Online Supplementary Material on "Inference in Nonparametric Series Estimation with Specification Searches for the Number of Series Terms"

Byunghoon Kang*
Department of Economics, Lancaster University

February 21, 2020

Abstract

This Online Supplementary Material contains the proofs of Lemmas in the main paper. Supplementary material also includes additional simulation results.

Section A contains useful lemmas used in the main paper. Section B contains the proofs of the main supporting lemmas used in the proofs of Theorems 3.1 and 4.1. Section C reports additional simulation results in addition to the model considered in the main paper.

Appendix A Preliminaries and Useful Lemmas

We use standard notations for the empirical process theory used in the proof of Theorem 4.1 in the main paper. Given measurable space $(S, \mathcal{S})$, let $\mathcal{F}$ as a class of measurable functions $f : \mathcal{S} \to \mathbb{R}$. For any probability measure Q on $(S, \mathcal{S})$, we define $N(\epsilon, \mathcal{F}, L_2(Q))$ as covering numbers, which is the minimal number of the $L_2(Q)$ balls of radius ϵ to cover $\mathcal{F}$ with $L_2(Q)$ norms $||f||_{Q,2} = (\int |f|^2 dQ)^{1/2}$. The uniform entropy numbers relative to the $L_2(Q)$ norms are defined as $\sup_Q \log N(\epsilon||F||_{Q,2}, \mathcal{F}, L_2(Q))$ where the supremum is over all discrete probability measures with an envelope function F. For $\alpha > 0$, we define $||X_i||_{\psi_\alpha} = \inf\{C > 0 : E[\psi_\alpha(|X_i|/C)] \leq 1\}$ with $\psi_\alpha(x) = \exp(x^\alpha) - 1$. For $\alpha \in [1, \infty)$, $||\cdot||_{\psi_\alpha}$ is an Orlicz norm, but for $\alpha \in (0, 1)$, $||\cdot||_{\psi_\alpha}$ is a quasi-norm. We define $\mathcal{F}$ as a *VC type* with envelope F if there are constants $A, v > 0$ such that

$$\sup_Q N(\epsilon||F||_{Q,2}, \mathcal{F}, L_2(Q)) \leq (A/\epsilon)^v$$

for all $0 < \epsilon \leq 1$. For notational convenience, we avoid discussing nonmeasurability issues and outer expectations, see van der Vaart and Wellner (1996) for the related issues. Throughout the proofs, we denote $c, C > 0$ as universal constants that do not depend on n.

*Email: b.kang1@lancaster.ac.uk, Homepage: https://sites.google.com/site/davidbhkang

To provide bounds in Lemma 1 and 2 in the main paper (Lemma B.2 and B.3), we first introduce matrix Bernstein inequality from Theorem 6.1.1 in Tropp (2015).

Lemma A.1. *Consider a finite sequence $\{S_i\}$ of independent, random matrices with common dimension $d_1 \times d_2$. Assume that $ES_i = 0$, $||S_i|| \leq L$ for each i. Let $Z = \sum_i S_i$, and define $v(Z) = \max\{||E(ZZ')||, ||E(Z'Z)||\}$. Then,*

$$P(||Z|| \geq t) \leq (d_1 + d_2)\exp(\frac{-t^2/2}{v(Z)Lt/3}), \qquad \forall t \geq 0,$$
$$E||Z|| \leq \sqrt{2v(Z)\log(d_1+d_2)} + \frac{1}{3}L\log(d_1+d_2).$$

The following coupling inequality is used in the proof of Theorem 3.1 in the main paper. See Theorem 4.1 and Corollary 4.1 in Chernozhukov, Chetverikov, and Kato (2014a).

Lemma A.2. *Let $X_1, ..., X_n$ be independent random vectors in $\mathbb{R}^p$ with mean zero and finite absolute third moments, that is, $E[X_{ij}] = 0, E[|X_{ij}|^3] < \infty$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$. Consider the statistic $Z = \max_{1\leq j \leq p}\sum_{i=1}^n X_{ij}$. Let $Y_1, ..., Y_n$ be independent random vectors in $\mathbb{R}^p$ with $Y_i \sim N(0, E[X_iX_i']), 1 \leq i \leq n$. Then for every $\delta > 0$, there exists a random variable $\widetilde{Z} = \max_{1\leq j\leq p}\sum_{i=1}^n Y_{ij}$ such that*

$$P(|Z - \widetilde{Z}| > 16\delta) \lesssim \delta^{-2}\{D_1 + \delta^{-1}(D_2 + D_3)\log(p \vee n)\}\log(p\vee n) + n^{-1}\log n$$

where

$$D_1 = E\big[\max_{1\leq j,l\leq p}|\sum_{i=1}^n (X_{ij}X_{il} - E[X_{ij}X_{il}])|\big], \quad D_2 = E\big[\max_{1\leq j\leq p}\sum_{i=1}^n |X_{ij}|^3\big],$$
$$D_3 = \sum_{i=1}^n E\big[\max_{1\leq j\leq p}|X_{ij}|^3 1\big(\max_{1\leq j\leq p}|X_{ij}| > \delta/\log(p\vee n)\big)\big].$$

The following maximal inequalities are used in the proof of Theorem 3.1 and 4.1 in the main paper. See also Lemmas 1, 8 and 9 in Chernozhukov, Chetverikov, and Kato (2015).

Lemma A.3. *Let $X_1, ..., X_n$ be independent centered random vectors in $\mathbb{R}^p$ with $p \geq 2$. Then, there exists a universal constant $C > 0$ such that*

$$E\Big[\max_{1\leq j,k\leq p}|\frac{1}{n}\sum_{i=1}^n (X_{ij}X_{ik} - E[X_{ij}X_{ik}])|\Big]$$
$$\leq C\Big[\sqrt{\frac{\log p}{n}}\max_{1\leq j\leq p}(\frac{1}{n}\sum_{i=1}^n E[X_{ij}^4])^{1/2} + \frac{\log p}{n}(E[\max_{1\leq i\leq n}\max_{1\leq j\leq p}X_{ij}^4])^{1/2}\Big].$$

Lemma A.4. *Let $X_1, ..., X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$. Define $M \equiv$*

$\max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|$ *and* $\sigma^2 \equiv \max_{1\le j\le p}\sum_{i=1}^n E[X_{ij}^2]$. *Then,*

$$E\Big[\max_{1\le j\le p}|\sum_{i=1}^n (X_{ij}-E[X_{ij}])|\Big] \lesssim (\sigma\sqrt{\log p}+\sqrt{E[M^2]}\log p).$$

Lemma A.5. *Let* $X_1, ..., X_n$ *be independent random vectors in* $\mathbb{R}^p$ *with* $p \ge 2$ *such that* $X_{ij} \ge 0$ *for all* $1 \le i \le n$ *and* $1 \le j \le p$. *Then,*

$$E\Big[\max_{1\le j\le p}\sum_{i=1}^n X_{ij}\Big] \lesssim \max_{1\le j\le p} E[\sum_{i=1}^n X_{ij}] + E[\max_{1\le i\le n}\max_{1\le j\le p} X_{ij}]\log p.$$

The following inequalities are Lemma C.1 in Chernozhukov, Chetverikov, and Kato (2017) and the Gaussian deviation inequality in Lemma 7 in Chernozhukov, Chetverikov, and Kato (2015).

