alpha(h))\psi_\alpha(X_i,Y_i,h)k^*(X_i/h) \\ &- \left[D_{g,u}(\alpha(0))f_{X,+}^{-1}(Y_i-\mu_+)I(X_i\geq 0) + D_{g,\ell}(\alpha(0))f_{X,-}^{-1}(Y_i-\mu_-)I(X_i<0)\right]\cdot \\ &e_1'M^{-1}p(|X_i/h|)k^*(X_i/h) \end{aligned}$$

$$= D_g(\alpha(h))\psi_\alpha(X_i, Y_i, h)k^*(X_i/h) - D_g(\alpha(0))\tilde{\psi}_\alpha(X_i, Y_i, h)k^*(X_i/h)$$

where the first $d_Y$ columns of $\tilde{\psi}_\alpha(X_i, Y_i, h)$ are given by $e_1' M^{-1} p(|X_i/h|) f_{X,+}^{-1}(Y_i - \mu_+) I(X_i \geq 0)$ and the remaining $d_Y$ columns are given by $e_1' M^{-1} p(|X_i/h|) f_{X,-}^{-1}(Y_i - \mu_-) I(X_i < 0)$. Note that the above expression can be written as

$$T(X_i/h, Y_i, \nu_u(h), \nu_\ell(h), \alpha(h), \{\beta_{u,j,m}(h)\}_{1 \leq j \leq d_Y, 1 \leq m \leq r+1}, \{\beta_{\ell,j,m}(h)\}_{1 \leq j \leq d_Y, 1 \leq m \leq r+1})$$
$$- T(X_i/h, Y_i, \nu_u(0), \nu_\ell(0), \alpha(0), \{\beta_{u,j,m}(0)\}_{1 \leq j \leq d_Y, 1 \leq m \leq r+1}, \{\beta_{\ell,j,m}(0)\}_{1 \leq j \leq d_Y, 1 \leq m \leq r+1})$$

for a function $T$ that is Lipschitz in its remaining arguments uniformly over $X_i/h, Y_i$ on bounded sets. Combining this with the previous lemmas gives the result. □

It follows from Lemmas S2.7 and S1.12 that the conclusion of Lemma S2.7 also holds with $\sigma(h)\psi(W_i, h)$ replaced by $\psi(W_i, h)$, so long as the remaining conditions of Lemma S1.12 (those involving the conditional expectation and variance of $\psi(W_i, 0)$) hold. We have

$$E[\psi(W_i, 0)|X_i = x]$$
$$= \frac{1}{\sigma(0)} \left\{ D_{g,u}(\alpha(0)) f_{X,+}^{-1} [\tilde{\mu}(x) - \tilde{\mu}_+] I(x \geq 0) + D_{g,\ell}(\alpha(0)) f_{X,-}^{-1} [\tilde{\mu}(x) - \tilde{\mu}_-] I(x < 0) \right\}$$

and

$$var[\psi(W_i, 0)|X_i = x] = \frac{1}{\sigma^2(0)} \left\{ D_{g,u}(\alpha(0)) \tilde{\Sigma}(x) D_{g,u}(\alpha(0))' f_{X,+}^{-2} I(x \geq 0) \right.$$
$$\left. + D_{g,\ell}(\alpha(0)) \tilde{\Sigma}(x) D_{g,\ell}(\alpha(0))' f_{X,-}^{-2} I(x < 0) \right\}$$

By the conditions on $\tilde{\mu}(x)$ and $\tilde{\Sigma}(x)$, it follows that these expressions are left and right continuous in $x$ at 0 with modulus $\ell(x)$ satisfying the necessary conditions. By this and the conditions on $f_X$, it follows that the same holds for $E[\psi(W_i, 0)||X_i| = x]$ and $var[\psi(W_i, 0)||X_i| = x]$. In addition, the assumptions guarantee that $var[\psi(W_i, 0)||X_i| = x]$ is bounded away from zero for small $x$ so that $\sigma(0) > 0$.

Thus, for $\psi(W_i, h)$ defined above,

$$\begin{aligned}
&\sup_{\underline{h}_n \le h \le \overline{h}_n} \left\| \frac{\sqrt{nh}(\hat{\theta}(h) - \theta(h))}{\hat{\sigma}(h)} - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \psi(W_i, h) k(X_i/h) \right\| \\
&\le \sup_{\underline{h}_n \le h \le \overline{h}_n} \left\| \frac{\sqrt{nh}(\hat{\theta}(h) - \theta(h))}{\sigma(h)} - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \psi(W_i, h) k(X_i/h) \right\| \\
&+ \sup_{\underline{h}_n \le h \le \overline{h}_n} \left\| \sqrt{nh}(\hat{\theta}(h) - \theta(h)) \right\| \cdot \left\| \frac{1}{\sigma(h)} - \frac{1}{\hat{\sigma}(h)} \right\|.
\end{aligned}$$

By Lemma S2.3, the first term is of the order $\mathcal{O}_P(1/\sqrt{\log\log \underline{h}_n^{-1}})$, and the last term is of the order $\mathcal{O}_P(\sqrt{\log\log \underline{h}_n^{-1}} \cdot \sqrt{\log\log \underline{h}_n^{-1}}/\sqrt{n\underline{h}_n})$. Thus, for $(\log\log \underline{h}_n^{-1})^3/n\underline{h}_n \to 0$, both terms will be $o_P(1/\sqrt{\log\log \underline{h}_n^{-1}})$ as required. This completes the proof of Theorem S2.1.

### S2.1 Equivalent Kernels for Local Linear Regression

Thus section gives the equivalent kernels for local polynomial regression at the boundary and in the interior, and outlines how our results can be extended to cover local polynomial regression at local-to-boundary points. Let

$$k(u;t) = e_1' M(t)^{-1} p(u) k^*(u),$$

where

$$M(t) = \int_{u=0}^{\infty} p(u-t)p(u-t)'k^*(u-t)\,du = \int_{u=-t}^{\infty} p(u)p(u)'k^*(u)\,du. \tag{2}$$

Then the equivalent kernel for local polynomial regression at the boundary is given by $k(u;0)$. For $r = 1$, we have

$$e_1' M(0)^{-1} p(u) = e_1' \begin{pmatrix} \mu_{k^*,0} & \mu_{k^*,1} \\ \mu_{k^*,1} & \mu_{k^*,2} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ |u| \end{pmatrix} = \frac{\mu_{k^*,2} - \mu_{k^*,1}|u|}{\mu_{k^*,0}\mu_{k^*,2} - \mu_{k^*,1}^2}.$$

For $r = 2$, we have

$$e_1 M^{-1} p(u) = \frac{1}{D}\left( (\mu_{k^*,4}\mu_{k^*,2} - \mu_{k^*,3}^2) + (\mu_{k^*,1}\mu_{k^*,4} - \mu_{k^*,2}\mu_{k^*,3})\,|u| + (\mu_{k^*,2}^2 - \mu_{k^*,1}\mu_{k^*,3})\,u^2 \right),$$

where $D = \det(M) = \mu_{k^*,0}(\mu_{k^*,2}\mu_{k^*,4} - \mu_{k^*,3}^2) - \mu_{k^*,1}(\mu_{k^*,1}\mu_{k^*,4} - \mu_{k^*,2}\mu_{k^*,3}) + \mu_{k^*,2}(\mu_{k^*,1}\mu_{k^*,3} - \mu_{k^*,2}^2)$. The moments $\mu_{k^*,j}$ for the uniform, triangular, and Epanechnikov kernel are given by

| Name | $\mu_0$ | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ |
|---|---|---|---|---|---|
| Uniform | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{6}$ | $\frac{1}{8}$ | $\frac{1}{10}$ |
| Triangular | $\frac{1}{2}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{20}$ | $\frac{1}{30}$ |
| Epanechnikov | $\frac{1}{2}$ | $\frac{3}{16}$ | $\frac{1}{10}$ | $\frac{1}{16}$ | $\frac{3}{70}$ |

Plugging these moments into the definitions of equivalent kernels in the two displays above then yields the definitions of equivalent kernels for local linear and local quadratic regressions. These definitions are summarized in Table S3.

