Econometrica Supplementary Material

SUPPLEMENT TO “INTERSECTION BOUNDS: ESTIMATION AND INFERENCE”

(*Econometrica*, Vol. 81, No. 2, March 2013, 667–737)

BY VICTOR CHERNOZHUKOV, SOKBAE LEE, AND ADAM M. ROSEN

This supplement provides appendices not included in the main text. Appendix E provides proofs omitted from the main text. Appendices F–H concern kernel-type estimators, providing primitive conditions for their application to conditional moment inequalities, strong approximation results, and proofs. Appendix I provides additional details on the use of primitive conditions to verify an asymptotic linear expansion needed for strong approximation of series estimators and Appendix J gives some detailed arguments for local polynomial estimation of conditional moment inequalities. Appendix K provides local asymptotic power analysis that supports the findings of our Monte Carlo experiments. Appendix L provides further Monte Carlo evidence.

APPENDIX E: PROOFS OMITTED FROM THE MAIN TEXT

E.1. *Proof of Lemma 2—Estimation of V_n*

There is a single proof for both analytical and simulation methods, but it is convenient for clarity to split the first step of the proof into separate cases. There are four steps in total.

Step 1a—Bounds on $k_{n,\mathcal{V}}(\gamma_n)$ in the Analytical Case. We have that for some constant $\eta > 0$,

$$k_{n,\mathcal{V}}(\gamma_n) := \left(\bar{a}_n + \frac{c(\gamma_n)}{\bar{a}_n}\right),$$

$$\kappa_n := \kappa_n(\gamma_n') := Q_{\gamma_n'}\Big(\sup_{v\in\mathcal{V}} Z_n^*(v)\Big), \quad \bar{\kappa}_n := 7\left(\bar{a}_n + \frac{\eta\ell\ell_n}{\bar{a}_n}\right).$$

The claim of this step is that given the sequence γ_n, we have, for all large n,

$$k_{n,\mathcal{V}}(\gamma_n) \geq \kappa_n(\gamma_n), \tag{E.1}$$

$$6k_{n,\mathcal{V}}(\gamma_n) < \bar{\kappa}_n. \tag{E.2}$$

Inequality (E.2) follows from (B.2) in Step 2 of the proof of Lemma 1 (with γ_n in place of γ_n'); (E.1) follows immediately from Condition C.3.

Step 1b—Bounds on $k_{n,\mathcal{V}}(\gamma_n)$ in the Simulation Case. We have

$$k_{n,\mathcal{V}}(\gamma_n) := Q_{\gamma_n}\Big(\sup_{v\in\mathcal{V}} Z_n^\star(v)\big|\mathcal{D}_n\Big),$$

$$\kappa_n = \kappa_n(\gamma_n') := Q_{\gamma_n'}\Big(\sup_{v\in\mathcal{V}} \bar{Z}_n^*(v)\Big), \quad \bar{\kappa}_n := 7\left(\bar{a}_n + \frac{\eta\ell\ell_n}{\bar{a}_n}\right).$$

 DOI: 10.3982/ECTA8718

The claim of this step is that given γ_n, there is $\gamma_n \geq \gamma_n' = \gamma_n - o(1)$ such that, wp $\to 1$,

$$k_{n,\mathcal{V}}(\gamma_n) \geq \kappa_n(\gamma_n'), \tag{E.3}$$

$$6k_{n,\mathcal{V}}(\gamma_n) < \bar{\kappa}_n. \tag{E.4}$$

To show inequality (E.3), note that by Condition C.2 and Lemma 11, wp $\to 1$,

$$\kappa_{n,\mathcal{V}}\big(\gamma_n + o(1/\ell_n)\big) + o(\delta_n) \geq k_{n,\mathcal{V}}(\gamma_n) \geq \kappa_{n,\mathcal{V}}\big(\gamma_n - o(1/\ell_n)\big) - o(\delta_n). \tag{E.5}$$

Hence (E.3) follows from

$$\begin{aligned}
&\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} \bar{Z}_n^*(v) \leq x\Big)\Big|_{x=k_{n,\mathcal{V}}(\gamma_n)} \\
&\quad \geq_{(1)} \mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} \bar{Z}_n^*(v) \leq \kappa_{n,\mathcal{V}}\big(\gamma_n - o(1/\ell_n)\big) - o(\delta_n)\Big) - o(1) \quad \text{wp} \to 1 \\
&\quad \geq_{(2)} \mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} \bar{Z}_n^*(v) \leq \kappa_{n,\mathcal{V}}\big(\gamma_n - o(1/\ell_n)\big)\Big) - o(1) \\
&\quad = \gamma_n - o(1/\ell_n) - o(1) =: \gamma_n', \quad (\gamma_n \geq \gamma_n')
\end{aligned}$$

where inequality (1) holds by (E.5) and inequality (2) holds by anti-concentration Corollary 2.

To show inequality (E.4), note that by Condition C.3, we have

$$\begin{aligned}
\kappa_{n,\mathcal{V}}\big(\gamma_n + o(1/\ell_n)\big) + o(\delta_n) &\leq \bar{a}_n + \frac{c(\gamma_n + o(1/\ell_n))}{\bar{a}_n} + o(\delta_n) \\
&\leq \bar{a}_n + \frac{\eta\ell\ell_n + \eta\log 10}{\bar{a}_n} + o(\delta_n),
\end{aligned}$$

where the last inequality relies on

$$c\big(\gamma_n + o(1/\ell_n)\big) \leq -\eta\log\big(1 - \gamma_n - o(1/\ell_n)\big) \leq \eta o(\ell\ell_n) + \eta\log 10$$

holding for large n by Condition C.3. From this we deduce (E.4).

Step 2—Lower Containment. We have that for all $v \in V_n$,

$$\begin{aligned}
A_n(v) &:= \hat{\theta}_n(v) - \inf_{v\in\mathcal{V}}\big(\hat{\theta}_n(v) + k_{n,\mathcal{V}}(\gamma_n)s_n(v)\big) \\
&\leq -Z_n(v)\sigma_n(v) + \kappa_n\sigma_n(v) \\
&\quad + \sup_{v\in\mathcal{V}}\big\{\theta_{n0} - \hat{\theta}_n(v) - k_{n,\mathcal{V}}(\gamma_n)s_n(v)\big\} := B_n(v),
\end{aligned}$$

since $\theta_n(v) \leq \theta_{n0} + \kappa_n \sigma_n(v)\ \forall v \in V_n$ and $\hat{\theta}_n(v) - \theta_n(v) = -Z_n(v)\sigma_n(v)$. Therefore,

$$\begin{aligned}
&\mathrm{P}_n\{V_n \subseteq \hat{V}_n\} \\
&\quad = \mathrm{P}_n\big\{A_n(v) \leq 2k_{n,\mathcal{V}}(\gamma_n)s_n(v), \forall v \in V_n\big\} \\
&\quad \geq \mathrm{P}_n\big\{B_n(v) \leq 2k_{n,\mathcal{V}}(\gamma_n)s_n(v), \forall v \in V_n\big\} \\
&\quad \geq \mathrm{P}_n\big\{-Z_n(v)\sigma_n(v) \leq 2k_{n,\mathcal{V}}(\gamma_n)s_n(v) - \kappa_n\sigma_n(v), \forall v \in V_n\big\} \\
&\qquad - \mathrm{P}_n\left\{\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \geq k_{n,\mathcal{V}}(\gamma_n)\right\} \\
&\quad := a - b = \gamma_n' - o(1) = 1 - o(1),
\end{aligned}$$

where $b = o(1)$ follows similarly to the proof of Theorem 1 (analytical case) and Theorem 2 (simulation case), using that $k_{n,\mathcal{V}}(\gamma_n) \geq k_{n,V_n}(\gamma_n)$ for sufficiently large n, and $a = 1 - o(1)$ follows from the argument

$$\begin{aligned}
a &\geq_{(1)} \mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} -Z_n(v) \leq 2k_{n,\mathcal{V}}(\gamma_n)\big[1 - o_{\mathrm{P}_n}\big(\delta_n/(\bar{a}_n + \ell\ell_n)\big)\big] - \kappa_n\Big) \\
&\geq_{(2)} \mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} -Z_n^*(v) \leq 2k_{n,\mathcal{V}}(\gamma_n) - \kappa_n - o_{\mathrm{P}_n}(\delta_n)\Big) - o(1) \\
&\geq_{(3)} \mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} -Z_n^*(v) \leq \kappa_n - o_{\mathrm{P}_n}(\delta_n)\Big) - o(1) \\
&\geq_{(4)} \gamma_n' - o(1) = 1 - o(1),
\end{aligned}$$

where terms $o(\delta_n)$ are different in different places; where inequality (1) follows by Condition C.4, inequality (2) is by Condition C.2 and by Step 1, namely by $k_{n,\mathcal{V}}(\gamma_n') \leq \bar{\kappa}_n \lesssim \bar{a}_n + \ell\ell_n$ wp $\to 1$, inequality (3) follows by Step 1, and inequality (4) follows by the anti-concentration Corollary 2 and definition of κ_n.

Step 3—Upper Containment. We have that for all $v \notin \overline{V}_n$,

$$\begin{aligned}
A_n(v) &:= \hat{\theta}_n(v) - \theta_{n0} - \inf_{v\in\mathcal{V}}\big(\hat{\theta}_n(v) - \theta_{n0} + k_{n,\mathcal{V}}(\gamma_n)s_n(v)\big) \\
&> -Z_n(v)\sigma_n(v) + \bar{\kappa}_n\bar{\sigma}_n \\
&\quad + \sup_{v\in\mathcal{V}}\big\{\theta_{n0} - \hat{\theta}_n(v) - k_{n,\mathcal{V}}(\gamma_n)s_n(v)\big\} := C_n(v),
\end{aligned}$$

since $\theta_n(v) > \theta_{n0} + \bar{\kappa}_n\bar{\sigma}_n, \forall v \notin \bar{V}_n$, and $\hat{\theta}_n(v) - \theta_n(v) = -Z_n(v)\sigma_n(v)$. Hence

$$\begin{aligned}
&\mathrm{P}_n(\hat{V}_n \not\subseteq \bar{V}_n) \\
&\quad = \mathrm{P}_n\big\{A_n(v) \leq 2k_{n,\mathcal{V}}(\gamma_n)s_n(v), \exists v \notin \bar{V}_n\big\}
\end{aligned}$$

$$
\begin{aligned}
&\le \mathrm{P}_n\{C_n(v) < 2k_{n,\mathcal{V}}(\gamma_n)s_n(v), \exists v \notin \bar{V}_n\} \\
&\le \mathrm{P}_n\Big\{-\sup_{v\in\mathcal{V}} 2|Z_n(v)|\bar{\sigma}_n < 3k_{n,\mathcal{V}}(\gamma_n)\bar{s}_n - \bar{\kappa}_n\bar{\sigma}_n\Big\} := a = o(1),
\end{aligned}
$$

where we used elementary inequalities to arrive at the last conclusion. Then $a = o(1)$, since

$$
\begin{aligned}
a &\le_{(1)} \mathrm{P}_n\big(-2|Z_n(v)| < 3k_{n,\mathcal{V}}(\gamma_n)\big[1 + o_{\mathrm{P}_n}\big(\delta_n/(\bar{a}_n + \ell\ell_n)\big)\big] - \bar{\kappa}_n, \\
&\qquad \exists v \in \mathcal{V}\big) \\
&\le_{(2)} \mathrm{P}_n\big(-|Z_n^*(v)| < \big(3k_{n,\mathcal{V}}(\gamma_n) - \bar{\kappa}_n\big)/2 + o(\delta_n), \exists v \in \mathcal{V}\big) + o(1) \\
&\le_{(3)} \mathrm{P}_n\big(-|Z_n^*(v)| < -k_{n,\mathcal{V}}(\gamma_n) + o(\delta_n), \exists v \in \mathcal{V}\big) + o(1) \\
&\le_{(4)} 2\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} Z_n^*(v) > k_{n,\mathcal{V}}(\gamma_n) - o(\delta_n)\Big) + o(1) \\
&\le_{(5)} 2\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}} Z_n^*(v) > \kappa_n - o(\delta_n)\Big) + o(1) \le_{(6)} 2\big(1 - \gamma_n'\big) + o(1),
\end{aligned}
$$

where inequality (1) follows by Condition C.4, inequality (2) follows by Condition C.2 and Step 1, namely by $k_{n,\mathcal{V}}(\gamma_n') \le \bar{\kappa}_n \lesssim \bar{a}_n + \ell\ell_n$ wp $\to 1$, inequality (3) follows by Step 1 and the union bound, inequality (4) holds by the union bound and symmetry, inequality (5) holds by Step 1, and inequality (6) holds by the definition of κ_n and the anti-concentration Corollary 2.

Step 4—Rate. We have that wp $\to 1$,

$$
\begin{aligned}
d_H(\hat{V}_n, V_0) &\le_{(1)} d_H(\hat{V}_n, V_n) + d_H(V_n, V_0) \le_{(2)} 2d_H(\overline{V}_n, V_0) \\
&\le_{(3)} 2(\bar{\sigma}_n\bar{\kappa}_n)^{1/\rho_n}/c_n,
\end{aligned}
$$

where inequality (1) holds by the triangle inequality, inequality (2) follows by the containment $V_0 \subseteq V_n \subseteq \hat{V}_n \subseteq \bar{V}_n$ holding wp $\to 1$, and inequality (3) follows from $\bar{\kappa}_n\bar{\sigma}_n \to 0$ holding by assumption and from the following relation that holds by Condition V:

$$
\begin{aligned}
d_H(\bar{V}_n, V_0) &= \sup_{v\in\bar{V}_n} d(v, V_0) \le \sup\{d(v, V_0) : \theta_n(v) - \theta_{n0} \le \bar{\kappa}_n\bar{\sigma}_n\} \\
&\le \sup\{d(v, V_0) : \big(c_n d(v, V_0)\big)^{\rho_n} \wedge \delta \le \bar{\kappa}_n\bar{\sigma}_n\} \\
&\le \sup\{t : (c_n t)^{\rho_n} \wedge \delta \le \bar{\kappa}_n\bar{\sigma}_n\} \\
&\le c_n^{-1}(\bar{\kappa}_n\bar{\sigma}_n)^{1/\rho_n} \quad \text{for all } 0 \le \bar{\kappa}_n\bar{\sigma}_n \le \delta. \qquad \textit{Q.E.D.}
\end{aligned}
$$

E.2. *Proof of Lemma 4*

Step 1—Verification of Condition C.1. This condition holds by inspection in view of continuity of $v \mapsto p_n(v, \beta_n)$ and $v \mapsto p_n(v, \hat{\beta})$ implied by Condition P(ii) and by Ω_n and $\hat{\Omega}_n$ being positive definite.

Step 2—Verification of Condition C.2.

(a) By Condition P, uniformly in $v \in \mathcal{V}$, for $\beta_n^*(v)$ denoting an intermediate value between β_n and $\hat{\beta}_n$,

$$\begin{aligned} Z_n(v) &= \frac{p_n(v, \beta_n^*(v))'}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|}\sqrt{n}(\hat{\beta}_n - \beta_n) \\ &= \frac{p_n(v, \beta_n)'}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|}\sqrt{n}(\hat{\beta}_n - \beta_n) \\ &\quad + \frac{L_n\sqrt{n}\|\hat{\beta}_n - \beta_n\|^2}{\min_{v\in\mathcal{V}}\|p_n(v, \beta_n)\|}\frac{1}{\lambda_{\min}(\Omega_n^{1/2})} \\ &= \frac{p_n(v, \beta_n)'\Omega_n^{1/2}}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|}\mathcal{N}_k + o_{\mathrm{P}_n}(\delta_n') + O_{\mathrm{P}_n}(n^{-1/2}). \end{aligned}$$

(b) First note, using the inequality

$$\left\|\frac{a}{\|a\|} - \frac{b}{\|b\|}\right\| \le \left(2\frac{\|a-b\|}{\|a\|}\right) \wedge \left(2\frac{\|a-b\|}{\|b\|}\right), \tag{E.6}$$

that we have

$$\begin{aligned} M_n &= \left\| \frac{p_n(v, \beta_n)'\Omega_n^{1/2}}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|} - \frac{p_n(v, \hat{\beta}_n)'\hat{\Omega}_n^{1/2}}{\|p_n(v, \hat{\beta}_n)'\hat{\Omega}_n^{1/2}\|} \right\| \\ &\le 2\frac{\|p_n(v, \beta_n)'\Omega_n^{1/2} - p_n(v, \hat{\beta}_n)'\hat{\Omega}_n^{1/2}\|}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|} \\ &\le 2\frac{\|p_n(v, \beta_n)'\Omega_n^{1/2}(I - \Omega_n^{-1/2}\hat{\Omega}_n^{1/2})\|}{\|p_n(v, \beta_n)'\Omega_n^{1/2}\|} \\ &\quad + 2\frac{L_n\|\hat{\beta}_n - \beta_n\|}{\min_{v\in\mathcal{V}}\|p_n(v, \beta_n)\|}\frac{\lambda_{\max}(\hat{\Omega}_n^{1/2})}{\lambda_{\min}(\Omega_n^{1/2})} \\ &\le 2\|\Omega_n^{-1/2}\|\|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\| + O_{\mathrm{P}_n}(n^{-1/2}) \\ &\le O_{\mathrm{P}_n}(n^{-b}) + O_{\mathrm{P}_n}(n^{-1/2}) = O_{\mathrm{P}_n}(n^{-b}), \end{aligned}$$

for some $b > 0$. We have that

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}}\big|Z_n^*(v) - Z_n^\star(v)\big|\big|\mathcal{D}_n\Big) \leq M_n E_{\mathrm{P}_n}\|\mathcal{N}_k\| \lesssim M_n\sqrt{k}.$$

Hence for any $\delta_n'' \propto n^{-b'}$ with a constant $0 < b' < b$, we have by Markov's inequality that

$$\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}}\big|Z_n^*(v) - Z_n^\star(v)\big| > \delta_n\ell_n\big|\mathcal{D}_n\Big) \leq \frac{O_{\mathrm{P}_n}(n^{-b})}{\delta_n''\ell_n} = o_{\mathrm{P}_n}\Big(\frac{1}{\ell_n}\Big).$$

Now select $\delta_n = \delta_n' \vee \delta_n''$.