Lemma A.6. *Let* X *be a nonnegative random variable such that* $P(X > x) \le A\exp(-x/B)$ *for all* $x \ge 0$ *and for some constants* $A, B > 0$. *Then for every* $t > 0$,

$$E[X^3 1(X > t)] \le 6A(t+B)^3\exp(-t/B).$$

Lemma A.7. *Let* $(Y_1, ..., Y_p)'$ *be centered Gaussian random vectors in* $\mathbb{R}^p$ *with* $\max_{1\le j\le p} E[Y_j^2] \le \sigma^2$ *for some* $\sigma^2 > 0$. *Then for every* $r > 0$,

$$P(\max_{1\le j\le p} Y_j \ge E[\max_{1\le j\le p} Y_j] + r) \le e^{-r^2/(2\sigma^2)}.$$

The following is the anti-concentration inequality of the maximum of Gaussian random vectors from Theorem 3 in Chernozhukov, Chetverikov, and Kato (2015).

Lemma A.8. *Let* $(Y_1, ..., Y_p)'$ *be centered Gaussian random vectors in* $\mathbb{R}^p$ *with* $\sigma_j^2 \equiv E[Y_j^2] > 0$ *for all* $1 \le j \le p$. *Let* $\underline{\sigma} \equiv \min_{1\le j\le p}\sigma_j, \overline{\sigma} \equiv \max_{1\le j\le p}\sigma_j$, *and* $a_p \equiv E[\max_{1\le j\le p}(Y_j/\sigma_j)]$. *(i) If the variances are all equal (i.e.,* $\underline{\sigma} = \overline{\sigma} = \sigma$*), then for every* $\epsilon > 0$,

$$\sup_{x\in\mathbb{R}} P(|\max_{1\le j\le p} Y_j - x| \le \epsilon) \le 4\epsilon(a_p+1)/\sigma.$$

(ii) If the variances are not equal ($\underline{\sigma} < \overline{\sigma}$*), then for every* $\epsilon > 0$,

$$\sup_{x\in\mathbb{R}} P(|\max_{1\le j\le p} Y_j - x| \le \epsilon) \le C\epsilon\{a_p + \sqrt{1\vee\log(\underline{\sigma}/\epsilon)}\},$$

where $C > 0$ *depends only on* $\underline{\sigma}$ *and* $\overline{\sigma}$.

The following lemmas are coupling inequalities for the supremum of the empirical process and the multiplier bootstrap process in Theorems 2.1 and 2.2 of Chernozhukov, Chetverikov, and Kato (2016).

Lemma A.9. *For a class of measurable functions $\mathcal{F}$, let $B : \mathcal{F} \to \mathbb{R}$ be a given functional, and for $\eta > 0$, let $N_B(\eta)$ be the minimal integer N such that there exists $f_1, \cdots, f_N \in \mathcal{F}$ with the property that for every $f \in \mathcal{F}$, there exists $1 \le j \le N$ with $|B(f) - B(f_j)| < \eta$. Suppose that the following assumptions hold; (a) there exists a countable subset $\mathcal{G}$ of $\mathcal{F}$ such that for any $f \in \mathcal{F}$, there exists a sequence $g_m \in \mathcal{G}$ with $g_m \to f$ pointwise and $B(g_m) \to B(f)$; (b) $\mathcal{F}$ is VC type with a measurable envelope F and constants $A \ge e$ and $v \ge 1$; (c) there exist constants $b \ge \sigma > 0, q \in [4, \infty)$ such that $\sup_{f\in\mathcal{F}} E[|f(X)|^k] \le \sigma^2 b^{k-2}$ for $k = 2, 3, 4$, and $||F||_{P,q} \le b$. Suppose that $K_n^3 \le n$ with $K_n = \log N_B(\eta) + v(\log n \vee \log(Ab/\sigma))$, and let $Z = \sup_{f\in\mathcal{F}}(B(f) + n^{-1/2}\sum_{i=1}^n (f(X_i) - E[f(X_i)]))$. Then for every $\gamma \in (0,1)$, there exists a random variable $\widetilde{Z} \stackrel{d}{=} \sup_{f\in\mathcal{F}}(B(f) + G_P f)$ with a centered Gaussian process G_P indexed by $\mathcal{F}$ with covariance function $E[G_P(f)G_P(g)] = Cov(f(X), g(X))$, $f, g \in \mathcal{F}$ such that*

$$P\big(|Z - \widetilde{Z}| > C_1(\eta + \delta_n)\big) \le C_2(\gamma + n^{-1})$$

where C_1, C_2 are positive constants that depend only on q, and

$$\delta_n = \frac{bK_n}{\gamma^{1/q} n^{1/2-1/q}} + \frac{(b\sigma^2 K_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}}.$$

Lemma A.10. *Suppose assumptions (a)-(c) in Lemma A.9 are satisfied and in addition suppose that $K_n \le n$. Let $Z^e = \sup_{f\in\mathcal{F}}(B(f) + n^{-1/2}\sum_{i=1}^n e_i(f(X_i) - E[f(X_i)]))$ where $e_1, ..., e_n$ are independent standard Gaussian random variables independent of $X = \{X_1, ..., X_n\}$. Then for every $\gamma \in (0,1)$, there exists a random variable $\widetilde{Z}^e \stackrel{d|X}{=} \sup_{f\in\mathcal{F}}(B(f) + G_P f)$ with a centered Gaussian process G_P defined in Lemma A.9 such that*

$$P\big(|Z^e - \widetilde{Z}^e| > C_1(\eta + \delta_n)\big) \le C_2(\gamma + n^{-1})$$

where C_1, C_2 are positive constants that depend only on q, and

$$\delta_n = \frac{bK_n}{\gamma^{1+1/q} n^{1/2-1/q}} + \frac{(b\sigma^2 K_n^{3/2})^{1/2}}{\gamma^{1+1/q} n^{1/4}}.$$

Lemma A.11. *Let V, W be real-valued random variables such that $P(|V - W| > r_1) \le r_2$ for some constants $r_1, r_2 > 0$. Then we have*

$$\sup_{t\in\mathbb{R}} |P(V \le t) - P(W \le t)| \le \sup_{t\in\mathbb{R}} P(|W - t| \le r_1) + r_2.$$

The following is the maximal inequality for uniformly bounded classes of functions derived in Giné and Koltchinskii (2006) and Chernozhukov, Chetverikov, and Kato (2014a). See also Proposition 6.1 in Belloni, Chernozhukov, Chetverikov, and Kato (2015).

Lemma A.12. *Let $X_1, ..., X_n$ be i.i.d random variables taking values in a measurable space $(S, \mathcal{S})$ with common distribution P, defined on the underlying n-fold product probability space. Let $\mathcal{F}$ be a suitably measurable class of functions mapping S to $\mathbb{R}$ with a measurable envelope F. Let σ^2 be* *a constant such that* $\sup_{f\in\mathcal{F}} var(f) \le \sigma^2 \le ||F||_{P,2}^2$*. Suppose that there exist constants* $A > e^2$ *and* $V \ge 2$ *such that* $\sup_Q \log N(\epsilon||F||_{Q,2}, \mathcal{F}, L_2(Q)) \le (A/\epsilon)^V$ *for all* $0 < \epsilon \le 1$*. Then,*

$$E\big[\sup_{f\in\mathcal{F}} |\sum_{i=1}^n f(X_i) - E[f(X_i)]|\big] \lesssim C\big[\sqrt{n\sigma^2 V \log\frac{A||F||_{P,2}}{\sigma}} + V||F||_\infty \log\frac{A||F||_{P,2}}{\sigma}\big].$$

