Supplemental Material for "*Testing (Infinitely) Many Zero Restrictions*"

Jonathan B. Hill*
University of North Carolina – Chapel Hill

April 9, 2022

Contents

List of Tables

*Dept. of Economics, University of North Carolina, Chapel Hill; www.unc.edu/∼jbhill; jbhill@email.unc.edu.

A Outline

This appendix contains omitted proofs from the main paper, and complete simulation results. Appendix B presents the assumptions. Appendix C contains the omitted proofs of Theorem 3.1 (consistency) and supporting Lemmas A.1-A.6. Appendix D contains examples involving linear and logistic regression models with verification of the assumptions. Complete simulation results are presented in Appendix E.

B Assumptions

We assume all random variables exist on a complete measure space, cf. Pollard (1984, Appendix C: permissibility criteria) and Dudley (1984, p. 101: admissible Suslin). $||\cdot||$ denotes the spectral norm for finite dimensional square matrices (and the Euclidean norm for vectors), $|\cdot|$ the l_1-norm ($|x| = \sum_{i,j} |x_{i,j}|$), $|\cdot|_2$ the Euclidean norm ($|x|_2 = (\sum_{i,j} x_{i,j}^2)^{1/2}$), and $||\cdot||_p$ is the L_p-norm. *a.s.* is *almost surely*. $\mathbf{0}_k$ denotes a zero vector with dimension $k \geq 1$. Write r-vectors as $x \equiv [x_i]_{i=1}^r$. $[x]$ rounds x to the nearest integer. $K > 0$ is non-random and finite, and may take different values in different places. Derivatives of functions $f : \mathbb{X} \to \mathbb{R}$ with infinite dimensional $\mathbb{X}$, denoted $(\partial/\partial x)f(x)$, are partial derivatives: $(\partial/\partial x)f(x) \equiv [(\partial/\partial x_i)f(x)]_{i=1}^{\infty}$. $awp1$ = *asymptotically with probability approaching one.*

B.1 Assumptions 1 and 2

Assumption 1 (Identification)**.**

a. The number of parameterizations $k_{\theta,n} \to k_\theta \in \mathbb{N} \cup \infty$.

b. Loss $\mathcal{L} : \mathcal{D} \times \Theta \to [0, \infty)$ *is continuous and differentiable on* $\mathcal{B} \equiv \mathcal{D} \times \Theta$*, where* $\mathcal{D}$ *is a compact subset of* $\mathbb{R}^{k_\delta}$*,* $k_\delta \in \mathbb{N}$*, and* $\Theta = \times_{i=1}^{k_\theta} \Theta_i$*, with non-empty compact* $\Theta_i \subset \mathbb{R}$*. Moreover:*

$$\frac{\partial}{\partial \beta}\mathcal{L}(\beta) = 0 \text{ if and only if } \beta = \beta_0 = [\delta_0', \theta_0']' \in \mathcal{B}, \tag{B.1}$$

where $\beta_0 = [\delta_0', \theta_0']'$ *is unique, and* δ_0 *and* $\theta_{0,i}$ *are interior points of* $\mathcal{D}$ *and* Θ_i*.*

c. Parsimonious loss $\mathcal{L}_{(i)} : \mathcal{B}_{(i)} \to [0, \infty)$ *are continuous and differentiable on compact* $\mathcal{B}_{(i)} \equiv \mathcal{D} \times \Theta_i$*.* $\beta_{(i)}^*$ *are the unique interior points of* $\mathcal{D} \times \Theta_i$ *that satisfy*

$$\frac{\partial}{\partial \beta_{(i)}}\mathcal{L}_{(i)}(\beta_{(i)}) = 0 \text{ if and only if } \beta_{(i)} = \beta_{(i)}^* = [\delta_{(i)}^{*\prime}, \theta_i]' \in \mathcal{D} \times \Theta. \tag{B.2}$$

d. The loss functions $\mathcal{L}(\beta)$ and $\mathcal{L}_{(i)}(\beta_{(i)})$ are linked:

$$\mathcal{L}_{(i)}(\beta_{(i)}) = \mathcal{L}(\delta, [0, ..., \theta_i, 0, ...]') \ \forall \beta_{(i)}. \tag{B.3}$$

Assumption 2 (Consistency)**.**

a. Each estimator $\hat{\beta}_{(i)}$ satisfies $\hat{\mathcal{L}}_{(i)}(\hat{\beta}_{(i)}) = \inf_{\beta_{(i)} \in \mathcal{B}_{(i)}} \{\hat{\mathcal{L}}_{(i)}(\beta_{(i)})\}$.

b. $\sup_{\beta_{(i)} \in \mathcal{B}_{(i)}} |\hat{\mathcal{L}}_{(i)}(\beta_{(i)})/n - \mathcal{L}_{(i)}(\beta_{(i)})| \xrightarrow{p} 0$ *for each i.*

B.2 Assumption 3

Now let $\hat{\mathcal{L}}_{(i)}(\beta_{(i)})$ be twice continuously differentiable. Define gradient and Hessian functions:

$$\widehat{\mathcal{G}}_{(i)}(\beta_{(i)}) \equiv \frac{\partial}{\partial \beta_{(i)}} \hat{\mathcal{L}}_{(i)}(\beta_{(i)}) \text{ and } \widehat{\mathcal{H}}_{(i)}(\beta_{(i)}) \equiv \frac{\partial^2}{\partial \beta_{(i)} \partial \beta'_{(i)}} \hat{\mathcal{L}}_{(i)}(\beta_{(i)}). \tag{B.4}$$

Write

$$\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) \equiv \frac{1}{n} \widehat{\mathcal{H}}_{(i)}(\beta_{(i)}) \text{ and } \mathfrak{H}_{(i)}(\beta_{(i)}) \equiv \text{plim}_{n \to \infty} \frac{1}{n} \widehat{\mathcal{H}}_{(i)}(\beta_{(i)}). \tag{B.5}$$

The asymptotic expansion and Gaussian approximation are derived under H_0, so let H_0 hold for the next two assumptions.

Let $\{\varpi_n\}_{n \in \mathbb{N}}$ be a sequence of positive real numbers, $\varpi_n \to 0$, that may be different in different places; and $\mathcal{B}_{n,(i)} \equiv \{\mathcal{B}_{(i)} : ||\beta_{(i)} - \beta^*_{(i)}|| \leq \varpi_n\}$. Let $\{k_{\theta,n}\}$ be an arbitrary monotonic sequence of positive integers.

Assumption 3 (Asymptotic Expansion)**.** *Let H_0 hold.*

a. $||\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \widehat{\mathfrak{H}}_{(i)}(\beta^*_{(i)})|| \leq \widehat{\mathcal{C}}_{(i)} ||\beta_{(i)} - \beta^*_{(i)}|| \ \forall \beta_{(i)} \in \mathcal{B}_{n,(i)}$ *for every $\{\varpi_n\}$ and some positive stochastic $\widehat{\mathcal{C}}_{(i)}$ with* $\max_{1 \leq i \leq k_{\theta,n}} \widehat{\mathcal{C}}_{(i)} = O_p(1)$. $||\mathfrak{H}_{(i)}(\beta_{(i)}) - \mathfrak{H}_{(i)}(\beta^*_{(i)})|| \leq \mathcal{C}_{(i)} ||\beta_{(i)} - \beta^*_{(i)}|| \ \forall \beta_{(i)} \in \{\mathcal{B}_{(i)} : ||\beta_{(i)} - \beta^*_{(i)}|| \leq \varepsilon\}$, *some $\varepsilon > 0$, and nonstochastic $\mathcal{C}_{(i)} > 0$ with* $\max_{i \in \mathbb{N}} \mathcal{C}_{(i)} < \infty$.

b.

(i) $\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)})$ is symmetric, and positive definite and bounded uniformly awp1*. Specifically:*

$$\inf_{\lambda'\lambda = 1} \min_{1 \leq i \leq k_{\theta,n}} \left\{ \inf_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \left\{ \lambda' \widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) \lambda \right\} \right\} > 0 \ awp1.$$

$$\max_{1 \leq i \leq k_{\theta,n}} \left\{ \sup_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \left\| \widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) \right\| \right\} = O_p(k_{\theta,n}).$$

(ii) For some $\zeta, \xi > 0$,

$$\max_{1 \leq i \leq k_{\theta,n}} E \left[\exp \left\{ \zeta \sup_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \sqrt{n} \left| \widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \mathfrak{H}_{(i)}(\beta_{(i)}) \right| \right\} \right] = O\left(n^{\xi \ln(k_{\theta,n})} \right),$$

where $\mathfrak{H}_{(i)}(\beta_{(i)})$ *is non-random, symmetric, and uniformly positive definite on* $\{\beta_{(i)} : ||\beta_{(i)} - \beta^*_{(i)}|| \leq \varepsilon\}$ *for some* $\varepsilon > 0$. *In particular* $\max_{i\in\mathbb{N}}\{\sup_{\beta_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon} ||\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})||\} < \infty$.

c. $\max_{1\leq i\leq k_{\theta,n}} E[\exp\{\zeta|\widehat{\mathcal{G}}_{(i)}/\sqrt{n}|\}] = O(n^{\xi\ln(k_{\theta,n})})$ *for some* $\zeta, \xi > 0$.

We discuss (c) first. The condition implies each $|\widehat{\mathcal{G}}_{(i)}/\sqrt{n}|$ has a Laplace transform and therefore moment generating function $[mgf]$ that is uniformly $O((\ln k_{\theta,n})^2)$. A well known necessary and sufficient condition for a random variable z to have a mgf is the exponential tail bound $P(|z| > g) \leq \mathcal{K}e^{-cg}$ for all g and some finite $c, \mathcal{K} > 0$.

Now suppose $\widehat{\mathcal{G}}_i \equiv \sum_{t=1}^n \epsilon_{i,t} w_{i,t}$ where $\{\epsilon_{i,t}, w_{i,t}\}_{t\in\mathbb{Z}}$ with $w_{i,t} = [w_{l,i,t}]_{l=1}^{k_\delta+1}$ are random variables on a common probability space, and $E[\epsilon_{i,t}] = 0$. In view of $E[\widehat{\mathcal{G}}_i] = 0$ under H_0, we may assume $E[w_{i,t}] = 0$ without loss of generality. The following provides a sufficient condition for (c) by exploiting sub-Gaussianicity for $\{\epsilon_{i,t}, w_{i,t}\}$. Recall a random variable z is *sub-Gaussian* [subG] when the Laplace transform $E[\exp\{\zeta z\}] \leq \exp\{\zeta^2\sigma^2/2\}$ $\forall\zeta \in \mathbb{R}$ and some finite $\sigma > 0$, hence necessarily $E[z] = 0$ (see, e.g., Kahane, 1960, Buldygin and Kozachenko, 1980, Stromberg, 1994, Vershynin, 2018).

Assumption 3.c† *Let* $\widehat{\mathcal{G}}_{(i)} \equiv \sum_{t=1}^n \epsilon_{i,t} w_{i,t}$, $w_{i,t} = [w_{l,i,t}]_{l=1}^{k_\delta+1}$, *where* $(\epsilon_{i,t}, w_{i,t})$ *are random variables on a common probability space, independent across* t, $E[\epsilon_{i,t}] = 0$, *and* $\epsilon_{i,s}$ *and* $w_{l,i,t}$ *are mutually independent* $\forall i, l, s, t$. *Further,* $\epsilon_{i,t}$ *and* $w_{l,i,t}$ *are uniformly* subG*:* $\max_{i,t\in N} E[\exp\{\zeta\epsilon_{i,t}\}] \leq \exp\{\zeta^2\sigma^2_\epsilon/2\}$ *and* $\max_{1\leq l\leq k_\delta+1}\max_{i,t\in N} E[\exp\{\zeta w_{l,i,t}\}] \leq \exp\{\zeta^2\sigma^2_w/2\}$ $\forall\zeta \in \mathbb{R}$ *and some universal constants* $\sigma_\epsilon, \sigma_w \in (0, \infty)$.

Lemma B.1. *Assumption 3.c*† *implies Assumption 3.c holds for any sequence of positive integers* $\{k_{\theta,n}\}$.

Remark 1. The proof does not make use of independence for $w_{i,t}$ over t, thus the serial dependence properties of $w_{i,t}$ can be left unrestricted. Mutual independence effectively imposes strict exogeneity on $w_{i,t}$. A broader dependence setting like a mixing property on $\{\epsilon_{i,t}w_{i,t}\}$ is left for future work.

Remark 2. Bounded random variables are *subG* by the Azuma–Hoeffding inequality, and Gaussian and Gaussian mixtures are *subG*. The *subG* property is closed under zero mean linear combinations, and under product convolutions, both for independent random variables. See, for example, Vershynin (2018), cf. Buldygin and Kozachenko (1980) and Stromberg (1994), and see the proof below.

Proof. In order to reduce notation, and without loss of generality, assume $w_{i,t} \in \mathbb{R}$ (i.e. $k_\delta = 0$).

It is easy to verify $E[\exp\{\zeta|z|\}] \leq E[\exp\{\zeta z\}] + E[\exp\{-\zeta z\}]$ for any $\zeta \in \mathbb{R}$. It therefore suffices to prove for some $\zeta, \xi > 0$:

$$\max_{1\leq i\leq k_{\theta,n}} E\left[\exp\left\{\zeta\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_{i,t}w_{i,t}\right\}\right] = O\left(n^{\xi\ln(k_{\theta,n})}\right).$$

Independence and sub-Gaussianicity for $\epsilon_{i,t}$, mutual independence, and Young's inequality yield:

$$\begin{aligned}\max_{1\leq i\leq k_{\theta,n}} E\left[\exp\left\{\zeta\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_{i,t}w_{i,t}\right\}\right] &= \max_{1\leq i\leq k_{\theta,n}} E\left(\prod_{t=1}^{n} E\left[\exp\left\{\zeta\frac{1}{\sqrt{n}}\epsilon_{i,t}w_{i,t}\right\} \mid \{w_{i,t}\}_{t=1}^{n}\right]\right) \\ &\leq \max_{1\leq i\leq k_{\theta,n}} E\left[\prod_{t=1}^{n}\exp\left\{\frac{1}{2}\sigma_{\epsilon}^{2}\zeta^{2}\frac{1}{n}w_{i,t}^{2}\right\}\right] \\ &\leq \max_{1\leq i\leq k_{\theta,n}} E\left[\frac{1}{n}\sum_{t=1}^{n}\exp\left\{n\frac{1}{2}\sigma_{\epsilon}^{2}\zeta^{2}\frac{1}{n}w_{i,t}^{2}\right\}\right] \\ &\leq \max_{1\leq i\leq k_{\theta,n}}\max_{1\leq t\leq n} E\left[\exp\left\{\frac{1}{2}\sigma_{\epsilon}^{2}\zeta^{2}w_{i,t}^{2}\right\}\right]. \end{aligned} \tag{B.6}$$

By assumption $w_{i,t}$ is uniformly *subG*, hence $w_{i,t}^2$ is uniformly *locally sub-exponential*. Simply note that for some universal constant $\sigma_w^2 > 0$:

$$\begin{aligned}\max_{1\leq i\leq k_{\theta,n}}\max_{1\leq t\leq n} E\left[\exp\left\{\frac{1}{2}\sigma_{\epsilon}^{2}\zeta^{2}w_{i,t}^{2}\right\}\right] &= 1 + \int_{1}^{\infty}\max_{1\leq i\leq k_{\theta,n}}\max_{1\leq t\leq n} P\left(\exp\left\{\frac{1}{2}\sigma_{\epsilon}^{2}\zeta^{2}w_{i,t}^{2}\right\} > u\right)du \\ &= 1 + \int_{1}^{\infty}\max_{1\leq i\leq k_{\theta,n}}\max_{1\leq t\leq n} P\left(|w_{i,t}| > \frac{\sqrt{2}}{\sigma_{\epsilon}\zeta}(\ln(u))^{1/2}\right)du \\ &\leq 1 + 2\int_{1}^{\infty}\exp\left\{-\frac{2\sigma_{w}^{2}}{\sigma_{\epsilon}^{2}\zeta^{2}}\ln(u)\right\}du \\ &= 1 + 2\int_{1}^{\infty}u^{-2\sigma_{w}^{2}/(\sigma_{\epsilon}^{2}\zeta^{2})}du.\end{aligned}$$

Now restrict $\zeta \leq \sigma_w/\sigma_\epsilon$ to yield for some $C \geq 2$ and $K \geq \sigma_\epsilon^2/\sigma_w^2$:

$$\begin{aligned}\max_{1\leq i\leq k_{\theta,n}}\max_{1\leq t\leq n} E\left[\exp\left\{\zeta^{2}w_{i,t}^{2}\right\}\right] &\leq 1 + \frac{2}{2\sigma_{w}^{2}/(\sigma_{\epsilon}^{2}\zeta^{2}) - 1} \\ &= \frac{2\sigma_{w}^{2}/\sigma_{\epsilon}^{2} + \zeta^{2}}{2\sigma_{w}^{2}/\sigma_{\epsilon}^{2} - \zeta^{2}} \leq \frac{2\sigma_{w}^{2}/\sigma_{\epsilon}^{2} + \zeta^{2}}{\sigma_{w}^{2}/\sigma_{\epsilon}^{2}} \\ &= 2 + \frac{\sigma_{\epsilon}^{2}}{\sigma_{w}^{2}}\zeta^{2} \leq C\exp\left\{K\zeta^{2}\right\}.\end{aligned} \tag{B.7}$$

The final inequality follows by noting:

$$\text{at } \zeta = 0 \Longrightarrow 2 + \frac{\sigma_\epsilon^2}{\sigma_w^2}\zeta^2 = 2 \leq C \exp\left\{K\zeta^2\right\} = C$$

and for all $\zeta > 0$:

$$\frac{\partial}{\partial \zeta}\left(2 + \frac{\sigma_\epsilon^2}{\sigma_w^2}\zeta^2\right) = 2\frac{\sigma_\epsilon^2}{\sigma_w^2}\zeta < \frac{\partial}{\partial \zeta} C \exp\left\{K\zeta^2\right\} = 2CK\zeta \exp\left\{K\zeta^2\right\}.$$

Combine (B.6) and (B.7) to deduce for any $0 < \zeta \leq \sigma_w/\sigma_\epsilon$:

$$\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_{i,t} w_{i,t}\right\}\right] \leq C \exp\left\{K\zeta^2\right\}.$$

This completes the proof. $\mathcal{QED}$.