Theorem S2.1 can be extended to apply to local polynomial estimation in the interior, provided that the definition of the equivalent kernel is appropriately altered to $k(u; \infty)$ (so that the integral on the right-hand side of Equation (2) is over the whole real line rather than the interval $(0, \infty)$ as in the boundary case). Our package `BWSnooping` can be used to calculate the appropriate critical values in this case. Note that for $r = 1$, the equivalent kernel and the original kernel coincide, so that one can use Table 1 to look up the appropriate critical value.

Finally, let us outline how our results can be extended to cover estimating a conditional mean at a point that is local to the boundary of the support of the distribution of the conditioning variable. Here we can use the local-to-boundary formulation of the problem as in Section 3.2.5 of Fan and Gijbels (1996). In particular, consider local polynomial estimation of $E(Y_i \mid X_i = x_0)$ where $x_0 = c\underline{h}_n$ and the lower support point of the density of $X_i$ is zero. Letting $\hat{\theta}(h)$ denote the $r$th order local polynomial estimator based on a kernel $k^*$, it can be shown that under regularity conditions, $\sup_{h\in[\underline{h}_n, \overline{h}_n]} \sqrt{nh}|\hat{\theta}(h) - \theta(h)|/\hat{\sigma}(h)$ can be approximated by $\sup_{t\in[1, \overline{h}_n/\underline{h}_n]} |\mathbb{H}(t)|$, where $\mathbb{H}(t)$ is a Gaussian process with covariance function $cov(\mathbb{H}(s), \mathbb{H}(t)) = \rho(s, t; c)$, with

$$\rho(s, t; c) = \frac{\int_{u=-c}^{\infty} k(u/s; c/s)k(u/t; c/t)\, du}{\sqrt{\int_{u=-c}^{\infty} k(u/s; c/s)^2\, du}\sqrt{\int_{u=-c}^{\infty} k(u/t; c/t)^2\, du}}.$$

Note that the critical value depends only on $h/\underline{h}_n$ and $c$ (along with the kernel and order of the local polynomial). Similar result obtains for one-sided $t$-statistics.

## S3 Proofs for Theorems in Appendix B

### S3.1 Regression Discontinuity/LATEs for Largest Sets of Compliers

This section proves Theorems B.1 and B.3. First, note that the regression discontinuity and LATE applications can both be written as functions of local polynomial estimators in the above setup, with $d_Y = 2$ and $Y_i$ playing the role of $Y_{i,1}$ and $D_i$ playing the role of $Y_{i,2}$. For the LATE application, we define $X_i = -(Z_i - \underline{z})I(|Z_i - \underline{z}| \leq |Z_i - \overline{z}|) + (\overline{z} - Z_i)I(|Z_i - \underline{z}| > |Z_i - \overline{z}|)$. Both of these applications fit into the setup of Section S2 with, letting $\alpha(h) = (\alpha_u(h)', \alpha_\ell(h)') = (\alpha_{u,Y}(h), \alpha_{u,D}(h), \alpha_{\ell,Y}(h), \alpha_{\ell,D}(h))'$ (where we use the suggestive subscripts "$Y$" and "$D$" rather than 1 and 2), $g(\alpha) = \frac{\alpha_{u,Y} - \alpha_{\ell,Y}}{\alpha_{u,D} - \alpha_{\ell,D}}$. Then, letting $\Delta_D = \alpha_{u,D} - \alpha_{\ell,D}$, we have

$$D_g(\alpha) = \begin{bmatrix} \frac{1}{\Delta_D} & \frac{-g(\alpha)}{\Delta_D} & \frac{-1}{\Delta_D} & \frac{g(\alpha)}{\Delta_D} \end{bmatrix}.$$

This is Lipschitz continuous and bounded over bounded sets with $\alpha_{u,D} - \alpha_{\ell,D}$ bounded away from zero.

For the last condition (non-degeneracy of the conditional variance), note that $D_{g,u}(\alpha(0))\tilde{\Sigma}_+ \cdot D_{g,u}(\alpha(0)) = \frac{1}{\Delta_D(0)^2} var[Y_i - g(\alpha(0))D_i | X_i = 0_+]$, which will be nonzero so long as $corr(D_i, Y_i \mid X_i = 0_+) < 1$ and $var(Y_i \mid X_i = 0_+) > 0$. A sufficient condition for this is that $var(Y_i \mid D_i = d, X_i = 0_+) > 0$ is nonzero for $d = 0$ or $d = 1$, and this (or the corresponding statement with $+$ replaced by $-$) holds under the conditions of the theorem.

### S3.2 Trimmed Average Treatment Effects under Unconfoundedness

This section proves Theorem B.2. We first give an intuitive derivation of the critical value, which explains why it differs in this setting, and provide the technical details at the end.

To derive the form of the correction in this case, note that, under the conditions of the theorem, $\frac{\sqrt{n}(\hat{\theta}(h) - \theta(h))}{\hat{\sigma}(h)}$ will converge to a Gaussian process $\mathbb{G}(h)$ with covariance

$$cov(\mathbb{G}(h), \mathbb{G}(h')) = \frac{cov\left\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h), [\tilde{Y}_i - \theta(h')]I(X_i \in \mathcal{X}_{h'})\right\}}{\sqrt{var\left\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h)\right\} var\left\{[\tilde{Y}_i - \theta(h')]I(X_i \in \mathcal{X}_{h'})\right\}}}.$$

Let $v(h) = var\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h)\}$ as defined in the statement of the theorem. Note that, for

$h \geq h'$,

$$
\begin{aligned}
&cov\left\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h), [\tilde{Y}_i - \theta(h')]I(X_i \in \mathcal{X}_{h'})\right\} = E\left\{[\tilde{Y}_i - \theta(h)][\tilde{Y}_i - \theta(h')]I(X_i \in \mathcal{X}_h)\right\} \\
&\quad = E\left\{[\tilde{Y}_i - \theta(h)]^2 I(X_i \in \mathcal{X}_h)\right\} + [\theta(h) - \theta(h')]E\left\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h)\right\} = v(h)
\end{aligned}
$$

where the last step follows since $E\left\{[\tilde{Y}_i - \theta(h)]I(X_i \in \mathcal{X}_h)\right\} = 0$. Note also that $v(h)$ is weakly decreasing in $h$, which can be seen by noting that $v(h) = \inf_a E\left\{\left[\tilde{Y}_i - a\right]^2 I(X_i \in \mathcal{X}_h)\right\}$, since $\theta(h)$ is the conditional expectation of $\tilde{Y}_i$ given $X_i \in \mathcal{X}_h$. Thus,

$$
cov(\mathbb{G}(h), \mathbb{G}(h')) = \frac{v(h \vee h')}{\sqrt{v(h)v(h')}} = \frac{v(h) \wedge v(h')}{\sqrt{v(h)v(h')}},
$$

so $\mathbb{G}(h) \stackrel{d}{=} \frac{\mathbb{B}(v(h))}{\sqrt{v(h)}}$ where $\mathbb{B}$ is a Brownian motion. Thus, the distribution of $\sup_{\underline{h} \leq h \leq \overline{h}} \frac{\sqrt{n}(\hat{\theta}(h) - \theta(h))}{\hat{\sigma}(h)}$ can be approximated by the distribution of $\sup_{v(\overline{h}) \leq t \leq v(\underline{h})} \frac{\mathbb{B}(t)}{\sqrt{t}} \stackrel{d}{=} \sup_{1 \leq t \leq v(\underline{h})/v(\overline{h})} \frac{\mathbb{B}(t)}{\sqrt{t}}$. Note that $v(h) = \sigma(h)^2 P(X_i \in \mathcal{X}_h)^2$, so that

$$
\frac{v(\underline{h})}{v(\overline{h})} = \frac{\sigma(\underline{h})^2 P(X_i \in \mathcal{X}_{\underline{h}})^2}{\sigma(\overline{h})^2 P(X_i \in \mathcal{X}_{\overline{h}})^2}.
$$

Thus, $\hat{t}$ is a consistent estimator for $\frac{v(\underline{h})}{v(\overline{h})}$ under the conditions of the theorem.