Lemma A.13. *Let* $(\varepsilon_1, X_1), ..., (\varepsilon_n, X_n)$ *be i.i.d. random vectors, defined on an underlying n-fold product probability space in* $\mathbb{R}^{d+1}$ *with* $E[\varepsilon_i|X_i] = 0$ *and* $\sigma^2 = \sup_{x\in\mathcal{X}} E[\varepsilon_i^2|X_i] < \infty$ *where* $\mathcal{X}$ *denotes the support of* X_1*. Let* $\mathcal{F}$ *be a class of functions on* $\mathbb{R}^d$ *such that* $E[f(X_1)^2] = 1$ *and* $||f||_\infty \le b$ *for all* $f \in \mathcal{F}$*. Let* $\mathcal{G} = \{(\varepsilon, x) \mapsto \varepsilon f(x) : f \in \mathcal{F}\}$*. Suppose that there exist constants* $A > e^2$ *and* $V \ge 2$ *such that* $\sup_Q N(\epsilon||G||_{Q,2}, \mathcal{G}, L_2(Q)) \le (A/\epsilon)^V$ *for all* $0 < \epsilon \le 1$ *for the envelope* $G(\varepsilon, x) = |\varepsilon| b$*. If for some* $q > 2, E[|\varepsilon_1|^q] < \infty$*, then*

$$E\big[\sup_{f\in\mathcal{F}} |\sum_{i=1}^n \varepsilon_i f(X_i)|\big] \lesssim C\big[(\sigma + \sqrt{E[|\varepsilon_1|^q]})\sqrt{nV\log(Ab)} + Vb^{q/(q-2)}\log(Ab)\big].$$

The following lemma is the anti-concentration for the Separable Gaussian process, which can be found in Theorem 2.1 in Chernozhukov, Chetverikov, and Kato (2014b). See also Lemma B.1 in Chernozhukov, Chetverikov, and Kato (2014b).

Lemma A.14. *Let* $Y = \{Y(t) : t \in T\}$ *be a separable Gaussian process indexed by a semimetric space* T *such that* $E[Y(t)] = 0$ *and* $E[Y(t)^2] = 1$ *for all* $t \in T$*. Assume that* $\sup_{t\in T} |Y(t)| < \infty a.s.$ *Then* $E[\sup_{t\in T} Y(t)] < \infty$ *and*

$$\sup_{x\in\mathbb{R}} P\big(\big|\sup_{t\in T} Y(t) - x\big| \le \epsilon\big) \le 4\epsilon(E[\sup_{t\in T} Y(t)] + 1)$$

for all $\epsilon \ge 0$.

Lemma A.15. *Let* $Y = \{Y(t) : t \in T\}$ *be a separable, centered Gaussian process such that* $E[Y(t)^2] \le 1$ *for all* $t \in T$*. Let* $c(\alpha)$ *denote the* $(1-\alpha)-$*quantile of* $\sup_{t\in T} |Y(t)|$ *and assume that* $E[\sup_{t\in T} |Y(t)|] < \infty$*. Then* $c(\alpha) \le E[\sup_{t\in T} |Y(t)|] + \sqrt{2|\log\alpha|}$ *and* $c(\alpha) \le M(\sup_{t\in T} |Y(t)|) + \sqrt{2|\log\alpha|}$ *where* $M(\sup_{t\in T} |Y(t)|)$ *is the median of* $\sup_{t\in T} |Y(t)|$*.*

Appendix B Supporting Lemmas

We first recall notations used in the main paper for Theorems 3.1 and 4.1. Let the data $z_i = (\varepsilon_i, x_i)$ be i.i.d. random vectors defined on the probability space $(\mathcal{Z} = \mathcal{E} \times \mathcal{X}, \mathcal{A}, P)$ with common probability distribution $P \equiv P_{\varepsilon,x}$. We think of $(\varepsilon_1, x_1), \cdots (\varepsilon_n, x_n)$ as the coordinates of the infinite product probability space. For any sequence $\{K = K_n : n \ge 1\} \in \prod_{n=1}^\infty \mathcal{K}_n$ under Assumption 2.1, define the orthonormalized vector of basis functions

$$\tilde{P}(K, x) \equiv Q_K^{-1/2} P(K, x) = E[P_{Ki}P_{Ki}']^{-1/2} P(K, x),\ \tilde{P}_{Ki} = \tilde{P}(K, x_i),\ \tilde{P}^K = [\tilde{P}_{K1}, \cdots, \tilde{P}_{Kn}]',$$

and observe that

$$\widehat{g}_n(K,x) = \tilde{P}(K,x)'(\tilde{P}^{K\prime}\tilde{P}^K)^{-1}\tilde{P}^{K\prime}Y, \quad V_n(K,x) = \tilde{P}(K,x)'\tilde{\Omega}_K\tilde{P}(K,x), \quad \tilde{\Omega}_K = E(\tilde{P}_{Ki}\tilde{P}'_{Ki}\varepsilon_i^2).$$

We define pseudo true value β_K such that $y_i = \tilde{P}'_{Ki}\beta_K + \varepsilon_{Ki}, E[\tilde{P}_{Ki}\varepsilon_{Ki}] = 0$ where $\varepsilon_{Ki} = r_{Ki} + \varepsilon_i$, $r_n(K,x) = g_0(x) - \tilde{P}(K,x)'\beta_K, r_{Ki} = r_n(K,x_i)$, and $r_K \equiv (r_{K1}, \cdots r_{Kn})'$. We also define $\widehat{Q}_K \equiv \frac{1}{n}\tilde{P}^{K\prime}\tilde{P}^K$, $\underline{\sigma}^2 \equiv \inf_x E[\varepsilon_i^2|x_i = x], \bar{\sigma}^2 \equiv \sup_x E[\varepsilon_i^2|x_i = x]$.

We first provide the coupling inequalities used in the proof of Theorem 3.1 in the main paper.

Lemma B.1. *Suppose that Assumptions 2.1, 3.1, and 3.2 hold. Let $t_n(K,x) = n^{-1/2}\sum_{i=1}^n \tilde{P}(K,x)'\tilde{P}_{Ki}\varepsilon_i/V_n(K,x)^{1/2}$ and let $Z_i = (Z_{i1},...,Z_{ip})' \sim N(0, \frac{1}{n}\Sigma_n)$ be a $p \times 1$ Gaussian random vector provided Σ_n exists and is a finite positive definite matrix with (j,l) elements defined as $\Sigma_n(j,l) = E[t_n(K_j,x)t_n(K_l,x))]$, $p = |\mathcal{K}_n|$. Then, there exists a sequence of random variables $\max_{1\le j\le p}\sum_{i=1}^n |Z_{ij}|$ such that*

$$P(|\max_{1\le j\le p}|t_n(K_j,x)| - \max_{1\le j\le p}\sum_{i=1}^n |Z_{ij}|| > 16\delta) \lesssim \frac{\log(p\vee n)}{\delta^2}D_1 + \frac{\log^2(p\vee n)}{\delta^3 n^{3/2}}(D_2 + D_3) + \frac{\log n}{n}$$

where under the case (a) in Assumption 3.2 (ii), we have

$$D_1 \lesssim \sqrt{\frac{(\max_K \zeta_K)^2 \log p}{n}} + \frac{(\max_K \zeta_K)^2 \log p}{n^{1-2/q}}, \quad D_2 \lesssim n\max_K \zeta_K + (\max_K \zeta_K)^3 n^{3/q}\log p,$$
$$D_3 \lesssim \frac{(\max_K \zeta_K)^q \log^{q-3}(p\vee n)}{n^{q/2-5/2}\delta^{q-3}}$$

while under the case (b), we have

$$D_1 \lesssim \sqrt{\frac{(\max_K \zeta_K)^2 \log p}{n}} + \frac{(\max_K \zeta_K)^2 \log^2(pn)\log p}{n}, \quad D_2 \lesssim n\max_K \zeta_K + (\max_K \zeta_K)^3 \log^3(pn)\log p,$$
$$D_3 \lesssim 12n(\delta\sqrt{n}/\log(p\vee n) + C\max_K \zeta_K \log p)^3 \exp(-\frac{\delta\sqrt{n}}{C\max_K \zeta_K \log p \log(p\vee n)}).$$