Step 3—Verification of Condition C.3. We shall employ Lemma 12, which has the required notation in place. We only need to compute an upper bound on the covering numbers $N(\varepsilon, \mathsf{V}, \rho)$ for the process Z_n. We have that

$$\begin{aligned}
&\sigma\big(Z_n^*(v) - Z_n^*(\tilde{v})\big)\\
&\quad\leq \left\|\frac{p_n(v,\beta_n)'\Omega_n^{1/2}}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|} - \frac{p_n(\tilde{v},\beta_n)'\Omega_n^{1/2}}{\|p_n(\tilde{v},\beta_n)'\Omega_n^{1/2}\|}\right\|\\
&\quad\leq 2\left\|\frac{(p_n(v,\beta_n) - p_n(\tilde{v},\beta_n))'\Omega_n^{1/2}}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|}\right\|\\
&\quad\leq 2\frac{L_n}{\min_{v\in\mathcal{V}}\|p_n(v,\beta_n)\|}\frac{\lambda_{\max}(\Omega_n^{1/2})}{\lambda_{\min}(\Omega_n^{1/2})}\|v-\tilde{v}\| \leq CL\|v-\tilde{v}\|,
\end{aligned}$$

where C is some constant that does not depend on n, by the eigenvalues of Ω_n bounded away from zero and from above. Hence by the standard volumetric argument,

$$N(\varepsilon, \mathsf{V}, \rho) \leq \left(\frac{1 + CL\,\mathrm{diam}(\mathsf{V})}{\varepsilon}\right)^d, \quad 0 < \varepsilon < 1,$$

where the diameter of V is measured by the Euclidean metric. Condition C.3 now follows by Lemma 12, with $a_n(\mathsf{V}) = (2\sqrt{\log L_n(\mathsf{V})}) \vee (1 + \sqrt{d})$, and $L_n(\mathsf{V}) = C'(1 + CL\,\mathrm{diam}(\mathsf{V}))^d$, where C' is some positive constant.

Step 4—Verification of Condition C.4. Under Condition P, we have that $1 \leq a_n(\mathsf{V}) \leq \bar{a}_n := a_n(\mathcal{V}) \lesssim 1$, so that Condition C.4(a) follows, since by Condition P,

$$\begin{aligned}
\bar{\sigma}_n &= \sqrt{\max_{v\in\mathcal{V}}\|p_n(v,\beta_n)\Omega_n^{1/2}\|/n}\\
&\leq \sqrt{\max_{v\in\mathcal{V}}\|p_n(v,\beta_n)\|\,\|\Omega_n^{1/2}\|/n} \lesssim \sqrt{1/n}.
\end{aligned}$$

To verify Condition C.4(b), note that uniformly in $v \in \mathcal{V}$,

$$\begin{aligned}
&\left|\frac{\|p_n(v,\beta_n)'\hat{\Omega}_n^{1/2}\|}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|}-1\right| \\
&\quad\le \left|\frac{\|p_n(v,\beta_n)'\hat{\Omega}_n^{1/2}\| - \|p_n(v,\beta_n)'\Omega_n^{1/2}\|}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|}\right| \\
&\quad\le \frac{\|p_n(v,\beta_n)'(\hat{\Omega}_n^{1/2}-\Omega_n^{1/2})\|}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|} \le \frac{\|p_n(v,\beta_n)'\Omega^{1/2}(\Omega_n^{-1/2}\hat{\Omega}_n^{1/2}-I)\|}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|} \\
&\quad\le \|\Omega_n^{-1/2}\hat{\Omega}_n^{1/2}-I\| \le \|\Omega_n^{-1/2}\|\|\hat{\Omega}_n^{1/2}-\Omega_n^{1/2}\| = o_{\mathrm{P}_n}(\delta_n),
\end{aligned}$$

since $\|\hat{\Omega}_n^{1/2}-\Omega_n^{1/2}\| = O_{\mathrm{P}_n}(n^{-b})$ for some $b>0$, and since $\|\Omega_n^{-1/2}\|$ is uniformly bounded, both implied by the assumptions.

Step 5—Verification of Condition S. Under Condition V, for large enough n, since $r_n \lesssim c_n^{-1}(1/\sqrt{n})^{1/\rho_n} = o(1)$, we have that $r_n \le \varphi_n$ for large n for some $\varphi_n = o(1)$. Condition S then follows by noting that for any positive $o(1)$ term, $\sup_{\|v-\tilde{v}\|\le o(1)} |Z_n(v) - Z_n(\tilde{v})| \lesssim o(1)\|\mathcal{N}_k\| = o_{\mathrm{P}_n}(1)$. *Q.E.D.*

APPENDIX F: KERNEL-TYPE ESTIMATION OF THE BOUNDING FUNCTION FROM CONDITIONAL MOMENT INEQUALITIES

In this section, we provide primitive conditions that justify application of kernel-type estimation methods covered in Section 4.3 for models characterized by conditional moment inequalities.

EXAMPLE 7—Bounding Function From Conditional Moment Inequalities: Suppose that we have an i.i.d. sample of (X_i, Z_i), $i = 1, \dots, n$, defined on the probability space $(A, \mathcal{A}, \mathrm{P})$, where we take P fixed in this example. Suppose that support$(Z_i) = \mathcal{Z} \subseteq [0,1]^d$ and

$$\theta_{n0} = \min_{v\in\mathcal{V}} \theta_n(v)$$

for $\theta_n(v) = E_{\mathrm{P}}[m(X_i,\mu,j)|Z_i = z]$, $v = (z,j)$, where $\mathcal{V} \subseteq \mathcal{Z} \times \{1,\dots,J\}$ is the set of interest. Suppose the first J_0 functions correspond to equalities treated as inequalities, so that $m(X_i,\mu,j) = -m(X_i,\mu,j+1)$ for $j \in \mathcal{J}_0 = \{1,3,\dots,J_0-1\}$. Hence $\theta_n(z,j) = -\theta_n(z,j+1)$ for $j \in \mathcal{J}_0$, and we only need to estimate functions $\theta_n(z,j)$ with the index $j \in \mathcal{J} := \mathcal{J}_0 \cup \{J_0+1, J_0+2, \dots, J\}$. Suppose we use the local polynomial approach to approximating and estimating $\theta_n(z,j)$. For $u \equiv (u_1,\dots,u_d)$, a d-dimensional vector of nonnegative integers, let $[u] = u_1 + \cdots + u_d$. Let A_p be the set of all d-dimensional vectors u such that $[u] \le p$ for some integer $p \ge 0$ and let $|A_p|$ denote the number of elements in A_p. For $z \in \mathbb{R}^d$ with $u \in A_p$, let $z^u = \prod_{i=1}^d z_i^{u_i}$. Now define

$$\mathbf{p}(b, z) = \sum_{u \in A_p} b_u z^u, \tag{F.1}$$

where $b = (b_u)_{u \in A_p}$ is a vector of dimension $|A_p|$. For each $v = (z, j)$ and $Y_i(j) := m(X_i, \mu, j)$, define

$$S_n(b) := \sum_{i=1}^n \left[Y_i(j) - \mathbf{p}\left(b, \frac{Z_i - z}{h_n}\right)\right]^2 K_{h_n}(Z_i - z),$$

where $K_h(u) := K(u/h)$, $K(\cdot)$ is a d-dimensional kernel function, and h_n is a sequence of bandwidths. The local polynomial estimator $\hat{\theta}_n(v)$ of the regression function is the first element of $\hat{b}(z, j) := \arg\min_{b \in \mathbb{R}^{|A_p|}} S_n(b)$.

We impose the following conditions:

(i) For each $j \in \mathcal{J}$, $\theta(z, j)$ is $(p+1)$ times continuously differentiable with respect to $z \in \mathcal{Z}$, where $\mathcal{Z}$ is convex;

(ii) the probability density function f of Z_i is bounded above and bounded below from zero with continuous derivatives on $\mathcal{Z}$;

(iii) $Y_i(j) := m(X_i, \mu, j)$, $Y_i := (Y_i(j), j \in \mathcal{J})'$, and $U_i := Y_i - E_{\mathrm{P}}[Y_i | Z_i]$, and U_i is a bounded random vector;

(iv) for each j, the conditional on Z_i density of U_i exists and is uniformly bounded from above and below or, more generally, Condition R stated in Appendix G holds;

(v) $K(\cdot)$ has support on $[-1, 1]^d$, is twice continuously differentiable, $\int uK(u)\,du = 0$, and $\int K(u)\,du = 1$;

(vi) $h_n \to 0$, $nh_n^{d+|\mathcal{J}|+1} \to \infty$, $nh_n^{d+2(p+1)} \to 0$, and $\sqrt{n^{-1}h^{-2d}} \to 0$ at polynomial rates in n.

These conditions are imposed to verify Assumptions A1–A7 in Kong, Linton, and Xia (2010). Details of verification are given in Supplementary Appendix J. Note that $p > |\mathcal{J}|/2 - 1$ is necessary to satisfy bandwidth conditions in (vi). The assumption that U_i is bounded is technical and is made to simplify exposition and proofs.

Let $\delta_n = 1/\log n$. Then it follows from Corollary 1 and Lemmas 8 and 10 of Kong, Linton, and Xia (2010) that

$$\begin{aligned}\hat{\theta}_n(z, j) - \theta(z, j) = {} & \frac{1}{nh_n^d f(z)} \mathbf{e}_1' S_p^{-1} \sum_{i=1}^n (e_j' U_i) K_h(Z_i - z) \mathbf{u}_p\left(\frac{Z_i - z}{h_n}\right) \\ & + B_n(z, j) + R_n(z, j),\end{aligned} \tag{F.2}$$

where $\mathbf{e}_1$ is an $|A_p| \times 1$ vector whose first element is 1 and all others are 0's, S_p is an $|A_p| \times |A_p|$ matrix such that $S_p = \{\int z^u (z^v)' \, du : u \in A_p, v \in A_p\}$, $\mathbf{u}_p(z)$ is an $|A_p| \times 1$ vector such that $\mathbf{u}_p(z) = \{z^u : u \in A_p\}$,

$$B_n(z, j) = O\big(h_n^{p+1}\big) \quad \text{and} \quad R_n(z, j) = o_{\mathrm{P}}\bigg(\frac{\delta_n}{(nh_n^d)^{1/2}}\bigg)$$

uniformly in $(z, j) \in \mathcal{Z} \times \{1, \dots, J\}$. The exact form of $B_n(z, j)$ is given in equation (12) of Kong, Linton, and Xia (2010). The result that $B_n(z, j) = O(h_n^{p+1})$ uniformly in (z, j) follows from the standard argument based on Taylor expansion given in Fan and Gijbels (1996), Kong, Linton, and Xia (2010), or Masry (1996). The condition that $nh_n^{d+2(p+1)} \to 0$ at a polynomial rate in n corresponds to the undersmoothing condition.

Now set $\mathbf{K}(z/h) \equiv \mathbf{e}_1' S_p^{-1} K_h(z) \mathbf{u}_p(z/h)$, which is a kernel of order $(p+1)$ (see Section 3.2.2 of Fan and Gijbels (1996)). Let

$$g_v(U, Z) := \frac{e_j' U}{(h_n^d)^{1/2} f(z)} \mathbf{K}\bigg(\frac{Z - z}{h_n}\bigg).$$

Then it follows from Lemma 15 in Appendix J that uniformly in $v \in \mathcal{V}$,

$$\big(nh_n^d\big)^{1/2}\big(\hat{\theta}_n(z, j) - \theta_n(z, j)\big) = \mathbb{G}_n(g_v) + o_{\mathrm{P}}(\delta_n).$$

Application of Theorems 8 and 9 in Appendix G, based on the Rio–Massart coupling, verifies condition NK(i)(a) and (b). Finally, Condition NK(ii) holds if we take $\hat{f}_n(z)$ to be the standard kernel density estimator with kernel K and let $e_j' \hat{U}_i = Y_i(j) - \hat{\theta}_n(z, j)$.

APPENDIX G: STRONG APPROXIMATION FOR KERNEL-TYPE METHODS

To establish our strong approximation for kernel-type estimators, we use Theorem 1.1 in Rio (1994), stated below, which builds on the earlier results of Massart (1989). After the statement of the Rio–Massart coupling, we provide our strong approximation result, which generalizes the previous results to kernel-type estimators for regression models with multivariate outcomes. We then provide a novel multiplier method to approximate the distribution of such estimators. The proofs for these results are provided in Appendix H.

G.1. *Rio–Massart Coupling*

Consider a sufficiently rich probability space $(A, \mathcal{A}, \mathrm{P})$. Indeed, we can always enrich an original space by taking the product with $[0, 1]$ equipped with the uniform measure over Borel sets of $[0, 1]$. Consider a suitably measurable, namely image admissible Suslin, function class $\mathcal{F}$ containing functions

$f : I^d \to I$ for $I = (-1, 1)$. A function class $\mathcal{F}$ is of uniformly bounded variation of at most $K(\mathcal{F})$ if

$$\mathrm{TV}(\mathcal{F}) := \sup_{f\in\mathcal{F}} \sup_{g\in\mathcal{D}_c(I^d)} \left(\int_{\mathbb{R}^d} f(x)\,\mathrm{div}\,g(x)/\|g\|_\infty\,dx\right) \le K(\mathcal{F}),$$

where $\mathcal{D}_c(I^d)$ is the space of C^∞ functions taking values in $\mathbb{R}^d$ with compact support included in I^d, and where $\mathrm{div}\,g(x)$ is the divergence of $g(x)$. Suppose the function class $\mathcal{F}$ obeys the uniform L_1 covering condition

$$\sup_Q N\big(\epsilon, \mathcal{F}, L_1(Q)\big) \le C(\mathcal{F})\epsilon^{d(\mathcal{F})},$$

where sup is taken over probability measures with finite support and $N(\epsilon, \mathcal{F}, L_1(Q))$ is the covering number under the $L_1(Q)$ norm on $\mathcal{F}$. Let $X_1, \dots, X_n$ be an i.i.d. sample on the probability space $(A, \mathcal{A}, P)$ from density f_X with support on I^d, bounded from above and away from zero. Let P_X be the measure induced by f_X. Then there exists a P_X-Brownian bridge $\mathbb{B}_n$ with almost sure continuous paths with respect to the $L^1(P_X)$ metric such that for any positive $t \ge C\log n$,

$$\mathrm{P}\Big(\sqrt{n}\sup_{f\in\mathcal{F}}\big|\mathbb{G}_n(f) - \mathbb{B}_n(f)\big| \ge C\sqrt{tn^{(d-1)/d}K(\mathcal{F})} + Ct\sqrt{\log n}\Big) \le e^{-t},$$

where constant C depends only on d, $C(\mathcal{F})$, and $d(\mathcal{F})$.

G.2. *Strong Approximation for Kernel-Type Estimators*

We shall use the following technical condition in what follows.

CONDITION R: The random $(J + d)$ vector (U_i, Z_i) obeys $U_i = (U_{i,1}, \dots, U_{i,J}) = \varphi_n(X_{i,1})$, and $Z_i = \tilde{\varphi}_n(X_{2i})$, where $X_i = (X_{1i}', X_{2i}')'$ is a $(d_1 + d)$ vector with $1 \le d_1 \le J$, which has density bounded away from zero by $\underline{f}$ and bounded above by $\bar{f}$ on the support I^{d_1+d}, where $\varphi_n : I^{d_1} \mapsto I^J$ and $\sum_{l=1}^{d_1}\int_{I^{d_1}} |D_{x_{1l}}\varphi_n(x_1)|\,dx_1 \le B$, where $D_{x_{1l}}\varphi_n(x_1)$ denotes the weak derivative with respect to the lth component of x_1, and $\tilde{\varphi}_n : I^d \mapsto I^d$ is continuously differentiable such that $\max_{k\le d}\sup_{x_2} |\partial\tilde{\varphi}_n(x_2)/\partial x_{2k}| \le B$ and $|\det \partial\tilde{\varphi}_n(x_2)/\partial x_2| \ge c > 0$, where $\partial\tilde{\varphi}_n(x_2)/\partial x_{2k}$ denotes the partial derivative with respect to the kth component of x_2. The constants $J, B, \underline{f}, \bar{f}$, and c, and vector dimensions do not depend on n. ($|\cdot|$ denotes the ℓ_1 norm.)

A simple example of (U_i, Z_i) satisfying this condition is given in Corollary 3 below.

THEOREM 8—Strong Approximation for Kernel-Type Estimators: *Consider a suitably enriched probability space* $(A, \mathcal{A}, \mathrm{P}_n)$ *for each* n. *Let* $n \to \infty$. *Assume the following conditions hold for each* n: (a) *There are* n *i.i.d.* $(J+d)$*-dimensional random vectors of the form* (U_i, Z_i) *that obey Condition* R, *and the density* f_n *of* Z *is bounded from above and away from zero on the set* $\mathcal{Z}$, *uniformly in* n. (b) *Let* $v = (z, j)$ *and* $\mathcal{V} = \mathcal{Z} \times \{1, \ldots, J\}$, *where* $\mathcal{Z} \subseteq I^d$. *The kernel estimator* $v \mapsto \hat{\theta}_n(v)$ *of some target function* $v \mapsto \theta_n(v)$ *has an asymptotic linear expansion uniformly in* $v \in \mathcal{V}$,

$$\left(nh_n^d\right)^{1/2}\left(\hat{\theta}_n(v) - \theta_n(v)\right) = \mathbb{G}_n(g_v) + o_{P_n}(\delta_n),$$
$$g_v(U_i, Z_i) := \frac{1}{(h_n^d)^{1/2} f_n(z)} e_j' U_i \mathbf{K}\left(\frac{z - Z_i}{h_n}\right),$$

where $e_j' U_i \equiv U_{ij}$, $\mathbf{K}$ *is a twice continuously differentiable product kernel function with support on* I^d, $\int \mathbf{K}(u)\,du = 1$, *and* h_n *is a sequence of bandwidths that converges to zero*. (c) *For a given* $\delta_n \searrow 0$, *the bandwidth sequence obeys* $(n^{-1/(d+d_1)} h_n^{-1} \log n)^{1/2} + (nh_n^d)^{-1/2} \log^{3/2} n = o(\delta_n)$. *Then there exists a sequence of centered* P_n*-Gaussian bridges* $\mathbb{B}_n$ *such that*

$$\sup_{v \in \mathcal{V}} \left| \left(nh_n^d\right)^{1/2}\left(\hat{\theta}_n(v) - \theta_n(v)\right) - \mathbb{B}_n(g_v)\right| = o_{\mathrm{P}_n}(\delta_n).$$

Moreover, the paths of $v \mapsto \mathbb{B}_n(g_v)$ *can be chosen to be continuous almost surely* (*a.s.*).