The formal result then obtains by noting that, by Theorem 19.5 in van der Vaart (1998), $\frac{\sqrt{n}(\hat{\theta}(h) - \theta(h))}{\hat{\sigma}(h)} \stackrel{d}{\to} \mathbb{G}(h)$, taken as processes over $h \in [\underline{h}, \overline{h}]$ with the supremum norm. By the calculations above,

$$
\sup_{h \in [\underline{h}, \overline{h}]} |\mathbb{G}(h)| \stackrel{d}{=} \sup_{h \in [\underline{h}, \overline{h}]} \left| \frac{\mathbb{B}(v(h))}{\sqrt{v(h)}} \right|,
$$

where $\mathbb{B}$ is a Brownian motion. The result then follows since $\{t | v(h) = t \text{ some } h \in [\underline{h}, \overline{h}]\} \subseteq [v(\overline{h}), v(\underline{h})]$, and the two sets are equal if $v(h)$ is continuous.

## S4 Additional details for critical values

We first give a one-sided version of Theorem 3.1 and Corollary 3.1. To state the result, recall the definitions of $b(t, k)$ and the Gaussian process $\mathbb{H}(h)$ as defined in the statement of Theorem 3.1.

**Theorem S4.1.** *Let $c^{\text{os}}_{1-\alpha}(t,k)$ be the $1-\alpha$ quantile of $\sup_{1\leq h\leq t}\mathbb{H}(h)$. Suppose that $\underline{h}_n \to 0$, $\overline{h}_n = \mathcal{O}_P(1)$, and $n\underline{h}_n/[(\log\log n)(\log\log\log n)]^2 \to \infty$. Then, under Assumptions 3.1 and 3.2,*

$$P\left(\theta(\theta) \in [\hat{\theta}(h) - \hat{\sigma}(h)\cdot c^{\text{os}}_{1-\alpha}\overline{h}_n/\underline{h}_n k/\sqrt{nh}, \infty) \text{ all } h \in [\underline{h}_n \leq h \leq \overline{h}_n]\right) \overset{n\to\infty}{\to} 1-\alpha$$

*The above display also holds with $c^{\text{os}}_{1-\alpha}(\overline{h}_n/\underline{h}_n, k)$ replaced by*

$$\frac{-\log\left(-\log(1-\alpha)\right) + b(\overline{h}_n/\underline{h}_n, k)}{\sqrt{2\log\log(\overline{h}_n/\underline{h}_n)}} + \sqrt{2\log\log(\overline{h}_n/\underline{h}_n)},$$

*provided $\overline{h}_n/\underline{h}_n \to \infty$. If $\sup_{h\in[\underline{h}_n,\overline{h}_n]} \frac{\sqrt{nh}(\theta(h)-\theta(0))}{\hat{\sigma}(h)} \leq o_P\left((\log\log(\overline{h}_n/\underline{h}_n))^{-1/2}\right)$, then*

$$\liminf_{n\to\infty} P\left(\theta(0) \in [\hat{\theta}(h) - \hat{\sigma}(h)\cdot c^{\text{os}}_{1-\alpha}\overline{h}_n/\underline{h}_n k/\sqrt{nh}, \infty) \text{ all } h \in [\underline{h}_n \leq h \leq \overline{h}_n]\right) \geq 1-\alpha.$$

Unlike in the two-sided case, the bias does not have to be negligible so long as it can be signed: if $\theta(h) - \theta(0)$ is known to be weakly negative (positive), then bias can only improve the coverage of a lower (upper) one-sided CI (see Section 4.1.2). The proof of Theorem S4.1 is analogous to the proof of Theorem 3.1 given in Appendix A.

Tables S1 and S2 give two- and one-sided critical values $c^{\text{os}}_{1-\alpha}(\overline{h}_n/\underline{h}_n, k)$ and $c_{1-\alpha}(\overline{h}_n/\underline{h}_n, k)$ for several kernel functions $k$, $\alpha$ and a selected of values of $\overline{h}_n/\underline{h}_n$ for 90%, 95%, and 99% confidence intervals. The critical values can also be obtained using our R package `BWSnooping`, which can be downloaded from `https://github.com/kolesarm/BWSnooping`. The package also includes critical values for local quadratic regression, and computes critical values for other significance levels and other ratios of maximum to minimum bandwidth $\overline{h}/\underline{h}$.

For comparison, Figure S1 plots critical values based on the extreme value approximation (given in the second part of Theorem 3.1) along with those based directly on the Gaussian process.

## S5 Monte Carlo evidence

We conduct a small Monte Carlo study of inference in a sharp regression discontinuity design to further illustrate our method and to examine how well it works in practice. In each replication, we generated a random sample $\{X_i, \varepsilon_i\}_{i=1}^n$, with size $n = 500$, $X_i = 2Z_i - 1$, where $Z_i$ has Beta distribution with parameters 2 and 4, and $\varepsilon_i \sim \mathcal{N}(0, 0.1295^2)$. The regression discontinuity point is normalized to zero. The outcome $Y_i$ is given by $Y_i = g_j(X_i) + \varepsilon_i$, where the regression function $g_j$ depends on the design. We consider two regression functions. The first one corresponds to a polynomial fit to the Lee (2008) data,

$$g_1(x) = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0, \\ 0.52 + 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{otherwise.} \end{cases}$$

This design corresponds exactly to the data generating process in Imbens and Kalyanaraman (2012, IK) and Calonico et al. (2014, CCT). The second regression function corresponds to another design in IK, and is given by

$$g_2(x) = 0.42 + 0.1I(x \geq 0) + 0.84x + 7.99x^3 - 9.01x^4 + 3.56x^5.$$

Figure S2 plots the conditional mean functions $g_1$ and $g_2$ that generate the data in Designs 1 and 2. The results for designs in which the error term $\varepsilon_i$ is heteroscedastic are very similar, and reported in an earlier version of the paper (Armstrong and Kolesár, 2015).

In each design, we consider estimates based on local linear regression using the uniform and the triangular kernel. We use the bandwidth selector proposed by IK to select a baseline bandwidth, and then construct confidence bands for estimators in bandwidth range around this baseline bandwidth. We also consider the robust bias correction method of CCT discussed in Section 4.1 by running a local quadratic regression at the same bandwidths. To define these estimators, let $p(x) = (1, x, \ldots, x^r)$ denote a polynomial expansion of order $r$. Given an i.i.d. sample $\{Y_i, X_i\}_{i=1}^n$, the RD estimator is given by the difference between the intercepts of polynomial linear regressions of order $r$ with the same bandwidth on either side of the cutoff,

$$\hat{\theta}(h) = \hat{\alpha}_u(h) - \hat{\alpha}_\ell(h),$$

where $\hat{\alpha}_u(h) = e_1'\beta_u(h)$, $\hat{\alpha}_\ell(h) = e_1'\beta_\ell(h)$,

$$\hat{\beta}_u(h) = \hat{\Gamma}_u(h)^{-1} \sum_{i=1}^n I(X_i \geq 0) k^*(X_i/h) p(|X_i|) Y_i,$$

$$\hat{\beta}_\ell(h) = \hat{\Gamma}_l(h)^{-1} \sum_{i=1}^{n} I(X_i < 0) k^*(X_i/h) p(|X_i|) Y_i,$$

$k^*$ is a kernel, and

$$\hat{\Gamma}_u(h) = \sum_i I(X_i \geq 0) k^*(X_i/h) p(|X_i|) p(|X_i|)',$$
$$\hat{\Gamma}_\ell(h) = \sum_i I(X_i < 0) k^*(X_i/h) p(|X_i|) p(|X_i|)'.$$

The corresponding function $\theta(h)$ is plotted in Figures S3 and S4 for the local linear and local quadratic estimators.