Proof. First consider

$$t_n \equiv (t_n(K_1,x), \cdots, t_n(K_p,x))' = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i$$

where $\xi_i = (\xi_{i1}, \xi_{i2}, \cdots, \xi_{ip})' \in \mathbb{R}^p$ with $\xi_{ij} = \frac{\tilde{P}(K_j,x)'\tilde{P}_{K_j i}\varepsilon_i}{V_n(K_j,x)^{1/2}}$. Applying Lemma A.2 to the p dimensional random vectors $\{\frac{1}{\sqrt{n}}\xi_i\}_{i=1}^n$, for any $\delta > 0$, there exists a random variable $\max_{1\le j\le p}\sum_{i=1}^n Z_{ij}$ with independent random vectors $\{Z_i\}_{i=1}^n \in \mathbb{R}^p$, $Z_i \sim N(0, \frac{1}{n}E[\xi_i\xi_i']), 1 \le i \le n$, such that

$$P(|\max_{1\le j\le p}|t_n(K_j,x)| - \max_{1\le j\le p}\sum_{i=1}^n |Z_{ij}|| > 16\delta) \lesssim \frac{\log(p\vee n)}{\delta^2}D_1 + \frac{\log^2(p\vee n)}{\delta^3 n^{3/2}}(D_2 + D_3) + \frac{\log n}{n}$$

where

$$D_1 = E\big[\max_{1\le j,l\le p}|\frac{1}{n}\sum_{i=1}^n(\xi_{ij}\xi_{il} - E[\xi_{ij}\xi_{il}])|\big], \quad D_2 = E\big[\max_{1\le j\le p}\sum_{i=1}^n|\xi_{ij}|^3\big],$$
$$D_3 = \sum_{i=1}^n E\big[\max_{1\le j\le p}|\xi_{ij}|^3 1\big(\max_{1\le j\le p}|\xi_{ij}| > \delta\sqrt{n}/\log(p\vee n)\big)\big].$$

Next, we consider D_1, D_2, D_3 in either case (a) or (b).

(Case (a)) By Lemma A.3, we have

$$D_1 \lesssim \sqrt{\frac{\log p}{n}}\max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n E[\xi_{ij}^4])^{1/2} + \frac{\log p}{n}(E[\max_{1\le i\le n}\max_{1\le j\le p}\xi_{ij}^4])^{1/2}.$$

Also, $E[\max_{1\le i\le n}|\varepsilon_i|^4|X=x] \lesssim n^{4/q}$ where $X=(x_1,\cdots,x_n)'$ by Assumption $\sup_x E[|\varepsilon_i|^q|x_i = x] \lesssim 1$. Hence,

$$E[\max_{1\le i\le n}\max_{1\le j\le p}\xi_{ij}^4] \lesssim (\max_K \zeta_K)^4 n^{4/q}$$

since $\max_{1\le i\le n}\max_{1\le j\le p}|\frac{\tilde{P}(K_j,x)'\tilde{P}_{K_j i}}{V_n(K_j,x)^{1/2}}| \le \max_K \zeta_K$. Next,

$$\max_{1\le j\le p}(\frac{1}{n}\sum_{i=1}^n E[\xi_{ij}^4]) = \max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n E[|\frac{\tilde{P}(K_j,x)'\tilde{P}_{K_j i}}{V_n(K_j,x)^{1/2}}|^4 E[|\varepsilon_i|^4|x_i = x]] \lesssim (\max_K \zeta_K)^2$$

since $n^{-1}\sum_{i=1}^n E[|\tilde{P}(K_j,x)'\tilde{P}_{K_j i}/V_n(K_j,x)^{1/2}|^2] \lesssim 1$, for all $1\le j\le p$, and by Assumption $E[|\varepsilon_i|^4|x_i = x]] \lesssim 1$. Thus,

$$D_1 \lesssim \sqrt{\frac{(\max_K \zeta_K)^2\log p}{n}} + \frac{(\max_K \zeta_K)^2\log p}{n^{1-2/q}}.$$

Similarly, we use Lemma A.5 to bound D_2,

$$D_2 \lesssim \max_{1\le j\le p}E[\sum_{i=1}^n|\xi_{ij}|^3] + E[\max_{1\le i\le n}\max_{1\le j\le p}|\xi_{ij}|^3]\log p \lesssim n\max_K \zeta_K + (\max_K \zeta_K)^3 n^{3/q}\log p.$$

Note that for any real-valued random variable Z and any $t>0$, we have $E[|Z|^3 1(|Z|>t)] \le E[|Z|^3(|Z|/t)^{q-3}1(|Z|>t)] \le t^{3-q}E[|Z|^q]$. Thus,

$$\begin{aligned}&E\big[\max_{1\le j\le p}|\xi_{ij}|^3 1\big(\max_{1\le j\le p}|\xi_{ij}| > \delta\sqrt{n}/\log(p\vee n)\big)\big]\\ &\qquad \le \frac{\log^{q-3}(p\vee n)}{n^{q/2-3/2}\delta^{q-3}}E[\max_{1\le j\le p}|\xi_{ij}|^q] \le \frac{(\max_K \zeta_K)^q\log^{q-3}(p\vee n)}{n^{q/2-3/2}\delta^{q-3}}.\end{aligned}$$

(case (b)) By Lemma A.4, we have

$$D_1 \lesssim n^{-1}(\sigma\sqrt{\log p} + \sqrt{E[M^2]}\log p)$$

where $\sigma^2 = \max_{1\le j,l\le p}\sum_{i=1}^n E[(\xi_{ij}\xi_{il} - E[\xi_{ij}\xi_{il}])^2]$, and $M = \max_{1\le i\le n}\max_{1\le j,l\le p}|\xi_{ij}\xi_{il} - E[\xi_{ij}\xi_{il}]|$. By Hölder's inequality and Assumption $E[\varepsilon_i^2|X_i = x] < \infty$,

$$\sigma^2 \le \max_{1\le j,l\le p}\sum_{i=1}^n E[|\xi_{ij}\xi_{il}|^2] \le n(\max_K \zeta_K)^2.$$

Observe that there exists a constant $C > 0$,

$$\begin{aligned}
||\max_{1\le i\le n}\max_{1\le j,l\le p}|\xi_{ij}\xi_{il} - E[\xi_{ij}\xi_{il}]|||_{\psi_{1/2}} &\le C(||\max_{1\le i\le n}\max_{1\le j,l\le p}|\xi_{ij}\xi_{il}|||_{\psi_{1/2}} + \max_{1\le i\le n}\max_{1\le j,l\le p}E[|\xi_{ij}\xi_{il}|]) \\
&\le C(||\max_{1\le i\le n}\max_{1\le j\le p}|\xi_{ij}|\ ||_{\psi_1}^2 + \max_{1\le i\le n}\max_{1\le j,l\le p}E[|\xi_{ij}\xi_{il}|])
\end{aligned}$$

because $||\cdot||_{\psi_{1/2}}$ is a quasi-norm and $||\max_{1\le i\le n,1\le j,l\le p}|\xi_{ij}\xi_{il}|||_{\psi_{1/2}} = ||\max_{1\le i\le n,1\le j\le p}|\xi_{ij}|^2||_{\psi_{1/2}}$. By Lemma 2.2.2 in van der Vaart and Wellner (1996), we have

$$||\max_{1\le i\le n}\max_{1\le j\le p}|\xi_{ij}|\ ||_{\psi_1} \lesssim \log(pn)\max_{1\le i\le n}\max_{1\le j\le p}||\xi_{ij}||_{\psi_1} \lesssim (\max_K \zeta_K)\log(pn)$$

since $|\frac{\tilde{P}(K_j,x)'\tilde{P}_{K_j i}}{V_n(K_j,x)^{1/2}}| \le \max_K \zeta_K$ for all $1\le i\le n, 1\le j\le p$, and Assumption $\sup_x E[\exp(|\varepsilon_i|/C)|X_i = x] \le 2$. Using the inequalities for L_p and the Orlicz norm, $E|X|^2 \le 2!||X||_{\psi_1}^2$ for a random variable X with $p \ge 1$, we have $\sqrt{E[M^2]} \lesssim (\max_K \zeta_K)^2\log^2(pn)$. Thus,

$$D_1 \lesssim \sqrt{\frac{(\max_K \zeta_K)^2\log p}{n}} + \frac{(\max_K \zeta_K)^2\log^2(pn)\log p}{n}.$$