REMARK 8: Conditions (a) and (b) cover standard conditions in the literature, imposing a uniform Bahadur expansion for kernel-type estimators, which have been shown in Masry (1996) and Kong, Linton, and Xia (2010) for kernel mean regression estimators and also local polynomial estimators under fairly general conditions. Implicit in the expansion above is that the asymptotic bias is negligible, which can be achieved by undersmoothing, that is, choosing the bandwidth to be smaller than the rate-optimal bandwidth.

COROLLARY 3—A Simple Leading Case for Moment Inequalities Application: *Suppose that* (U_i, Z_i) *has bounded support, which we then take to be a subset of* I^{J+d} *without loss of generality. Suppose that* $U_i = (U_{ij}, j = 1, \ldots, J)$, *where for the first* $J_0/2$ *pairs of terms, we have* $U_{ij} = -U_{ij+1}$, $j = 1, 3, \ldots, J_0 - 1$. *Let* $\mathcal{J} = \{1, 3, \ldots, J_0 - 1, J_0 + 1, J_0 + 2, \ldots\}$. *Suppose that* $(U_{ij}, Z_i, j \in \mathcal{J})$ *have joint density bounded from above and below by some constants* $\bar{f}$ *and* $\underline{f}$. *Suppose these constants and* d, J, *and* $d_1 = |\mathcal{J}|$ *do not depend on* n. *Then Condition* R *holds and the conclusions of Theorem* 8 *then hold under the additional conditions imposed in the theorem*.

Note that Condition R allows for much more general error terms and regressors. For example, it allows error terms U_i not to have a density at all, and Z_i only to have density bounded from above.

The next theorem shows that the Brownian bridge $\mathbb{B}_n(g_v)$ can be approximately simulated via the Gaussian multiplier method. That is, consider the symmetrized process

$$\mathbb{G}_n^o(g_v) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i g_v(U_i, Z_i) = \mathbb{G}_n(\xi g_v), \tag{G.1}$$

where $\xi_1, \dots, \xi_n$ are i.i.d. $N(0,1)$, independent of the data $\mathcal{D}_n$ and of $\{(U_i, Z_i)\}_{i=1}^n$, which are i.i.d. copies of (U, Z). Conditional on the data, this is a Gaussian process with a covariance function that is a consistent estimate of the covariance function of $v \mapsto \mathbb{B}_n(g_v)$. Theorem 9 below shows that the uniform distance between a copy of $\mathbb{B}_n(g_v)$ and $\mathbb{G}_n^o(g_v)$ is small with an explicit probability bound. Note that if the function class $\{g_v, v \in \mathcal{V}\}$ were Donsker, then such a result would follow from the multiplier functional central limit theorem. In our case, this function class is not Donsker, so we require a different argument.

For the following theorem, consider now a sufficiently rich probability space $(A, \mathcal{A}, \mathrm{P}_n)$. Note that we can always enrich the original space if needed by taking the product with $[0, 1]$ equipped with the uniform measure over Borel sets of $[0, 1]$.

THEOREM 9—Multiplier Method for Kernels: *Let $v = (z, j)$ and $\mathcal{V} \subseteq \mathcal{Z} \times \{1, \dots, J\}$, where $\mathcal{Z}$ is a compact convex set that does not depend on n. The estimator $v \mapsto \hat{\theta}_n(v)$ and the function $v \mapsto \theta_n(v)$ are continuous in v. In what follows, let e_j denote the J vector with jth element 1 and all other elements 0. Suppose that (U, Z) is a $(J + d)$-dimensional random vector, where U is a generalized residual such that $E[U|Z] = 0$ a.s. and Z is a covariate; the density f_n of Z is continuous and bounded away from zero and from above on $\mathcal{Z}$, uniformly in n; and the support of U is bounded uniformly in n. $\mathbf{K}$ is a twice continuously differentiable, possibly higher order, product kernel function with support on $[-1, 1]^d$, $\int \mathbf{K}(u)\, du = 1$, and h_n is a sequence of bandwidths such that $h_n \to 0$ and $nh^d \to \infty$ such that $\sqrt{n^{-1}h^{-2d}} = o((\delta_n/[\ell_n\sqrt{\log n}])^{d+1})$. Let $\{(U_i, Z_i)\}_{i=1}^n$ be i.i.d. copies of (U, Z), where $\{Z_i\}_{i=1}^n$ are a part of the data $\mathcal{D}_n$, and $\{U_i\}$ are a measurable transformation of data. Let $\mathbb{B}_n$ denote the P_n-Brownian bridge and let*

$$g_v(U, Z) := \frac{e_j' U}{(h_n^d)^{1/2} f_n(z)} \mathbf{K}\left(\frac{z - Z}{h_n}\right).$$

Then there exists an independent from data $\mathcal{D}_n$, identically distributed copy $v \mapsto \bar{\mathbb{B}}_n(g_v)$ of the process $v \mapsto \mathbb{B}_n(g_v)$, such that for some $o(\delta_n)$ and $o(1/\ell_n)$ sequences,

$$\mathrm{P}_n\Big(\sup_{v \in \mathcal{V}} \big|\mathbb{G}_n^o(g_v) - \bar{\mathbb{B}}_n(g_v)\big| > o(\delta_n) \big| \mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1/\ell_n). \tag{G.2}$$

APPENDIX H: PROOFS FOR NONPARAMETRIC ESTIMATION OF $\theta(v)$ VIA KERNEL-TYPE METHODS

H.1. *Proof of Lemma 7*

There are six steps, with the first four verifying Conditions C.1–C.4 and the last two providing auxiliary calculations. Let $U_{ij} \equiv e_j' U_i$.

Step 1—Verification of Condition C.1. Condition C.1 holds by inspection, in view of the continuity of $v \mapsto \hat{\theta}_n(v)$, $v \mapsto \theta_n(v)$, $v \mapsto \sigma_n(v)$, and $v \mapsto s_n(v)$.

Step 2—Verification of Condition C.3. Note that

$$\frac{g_v(U_i, Z_i)}{\sigma_n(v)\sqrt{nh_n^d}} = \frac{\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}}{\Big\|\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}}.$$

We shall employ Lemma 12, which has the required notation in place. We only need to compute an upper bound on the covering numbers $N(\varepsilon, \mathsf{V}, \rho)$ of V under the metric $\rho(v, \bar{v}) = \sigma(Z_n^*(v) - Z_n^*(\bar{v}))$. We have that for $v = (z, j)$ and $\bar{v} = (\bar{z}, j)$,

$$\sigma\big(Z_n^*(v) - Z_n^*(\bar{v})\big) \leq Y_n\|z - \bar{z}\|,$$

$$Y_n := \sup_{v \in \mathcal{V}, 1 \leq k \leq d} \left\| \nabla_{z_k} \frac{\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}}{\Big\|\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}} \right\|_{\mathrm{P}_n,2}.$$

We have that

$$Y_n \leq \sup_{v \in \mathcal{V}, 1 \leq k \leq d} \frac{\Big\|\nabla_{z_k}\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}}{\Big\|\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}} + \frac{\Big|\nabla_{z_k}\Big\|\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}\Big|}{\Big\|\frac{1}{h_n^{d/2}}\mathbf{K}\Big(\frac{z - Z_i}{h_n}\Big)U_{ij}\Big\|_{\mathrm{P}_n,2}},$$

which is bounded by $C(1+h_n^{-1})$ for large n by Step 6. Since J is finite, it follows that for all large $n > n_0$, for all nonempty subsets of $\mathsf{V} \subseteq \mathcal{V}$,

$$N(\varepsilon, \mathsf{V}, \rho) \leq \left(\frac{J^{1/d}(1 + C(1+h_n^{-1}) \operatorname{diam}(\mathsf{V}))}{\varepsilon} \right)^d, \quad 0 < \varepsilon < 1.$$

Condition C.3 now follows for all $n > n_0$ by Lemma 12, with

$$a_n(\mathsf{V}) = \big(2\sqrt{\log L_n(\mathsf{V})}\big) \vee (1 + \sqrt{d}),$$
$$L_n(\mathsf{V}) = C'\big(1 + C\big(1 + h_n^{-1}\big) \operatorname{diam}(\mathsf{V})\big)^d,$$

where C' is some positive constant.

Step 3—Verification of Condition C.4. Under Condition NK, we have that

$$a_n(\mathsf{V}) \leq \bar{a}_n := a_n(\mathcal{V}) \lesssim \sqrt{\log \ell_n + \log n} \lesssim \sqrt{\log n},$$

so that Condition C.4(a) follows if $\sqrt{\log n/(nh_n^d)} \to 0$.

To verify Condition C.4(b), note that

$$\left| \frac{s_n(v)}{\sigma_n(v)} - 1 \right| = \left| \underbrace{\left(\frac{f_n(z)}{\hat{f}_n(z)} \right)}_{a} \left(\underbrace{\frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \hat{U}_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}}}_{b/c} \right) - 1 \right|.$$

Since $|a(b/c) - 1| \leq 2|a - 1| + |(b-c)/c|$ when $|(b-c)/c| \leq 1$, the result follows from $|a - 1| = O_{\mathrm{P}_n}(n^{-b}) = o_p(\delta_n/(\bar{a}_n + \ell_n))$ holding by Condition NK(ii) for some $b > 0$ and from

$$\begin{aligned}
\left| \frac{(b-c)}{c} \right| &\leq \max_{1 \leq i \leq n} \|\hat{U}_i - U_i\| \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}} \\
&\quad + \left| \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}} - 1 \right| \\
&\leq O_{\mathrm{P}_n}\big(n^{-b}\big) O_{\mathrm{P}_n}(1) + O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{nh^d}} \right)
\end{aligned}$$

$$= O_{P_n}(n^{-b}) = o_{P_n}\left(\frac{\delta_n}{\bar{a}_n + \ell_n}\right)$$

for some $b > 0$, where we used Condition NK(ii), the results of Step 6, and the condition that $nh_n^d \to \infty$ at a polynomial rate.

Step 4—Verification of Condition C.2. Condition NK(i) and $1 \lesssim E_{P_n}[g_v^2] \lesssim 1$ uniformly in $v \in \mathcal{V}$ holding by Step 6 give

$$\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n(g_v)}{\sqrt{E_{P_n}[g_v^2]}} - \frac{\mathbb{B}_n(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| = O_{P_n}(\delta_n),$$

where $v \mapsto \mathbb{B}_n(g_v)$ is a zero-mean P_n-Brownian bridge, with a.s. continuous sample paths. This and the condition on the remainder term in Condition NK(i) in turn imply Condition C.2(a).

To show Condition C.2(b), we need to show that for any $C > 0$,

$$P_n\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v)}{\sqrt{\mathbb{E}_n[\hat{g}_v^2]}} - \frac{\bar{\mathbb{B}}_n(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| > C\delta_n \middle| \mathcal{D}_n\right) = o_{P_n}(1/\ell_n),$$

where $\bar{\mathbb{B}}_n$ is a copy of $\mathbb{B}_n$, which is independent of the data. First, Condition NK(i), with the fact that $1 \lesssim E_{P_n}[g_v^2] \lesssim 1$ uniformly in $v \in \mathcal{V}$, implies that

$$P_n\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}} - \frac{\bar{\mathbb{B}}_n(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| > C\delta_n \middle| \mathcal{D}_n\right) = o_{P_n}\left(\frac{1}{\ell_n}\right).$$

Therefore, in view of the triangle inequality and the union bound, it remains to show that

$$P_n\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v)}{\sqrt{\mathbb{E}_n[\hat{g}_v^2]}} - \frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| > C\delta_n \middle| \mathcal{D}_n\right) = o_{P_n}\left(\frac{1}{\ell_n}\right). \tag{H.1}$$

We have that

$$\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v)}{\sqrt{\mathbb{E}_n[\hat{g}_v^2]}} - \frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right|$$
$$\le \sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v - g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| + \sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| \sup_{v\in\mathcal{V}}\left|\frac{\sigma_n(v)}{s_n(v)} - 1\right|.$$

We observe that

$$E_{P_n}\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| \sup_{v\in\mathcal{V}}\left|\frac{\sigma_n(v)}{s_n(v)} - 1\right| \middle| \mathcal{D}_n\right)$$
$$= E_{P_n}\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{P_n}[g_v^2]}}\right| \middle| \mathcal{D}_n\right) \sup_{v\in\mathcal{V}}\left|\frac{\sigma_n(v)}{s_n(v)} - 1\right|$$

$$
\begin{aligned}
&= O_{\mathrm{P}_n}\big(\sqrt{\log n} n^{-b}\big) \\
&= O_{\mathrm{P}_n}(\delta_n/\ell_n),
\end{aligned}
$$

where the last equality follows from Steps 5 and 3. Also we note that

$$
\begin{aligned}
&E_{\mathrm{P}_n}\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v - g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}\right| \Big| \mathcal{D}_n\right) \\
&\leq_{(1)} O_{\mathrm{P}_n}(\sqrt{\log n}) \sup_{v\in\mathcal{V}} \frac{\left\| \left(\frac{U_{ij}}{f_n(z)} - \frac{\hat{U}_{ij}}{\hat{f}_n(z)}\right) \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right)\right\|_{\mathbb{P}_n,2}}{\frac{1}{f_n(z)} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij}\right\|_{\mathrm{P}_n,2}} \\
&\lesssim_{(2)} O_{\mathrm{P}_n}(\sqrt{\log n}) \sup_{v\in\mathcal{V}} \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) (1 + |U_{ij}|)\right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij}\right\|_{\mathrm{P}_n,2}} \\
&\quad \times \left(\left|\frac{f_n(z)}{\hat{f}_n(z)} - 1\right| \vee \max_{1\leq i\leq n} \|\hat{U}_i - U_i\|\right) \\
&\leq_{(3)} O_{\mathrm{P}_n}(\sqrt{\log n}) O_{\mathrm{P}_n}(1) O_{\mathrm{P}_n}\big(n^{-b}\big) \\
&= o_{\mathrm{P}_n}\left(\frac{\delta_n}{\ell_n}\right),
\end{aligned}
$$

where inequality (1) follows from Step 5, inequality (2) follows by elementary inequalities, and inequality (3) follows by Step 6 and Condition NK(ii). It follows that (H.1) holds by Markov's inequality.

Step 5—This step shows that

$$
E_{\mathrm{P}_n}\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}\right| \Big| \mathcal{D}_n\right) = O_{\mathrm{P}_n}(\sqrt{\log n}), \tag{H.2}
$$

$$
\begin{aligned}
&E_{\mathrm{P}_n}\left(\sup_{v\in\mathcal{V}}\left|\frac{\mathbb{G}_n^o(\hat{g}_v - g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}\right| \Big| \mathcal{D}_n\right) \\
&\quad \leq O_{\mathrm{P}_n}(\sqrt{\log n}) \sup_{v\in\mathcal{V}} \frac{\left\| \left(\frac{U_{ij}}{f_n(z)} - \frac{\hat{U}_{ij}}{\hat{f}_n(z)}\right) \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right)\right\|_{\mathbb{P}_n,2}}{\frac{1}{f_n(z)} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij}\right\|_{\mathrm{P}_n,2}}.
\end{aligned} \tag{H.3}
$$

To show (H.2), we use Lemma 13 applied to $X_v = \frac{\mathbb{G}_n^o(g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}$ conditional on $\mathcal{D}_n$. First, we compute

$$\sigma(X) = \sup_{v\in\mathcal{V}} \bigl(E_{\mathrm{P}_n}\bigl(X_v^2|\mathcal{D}_n\bigr)\bigr)^{1/2} = \sup_{v\in\mathcal{V}} \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}}$$
$$= 1 + o_{\mathrm{P}_n}(1),$$

where the last equality holds by Step 6. Second, we observe that for $v = (z, j)$ and $\bar{v} = (\bar{z}, j)$,

$$\sigma(X_v - X_{\bar{v}}) \le Y_n \|z - \bar{z}\|,$$
$$Y_n := \sup_{v\in\mathcal{V}, 1\le k\le d} \left\| \nabla_{z_k} \frac{\frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}} \right\|_{\mathbb{P}_n,2}.$$

We have that

$$Y_n \le \sup_{v\in\mathcal{V}, 1\le k\le d} \frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}}$$
$$+ \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}} \cdot \frac{\left| \nabla_{z_k} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2} \right|}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z-Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}},$$

which is bounded with probability converging to 1 by $C(h_n^{-1} + 1)$ for large n by Step 6 and Condition NK(ii). Since J is finite, it follows that for all large $n > n_0$, the covering number for $\mathcal{V}$ under $\rho(v, \bar{v}) = \sigma(X_v - X_{\bar{v}})$ obeys, with probability converging to 1,

$$N(\varepsilon, \mathcal{V}, \rho) \le \left(\frac{J^{1/d}(1 + C(1 + h_n^{-1})\,\mathrm{diam}(\mathcal{V}))}{\varepsilon} \right)^d, \quad 0 < \varepsilon < \sigma(X).$$

Hence $\log N(\varepsilon, \mathcal{V}, \rho) \lesssim \log n + \log(1/\varepsilon)$. Hence by Lemma 13, we have that

$$E_{\mathrm{P}_n}\Bigl(\sup_{v\in\mathcal{V}} |X_v| \big| \mathcal{D}_n\Bigr) \le \sigma(X) + \int_0^{2\sigma(X)} \sqrt{\log(n/\varepsilon)}\, d\varepsilon = O_{\mathrm{P}_n}(\sqrt{\log n}).$$

To show (H.3), we use Lemma 13 applied to $X_v = \frac{\mathbb{G}_n^o(\hat{g}_v - g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}$ conditional on $\mathcal{D}_n$. First, we compute

$$\sigma(X) = \sup_{v \in \mathcal{V}} \big(E_{\mathrm{P}_n}\big(X_v^2 | \mathcal{D}_n\big)\big)^{1/2} = \sup_{v \in \mathcal{V}} \frac{\left\| \left(\frac{U_{ij}}{f_n(z)} - \frac{\hat{U}_{ij}}{\hat{f}_n(z)} \right) \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \right\|_{\mathbb{P}_n,2}}{\frac{1}{f_n(z)} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}}.$$

Second, we observe that for $v = (z, j)$ and $\bar{v} = (\bar{z}, j)$,

$$\sigma(X_v - X_{\bar{v}}) \leq (Y_n + \hat{Y}_n) \|z - \bar{z}\|,$$

where

$$\hat{Y}_n := \sup_{v \in \mathcal{V}, 1 \leq k \leq d} \left\| \nabla_{z_k} \frac{\frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \hat{U}_{ij}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}} \right\|_{\mathbb{P}_n,2}$$

and Y_n is the same as defined above.