Define an arbitrary neighborhood of zero $\mathcal{N}_b \equiv (-b, b)$ for some $b \in (0, \infty)$. Under *subG* the Laplace transform bound holds *for all* $\zeta \in \mathbb{R}$, but Assumption 3.c only needs the key bound to hold *for some* $\zeta > 0$. The *locally subG* property $E[\exp\{\zeta z\}] \leq \exp\{\zeta \upsilon + \zeta^2\sigma^2/2\}$ $\forall \zeta \in \mathcal{N}_b$, however, trivially allows for a non-zero mean ($E[z] = \upsilon$) and ζ in only a neighborhood of zero. *Local subG* therefore nests *subG*, and has the same characteristics, including being closed under linear combinations and product convolutions for independent random variables, and having exponentially bounded tails $P(|z - \upsilon| > c) \leq \exp\{-c^2\sigma^2/2\}$ (see Chareka, Chareka, and Kennedy, 2006). Further, all random variables with a bounded *mgf* in some neighborhood of zero are *locally subG* (e.g. Binomial, exponential, Gamma, Laplace, Poisson) (Chareka, Chareka, and Kennedy, 2006, Theorem 1). The existence of a *mgf* for some random variable z is equivalent to $P(|z| > c) \leq \mathcal{K}e^{-\delta c}$ $\forall c > 0$ and some finite $(\delta, \mathcal{K})$ > 0 by Chernoff's bound. The subsequent corollary now follows (and a proof is therefore omitted).

Assumption 3.c* *Let* $\widehat{\mathcal{G}}_{(i)} \equiv \sum_{t=1}^{n} \epsilon_{i,t} w_{i,t}$, $w_{i,t} = [w_{l,i,t}]_{l=1}^{k_\delta+1}$, *where* $(\epsilon_{i,t}, w_{i,t})$ *are random variables on a common probability space, independent across* t, $E[\epsilon_{i,t}] = 0$, *and* $\epsilon_{i,s}$ *and* $w_{l,i,t}$ *are mutually independent* $\forall i, l, s, t$. *Further,* $\max_{i,t \in N} P(|\epsilon_{i,t}| > c) \leq \mathcal{K}e^{-\delta c}$ *and* $\max_{i,t \in N} P(|w_{i,t} - E[w_{i,t}]| > c) \leq \mathcal{K}e^{-\delta c}$ $\forall c > 0$ *for finite constants* $(\delta, \mathcal{K}) > 0$ *that may be different in different places.*

Corollary B.2. *Assumption 3.c* implies Assumption 3.c.*

Remark 3. As with Lemma B.1, we do not technically need to restrict dependence for $w_{i,t}$ over t.

The same idea extends to Assumption 3.b(ii). In this case, however, we are not working with a product convolution involving mutually independent random variables, and an independent sequence (e.g. $\epsilon_{i,t}$). We therefore impose independence on the key Hessian process.

Assumption 3.b(ii)* *Let* $\widehat{\mathfrak{H}}_{(i)} \equiv 1/n \sum_{t=1}^{n} h_{i,t}(\beta_{(i)})$ *and* $\mathfrak{H}_{(i)}(\beta_{(i)}) \equiv \lim_{n\to\infty} 1/n \sum_{t=1}^{n} E[h_{i,t}(\beta_{(i)})]$, *where* $\{h_{i,t}(\cdot) = [h_{l,m,i,t}(\cdot)]_{l,m=1}^{k_\delta+1}\}$ *are measurable random variables on a common probability space, independent across* t. *For some finite universal constants* $(\delta, \mathcal{K}) > 0$:

$$\max_{1\leq l,m\leq k_\delta+1} \max_{i,t\in N} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \left|h_{l,m,i,t}(\beta_{(i)}) - \mathfrak{H}_{l,m,(i)}(\beta_{(i)})\right| > c\right) \leq \mathcal{K}e^{-\delta c} \;\forall c > 0. \tag{B.8}$$

Corollary B.3. *Assumption 3.b(ii)* implies Assumption 3.b(ii)..*

Proof. Write

$$\mathcal{J}_{l,m,i,t} \equiv \sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \left|h_{l,m,i,t}(\beta_{(i)}) - \mathfrak{H}_{l,m,(i)}(\beta_{(i)})\right|.$$

Under (B.8) $\mathcal{J}_{l,m,i,t}$ satisfies $\forall\zeta \in [\delta^{a-1}, \delta/2]$, any $a > 0$ such that that $\delta^{a-1} \leq \delta/2$, and some $K > 0$:

$$\begin{aligned}
&\max_{1\leq l,m\leq k_\delta+1} \max_{i,t\in N} E\left[\exp\left\{\zeta\mathcal{J}_{l,m,i,t}\right\}\right] \qquad\qquad (B.9)\\
&\quad = 1 + \max_{1\leq l,m\leq k_\delta+1} \max_{i,t\in N} \int_1^\infty P\left(\mathcal{J}_{l,m,i,t} > \frac{1}{\zeta}\ln(c)\right) dc \\
&\quad = 1 + \mathcal{K}\frac{\zeta}{\delta - \zeta} \leq 1 + 2\mathcal{K}\frac{\zeta}{\delta} \leq 1 + 2\mathcal{K}\frac{\zeta^2}{\delta^a} \leq \exp\left\{K\zeta^2\right\}.
\end{aligned}$$

The first inequality uses $\zeta \leq \delta/2$, the second uses $\zeta \geq \delta^{a-1}$. Now use independence with (B.9) to yield:

$$\begin{aligned}
\max_{1\leq i\leq k_{\theta,n}} E\left[\exp\left\{\zeta\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{J}_{l,m,i,t}\right\}\right] &= \max_{1\leq i\leq k_{\theta,n}} \prod_{t=1}^{n} E\left[\exp\left\{\zeta\frac{1}{\sqrt{n}}\mathcal{J}_{l,m,i,t}\right\}\right] \\
&\leq \left(\exp\left\{K^2\zeta^2\frac{1}{n}\right\}\right)^n = \exp\left\{K^2\zeta^2\right\}.
\end{aligned}$$

The triangle inequality therefore yields:

$$\max_{1\leq i\leq k_{\theta,n}} E\left[\exp\left\{\zeta \sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \sqrt{n}\left|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \mathfrak{H}_{(i)}(\beta_{(i)})\right|\right\}\right] = O(1),$$

hence Assumption 3.b(ii) holds. $\mathcal{QED}$.

B.3 Assumption 4

Recall

$$\hat{\mathcal{Z}}_{(i)} = -\mathfrak{H}_{(i)}^{-1}(\beta_{(i)}^*)\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}(\beta_{(i)}^*) = -\mathfrak{H}_{(i)}^{-1}\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}.$$

Define a long run covariance kernel

$$\mathcal{S}_n(i,j) \equiv \frac{1}{n}E\left[\widehat{\mathcal{G}}_{(i)}\widehat{\mathcal{G}}_{(j)}'\right] \text{ and } \mathcal{S}(i,j) = \lim_{n\to\infty}\mathcal{S}_n(i,j).$$

Assumption 4.a below implies $\mathcal{S}\left(i,j\right)$ exists. Define

$$\begin{aligned} &\mathcal{V}_{n,i} \equiv \mathfrak{H}_{(i)}^{-1}\mathcal{S}_n(i,i)\mathfrak{H}_{(i)}^{-1} \text{ and } \mathcal{V}_{(i)} \equiv \mathfrak{H}_{(i)}^{-1}\mathcal{S}(i,i)\mathfrak{H}_{(i)}^{-1} \\ &\sigma_{n,i}^2(\lambda) \equiv \lambda'\mathcal{V}_{n,i}\lambda \text{ and } \sigma_i^2(\lambda) \equiv \lambda'\mathcal{V}_{(i)}\lambda \end{aligned} \tag{B.10}$$

where $\lambda \in \mathbb{R}^{k_\delta+1}$. By construction $\sigma_{n,i}^2(\lambda) = E[(\lambda'\hat{\mathcal{Z}}_{(i)})^2]$. Our specific interest is in the case $\lambda = [\mathbf{0}_{k_\delta}', 1]'$

Recall for a Gaussian approximation theory, we assume the loss gradient $\widehat{\mathcal{G}}_{(i)}$ has the form

$$\widehat{\mathcal{G}}_{(i)} = \sum_{t=1}^{n} G_{(i),t}$$

where $\{G_{i,t}\}_{t=1}^n$ for each i are properly defined independent random variables, with $E[G_{i,t}] = 0$. Then

$$\lambda'\hat{\mathcal{Z}}_{(i)} = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\lambda'\mathfrak{H}_{(i)}^{-1}G_{i,t} \equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{(i),t}(\lambda).$$

By independence:

$$\sigma_{n,(i)}^2(\lambda) = E\left[(\lambda'\hat{\mathcal{Z}}_{(i)})^2\right] = \frac{1}{n}\sum_{t=1}^{n}\lambda'\mathfrak{H}_{(i)}^{-1}E\left[G_{i,t}G_{i,t}'\right]\mathfrak{H}_{(i)}^{-1}\lambda = \lambda'\mathcal{V}_{n,(i)}\lambda.$$

Assumption 4 (Limit Distribution)**.** *Let H_0 hold, and let $\max_{i,t\in\mathbb{N}}||\mathfrak{H}_{(i)}^{-1}|| < \infty$.*

a. $G_{i,t}$ is independent across t, and $\max_{i,t\in\mathbb{N}}||E[G_{i,t}G_{i,t}']|| < \infty$.

b. Let (i) $\liminf_{n\to\infty}\min_{i\in\mathbb{N}}\inf_{\lambda'\lambda=1}\sigma_{n,(i)}^2(\lambda) \geq \underline{c}$ and (ii) $\limsup_{n\to\infty}\max_{i\in\mathbb{N}}\sup_{\lambda'\lambda=1}\sigma_{n,(i)}^2(\lambda) \leq \bar{c}$ for some $0 < \underline{c} \leq \bar{c} < \infty$.

c. $\max_{\gamma=1,2}\{E|z_{(i),t}(\lambda)|^{2+\gamma}/B_n^\gamma\} + E[\exp\{|z_{(i),t}(\lambda)|/B_n\}] \leq 4$ uniformly in $i,t \in \mathbb{N}$ and $\lambda'\lambda = 1$, for some sequence of positive non-random numbers $\{B_n\}$, $B_n \geq 1$ where $B_n \to \infty$ is possible.

B.4 Assumption 5

In the parametric bootstrap setting, we work with a nonlinear regression model:

$$y_t = f(x_t, \beta_0) + \epsilon_t$$

where f is a known response function and $\beta_0 = [\delta_0', \theta_0']' \in \mathcal{B}$. The null hypothesis is $H_0 : \theta_0 = 0$.

Let $f_{(i)}(x_t, \beta_{(i)})$ be $f(x_t, \delta, \theta)$ when the i^{th} element of θ is θ_i and the remaining $\theta_j = 0$ for $j \neq i$:

$$f_{(i)}(x_t, \beta_{(i)}) \equiv f(x_t, \delta_{(i)}, [0, ..., \theta_i, 0, , ,]')$$

where $[0, ..., \theta_i, 0, , ,]'$ is a zero vector with i^{th} element θ_i. Define the response gradient

$$g_{(i)}(x_t, \beta_{(i)}) \equiv \frac{\partial}{\partial \beta_{(i)}} f_{(i)}(x_t, \beta_{(i)}) \text{ and } h_{(i)}(x_t, \beta_{(i)}) \equiv \left(\frac{\partial}{\partial \beta_{(i)}}\right)^2 f_{(i)}(x_t, \beta_{(i)}).$$

The parsimonious models are

$$y_t = f_{(i)}(x_t, \beta_{(i)}^*) + v_{(i),t} \text{ where } \beta_{(i)}^* = [\delta_{(i)}^{*\prime}, \theta_i^*]' \text{ is unique, } E\left[v_{(i),t} g_{(i)}(x_t, \beta_{(i)}^*)\right] = 0.$$

Now define the restricted estimator $\hat{\beta}^{(0)} \equiv [\hat{\delta}^{(0)\prime}, \mathbf{0}_{k_\theta}']'$ under $H_0 : \theta_0 = 0$, where $\hat{\delta}^{(0)}$ uniquely minimizes $.5 \sum_{t=1}^n \{y_t - f(x_t, [\delta_0, \mathbf{0}_{k_\theta}])\}^2$ on $\mathcal{D}$. Define residuals

$$\epsilon_{n,t}^{(0)} \equiv y_t - f(x_t, \hat{\beta}^{(0)}).$$

Let $\{\eta_t\}_{t=1}^n$ be iid draws from $N(0, 1)$, and generate the array

$$y_{n,t}^* \equiv f(x_t, \hat{\beta}^{(0)}) + \epsilon_{n,t}^{(0)} \eta_t.$$

Now construct $k_{\theta,n}$ parsimonious regression models

$$y_{n,t}^* = f_{(i)}(x_t, \tilde{\beta}_{(i)}) + v_{n,(i),t}.$$

The bootstrap estimator of $\tilde{\beta}_{(i)}$ is $\widehat{\tilde{\beta}}_{(i)} = [\widehat{\tilde{\delta}}_{(i)}', \widehat{\tilde{\theta}}_i]'$ of $\tilde{\beta}_{(i)}$ on $\mathcal{B}_{(i)} \equiv \mathcal{D} \times \Theta_i$.

Write $\beta^{(0)} = [\delta^{(0)\prime}, \mathbf{0}_{k_\theta}']'$ and $\beta_{(i)}^{(0)} = [\delta^{(0)\prime}, 0]'$, where $\delta^{(0)}$ minimizes $.5E[(y_t - f(x_t[\delta, \mathbf{0}_{k_\theta}]))^2]$. Write $g_{(i),t}^{(0)} = [g_{(i),t,l}^{(0)}]_{l=1}^{k_\delta+1} \equiv g_{(i)}(x_t, \beta_{(i)}^{(0)})$, $h_{(i)}(x_t, \beta_{(i)}) = [h_{(i),l,m}(x_t, \beta_{(i)})]_{l,m=1}^{k_\delta+1} \equiv (\partial/\partial\beta_{(i)})^2 f(x_t, \beta_{(i)})$ and

$$G_{(i),t}^{(0)} \equiv \left\{y_t - f(x_t, \beta^{(0)})\right\} g_{(i),t}^{(0)}.$$

Define:

$$\bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}, \tilde{\beta}_{(i)}) \equiv \frac{1}{n}\sum_{t=1}^{n} g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \tilde{\beta}_{(i)})' \text{ and } \bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}) \equiv \bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}, \beta_{(i)})$$

$$\bar{\mathcal{H}}_{n.(i)}(\beta_{(i)}) \equiv \frac{1}{n}\sum_{t=1}^{n} E[g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \beta_{(i)})'] \text{ and } \bar{\mathcal{H}}_{(i)}(\beta_{(i)}) \equiv \max_{t\in\mathbb{N}} E\left[g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \beta_{(i)})'\right]$$

$$\bar{\mathcal{H}}_{n,(i)}^{(0)} \equiv \frac{1}{n}\sum_{t=1}^{n} E[g_{(i),t}^{(0)} g_{(i),t}^{(0)\prime}] \text{ and } \bar{\mathcal{H}}_{(i)}^{(0)} \equiv \lim_{n\to\infty} \bar{\mathcal{H}}_{n,(i)}^{(0)}.$$

Assumption 5 (Bootstrap)**.**

a. Response Function*: $y_t = f(x_t, \beta_0) + \epsilon_t$, $\{x_t, \epsilon_t\}$ are independent over t and mutually independent, $E[\epsilon_t] = 0$, $E[\epsilon_t^2] \in (0, \infty)$ uniformly in $t \in \mathbb{N}$; $f : \mathbb{R}^{k_x} \times \mathcal{B} \to \mathbb{R}$, where $\mathcal{B} \equiv \mathcal{D} \times \Theta \subset \mathbb{R}^{k_\beta}$ with $\Theta = \times_{i=1}^{k_\theta}\Theta_i$; $\mathcal{D} \subset \mathbb{R}^{k_\delta}$ and $\Theta_i \subset \mathbb{R}$ are compact. $f(x, \cdot)$ is for each x Borel measurable, and $f(\cdot, \beta)$ is three times continuously differentiable on $\mathcal{B}$.*

b. Identification*: $\beta_0 \equiv [\delta_0', \theta_0']'$ uniquely minimizes $E[(y_t - f(x_t, \beta))^2]$ on $\mathcal{B}$ $\forall t$, where δ_0 and $\theta_{0,i}$ are interior points of $\mathcal{D}$ and Θ_i. $E[(y_t - f(x_t, [\delta, \mathbf{0}_{k_\theta}]))^2]$ has $\forall t$ a unique minimum on $\mathcal{D}$.*

c. Sub-Exponential Tails*: Let $w_{i,t}$ denote $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |g_{(i)}(x_t, \beta_{(i)})|$, $\sup_{\delta\in\mathcal{D}} |f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])|$, $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\} g_{(i)}(x_t, \beta_{(i)})|$ or $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\} h_{(i)}(x_t, \beta_{(i)})|$. Then $\max_{t\in\mathbb{N}} P(|\epsilon_t| > c) \leq C\exp\{-\mathcal{K}c\}$ and $\max_{i,t\in\mathbb{N}} P(|w_{i,t}| > c) \leq C\exp\{-\mathcal{K}c\}$ for some finite $C, \mathcal{K} > 0$ that may be different for different $(\epsilon_t, w_{i,t})$.*

d. Envelope Bounds*: Define $v_{(i),t}(\beta_{(i)}) \equiv y_t - f_{(i)}(x_t, \beta_{(i)})$. For some $p > 4$, each i, and each $l, m = 1, ..., k_\delta + 1$ and $s = 0, 1, 2, 3$: $E[\sup_{\beta\in\mathcal{B}} |y_t - f(x_t, \beta)|^p] < \infty$, and*

$$\max_{i,t\in\mathbb{N}} E\left[\left\{\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\left(\frac{\partial}{\partial\beta_{(i),l}}\right)^s f_{(i)}(x_t, \beta_{(i)})\right|^p\right\}\right] < \infty$$

$$\min_{i,t\in\mathbb{N}} \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}} \inf_{\lambda'\lambda=1} E\left[\left(\lambda' g_{(i)}(x_t, \beta_{(i)})\right)^2\right] > 0$$

$$\min_{i,t\in\mathbb{N}} \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}} \inf_{\lambda'\lambda=1} E\left[v_{(i),t}^2(\beta_{(i)}) \left(\lambda' g_{(i)}(x_t, \beta_{(i)})\right)^2\right] > 0$$

$$\max_{i,l,t\in\mathbb{N}} E\left[\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left\{g_{(i),l}^4(x_t, \beta_{(i)}) + h_{(i),l,m}^4(x_t, \beta_{(i)})\right\}\right] < \infty$$

$$\max_{i,l,t\in\mathbb{N}} E\left[\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left\{v_{(i),t}^4(\beta_{(i)}) \left(g_{(i),l}^4(x_t, \beta_{(i)}) + h_{(i),l,m}^4(x_t, \beta_{(i)})\right)\right\}\right] < \infty$$