Further, by Lemma A.5 and using similar calculations above gives

$$D_2 \lesssim n\max_K \zeta_K + E[\max_{1\le i\le n}\max_{1\le j\le p}|\xi_{ij}|^3]\log p \lesssim n\max_K \zeta_K + (\max_K \zeta_K)^3\log^3(pn)\log p.$$

Using the Markov's inequality, for every $t > 0$,

$$\begin{aligned}
P\big(\max_j|\xi_{ij}| > t\big) &= P\big(\exp(\max_j|\xi_{ij}|/C\max_K \zeta_K\log p) > \exp(t/C\max_K \zeta_K\log p)\big) \\
&\le 2\exp\big(-\frac{t}{C\max_K \zeta_K\log p}\big)
\end{aligned}$$

since $||\max_{1\le j\le p}|\xi_{ij}|||_{\psi_1} \le C\log p\max_{1\le j\le p}|||\xi_{ij}|||_{\psi_1} \le C(\max_K \zeta_K)\log p$ for some constant $C > 0$. Combined with Lemma A.6 we have

$$\begin{aligned}
&E\big[\max_{1\le j\le p}|\xi_{ij}|^3 1\big(\max_{1\le j\le p}|\xi_{ij}| > \delta\sqrt{n}/\log(p\vee n)\big)\big] \\
&\le 12(\delta\sqrt{n}/\log(p\vee n) + C\max_K \zeta_K\log p)^3\exp(-\frac{\delta\sqrt{n}}{C\max_K \zeta_K\log p\log(p\vee n)}).
\end{aligned}$$

■

Next, we provide useful lemmas which will be used in the proof of Theorem 3.1 and 4.1. The versions of proofs of Lemma B.2 and B.3 with $\mathcal{K}_n = \{K\}$ are available in the literature, such as Belloni et al. (2015) and Chen and Christensen (2015), among many others. Note that different rate conditions of $K = K_n$ such as those in Newey (1997) can be used here, but lead to different bounds (B.1)-(B.2).

Lemma B.2. *Suppose that Assumptions 2.1, 3.1, and 3.2 hold, then* $||\widehat{Q}_K - I_K|| = O_p(\sqrt{\lambda_K^2 \zeta_K^2 \log K/n})$ *for any* $K \in \mathcal{K}_n$*, and the following holds*

$$\max_{K\in\mathcal{K}_n} |R_1(K,x)| = O_p(\max_{K\in\mathcal{K}_n} \sqrt{\frac{\lambda_K^2 \zeta_K^2 \log K \log p}{n}}(1 + \ell_K c_K \sqrt{K})), \tag{B.1}$$

$$\max_{K\in\mathcal{K}_n} |R_2(K,x)| = O_p(\max_{K\in\mathcal{K}_n} (\ell_K c_K)\sqrt{\log p}), \tag{B.2}$$

where $R_1(K,x) \equiv \sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'(\widehat{Q}_K^{-1} - I_K)\tilde{P}^{K'}(\varepsilon + r_K)$, $R_2(K,x) \equiv \sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\tilde{P}^{K'} r_K$.

Proof. We first define $S_i = \frac{1}{n}(\tilde{P}_{Ki}\tilde{P}_{Ki}' - E(\tilde{P}_{Ki}\tilde{P}_{Ki}'))$. Note that $\mathbb{E}\, S_i = 0$, $||S_i|| \le L = \frac{1}{n}(\lambda_K^2\zeta_K^2 + 1)$, and $v(Z) = \frac{1}{n}||E(\tilde{P}_{Ki}\tilde{P}_{Ki}'\tilde{P}_{Ki}\tilde{P}_{Ki}') - E(\tilde{P}_{Ki}\tilde{P}_{Ki}')E(\tilde{P}_{Ki}\tilde{P}_{Ki}')|| \le \frac{1}{n}(\lambda_K^2\zeta_K^2 + 1)$ by definition of λ_K, ζ_K and $E(\tilde{P}_{Ki}\tilde{P}_{Ki}') = I_K$. By Lemma A.1, we have

$$E||\widehat{Q}_K - I_K|| = E||\sum_i \frac{1}{n}(\tilde{P}_{Ki}\tilde{P}_{Ki}' - I_K)|| \le C(\sqrt{\lambda_K^2\zeta_K^2 \log(K)/n} + \lambda_K^2\zeta_K^2 \log(K)/n),$$

and $||\widehat{Q}_K - I_K|| = O_P(\sqrt{\lambda_K^2\zeta_K^2 \log(K)/n})$ by the Markov inequality. For (B.1), we first look at the terms $\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K'}\varepsilon$. For any $K \in \mathcal{K}_n$, conditional on the sample $X^n = [x_1, \cdots, x_n]$, this term has mean zero and variance,

$$\begin{aligned}
&\frac{1}{nV_n(K,x)}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K'}E(\varepsilon\varepsilon'|X)\tilde{P}^K\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}(K,x)\\
&\le \frac{\bar{\sigma}^2}{V_n(K,x)}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1} - I_K\right)\widehat{Q}_K\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}(K,x)\\
&\le \frac{\bar{\sigma}^2\tilde{P}_K(x)'\tilde{P}_K(x)}{V_n(K)}\lambda_{max}(\widehat{Q}_K^{-1})||\left(\widehat{Q}_K - I_K\right)||^2 = O_P(\lambda_K^2\zeta_K^2\log(K)/n)
\end{aligned}$$

where the first and the second inequality uses $V_n(K,x) \le \bar{\sigma}^2\tilde{P}(K,x)'\tilde{P}(K,x)$, $V_n(K,x) \ge \underline{\sigma}^2\tilde{P}(K,x)'\tilde{P}(K,x)$ by Assumption 3.2(ii), and $\lambda_{max}(\widehat{Q}_K^{-1}) = (\lambda_{max}(\widehat{Q}_K))^{-1} = O_p(1)$ since all eigenvalues of $\widehat{Q}_K$ are bounded away from zero as $|\lambda_{min}(\widehat{Q}_K) - 1| \le ||\widehat{Q}_K - I_K|| = o_p(1)$ and Assumption 3.1(ii) and rate conditions. By using $||\frac{\tilde{P}(K,x)}{V_n(K,x)^{1/2}}|| \asymp 1$ and $||\widehat{Q}_K^{-1}|| = O_p(1)$,

$$\begin{aligned}
|\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1} - I_K\right)\tilde{P}^{K'}\varepsilon| &\le C||\widehat{Q}_K^{-1}|| \cdot ||\left(\widehat{Q}_K - I_K\right)|| \cdot ||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}\varepsilon_i||\\
&= O_p(\sqrt{\frac{\lambda_K^2\zeta_K^2 K\log(K)}{n}})
\end{aligned}$$

where $||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}\varepsilon_i|| = O_p(\sqrt{K})$ since $E[||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}\varepsilon_i||^2] = E[\sum_{j=1}^K \tilde{P}_{Ki,j}^2\varepsilon_i^2] \lesssim E[||\tilde{P}_{Ki}||^2] = K$.