We have that

$$\hat{Y}_n \leq \sup_{v \in \mathcal{V}, 1 \leq k \leq d} \frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \hat{U}_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}}$$
$$+ \frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \hat{U}_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}} \cdot \frac{\left| \nabla_{z_k} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathbb{P}_n,2} \right|}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}}.$$

The first term is bounded by

$$\sup_{v \in \mathcal{V}, 1 \leq k \leq d} \frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}}$$
$$+ \max_{1 \leq i \leq n} \|\hat{U}_i - U_i\| \sup_{v \in \mathcal{V}, 1 \leq k \leq d} \frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n} \right) U_{ij} \right\|_{\mathrm{P}_n,2}},$$

which is bounded by $C(1+h_n^{-1})+O_{\mathrm{P}_n}(n^{-b})O_{\mathrm{P}_n}(1)$ for large n by Step 6 and Condition NK(ii). In the second term, the left term of the product is bounded by

$$\sup_{v\in\mathcal{V},1\le k\le d}\frac{\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)U_{ij}\right\|_{\mathbb{P}_n,2}}{\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)U_{ij}\right\|_{\mathrm{P}_n,2}}$$

$$+\max_{1\le i\le n}\|\hat U_i-U_i\|\sup_{v\in\mathcal{V},1\le k\le d}\frac{\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)\right\|_{\mathbb{P}_n,2}}{\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)U_{ij}\right\|_{\mathrm{P}_n,2}},$$

which is bounded by $C(1+o_{\mathrm{P}_n}(1))+O_{\mathrm{P}_n}(n^{-b})O_{\mathrm{P}_n}(1)$ for large n by Step 6 and Condition NK(ii); the right term of the product is bounded by $C(1+h_n^{-1}+o_{\mathrm{P}_n}(1))$ by Step 6. We conclude that $\hat Y_n\le C(1+h_n^{-1})$ for some constant $C>0$ with probability converging to 1.

Since J is finite, it follows that for all large $n>n_0$, the covering number for $\mathcal{V}$ under $\rho(v,\bar v)=\sigma(X_v-X_{\bar v})$ obeys, with probability converging to 1,

$$N(\varepsilon,\mathcal{V},\rho)\le\left(\frac{J^{1/d}(1+C(1+h_n^{-1})\,\mathrm{diam}(\mathcal{V}))}{\varepsilon}\right)^d,\quad 0<\varepsilon<\sigma(X).$$

Hence

$$\log N(\varepsilon,\mathcal{V},\rho)\lesssim\log n+\log(1/\varepsilon).$$

Hence by Lemma 13, we have that

$$\begin{aligned}E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}}|X_v|\big|\mathcal{D}_n\Big)&\lesssim\sigma(X)+\int_0^{2\sigma(X)}\sqrt{\log(n/\varepsilon)}\,d\varepsilon\\&=O_{\mathrm{P}_n}(\sqrt{\log n})\sigma(X).\end{aligned}$$

Step 6. The claims of this step are the relations, uniformly in $v\in\mathcal{V}$, $1\le k\le d$,

$$1\lesssim\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)U_{ij}\right\|_{\mathrm{P}_n,2}\lesssim 1, \tag{H.4}$$

$$1\lesssim\left\|\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)\right\|_{\mathrm{P}_n,2}\lesssim 1, \tag{H.5}$$

$$h_n^{-1}\lesssim\left\|\nabla_{z_k}\frac{1}{h_n^{d/2}}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right)U_{ij}\right\|_{\mathrm{P}_n,2}\lesssim h_n^{-1}, \tag{H.6}$$

$$h_n^{-1} \lesssim \left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) \right\|_{\mathrm{P}_n,2} \lesssim h_n^{-1}, \tag{H.7}$$

$$h_n^{-1} \lesssim \left| \nabla_{z_k} \left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2} \right| \lesssim h_n^{-1} \tag{H.8}$$

and

$$\frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}} = 1 + O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{n h_n^d}}\right), \tag{H.9}$$

$$\frac{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) \right\|_{\mathbb{P}_n,2}}{\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) \right\|_{\mathrm{P}_n,2}} = 1 + O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{n h_n^d}}\right), \tag{H.10}$$

$$\frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathbb{P}_n,2}}{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}} = 1 + O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{n h_n^d}}\right), \tag{H.11}$$

$$\frac{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) \right\|_{\mathbb{P}_n,2}}{\left\| \nabla_{z_k} \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) \right\|_{\mathrm{P}_n,2}} = 1 + O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{n h_n^d}}\right). \tag{H.12}$$

The proofs of (H.4)–(H.8) are all similar to one another, as are those of (H.9)–(H.12), and are standard in the kernel estimator literature. We therefore prove only (H.4) and (H.9) to demonstrate the argument. To establish (H.4), we have

$$\begin{aligned}
&\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}^2 \\
&= h_n^{-d} \int \mathbf{K}^2\big((z - \bar{z})/h_n\big) E\big[U_{ij}^2 | \bar{z}\big] f_n(\bar{z})\, d\bar{z} \\
&\leq_{(1)} h_n^{-d} \int \mathbf{K}^2\big((z - \bar{z})/h_n\big) C\, d\bar{z} \leq_{(2)} \int \mathbf{K}^2(u) C\, du
\end{aligned}$$

for some constant $0 < C < \infty$, where in inequality (1), we use the assumption that $E[U_{ij}^2|z]$ and $f_n(z)$ are bounded uniformly from above, and in inequality

(2), change of variables. On the other hand,

$$\begin{aligned}
&\left\| \frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right) U_{ij} \right\|_{\mathrm{P}_n,2}^2 \\
&\quad = h_n^{-d} \int \mathbf{K}^2\big((z - \bar{z})/h_n\big) E\big[U_{ij}^2 | \bar{z}\big] f_n(\bar{z})\, d\bar{z} \\
&\quad \geq_{(1)} h_n^{-d} \int \mathbf{K}^2\big((z - \bar{z})/h_n\big) C\, d\bar{z} \geq_{(2)} \int \mathbf{K}^2(u) C\, du
\end{aligned}$$

for some constant $0 < C < \infty$, where in inequality (1), we use the assumption that $E[U_{ij}^2|z]$ and $f_n(z)$ are bounded away from zero uniformly in n, and in inequality (2), change of variables.

Moving to (H.9), it suffices to show that, uniformly in $v \in \mathcal{V}$,

$$\begin{aligned}
&\mathbb{E}_n\left(\left(\frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right)\right)^2 U_{ij}^2\right) - E_{\mathrm{P}_n}\left(\left(\frac{1}{h_n^{d/2}} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right)\right)^2 U_{ij}^2\right) \\
&\quad = O_{\mathrm{P}_n}\left(\sqrt{\frac{\log n}{n h_n^d}}\right)
\end{aligned}$$

or, equivalently,

$$\mathbb{E}_n\left(\mathbf{K}^2\left(\frac{z - Z_i}{h_n}\right) U_{ij}^2\right) - E_{\mathrm{P}_n}\left(\mathbf{K}^2\left(\frac{z - Z_i}{h_n}\right) U_{ij}^2\right) = O_{\mathrm{P}_n}\left(\sqrt{\frac{h_n^d \log n}{n}}\right). \tag{H.13}$$

Given the boundedness of U_{ij} imposed by Condition R, this is, in fact, a standard result on local empirical processes, using Pollard's empirical process methods. Specifically, (H.13) follows by the application of Theorem 37 in Chapter II of Pollard (1984). *Q.E.D.*

H.2. *Proof of Lemma 8*

To show claim (i), we need to establish that for

$$\varphi_n = o(1) \cdot \left(\frac{h_n}{\sqrt{\log n}}\right)$$

for any $o(1)$ term, we have that

$$\sup_{\|v - v'\| \leq \varphi_n} \big| Z_n^*(v) - Z_n^*(v') \big| = o_{\mathrm{P}_n}(1).$$

Consider the stochastic process $X = \{Z_n(v), v \in \mathcal{V}\}$. We shall use the standard maximal inequality stated in Lemma 13. From the proof of Lemma 7, we have that for $v=(z,j)$ and $v'=(z',j)$, $\sigma(Z_n^*(v)-Z_n^*(v'))\le C(1+h_n^{-1})\|z-z'\|$, where C is some constant that does not depend on n, and $\log N(\varepsilon,\mathsf{V},\rho)\lesssim \log n+\log(1/\varepsilon)$. Since

$$\|v-v'\|\le\varphi_n \quad\Longrightarrow\quad \sigma\big(Z_n^*(v)-Z_n^*(v')\big)\le C\frac{o(1)}{\sqrt{\log n}},$$

we have

$$\begin{aligned}E\sup_{\|v-v'\|\le\varphi_n}|X_v-X_{v'}| &\lesssim \int_0^{Co(1)/\sqrt{\log n}}\sqrt{\log(n/\varepsilon)}\,d\varepsilon\\ &\lesssim \frac{o(1)}{\sqrt{\log n}}\sqrt{\log n}=o(1).\end{aligned}$$

Hence the conclusion follows from Markov's inequality.

Under Condition V, by Lemma 2,

$$r_n\lesssim\left(\sqrt{\frac{\log n}{nh_n^d}}\log n\right)^{1/\rho_n}c_n^{-1},$$

so $r_n=o(\varphi_n)$ if

$$\left(\sqrt{\frac{\log n}{nh_n^d}}\log n\right)^{1/\rho_n}c_n^{-1}=o\left(\frac{h_n}{\sqrt{\log n}}\right).$$

Thus, Condition S is satisfied. *Q.E.D.*

H.3. *Proof of Theorem 8*

To prove this theorem, we use the Rio–Massart coupling. First we note that

$$\begin{aligned}\mathcal{M}=\big\{&h_n^{d/2}f_n(z)g_v(U_i,Z_i)=e_j'U_i\mathbf{K}\big((z-Z_i)/h_n\big),\\ &z\in\mathcal{Z},j\in\{1,\dots,J\}\big\}\end{aligned}$$

is the product of $\{e_j'U_i, j\in 1,\dots,J\}$ with covering number trivially bounded above by J and $\mathcal{K}:=\{\mathbf{K}((z-Z_i)/h_n), z\in\mathcal{Z}\}$ obeys $\sup_Q N(\epsilon,\mathcal{K},L_1(Q))\lesssim\epsilon^{-\nu}$ for some finite constant ν; see Lemma 4.1 of Rio (1994). Therefore, by Lemma A.1 in Ghosal, Sen, and van der Vaart (2000), we have that

$$\sup_Q N\big(\epsilon,\mathcal{M},L_1(Q)\big)\lesssim J(\epsilon/2)^{-\nu}\lesssim\epsilon^{-\nu}. \tag{H.14}$$

Next we bound, for $\mathbf{K}_l(u) = \partial \mathbf{K}(u)/\partial u_l$,

$$
\begin{aligned}
\mathrm{TV}(\mathcal{M}) &\leq \sup_{f \in \mathcal{M}} \int \left| D_{(x_1', x_2')'} f(x_1, x_2) \right| dx_1\, dx_2 \\
&\leq \sup_{v \in \mathcal{V}} \int_{I^d} \int_{I^{d_1}} \Bigg(\sum_{l=1}^{d_1} \left| e_j' D_{x_{1l}} \varphi_n(x_1) \mathbf{K}\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \right| \\
&\quad + \sum_{l=1}^{d} \Big| e_j' \varphi_n(x_1) \mathbf{K}_l\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \\
&\quad \times h_n^{-1} \partial \tilde{\varphi}(x_2)/\partial x_{2k} \Big| \Bigg) dx_1\, dx_2 \\
&\leq C \max_{1 \leq l \leq d} \sup_{v \in \mathcal{V}} \int_{I^d} \Big(\left| \mathbf{K}\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \right| \\
&\quad + h_n^{-1} \left| \mathbf{K}_l\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \right| B \Big) dx_2 \\
&\leq C h_n^d + C h_n^{-1} h_n^d \leq C h_n^{d-1} =: K(\mathcal{M}),
\end{aligned}
$$

where C is a generic constant, possibly different in different places, and where we rely on

$$
\begin{aligned}
&\int_{I^{d_1}} \left| D_{x_{1l}} \varphi_n(x_1) \right| dx_1 \leq B, \\
&\sup_{x_1} \left| e_j' \varphi_n(x_1) \right| \leq B, \\
&\sup_{x_2} \left| \partial \tilde{\varphi}(x_2)/\partial x_{2k} \right| \leq B
\end{aligned}
$$

as well as on

$$
\begin{aligned}
&\int_{I^d} \left| \mathbf{K}\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \right| dx_2 \leq C h^d, \\
&\int_{I^d} \left| \mathbf{K}_l\big((z - \tilde{\varphi}_n(x_2))/h_n \big) \right| dx_2 \leq C h^d.
\end{aligned}
$$

To see how the latter relationships holds, note that $Y = \tilde{\varphi}_n(v)$ when $v \sim U(I^d)$ has a density bounded uniformly from above: $f_Y(y) \lesssim 1/|\det \partial \tilde{\varphi}_n(v)/\partial v| \lesssim 1/c$. Moreover, the functions $|\mathbf{K}((z - y)/h_n)|$ and $|\mathbf{K}_l((z - y)/h_n)|$ are bounded above by some constant $\bar{K}$ and are nonzero only over a y belonging to a cube centered at z of volume $(2h)^d$. Hence

$$\int_{I^d} \left|\mathbf{K}\big((z-\tilde{\varphi}_n(x_2))/h_n\big)\right| dx_2 \le \int_{I^d} \left|\mathbf{K}\big((z-y)/h_n\big)\right| f_Y(y)\,dy$$
$$\le \bar{K}(2h)^d(1/c) \le Ch^d$$

and similarly for the second term.

By the Rio–Massart coupling we have that for some constant C and $t \ge C\log n$,

$$\mathrm{P}_n\Big(\sqrt{n}\sup_{f\in\mathcal{M}}\left|\mathbb{G}_n(f)-\mathbb{B}_n(f)\right| \ge C\sqrt{tn^{(d+d_1-1)/(d+d_1)}K(\mathcal{M})} + Ct\sqrt{\log n}\Big)$$
$$\le e^{-t},$$

which implies that

$$\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}}\left|\mathbb{G}_n(g_v)-\mathbb{B}_n(g_v)\right|$$
$$\ge n^{-1/2}C\sqrt{tn^{(d+d_1-1)/(d+d_1)}h_n^{d-1}}h_n^{-d/2} + n^{-1/2}h_n^{-d/2}Ct\sqrt{\log n}\Big) \le e^{-t},$$

which upon inserting $t = C\log n$ gives

$$\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}}\left|\mathbb{G}_n(g_v)-\mathbb{B}_n(g_v)\right|$$
$$\ge C\big[n^{-1/2(d+d_1)}\big(h_n^{-1}\log n\big)^{1/2} + \big(nh_n^d\big)^{-1/2}\log^{3/2} n\big]\Big) \lesssim 1/n.$$

This implies the required conclusion. Note that $g_v \mapsto \mathbb{B}_n(g_v)$ is continuous under the $L_1(f_X)$ metric by the Rio–Massart coupling, which implies continuity of $v \mapsto \mathbb{B}_n(g_v)$, since $v - v' \to 0$ implies $g_v - g_{v'} \to 0$ in the $L_1(f_X)$ metric.

Q.E.D.

H.4. *Proof of Theorem 9*

Step 1. First we note that (G.2) is implied by

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}}\left|\mathbb{G}_n^o(g_v)-\bar{\mathbb{B}}_n(g_v)\right|\Big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(\delta_n/\ell_n) \tag{H.15}$$

in view of the Markov inequality. Using calculations similar to those in Step 5 in the proof of Lemma 7, we can conclude that for $X_v := \mathbb{G}_n^o(g_v)$, $v = (z, j)$, and $\bar{v} = (\bar{z}, j)$,

$$\sigma(X_v - X_{\bar{v}}) \le \|v-\bar{v}\| O_{\mathrm{P}_n}\big(1+h_n^{-1}\big),$$

where $\sigma^2(X_v - X_{\bar v}) := E_{P_n}((X_v - X_{\bar v})^2|\mathcal{D}_n)$. Application of the Gaussian maximal inequality quoted in Lemma 13, similarly Step 6 in the proof of Lemma 7, then gives

$$E_{\mathrm{P}_n}\Big(\sup_{\|v-\bar v\|\le\varepsilon} |X_v - X_{\bar v}|\big|\mathcal{D}_n\Big) = \varepsilon O_{\mathrm{P}_n}\big((1+h_n^{-1})\sqrt{\log n}\big), \tag{H.16}$$

where

$$\varepsilon \propto o\left(\frac{\delta_n/\ell_n}{h_n^{-1}\sqrt{\log n}}\right),$$

whence

$$E_{\mathrm{P}_n}\Big(\sup_{\|v-\bar v\|\le\varepsilon} |X_v - X_{\bar v}|\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(\delta_n/\ell_n). \tag{H.17}$$

Next we set up a regular mesh $\mathcal{V}_0 \subset \mathcal{V}$ with mesh width ε. The cardinality of the mesh is given by

$$K_n \propto (1/\varepsilon)^d \propto h_n^{-d}\lambda_n^d, \qquad \lambda_n = \frac{\sqrt{\log n}}{o(\delta_n/\ell_n)}.$$

With such mesh selection, we have that

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}} |X_v - X_{\pi(v)}|\big|\mathcal{D}_n\Big) \le E_{\mathrm{P}_n}\Big(\sup_{\|v-\bar v\|\le\varepsilon} |X_v - X_{\bar v}|\big|\mathcal{D}_n\Big) \le o_{\mathrm{P}_n}(\delta_n/\ell_n), \tag{H.18}$$

where $\pi(v)$ denotes a point in $\mathcal{V}_0$ that is closest to v.