Let $\mathcal{B}_{(i)}(\varsigma) \equiv \{\mathcal{B}_{(i)} : ||\beta_{(i)} - \beta_{(i)}^*|| \leq \varsigma\}$ for some fixed $\varsigma > 0$. Let $\bar{\lambda}_{n.(i)}^{(\min)}(\beta_{(i)})$ and $\bar{\lambda}_{n.(i)}^{(\max)}(\beta_{(i)})$

denote minimum and maximum eigenvalues of $\bar{\mathcal{H}}_{n.(i)}(\beta_{(i)})$*; let* $\bar{\lambda}_{(i)}^{(\min)}(\beta_{(i)})$ *and* $\bar{\lambda}_{(i)}^{(\max)}(\beta_{(i)})$ *be the minimum and maximum eigenvalues for* $\bar{\mathcal{H}}_{(i)}(\beta_{(i)})$*. For some* ς*:*

$$\min_{i\in\mathbb{N}} \inf_{\lambda'\lambda=1} \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}(\varsigma)} \left\{\lambda'\bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)})\lambda\right\} > 0 \ a.s.$$

$$\min_{i\in\mathbb{N}} \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}(\varsigma)} \bar{\lambda}_{n.(i)}^{(\min)}(\beta_{(i)}) > 0 \quad and \quad \limsup_{n\to\infty} \max_{i\in\mathbb{N}} \sup_{\beta_{(i)}\in\mathcal{B}_{(i)}(\varsigma)} \bar{\lambda}_{n.(i)}^{(\max)}(\beta_{(i)}) < \infty$$

$$\min_{i\in\mathbb{N}} \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}(\varsigma)} \bar{\lambda}_{(i)}^{(\min)}(\beta_{(i)}) > 0 \quad and \quad \max_{i\in\mathbb{N}} \sup_{\beta_{(i)}\in\mathcal{B}_{(i)}(\varsigma)} \bar{\lambda}_{(i)}^{(\max)}(\beta_{(i)}) < \infty$$

$$\max_{1\leq i\leq k_{\theta,n}} \left\|\bar{\mathcal{H}}_{n.(i)}^{(0)-1}\right\| = O(1) \quad and \ \max_{i\in\mathbb{N}} \left\|\bar{\mathcal{H}}_{(i)}^{(0)-1}\right\| < \infty$$

$$\max_{1\leq i\leq k_{\theta,n}} \left\|\bar{\mathcal{H}}_{n,(i)}^{(0)-1} - \bar{\mathcal{H}}_{(i)}^{(0)-1}\right\| = O(k_{\theta,n}/\sqrt{n}) \tag{B.11}$$

and:

$$\liminf_{n\to\infty} \min_{i\in\mathbb{N}} \inf_{\lambda'\lambda=1} \inf_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}(\varsigma)\times\tilde{\mathcal{B}}_{(i)}(\varsigma)} \left\{\lambda'\bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)},\tilde{\beta}_{(i)})\lambda\right\} > 0 \ a.s.$$

$$\max_{1\leq i\leq k_{\theta,n}} \sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}(\varsigma)\times\tilde{\mathcal{B}}_{(i)}(\varsigma)} \left\|\bar{\mathfrak{H}}_{n,(i)}^{-1}(\beta_{(i)},\tilde{\beta}_{(i)})\right\| = O_p(k_{\theta,n})$$

Finally, for any monotonic sequence of positive integers $\{k_{\theta,n}\}$*:*

$$\max_{1\leq i,j\leq k_{\theta,n}} \left\|\frac{1}{n}\sum_{t=1}^{n} E\left[G_{(i),t}^{(0)}G_{(j),t}^{(0)\prime}\right] - \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E\left[G_{(i),t}^{(0)}G_{(j),t}^{(0)\prime}\right]\right\| = O(k_{\theta,n}/\sqrt{n}) \tag{B.12}$$

and $\max_{i,j\in\mathbb{N}} ||\lim_{n\to\infty} 1/n\sum_{t=1}^{n} E[G_{i,t}^{(0)}G_{j,t}^{(0)\prime}]|| < \infty$.

Remark 4. Bounds (B.11) and (B.12) effectively restrict heterogeneity (non-identical distributedness). Under stationarity the bounds are trivial since, e.g., $1/n\sum_{t=1}^{n} E[G_{(i),t}^{(0)}G_{(j),t}^{(0)\prime}]$ $= E[G_{(i),1}^{(0)}G_{(j),1}^{(0)\prime}]$. In general the bounds are mild since, e.g., for iid $\eta_t \sim N(0,1)$ that is independent of $\{x_t, y_t\}_{t=1}^{\infty}$:

$$\max_{1\leq i,j\leq k_{\theta,n}} \left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\eta_t^2 G_{(i),t}^{(0)}G_{(j),t}^{(0)\prime} - E\left[G_{(i),t}^{(0)}G_{(j),t}^{(0)\prime}\right]\right\}\right\| = O_p(k_{\theta,n}/\sqrt{n}).$$

See the proof of Lemma A.6.

C Omitted Proofs

C.1 Theorem 3.1

Theorem 3.1. *Under Assumptions 1.c and 2,* $\hat{\beta}_{(i)} \xrightarrow{p} \beta^*_{(i)}$ *for each* $1 \leq i \leq \mathring{k}$ *and any* $\mathring{k} \in \mathbb{N}$, $\mathring{k} \leq \lim_{n\to\infty} k_{\theta,n}$. *Hence* $\hat{\theta}_i \xrightarrow{p} 0$ if and only if H_0 *is true.*

Proof. Define

$$\tilde{\mathcal{L}}_i\left(\beta_{(i)}\right) \equiv \mathcal{L}_{(i)}(\beta_{(i)}) - \mathcal{L}_{(i)}(\beta^*_{(i)}) \text{ and } \tilde{\mathcal{L}}_{n,i}\left(\beta_{(i)}\right) \equiv \frac{1}{n}\left(\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) - \hat{\mathcal{L}}_{(i)}(\beta^*_{(i)})\right).$$

By identification Assumption 1.c it follows $\tilde{\mathcal{L}}_i(\beta_{(i)}) = 0$ *if and only if* $\beta_{(i)} = \beta^*_{(i)}$. We will prove $\hat{\beta}_{(i)} \xrightarrow{p} \beta^*_{(i)}$ by verifying conditions (i)-(iii) of Theorem 3.1 in Pakes and Pollard (1989). The second claim, $\hat{\theta}_i \xrightarrow{p} 0$ for each $1 \leq i \leq k_{\theta,n}$ *if and only if* H_0 is true, then follows by Corollary 2.2.

By construction and Assumption 2.a

$$\tilde{\mathcal{L}}_{n,i}(\hat{\beta}_{(i)}) = \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}}\left\{\tilde{\mathcal{L}}_{n,i}\left(\beta_{(i)}\right)\right\},$$

and trivially by construction $\tilde{\mathcal{L}}_{n,i}(\beta^*_{(i)}) = 0$, hence Pakes and Pollard (1989)'s (i) and (ii) hold.

Condition (iii) holds if $\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi} |\tilde{\mathcal{L}}_{n,i}(\beta_{(i)})| \xrightarrow{p} (0,\infty]$ for each $\xi > 0$, where $\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}$ $= \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||>\xi}$. By the triangle inequality and Assumption 2.b

$$\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{n,i}(\beta_{(i)})\right| - \inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{i}(\beta_{(i)})\right| \geq \inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{n,i}(\beta_{(i)}) - \tilde{\mathcal{L}}_{i}(\beta_{(i)})\right| \xrightarrow{p} 0.$$

We may therefore write for some sequence of random variables $\{r_n\}$, $r_n \geq 0$, $r_n \xrightarrow{p} 0$:

$$\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{n,i}(\beta_{(i)})\right| = \inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{i}(\beta_{(i)})\right| + r_n. \tag{C.1}$$

By the definition of $\beta^*_{(i)}$ it follows $\tilde{\mathcal{L}}_i\left(\beta_{(i)}\right) \equiv \mathcal{L}_{(i)}(\beta_{(i)}) - \mathcal{L}_{(i)}(\beta^*_{(i)}) \geq 0$, hence

$$\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left|\tilde{\mathcal{L}}_{i}(\beta_{(i)})\right| = \inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left\{\tilde{\mathcal{L}}_{i}(\beta_{(i)})\right\} = \inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left\{\mathcal{L}_{(i)}(\beta_{(i)})\right\} - \mathcal{L}_{(i)}(\beta^*_{(i)}). \tag{C.2}$$

Identification Assumption 1.c implies for each $\xi > 0$:

$$\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi}\left\{\mathcal{L}_{(i)}(\beta_{(i)})\right\} > \inf_{\beta_{(i)}\in\mathcal{B}_{(i)}}\left\{\mathcal{L}_{(i)}(\beta_{(i)})\right\}. \tag{C.3}$$

Combine (C.1)-(C.3) to conclude $\inf_{||\beta_{(i)}-\beta^*_{(i)}||>\xi} |\tilde{\mathcal{L}}_{n,i}(\beta_{(i)})| \xrightarrow{p} (0,\infty]$ as required. $\mathcal{QED}$.

C.2 Supporting Lemmas

Lemma A.1. *Let $\{\mathcal{X}_{n,i}\}_{n\geq 1}$ be random variables on a common probability space, $i = 1, 2, ...$ Then*

$$\max_{1\leq i\leq k_n} |\mathcal{X}_{n,i}| = O_p(\ln(k_n)\ln(n))$$

for any sequence of positive integers $\{k_n\}$ provided $\max_{1\leq i\leq k_n} E[\exp\{\vartheta |\mathcal{X}_{n,i}|\}] = O(n^{\xi \ln(k_n)})$ for some $\vartheta, \xi > 0$.

Proof. We need the following mapping:

$$F_\vartheta(z) \equiv \frac{1}{\vartheta}\ln\left(\sum_{i=1}^{k}\exp\{\vartheta |z_i|\}\right), \vartheta > 0. \tag{C.4}$$

Trivially $\max_{1\leq i\leq k}|z_i| \leq F_\vartheta(|z|) \leq \max_{1\leq i\leq k}|z_i| + \frac{1}{\vartheta}\ln k$ for any $\vartheta > 0$, hence $F_\vartheta(z)$ acts as a "smooth max function".[1] Then:

$$\begin{aligned} P\left(\max_{1\leq i\leq k}|\mathcal{X}_{n,i}| > \varepsilon\right) &\leq P\left(\frac{1}{\vartheta}\ln\left(\sum_{i=1}^{k}\exp\{\vartheta|\mathcal{X}_{n,i}|\}\right) > \varepsilon\right) \\ &\leq \frac{1}{\varepsilon}\frac{1}{\vartheta}E\left[\ln\left(\sum_{i=1}^{k}\exp\{\vartheta|\mathcal{X}_{n,i}|\}\right)\right] \\ &\leq \frac{1}{\varepsilon}\frac{1}{\vartheta}\ln\left(\sum_{i=1}^{k}E\left[\exp\{\vartheta|\mathcal{X}_{n,i}|\}\right]\right) \\ &\leq \frac{1}{\varepsilon}\frac{1}{\vartheta}\left(\ln(k) + \ln\left(\max_{1\leq i\leq k}E\left[\exp\{\vartheta|\mathcal{X}_{n,i}|\}\right]\right)\right). \end{aligned} \tag{C.5}$$

The second inequality is Markov's, and $|\ln(a)| = \ln(a)$ with $a \equiv \sum_{i=1}^{k}\exp\{\vartheta|\mathcal{X}_{n,i}|\} \geq k \geq 1$. The third is Jensen's inequality.

Now let $\{k_n\}$ be an arbitrary sequence of positive integers. By assumption $\max_{1\leq i\leq k_n} E[\exp\{\vartheta|\mathcal{X}_{n,i}|\}] = O(n^{\xi\ln(k_n)})$ for some $\vartheta, \xi > 0$. Combine that with (C.5) to

[1] Typically $F_\vartheta(z)$ appears under the name "*log-sum-exp*", and is used in the literatures on spin glasses (cf. Talagrand, 2011), machine learning and convex optimization (see, e.g., Boyd and Vanderberghe, 2004), and high dimensional Gausssian approximations (Chernozhukov, Chetverikov, and Kato, 2013, 2015, Chang, Chen, and Wu, 2021). However, unlike (Chernozhukov, Chetverikov, and Kato, 2013, 2015, Chang, Chen, and Wu, 2021) who seek Gaussian approximations, we do not need $\max_{1\leq i\leq k}|z_i| \approx F_\vartheta(|z|)$ hence we do not impose $\vartheta = \vartheta_n \to \infty$ as $n \to \infty$. We only use $\max_{1\leq i\leq k}|z_i| \leq F_\vartheta(|z|)$.

yield:

$$P\left(\max_{1\le i\le k_n}|\mathcal{X}_{n,i}|>\varepsilon\right)\le\frac{1}{\varepsilon}\frac{1}{\vartheta}\left(\ln(k_n)+O\left(\ln(k_n)\ln(n)\right)\right),$$

hence $\max_{1\le i\le k_n}|\mathcal{X}_{n,i}|=O_p(\ln(k_n)\ln(n))$. $\mathcal{QED}$.

Lemma A.2. *Let $\{k_n\}$ be any sequence of positive integers, $k_n=O(n)$. Let $\{w_{i,t}\}_{t=1}^{\infty}$ be random variables on a common probability space, independent across t, for $i=1,2,\ldots.$*

a. Assume either (i) $\max_{i\in\mathbb{N},1\le t\le n}|w_{i,t}|\le\mathcal{M}_n$ a.s. for non-random positive $\mathcal{M}_n=O(\ln(n))$; or (ii) $\max_{i,t\in\mathbb{N}}P(|w_{i,t}|>c)\le C\exp\{-\mathcal{K}c\}$ for some finite $C,\mathcal{K}>0$. Let $\{\eta_t\}_{t=1}^{n}$ be iid $N(0,1)$ random variables, independent of $\{w_{i,t}\}_{t=1}^{n}$, $i=1,2,\ldots$ Then

$$\max_{1\le i\le k_n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t w_{i,t}\right|=O_p\left(\ln(k_n)\ln(n)\right).$$

b. If $\max_{i,t\in\mathbb{N}}E[w_{i,t}^2]<\infty$ then

$$\max_{1\le i\le k_n}\left|\frac{1}{n}\sum_{t=1}^{n}(\eta_t^2 w_{i,t}-E[w_{i,t}])\right|=O_p(k_n/\sqrt{n}).$$

Proof.

Claim (a). Define

$$\mathcal{X}_{n,i}\equiv\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t w_{i,t}.$$

By the Cauchy-Schwartz inequality:

$$\begin{aligned}E\left[\exp\left\{\vartheta|\mathcal{X}_{n,i}|\right\}\right] &= E\left[\exp\left\{\vartheta\mathcal{X}_{n,i}\right\}I\left(\mathcal{X}_{n,i}>0\right)+\exp\left\{-\vartheta\mathcal{X}_{n,i}\right\}I\left(\mathcal{X}_{n,i}\le 0\right)\right]\\ &\le \left(E\left[\exp\left\{2\vartheta\mathcal{X}_{n,i}\right\}\right]\right)^{1/2}+\left(E\left[\exp\left\{-2\vartheta\mathcal{X}_{n,i}\right\}\right]\right)^{1/2}\end{aligned}$$

By Lemma A.1 it therefore suffices to prove $\max_{1\le i\le k_n}E[\exp\{\vartheta\mathcal{X}_{n,i}\}]=O(n^{\xi\ln(k_n)})$ for some $\vartheta,\xi>0$. The proof for $E[\exp\{-\vartheta\mathcal{X}_{n,i}\}]$ is analogous.

Let $\vartheta>0$ be an arbitrary finite constant.

Case i: Let $\max_{i\in\mathbb{N},1\le t\le n}|w_{i,t}|\le\mathcal{M}_n$ *a.s.*, $\mathcal{M}_n=O(\ln(n))$. Use $\eta_t\overset{iid}{\sim}N(0,1)$, iterated expectations, and independence for $w_{\cdot,t}$ to yield:

$$\max_{1\le i\le k_n}E\left[\exp\left\{\vartheta\mathcal{X}_{n,i}\right\}\right] = \max_{1\le i\le k_n}E\left(\prod_{t=1}^{n}E\left[\exp\left\{\vartheta\frac{1}{\sqrt{n}}\eta_t w_{i,t}\right\}|w_{i,t}\right]\right)$$

$$= \prod_{i=1}^{n} \max_{1\leq i\leq k_n} E\left[\exp\left\{2\vartheta^2 \frac{1}{n} w_{i,t}^2\right\}\right] \tag{C.6}$$

$$\leq \prod_{i=1}^{n} \exp\left\{2\vartheta^2 \frac{1}{n}\mathcal{M}_n^2\right\} = \exp\left\{2\vartheta^2\mathcal{M}_n^2\right\}.$$

Now use $\mathcal{M}_n = O(\ln(n))$ and $k_n = O(n)$ to yield $\exp\{2\vartheta^2\mathcal{M}_n^2\} = O(n^{\xi \ln(k_n)})$ for some $\xi > 0$.

Case ii: Now let $\max_{i,t\in\mathbb{N}} P(|w_{i,t}| > c) \leq C\exp\{-\mathcal{K}c\}$ for some $\mathcal{K} > 0$. Then $w_{i,t}$ has a moment generating function (uniformly in i, t), and is therefore uniformly locally sub-Gaussian (Chareka, Chareka, and Kennedy, 2006, Theorem 1). Therefore $w_{i,t}^2$ is uniformly sub-exponential by straightforward modifications to the proof of Proposition 2.5.2 in Vershynin (2018), and properties of local sub-Gaussian distributions (cf Chareka, Chareka, and Kennedy, 2006). Now use (C.6) to yield for some $\vartheta > 0$:

$$\max_{1\leq i\leq k_n} E\left[\exp\left\{\vartheta \mathcal{X}_{n,i}\right\}\right] \leq \prod_{i=1}^{n} \exp\left\{2\vartheta^2 \frac{1}{n}\mathcal{K}\right\} = \exp\left\{2\vartheta^2\mathcal{K}\right\}$$

Claim (b). Minkowski's inequality yields:

$$\left\|\max_{1\leq i\leq k_n}\left|\frac{1}{n}\sum_{t=1}^{n}(\eta_t^2 w_{i,t} - E[w_{i,t}])\right|\right\|_2 \leq k_n \max_{1\leq i\leq k_n}\left\|\frac{1}{n}\sum_{t=1}^{n}(\eta_t^2 w_{i,t} - E[w_{i,t}])\right\|_2.$$

We will prove

$$\max_{1\leq i\leq k_n}\left\|\frac{1}{n}\sum_{t=1}^{n}(\eta_t^2 w_{i,t} - E[w_{i,t}])\right\|_2 = O(1/\sqrt{n}). \tag{C.7}$$

The claim then follows by Chebyshev's inequality.