By applying Lemma A.4, we have

$$
\begin{aligned}
&E\big[\max_{K\in\mathcal{K}_n}|\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1}-I_K\right)\tilde{P}^{K\prime}\varepsilon|\big|X\big] \\
&\lesssim_P \max_{K\in\mathcal{K}_n}\sqrt{\frac{\lambda_K^2\zeta_K^2 K\log(K)}{n}}\frac{\log p}{\sqrt{n}} + \max_{K\in\mathcal{K}_n}\sqrt{\frac{\lambda_K^2\zeta_K^2\log(K)}{n}}\sqrt{\log p} \lesssim \max_{K\in\mathcal{K}_n}\sqrt{\frac{\lambda_K^2\zeta_K^2\log(K)}{n}}\sqrt{\log p}
\end{aligned}
$$

where the last inequality uses $\sqrt{\max_K K\log p/n} = o(1)$.

Next, consider the terms $\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1}-I_K\right)\tilde{P}^{K\prime}r_K$. Observe that $E[\tilde{P}_{Ki}r_{Ki}] = 0$, $||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}r_{Ki}|| = O_p(\ell_K c_K\sqrt{K})$ since

$$
E[||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}r_{Ki}||^2] = E[\sum_{j=1}^K \tilde{P}_{ji}^2 r_{Ki}^2] \leq \ell_K^2 c_K^2 E[||\tilde{P}_{Ki}||^2] = \ell_K^2 c_K^2 K. \tag{B.3}
$$

By using (B.3), $||\frac{\tilde{P}(K,x)}{V_n(K,x)^{1/2}}|| \asymp 1$ and $||\widehat{Q}_K^{-1}|| = O_p(1)$, we have

$$
\begin{aligned}
|\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1}-I_K\right)\tilde{P}^{K\prime}r_K| &\leq C||\widehat{Q}_K^{-1}||\cdot||\left(\widehat{Q}_K-I_K\right)||\cdot||\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{P}_{Ki}r_{Ki}|| \\
&= O_p(\sqrt{\frac{\lambda_K^2\zeta_K^2\log(K)}{n}}\ell_K c_K\sqrt{K}).
\end{aligned}
$$

Similarly,

$$
E\big[\max_{K\in\mathcal{K}_n}|\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\left(\widehat{Q}_K^{-1}-I_K\right)\tilde{P}^{K\prime}r_K|\big] \lesssim_P \max_{K\in\mathcal{K}_n}\sqrt{\frac{\lambda_K^2\zeta_K^2 K\log(K)}{n}}\ell_K c_K\sqrt{\log p}
$$

by Lemma A.4 and (B.1) follows by Chebyshev's inequality. Lastly, we consider (B.2).

$$
E[(\sqrt{\frac{1}{nV_n(K,x)}}\tilde{P}(K,x)'\tilde{P}^{K\prime}r_K)^2] = E[(\frac{\tilde{P}(K,x)'\tilde{P}_{Ki}}{V_n(K,x)^{1/2}}r_{Ki})^2] \leq (c_K\ell_K)^2
$$

since $E[(\frac{\tilde{P}(K,x)'\tilde{P}_{Ki}}{V_n(K,x)^{1/2}})^2] \asymp 1$ by Assumption 3.2(ii) and $E(r_{Ki})^2 \leq (\ell_K c_K)^2$ by Assumption 3.1(ii). Furthermore, $|\frac{\tilde{P}(K,x)'\tilde{P}_{Ki}r_{Ki}}{V_n(K,x)^{1/2}}| \leq \max_{K\in\mathcal{K}_n}(\ell_K c_K)\zeta_K$. Again, by Lemma A.4,

$$
E[\max_{K\in\mathcal{K}_n} R_2(K,x)] \lesssim \max_K(\ell_K c_K)\zeta_K\frac{\log p}{\sqrt{n}} + \max_K(\ell_K c_K)\sqrt{\log p} \lesssim \max_K(\ell_K c_K)\sqrt{\log p}
$$

where the last inequality uses $\sqrt{\max_K \zeta_K^2\log p/n} = o(1)$. Thus, (B.2) follows and this completes the proof. ■

Lemma B.3. *Suppose that Assumptions 2.1, 3.1 and 4.1 hold, then the following holds*

$$\sup_{K\in\mathcal{K}_n, x\in\mathcal{X}} |R_1(K,x)| = O_p(\max_{K\in\mathcal{K}_n} \sqrt{\frac{\lambda_K^2 \zeta_K^2 \log K \log n}{n}}(n^{1/q} + \ell_K c_K \sqrt{K})), \tag{B.4}$$

$$\sup_{K\in\mathcal{K}_n, x\in\mathcal{X}} |R_2(K,x)| = O_p(\max_{K\in\mathcal{K}_n} (\ell_K c_K)\sqrt{\log n}), \tag{B.5}$$

where $R_1(K,x), R_2(K,x)$ are defined in Lemma B.2.

Proof. The proof follows from the same arguments to those used in Lemma 4.2 in Belloni et al. (2015) using the maximal inequalities (Lemmas A.12 and A.13). To provide the bounds in (B.4) and (B.5), we use similar calculations on $|R_1(K,x)|, |R_2(K,x)|$ uniformly in $K\in\mathcal{K}_n$ in the proof of Lemma B.2, and use similar derivations as in the proof of Theorem 4.1 in the main paper such that the class of functions $\mathcal{F}_n = \{f_{n,K,x} : (K,x)\in\mathcal{K}_n\times\mathcal{X}\}$ is a VC type with the envelope function $F_n(\varepsilon,t) \equiv C|\varepsilon|\max_K \zeta_K \vee 1$, where

$$f_{n,K,x}(\varepsilon,t) = \frac{\tilde{P}(K,x)'\tilde{P}(K,t)\varepsilon}{V_n(K,x)^{1/2}}, (\varepsilon,t)\in\mathcal{E}\times\mathcal{X}. \tag{B.6}$$

for given $n\geq 1$, $K\in\mathcal{K}_n, x\in\mathcal{X}$ because $|f_{n,K,x} - f_{n,K',x'}| \leq |\varepsilon| A \max_K \zeta_K L_n(||x-x'|| + |K-K'|)$ for all $x,x'\in\mathcal{X}, K,K'\in\mathcal{K}_n$ where $L_n = \zeta^{L_1}\vee\zeta^{L_2}$ under Assumption 4.1. This completes the proof of the Lemma. ■

Appendix C Additional Simulations

Section C reports additional simulation results in addition to the model considered in the main paper. The main specification we consider is:

$$y_i = g(x_i) + \varepsilon_i,$$
$$x_i = \Phi(x_i^*), \begin{pmatrix} x_i^* \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2(x_i^*) \end{pmatrix}\right)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function needed to ensure compact support, and $\sigma^2(x_i^*) = (\frac{1+2x_i^*}{2})^2$. As in the main paper, we investigate the following three functions for $g(x)$: $g_1(x) = \ln(|6x-3|+1)sgn(x-1/2)$, $g_2(x) = \frac{\sin(7\pi x/2)}{1+2x^2(sgn(x)+1)}$, and $g_3(x) = x - 1/2 + 5\phi(10(x-1/2))$, where $\phi(\cdot)$ is the standard normal probability density function, and $sgn(\cdot)$ is the sign function. We generate 2000 simulation replications for each design.

We calculate a pointwise coverage rate (COV) and the average length (AL) of various 95% nominal CIs, as well as analogous uniform CBs for the grid points of x on the support $\mathcal{X} = [0.05, 0.95]$. To be specific, we consider (1) the standard CI with $\widehat{K}_{\texttt{cv}}\in\mathcal{K}_n$ selected to minimize leave-one-out cross-validation; (2) robust CI with $\widehat{K}_{\texttt{cv}}$ using the critical value $\widehat{c}_{1-\alpha}(x)$; (3) robust CI using $\widehat{K}_{\texttt{cv+}} = \widehat{K}_{\texttt{cv}} + 2$ and analogous uniform inference results. The critical values, $\widehat{c}_{1-\alpha}(x)$ and

$\widehat{c}_{1-\alpha}$ are constructed using Monte Carlo methods and weighted bootstrap methods, respectively. 1000 additional Monte Carlo or bootstrap replications are performed on each simulation iteration to calculate critical values.