To estimate the variance of the estimator, we use the Eicker-Huber-White (EHW) robust variance estimator that treats the two linear linear regressions on either side of the cutoff as a weighted linear regression. In Theorem B.1 below, we show formally that using this estimator leads to uniformly valid confidence intervals. We also consider a modification of the EHW estimator that uses a nearest neighbor (NN) estimator to estimate $var(Y_i \mid X_i)$ in the middle part of the Eicker-Huber-White "sandwich", rather than using the regression residuals. This estimator was introduced by Abadie and Imbens (2006) and Abadie et al. (2014), and it was studied by Calonico et al. (2014) in an RD context. The nearest neighbor (NN) and EHW variance estimators have the form

$$\hat{\sigma}^2(h) = nh \left( \widehat{var}(\hat{\alpha}_u(h)) + \widehat{var}(\hat{\alpha}_\ell(h)) \right),$$

where

$$\widehat{var}(\hat{\alpha}_u(h)) = e_1' \hat{\Gamma}_u(h)^{-1} \left( \sum_{i=1}^{n} I(X_i \geq 0) \hat{\sigma}_u^2(X_i) k^*(X_i/h) p(|X_i|) p(|X_i|)' \right) \hat{\Gamma}_u(h)^{-1} e_1$$

and similarly for $\widehat{var}(\hat{\alpha}_u(h))$, where $\hat{\sigma}_u^2(X_i)$ and $\hat{\sigma}_\ell^2(X_i)$ are some estimators of $var(Y_i \mid X_i)$. The EHW estimator sets $\hat{\sigma}_u^2(X_i) = (Y_i - X_i'\hat{\beta}_u)^2$, and the NN estimators sets

$$\hat{\sigma}_u^2(X_i) = I(X_i \geq 0) \frac{J}{J+1} \left( Y_i - \sum_{j=1}^{J} Y_{\ell_{u,j}(i)} \right)^2,$$

where $\ell_{u,j}(i)$ is the $j$th closest unit to $i$ among $\{k \neq i \colon X_k \geq 0\}$, and $J = 3$.

Table S4 reports empirical coverage of the confidence bands for $\theta(h)$ for the two designs we consider. Our adjustment works well overall, with the empirical coverage being close to 95% for almost all specifications, in contrast with the naive confidence bands (using the unadjusted 1.96 critical value), which undercover. As plotted in Figure 2, Theorem 3.1 predicts that with $\bar{h}/\underline{h} = 2$, the coverage should be 91.6% for the triangular kernel, and 83.9% for the uniform kernel. When $\bar{h}/\underline{h} = 4$, the coverage of the naive confidence bands should drop to 88.5% and 76.8%, respectively. The Monte Carlo results match these predictions closely. There are a few specifications in which the adjusted confidence bands based on EHW standard errors undercover. This happens when small bandwidths are considered, and is due to the well-known downward bias of EHW standard errors in small samples, so that the pointwise confidence intervals fail to achieve nominal coverage in the first place. Since our method only corrects for the multiple comparisons, it cannot solve this problem. Overall, the adjusted confidence bands have coverage that is as good as the coverage of the underlying pointwise confidence intervals.

Typically in regression discontinuity studies, the primary object of interest is $\theta(0)$, the average treatment effect conditional on $X = 0$. We therefore also report empirical coverage of the confidence bands for $\theta(0)$ in Table S5. Confidence bands around undersmoothed local linear estimator, (that correspond to the bandwidth range $[\hat{h}_{IK}/4, \hat{h}_{IK}/2]$) perform well, provided NN standard errors, which perform better in small samples, are used. At larger values of the bandwidth, $\hat{\theta}(h)$ is a biased estimator of $\theta(0)$. The pointwise confidence intervals based on the local linear regression do not take this bias into account, and they fail to achieve proper coverage. Consequently, although our adjustment ensures that the coverage of the adjusted confidence band is within the range of the pointwise confidence intervals, it still falls short of 95% due to the pointwise confidence intervals performing poorly. On the other hand, confidence bands around the bias-adjusted confidence intervals (that correspond to local quadratic regression) perform well, especially when the NN standard errors are used.

In conclusion, our adjustment performs well in terms of coverage of $\theta(h)$, with empirical coverage close to nominal coverage, especially when combined with NN standard errors. If our method is combined with undersmoothing (corresponding to bandwidth ranges smaller than $\hat{h}_{IK}$), or bias-correction (such as when the CCT method for constructing confidence intervals is used), so that the underlying pointwise confidence intervals achieve good coverage of $\theta(0)$, our method also achieves good coverage of $\theta(0)$.

## References


Abadie, A. and G. W. Imbens (2006): "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 74, 235–267.

Abadie, A., G. W. Imbens, and F. Zheng (2014): "Inference for Misspecified Models With Fixed Regressors," *Journal of the American Statistical Association*, 109, 1601–1614.

Armstrong, T. B. and M. Kolesár (2015): "A Simple Adjustment for Bandwidth Snooping," ArXiv: 1412.0267v3.

Bickel, P. J. and M. Rosenblatt (1973): "On some global measures of the deviations of density function estimates," *The Annals of Statistics*, 1, 1071–1095.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Fan, J. and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*, CRC Press.

Imbens, G. and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 79, 933–959.

Klein, T. and E. Rio (2005): "Concentration around the mean for maxima of empirical processes," *The Annals of Probability*, 33, 1060–1077.

Leadbetter, M. R., G. Lindgren, and H. Rootzen (1983): *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer.

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142, 675–697.

Massart, P. (2000): "About the Constants in Talagrand's Concentration Inequalities for Empirical Processes," *The Annals of Probability*, 28, 863–884.

Mörters, P. and Y. Peres (2010): *Brownian Motion*, Cambridge, UK; New York, NY: Cambridge University Press.

Pollard, D. (1984): *Convergence of stochastic processes*, New York, NY: Springer.

Sakhanenko, A. I. (1985): "Convergence rate in the invariance principle for non-identically distributed variables with exponential moments," in *Advances in Probability Theory: Limit Theorems for Sums of Random Variables*, ed. by A. A. Borovkov, Springer, 2–73.

Shao, Q.-M. (1995): "Strong Approximation Theorems for Independent Random Variables and Their Applications," *Journal of Multivariate Analysis*, 52, 107–130.

Talagrand, M. (1996): "New concentration inequalities in product spaces," *Inventiones mathematicae*, 126, 505–563.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge, UK; New York, NY: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak convergence and empirical processes*, Springer.