The steps given below will show that there is a Gaussian process $\{Z_v, v\in\mathcal{V}\}$, which is independent of $\mathcal{D}_n$, having the same law as $\{\mathbb{B}_n(g_v), v\in\mathcal{V}\}$ and having the two key properties

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}_0} |X_v - Z_v|\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(\delta_n/\ell_n), \tag{H.19}$$

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}} |Z_v - Z_{\pi(v)}|\Big) = o(\delta_n/\ell_n). \tag{H.20}$$

The claim of the lemma then follows by setting $\{\bar{\mathbb{B}}_n(g_v), v\in\mathcal{V}\} = \{Z_v, v\in\mathcal{V}\}$ and then noting that

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}} |X_v - Z_v|\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(\delta_n/\ell_n)$$

holds by the triangle inequality for the sup norm and (H.18)–(H.20). Note that the last display is equivalent to (H.15). We now prove these assertions in the followings steps.

Step 2. In this step, we construct the process Z_v on points $v \in \mathcal{V}_0$, and show that (H.19) holds. In what follows, we use the notation $(X_v)_{v\in\mathcal{V}_0}$ to denote a K_n vector collecting X_v with indices $v \in \mathcal{V}_0$. We have that conditional on the data $\mathcal{D}_n$,

$$(X_v)_{v\in\mathcal{V}_0} = \hat{\Omega}_n^{1/2}\mathcal{N}, \quad \mathcal{N} \sim N(0, I),$$

where $\mathcal{N}$ is independent of $\mathcal{D}_n$ and

$$\hat{\Omega} = \mathbb{E}_n\big[p_i p_i'\big] \quad \text{and} \quad p_i = \big(g_v(U_i, Z_i)\big)_{v\in\mathcal{V}_0}.$$

We then set $(Z_v)_{v\in\mathcal{V}_0} = \Omega_n^{1/2}\mathcal{N}$ for $\Omega = E_{\mathrm{P}_n}[p_i p_i']$ and the same $\mathcal{N}$ as defined above.

Before proceeding further, we note that by construction the process $\{Z_v, v \in \mathcal{V}_0\}$ is independent of the data $\mathcal{D}_n$. This is facilitated by suitably enlarging the probability space as needed.[32]

Since the support of $\mathbf{K}$ is compact and points of the grid $\mathcal{V}_0$ are equally spaced, we have that

$$N_i := \big|\big\{v \in \mathcal{V}_0 : g_v(U_i, Z_i) \neq 0\big\}\big| \lesssim (h_n/\varepsilon)^d \lesssim \lambda_n^d.$$

Using the boundedness assumptions of the lemma, we have that

$$\|p_i\| \leq (\bar{U}/\underline{f})\sqrt{N_i}/h_n^{d/2} \lesssim (\lambda_n/h_n)^{d/2},$$

where $\bar{U}$ is the upper bound on U and $\underline{f}$ is the lower bound on the density f_n, both of which do not depend on n.

The application of Rudelson's (1999) law of large numbers for operators yields

$$E_{\mathrm{P}_n}\|\hat{\Omega}_n - \Omega_n\| \lesssim \sqrt{\log n/\big(n(h_n/\lambda_n)^d\big)}.$$

The application of the Gaussian maximal inequality quoted in Lemma 13 gives

$$E_{\mathrm{P}_n}\Big(\sup_{v\in\mathcal{V}_0}|X_v - Z_v||\mathcal{D}_n\Big) \lesssim \sqrt{\log K_n}\max_{v\in\mathcal{V}_0}\sigma(X_v - Z_v).$$

[32]Given the space $(A', \mathcal{A}', \mathrm{P}_n')$ that carries $\mathcal{D}_n$ and given a different space $(A'', \mathcal{A}'', \mathrm{P}_n'')$ that carries $\{Z_v, v \in \mathcal{V}_0\}$ as well as its complete version $\{Z_v, v \in \mathcal{V}\}$, we can take $(A, \mathcal{A}, \mathrm{P}_n)$ as the product of the two spaces, thereby maintaining independence between the data and the constructed process. Since $\{Z_v, v \in \mathcal{V}\}$ constructed below takes values in a separable metric space, it suffices to take $(A'', \mathcal{A}'', \mathrm{P}_n'')$ as the canonical probability space, as noted in Appendix A.

Since $(X_v)_{v\in\mathcal{V}_0} - (Z_v)_{v\in\mathcal{V}_0} = (\hat{\Omega}_n^{1/2} - \Omega_n^{1/2})'\mathcal{N}$, we have that

$$\max_{v\in\mathcal{V}_0} \sigma(X_v - Z_v)^2 \le \left\| \left(\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\right)^2 \right\| \le \left\| \hat{\Omega}_n^{1/2} - \Omega_n^{1/2} \right\|^2 \le K_n \|\hat{\Omega}_n - \Omega_n\|^2,$$

where the last inequality follows by a useful matrix inequality derived in Chetverikov (2011). Putting bounds together and using $\log K_n \lesssim \log n$ gives

$$E_{P_n}\Big(\sup_{v\in\mathcal{V}_0} |X_v - Z_v| \big| \mathcal{D}_n\Big) = O_{P_n}\left(\sqrt{\log n}\sqrt{K_n}\sqrt{\log n / \left(n(h_n/\lambda_n)^d\right)}\right) \tag{H.21}$$

$$= O_{P_n}\left(\log n / \sqrt{n(h_n/\lambda_n)^{2d}}\right) = o_{P_n}(\delta_n/\ell_n), \tag{H.22}$$

where the last condition holds by the conditions on the bandwidth.

Step 3. In this step, we complete the construction of the process $\{Z_v, v\in\mathcal{V}\}$. We have defined the process Z_v for all $v\in\mathcal{V}_0$. We want to embed these random variables into a path of a Gaussian process $\{Z_v, v\in\mathcal{V}\}$, whose covariance function is given by $(v,\bar{v}) \mapsto E_{P_n}[g_v g_{\bar{v}}]$. We want to maintain the independence of the process from $\mathcal{D}_n$. The construction follows by Lemma 11 in Belloni, Chernozhukov, and Fernandez-Val (2011). This lemma requires that a version of $\{Z_v, v\in\mathcal{V}\}$ has a.s. uniformly continuous sample paths, which follows from the Gaussian maximal inequalities and entropy calculations similar to those given in Step 3 in the proof of Lemma 7. Indeed, we can conclude that

$$\sigma(Z_v - Z_{\bar{v}}) \le \|v - \bar{v}\| C\left(1 + h_n^{-1}\right),$$

which establishes total boundedness of $\mathcal{V}$ under the standard deviation pseudometric. Moreover, application of the Gaussian maximal inequality Lemma 13 to Z_v gives

$$E_{P_n}\Big(\sup_{\|v-\bar{v}\|\le\varepsilon} |Z_v - Z_{\bar{v}}|\Big) \le C\varepsilon\left(\left(1 + h_n^{-1}\right)\sqrt{\log n}\right).$$

By a standard argument (e.g., van der Vaart and Wellner (1996)), these facts imply that the paths of Z_v are a.s. uniformly continuous.[33]

The last claim (H.20) follows from the preceding display, the choice of mesh width ε, and the inequality:

$$E_{P_n}\Big(\sup_{v\in\mathcal{V}} |Z_v - Z_{\pi(v)}|\Big) \le E_{P_n}\Big(\sup_{\|v-\bar{v}\|\le\varepsilon} |Z_v - Z_{\bar{v}}|\Big) = o(\delta_n/\ell_n). \qquad \textit{Q.E.D.}$$

[33]Note, however, that the process depends on n, and the statement here is a nonasymptotic statement, holding for any n. This property should not be confused with asymptotic equicontinuity, which does not hold here.

APPENDIX I: ASYMPTOTIC LINEAR REPRESENTATION FOR SERIES ESTIMATOR OF A CONDITIONAL MEAN

In this section, we use the primitive conditions set out in Example 5 of the main text to verify the required asymptotically linear representation for $\sqrt{n}(\hat{\beta}_n - \beta_n)$ using Newey (1997). This representation is also condition (b) of Theorem 7. We now reproduce the imposed conditions from the example for clarity. We note that it is also possible to develop similar conditions for nonlinear estimators; see, for example, Theorem 1(d) of Horowitz and Mammen (2004).

We have that $\theta_n(v) = E_{\mathrm{P}_n}[Y_i|V_i = v]$, assumed to be a continuous function. There is an i.i.d. sample (Y_i, V_i), $i = 1, \dots, n$, with $\mathcal{V} \subseteq \text{support}(V_i) \subseteq [0, 1]^d$ for each n. Here d does not depend on n, but all other parameters, unless stated otherwise, can depend on n. Then we have $\theta_n(v) = p_n(v)'\beta_n + A_n(v)$ for $p_n : \text{support}(V_i) \mapsto \mathbb{R}^{K_n}$ representing the series functions, β_n is the coefficient of the best least squares approximation to $\theta_n(v)$ in the population, and $A_n(v)$ is the approximation error. The number of series terms K_n depends on n.

Recall that we have imposed the following technical conditions in the main text:

Uniformly in n:

(i) p_n are either B-splines of a fixed order or trigonometric series terms or any other terms $p_n = (p_{n1}, \dots, p_{nK_n})'$ such that $\|p_n(v)\| \lesssim \zeta_n = \sqrt{K_n}$ for all $v \in \text{support}(V_i)$, $\|p_n(v)\| \gtrsim \zeta_n' \geq 1$ for all $v \in \mathcal{V}$, and $\log \text{lip}(p_n) \lesssim \log K_n$;

(ii) the mapping $v \mapsto \theta_n(v)$ is sufficiently smooth, namely $\sup_{v \in \mathcal{V}} |A_n(v)| \lesssim K_n^{-s}$ for some $s > 0$;

(iii) $\lim_{n\to\infty} (\log n)^c \sqrt{n} K_n^{-s} = 0$ for each $c > 0$[34];

(iv) for $\epsilon_i = Y_i - E_{\mathrm{P}_n}[Y_i|V_i]$, $E_{\mathrm{P}_n}[\epsilon_i^2|V_i = v]$ is bounded away from zero uniformly in $v \in \text{support}(V_i)$;

(v) eigenvalues of $Q_n = E_{\mathrm{P}_n}[p_n(V_i)p_n(V_i)']$ are bounded away from zero and from above;

(vi) $E_{\mathrm{P}_n}[|\epsilon_i|^4|V_i = v]$ is bounded from above uniformly in $v \in \text{support}(V_i)$;

(vii) $\lim_{n\to\infty} (\log n)^c K_n^5/n = 0$ for each $c > 0$.

We impose condition (i) directly through the choice of basis functions. Condition (ii) is a standard condition on the error of the series approximation and is the same as Assumption A3 of Newey (1997), which also was used by Chen (2007). Condition (v) is Assumption 2(i) of Newey (1997). The constant s will depend on the choice of basis functions. For example, if splines are used, then $s = \alpha/d$, where α is the number of continuous derivatives of $\theta_n(v)$ and d is the dimension of v. Restrictions on K_n in conditions (iii) and (vii) require that $\alpha > 5d/2$. Conditions (i), (vi), and (vii) and Theorem 7, namely Corollary 1,

[34]This condition, which is based on Newey (1997), can be relaxed to $(\log n)^c K_n^{-s+1} \to 0$ and $(\log n)^c \sqrt{n} K_n^{-s}/\zeta_n' \to 0$, using the recent results of Belloni, Chernozhukov, and Kato (2010) for least squares series estimators.

ensure that the constraint on the growth rate for the number of series terms is satisfied.

Define $S_n \equiv E[\epsilon_i^2 p_n(V_i) p_n(V_i)']$ and $\Omega_n \equiv Q_n^{-1} S_n Q_n^{-1}$. Arguments based on Newey (1997) give the following lemma, which verifies the linear expansion required in condition (b) of Theorem 7 with $\delta_n = 1/\log n$.

LEMMA 14—Asymptotically Linear Representation of the Series Estimator: *Suppose conditions* (i)–(vii) *hold. Then we have the asymptotically linear representation*

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) = \Omega_n^{-1/2} Q_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n p_n(Z_i)\epsilon_i + o_{\mathrm{P}_n}(1/\log n).$$

PROOF: As in Newey (1997), we have the representation, with probability approaching 1,

$$\hat{\beta}_n - \beta_n = n^{-1}\hat{Q}_n^{-1} \sum_{i=1}^n p_n(V_i)\epsilon_i + \nu_n, \tag{I.1}$$

where $\hat{Q}_n \equiv \mathbb{E}_n[p_n(V_i)p_n(V_i)']$, $\epsilon_i \equiv Y_i - E_{\mathrm{P}_n}[Y|V = V_i]$, and $\nu_n \equiv n^{-1}\hat{Q}_n^{-1} \times \sum_{i=1}^n p_n(V_i)A_n(V_i)$, where $A_n(v) := \theta_n(v) - p_n(v)'\beta_n$. As shown in the proof of Theorem 1 of Newey (1997), we have $\|\nu_n\| = O_{\mathrm{P}_n}(K_n^{-s})$. In addition, write

$$\begin{aligned}\bar{R}_n &:= [\hat{Q}_n^{-1} - Q_n^{-1}]n^{-1} \sum_{i=1}^n p_n(V_i)\epsilon_i \\ &= Q_n^{-1}[Q_n - \hat{Q}_n]n^{-1}\hat{Q}_n^{-1} \sum_{i=1}^n p_n(V_i)\epsilon_i.\end{aligned}$$

Then it follows from the proof of Theorem 1 of Newey (1997) that

$$\|\bar{R}_n\| = O_{\mathrm{P}_n}(\zeta_n K_n/n),$$

where $\zeta_n = \sqrt{K_n}$ by condition (i). Combining the results above gives

$$\hat{\beta}_n - \beta_n = n^{-1} Q_n^{-1} \sum_{i=1}^n p_n(V_i)\epsilon_i + R_n, \tag{I.2}$$

where the remainder term R_n satisfies

$$\|R_n\| = O_{\mathrm{P}_n}\left(\frac{K_n^{3/2}}{n} + K_n^{-s}\right).$$

Note that by condition (iv), eigenvalues of S_n^{-1} are bounded above. In other words, using the notation in Corollary 1 in the main text, we have that $\tau_n \lesssim 1$. Then

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) = n^{-1/2}\sum_{i=1}^{n} u_{i,n} + r_n, \tag{I.3}$$

where

$$u_{i,n} := \Omega_n^{-1/2} Q_n^{-1} p_n(V_i)\epsilon_i \tag{I.4}$$

and the new remainder term r_n satisfies

$$\|r_n\| = O_{\mathrm{P}_n}\big[n^{1/2}\big(K_n^{3/2}/n + K_n^{-s}\big)\big].$$

Therefore, $r_n = O_{\mathrm{P}_n}(1/\log n)$ if

$$(\log n) n^{1/2}\big(K_n^{3/2}/n + K_n^{-s}\big) \to 0, \tag{I.5}$$

which is satisfied under conditions (iii) and (vii), and we have proved the lemma. *Q.E.D.*

APPENDIX J: ASYMPTOTIC LINEAR REPRESENTATION FOR THE LOCAL POLYNOMIAL ESTIMATOR OF A CONDITIONAL MEAN

In this section, we provide details of Example 7 that were omitted in Appendix F. Results obtained in Kong, Linton, and Xia (2010) give the following lemma, which verifies the linear expansion required in condition (b) of Theorem 8 with $\delta_n = 1/\log n$.

LEMMA 15—Asymptotically Linear Representation of the Local Polynomial Estimator: *Suppose conditions* (i)–(vi) *hold*. *Then we have the asymptotically linear representation*, *uniformly in* $v = (z, j) \in \mathcal{V} \subseteq \mathcal{Z} \times \mathcal{J}$,

$$\big(nh_n^d\big)^{1/2}\big(\hat{\theta}_n(v) - \theta_n(v)\big) = \mathbb{G}_n(g_v) + o_{\mathrm{P}}(1/\log n).$$

PROOF: We first verify Assumptions A1–A7 in Kong, Linton, and Xia (2010; KLX hereafter). In our example, $\rho(y; \theta) = \frac{1}{2}(y - \theta)^2$ using the notation in KLX. Then $\varphi(y; \theta)$ in Assumptions A1 and A2 in KLX is $\varphi(y; \theta) = \varphi(y - \theta) = -(y - \theta)$. Then Assumption A1 is satisfied since the probability density function (p.d.f.) of U_i is bounded and U_i is a bounded random vector. Assumption A2 is trivially satisfied since $\varphi(u) = -u$. Assumption A3 follows since $K(\cdot)$ has compact support and is twice continuously differentiable. Assumption A4 holds by condition (ii) since X_i and X_j are independent in our example ($i \neq j$). Assumption A5 is implied directly by condition (i). Since we have i.i.d. data, mixing coefficients ($\gamma[k]$ using the notation of KLX) are identically zeros for any $k \geq 1$. The regression error U_i is assumed to be bounded, so that ν_1 in KLX can be arbitrarily large. Hence, to verify Assumption A6 of KLX, it suffices to check that for some $\nu_2 > 2$, $h_n \to 0$, $nh_n^d/\log n \to \infty$, $h_n^{d+2(p+1)}/\log n < \infty$, and $n^{-1}(nh_n^d/\log n)^{\nu_2/8} d_n \log n/M_n^{(2)} \to \infty$, where $d_n = (nh_n^d/\log n)^{-1/2}$ and $M_n^{(2)} = M^{1/4}(nh_n^d/\log n)^{-1/2}$ for some $M > 2$, by choosing $\lambda_2 = 1/2$ and $\lambda_1 = 3/4$ on page 1540 in KLX. By choosing a sufficiently large ν_2 (at least greater than 8), then $n^{-1}(nh_n^d)^{\nu_2/8} \to \infty$ holds. Then condition (vi) implies Assumption A6. Finally, condition (iv) implies Assumption A7 since we have i.i.d. data. Thus, we have verified all the conditions in KLX.