We now prove (C.7). By mutual independence between η_t and $w_{i,t}$, independence over t, $E[\eta_t^4] = 3$ and iterated expectations, it is straightforward to verify

$$\begin{aligned} E\left[\left(\frac{1}{n}\sum_{t=1}^{n}(\eta_t^2 w_{i,t} - E[w_{i,t}])\right)^2\right] &= \frac{1}{n^2}\sum_{t=1}^{n} E\left[\left(\eta_t^2 w_{i,t} - E[w_{i,t}]\right)^2\right] \\ &\quad + \frac{1}{n^2}\sum_{s\neq t}^{n} E\left[\left(\eta_s^2 w_{i,s} - E[w_{i,s}]\right)\left(\eta_t^2 w_{i,t} - E[w_{i,t}]\right)\right] \\ &= \frac{1}{n^2}\sum_{t=1}^{n}\left\{3E\left[w_{i,t}^2\right] - (E[w_{i,t}])^2\right\} \end{aligned}$$

$$\leq \ 4\frac{1}{n}\max_{1\leq i\leq k_n} E\left[w_{i,t}^2\right] \leq K\frac{1}{n}.$$

The last line uses $\max_{i,t\in\mathbb{N}} E[w_{i,t}^2] < \infty$ with the Cauchy-Schwartz inequality. This proves (C.7) as required. $\mathcal{QED}$.

Lemma A.3. *Let $\{x_{i,t}\}_{t=1}^{\infty}$ be random variables on a common probability space, $i = 1, 2, \ldots$ Assume $x_{i,t}$ are independent across t, $E[x_{i,t}] = 0\ \forall i, t$, and $1/n\sum_{t=1}^{n}\max_{i\in\mathbb{N}}\{E[x_{i,t}^2]\} = O(1)$. Then $\max_{1\leq i\leq k_n}|1/\sqrt{n}\sum_{t=1}^{n} x_{i,t}| = O_p(k_n)$ for any sequence of positive integers $\{k_n\}$.*

Proof. Under independence and $1/n\sum_{t=1}^{n}\max_{i\in\mathbb{N}}\{E[x_{i,t}^2]\} = O(1)$

$$\max_{i\in\mathbb{N}} E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right)^2\right] = O(1). \tag{C.8}$$

Exploit $\max_{1\leq i\leq k_n}|a_i| \leq \sum_{i=1}^{k_n}|a_i|$, and Chebyshev and Minkowski inequalities to yield for any $\epsilon > 0$:

$$\begin{aligned}
P\left(\max_{1\leq i\leq k_n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right| > \epsilon\right) &\leq P\left(\sum_{i=1}^{k_n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right| > \epsilon\right) \\
&\leq \frac{1}{\epsilon^2}E\left[\left(\sum_{i=1}^{k_n}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right|\right)^2\right] \\
&\leq \frac{1}{\epsilon^2}\left(\sum_{i=1}^{k_n}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right\|_2\right)^2 \\
&\leq k_n^2\frac{1}{\epsilon^2}\max_{i\in\mathbb{N}} E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{i,t}\right)^2\right] = O(k_n^2).
\end{aligned}$$

Reversing Chebyshev's inequality proves $\max_{1\leq i\leq k_n}|1/\sqrt{n}\sum_{t=1}^{n} x_{i,t}| = O_p(k_n)$ as claimed. $\mathcal{QED}$.

Lemma A.4. *Under Assumptions 2 and 3* $\max_{1\leq i\leq k_{\theta,n}}|\sqrt{n}(\hat{\beta}_{(i)} - \beta_{(i)}^{*})| = O_p((\ln(n))^2)$ *provided $k_{\theta,n} = O(\sqrt{n}/(\ln(n))^2)$.*

Proof. Recall

$$\widehat{\mathcal{H}}_{(i)}(\beta_{(i)}) \equiv \frac{\partial^2}{\partial\beta_{(i)}\partial\beta_{(i)}'}\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) \text{ and } \widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) \equiv \frac{1}{n}\widehat{\mathcal{H}}_{(i)}(\beta_{(i)}),$$

and

$$\widehat{\mathcal{G}}_{(i)} + \widehat{\mathcal{H}}_{(i)}(\ddot{\beta}_{(i)})(\hat{\beta}_{(i)} - \beta^*_{(i)}) = 0 \tag{C.9}$$

for some sequence $\{\ddot{\beta}_{(i)}\}$, $||\ddot{\beta}_{(i)} - \beta^*_{(i)}|| \leq ||\hat{\beta}_{(i)} - \beta^*_{(i)}||$. Therefore:

$$\max_{1\leq i\leq k_{\theta,n}} \left|\sqrt{n}\left(\hat{\beta}_{(i)} - \beta^*_{(i)}\right)\right| \leq \max_{1\leq i\leq k_{\theta,n}} \left|\widehat{\mathfrak{H}}^{-1}_{(i)}(\ddot{\beta}_{(i)})\right| \times \max_{1\leq i\leq k_{\theta,n}} \left|\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}\right|.$$

Recall $||\ddot{\beta}_{(i)} - \beta^*_{(i)}|| \leq ||\hat{\beta}_{(i)} - \beta^*_{(i)}|| \xrightarrow{p} 0$ by Theorem 3.1, hence $\widehat{\mathfrak{H}}^{-1}_{(i)}(\ddot{\beta}_{(i)})$ exists under Assumption 3.b *awp1*. In view of $||\ddot{\beta}_{(i)} - \beta^*_{(i)}|| \xrightarrow{p} 0$, we may write *awp1* for any $\varepsilon > 0$:

$$\begin{aligned}\left|\sqrt{n}\left(\hat{\beta}_{(i)} - \beta^*_{(i)}\right)\right| \leq & \sup_{\beta_{(i)}\in\{\mathcal{B}_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon\}} \left\{\left|\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right|\right\} \times \left|\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}\right| \\ & + \sup_{\beta_{(i)}\in\{\mathcal{B}_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon\}} \left\{\left|\widehat{\mathfrak{H}}^{-1}_{(i)}(\beta_{(i)}) - \mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right|\right\} \times \left|\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}\right|.\end{aligned} \tag{C.10}$$

Under Assumption 3.b(ii) $\max_{i\in\mathbb{N}}\{\sup_{\beta_{(i)}\in\{\mathcal{B}_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon\}} |\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})|\} < \infty$. Under Assumption 3.c, Lemma A.1 applies to $\widehat{\mathcal{G}}_{(i)}/\sqrt{n}$, hence:

$$\max_{1\leq i\leq k_{\theta,n}} \left|\frac{1}{\sqrt{n}}\widehat{\mathcal{G}}_{(i)}\right| = O_p\left(\ln\left(k_{\theta,n}\right)\ln(n)\right). \tag{C.11}$$

Now define $\mathcal{B}_{n,(i)} \equiv \{\mathcal{B}_{(i)} : ||\beta_{(i)} - \beta^*_{(i)}|| \leq \varpi_n\}$ for any sequence of positive numbers $\{\varpi_n\}_{n\geq 1}$ with $\varpi_n \to 0$. We prove below:

$$\max_{1\leq i\leq k_{\theta,n}} \left\{\sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \left|\widehat{\mathfrak{H}}^{-1}_{(i)}(\beta_{(i)}) - \mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right|\right\} = O_p\left(\frac{k_{\theta,n}\ln\left(k_{\theta,n}\right)\ln(n)}{\sqrt{n}}\right). \tag{C.12}$$

The above bounds (C.10)-(C.12) yield the desired result: for any $\{k_{\theta,n}\}$, $k_{\theta,n} = O(\sqrt{n}/(\ln(n))^2)$:

$$\begin{aligned}\max_{1\leq i\leq k_{\theta,n}} \left|\sqrt{n}\left(\hat{\beta}_{(i)} - \beta^*_{(i)}\right)\right| &= O_p\left(\ln\left(k_{\theta,n}\right)\ln(n)\right) + O_p\left(\frac{k_{\theta,n}\left(\ln\left(k_{\theta,n}\right)\right)^2\left(\ln(n)\right)^2}{\sqrt{n}}\right) \\ &= O_p\left(\ln\left(k_{\theta,n}\right)\ln(n)\right) + O_p(1) = O_p((\ln(n))^2).\end{aligned}$$

Finally, consider (C.12). We have for $\beta_{(i)} \in \mathcal{B}_{n,(i)}$:

$$\left|\widehat{\mathfrak{H}}^{-1}_{(i)}(\beta_{(i)}) - \mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right| \leq \left|\widehat{\mathfrak{H}}^{-1}_{(i)}(\beta_{(i)})\right| \times \left|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \mathfrak{H}_{(i)}(\beta_{(i)})\right| \times \left|\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right|. \tag{C.13}$$

Under Assumption 3.b it follows

$$\max_{1\leq i\leq k_{\theta,n}}\left\{\sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}}\left|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)})\right|\right\}=O_p(k_{\theta,n}) \tag{C.14}$$

$$\max_{i\in\mathbb{N}}\left\{\sup_{\beta_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon}\left|\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})\right|\right\}<\infty.$$

Further, from Assumption 3.b(ii) Lemma A.1 applies to $\sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}}\sqrt{n}|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)})-\mathfrak{H}_{(i)}(\beta_{(i)})|$:

$$\max_{1\leq i\leq k_{\theta,n}}\left\{\sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}}\left|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)})-\mathfrak{H}_{(i)}(\beta_{(i)})\right|\right\}=O_p\left(\ln\left(k_{\theta,n}\right)\ln(n)/\sqrt{n}\right). \tag{C.15}$$

Combine (C.13)-(C.15) to yield C.12, which completes the proof. $\mathcal{QED}$.

Lemma A.5. *Let $w_t:\mathcal{A}\to\mathbb{R}$, $t\in\mathbb{N}$, be well defined random functions on a common probability measure space, with compact $\mathcal{A}\subset\mathbb{R}^k$, $k\in\mathbb{N}$. Assume $w_t(\alpha)$ is almost surely continuously differentiable on $\mathcal{A}$ for all t, independent, and $\max_{t\in\mathbb{N}}E[\sup_{\alpha\in\mathcal{A}}|(\partial/\partial\alpha)^j w_t(\alpha)|^2]<\infty$ for $j=0,1$. Then $\sup_{\alpha\in\mathcal{A}}|1/n\sum_{t=1}^n\{w_t(\alpha)-E[w_t(\alpha)]\}|\xrightarrow{p}0$.*

Proof. Uniform convergence requires pointwise convergence and stochastic equicontinuity (see, e.g., Andrews, 1992). Independence and $\max_{t\in\mathbb{N}}E[\sup_{\alpha\in\mathcal{A}}|w_t(\alpha)|^2]<\infty$ yield pointwise convergence $1/n\sum_{t=1}^n\{w_t(\alpha)-E[w_t(\alpha)\}]\xrightarrow{p}0$ by Chebyshev's inequality. Stochastic equicontinuity holds by differentiability and the squared envelope bound. Let $K\equiv E[\sup_{\alpha\in\mathcal{A}}|(\partial/\partial\alpha)w_t(\alpha)|^2]$. Then $\forall\epsilon,\eta>0$ there exists $\delta\in(0,\eta\sqrt{\epsilon/K})$ such that:

$$\begin{aligned}P\left(\sup_{\|\alpha-\tilde{\alpha}\|\leq\delta}\left|\frac{1}{n}\sum_{t=1}^n(w_t(\alpha)-w_t(\tilde{\alpha}))\right|>\eta\right)&\leq P\left(\sup_{\alpha\in\mathcal{A}}\left|\frac{1}{n}\sum_{t=1}^n\frac{\partial}{\partial\alpha}w_t(\alpha)\right|>\frac{\eta}{\delta}\right)\\&\leq\frac{\delta^2}{\eta^2}E\left[\sup_{\alpha\in\mathcal{A}}\left|\frac{1}{n}\sum_{t=1}^n\frac{\partial}{\partial\alpha}w_t(\alpha)\right|^2\right]\\&\leq\frac{\delta^2}{\eta^2}\frac{1}{n^2}\left(\sum_{t=1}^n\left\|\sup_{\alpha\in\mathcal{A}}\left|\frac{\partial}{\partial\alpha}w_t(\alpha)\right|\right\|_2\right)^2=K\frac{\delta^2}{\eta^2}<\epsilon.\end{aligned}$$

The inequalities in order use the mean-value-theorem, and Chebyshev's and Minkowski's inequalities. This proves the claim. $\mathcal{QED}$.

Lemma A.6. *Under Assumption 5: (a) $\hat{\beta}^{(0)}\xrightarrow{p}\beta^{(0)}$; (b) $\hat{\beta}^{(0)}=\beta^{(0)}+O_p(1/\sqrt{n})$; and (c) $\max_{1\leq i\leq k_{\theta,n}}||\widehat{\tilde{\beta}}_{(i)}-\hat{\beta}^{(0)}_{(i)}||=O_p((\ln(n))^2/\sqrt{n})$ for any monotonically increasing sequence of positive integers $\{k_{\theta,n}\}$, $k_{\theta,n}=o(\sqrt{n})$.*

Proof. Define

$$f_0(x_t,\delta) = f(x_t,[\delta,\mathbf{0}_{k_\theta}]), \;\; g_0(x_t,\delta) \equiv \frac{\partial}{\partial\delta} f_0(x_t,\delta), \;\; h_0(x_t,\delta) \equiv \frac{\partial^2}{\partial\delta\partial\delta'} f_0(x_t,\delta)$$

Recall $\beta^{(0)} = [\delta^{(0)\prime},\mathbf{0}_{k_\theta}']'$ and $\beta_{(i)}^{(0)} = [\delta^{(0)\prime},0]'$, where $\delta^{(0)} \equiv \arg\inf_{\delta\in\mathcal{D}} E[(y_t - f(x_t[\delta,\mathbf{0}_{k_\theta}]))^2]$.

Claim (a). Define $\mathcal{L}_n^{(0)}(\delta) \equiv 1/n\sum_{t=1}^n\{y_t - f(x_t,[\delta,\mathbf{0}_{k_\theta}])\}^2$. Under Assumption 5.c, $\{y_t - f(x_t,[\delta,\mathbf{0}_{k_\theta}])\}^2$ (and therefore $\mathcal{L}_n^{(0)}(\delta)$) satisfies the conditions of ULLN Lemma A.5:

$$\sup_{\delta\in\mathcal{D}}\left|\frac{1}{n}\sum_{t=1}^n\left(\{y_t - f_0(x_t,\delta)\}^2 - E\left[\{y_t - f_0(x_t,\delta)\}^2\right]\right)\right| \xrightarrow{p} 0.$$

In order to see this, by the triangle inequality:

$$|y_t - f_0(x_t,\delta)| \leq |\epsilon_t| + |f(x_t,\beta_0) - f_0(x_t,\delta)|\,.$$

Now apply sub-additivity and Assumption 5.c to yield for some finite $C,\mathcal{K} > 0$:

$$\begin{aligned}&\max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|y_t - f_0(x_t,\delta)| > c\right)\\ &\qquad \leq \max_{t\in\mathbb{N}} P\left(|\epsilon_t| > c/2\right) + \max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|f(x_t,\beta_0) - f_0(x_t,\delta)| > c/2\right)\\ &\qquad \leq C\exp\{-\mathcal{K}c\}.\end{aligned}$$

Thus $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|y_t - f_0(x_t,\delta)|$ has sub-exponential tails and therefore a moment generating function in a neighborhood of zero, hence it is L_p-bounded for any $p > 0$. In particular, for $p > 2$:

$$\begin{aligned}\max_{t\in\mathbb{N}} E\left[\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|y_t - f_0(x_t,\delta)|^p\right] &= \max_{t\in\mathbb{N}}\int_0^\infty P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|y_t - f_0(x_t,\delta)| > c\right)dc\\ &\leq \int_0^\infty C\exp\{-\mathcal{K}c\}dc < \infty.\end{aligned}$$

Similarly:

$$\frac{\partial}{\partial\delta}\left\{y_t - f_0(x_t,\delta)\right\}^2 = -2\left\{y_t - f_0(x_t,\delta)\right\}g_0(x_t,\delta),$$

hence

$$\left|\frac{\partial}{\partial\delta}\left\{y_t - f_0(x_t,\delta)\right\}^2\right| \leq |2\epsilon_t g_0(x_t,\delta)| + |2\epsilon_t\left(f(x_t,\beta_0) - f_0(x_t,\delta)\right)|\,.$$

Under Assumption 5.c and arguments in the proof of Lemma B.1, $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |\epsilon_t g_0(x_t,\delta)|$ and $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |\epsilon_t (f(x_t,\beta_0) - f_0(x_t,\delta))|$ are sub-exponential uniformly over $i,t\in\mathbb{N}$. Therefore, by sub-additivity for some finite $C,\mathcal{K}>0$:

$$\begin{aligned}
&\max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\frac{\partial}{\partial\delta}\{y_t - f_0(x_t,\delta)\}^2\right| > c\right) \\
&\quad\le \max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} |2\epsilon_t g_0(x_t,\delta)| > c/2\right) + \max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} 2\left|\epsilon_t (f(x_t,\beta_0) - f_0(x_t,\delta))\right| > c/2\right) \\
&\quad\le C\exp\{-\mathcal{K}c\},
\end{aligned}$$

hence for $p>2$:

$$\max_{t\in\mathbb{N}} E\left[\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\frac{\partial}{\partial\delta}\{y_t - f_0(x_t,\delta)\}^2\right|^p\right] < \infty.$$

The conditions of Lemma A.5 are thus satisfied.

Furthermore, by assumption $\delta^{(0)} \equiv \arg\inf_{\delta\in\mathcal{D}}\{E[(y_t - f(x_t[\delta,\mathbf{0}_{k_\theta}]))^2]\}$ is a unique interior point of compact $\mathcal{D}$ for every t, and $\hat{\delta}^{(0)} = \arg\inf_{\delta\in\mathcal{D}}\{\mathcal{L}_n^{(0)}(\delta)\}$ exists and is $\sigma(\{x_t,y_t\}_{t=1}^n)$-measurable (Jennrich, 1969, Lemma 2). Therefore $\hat{\delta}^{(0)} \xrightarrow{p} \delta^{(0)}$ (Amemiya 1973, Lemma 3, cf. Newey and McFadden 1994, Theorem 2.1).