| | NW / Loc. linear (interior) | | | | | | | | | Loc. linear (boundary) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Unif | | | Tri | | | Epa | | | Unif | | | Tri | | | Epa | | |
| $\overline{h}/\underline{h}$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 1.0 | 1.65 | 1.96 | 2.57 | 1.64 | 1.96 | 2.58 | 1.64 | 1.96 | 2.58 | 1.63 | 1.96 | 2.57 | 1.64 | 1.95 | 2.57 | 1.64 | 1.96 | 2.57 |
| 1.2 | 1.92 | 2.24 | 2.85 | 1.70 | 2.01 | 2.63 | 1.71 | 2.03 | 2.65 | 1.92 | 2.23 | 2.83 | 1.72 | 2.03 | 2.64 | 1.73 | 2.05 | 2.66 |
| 1.4 | 2.02 | 2.33 | 2.93 | 1.74 | 2.05 | 2.67 | 1.77 | 2.08 | 2.69 | 2.02 | 2.33 | 2.93 | 1.77 | 2.08 | 2.69 | 1.80 | 2.11 | 2.72 |
| 1.6 | 2.09 | 2.40 | 2.98 | 1.78 | 2.09 | 2.70 | 1.81 | 2.12 | 2.72 | 2.09 | 2.39 | 3.00 | 1.80 | 2.12 | 2.73 | 1.84 | 2.15 | 2.76 |
| 1.8 | 2.14 | 2.45 | 3.03 | 1.81 | 2.11 | 2.72 | 1.85 | 2.15 | 2.75 | 2.14 | 2.44 | 3.04 | 1.84 | 2.16 | 2.76 | 1.88 | 2.19 | 2.81 |
| 2 | 2.18 | 2.48 | 3.07 | 1.83 | 2.14 | 2.75 | 1.87 | 2.17 | 2.78 | 2.18 | 2.48 | 3.08 | 1.87 | 2.18 | 2.78 | 1.91 | 2.22 | 2.83 |
| 3 | 2.30 | 2.60 | 3.18 | 1.91 | 2.22 | 2.83 | 1.96 | 2.27 | 2.86 | 2.30 | 2.60 | 3.18 | 1.96 | 2.27 | 2.86 | 2.01 | 2.32 | 2.91 |
| 4 | 2.37 | 2.66 | 3.24 | 1.96 | 2.26 | 2.86 | 2.02 | 2.31 | 2.92 | 2.36 | 2.66 | 3.24 | 2.01 | 2.32 | 2.90 | 2.06 | 2.37 | 2.95 |
| 5 | 2.41 | 2.70 | 3.28 | 2.00 | 2.30 | 2.90 | 2.05 | 2.35 | 2.95 | 2.41 | 2.71 | 3.27 | 2.05 | 2.35 | 2.94 | 2.11 | 2.41 | 2.99 |
| 6 | 2.44 | 2.73 | 3.31 | 2.02 | 2.32 | 2.92 | 2.08 | 2.37 | 2.97 | 2.44 | 2.73 | 3.31 | 2.08 | 2.37 | 2.96 | 2.13 | 2.43 | 3.01 |
| 7 | 2.47 | 2.75 | 3.34 | 2.04 | 2.34 | 2.94 | 2.10 | 2.39 | 2.99 | 2.47 | 2.76 | 3.33 | 2.10 | 2.39 | 2.98 | 2.16 | 2.45 | 3.04 |
| 8 | 2.49 | 2.77 | 3.35 | 2.06 | 2.35 | 2.95 | 2.12 | 2.41 | 3.01 | 2.49 | 2.78 | 3.35 | 2.12 | 2.41 | 2.99 | 2.18 | 2.47 | 3.05 |
| 9 | 2.51 | 2.79 | 3.37 | 2.07 | 2.37 | 2.96 | 2.14 | 2.42 | 3.02 | 2.50 | 2.79 | 3.37 | 2.14 | 2.43 | 3.00 | 2.20 | 2.48 | 3.06 |
| 10 | 2.52 | 2.80 | 3.38 | 2.08 | 2.38 | 2.97 | 2.15 | 2.44 | 3.04 | 2.52 | 2.81 | 3.39 | 2.15 | 2.44 | 3.01 | 2.21 | 2.50 | 3.07 |
| 20 | 2.61 | 2.89 | 3.45 | 2.16 | 2.45 | 3.03 | 2.23 | 2.51 | 3.10 | 2.61 | 2.89 | 3.45 | 2.23 | 2.52 | 3.08 | 2.29 | 2.58 | 3.14 |
| 50 | 2.70 | 2.97 | 3.51 | 2.24 | 2.53 | 3.10 | 2.31 | 2.59 | 3.15 | 2.70 | 2.98 | 3.52 | 2.32 | 2.60 | 3.16 | 2.38 | 2.66 | 3.21 |
| 100 | 2.75 | 3.02 | 3.56 | 2.29 | 2.57 | 3.14 | 2.36 | 2.64 | 3.20 | 2.76 | 3.02 | 3.56 | 2.37 | 2.65 | 3.20 | 2.44 | 2.71 | 3.25 |

Table S1: Critical values $c_{1-\alpha}(\overline{h}/\underline{h}, k)$ for level $\alpha = 0.1$, 0.05, and 0.01 for the Uniform (Unif, $k(u) = \frac{1}{2}I(|u| \leq 1)$), Triangular (Tri, $(1-|u|)I(|u| \leq 1)$) and Epanechnikov (Epa, $3/4(1-u^2)I(|u| \leq 1)$) kernels. "NW / Loc. linear (interior)" refers to Nadaraya-Watson (local constant) regression in the interior or at a boundary, as well as local linear regression in the interior. "Loc. linear (boundary)" refers to local linear regression at a boundary (including regression discontinuity designs).

| | NW / Loc. linear (interior) | | | | | | | | | Loc. linear (boundary) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Unif | | | Tri | | | Epa | | | Unif | | | Tri | | | Epa | | |
| $\overline{h}/\underline{h}$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 1.0 | 1.29 | 1.66 | 2.33 | 1.29 | 1.66 | 2.34 | 1.29 | 1.66 | 2.34 | 1.28 | 1.64 | 2.33 | 1.29 | 1.65 | 2.33 | 1.28 | 1.65 | 2.33 |
| 1.2 | 1.57 | 1.94 | 2.64 | 1.35 | 1.72 | 2.39 | 1.36 | 1.73 | 2.41 | 1.57 | 1.93 | 2.59 | 1.36 | 1.72 | 2.40 | 1.38 | 1.74 | 2.42 |
| 1.4 | 1.67 | 2.04 | 2.73 | 1.39 | 1.76 | 2.44 | 1.42 | 1.79 | 2.45 | 1.67 | 2.03 | 2.69 | 1.41 | 1.78 | 2.46 | 1.44 | 1.80 | 2.47 |
| 1.6 | 1.75 | 2.11 | 2.79 | 1.42 | 1.80 | 2.47 | 1.46 | 1.83 | 2.50 | 1.74 | 2.10 | 2.76 | 1.46 | 1.81 | 2.49 | 1.49 | 1.85 | 2.52 |
| 1.8 | 1.80 | 2.15 | 2.83 | 1.46 | 1.83 | 2.49 | 1.49 | 1.86 | 2.53 | 1.80 | 2.15 | 2.81 | 1.49 | 1.84 | 2.52 | 1.53 | 1.88 | 2.55 |
| 2 | 1.84 | 2.19 | 2.85 | 1.48 | 1.85 | 2.52 | 1.52 | 1.89 | 2.55 | 1.84 | 2.19 | 2.84 | 1.52 | 1.87 | 2.55 | 1.56 | 1.91 | 2.58 |
| 3 | 1.97 | 2.31 | 2.97 | 1.56 | 1.93 | 2.58 | 1.62 | 1.98 | 2.64 | 1.96 | 2.30 | 2.95 | 1.62 | 1.96 | 2.62 | 1.67 | 2.01 | 2.67 |
| 4 | 2.04 | 2.38 | 3.02 | 1.61 | 1.97 | 2.63 | 1.68 | 2.03 | 2.68 | 2.03 | 2.36 | 3.02 | 1.67 | 2.01 | 2.67 | 1.73 | 2.06 | 2.72 |
| 5 | 2.09 | 2.42 | 3.05 | 1.65 | 2.01 | 2.66 | 1.71 | 2.07 | 2.71 | 2.08 | 2.41 | 3.04 | 1.71 | 2.05 | 2.70 | 1.77 | 2.11 | 2.76 |
| 6 | 2.12 | 2.45 | 3.08 | 1.68 | 2.03 | 2.68 | 1.74 | 2.09 | 2.74 | 2.12 | 2.44 | 3.07 | 1.74 | 2.08 | 2.72 | 1.80 | 2.13 | 2.77 |
| 7 | 2.15 | 2.48 | 3.10 | 1.71 | 2.05 | 2.70 | 1.77 | 2.11 | 2.76 | 2.14 | 2.47 | 3.09 | 1.76 | 2.10 | 2.74 | 1.83 | 2.16 | 2.80 |
| 8 | 2.17 | 2.50 | 3.12 | 1.72 | 2.07 | 2.72 | 1.79 | 2.13 | 2.77 | 2.17 | 2.49 | 3.11 | 1.79 | 2.12 | 2.75 | 1.85 | 2.18 | 2.81 |
| 9 | 2.19 | 2.52 | 3.14 | 1.74 | 2.09 | 2.73 | 1.80 | 2.14 | 2.79 | 2.18 | 2.51 | 3.12 | 1.80 | 2.14 | 2.77 | 1.87 | 2.20 | 2.82 |
| 10 | 2.21 | 2.53 | 3.16 | 1.76 | 2.10 | 2.74 | 1.82 | 2.16 | 2.81 | 2.20 | 2.52 | 3.13 | 1.82 | 2.15 | 2.79 | 1.88 | 2.21 | 2.84 |
| 20 | 2.29 | 2.62 | 3.23 | 1.83 | 2.17 | 2.80 | 1.91 | 2.24 | 2.87 | 2.29 | 2.61 | 3.22 | 1.90 | 2.23 | 2.86 | 1.97 | 2.29 | 2.91 |
| 50 | 2.40 | 2.71 | 3.31 | 1.92 | 2.25 | 2.87 | 2.00 | 2.32 | 2.94 | 2.39 | 2.70 | 3.30 | 1.99 | 2.31 | 2.92 | 2.06 | 2.38 | 2.99 |
| 100 | 2.46 | 2.77 | 3.36 | 1.98 | 2.30 | 2.92 | 2.06 | 2.37 | 2.99 | 2.45 | 2.76 | 3.35 | 2.05 | 2.37 | 2.96 | 2.12 | 2.44 | 3.03 |