Let $\delta_n = 1/\log n$. Then it follows from Corollary 1 and Lemmas 8 and 10 of KLX that

$$\hat{\theta}_n(z,j) - \theta_n(z,j) = \frac{1}{nh_n^d f(z)} \mathbf{e}_1' S_p^{-1} \sum_{i=1}^n (e_j' U_i) K_h(Z_i - z) \mathbf{u}_p\left(\frac{Z_i - z}{h_n}\right) + B_n(z,j) + R_n(z,j), \tag{J.1}$$

where $\mathbf{e}_1$ is an $|A_p| \times 1$ vector whose first element is 1 and all others are 0's, S_p is a $|A_p| \times |A_p|$ matrix such that $S_p = \{\int z^u (z^v)' \, du : u \in A_p, v \in A_p\}$, $\mathbf{u}_p(z)$ is a $|A_p| \times 1$ vector such that $\mathbf{u}_p(z) = \{z^u : u \in A_p\}$,

$$B_n(z,j) = O\big(h_n^{p+1}\big), \quad \text{and} \quad R_n(z,j) = o_{\mathrm{P}}\left(\frac{\delta_n}{(nh_n^d)^{1/2}}\right)$$

uniformly in $(z,j) \in \mathcal{V}$. The exact form of $B_n(z,j)$ is given in equation (12) of KLX. The result that $B_n(z,j) = O(h_n^{p+1})$ uniformly in (z,j) follows from the standard argument based on Taylor expansion given in Fan and Gijbels (1996), KLX, or Masry (1996). The condition that $nh_n^{d+2(p+1)} \to 0$ at a polynomial rate in n corresponds to the undersmoothing condition. Now the lemma follows from (J.1) immediately since $\mathbf{K}(z/h) \equiv \mathbf{e}_1' S_p^{-1} K_h(z) \mathbf{u}_p(z/h)$ is a kernel of order $(p+1)$ (see Section 3.2.2 of Fan and Gijbels (1996)). *Q.E.D.*

APPENDIX K: LOCAL ASYMPTOTIC POWER COMPARISONS

We have shown in the main text that the test of $H_0 : \theta_{na} \leq \theta_{n0}$ of the form

$$\text{reject} \quad H_0 \quad \text{if} \quad \theta_{na} > \hat{\theta}_{n0}(p)$$

can reject all local alternatives θ_{na} that are more distant than $\bar{\sigma}_n \bar{a}_n$. We now provide a couple of examples of local alternatives against which our test has nontrivial power, but for which the CvM statistic of Andrews and Shi (2013; henceforth AS) does not. See also Armstrong (2011) for a comprehensive analysis of power properties of the KS (Kolmogorov–Smirnov) statistic of AS. It is evident from the results of AS on local asymptotic power that there are also models for which their CvM statistic will have power against some $n^{-1/2}$ alternatives, whereas our approach will not.[35] We conclude that neither approach dominates.

We consider two examples in which

$$Y_i = \theta_n(V_i) + U_i,$$

where U_i are i.i.d. with $E[U_i|V_i] = 0$ and V_i are i.i.d. random variables uniformly distributed on $[-1, 1]$. Suppose that for all $v \in [-1, 1]$, we have

$$\theta^* \le E[Y_i|V_i = v]$$

or, equivalently

$$\theta^* \le \theta_0 = \min_{v\in[-1,1]} \theta_n(v).$$

In the examples below, we consider two specifications of the bounding function $\theta_n(v)$, each with

$$\min_{v\in[-1,1]} \theta_n(v) = 0,$$

and we analyze asymptotic power against a local alternative $\theta_{na} > \theta_0$.

Following AS, consider the CvM test statistic

$$T_n(\theta) := \int \left[n^{1/2} \frac{\overline{m}_n(g;\theta)}{\hat{\sigma}_n(g;\theta) \vee \varepsilon} \right]_-^2 dQ(g) \tag{K.1}$$

for some $\varepsilon > 0$, where $[u]_- := -u1(u < 0)$ and θ is the parameter value being tested. In the present context, we have

$$\overline{m}_n(g;\theta) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta)g(V_i),$$

where $g \in \mathcal{G}$ are instrument functions used to transform the conditional moment inequality $E[Y - \theta|V = v]$ almost everywhere $v \in \mathcal{V}$ to unconditional inequalities, and $Q(\cdot)$ is a measure on the space $\mathcal{G}$ of instrument functions as described in AS, Section 3.4. $\hat{\sigma}_n(g;\theta)$ is a uniformly consistent estimator for $\sigma_n(g;\theta)$, the standard deviation of $n^{1/2}\overline{m}_n(g;\theta)$.

We can show that $T_n(\theta) = \tilde{T}_n(\theta) + o_p(1)$, where

$$\tilde{T}_n(\theta) := \int \big[\beta_n(\theta, g)/\big(\sigma_n(g;\theta) \vee \varepsilon\big) + w(\theta, g)\big]_-^2 \, dQ(g),$$

[35]For the formal results, see AS, Section 7, Theorem 4. In the examples that follow, their Assumption LA3′ is violated, as is also the case in the example covered in their Section 13.5.

where $w(\theta, g)$ is a mean-zero Gaussian process and $\beta_n(\theta, g)$ is a deterministic function of the form

$$\beta_n(\theta, g) \equiv \sqrt{n}E\{[\theta_n(V_i) - \theta]g(V_i)\}.$$

For any θ, the testing procedure based on the CvM statistic rejects $H_0 : \theta \leq \theta_{n0}$ if

$$T_n(\theta) > c(\theta, 1 - \alpha),$$

where $c(\theta, 1-\alpha)$ is a generalized moment selection (GMS) critical value that satisfies

$$c(\theta, 1-\alpha) = (1-\alpha)\text{-quantile of } \left(\int [\varphi_n(\theta, g)/(\sigma_n(g; \theta) \vee \varepsilon) + w(\theta, g)]_-^2 \, dQ(g) \right) + o_p(1)$$

and $\varphi_n(\theta, g)$ is a GMS function that satisfies $0 \leq \varphi_n(\theta, g) \leq \beta_n(\theta, g)$ with probability approaching 1 whenever $\beta_n(\theta, g) \geq 0$; see AS, Section 4.4, for further details. Relative to $\tilde{T}_n(\theta)$, in the integrand of the expression above, $\varphi_n(\theta, g)$ is replaced with $\beta_n(\theta, g)$. Hence if

$$\sup_{g \in \mathcal{G}} [\beta_n(\theta_{na}, g)]_- \to 0$$

for the sequence of local alternatives θ_{na}, then

$$\liminf_{n \to \infty} \mathrm{P}(T_n(\theta_{na}) > c(\theta_{na}, 1-\alpha)) \leq \alpha,$$

since asymptotically $c(\theta_{na}, 1-\alpha)$ exceeds the $(1-\alpha)$-quantile of $\tilde{T}_n(\theta)$. It follows that the CvM test has only trivial power against such a sequence of alternatives. The same conclusion holds using plug-in asymptotic critical values, since these are no smaller than GMS critical values.

In the following two examples, we now verify that $\sup_{g \in \mathcal{G}} [\beta_n(\theta_{na}, g)]_- \to 0$. We assume that instrument functions g are either indicators of boxes or cubes, defined in AS, Section 3.3, and hence bounded between 0 and 1.

EXAMPLE K.1—Unique, Well Defined Optimum: Let the function $\theta(\cdot)$ be specified as

$$\theta_n(v) = |v|^a$$

for some $a \geq 1$.

Let us now proceed to bound, using that $0 \leq g \leq 1$,

$$\begin{aligned}\big[\beta_n(\theta_{na}, g)\big]_- &= \sqrt{n}\big[E\big\{\big[\theta_n(V_i) - \theta_{na}\big]g(V_i)\big\}\big]_- \\ &\leq \sqrt{n}E\big\{\big[\theta_n(V_i) - \theta_{na}\big]_-\big\} \\ &= \sqrt{n}\int_{-1}^{1}\big(\theta_{na} - |v|^a\big)1\big\{|v|^a \leq \theta_{na}\big\}\,dv \\ &= 2\sqrt{n}\int_{0}^{1}\big(\theta_{na} - v^a\big)1\big\{v \leq \theta_{na}^{1/a}\big\}\,dv \\ &= \frac{2a}{a+1}\sqrt{n}\theta_{na}^{(a+1)/a} \\ &\equiv \overline{\beta}_n.\end{aligned}$$

Note that

$$\theta_{na} = o\big(n^{-a/[2(a+1)]}\big) \quad \Rightarrow \quad \overline{\beta}_n \to 0.$$

Thus, in this case the asymptotic rejection probability of the CvM test for the local alternative θ_{na} is bounded above by α. On the other hand, by Theorems 1 and 2 of the main text, our test rejects all local alternatives θ_{na} that are more distant than $\bar{\sigma}_n\bar{a}_n$ with probably at least α asymptotically. It suffices to find a sequence of local alternatives θ_{na} such that $\theta_{na} = o(n^{-a/[2(a+1)]})$ but $\theta_{na} \gg \bar{\sigma}_n\bar{a}_n$.

For instance, consider the case where $a = 2$. Then

$$\sqrt{n}\theta_{na}^{3/2} \to 0 \quad \Rightarrow \quad \overline{\beta}_n \to 0,$$

that is, $\theta_{na} = o(n^{-1/3}) \Rightarrow \overline{\beta}_n \to 0$, so the CvM test has trivial asymptotic power against θ_{na}. In contrast, since this is a very smooth case, our approach can achieve $\bar{\sigma}_n\bar{a}_n = O(n^{-\delta})$ for some δ that can be close to $1/2$, for instance, by using a series estimator with a slowly growing number of terms, or a higher order kernel or local polynomial estimator. Our test would then be able to reject any θ_{na} that converges to zero faster than $n^{-1/3}$ but more slowly than $n^{-\delta}$.

EXAMPLE K.2—Deviation With Small Support: Now suppose that the form of the conditional mean function, $\theta_n(v) \equiv E[Y_i|V_i = v]$, is given by

$$\theta_n(v) := \bar{\theta}(v) - \tau_n^a\big(\phi(v/\tau_n) - \phi(0)\big),$$

where τ_n is a sequence of positive constants converging to zero and $\phi(\cdot)$ is the standard normal density function. Let $\bar{\theta}(v)$ be minimized at zero so that

$$\theta_0 = \min_{v\in[-1,1]} \theta_n(v) = \min_{v\in[-1,1]} \bar{\theta}(v) = 0.$$

Let the alternative be $\tilde{\theta}_{na} \equiv \tau_n^a \phi(0)$. Again, the behavior of the AS statistic is driven by $[\beta_n(\tilde{\theta}_{na}, g)]_-$, which we bound from above as

$$\begin{aligned}[\beta_n(\tilde{\theta}_{na}, g)]_- &= \sqrt{n}\big[E\big\{[\theta_n(V_i) - \tilde{\theta}_{na}]g(V_i)\big\}\big]_- \\ &\leq \sqrt{n}E\{\tau_n^a \phi(V_i/\tau_n)\} \\ &= \frac{\sqrt{n}}{2}\int_{-1}^{1} \tau_n^a \phi(v/\tau_n)\,dv \\ &\leq \frac{\sqrt{n}}{2}\tau_n^{a+1} \equiv \overline{\beta}_n.\end{aligned}$$

Consider the case $a = 2$. If $\tau_n = o(n^{-1/6})$, then $\overline{\beta}_n \to 0$, so that again the CvM test has only trivial asymptotic power. If $\tau_n = n^{-1/6-c/2}$ for some small positive constant c, then $\tilde{\theta}_{na} \equiv n^{-1/3-c}\phi(0)$. Note that

$$f(v) := \tau_n^2 \phi(v/\tau_n) \quad \Rightarrow \quad f''(v) = \phi''(v/\tau_n) \leq \overline{\phi''} < \infty$$

for some constant $\overline{\phi''}$. Hence, if $\bar{\theta}(v)$ is twice continuously differentiable, we can use a series or kernel estimator to estimate $\theta_n(v)$ uniformly at the rate of $(\log n)^d n^{-2/5}$ for some $d > 0$, leading to nontrivial power against alternatives $\tilde{\theta}_{na}$ for sufficiently small c.

APPENDIX L: RESULTS OF ADDITIONAL MONTE CARLO EXPERIMENTS

In this section, we present the results of some additional Monte Carlo experiments to further illustrate the finite-sample performance of our method. We consider two types of additional data-generating processes (DGPs). The first set of DGPs, DGP5–DGP8, are motivated by Manski and Pepper (2009) and were discussed briefly in Example B of the main text. The second set, DGP9–DGP12, are from Section 10.3 of AS.

L.1. *Monte Carlo Designs*

In DGP5–DGP8, we consider the lower bound on $\theta^* = E[Y_i(t)|V_i = v]$ under the monotone instrumental variable (MIV) assumption, where t is a treatment, $Y_i(t)$ is the corresponding potential outcome, and V_i is a monotone instrumental variable. The lower bound on $E[Y_i(t)|V_i = v]$ can be written as

$$\max_{u \leq v} E\big[Y_i \cdot 1\{Z_i = t\} + y_0 \cdot 1\{Z_i \neq t\}|V_i = u\big], \tag{L.1}$$

where Y_i is the observed outcome, Z_i is a realized treatment, and y_0 is the left end point of the support of Y_i; see Manski and Pepper (2009). The parameter of interest is $\theta^* = E[Y_i(1)|V_i = 1.5]$.

In DGP5, $V_0 = \mathcal{V}$ and the MIV assumption has no identifying power. In other words, the bound-generating function is flat on $\mathcal{V}$, in which case the bias of the analog estimator is most acute; see Manski and Pepper (2009). In DGP6, the MIV assumption has identifying power and V_0 is a strict subset of $\mathcal{V}$. In DGP7–DGP8, we set V_0 to be a singleton set.

Specifically, for DGP5–DGP8 we generated 1000 independent samples as

$$V_i \sim \text{Unif}[-2, 2], \quad Z_i = 1\{\varphi_0(V_i) + \varepsilon_i > 0\}, \quad \text{and}$$
$$Y_i = \min\{\max\{-0.5, \sigma_0(V_i)U_i\}, 0.5\},$$

where $\varepsilon_i \sim N(0, 1)$, $U_i \sim N(0, 1)$, $\sigma_0(V_i) = 0.1 \times |V_i|$, and (V_i, U_i) are statistically independent ($i = 1, \dots, n$). The bounding function has the form

$$\theta(v) := E[Y_i \cdot 1\{Z_i = 1\} + y_0 \cdot 1\{Z_i \neq 1\} | V_i = v]$$
$$= -0.5\Phi[-\varphi_0(v)],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. For DGP5, we set $\varphi_0(v) \equiv 0$. In this case, the bounding function is completely flat ($\theta_l(v) = -0.25$ for each $v \in \mathcal{V} = [-2, 1.5]$). For DGP6, an alternative specification is considered:

$$\varphi_0(v) = v1(v \leq 1) + 1(v > 1).$$

In this case, $v \mapsto \theta(v)$ is strictly increasing on $[-2, 1]$ and is flat on $[1, 2]$, and $V_0 = [1, 1.5]$ is a strict subset of $\mathcal{V} = [-2, 1.5]$. For DGP7, we consider

$$\varphi_0(v) = -2v^2.$$

In this case, $v \mapsto \theta_l(v)$ has a unique maximum at $v = 0$; thus, $V_0 = \{0\}$ is singleton. For DGP8, we consider

$$\varphi_0(v) = -10v^2.$$

In this case, $v \mapsto \theta(v)$ has a unique maximum at $v = 0$ and is more peaked than that of DGP7. Figures S.1 and S.2, and Figures S.3 and S.4, show data realizations and bounding functions for DGP1–DGP4 described in Section 7 of the main text, and DGP5–DGP8 described above, respectively.

DGP9–DGP12 use the bounding functions in Section 10.3 of AS. DGP9 and DGP10 feature a roughly plateau-shaped bounding function given by

$$\theta(v) := L\phi(v^{10}) \tag{L.2}$$

instead of $\theta(v) := L\phi(v)$ as in Section 7 of the main text. DGP11 and DGP12 use the roughly double-plateau-shaped bounding function

$$\theta(v) := L \cdot \max\{\phi((v - 1.5)^{10}), \phi((v + 1.5)^{10})\}. \tag{L.3}$$

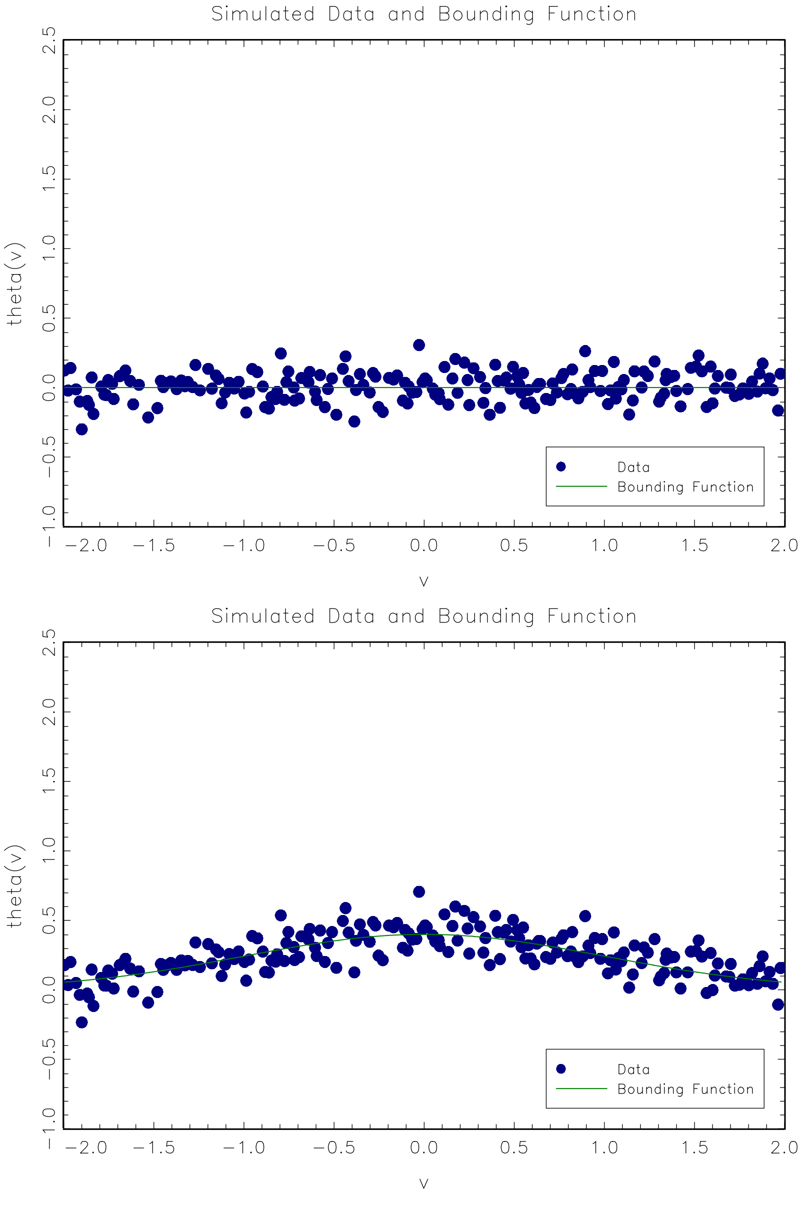

FIGURE S.1.—Simulated data and bounding functions: DGP1 and DGP2.