In the main paper, we report the results for quadratic splines with evenly placed knots where the number of knots K are selected among $\mathcal{K}_n = [\underline{K}, \overline{K}]$ by setting $\underline{K} = 2n^{1/5}$ and $\overline{K} = 2n^{1/3}$ rounded up to the nearest integer with a sample size $n = 200$. Table 1 reports a homoskedastic error case. Specifically, we set $\sigma^2(x_i^*) = 1$. In Tables 2-3, we consider different sample sizes $n \in \{100, 500\}$. Table 4 reports results for polynomial regressions with $\underline{K} = n^{1/5}$ and $\overline{K} = n^{1/3}$. The simulation results are qualitatively similar to the results in the main paper, except for poor coverage property of the polynomial regressions at particular points.

Finally, Table 5 explores the following bivariate specification as in Cattaneo and Farrell (2013) with bivariate and non-normal regressors:

$$y_i = (1 - (4x_{1i} - 2)^2)^2(\sin(5x_{2i})/5) + \varepsilon_i,$$

where $\varepsilon_i \sim N(0,1)$, and x_{1i}, x_{2i} are independently distributed as $Beta(\alpha, \beta)$ distributions truncated to $[0.05, 0.95]$. We consider quadratic splines with the basis $\{1, x_1, x_1^2, x_2, x_2^2, x_1x_2, \max(x_1 - \tau_1, 0)^2, ..., \max(x_1 - \tau_K, 0)^2, \max(x_2 - \tau_1, 0)^2, ..., \max(x_2 - \tau_K, 0)^2\}$, and the total $2K + 6$ number of series terms with $K \in \mathcal{K}_n = [2n^{1/5}, 2n^{1/3}]$. We consider the following cases: (1) $\alpha = \beta = 1$ (uniform); (2) $\alpha = \beta = 1/2$ (mass at the boundary). For this bivariate specification, we generate 1000 simulated data sets with $n = 200$.

Table 1: Coverage and Length of Nominal 95% CIs and CBs - Splines (homoskedastic)

	Pointwise								Uniform	
	$x = 0.2$		$x = 0.5$		$x = 0.8$		$x = 0.9$			
	COV	AL	COV	AL	COV	AL	COV	AL	COV	AL
Model 1: $g_1(x) = \ln(\lvert 6x - 3\rvert + 1)sgn(x - 1/2)$										
Standard	0.90	0.70	0.90	0.65	0.91	0.70	0.93	0.85	0.41	0.74
Robust ($\widehat{K}_{\texttt{cv}}$)	0.95	0.86	0.96	0.83	0.96	0.85	0.96	0.98	0.96	1.39
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.97	1.05	0.97	0.93	0.97	1.05	0.96	1.07	0.97	1.55
Model 2: $g_2(x) = \sin(7\pi x/2)/[1 + 2x^2(sgn(x) + 1)]$										
Standard	0.89	0.71	0.90	0.66	0.91	0.71	0.93	0.85	0.38	0.74
Robust ($\widehat{K}_{\texttt{cv}}$)	0.94	0.87	0.96	0.84	0.96	0.86	0.96	0.99	0.96	1.39
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.97	1.05	0.96	0.94	0.97	1.05	0.96	1.07	0.97	1.55
Model 3: $g_3(x) = x - 1/2 + 5\phi(10(x - 1/2))$										
Standard	0.90	0.76	0.81	0.72	0.91	0.76	0.93	0.88	0.27	0.76
Robust ($\widehat{K}_{\texttt{cv}}$)	0.95	0.93	0.90	0.92	0.95	0.93	0.95	1.02	0.93	1.43
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.97	1.03	0.96	1.03	0.97	1.03	0.96	1.05	0.96	1.59

Notes: "Pointwise" reports coverage (COV) and average length (AL) of (1) the standard 95% CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$; (2) robust CI with $\widehat{K}_{\texttt{cv}}$; (3) robust CI with $\widehat{K}_{\texttt{cv+}}$. "Uniform" reports analogous uniform inference results for confidence bands. $\widehat{K}_{\texttt{cv}}$ is selected to minimize leave-one-out cross-validation and $\widehat{K}_{\texttt{cv+}} = \widehat{K}_{\texttt{cv}} + 2$. Using quadratic spline regressions with evenly placed knots, and the number of knots $K \in \mathcal{K}_n = [2n^{1/3}, 2n^{1/5}]$, $n = 200$.

Table 2: Coverage and Length of Nominal 95% CIs and CBs - Splines ($n = 100$)

	Pointwise								Uniform	
	$x = 0.2$		$x = 0.5$		$x = 0.8$		$x = 0.9$			
	COV	AL	COV	AL	COV	AL	COV	AL	COV	AL
Model 1: $g_1(x) = \ln(\lvert 6x-3\rvert + 1)sgn(x - 1/2)$										
Standard	0.90	0.50	0.93	0.55	0.89	1.36	0.85	1.81	0.32	0.90
Robust ($\widehat{K}_{\mathtt{cv}}$)	0.96	0.65	0.97	0.66	0.94	1.62	0.90	2.12	0.92	1.78
Robust ($\widehat{K}_{\mathtt{cv+}}$)	0.97	0.57	0.96	0.72	0.95	1.79	0.91	2.71	0.93	1.92
Model 2: $g_2(x) = \sin(7\pi x/2)/[1 + 2x^2(sgn(x) + 1)]$										
Standard	0.87	0.50	0.92	0.55	0.89	1.35	0.86	1.82	0.15	0.90
Robust ($\widehat{K}_{\mathtt{cv}}$)	0.93	0.65	0.96	0.67	0.94	1.61	0.90	2.13	0.91	1.78
Robust ($\widehat{K}_{\mathtt{cv+}}$)	0.96	0.57	0.96	0.73	0.95	1.79	0.92	2.71	0.93	1.91
Model 3: $g_3(x) = x - 1/2 + 5\phi(10(x - 1/2))$										
Standard	0.90	0.51	0.90	0.55	0.89	1.36	0.85	1.81	0.28	0.90
Robust ($\widehat{K}_{\mathtt{cv}}$)	0.95	0.65	0.95	0.66	0.94	1.62	0.90	2.12	0.92	1.79
Robust ($\widehat{K}_{\mathtt{cv+}}$)	0.96	0.56	0.95	0.73	0.95	1.78	0.91	2.71	0.92	1.92

Notes: "Pointwise" reports coverage (COV) and average length (AL) of (1) the standard 95% CI with $\widehat{K}_{\mathtt{cv}} \in \mathcal{K}_n$; (2) robust CI with $\widehat{K}_{\mathtt{cv}}$; (3) robust CI with $\widehat{K}_{\mathtt{cv+}}$. "Uniform" reports analogous uniform inference results for confidence bands. $\widehat{K}_{\mathtt{cv}}$ is selected to minimize leave-one-out cross-validation and $\widehat{K}_{\mathtt{cv+}} = \widehat{K}_{\mathtt{cv}} + 2$. Using quadratic spline regressions with evenly placed knots, and the number of knots $K \in \mathcal{K}_n = [2n^{1/3}, 2n^{1/5}]$, $n = 100$.