Table S2: One-sided critical values $c^{\text{os}}_{1-\alpha}(\overline{h}/\underline{h}, k)$ for level $\alpha = 0.1$, 0.05, and 0.01 for the Uniform (Unif, $k(u) = \frac{1}{2}I(|u| \leq 1)$), Triangular (Tri, $(1-|u|)I(|u| \leq 1)$) and Epanechnikov (Epa, $3/4(1-u^2)I(|u| \leq 1)$) kernels. "NW / Loc. linear (interior)" refers to Nadaraya-Watson (local constant) regression in the interior or at a boundary, as well as local linear regression in the interior. "Loc. linear (boundary)" refers to local linear regression at a boundary (including regression discontinuity designs).

| Name | $k^*(u)$ | Order | $k(u)$ |
|---|---|---|---|
| | | 0 | $\frac{1}{2}I(\lvert u\rvert \leq 1)$ |
| Uniform | $\frac{1}{2}I(\lvert u\rvert \leq 1)$ | 1 | $(4-6\lvert u\rvert)I(\lvert u\rvert \leq 1)$ |
| | | 2 | $(9-36\lvert u\rvert+30u^2)I(\lvert u\rvert \leq 1)$ |
| | | 0 | $(1-\lvert u\rvert)_+$ |
| Triangular | $(1-\lvert u\rvert)_+$ | 1 | $6(1-2\lvert u\rvert)(1-\lvert u\rvert)_+$ |
| | | 2 | $12(1-5\lvert u\rvert+5u^2)(1-\lvert u\rvert)_+$ |
| | | 0 | $\frac{3}{4}(1-u^2)_+$ |
| Epanechnikov | $\frac{3}{4}(1-u^2)_+$ | 1 | $\frac{6}{19}(16-30\lvert u\rvert)(1-u^2)_+$ |
| | | 2 | $\frac{1}{8}(85-400\lvert u\rvert+385u^2)(1-u^2)_+$ |

Table S3: Definitions of kernels and equivalent kernels for regression discontinuity / estimation at a boundary. Order refers to the order of the local polynomial.

| $(\underline{h}, \bar{h})$ | $\hat{\sigma}(h)$ | Uniform kernel | | | Triangular kernel | | |
|---|---|---|---|---|---|---|---|
| | | Pointwise | Naive | Adj. | Pointwise | Naive | Adj. |
| Design 1: Local Linear regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (92.7, 94.4) | 83.7 | 93.9 | (91.8, 94.0) | 88.5 | 92.3 |
| | NN | (94.6, 95.8) | 87.3 | 95.3 | (94.2, 95.3) | 91.2 | 94.2 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (94.2, 94.7) | 85.0 | 95.0 | (93.9, 94.5) | 90.5 | 94.0 |
| | NN | (95.3, 96.1) | 87.9 | 96.3 | (94.9, 95.9) | 92.3 | 95.3 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (90.4, 94.7) | 74.8 | 93.4 | (92.1, 94.5) | 85.6 | 93.0 |
| | NN | (91.8, 96.1) | 77.4 | 94.4 | (93.4, 95.9) | 88.2 | 94.4 |
| Design 1: Local quadratic regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (89.5, 92.6) | 78.1 | 90.2 | (88.5, 91.9) | 83.3 | 88.7 |
| | NN | (93.8, 94.8) | 85.2 | 94.3 | (93.2, 94.5) | 89.0 | 93.0 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (92.7, 94.3) | 82.7 | 93.5 | (92.0, 94.0) | 88.0 | 92.4 |
| | NN | (94.8, 95.7) | 87.1 | 95.5 | (94.5, 95.4) | 91.3 | 94.6 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (84.4, 95.1) | 68.6 | 90.1 | (89.4, 94.8) | 80.1 | 89.7 |
| | NN | (87.1, 96.2) | 74.9 | 92.9 | (91.3, 96.0) | 84.5 | 92.5 |
| Design 2: Local Linear regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (85.6, 91.4) | 73.7 | 86.5 | (83.0, 90.0) | 78.3 | 83.2 |
| | NN | (93.3, 94.3) | 84.6 | 93.9 | (92.7, 93.5) | 88.7 | 92.1 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (91.3, 92.7) | 80.3 | 91.9 | (90.2, 92.1) | 86.0 | 90.3 |
| | NN | (94.0, 94.6) | 85.3 | 94.3 | (93.5, 94.2) | 90.2 | 93.2 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (88.5, 92.8) | 70.6 | 90.4 | (86.3, 92.1) | 78.8 | 87.7 |
| | NN | (85.0, 94.8) | 73.0 | 91.3 | (81.0, 94.2) | 76.0 | 85.3 |
| Design 2: Local quadratic regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (74.6, 86.9) | 57.6 | 72.9 | (74.3, 85.4) | 65.5 | 72.0 |
| | NN | (92.2, 93.1) | 80.7 | 91.6 | (91.5, 92.4) | 85.1 | 89.6 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (87.5, 91.5) | 74.1 | 87.6 | (85.8, 90.9) | 80.3 | 85.9 |
| | NN | (92.2, 93.3) | 81.6 | 92.1 | (91.6, 92.8) | 86.8 | 90.6 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (87.5, 94.6) | 67.6 | 88.3 | (85.8, 94.0) | 76.8 | 86.6 |
| | NN | (92.2, 95.2) | 74.8 | 92.2 | (91.6, 94.5) | 83.2 | 91.0 |

Table S4: Monte Carlo study of regression discontinuity. Empirical coverage of $\theta(h)$ for nominal 95% confidence bands around IK bandwidth. "Pointwise" refers to range of coverage of pointwise confidence intervals. "Naive" refers to the coverage of the naive confidence band that uses the unadjusted critical value equal to 1.96. "Adj." refers to confidence bands using adjusted critical values based on Theorem 3.1. Variance estimators are described in the text. 10,000 Monte Carlo draws, 100 grid points for $h$.