Claim (b). By construction $\hat{\beta}_n^{(0)} \equiv [\hat{\delta}^{(0)\prime}, \mathbf{0}_{k_\theta}']'$ where $\hat{\delta}^{(0)}$ minimizes $\sum_{t=1}^n \{y_t - f_0(x_t,\delta)\}^2$. Hence, given differentiability, $0 = \sum_{t=1}^n \{y_t - f_0(x_t,\hat{\delta}^{(0)})\} g_0(x_t,\hat{\delta}^{(0)})$. By the mean value theorem, for some $\ddot{\delta}^{(0)}$, $||\delta^{(0)} - \ddot{\delta}^{(0)}|| \le ||\delta^{(0)} - \hat{\delta}^{(0)}||$:

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{t=1}^n \left\{y_t - f_0(x_t,\delta^{(0)})\right\} g(x_t,\delta^{(0)}) \\
&\quad + \left\{-\frac{1}{n}\sum_{t=1}^n g_0(x_t,\ddot{\delta}^{(0)}) g_0(x_t,\ddot{\delta}^{(0)})' + \frac{1}{n}\sum_{t=1}^n \left(y_t - f_0(x_t,\ddot{\delta}^{(0)})\right) h_0(x_t,\ddot{\delta}^{(0)})\right\}\left(\hat{\delta}^{(0)} - \delta^{(0)}\right).
\end{aligned}$$

Under Assumption 5, Lemma A.5 yields

$$\sup_{\delta\in\mathcal{D}} \left\|\frac{1}{n}\sum_{t=1}^n \left\{(y_t - f_0(x_t,\delta))\, h_0(x_t,\delta) - E\left[(y_t - f_0(x_t,\delta))\, h_0(x_t,\delta)\right]\right\}\right\| \xrightarrow{p} 0.$$

In view of (a), $||\delta^{(0)} - \ddot{\delta}^{(0)}|| \le ||\delta^{(0)} - \hat{\delta}^{(0)}|| \xrightarrow{p} 0$. Moreover, $E[(y_t - f_0(x_t,\delta^{(0)}))h_0(x_t,\delta^{(0)})] = 0$ under Assumption 5.b. Hence $1/n\sum_{t=1}^n (y_t - f_0(x_t,\ddot{\delta}^{(0)}))h_0(x_t,\ddot{\delta}^{(0)}) \xrightarrow{p} 0$ by the mapping theorem and three times continuously differentiability of $f(\cdot,\beta)$. Under Assumption 5 and Lemma A.5, $1/n\sum_{t=1}^n g_0(x_t,\ddot{\delta}^{(0)}) g_0(x_t,\ddot{\delta}^{(0)})' \xrightarrow{p} E[g_0(x_t,\delta^{(0)}) g_0(x_t,\delta^{(0)})']$, and

$E[g_0(x_t, \delta^{(0)}) g_0(x_t, \delta^{(0)})']$ is nonsingular. Hence, *awp1* we may write:

$$\hat{\delta}^{(0)} - \delta^{(0)} = \left\{E\left[g_0(x_t, \delta^{(0)}) g_0(x_t, \delta^{(0)})'\right] + o_p(1)\right\}^{-1} \frac{1}{n}\sum_{t=1}^{n}\left\{y_t - f_0(x_t, \delta^{(0)})\right\} g(x_t, \delta^{(0)}).$$

Independence, $E[\{y_t - f_0(x_t, \delta^{(0)})\} g(x_t, \delta^{(0)})] = 0$, Assumption 5.c, and Chebyschev's inequality yield:

$$\frac{1}{n}\sum_{t=1}^{n}\left\{y_t - f_0(x_t, \delta^{(0)})\right\} g(x_t, \delta^{(0)}) = O_p(1/\sqrt{n}),$$

hence $\hat{\delta}^{(0)} - \delta^{(0)} = O_p(1/\sqrt{n})$ which proves the claim.

Claim (c).

Step 1. We first show $||\widehat{\hat{\beta}}_{(i)} - \hat{\beta}^{(0)}_{(i)}|| \xrightarrow{p} 0$ pointwise. In view of (a) it suffices to prove

$$\left\|\widehat{\hat{\beta}}_{(i)} - \beta^{(0)}_{(i)}\right\| \xrightarrow{p} 0. \tag{C.16}$$

By construction $f(x_t, \hat{\beta}^{(0)}_n) = f_{(i)}(x_t, \hat{\beta}^{(0)}_{(i)})$ where $\hat{\beta}^{(0)}_{(i)} = [\hat{\delta}^{(0)\prime}, 0]'$. Similarly $f(x_t, \beta^0) = f_{(i)}(x_t, \beta^0_{(i)})$. The bootstrap criterion therefore reduces to

$$\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) \equiv \sum_{t=1}^{n}\left\{f_{(i)}(x_t, \hat{\beta}^{(0)}_{(i)}) + \left(y_t - f_{(i)}(x_t, \hat{\beta}^{(0)}_{(i)})\right)\eta_t - f_{(i)}(x_t, \beta_{(i)})\right\}^2.$$

By the mean value theorem, for some sequence $\{\tilde{\beta}_{(i)}\}, ||\hat{\beta}^{(0)}_{(i)} - \tilde{\beta}_{(i)}|| \leq ||\hat{\beta}^{(0)}_{(i)} - \beta^{(0)}_{(i)}||$:

$$\begin{aligned}
\frac{1}{n}\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) &= \frac{1}{n}\sum_{t=1}^{n}\left\{f(x_t, \beta^0) + \left(y_t - f(x_t, \beta^0)\right)\eta_t - f_{(i)}(x_t, \beta_{(i)})\right\}^2 \\
&\quad + 2\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t, \tilde{\beta}_{(i)}) + \left(y_t - f_{(i)}(x_t, \tilde{\beta}_{(i)})\right)\eta_t - f_{(i)}(x_t, \beta_{(i)})\right\} \\
&\qquad \times \left(g_{(i)}(x_t, \tilde{\beta}_{(i)}) - g_{(i)}(x_t, \tilde{\beta}_{(i)})\eta_t\right)' \times \left(\hat{\beta}^{(0)}_{(i)} - \beta^{(0)}_{(i)}\right) \\
&= \tilde{\mathcal{L}}_{(i)}(\beta_{(i)}) + \frac{1}{n}\sum_{t=1}^{n} m_{i,t}(\tilde{\beta}_{(i)}, \beta_{(i)}) \times \left(\hat{\beta}^{(0)}_{(i)} - \beta^{(0)}_{(i)}\right),
\end{aligned}$$

say. Observe that:

$$\left\|\frac{1}{n}\sum_{t=1}^{n} m_{i,t}(\tilde{\beta}_{(i)}, \beta_{(i)})\right\| \leq \left|\frac{1}{n}\sum_{t=1}^{n}\left(\sup_{\beta_{(i)}, \tilde{\beta}_{(i)} \in \mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)}, \beta_{(i)})\right\|\right.\right.$$

$$-E\left[\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|\right]\Bigg)\Bigg|$$
$$+\left|\frac{1}{n}\sum_{t=1}^{n}E\left[\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|\right]\right|.$$

Under Assumption 5.d

$$\max_{1\leq i\leq k_{\theta,n}}\left|\frac{1}{n}\sum_{t=1}^{n}E\left[\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|\right]\right|\leq\max_{i,t\in\mathbb{N}}\left|E\left[\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|\right]\right|<\infty,$$

and by Lemma A.3

$$\max_{1\leq i\leq k_{\theta,n}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|-E\left[\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}}\left\|m_{i,t}(\tilde{\beta}_{(i)},\beta_{(i)})\right\|\right]\right)\right|=O_p(k_{\theta,n}/\sqrt{n}).$$

Hence given $k_{\theta,n}/\sqrt{n}\rightarrow 0$:

$$\max_{1\leq i\leq k_{\theta,n}}\left|\frac{1}{n}\hat{\mathcal{L}}_{(i)}(\beta_{(i)})-\tilde{\mathcal{L}}_{(i)}(\beta_{(i)})\right|\xrightarrow{p}0.$$

Repeat the proof of (a) to yield $||\widehat{\tilde{\beta}}_{(i)}-\hat{\beta}_{(i)}^{(0)}||\xrightarrow{p}0$ as desired.

Step 2. Recall $\epsilon_{n,t}^{(0)}\equiv y_t-f(x_t,\hat{\beta}_n^{(0)})$ and $y_{n,t}^{*}\equiv f(x_t,\hat{\beta}_n^{(0)})+\epsilon_{n,t}^{(0)}\eta_t$. Hence $\widehat{\tilde{\beta}}_{(i)}$ satisfies by construction:

$$0=\frac{1}{n}\sum_{t=1}^{n}\eta_t\left(y_t-f(x_t,\hat{\beta}_n^{(0)})\right)g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-f(x_t,\hat{\beta}_n^{(0)})\right\}g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)}).$$

We first expand around $\hat{\beta}_n^{(0)}=[\hat{\delta}^{(0)\prime},\mathbf{0}_{k_{\theta,n}}^{\prime}]^{\prime}$. Recall $f(x_t,\hat{\beta}_n^{(0)})=f_0(x_t,\hat{\delta}_{(i)}^{(0)})$. By the mean value theorem, for some $\ddot{\delta}^{(0)}$, $||\delta^{(0)}-\ddot{\delta}^{(0)}||\leq||\delta^{(0)}-\hat{\delta}^{(0)}||$:

$$\begin{aligned}0\;&=\;\frac{1}{n}\sum_{t=1}^{n}\eta_t\left(y_t-f_0(x_t,\hat{\delta}^{(0)})\right)g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-f_0(x_t,\hat{\delta}^{(0)})\right\}g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})\\&=\;\frac{1}{n}\sum_{t=1}^{n}\eta_t\left(y_t-f_0(x_t,\delta^{(0)})\right)g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})-f_0(x_t,\delta^{(0)})\right\}g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})\\&\quad+\left\{-\frac{1}{n}\sum_{t=1}^{n}\eta_t g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})^{\prime}+\frac{1}{n}\sum_{t=1}^{n}g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})^{\prime}\right\}\left(\hat{\delta}^{(0)}-\delta^{(0)}\right).\end{aligned}$$

Under Assumption 5.d,

$$E\left|\frac{1}{n}\sum_{t=1}^{n} g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})'\right| \leq \max_{i,t\in\mathbb{N}} \sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} E\left|g_{(i)}(x_t,\beta_{(i)})g_{(i)}(x_t,\beta_{(i)})'\right| < \infty$$

hence by Markov's inequality:

$$\frac{1}{n}\sum_{t=1}^{n} g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})' = O_p(1).$$

Furthermore, by Assumption 5.d $g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})'$ satisfies the conditions of Lemma A.2.a. Combine that with $k_{\theta,n} = o(\sqrt{n})$ to yield:

$$\left\|\frac{1}{n}\sum_{t=1}^{n} \eta_t g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)})g_0(x_t,\ddot{\delta}^{(0)})'\right\| = O_p\left(\ln(k_{\theta,n})\ln(n)/\sqrt{n}\right) = O_p\left((\ln(n))^2/\sqrt{n}\right),$$

where $O_p(\cdot)'s$ terms here and below are not a function of $\widehat{\tilde{\beta}}_{(i)}$ or $i = 1,...,k_{\theta,n}$. Now use $\hat{\delta}^{(0)} - \delta^{(0)} = O_p(1/\sqrt{n})$ from (b) to yield:

$$\begin{aligned}0 &= \frac{1}{n}\sum_{t=1}^{n} \eta_t\left(y_t - f_0(x_t,\delta^{(0)})\right) g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)}) \\ &\quad -\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)}) - f_0(x_t,\delta^{(0)})\right\} g_{(i)}(x_t,\widehat{\tilde{\beta}}_{(i)}) + O_p(1/\sqrt{n}).\end{aligned} \tag{C.17}$$

Next, recall $g_{(i)}(x_t,\beta_{(i)}) \equiv (\partial/\partial\beta_{(i)})f(x_t,\beta_{(i)})$ and define $h_{(i)}(x_t,\beta_{(i)}) \equiv (\partial/\partial\beta_{(i)})^2 f(x_t,\beta_{(i)})$. Expand terms in (C.17) with $\widehat{\tilde{\beta}}_{(i)}$ around $\hat{\beta}_{(i)}^{(0)}$ yielding for some $\tilde{\beta}_{(i)}^*$, $||\hat{\beta}_{(i)}^{(0)} - \tilde{\beta}_{(i)}^*|| \leq ||\hat{\beta}_{(i)}^{(0)} - \widehat{\tilde{\beta}}_{(i)}||$:

$$\begin{aligned}0 &= \frac{1}{n}\sum_{t=1}^{n} \eta_t\left(y_t - f_0(x_t,\delta^{(0)})\right) g_{(i)}(x_t,\hat{\beta}_{(i)}^{(0)}) \\ &\quad -\frac{1}{n}\sum_{t=1}^{n}\left\{f_{(i)}(x_t,\hat{\beta}_{(i)}^{(0)}) - f_0(x_t,\delta^{(0)})\right\} g_{(i)}(x_t,\hat{\beta}_{(i)}^{(0)}) \\ &\quad +\left\{\frac{1}{n}\sum_{t=1}^{n}\eta_t\left(y_t - f_{(i)}(x_t,\hat{\beta}_{(i)}^{(0)})\right) h_{(i)}(x_t,\tilde{\beta}_{(i)}^*)\right. \\ &\qquad \left.-\frac{1}{n}\sum_{t=1}^{n} g_{(i)}(x_t,\hat{\beta}_{(i)}^{(0)})g_{(i)}(x_t,\tilde{\beta}_{(i)}^*)\right\}\left(\widehat{\tilde{\beta}}_{(i)} - \hat{\beta}_{(i)}^{(0)}\right) + O_p(1)/\sqrt{n})\end{aligned} \tag{C.18}$$

$$= \widehat{\mathfrak{W}}_{(i)} - \widehat{\mathfrak{X}}_{(i)} + \left\{ \widehat{\mathfrak{Y}}_{(i)} - \widehat{\mathfrak{G}}_{(i)} \right\} \left(\hat{\tilde{\beta}}_{(i)} - \hat{\beta}_{(i)}^{(0)} \right) + O_p\left(1/\sqrt{n}\right).$$

Note η_t is iid $N(0,1)$ distributed, and independent of the sample $\{x_t, y_t\}_{t=1}^n$. Under the Assumption 5.c bounds, $|y_t - f_0(x_t, \delta^{(0)})| \times \sup_{\beta_{(i)} \in \mathcal{B}_{(i)}} ||g_{(i)}(x_t, \beta_{(i)})||$ and $\sup_{\delta \in \mathcal{D}} |y_t - f_0(x_t, \delta)| \times \sup_{\beta_{(i)} \in \mathcal{B}_{(i)}} ||h_{(i)}(x_t, \beta_{(i)})||$ satisfy the conditions of Lemma A.2.a. Hence with $k_{\theta,n} = o(\sqrt{n})$ it follows $\max_{1 \leq i \leq k_{\theta,n}} ||\widehat{\mathfrak{W}}_{(i)}||$ and $\max_{1 \leq i \leq k_{\theta,n}} ||\widehat{\mathfrak{Y}}_{(i)}||$ are $O_p((\ln(n))^2/\sqrt{n})$.

Next $\widehat{\mathfrak{X}}_{(i)}$. Recall:

$$\bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}, \tilde{\beta}_{(i)}) \equiv \frac{1}{n} \sum_{t=1}^n g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \tilde{\beta}_{(i)})' \text{ and } \bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}) \equiv \bar{\mathfrak{H}}_{n,(i)}(\beta_{(i)}, \beta_{(i)})$$

$$\bar{\mathcal{H}}_{n.(i)}(\beta_{(i)}) \equiv \frac{1}{n} \sum_{t=1}^n E[g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \beta_{(i)})']$$

$$\bar{\mathcal{H}}_{(i)}(\beta_{(i)}) \equiv \max_{t \in \mathbb{N}} E\left[g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \beta_{(i)})'\right]$$

By construction $f_0(x_t, \delta^{(0)}) = f_{(i)}(x_t, \beta_{(i)}^{(0)})$ with $\beta_{(i)}^{(0)} = [\delta^{(0)\prime}, 0]'$. Further, $\max_{1 \leq i \leq k_{\theta,n}} ||\hat{\beta}_{(i)}^{(0)} - \beta_{(i)}^{(0)}|| = O_p(1/\sqrt{n})$ by (b) coupled with $\hat{\beta}_{(i)}^{(0)} - \beta_{(i)}^{(0)} = \hat{\beta}_{(j)}^{(0)} - \beta_{(j)}^{(0)}$ $\forall i, j$. Define $w_{t,i} \equiv \sup_{\beta_{(i)}, \tilde{\beta}_{(i)} \in \mathcal{B}_{(i)}} ||g_{(i)}(x_t, \beta_{(i)}) g_{(i)}(x_t, \tilde{\beta}_{(i)})||$. Under Assumption 5.d $w_{t,i} - E[w_{t,i}]$ satisfies Lemma A.3, and $\max_{1 \leq i \leq k_{\theta,n}} |1/n \sum_{t=1}^n E|w_{t,i}| \leq \max_{i,t \in \mathbb{N}} E|w_{t,i}| < \infty$. By a first order expansion, and adding and subtracting like terms, we therefore obtain:

$$\begin{aligned} \max_{1 \leq i \leq k_{\theta,n}} \left\| \widehat{\mathfrak{X}}_{(i)} \right\| &\leq \max_{1 \leq i \leq k_{\theta,n}} \left\| \frac{1}{n} \sum_{t=1}^n g_{(i)}(x_t, \hat{\beta}_{(i)}^{(0)}) g_{(i)}(x_t, \beta_{(i)}^{*(0)}) \right\| \\ &\quad \times \max_{1 \leq i \leq k_{\theta,n}} \left\| \hat{\beta}_{(i)}^{(0)} - \beta_{(i)}^{(0)} \right\| \\ &\leq \left(\max_{1 \leq i \leq k_{\theta,n}} \left| \frac{1}{n} \sum_{t=1}^n (w_{t,i} - E[w_{t,i}]) \right| + \max_{1 \leq i \leq k_{\theta,n}} \left| \frac{1}{n} \sum_{t=1}^n E[w_{t,i}] \right| \right) \\ &\quad \times \max_{1 \leq i \leq k_{\theta,n}} \left\| \hat{\beta}_{(i)}^{(0)} - \beta_{(i)}^{(0)} \right\| \\ &= O_p(1) \times O_p(1/\sqrt{n}) = O_p(1/\sqrt{n}). \end{aligned} \tag{C.19}$$

Therefore:

$$O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right) = \left\{ \frac{1}{n} \sum_{t=1}^n g_{(i)}(x_t, \hat{\beta}_{(i)}^{(0)}) g_{(i)}(x_t, \tilde{\beta}_{(i)}^*) \right. \tag{C.20}$$

$$+O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right)\Bigg\}\left(\widehat{\tilde{\beta}}_{(i)}-\hat{\beta}^{(0)}_{(i)}\right).$$

By Assumption 5.d, (a) and (C.16), $\widehat{\mathfrak{G}}_{(i)}=1/n\sum_{t=1}^n g_{(i)}(x_t,\hat{\beta}^{(0)}_{(i)})g_{(i)}(x_t,\tilde{\beta}^*_{(i)})$ is positive definite $a.s.$ as $n\to\infty$, uniformly over $1\leq i\leq k_{\theta,n}$. This yields *awp1*:

$$\begin{aligned}\widehat{\tilde{\beta}}_{(i)}-\hat{\beta}^{(0)}_{(i)} &= \left\{\frac{1}{n}\sum_{t=1}^n g_{(i)}(x_t,\hat{\beta}^{(0)}_{(i)})g_{(i)}(x_t,\tilde{\beta}^*_{(i)})+O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right)\right\}^{-1}\times O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right) \quad \text{(C.21)}\\ &= \left\{\bar{\mathfrak{H}}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})+O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right)\right\}^{-1}\times O_p\left(\frac{(\ln(n))^2}{\sqrt{n}}\right).\end{aligned}$$