Table 3: Coverage and Length of Nominal 95% CIs and CBs - Splines ($n = 500$)

	Pointwise								Uniform	
	$x = 0.2$		$x = 0.5$		$x = 0.8$		$x = 0.9$			
	COV	AL	COV	AL	COV	AL	COV	AL	COV	AL
Model 1: $g_1(x) = \ln(\lvert 6x-3\rvert+1)sgn(x-1/2)$										
Standard	0.93	0.19	0.94	0.28	0.93	0.69	0.93	1.10	0.39	0.47
Robust ($\widehat{K}_{\texttt{cv}}$)	0.98	0.27	0.98	0.37	0.98	0.87	0.97	1.37	0.99	0.90
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.99	0.38	0.98	0.40	0.98	1.00	0.97	1.46	0.99	0.97
Model 2: $g_2(x) = \sin(7\pi x/2)/[1+2x^2(sgn(x)+1)]$										
Standard	0.92	0.19	0.95	0.27	0.91	0.69	0.94	1.10	0.39	0.47
Robust ($\widehat{K}_{\texttt{cv}}$)	0.98	0.27	0.99	0.36	0.95	0.87	0.97	1.36	0.99	0.90
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.99	0.38	0.99	0.39	0.98	1.01	0.98	1.45	0.98	0.96
Model 3: $g_3(x) = x - 1/2 + 5\phi(10(x-1/2))$										
Standard	0.91	0.19	0.94	0.27	0.90	0.69	0.93	1.10	0.23	0.47
Robust ($\widehat{K}_{\texttt{cv}}$)	0.98	0.27	0.99	0.36	0.96	0.87	0.97	1.36	0.97	0.90
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.99	0.38	0.98	0.39	0.98	1.01	0.98	1.45	0.97	0.97

Notes: "Pointwise" reports coverage (COV) and average length (AL) of (1) the standard 95% CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$; (2) robust CI with $\widehat{K}_{\texttt{cv}}$; (3) robust CI with $\widehat{K}_{\texttt{cv+}}$. "Uniform" reports analogous uniform inference results for confidence bands. $\widehat{K}_{\texttt{cv}}$ is selected to minimize leave-one-out cross-validation and $\widehat{K}_{\texttt{cv+}} = \widehat{K}_{\texttt{cv}} + 2$. Using quadratic spline regressions with evenly placed knots, and the number of knots $K \in \mathcal{K}_n = [2n^{1/3}, 2n^{1/5}]$, $n = 500$.

Table 4: Coverage and Length of Nominal 95% CIs and CBs - Polynomial

	Pointwise								Uniform	
	$x = 0.2$		$x = 0.5$		$x = 0.8$		$x = 0.9$			
	COV	AL	COV	AL	COV	AL	COV	AL	COV	AL
Model 1: $g_1(x) = \ln(\lvert 6x - 3\rvert + 1)sgn(x - 1/2)$										
Standard	0.91	0.35	0.93	0.34	0.91	0.69	0.92	1.09	0.27	0.55
Robust ($\widehat{K}_{\texttt{cv}}$)	0.95	0.40	0.95	0.38	0.95	0.80	0.94	1.21	0.93	0.98
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.96	0.42	0.95	0.40	0.96	0.99	0.95	1.22	0.97	1.09
Model 2: $g_2(x) = \sin(7\pi x/2)/[1 + 2x^2(sgn(x) + 1)]$										
Standard	0.64	0.37	0.86	0.36	0.88	0.85	0.91	1.12	0.21	0.61
Robust ($\widehat{K}_{\texttt{cv}}$)	0.70	0.42	0.89	0.40	0.93	0.99	0.93	1.24	0.70	1.09
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.77	0.43	0.95	0.40	0.96	1.04	0.94	1.25	0.84	1.12
Model 3: $g_3(x) = x - 1/2 + 5\phi(10(x - 1/2))$										
Standard	0.84	0.39	0.00	0.40	0.88	0.83	0.82	1.10	0.00	0.62
Robust ($\widehat{K}_{\texttt{cv}}$)	0.87	0.44	0.00	1.00	0.93	0.96	0.87	1.22	0.00	1.11
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.94	0.45	0.00	1.02	0.96	1.07	0.88	1.28	0.00	1.17

Notes: "Pointwise" reports coverage (COV) and average length (AL) of (1) the standard 95% CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$; (2) robust CI with $\widehat{K}_{\texttt{cv}}$; (3) robust CI with $\widehat{K}_{\texttt{cv+}}$. "Uniform" reports analogous uniform inference results for confidence bands. $\widehat{K}_{\texttt{cv}}$ is selected to minimize leave-one-out cross-validation and $\widehat{K}_{\texttt{cv+}} = \widehat{K}_{\texttt{cv}} + 2$. Using polynomial regressions with the order of polynomial $K \in \mathcal{K}_n = [n^{1/5}, n^{1/3}]$, $n = 200$.

Table 5: Coverage and Length of Nominal 95% CIs and CBs - Multivariate

	Pointwise						Uniform	
	$(x_1, x_2) = (0.5, 0.5)$		$(x_1, x_2) = (0.1, 0.5)$		$(x_1, x_2) = (0.1, 0.1)$			
	COV	AL	COV	AL	COV	AL	COV	AL
Model : $g(x_1, x_2) = (1 - (4x_1 - 2)^2)^2(\sin(5x_2)/5)$, $x_{1i}, x_{2i} \sim Beta(0.5, 0.5)$								
Standard	0.85	1.09	0.91	1.09	0.84	1.18	0.00	1.17
Robust ($\widehat{K}_{\texttt{cv}}$)	0.94	1.40	0.96	1.33	0.91	1.39	0.82	2.39
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.94	1.61	0.95	1.60	0.93	1.67	0.86	2.78
Model : $g(x_1, x_2) = (1 - (4x_1 - 2)^2)^2(\sin(5x_2)/5)$, $x_{1i}, x_{2i} \sim Beta(1, 1)$								
Standard	0.91	0.84	0.89	1.07	0.86	1.41	0.02	1.03
Robust ($\widehat{K}_{\texttt{cv}}$)	0.97	1.08	0.95	1.30	0.92	1.70	0.92	2.12
Robust ($\widehat{K}_{\texttt{cv+}}$)	0.98	1.24	0.96	1.65	0.93	2.11	0.93	2.51

Notes: "Pointwise" reports coverage (COV) and average length (AL) of (1) the standard 95% CI with $\widehat{K}_{\texttt{cv}} \in \mathcal{K}_n$; (2) robust CI with $\widehat{K}_{\texttt{cv}}$; (3) robust CI with $\widehat{K}_{\texttt{cv+}}$. "Uniform" reports analogous uniform inference results for confidence bands. $\widehat{K}_{\texttt{cv}}$ is selected to minimize leave-one-out cross-validation and $\widehat{K}_{\texttt{cv+}} = \widehat{K}_{\texttt{cv}} + 2$. Using quadratic spline regressions with evenly placed knots.

References

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345-366.

Cattaneo, M. D. and M. H. Farrell (2013): "Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators," *Journal of Econometrics*, 174, 127-143.

Chernozhukov V, D. Chetverikov, and K. Kato (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *The Annals of Statistics*, 41(6), 2786-2819.

Chernozhukov V, D. Chetverikov, and K. Kato (2014a): "Gaussian approximation of suprema of empirical processes," *The Annals of Statistics*, 42(4), 1564-1597.

Chernozhukov V, D. Chetverikov, and K. Kato (2014b): "Anti-Concentration and Honest, Adaptive Confidence Bands," *The Annals of Statistics*, 42(5), 1787-1818.

Chernozhukov V, D. Chetverikov, and K. Kato (2015): "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47-70.

Chernozhukov V, D. Chetverikov, and K. Kato (2016): "Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings," *Stochastic Processes and their Applications*, 126(12), 3632-3651.

Chernozhukov V, D. Chetverikov, and K. Kato (2017): "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45(4), 2309-2352.

Giné, E. and V. Koltchinskii (2006): "Concentration inequalities and asymptotic results for ratio type empirical processes," *The Annals of Probability*, 34(3), 1143-1216.

Tropp, J. A. (2015): *An Introduction to Matrix Concentration Inequalities*, Foundations and Trends in Machine Learning, Vol. 8: No.1-2, 1-230.

Van Der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, Springer.