| $(\underline{h}, \bar{h})$ | $\hat{\sigma}(h)$ | Uniform kernel | | | Triangular kernel | | |
|---|---|---|---|---|---|---|---|
| | | Pointwise | Naive | Adj. | Pointwise | Naive | Adj. |
| Design 1: Local Linear regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (90.3, 92.7) | 79.9 | 92.0 | (90.0, 92.1) | 85.5 | 90.1 |
| | NN | (92.4, 94.7) | 84.5 | 94.4 | (92.4, 94.1) | 89.2 | 92.7 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (73.7, 89.8) | 62.0 | 80.7 | (76.7, 89.5) | 74.0 | 80.9 |
| | NN | (77.1, 92.0) | 66.9 | 84.2 | (80.1, 91.9) | 78.0 | 84.1 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (73.2, 89.8) | 54.3 | 80.9 | (76.5, 89.5) | 69.4 | 81.1 |
| | NN | (76.7, 92.0) | 59.8 | 84.9 | (79.9, 91.9) | 74.2 | 84.9 |
| Design 1: Local quadratic regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (89.6, 92.7) | 78.3 | 90.0 | (88.6, 92.2) | 83.5 | 88.6 |
| | NN | (93.8, 94.7) | 85.1 | 94.4 | (93.2, 94.3) | 88.8 | 93.0 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (90.3, 93.6) | 80.2 | 92.2 | (89.3, 93.0) | 85.2 | 90.0 |
| | NN | (92.4, 95.3) | 84.8 | 94.7 | (91.5, 94.7) | 88.4 | 92.8 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (74.8, 93.6) | 54.8 | 81.8 | (78.6, 93.0) | 70.3 | 82.5 |
| | NN | (78.6, 95.3) | 62.2 | 86.3 | (82.2, 94.7) | 75.3 | 86.5 |
| Design 2: Local Linear regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (85.6, 91.4) | 73.7 | 86.5 | (83.0, 90.1) | 78.3 | 83.2 |
| | NN | (93.3, 94.3) | 84.6 | 93.9 | (92.7, 93.5) | 88.7 | 92.1 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (91.3, 92.6) | 80.2 | 91.8 | (90.2, 91.9) | 85.7 | 90.1 |
| | NN | (94.0, 94.6) | 85.1 | 94.3 | (93.5, 94.1) | 89.9 | 93.1 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (59.3, 92.6) | 47.8 | 75.3 | (53.9, 91.9) | 47.4 | 60.7 |
| | NN | (63.1, 94.6) | 54.0 | 79.8 | (57.6, 94.1) | 52.9 | 65.5 |
| Design 2: Local quadratic regression | | | | | | | |
| $(\hat{h}_{IK}/4, \hat{h}_{IK}/2)$ | EHW | (74.6, 86.9) | 57.6 | 72.9 | (74.3, 85.4) | 65.5 | 72.0 |
| | NN | (93.6, 94.8) | 84.4 | 93.6 | (92.8, 94.0) | 87.7 | 91.6 |
| $(\hat{h}_{IK}/2, \hat{h}_{IK})$ | EHW | (87.5, 91.5) | 74.2 | 87.6 | (85.8, 90.9) | 80.3 | 85.9 |
| | NN | (93.5, 94.4) | 84.1 | 93.8 | (92.9, 93.8) | 88.4 | 92.4 |
| $(\hat{h}_{IK}/2, 2\hat{h}_{IK})$ | EHW | (87.5, 94.3) | 67.2 | 88.0 | (85.8, 93.6) | 75.9 | 86.1 |
| | NN | (93.5, 95.8) | 78.3 | 93.8 | (92.9, 95.1) | 84.6 | 92.4 |

Table S5: Monte Carlo study of regression discontinuity. Empirical coverage of $\theta(0)$ for nominal 95% confidence bands around IK bandwidth. "Pointwise" refers to range of coverage of pointwise confidence intervals. "Naive" refers to the coverage of the naive confidence band that uses the unadjusted critical value equal to 1.96. "Adj." refers to confidence bands using adjusted critical values based on Theorem 3.1. Variance estimators are described in the text. 10,000 Monte Carlo draws, 100 grid points for $h$.

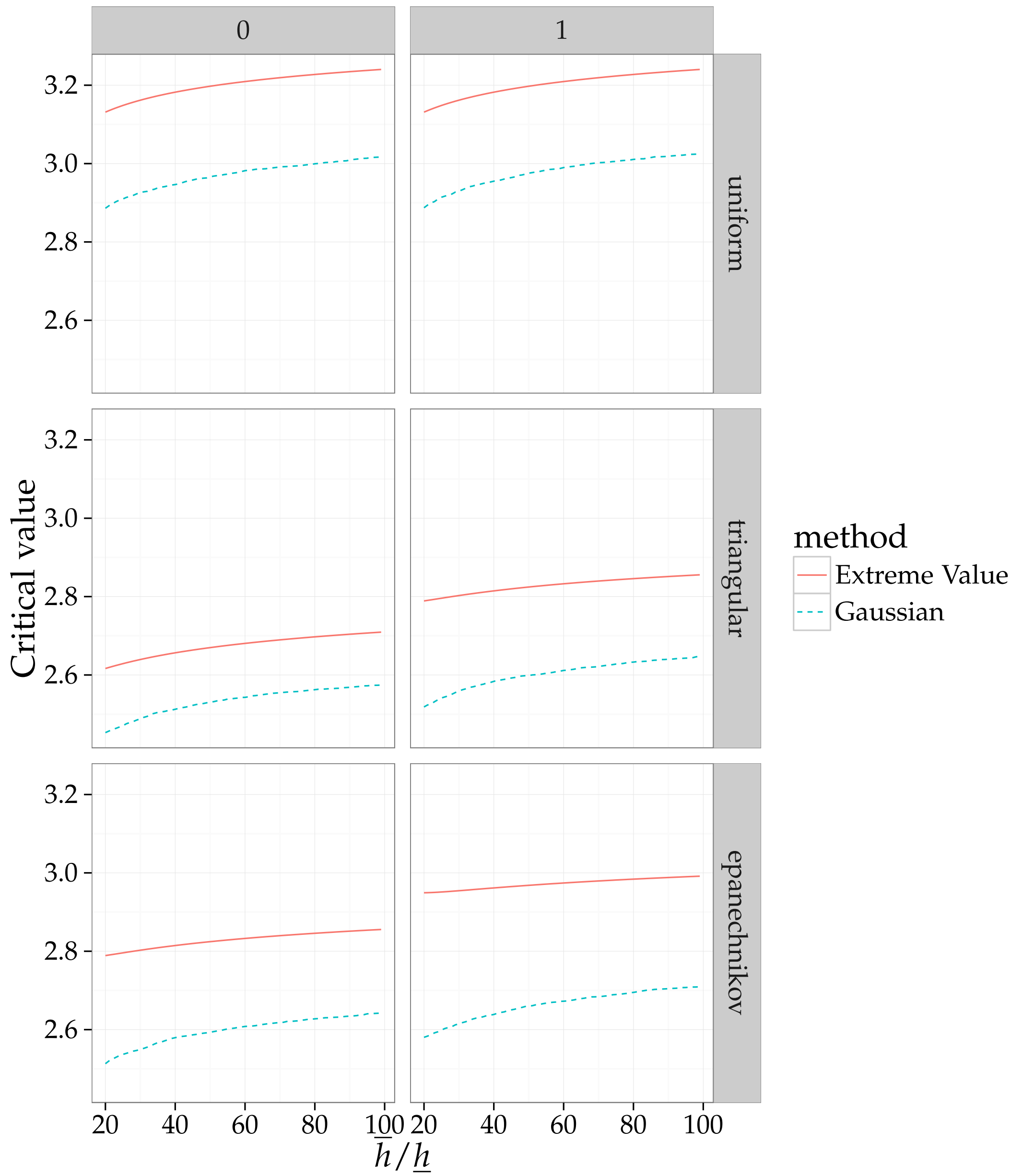


Figure S1: Comparison of critical values based on Gaussian approximation and extreme value approximation (i.e. asymptotic approximation as $\overline{h}/\underline{h} \to \infty$). Order "0" corresponds to Nadaraya-Watson interior or boundary regression, and to local linear regression in the interior, and order "1" to local linear regression at a boundary.