It remains to tackle $\bar{\mathfrak{H}}_{n,(i)}(\cdot,\cdot)$. Observe:

$$\begin{aligned}\left\|\bar{\mathfrak{H}}^{-1}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})\right\| &\leq \left\|\bar{\mathfrak{H}}^{-1}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})\right\|\left\|\bar{\mathfrak{H}}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})-\bar{\mathfrak{H}}_{n,(i)}(\beta^{(0)}_{(i)})\right\|\left\|\bar{\mathfrak{H}}^{-1}_{n,(i)}(\beta^{(0)}_{(i)})\right\|\\ &\quad+\left\|\bar{\mathcal{H}}^{-1}_{(i)}(\beta^{(0)}_{(i)})\right\|\left\|\bar{\mathcal{H}}^{-1}_{(i)}(\beta^{(0)}_{(i)})\right\|\left\|\bar{\mathfrak{H}}_{n,(i)}(\beta^{(0)}_{(i)},\beta^{(0)}_{(i)})-\bar{\mathcal{H}}_{(i)}(\beta^{(0)}_{(i)})\right\|\\ &\quad+\left\|\bar{\mathcal{H}}^{-1}_{(i)}(\beta^{(0)}_{(i)})\right\|.\end{aligned}$$

Under Assumption 5.d $\max_{i\in\mathbb{N}}||\bar{\mathcal{H}}^{-1}_{(i)}(\beta^{(0)}_{(i)},\beta^{(0)}_{(i)})||<\infty$ and:

$$\max_{1\leq i\leq k_{\theta,n}}\left\|\bar{\mathfrak{H}}^{-1}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})\right\|\leq\max_{1\leq i\leq k_{\theta,n}}\sup_{\beta_{(i)},\tilde{\beta}_{(i)}\in\mathcal{B}_{(i)}\times\tilde{\mathcal{B}}_{(i)}}\left\|\bar{\mathfrak{H}}^{-1}_{n,(i)}(\beta_{(i)},,\tilde{\beta}_{(i)})\right\|=O_p(k_{\theta,n})$$

Finally, invoke Lemma A.3, $k_{\theta,n}=o(\sqrt{n})$, and the argument leading to (C.19) to obtain respectively

$$\max_{1\leq i\leq k_{\theta,n}}\left\|\bar{\mathfrak{H}}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})-\bar{\mathfrak{H}}_{n,(i)}(\beta^{(0)}_{(i)})\right\|=O_p(k_{\theta,n}/\sqrt{n})=o_p(1)$$

$$\max_{1\leq i\leq k_{\theta,n}}\left\|\bar{\mathfrak{H}}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})-\bar{\mathfrak{H}}_{n,(i)}(\beta^{(0)}_{(i)})\right\|=O_p(1).$$

Therefore $\max_{1\leq i\leq k_{\theta,n}}||\bar{\mathfrak{H}}^{-1}_{n,(i)}(\hat{\beta}^{(0)}_{(i)},\tilde{\beta}^*_{(i)})||=O_p(1)$. In view of (C.21) this proves as claimed $\max_{1\leq i\leq k_{\theta,n}}||\widehat{\tilde{\beta}}_{(i)}-\hat{\beta}^{(0)}_{(i)}||=O_p((\ln(n))^2/\sqrt{n})$. $\mathcal{QED}$.

D Examples

We present the linear regression example from the main paper with a proof of Lemma 5.1. We then give a logistic regression example.

D.1 Linear Regression

We verify all assumptions for a linear regression model:

$$y_t = \delta_0' x_{\delta,t} + \theta_0' x_{\theta,t} + \epsilon_t = \beta_0' x_t + \epsilon_t.$$

Let $x_{\delta,t} \in \mathbb{R}^{k_\delta}$, $x_{\theta,t} \in \mathbb{R}^{k_\theta}$, and $E[\epsilon_t] = 0$. Assume $\min_{t\in\mathbb{N}} E[\epsilon_t^2] > 0$ and $\min_{i,t\in\mathbb{N}} E[x_{i,t}^2] > 0$. Assume $\{x_t, \epsilon_t\}$ are independent across t and mutually independent for unique $\beta_0 = [\delta_0', \theta_0']'$. δ_0 and $\theta_{0,i}$ are interior points of compact $\mathcal{D} \subset \mathbb{R}^{k_\delta}$, $k_\delta \in \mathbb{N}$, and $\Theta_i \subset \mathbb{R}$. Assume $\Theta \subset \{\times_{i=1}^{k_\theta} \Theta_i : |\theta| < \infty\}$. In a linear framework we need $|\theta| < \infty$ to ensure $\sup_{\beta\in\mathcal{B}} ||\beta' x_t||_p < \infty$ for some $p > 4$ under general conditions on x_t, which is used to verify several conditions in bootstrap Assumption 5.d. The bound is trivial when $k_\theta < \infty$, and covers sparse (i.e. $\sum_{i=1}^{k_\theta} I(\theta_i \neq 0) < \infty$) and nonsparse (e.g. $\theta_i = O(\rho^i)$, $|\rho| < 1$) cases. However, we do not require $|\theta| < \infty$ with logistic or similar "squash"-like response. See Appendix D.2, below.

Let ϵ_t and $x_{i,t}$ be uniformly sub-exponential $\max_{t\in\mathbb{N}} P(|\epsilon_t| > c) \leq \mathcal{C}\exp\{-\mathcal{K}c\}$ and $\max_{i,t\in\mathbb{N}} P(|x_{i,t}| > c) \leq \exp\{-\mathcal{K}c\}$ for some finite $\mathcal{C}, \mathcal{K} > 0$ that may be different in different places. Assume all pairs $(x_{i,t}, x_{j,t} : i \neq j)$ are conditionally sub-exponential: $\max_{i,j,t\in\mathbb{N}} P(|x_{i,t}| > c|x_{j,t}) \leq \max_{j,t\in\mathbb{N}} \mathcal{C}(x_{j,t}) \exp\{-Kc\}$ $a.s.$ for some $\sigma(x_{j,t})$-measurable random variable $\mathcal{C}(x_{j,t})$ with $E[\max_{j,t\in\mathbb{N}} \mathcal{C}(x_{j,t})^2] < \infty$.

We want to test $H_0 : \theta_0 = 0$. The parsimonious models are for $i = 1, ..., k_{\theta,n}$:

$$y_t = \delta_{(i)}^{*\prime} x_{\delta,t} + \theta_i^* x_{\theta,i,t} + v_{(i),t} = \beta_{(i)}^{*\prime} x_{(i),t} + v_{(i),t},$$

where $E[v_{(i),t} x_{(i),t}] = 0$ for unique $\beta_{(i)}^*$ in the interior of $\mathcal{B}_{(i)} \equiv \mathcal{D} \times \Theta_i$. Squared error loss is used: $\mathcal{L}(\beta) = .5E[(y_t - \beta' x_t)^2]$ and $\mathcal{L}_{(i)}(\beta) = .5E[(y_t - \beta_{(i)}' x_{(i).t})^2]$. The estimator is computed by least squares with criterion $\hat{\mathcal{L}}_{(i)}(\beta) \equiv \sum_{t=1}^n (y_t - \beta_{(i)}' x_{(i).t})^2$, hence $\widehat{\mathcal{G}}_{(i)} = -\sum_{t=1}^n v_{(i),t} x_{(i),t}$, $\widehat{\mathcal{H}}_{(i)} = \sum_{t=1}^n x_{(i),t} x_{(i),t}'$, $\widehat{\mathfrak{H}}_{(i)} = \bar{\mathfrak{H}}_{n,(i)} = \widehat{\mathcal{H}}_{(i)}/n$, and therefore $\mathcal{H}_{(i)} = \bar{\mathcal{H}}_{(i)} = \mathfrak{H}_{(i)} = \lim_{n\to\infty} 1/n \sum_{t=1}^n E[x_{(i),t} x_{(i),t}']$.

Assume $\inf_{\lambda'\lambda=1} \min_{1\leq i\leq k_{\theta,n}} \{1/n \sum_{t=1}^n (\lambda' x_{(i),t})^2\} > 0$ $a.s.$, and $\max_{1\leq i\leq k_{\theta,n}} ||(1/n \sum_{t=1}^n x_{(i),t} x_{(i),t}')^{-1}|| = O_p(k_{\theta,n})$. The latter holds, for example, when there are $k_\delta = 0$ nuisance parameters and $\min_{i\in\mathbb{N}} \liminf_{n\to\infty} 1/n \sum_{t=1}^n x_{i,t}^2 > 0$ $a.s.$ Finally, we need to assume

$$\max_{1\leq i,j\leq k_{\theta,n}} \left| \frac{1}{n} \sum_{t=1}^n E\left[\epsilon_t^2\right] E\left[x_{i,t}^2\right] - \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n E\left[\epsilon_t^2\right] E\left[x_{i,t}^2\right] \right| = O(k_{\theta,n}/\sqrt{n}) \quad \text{(D.1)}$$

$$\max_{1\leq i,j\leq k_{\theta,n}} \left| \frac{1}{n} \sum_{t=1}^n E\left[(\beta_0' x_t)^2 x_{i,t}^2\right] - \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n E\left[(\beta_0' x_t)^2 x_{i,t}^2\right] \right| = O(k_{\theta,n}/\sqrt{n}).$$

Each restricts heterogeneity, and is trivial when $(\epsilon_t, x_{i,t})$ are identically distributed across t.

Lemma 5.1. *In the setting above, Assumptions 1-5 hold.*

Proof.

Assumption 1. $E[\epsilon_t|x_t] = 0$ *a.s.* implies $(\partial/\partial\beta)\mathcal{L}(\beta) = -E[(y_t - \beta' x_t)x_t] = 0$ *if and only if* $\beta = \beta_0$, hence (B.1). Similarly, $(\partial/\partial\beta_{(i)})\mathcal{L}_{(i)}(\beta_{(i)}) = -E[(y_t - \beta'_{(i)}x_{(i),t})x_{(i),t}] = 0$ *if and only if* $\beta_{(i)} = \beta_{(i)}^{(*)}$ by assumption, hence (B.2). Lastly, (B.3) holds by construction, hence each condition of Assumption 1 holds.

Assumption 2. Recall

$$\mathcal{L}_{(i)}(\beta_{(i)}) = \frac{1}{2}E\left[\left(y_t - \beta'_{(i)}x_{(i),t}\right)^2\right] \text{ and } \frac{1}{n}\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) = \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \beta'_{(i)}x_{(i),t}\right)^2,$$

Assumption 2.a $\hat{\mathcal{L}}_{(i)}(\hat{\beta}_{(i)}) = \min_{\beta_{(i)}\in\mathcal{B}_{(i)}} \hat{\mathcal{L}}_{(i)}(\beta_{(i)}\}$ is trivial by linearity of the regression response, and quadratic loss.

Consider Assumption 2.b. By construction of θ_i^*:

$$\begin{aligned} 2\left|\hat{\mathcal{L}}_{(i)}(\beta_{(i)})/n - \mathcal{L}_{(i)}(\beta_{(i)})\right| \leq & \left|\frac{1}{n}\sum_{t=1}^{n}\left(v_{(i),t}^2 - E\left[v_{(i),t}^2\right]\right)\right| + 2\left|\beta_{(i)}^* - \beta_{(i)}\right|\left|\frac{1}{n}\sum_{t=1}^{n}v_{(i),t}x_{(i),t}\right| \\ & + \left|\beta_{(i)}^* - \beta_{(i)}\right|^2\left|\frac{1}{n}\sum_{t=1}^{n}\left(x_{(i),t}x'_{(i),t} - E\left[x_{(i),t}x'_{(i),t}\right]\right)\right|. \end{aligned}$$

Independence and the moment properties, and compactness of $\mathcal{B}_{(i)}$, together yield:

$$\begin{aligned} 2\times\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}\left|\hat{\mathcal{L}}_{(i)}(\beta_{(i)})/n - \mathcal{L}_{(i)}(\beta_{(i)})\right| \leq & \left|\frac{1}{n}\sum_{t=1}^{n}\left(v_{(i),t}^2 - E\left[v_{(i),t}^2\right]\right)\right| + K\left|\frac{1}{n}\sum_{t=1}^{n}v_{(i),t}x_{(i),t}\right| \\ & + K\left|\frac{1}{n}\sum_{t=1}^{n}\left(x_{(i),t}x'_{(i),t} - E\left[x_{(i),t}x'_{(i),t}\right]\right)\right| \\ & \xrightarrow{p} 0. \end{aligned}$$

In order to reduce notation, for the remainder of the proof we assume $k_\delta = 0$, hence $y_t = \theta_i^* x_{(i),t} + v_{(i),t}$ where $x_{(i),t} = x_{i,t}$

Assumption 3. Let H_0 hold. Hessian Lipschitz property **(a)** is trivial under linearity. Hessian bounds **(b.i)** hold under the stated conditions and by linearity.

Consider exponential moment bounds **(b.ii)**. Observe for any random variable x, by the Cauchy-Schwartz inequality:

$$E\left[\exp\{a|x|\}\right] = E\left[\exp\{ax\}I\left(x > 0\right) + \exp\{-ax\}I\left(x \leq 0\right)\right]$$

$$\leq \quad \left(E\left[\exp\{2ax\}\right]\right)^{1/2} + \left(E\left[\exp\{-2ax\}\right]\right)^{1/2}.$$

It therefore suffices to prove for some $\zeta > 0$:

$$\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right] = O\left(n^{\xi \ln(k_n)}\right),$$

The proof where ζ is replaced with $-\zeta$ is analogous.

By independence:

$$\begin{aligned}\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right] &= \max_{1 \leq i \leq k_{\theta,n}} \prod_{t=1}^{n} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right] \\ &\leq \left(\max_{i,t \in \mathbb{N}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right]\right)^n\end{aligned}$$

If $x_{i,t}$ has bounded support then it is sub-Gaussian. Otherwise, if $\max_{i,t \in \mathbb{N}} P(|x_{i,t}| > c) \leq C \exp\{-\mathcal{K}c\}$ then $x_{j,t}$ has a moment generating function (uniformly in i, t), and is therefore locally sub-Gaussian (Chareka, Chareka, and Kennedy, 2006, Theorem 1). Either way, for some $\zeta > 0$, and some universal constant $\sigma^2 > 0$ (Vershynin, 2018, Proposition 2.5.2(iii)):

$$\max_{i,t \in \mathbb{N}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right] \leq \exp\left\{\zeta^2 \sigma^2 \frac{1}{n}\right\}.$$

Hence

$$\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{\zeta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(x_{i,t}^2 - E\left[x_{i,t}^2\right]\right)\right\}\right] \leq \exp\left\{\zeta^2 \sigma^2\right\},$$

which proves (b.ii).

Condition (c) holds under the stated conditions by Corollary B.2.

Assumption 4. Let H_0 hold.

a. Recall under least squares, and $k_\delta = 0$, $z_{(i),t}(\lambda) = -\mathfrak{H}_{(i)}^{-1} \epsilon_t x_{i,t}$. By assumption $z_{(i),t}(\lambda)$ is independent across t. Further, since ϵ_t an $x_{i,t}$ are mutually independent:

$$\max_{i,t \in \mathbb{N}} \left\| \mathfrak{H}_{(i)}^{-1} \epsilon_t x_{i,t} \right\|_2 \leq \left(\frac{1}{\min_{i \in \mathbb{N}} E\left[x_{i,t}^2\right]} \max_{t \in \mathbb{N}} E\left[\epsilon_t^2\right] \max_{i,t \in \mathbb{N}} E\left[x_{i,t}^2\right] \right)^{1/2}.$$

The (uniform) exponential tail bounds imply a moment generating exists (uniformly), hence $\max_{t \in \mathbb{N}} E\left[\epsilon_t^2\right] < \infty$ and $\max_{i,t \in \mathbb{N}} E\left[x_{i,t}^2\right] < \infty$. Moreover, $\min_{i \in \mathbb{N}} E\left[x_{i,t}^2\right] > 0$ by supposition. Hence $\max_{i,t \in \mathbb{N}} ||\mathfrak{H}_{(i)}^{-1} \epsilon_t x_{i,t}||_2 < \infty$.

b. By independence, supposition, $k_\delta = 0$ (hence $\lambda = 1$) and $\max_{i,t\in\mathbb{N}} E\left[x_{i,t}^2\right] < \infty$:

$$\begin{aligned}\limsup_{n\to\infty} \max_{i\in\mathbb{N}} \sup_{\lambda'\lambda=1} \sigma_{n,i}^2(\lambda) &= \liminf_{n\to\infty} \max_{i\in\mathbb{N}} E\left[\hat{\mathcal{Z}}_{(i)}^2\right] \\ &\leq \frac{1}{\left(\min_{i\in\mathbb{N}} E\left[x_{i,t}^2\right]\right)^2} \max_{t\in\mathbb{N}} E\left[\epsilon_t^2\right] \max_{i,t\in\mathbb{N}} E\left[x_{i,t}^2\right] \equiv \bar{c} < \infty,\end{aligned}$$

and

$$\begin{aligned}\liminf_{n\to\infty} \min_{i\in\mathbb{N}} \inf_{\lambda'\lambda=1} \sigma_{n,i}^2(\lambda) &= \liminf_{n\to\infty} \min_{i\in\mathbb{N}} \frac{1}{\mathfrak{H}_{(i)}^2} \frac{1}{n} \sum_{t=1}^n E\left[\epsilon_t^2\right] E\left[x_{i,t}^2\right] \\ &\geq \frac{1}{\max_{i\in\mathbb{N}} \mathfrak{H}_{(i)}^2} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^n E\left[\epsilon_t^2\right] \min_{i\in\mathbb{N}} E\left[x_{i,t}^2\right] \\ &\geq \frac{\min_{t\in\mathbb{N}} E\left[\epsilon_t^2\right] \min_{i,t\in\mathbb{N}} E\left[x_{i,t}^2\right]}{\max_{i,t\in\mathbb{N}} E\left[x_{i,t}^2\right]} \equiv \underline{c} > 0.\end{aligned}$$

c. We need to show for some sequence of non-random numbers $\{B_n\}$, $B_n \geq 1$:

$$\max_{i,t\in\mathbb{N}} \left\{ \max_{\gamma=1,2} \frac{1}{n} \sum_{t=1}^n \frac{1}{B_n^\gamma} E\left|z_{(i),t}(\lambda)\right|^{2+\gamma} + E\left[\exp\left\{\frac{\left|z_{(i),t}(\lambda)\right|}{B_n}\right\}\right]\right\} \leq 4, \tag{D.2}$$

where, as above, $\lambda = 1$. Write $z_{(i),t} = z_{(i),t}(\lambda)$. By mutual independence, uniform sub-exponential tail decay and iterated expectations, the moment generating function satisfies for some ζ (cf. Vershynin, 2018, Proposition 2.7.1):

$$\begin{aligned}\max_{i,t\in\mathbb{N}} E\left[\exp\left\{\zeta\mathfrak{H}_{(i)}^{-1}\left|\epsilon_t x_{i,t}\right|\right\}\right] &= \max_{i,t\in\mathbb{N}} E\left(E\left[\exp\left\{\zeta\mathfrak{H}_{(i)}^{-1}\left|\epsilon_t\right| \times \left|x_{i,t}\right|\right\}\left|x_{i,t}\right.\right]\right) \\ &\leq \max_{i,t\in\mathbb{N}} E\left[\exp\left\{K\zeta\left|x_{i,t}\right|\right\}\right] \\ &\leq \exp\left\{K\zeta\right\}.\end{aligned}$$

Therefore $z_{(i),t} = -\mathfrak{H}_{(i)}^{-1}\epsilon_t x_{i,t}$ has sub-exponential tails uniformly in i, t. Hence $\max_{t\in\mathbb{N}} E[\exp\{\left|z_{(i),t}\right|\}] < \infty$, and for any $\gamma > 0$:

$$\frac{1}{n}\sum_{t=1}^n E\left|z_{(i),t}\right|^{2+\gamma} \leq \max_{i,t\in\mathbb{N}} E\left|z_{(i),t}\right|^{2+\gamma} < \infty.$$

We may therefore set $B_n = 1$ to yield (D.2). See also Remark 4.1 following Assumption 4 in the main paper.