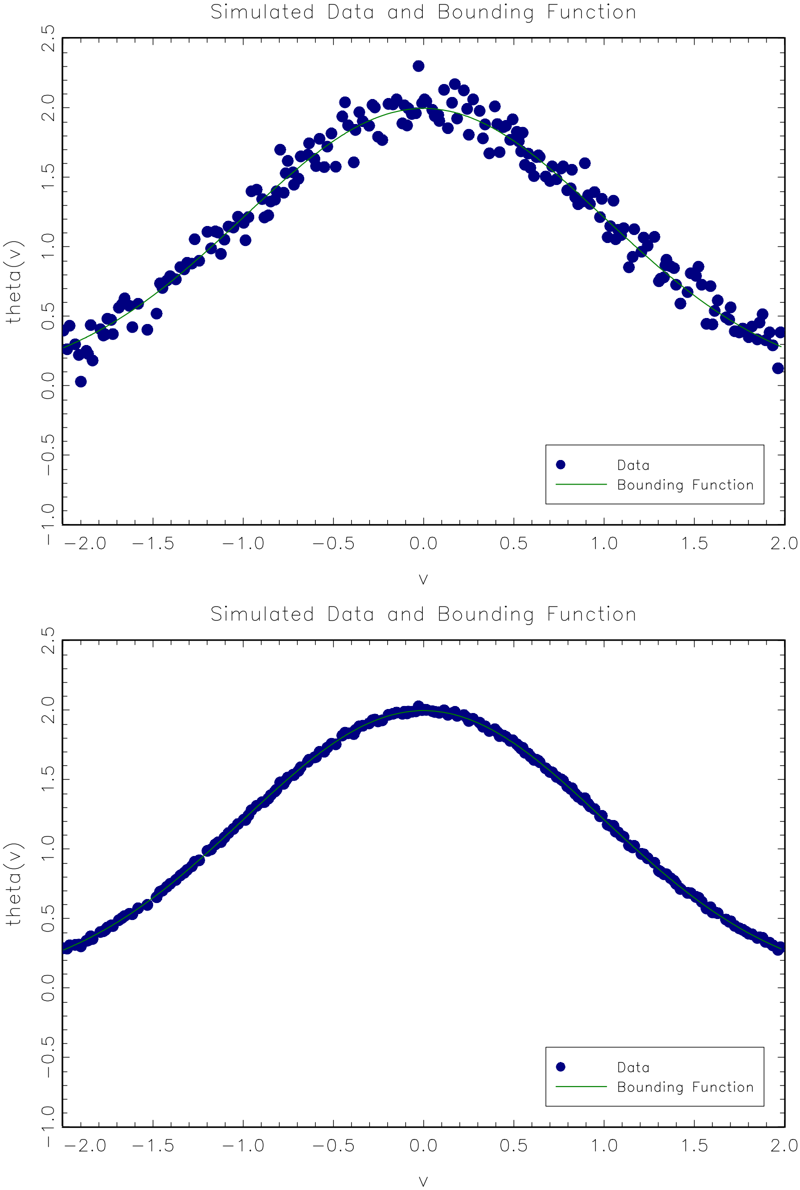

FIGURE S.2.—Simulated data and bounding functions: DGP3 and DGP4.

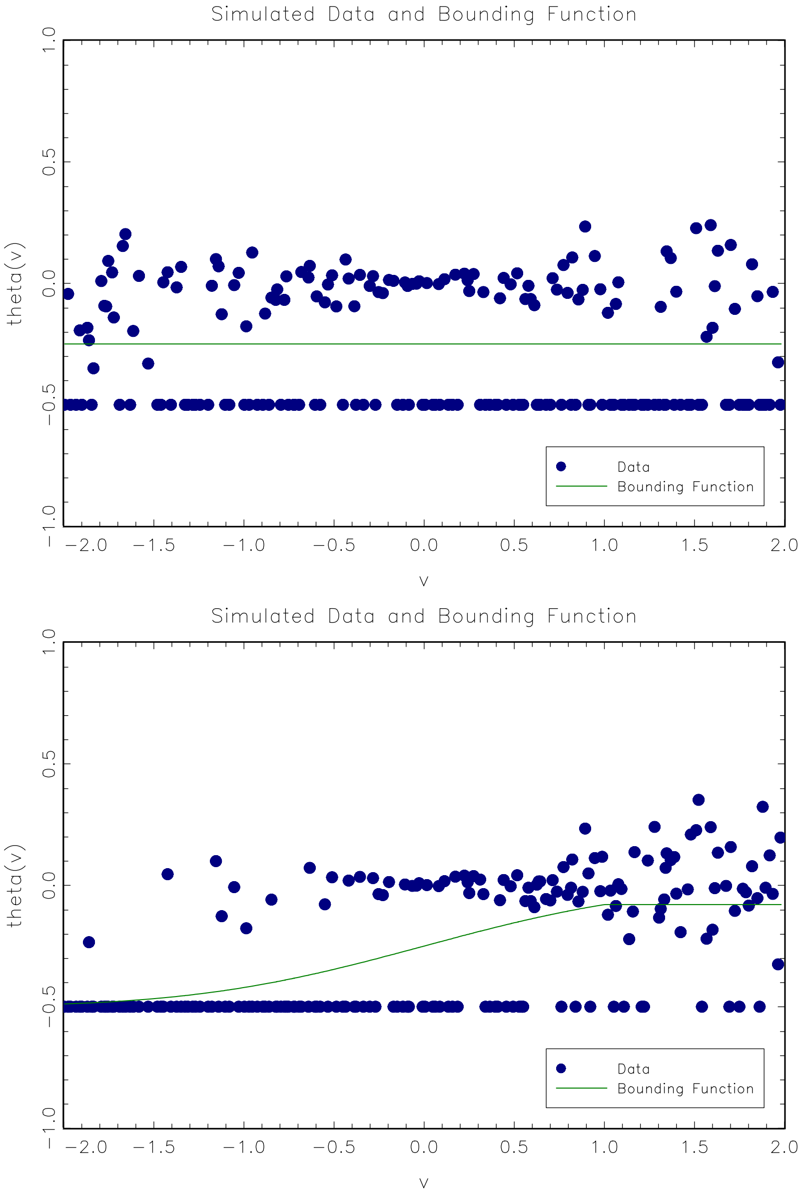

FIGURE S.3.—Simulated data and bounding functions: DGP5 and DGP6.

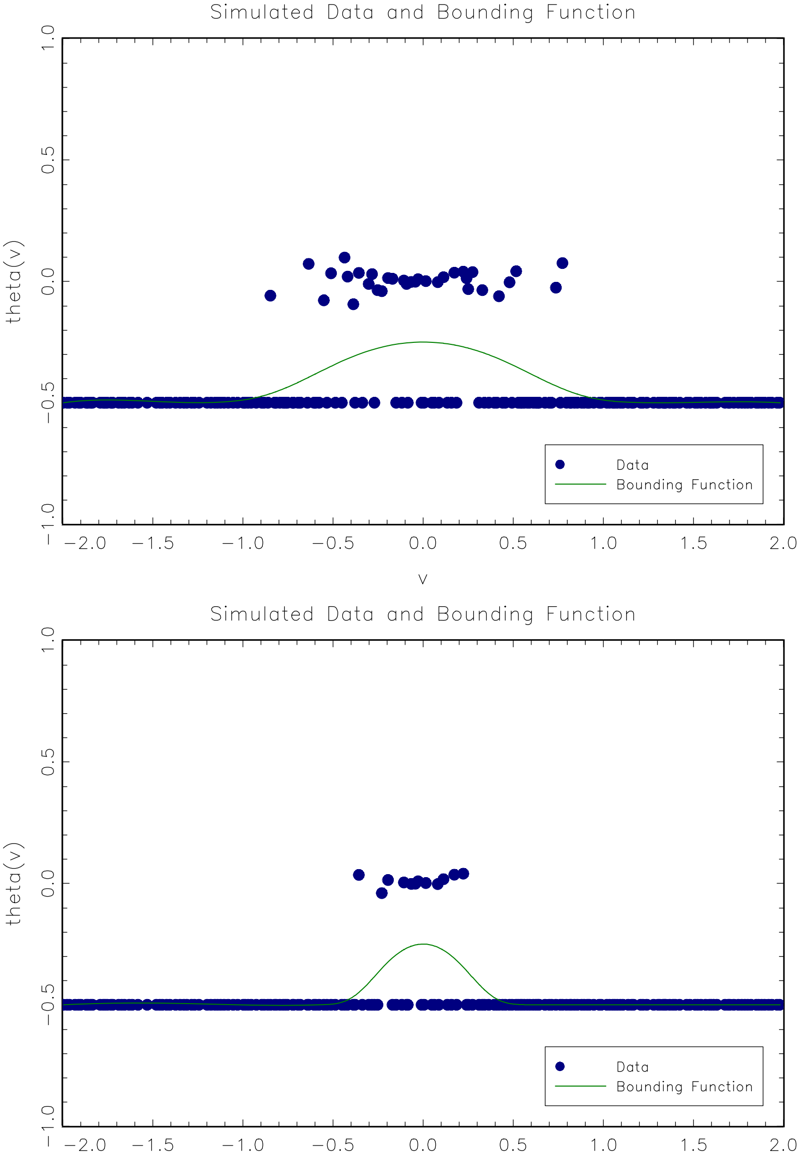

FIGURE S.4.—Simulated data and bounding functions: DGP7 and DGP8.

Specifically, we generated 1000 independent samples from the model,

$$V_i \sim \text{Unif}[-2, 2], \quad U_i = \min\{\max\{-3, \sigma\tilde{U}_i\}, 3\}, \quad \text{and}$$
$$Y_i = \theta(V_i) + U_i,$$

where $\tilde{U}_i \sim N(0, 1)$, with L and σ as follows:

$$\text{DGP9 and DGP11:} \quad L = 1 \text{ and } \sigma = 0.1,$$
$$\text{DGP10 and DGP12:} \quad L = 5 \text{ and } \sigma = 0.1.$$

Figure S.5 illustrates the bounding function and data realizations for DGP9 and DGP10; Figure S.6 provides the same for DGP11 and DGP12. Interest again lies in inference on $\theta_0 = \sup_{v \in \mathcal{V}} \theta(v)$, which in these DGPs is $\theta_0 = L\phi(0)$.

L.2. *Simulation Results*

To evaluate the relative performance of our inference method in DGP5–DGP8, we used our method with cubic B-splines with knots equally spaced over the sample quantiles of V_i, and we also implemented one of the inference methods proposed by AS, specifically their Cramér–von Mises-type (CvM) statistic with PA/Asy and GMS/Asy critical values. Implementation details for our method with B-splines are the same as in Section 7.2 of the main text. Tuning parameters for CvM were chosen exactly as in AS (see Section 9).[36] We considered sample sizes $n = 250$, $n = 500$, and $n = 1000$.

The coverage probability is evaluated at the true lower bound θ_0 (with the nominal level of 95%), and the false coverage probability (FCP) is evaluated at a θ value outside the identified set. For DGP5, we set $\theta = \theta_0 - 0.03$; for DGP6 and DGP7, $\theta = \theta_0 - 0.05$; for DGP8, $\theta = \theta_0 - 0.07$. These points are chosen differently across different DGPs to ensure that the FCPs have similar values. This type of FCP was reported in AS, along with a so-called CP correction (similar to size correction in testing). We did not do CP correction in our reported results. There were 1000 replications for each experiment. Table S.I summarizes the results of Monte Carlo experiments. CLR and AS refer to our inference method and that of AS, respectively.

First, we consider Monte Carlo results for DGP5. The discrepancies between nominal and actual coverage probabilities are not large across all methods, implying that all of them perform well in finite samples. For DGP5, since the true arg max set V_0 is equal to $\mathcal{V}$, an estimated V_0 should be the entire set $\mathcal{V}$. Thus the simulation results are the same whether or not we estimate V_0, since for most of simulation draws, $\hat{V}_n = \mathcal{V}$. Similar conclusions hold for AS with

[36]In DGP5–DGP8, our Monte Carlo design differs from that of AS, and alternative choices of tuning parameters could perform more or less favorably in our design. We did not examine sensitivity to the choice of tuning parameters for their method.

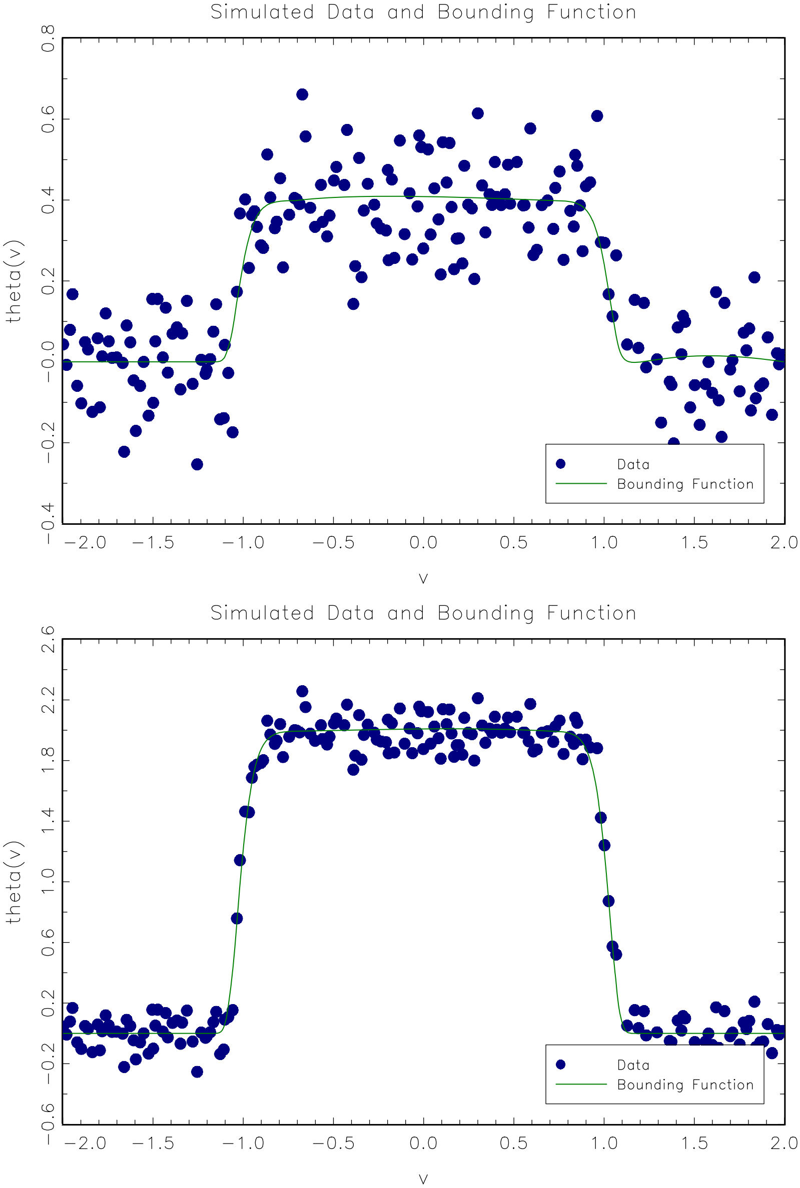

FIGURE S.5.—Simulated data and bounding functions: DGP9 and DGP10.