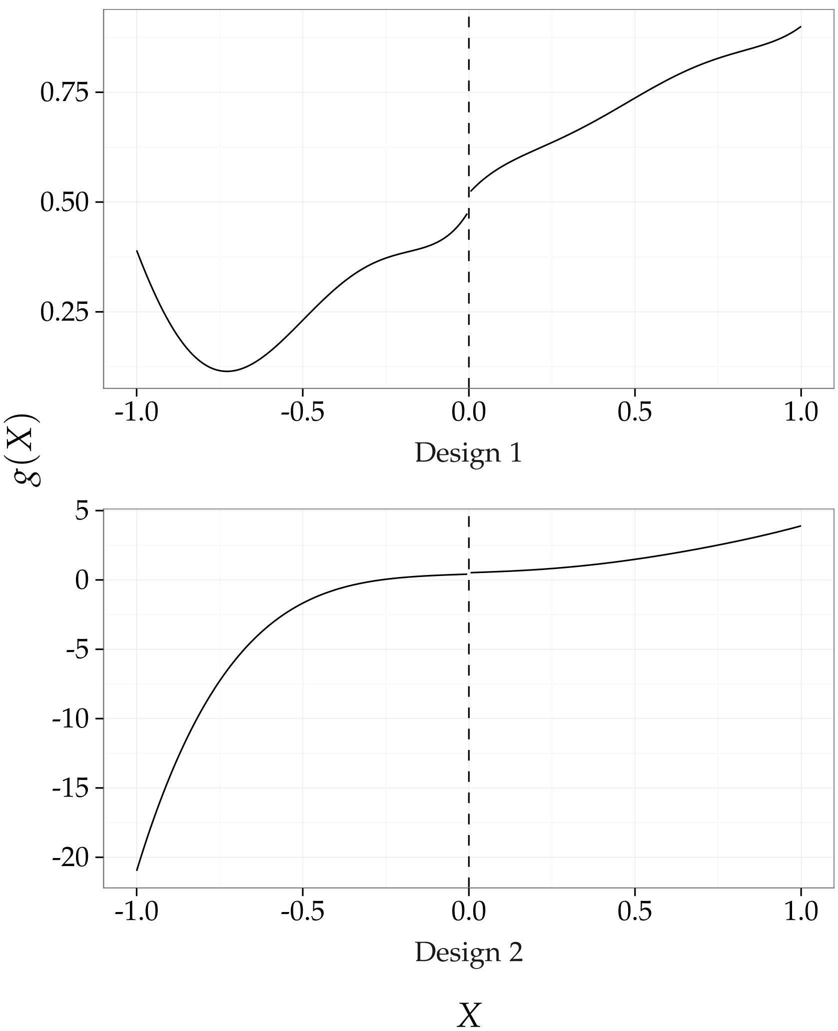


Figure S2: Monte Carlo study of regression discontinuity. Regression function $g(X)$ for designs 1 and 2.

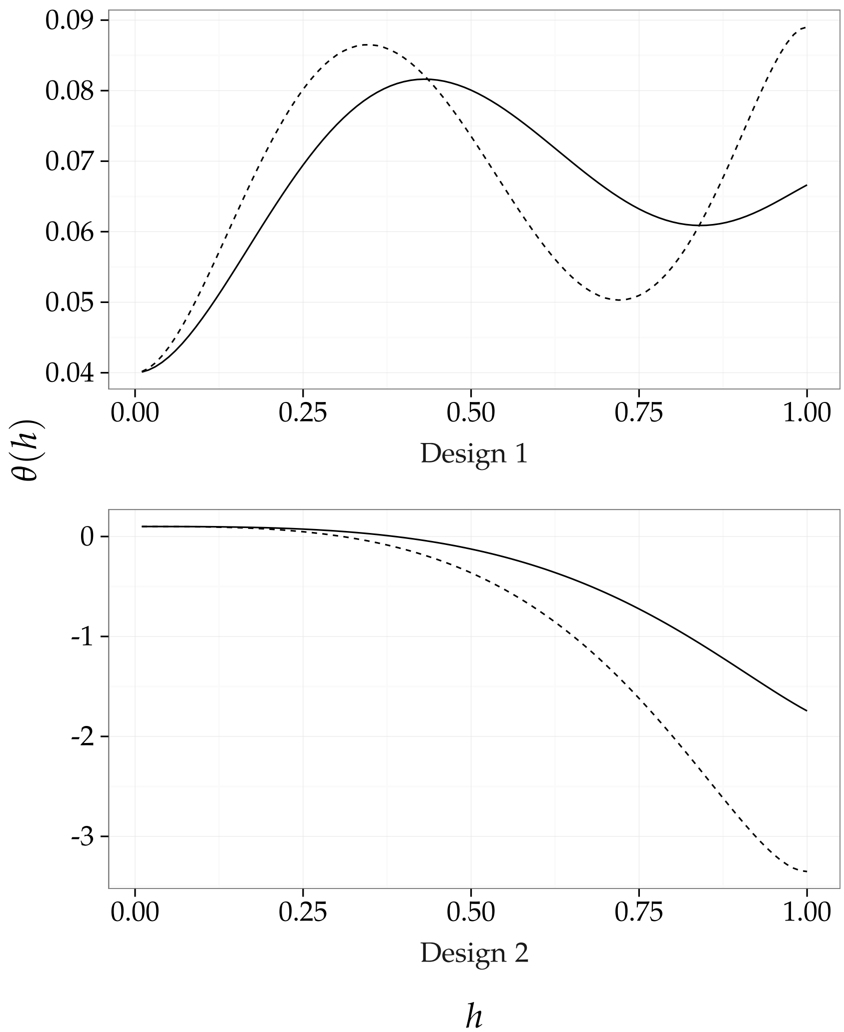


Figure S3: Monte Carlo study of regression discontinuity. Function $\theta(h)$ for local linear regression for designs 1 and 2. Solid lines correspond to the triangular kernel, dotted lines to the uniform kernel.

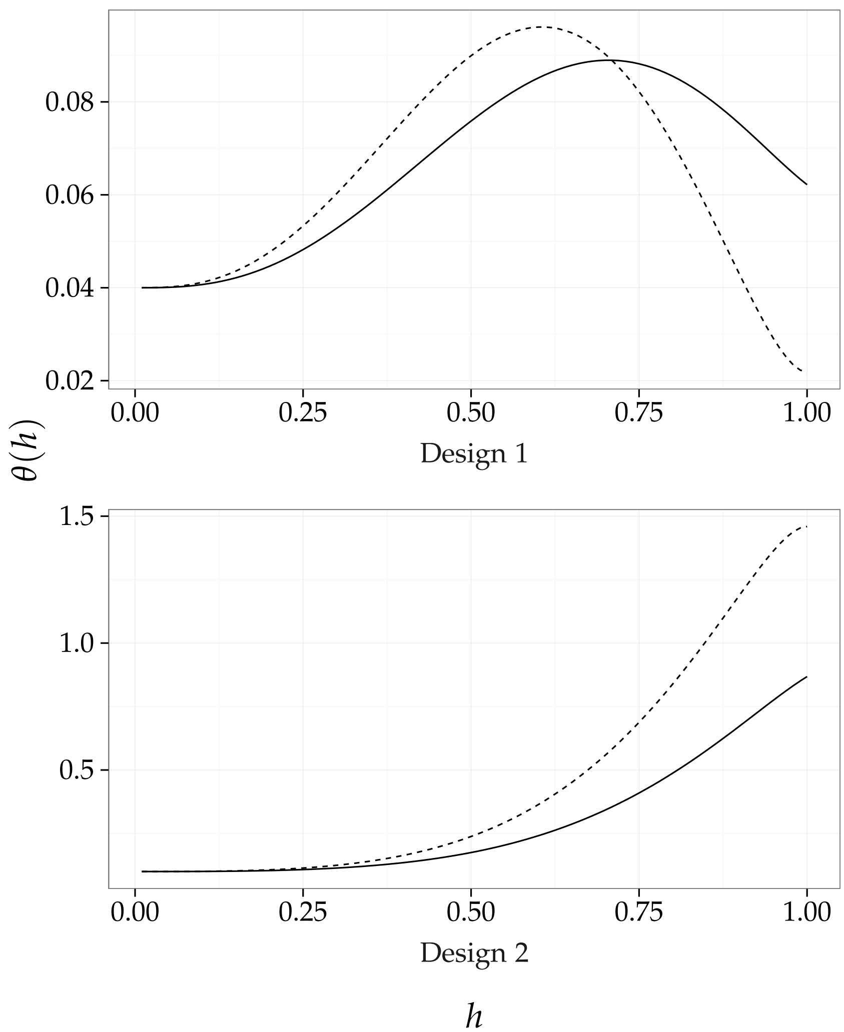


Figure S4: Monte Carlo study of regression discontinuity. Function $\theta(h)$ for local quadratic regression for designs 1 and 2. Solid lines correspond to the triangular kernel, dotted lines to the uniform kernel.