Assumption 5.

a. and **b.** hold by linearity, cf. verification of Assumption 1 above.

c. Sub-exponentiality for the error $\max_{t\in\mathbb{N}} P(|\epsilon_t| > c) \leq C\exp\{-\mathcal{K}c\}$ holds by supposition. Now let $w_{i,t}$ denote

$$\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|g_{(i)}(x_t, \beta_{(i)})\right|, \text{ or}$$

$$\sup_{\delta\in\mathcal{D}} |f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])|, \text{ or}$$

$$\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\} g_{(i)}(x_t, \beta_{(i)})\right|, \text{ or}$$

$$\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\} h_{(i)}(x_t, \beta_{(i)})\right|.$$

The final term is identically zero because $h_{(i)}(x_t, \beta_{(i)}) = 0$ under linearity.

We need to show the remaining $w_{i,t}$ satisfy for some finite $C, K, \mathcal{K} > 0$ that may be different in different places and for different $w_{i,t}$:

$$\max_{i,t\in\mathbb{N}} P(w_{i,t} > c) \leq C\exp\{-\mathcal{K}c\}. \tag{D.3}$$

First, $g_{(i)}(x_t, \beta_{(i)}) = x_{(i),t}$. By sub-additivity, and uniform sub-exponentiality by supposition, for some finite $K > 0$:

$$\begin{aligned}
\max_{i,t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|g_{(i)}(x_t, \beta_{(i)})\right| > c\right) &\leq \max_{i,t\in\mathbb{N}} P(|x_{\delta,t}| + |x_{i,\theta,t}| > c) \\
&\leq \sum_{j=1}^{k_\delta} \max_{t\in\mathbb{N}} P\left(|x_{j,\delta,t}| > \frac{c}{k_\delta + 1}\right) \\
&\qquad + \max_{i,t\in\mathbb{N}} P\left(|x_{i,\theta,t}| > \frac{c}{k_\delta + 1}\right) \\
&\leq (k_\delta + 1)\max_{i,t\in\mathbb{N}} P\left(|x_{i,t}| > \frac{c}{k_\delta + 1}\right) \\
&\leq \mathcal{C}\exp\{-\mathcal{K}c\}.
\end{aligned}$$

Second, by compactness of $\mathcal{D}$:

$$\sup_{\delta\in\mathcal{D}} |f(x_t,\beta_0) - f(x_t,[\delta,\mathbf{0}_{k_\theta}])| = \sup_{\delta\in\mathcal{D}} |\beta_0' x_t - \delta' x_{\delta,t}| \leq |\beta_0' x_t| + K|x_{\delta,t}|.$$

Sub-additivity, $|\beta_0| < \infty$ and sub-exponentiality therefore yield:

$$\max_{t\in\mathbb{N}} P\left(\sup_{\delta\in\mathcal{D}} |f(x_t,\beta_0) - f(x_t,[\delta,\mathbf{0}_{k_\theta}])| > c\right)$$

$$\leq \max_{t\in\mathbb{N}} P\left(|\beta_0' x_t| > c/2\right) + \max_{t\in\mathbb{N}} P\left(K|x_{\delta,t}| > c/2\right)$$

$$\leq \max_{i,t\in\mathbb{N}} P\left(|x_{i,t}| > \frac{c}{2|\beta_0|}\right) + \max_{1\leq j\leq k_\delta, t\in\mathbb{N}} P\left(|x_{j,\delta,t}| > \frac{c}{K\delta 2}\right)$$

$$\leq C\exp\{-\mathcal{K}c\} + C\exp\{-\mathcal{K}c\}$$

$$\leq C\exp\{-\mathcal{K}c\}.$$

Third, by the argument above:

$$\max_{t\in\mathbb{N}} P\left(\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\{f(x_t,\beta_0) - f(x_t,[\delta,\mathbf{0}_{k_\theta}])\} g_{(i)}(x_t,\beta_{(i)})\right| > c\right)$$

$$\leq \max_{i,t\in\mathbb{N}} P\left(|\beta_0' x_t| \times \left|x_{(i),t}\right| > c/2\right) + \max_{i,t\in\mathbb{N}} P\left(K|x_{\delta,t}| \times \left|x_{(i),t}\right| > c/2\right)$$

$$\leq \max_{i,t\in\mathbb{N}} P\left(|x_{i,t}| \times \left|x_{(i),t}\right| > \frac{c}{2|\beta_0|}\right) + \max_{1\leq j\leq k_\delta, i,t\in\mathbb{N}} P\left(|x_{j,\delta,t}| \times \left|x_{(i),t}\right| > \frac{c}{K\delta 2}\right)$$

$$\leq K \max_{i,j,t\in\mathbb{N}} P\left(|x_{i,t}| \times |x_{j,t}| > Kc\right).$$

It remains to prove $|x_{i,t}| \times |x_{j,t}|$ is uniformly sub-exponential. Let $i = j$. By uniform sub-exponentiality $x_{i,t}$ has a moment generating function in a neighborhood of zero. Hence $x_{i,t}$ is uniformly locally sub-Gaussian (Chareka, Chareka, and Kennedy, 2006, Theorem 1). It follows $x_{i,t}^2$ is uniformly locally sub-exponential by arguments analogous to Vershynin (2018, proof of Proposition 2.5.2(iii)) . Therefore, for $\zeta \in (0,b]$ and some finite $b > 0$:

$$\max_{i,t\in\mathbb{N}} E\left[\exp\left\{\zeta x_{i,t}^2\right\}\right] \leq \mathcal{C}\exp\{\mathcal{K}\zeta\} \leq \mathcal{C}\exp\{\mathcal{K}b\} < \infty.$$

Hence $x_{i,t}^2$ is uniformly sub-exponential by Chernoff's bound:

$$\max_{i,t\in\mathbb{N}} P\left(x_{i,t}^2 > c\right) \leq C\exp\left\{-\zeta c\right\} \text{ for some } \zeta > 0.$$

Now let $i \neq j$. By assumption $x_{i,t}$ is conditionally uniformly sub-exponential

$$\max_{i,j,t\in\mathbb{N}} P\left(|x_{i,t}| > c|x_{j,t}\right) \leq \max_{j,t\in\mathbb{N}} \mathcal{C}(x_{j,t})\exp\left\{-Kc\right\} \;\; a.s.$$

where $\mathcal{C}(x_{j,t})$ is $\sigma(x_{j,t})$-measurable, and $E[\max_{j,t\in\mathbb{N}} \mathcal{C}(x_{j,t})^2] < \infty$. Thus, by iterated expectations, the Cauchy-Schwartz inequality, and finite $K > 0$ that changes from line to line:

$$\begin{aligned}\max_{i,j,t\in\mathbb{N}} P\left(|x_{i,t}x_{j,t}| > c\right) &\leq \max_{i,t\in\mathbb{N}} E\left[\mathcal{C}(x_{i,t})\exp\left\{-Kc\frac{1}{|x_{i,t}|}\right\}\right] \\ &\leq K\left(\max_{i,t\in\mathbb{N}} E\left[\exp\left\{-Kc\frac{1}{|x_{i,t}|}\right\}\right]\right)^{1/2}.\end{aligned}$$

By Jensen's inequality, the global concavity of $e^{-1/x}$ on $[0,\infty)$, and $\max_{i,t\in\mathbb{N}} E|x_{i,t}| < \infty$:

$$\begin{aligned}\max_{i,t\in\mathbb{N}} E\left[\exp\left\{-2\frac{1}{|x_{i,t}|}Kc\right\}\right] &\leq \exp\left\{-Kc\min_{i,t\in\mathbb{N}} E\left[\frac{1}{|x_{i,t}|}\right]\right\} \\ &\leq \exp\left\{-Kc\frac{1}{\max_{i,t\in\mathbb{N}} E\left|x_{i,t}\right|}\right\} = \exp\left\{-Kc\right\}.\end{aligned}$$

Hence again we have sub-exponentiality:

$$\max_{i,j,t\in\mathbb{N}} P\left(|x_{i,t}x_{j,t}| > c\right) \leq K\left(\max_{i,t\in\mathbb{N}} E\left[\exp\left\{-Kc\frac{1}{|x_{i,t}|}\right\}\right]\right)^{1/2} \leq K\exp\left\{-Kc\right\},$$

d. The required moment bounds hold by linearity, $\min_{t\in\mathbb{N}} E[\epsilon_t^2] > 0$, $\min_{i,t\in\mathbb{N}} E[x_{i,t}^2] > 0$, $\sup_{\beta\in\mathcal{B}} |\beta| < \infty$, the stated moment lower bounds, limit bounds (D.1), and the fact that sub-exponentiality implies the existence of a moment generating function (in a neighborhood of zero) and therefore the existence of all moments. $\mathcal{QED}$.

D.2 Logistic Regression

We verify the assumptions for a logistic-like regression model:

$$y_t = \frac{1}{1+\exp\left\{\delta_0' x_{\delta,t} + \theta_0' x_{\theta,t}\right\}} + \epsilon_t = \frac{1}{1+\exp\left\{\beta_0' x_t\right\}} + \epsilon_t,$$

where $x_{\delta,t} \in \mathbb{R}^{k_\delta}$, $x_{\theta,t} \in \mathbb{R}^{k_\theta}$, and $E[\epsilon_t] = 0$. Typically $y_t \in \{0,1\}$ in applications, but here we do not impose a support restriction on y_t. A similar model that we do not treat here is the additively nonlinear $y_t = \delta_{0,1}' x_{\delta,t} + \delta_{0,2} F(\theta_0' x_{\theta,t}) + u_t$, with logistic $F(u) = 1/(1 + \exp\{u\})$, or any real analytic *squash* function (cf. White, 1989, Hornik, Stinchcombe, and White, 1989).

Assume $\{x_t, \epsilon_t\}$ are iid across t and mutually independent for a unique point $\beta_0 = [\delta_0', \theta_0']'$, where δ_0 and $\theta_{0,i}$ are interior points of compact $\mathcal{D} \subset \mathbb{R}^{k_\delta}$, $k_\delta \in \mathbb{N}$, and $\Theta_i \subset \mathbb{R}$. Assume $\Theta \subset \{\times_{i=1}^{k_\theta} \Theta_i : |\theta| < \infty\}$. Assume $\min_{t \in \mathbb{N}} E[\epsilon_t^2] > 0$ and $\min_{i,t \in \mathbb{N}} E[x_{i,t}^2] > 0$. Let ϵ_t and $x_{i,t}^2$ be uniformly sub-exponential $\max_{t \in \mathbb{N}} P(|\epsilon_t| > c) \leq \mathcal{C} \exp\{-\gamma c\}$ and $\max_{i,t \in \mathbb{N}} P(x_{i,t}^2 > c) \leq \mathcal{C} \exp\{-\gamma c\}$ for some finite $\gamma, \mathcal{C} > 0$ that may be different in different places. Identical distributedness is imposed merely to ease notation in a nonlinear setting.

The parsimonious models are for $i = 1, ..., k_{\theta,n}$:

$$\begin{aligned} y_t &= \frac{1}{1 + \exp\left\{\delta_{(i)}^{*\prime} x_{\delta,t} + \theta_i^* x_{\theta,i,t}\right\}} + v_{(i),t} \\ &= \frac{1}{1 + \exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}} + v_{(i),t} = f(\beta_{(i)}^{*\prime} x_{(i),t}) + v_{(i),t} \end{aligned}$$

where for unique $\beta_{(i)}^*$ in the interior of $\mathcal{B}_{(i)} \equiv \mathcal{D} \times \Theta_i$:

$$E\left[v_{(i),t} \frac{\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}{\left(1 + \exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}\right)^2} x_{(i),t}\right] = 0.$$

Squared error loss is used:

$$\mathcal{L}(\beta) = \frac{1}{2} E\left[\left(y_t - \frac{1}{1 + \exp\left\{\beta_0' x_t\right\}}\right)^2\right] \text{ and } \mathcal{L}_{(i)}(\beta) = \frac{1}{2} E\left[\left(y_t - \frac{1}{1 + \exp\left\{\beta_{(i)}' x_{(i),t}\right\}}\right)^2\right]$$

Define

$$\tilde{x}_{i,t} = \frac{\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}{\left[1 + \exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}\right]^2} x_{i,t} \text{ and } \tilde{x}_{(i),t} = \frac{\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}{\left[1 + \exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}\right]^2} x_{(i),t}.$$

The estimator is computed by least squares with criterion $\hat{\mathcal{L}}_{(i)}(\beta) \equiv .5 \sum_{t=1}^n \{y_t - f(\beta_{(i)}' x_{(i),t})\}^2$, hence

$$\widehat{\mathcal{G}}_{(i)} = -\sum_{t=1}^{n} v_{(i),t} \frac{\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}{\left[1+\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}\right]^2} x_{(i),t} = -\sum_{t=1}^{n} v_{(i),t} \tilde{x}_{(i),t}$$

$$G_{(i),t}^{(0)} = \epsilon_t \tilde{x}_{(i),t} + \left(\frac{1}{1+\exp\{\beta_0' x_t\}} - \frac{1}{1+\exp\left\{\delta_{(i)}^{*\prime} x_{\delta,t}\right\}}\right) \tilde{x}_{(i),t}$$

and

$$\widehat{\mathcal{H}}_{(i)} = \sum_{t=1}^{n} \tilde{x}_{(i),t} \tilde{x}_{(i),t}' - \sum_{t=1}^{n} v_{(i),t} \left(\frac{1-\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}{1+\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}}\right) \tilde{x}_{(i),t} x_{(i),t}'$$

with $\widehat{\mathfrak{H}}_{(i)} = \bar{\mathfrak{H}}_{n,(i)} = \widehat{\mathcal{H}}_{(i)}/n$.

Let $\{\varpi_n\}_{n\in\mathbb{N}}$ be a sequence of positive real numbers, $\varpi_n \to 0$, that may be different in different places; and $\mathcal{B}_{n,(i)} \equiv \{\mathcal{B}_{(i)} : ||\beta_{(i)} - \beta_{(i)}^*|| \leq \varpi_n\}$. Write

$$\tilde{x}_{(i),t}(\beta_{(i)}') \equiv \frac{\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}{\left[1+\exp\left\{\beta_{(i)}' x_{(i),t}\right\}\right]^2} x_{(i),t},$$

and assume

$$\inf_{\lambda'\lambda=1} \min_{1\leq i\leq k_{\theta,n}} \left\{ \inf_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \left\{ \frac{1}{n}\sum_{t=1}^{n} \left(\lambda' \tilde{x}_{(i),t}(\beta_{(i)})\right)^2 \right\} \right\} > 0 \ a.s.$$

$$\max_{1\leq i\leq k_{\theta,n}} \sup_{\beta_{(i)}\in\mathcal{B}_{n,(i)}} \left\| \left(\frac{1}{n}\sum_{t=1}^{n} \tilde{x}_{(i),t}(\beta_{(i)}) \tilde{x}_{(i),t}'(\beta_{(i)}) \right)^{-1} \right\| = O_p(k_{\theta,n})$$

$$\lim_{n\to\infty} \inf_{i\in\mathbb{N}} \min_{\tilde{\lambda}'\tilde{\lambda}=1} \inf E\left[\left(\tilde{\lambda}' \tilde{x}_{(i),t}\right)^2\right] \geq \underline{c} \text{ for some } \underline{c} \in (0,\infty).$$

Finally, define

$$\phi_t(\beta_{(i)}) \equiv \left(\frac{1-\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}{1+\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}\right) \frac{\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}{\left[1+\exp\left\{\beta_{(i)}' x_{(i),t}\right\}\right]^2}$$

$$\psi_t(\beta_{(i)}) \equiv \left(\frac{1}{1+\exp\left\{\beta_{(i)}^{*\prime} x_{(i),t}\right\}} - \frac{1}{1+\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}\right) \times \phi_t(\beta_{(i)}),$$

and let

$$\mathfrak{H}_{(i)}(\beta_{(i)}) \equiv E\left[\tilde{x}_{(i),t}(\beta_{(i)})\tilde{x}_{(i),t}(\beta_{(i)})'\right] - E\left[\psi_t(\beta_{(i)})x_{i,t}x_{i,t}'\right]$$

be uniformly positive definite on $\{\beta_{(i)} : ||\beta_{(i)} - \beta^*_{(i)}|| \leq \varepsilon\}$ and some $\varepsilon > 0$, in particular $\max_{i \in \mathbb{N}}\{\sup_{\beta_{(i)}:||\beta_{(i)}-\beta^*_{(i)}||\leq\varepsilon} ||\mathfrak{H}^{-1}_{(i)}(\beta_{(i)})||\} < \infty$. Notice for tiny ε we are supposing an infinitesimal shift in $E\left[\tilde{x}_{(i),t}(\beta_{(i)})\tilde{x}_{(i),t}(\beta_{(i)})'\right]$ is uniformly positive definite. This is not atypical in the literature when working with nonlinear response functions.

Lemma 5.2. *In the setting above, Assumptions 1-5 hold.*

Proof.

Assumption 1. Define

$$f(u) \equiv \frac{1}{1+\exp\{u\}}. \tag{D.1}$$

Then

$$y_t = f(\beta_0' x_t) + \epsilon_t \text{ and } y_t = f(\beta^{*\prime}_{(i)} x_{(i),t}) + v_{(i),t}.$$

The identification assumption therefore follows from the argument used in the proof of Lemma 5.1, and Lemma 2.3 in the main paper.