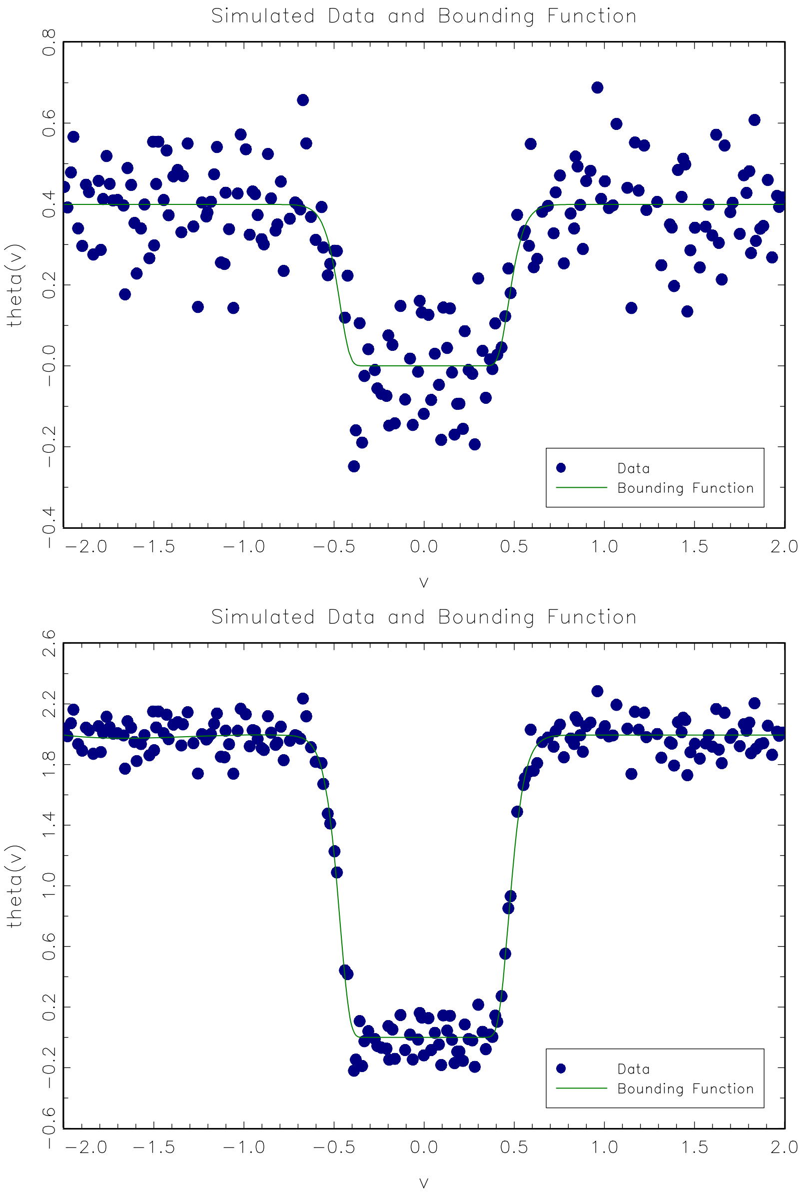

FIGURE S.6.—Simulated data and bounding functions: DGP11 and DGP12.

TABLE S.I
RESULTS FOR MONTE CARLO EXPERIMENTS[a]

DGP	Sample Size	Critical Value Estimating V_n?	Cov. Prob.	False Cov. Prob.
CLR With Series Estimation Using B-Splines				
5	250	No	0.914	0.709
5	250	Yes	0.914	0.709
5	500	No	0.935	0.622
5	500	Yes	0.935	0.622
5	1000	No	0.947	0.418
5	1000	Yes	0.947	0.418
6	250	No	0.953	0.681
6	250	Yes	0.942	0.633
6	500	No	0.967	0.548
6	500	Yes	0.941	0.470
6	1000	No	0.973	0.298
6	1000	Yes	0.957	0.210
7	250	No	0.991	0.899
7	250	Yes	0.980	0.841
7	500	No	0.996	0.821
7	500	Yes	0.994	0.697
7	1000	No	0.987	0.490
7	1000	Yes	0.965	0.369
8	250	No	0.999	0.981
8	250	Yes	0.996	0.966
8	500	No	1.000	0.984
8	500	Yes	0.999	0.951
8	1000	No	0.998	0.909
8	1000	Yes	0.995	0.787
AS With CvM-Type Statistic				
5	250	PA/Asy	0.951	0.544
5	250	GMS/Asy	0.945	0.537
5	500	PA/Asy	0.949	0.306
5	500	GMS/Asy	0.945	0.305
5	1000	PA/Asy	0.962	0.068
5	1000	GMS/Asy	0.956	0.068
6	250	PA/Asy	1.000	0.941
6	250	GMS/Asy	0.990	0.802
6	500	PA/Asy	1.000	0.908
6	500	GMS/Asy	0.980	0.674
6	1000	PA/Asy	1.000	0.744
6	1000	GMS/Asy	0.980	0.341

(*Continues*)

TABLE S.I—*Continued*

DGP	Sample Size	Critical Value Estimating V_n?	Cov. Prob.	False Cov. Prob.
7	250	PA/Asy	1.000	1.000
7	250	GMS/Asy	0.997	0.948
7	500	PA/Asy	1.000	0.997
7	500	GMS/Asy	0.997	0.916
7	1000	PA/Asy	1.000	0.993
7	1000	GMS/Asy	0.997	0.823
8	250	PA/Asy	1.000	1.000
8	250	GMS/Asy	1.000	0.988
8	500	PA/Asy	1.000	1.000
8	500	GMS/Asy	0.999	0.972
8	1000	PA/Asy	1.000	1.000
8	1000	GMS/Asy	1.000	0.942

[a]CLR and AS refer to our inference methods and those of Andrews and Shi (2013), respectively. There were 1000 replications per experiment.

CvM between PA/Asy and GMS/Asy critical values. In terms of false coverage probability, CvM with either critical value performs better than our method.

We now move to DGP6–DGP8. In DGP6, the true arg max set V_0 is $[1, 1.5]$, and in DGP7 and DGP8, V_0 is a singleton set. In these cases, the true arg max set V_0 is a strict subset of $\mathcal{V}$. Hence, we expect that it is important to estimate V_0. On average, for DGP6, the estimated sets were $[-0.793, 1.5]$ when $n = 250$, $[-0.359, 1.5]$ when $n = 500$, and $[-0.074, 1.5]$ when $n = 1000$; for DGP7, the estimated sets were $[-0.951, 0.943]$ when $n = 250$, $[-0.797, 0.798]$ when $n = 500$, and $[-0.684, 0.680]$ when $n = 1000$; for DGP8, the estimated sets were $[-1.197, 0.871]$ when $n = 250$, $[-0.662, 0.645]$ when $n = 500$, and $[-0.403, 0.402]$ when $n = 1000$.

Hence, an average estimated set is larger than V_0; however, it is still a strict subset of $\mathcal{V}$ and gets smaller as n gets larger. For all the methods, the Monte Carlo results are consistent with asymptotic theory. Unlike in DGP5, the CLR method performs better than the AS method in terms of false coverage probability.[37] As can be seen from the table, the CLR method performs better when V_0 is estimated in terms of making the coverage probability less conservative and also of making the false coverage probability smaller. Similar gains are obtained for the CvM with GMS/Asy critical values, relative to that with PA/Asy critical values.

We now turn to DGP9–DGP12. AS, Section 10.3, reported results for their approach using their CvM and KS statistics, and we refer the reader to their paper for results using their method. They also included results for our approach

[37]As in Section 7, this conclusion will remain valid even with CP correction as in AS, since our method performs better in DGP6–DGP8, where we have overcoverage.

TABLE S.II

RESULTS FOR MONTE CARLO EXPERIMENTS (CLR WITH SERIES ESTIMATION USING B-SPLINES)[a]

DGP	Sample Size	Critical Value Estimating V_n?	Ave. Smoothing Parameter	Cov. Prob.	False Cov. Prob.	Ave. Argmax Set Min.	Ave. Argmax Set Max.
9	500	No	35.680	0.920	0.562	−1.799	1.792
9	500	Yes	35.680	0.870	0.475	−1.001	1.001
9	1000	No	39.662	0.937	0.487	−1.801	1.797
9	1000	Yes	39.662	0.913	0.380	−0.977	0.977
10	500	No	39.090	0.887	0.534	−1.799	1.792
10	500	Yes	39.090	0.825	0.428	−0.912	0.912
10	1000	No	41.228	0.920	0.477	−1.801	1.797
10	1000	Yes	41.228	0.891	0.351	−0.902	0.903
11	500	No	35.810	0.880	0.462	−1.799	1.792
11	500	Yes	35.810	0.853	0.399	−1.799	1.792
11	1000	No	40.793	0.937	0.374	−1.801	1.797
11	1000	Yes	40.793	0.912	0.299	−1.801	1.797
12	500	No	39.474	0.836	0.459	−1.799	1.792
12	500	Yes	39.474	0.811	0.386	−1.799	1.792
12	1000	No	42.224	0.917	0.367	−1.801	1.797
12	1000	Yes	42.224	0.885	0.294	−1.801	1.797

[a] DGP9–DGP12 correspond to DGP1–DGP4 in Andrews and Shi (2013, Section 10.3). The last two columns report the average values of the minimum and maximum of the arg max set. The estimated set is allowed to be disconnected and so the interval between the minimum and maximum of the arg max set is just an outer set for the estimated arg max set.

using B-splines and local-linear estimation of the bounding function. Here we provide further investigation of the performance of our method in additional Monte Carlo simulations.

From Figures S.5 and S.6, we see that the bounding function is very nearly flat in some regions, including areas close to V_0, and also has very large derivatives a bit farther away from V_0. The functions are smooth, but the steep derivatives mimic discontinuity points and are challenging for nonparametric estimation methods. The AS approach does not rely on smoothness of the bounding function and performs better in most—though not all—of the comparisons of AS. The performance of our approach improves with the sample size, as expected.

Our Monte Carlo experiments for DGP9–DGP12 further examine the performance of our method in such a setup. In all experiments, we report coverage probabilities (CPs) at $\theta_0 = L\phi(0)$ and FCPs at $\theta_0 - 0.02$ as in AS. We provide results for sample sizes $n = 500$ and $n = 1000$, both with and without estimation of the contact set. In Table S.II, we report the results of series estimation via B-splines. We used cubic B-splines and our implementation was identical to that described in Section 7 of the main text. Compared to the results in Table I of the main text for DGP1–DGP4, we see that the average number of series terms is much higher. This is due to a higher number of terms selected during cross-validation, presumably because of the regions with very large derivatives. Our coverage probabilities are below the nominal level, but they improve with the sample size, as in AS, ranging from 0.885 to 0.937 across DGPs at $n = 1000$. Moreover, we see that our method using $\mathsf{V} = \mathcal{V}$ rather than $\mathsf{V} = \hat{V}_n$ actually performs better in this setup.

To further investigate the challenge of nonparametrically estimating a bounding function with steep derivatives, we implemented our method with a locally constant Nadaraya–Watson kernel estimator. The functions are in fact nearly locally constant at most points, with the exception of the relatively narrow regions with steep derivatives. The top half of Table S.III presents the results with a bandwidth selected the same way as for the local-linear estimator in Section 7.3, equation (7.3). When V_n is estimated, coverage probabilities in these DGPs range from 0.903 to 0.923 when $n = 500$ and 0.926 to 0.945 when $n = 1000$, becoming closer to the nominal level. The procedure exhibits good power in all cases, with FCPs decreasing with the sample size. These results are qualitatively similar to those reported in AS for the local-linear estimator. We also include results when $\mathcal{V}$ is used instead of estimating the contact set. This results in higher coverage probabilities for θ_0, in most cases closer to the nominal level, but also somewhat higher FCPs. Overall performance remains reasonably good.

The bottom half of Table S.III gives results for locally constant kernel estimation using an identical rule-of-thumb for bandwidth selection, $h = \hat{h}_{\mathrm{ROT}} \times \hat{s}_v$, but without applying an undersmoothing factor. The proofs of our asymptotic results use undersmoothing, but with a locally constant kernel estimator this does not appear to be essential. Our exploratory Monte Carlo results are very good, offering support to that view. Compared to the results with undersmoothing, all coverage probabilities increase and all FCPs decrease. This suggests that future research on the possibility of abandoning undersmoothing may be warranted.

The overall results of this section support the conclusions reached in Section 7 of the main text regarding comparisons to AS. In completely flat cases, the AS method outperforms our method, whereas in nonflat cases, our method outperforms the AS method. In this section, we also considered some intermediate cases. In DGP7, where the bounding function is partly flat, our method performed favorably. More generally, there is a wide range of intermediate cases that could be considered, and we would expect the approach of AS to perform favorably in some cases too. Indeed, in DGP9–DGP12 from AS, the bounding function exhibits areas with extremely steep derivatives. Their results indicate that in these DGPs, their approach performs better at smaller sample sizes ($n = 100, 250$) than does our approach, which is based on nonparametric estimation methods. However, at larger sample sizes ($n = 500, 1000$), even with the very steep derivatives of DGP9–DGP12, our approach performs well,

TABLE S.III

RESULTS FOR MONTE CARLO EXPERIMENTS (NADARAYA–WATSON KERNEL ESTIMATION)[a]

DGP	Sample Size	Critical Value Estimating V_n?	Ave. Smoothing Parameter	Cov. Prob.	False Cov. Prob.	Ave. Argmax Set Min.	Ave. Argmax Set Max.
Using Bandwidth $h = \hat{h}_{\mathrm{ROT}} \times \hat{s}_v \times n^{1/5} \times n^{-2/7}$							
9	500	No	0.206	0.944	0.496	−1.799	1.792
9	500	Yes	0.206	0.916	0.391	−0.955	0.954
9	1000	No	0.169	0.966	0.326	−1.801	1.796
9	1000	Yes	0.169	0.945	0.219	−0.941	0.940
10	500	No	0.166	0.945	0.523	−1.799	1.792
10	500	Yes	0.166	0.903	0.411	−0.879	0.880
10	1000	No	0.136	0.963	0.387	−1.801	1.796
10	1000	Yes	0.136	0.926	0.266	−0.868	0.868
11	500	No	0.195	0.938	0.403	−1.799	1.792
11	500	Yes	0.195	0.923	0.345	−1.799	1.792
11	1000	No	0.160	0.951	0.201	−1.801	1.796
11	1000	Yes	0.160	0.926	0.152	−1.801	1.796
12	500	No	0.169	0.937	0.439	−1.799	1.792
12	500	Yes	0.169	0.917	0.365	−1.799	1.792
12	1000	No	0.138	0.947	0.235	−1.801	1.796
12	1000	Yes	0.138	0.933	0.176	−1.801	1.796
Using Bandwidth $h = \hat{h}_{\mathrm{ROT}} \times \hat{s}_v$ (No Undersmoothing Factor)							
9	500	No	0.351	0.968	0.470	−1.799	1.792
9	500	Yes	0.351	0.944	0.360	−0.908	0.907
9	1000	No	0.306	0.977	0.211	−1.801	1.796
9	1000	Yes	0.306	0.956	0.138	−0.885	0.884
10	500	No	0.283	0.959	0.520	−1.799	1.792
10	500	Yes	0.283	0.931	0.409	−0.839	0.839
10	1000	No	0.247	0.977	0.285	−1.801	1.796
10	1000	Yes	0.247	0.959	0.188	−0.828	0.828
11	500	No	0.333	0.955	0.316	−1.799	1.792
11	500	Yes	0.333	0.939	0.261	−1.799	1.792
11	1000	No	0.290	0.960	0.118	−1.801	1.796
11	1000	Yes	0.290	0.943	0.079	−1.801	1.796
12	500	No	0.287	0.956	0.376	−1.799	1.792
12	500	Yes	0.287	0.944	0.295	−1.799	1.792
12	1000	No	0.250	0.960	0.154	−1.801	1.796
12	1000	Yes	0.250	0.948	0.111	−1.801	1.796

[a] DGP9–DGP12 correspond to DGP1–DGP4 in Andrews and Shi (2013, Section 10.3). The last two columns report the average values of the minimum and maximum of the arg max set. The estimated set is allowed to be disconnected and so the interval between the minimum and maximum of the arg max set is just an outer set for the estimated arg max set.

and in a handful of cases (e.g., DGP10 with kernel estimation), can even perform favorably. The main conclusions we draw from the full range of Monte Carlo experiments are that our inference method generally performs well both in coverage probabilities, and false coverage probabilities, and that in terms of a comparison between our approach and that of AS, each has their relative advantages and neither approach dominates.

REFERENCES

ANDREWS, D. W. K., AND X. SHI (2013): "Inference for Parameters Defined by Conditional Moment Inequalities," *Econometrica* (forthcoming). [31,45,46,48]

ARMSTRONG, T. B. (2011): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," Working Paper, Stanford University. [31]

BELLONI, A., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2011): "Conditional Quantile Processes Under Increasing Dimensions," Working Paper CWP19/11, CeMMAP. [27]

BELLONI, A., V. CHERNOZHUKOV, AND K. KATO (2010): "(Some) New Asymptotic Theory for Linear Series Estimators," Working Paper, MIT. [28]

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models (Chapter 76)," in *Handbook of Econometrics*,Vol. 6. Amsterdam: Elsevier, 5549–5632. [28]

CHETVERIKOV, D. (2011): "Adaptive Rate Optimal Test of Conditional Moment Inequalities," Working Paper, MIT. [27]

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall. [9,31]

GHOSAL, S., A. SEN, AND A. W. VAN DER VAART (2000): "Testing Monotonicity of Regression," *The Annals of Statistics*, 28, 1054–1082. [22]

HOROWITZ, J. L., AND E. MAMMEN (2004): "Nonparametric Estimation of an Additive Model With a Link Function," *The Annals of Statistics*, 32 (6), 2412–2443. [28]

KONG, E., O. LINTON, AND Y. XIA (2010): "Uniform Bahadur Representation for Local Polynomial Estimates of m-Regression and Its Application to the Additive Model," *Econometric Theory*, 26, 1529–1564. [8,9,11,30]

MANSKI, C. F., AND J. V. PEPPER (2009): "More on Monotone Instrumental Variables," *Econometrics Journal*, 12 (s1), S200–S216. [35,36]

MASRY, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571–599. [9,11,31]

MASSART, P. (1989): "Strong Approximation for Multivariate Empirical and Related Processes, via KMT Constructions," *The Annals of Probability*, 17 (1), 266–291. [9]

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [28,29]

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer. [21]

RIO, E. (1994): "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, 98, 21–45. [9,22]

RUDELSON, M. (1999): "Random Vectors in the Isotropic Position," *Journal of Functional Analysis*, 164 (1), 60–72. [26]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer-Verlag. [27]

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.; vchern@mit.edu; http://www.mit.edu/~vchern/,

Dept. of Economics, Seoul National University, Seoul 151-742, Korea; sokbae@gmail.com; https://sites.google.com/site/sokbae/, *and*

Centre for Microdata Methods and Practice, Institute for Fiscal Studies, and Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; adam.rosen@ucl.ac.uk; http://www.homepages.ucl.ac.uk/~uctparo/.

Manuscript received July, 2009; final revision received September, 2012.