Assumption 2.

a. $\hat{\mathcal{L}}_{(i)}(\hat{\beta}_{(i)}) = \min_{\beta_{(i)} \in \mathcal{B}_{(i)}} \hat{\mathcal{L}}_{(i)}(\beta_{(i)})$ follows from the real analytic response and least squares criterion (Jennrich, 1969, cf. Lemma 2).

b.Use (D.1). By mutual independence, independence, $E\left[\epsilon_t^2\right] < \infty$ and $f(\beta'_{(i)}x_{(i),t}) \leq 1$ $a.s.$ Kolmogorov's strong law yields:

$$\begin{aligned}
\left|\frac{1}{n}\hat{\mathcal{L}}_{(i)}(\beta_{(i)}) - \mathcal{L}_{(i)}(\beta_{(i)})\right| \leq\; & \frac{1}{2}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\epsilon_t^2 - E\left[\epsilon_t^2\right]\right)\right| + \frac{1}{n}\sum_{t=1}^{n}\epsilon_t f(\beta^{*\prime}_{(i)}x_{(i),t}) \\
& + \frac{1}{2}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{f(\beta^{*\prime}_{(i)}x_{(i),t}) - E\left[f(\beta^{*\prime}_{(i)}x_{(i),t})\right]\right\}\right| \\
& + \left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t f(\beta'_{(i)}x_{(i),t})\right| + \frac{1}{2}\left|\frac{1}{n}\sum_{t=1}^{n} f(\beta'_{(i)}x_{(i),t}) - E\left[f(\beta'_{(i)}x_{(i),t})\right]\right| \\
=\; & \left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t f(\beta'_{(i)}x_{(i),t})\right| + \frac{1}{2}\left|\frac{1}{n}\sum_{t=1}^{n} f(\beta'_{(i)}x_{(i),t}) - E\left[f(\beta'_{(i)}x_{(i),t})\right]\right| + o_p(1)
\end{aligned}$$

where $o_p(1)$ is not a function of $\beta_{(i)}$. In view of boundedness of $f(\beta'_{(i)}x_{(i),t})$ and $(\partial/\partial\beta_{(i)})f(\beta'_{(i)}x_{(i),t})$, $E[\epsilon_t f(\beta'_{(i)}x_{(i),t})] = 0$, $E[\epsilon_t^2] < \infty$, independence and homogeneity, Lemma A.5 yields (e.g.

Newey, 1991, Corollaries 2.2 and 31):

$$\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t f(\beta_{(i)}'x_{(i),t})\right| \xrightarrow{p} 0$$

$$\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}\left|\frac{1}{n}\sum_{t=1}^{n}f(\beta_{(i)}'x_{(i),t}) - E\left[f(\beta_{(i)}'x_{(i),t})\right]\right| \xrightarrow{p} 0$$

which suffices to validate $\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}}|\hat{\mathcal{L}}_{(i)}(\beta_{(i)})/n - \mathcal{L}_{(i)}(\beta_{(i)})| \xrightarrow{p} 0$ for each i.

Assumption 3. Let H_0 hold.

a. We have

$$\begin{aligned}
\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) &= \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_{(i),t}(\beta_{(i)})\tilde{x}_{(i),t}'(\beta_{(i)}) \\
&\quad -\frac{1}{n}\sum_{t=1}^{n}\epsilon_t\left(\frac{1-\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}{1+\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}\right)\tilde{x}_{(i),t}(\beta_{(i)})x_{(i),t}' \\
&\quad -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{1+\exp\left\{\beta_{(i)}^{*\prime}x_{(i),t}\right\}}-\frac{1}{1+\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}\right) \\
&\qquad\qquad \times\left(\frac{1-\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}{1+\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}\right)\tilde{x}_{(i),t}(\beta_{(i)})x_{(i),t}' \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathfrak{a}_i(\beta_{(i)})+\frac{1}{n}\sum_{t=1}^{n}\mathfrak{b}_i(\beta_{(i)})+\frac{1}{n}\sum_{t=1}^{n}\mathfrak{c}_i(\beta_{(i)}),
\end{aligned} \tag{D.2}$$

say (notice $\tilde{x}_{(i),t}(\beta_{(i)})x_{(i),t}'$ has $x_{(i),t}$), and

$$\begin{aligned}
\mathfrak{H}_{(i)}(\beta_{(i)}) &= E\left[\tilde{x}_{(i),t}(\beta_{(i)})\tilde{x}_{(i),t}'(\beta_{(i)})\right] \\
&\quad -E\Bigg[\left(\frac{1}{1+\exp\left\{\beta_{(i)}^{*\prime}x_{(i),t}\right\}}-\frac{1}{1+\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}\right) \\
&\qquad\qquad \times\left(\frac{1-\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}{1+\exp\left\{\beta_{(i)}'x_{(i),t}\right\}}\right)\tilde{x}_{(i),t}(\beta_{(i)})x_{(i),t}'\Bigg] \\
&= E\left[\mathfrak{a}_i(\beta_{(i)})\right]+E\left[\mathfrak{c}_i(\beta_{(i)})\right].
\end{aligned}$$

Clearly:

$$\left\|\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \widehat{\mathfrak{H}}_{(i)}(\beta^*_{(i)})\right\| \leq \frac{1}{n}\sum_{t=1}^{n}\left\|\mathfrak{a}_i(\beta_{(i)}) - \mathfrak{a}_i(\beta^*_{(i)})\right\| + \frac{1}{n}\sum_{t=1}^{n}\left\|\mathfrak{b}_i(\beta_{(i)}) - \mathfrak{b}_i(\beta^*_{(i)})\right\| + \frac{1}{n}\sum_{t=1}^{n}\left\|\mathfrak{c}_i(\beta_{(i)}) - \mathfrak{c}_i(\beta^*_{(i)})\right\|.$$

Response boundedness and the mean-value theorem yield:

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left\|\mathfrak{a}_i(\beta_{(i)}) - \mathfrak{a}_i(\beta^*_{(i)})\right\| \\
&\quad = \frac{1}{n}\sum_{t=1}^{n}\left\|\tilde{x}_{(i),t}(\beta_{(i)})\tilde{x}'_{(i),t}(\beta_{(i)}) - \tilde{x}_{(i),t}(\beta^*_{(i)})\tilde{x}'_{(i),t}(\beta^*_{(i)})\right\| \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\exp\left\{2\beta'_{(i)}x_{(i),t}\right\}}{\left[1+\exp\left\{\beta'_{(i)}x_{(i),t}\right\}\right]^4} - \frac{\exp\left\{2\beta^{*\prime}_{(i)}x_{(i),t}\right\}}{\left[1+\exp\left\{\beta^{*\prime}_{(i)}x_{(i),t}\right\}\right]^4}\right\| \times \left\|x_{(i),t}x'_{(i),t}\right\| \\
&\quad \leq 2\frac{1}{n}\sum_{t=1}^{n}\left\|x_{(i),t}\right\|^3 \times \left\|\beta_{(i)} - \beta^*_{(i)}\right\|
\end{aligned}$$

where independence, homogeneity and sub-exponential tails imply $1/n\sum_{t=1}^{n}||x_{(i),t}||^3 = O_p(1)$. Nearly identical derivations compiled together yield $||\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \widehat{\mathfrak{H}}_{(i)}(\beta^*_{(i)})|| \leq \widehat{\mathcal{C}}_{(i)}||\beta_{(i)} - \beta^*_{(i)}||$ $\forall \beta_{(i)}$ and some positive stochastic $\{\widehat{\mathcal{C}}_{(i)}\}_{n\geq 1}$ with $\max_{1\leq i\leq k_{\theta,n}}\widehat{\mathcal{C}}_{(i)} = O_p(1)$. Similarly, boundedness and the mean-value theorem yield:

$$\left\|E\left[\mathfrak{a}_i(\beta_{(i)})\right] - E\left[\mathfrak{a}_i(\beta^*_{(i)})\right]\right\| \leq 2E\left[\left\|x_{(i),t}\right\|^3\right] \times \left\|\beta_{(i)} - \beta^*_{(i)}\right\|,$$

where $\max_{i\in\mathbb{N}} E[||x_{(i),t}||^3] < \infty$ by the uniform exponential tail property. Similar arguments extend to $||E\left[\mathfrak{c}_i(\beta_{(i)})\right] - E[\mathfrak{c}_i(\beta^*_{(i)})]||$, yielding $||\mathfrak{H}_{(i)}(\beta_{(i)}) - \mathfrak{H}_{(i)}(\beta^*_{(i)})|| \leq \mathcal{C}_{(i)}||\beta_{(i)} - \beta^*_{(i)}||$ $\forall \beta_{(i)}$ and nonstochastic $\mathcal{C}_{(i)} > 0$ with $\max_{i\in\mathbb{N}}\mathcal{C}_{(i)} < \infty$.

b(i). First, $\widehat{\mathfrak{H}}_{(i)}(\beta_{(i)})$ is symmetric by construction. Second, by assumption

$$\inf_{\lambda'\lambda=1}\min_{1\leq i\leq k_{\theta,n}}\left\{\inf_{\beta_{(i)}\in\mathcal{B}_{n,(i)}}\left\{\frac{1}{n}\sum_{t=1}^{n}\left(\lambda'\tilde{x}_{(i),t}(\beta_{(i)})\right)^2\right\}\right\} > 0 \ a.s.$$

Third, under the stated assumptions, and by arguments above (see, e.g., Newey, 1991):

$$\sup_{\beta_{(i)} \in \mathcal{B}_{(i)}} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t \left(\frac{1 - \exp\left\{\beta'_{(i)} x_{(i),t}\right\}}{1 + \exp\left\{\beta'_{(i)} x_{(i),t}\right\}} \right) \tilde{x}_{(i),t}(\beta_{(i)}) x'_{(i),t} \right\| \xrightarrow{p} 0.$$

Furthermore, sub-exponential tails, boundedness, the mean-value theorem and $||\beta_{(i)} - \beta^*_{(i)}|| \leq \varpi_n \rightarrow 0$ in $\mathcal{B}_{n,(i)}$ yield:

$$\begin{aligned}
&\sup_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{1 + \exp\left\{\beta^{*\prime}_{(i)} x_{(i),t}\right\}} - \frac{1}{1 + \exp\left\{\beta'_{(i)} x_{(i),t}\right\}} \right) \right. \\
&\qquad\qquad \left. \times \left(\frac{1 - \exp\left\{\beta'_{(i)} x_{(i),t}\right\}}{1 + \exp\left\{\beta'_{(i)} x_{(i),t}\right\}} \right) \tilde{x}_{(i),t}(\beta_{(i)}) x'_{(i),t} \right\| \\
&\leq \sup_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \frac{1}{n} \sum_{t=1}^{n} \left| \frac{1}{1 + \exp\left\{\beta^{*\prime}_{(i)} x_{(i),t}\right\}} - \frac{1}{1 + \exp\left\{\beta'_{(i)} x_{(i),t}\right\}} \right| \left\| \tilde{x}_{(i),t}(\beta_{(i)}) x'_{(i),t} \right\| \\
&\leq \frac{1}{n} \sum_{t=1}^{n} \left| x_{(i),t} \right| \times \left\| \tilde{x}_{(i),t}(\beta_{(i)}) x'_{(i),t} \right\| \times \varpi_n \\
&= O_p(1) \times o(1) = o_p(1)
\end{aligned}$$

In view of (D.2) we have shown

$$\sup_{\beta_{(i)} \in \mathcal{B}_{n,(i)}} \left\| \widehat{\mathfrak{H}}_{(i)}(\beta_{(i)}) - \frac{1}{n} \sum_{t=1}^{n} \tilde{x}_{(i),t}(\beta_{(i)}) \tilde{x}'_{(i),t}(\beta_{(i)}) \right\| \xrightarrow{p} 0. \tag{D.3}$$

This, coupled with the imposed suppositions, yield (i).

b(ii). The required properties for $\mathfrak{H}_{(i)}(\beta_{(i)})$ hold by supposition. Now let $k_\delta = 0$ to ease notation, hence $\beta_{(i)} = \theta_i$ and $\mathcal{B}_{n,(i)} = \Theta_{n,i} \equiv \{\Theta_i : ||\theta_i - \theta^*_i|| \leq \varpi_n\}$. Define

$$\begin{aligned}
\phi_t(\theta_i) &\equiv \left(\frac{1 - \exp\{\theta_i x_{i,t}\}}{1 + \exp\{\theta_i x_{i,t}\}} \right) \frac{\exp\{\theta_i x_{i,t}\}}{[1 + \exp\{\theta_i x_{i,t}\}]^2} \\
\psi_t(\theta_i) &\equiv \left(\frac{1}{1 + \exp\{\theta^*_i x_{i,t}\}} - \frac{1}{1 + \exp\{\theta_i x_{i,t}\}} \right) \times \phi_t(\theta_i)
\end{aligned}$$

Then

$$\widehat{\mathfrak{H}}_{(i)}(\theta_i) = \frac{1}{n}\sum_{t=1}^{n} \tilde{x}_{i,t}^2(\theta_i) - \frac{1}{n}\sum_{t=1}^{n} \epsilon_t \phi_t(\theta_i) x_{i,t}^2 - \frac{1}{n}\sum_{t=1}^{n} \psi_t(\theta_i) x_{i,t}^2$$
$$\mathfrak{H}_{(i)}(\theta_i) = E\left[\tilde{x}_{i,t}^2(\theta_i)\right] - E\left[\psi_t(\theta_i) x_{i,t}^2\right]$$
$$\mathfrak{H}_{(i)}(\theta_{0,i}) = E\left[\tilde{x}_{i,t}^2(\theta_{0,i})\right] = E\left[\tilde{x}_{i,t}^2\right].$$

Multiple uses of the Cauchy-Schwartz inequality therefore yields:

$$\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{\zeta \sup_{\theta_i \in \Theta_{n,i}} \sqrt{n}\left|\widehat{\mathfrak{H}}_{(i)}(\theta_i) - \mathfrak{H}_{(i)}(\theta_i)\right|\right\}\right] \leq (\mathfrak{A}_{1,n} \times \mathfrak{A}_{2,n} \times \mathfrak{A}_{3,n})^{1/6}$$

where:

$$\mathfrak{A}_{1,n} \equiv \max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{6\zeta \sup_{\theta_i \in \Theta_{n,i}} \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{x}_{i,t}^2(\theta_i) - E\left[\tilde{x}_{i,t}^2(\theta_i)\right]\right)\right|\right\}\right]$$
$$\mathfrak{A}_{2,n} \equiv \max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{6\zeta \sup_{\theta_i \in \Theta_{n,i}} \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \epsilon_t \phi_t(\theta_i) x_{i,t}^2\right|\right\}\right]$$
$$\mathfrak{A}_{3,n} \equiv \max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{6\zeta \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\psi_t(\theta_i) x_{i,t}^2 - E\left[\psi_t(\theta_i) x_{i,t}^2\right]\right)\right\}\right].$$

Since $\tilde{x}_{i,t}(\theta_i)$ and $|\psi_t(\theta_i)|^{1/2} x_{i,t}$ have the same tail properties as $x_{i,t}(\theta_{(i)})$ in view of $\sup_{\theta_i \in \Theta_{n,i}} \max_{1 \leq i \leq k_{\theta,n} 1 \leq t \leq n} |\phi_t(\theta_i)| \leq 1$ *a.s.* and $\sup_{\theta_{(i)} \in \Theta_{(i)}} \max_{1 \leq i \leq k_{\theta,n} 1 \leq t \leq n} |\psi_t(\theta_i)| \leq 1$ *a.s.*, arguments in the proof of Lemma 5.1 carry over verbatim to yield $\mathfrak{A}_{1,n} = O(1)$ and $\mathfrak{A}_{3,n} = O(1)$.

For $\mathfrak{A}_{2,n}$ observe that $\phi_t(\theta_i) x_{i,t}^2 \leq x_{i,t}^2$ *a.s.* uniformly on Θ_i and over i. By assumption ϵ_t and $x_{i,t}^2$ are (uniformly) sub-exponential. Corollary B.2 therefore yields:

$$\max_{1 \leq i \leq k_{\theta,n}} E\left[\exp\left\{6\zeta \sup_{\theta_i \in \Theta_{n,i}} \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \epsilon_t \phi_t(\theta_i) x_{i,t}^2\right|\right\}\right] = O(1).$$

c. The proof that $\max_{1 \leq i \leq k_{\theta,n}} E[\exp\{\zeta|\widehat{\mathcal{G}}_{(i)}/\sqrt{n}|\}] = O(1)$ for some $\zeta > 0$ is essentially identical to the preceding argument, and therefore omitted.

Assumption 4. Let H_0 hold. $\max_{i,t \in \mathbb{N}} ||\mathfrak{H}_{(i)}^{-1}|| < \infty$ holds by supposition.

a. $G_{i,t}$ is independent across t by assumption, and $\max_{i,t \in \mathbb{N}} ||E[G_{i,t} G_{i,t}']|| < \infty$ by mutual independence and sub-exponentiality.

b. By mutual independence:

$$\sigma_{n,i}^2(\lambda) = E\left[\epsilon_t^2\right] \times E\left[\left(\lambda' \mathfrak{H}_{(i)}^{-1} \tilde{x}_{(i),t}\right)^2\right] = E\left[\epsilon_t^2\right] \times E\left[\left(\tilde{\lambda}' \tilde{x}_{(i),t}\right)^2\right],$$

say. Then $\liminf_{n\to\infty} \min_{i\in\mathbb{N}} \inf_{\tilde{\lambda}'\tilde{\lambda}=1} E[(\tilde{\lambda}'\tilde{x}_{(i),t})^2] \geq \underline{c} > 0$ by assumption, and the upper bound holds under sub-exponentiality.

c. Let $k_\delta = 0$ to ease notation. Then we may drop λ and use:

$$z_{(i),t} = -\mathfrak{H}_{(i)}^{-1} G_{i,t} = -\frac{1}{E\left[\tilde{x}_{i,t}^2\right]} \epsilon_t \tilde{x}_{i,t} = -\frac{1}{E\left[\tilde{x}_{i,t}^2\right]} \frac{\exp\left\{\theta_i x_{i,t}\right\}}{\left[1 + \exp\left\{\theta_i x_{i,t}\right\}\right]^2} \epsilon_t x_{i,t}.$$

Observe $|z_{(i),t}| \leq K|\epsilon_t x_{i,t}|$ $a.s.$ The required condition therefore follows from sub-exponentiality, cf. the proof of Lemma 5.1.

Assumption 5.

a. and **b.** The conditions hold by construction of the logistic response (see Jennrich, 1969, Lemma 2).

c. Sub-exponentiality for the error $\max_{t\in\mathbb{N}} P(|\epsilon_t| > c) \leq C\exp\{-\mathcal{K}c\}$ holds by supposition. Now let $w_{i,t}$ denote

$$\begin{aligned}
&\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|g_{(i)}(x_t, \beta_{(i)})\right|, \text{ or} \\
&\sup_{\delta\in\mathcal{D}} \left|f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right|, \text{ or} \\
&\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\left\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right\} g_{(i)}(x_t, \beta_{(i)})\right|, \text{ or} \\
&\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|\left\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right\} h_{(i)}(x_t, \beta_{(i)})\right|.
\end{aligned}$$

Observe *almost surely*:

$$\begin{aligned}
\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left|g_{(i)}(x_t, \beta_{(i)})\right| &= \sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} \left| \frac{\exp\left\{\beta_{(i)}' x_{(i),t}\right\}}{\left[1 + \exp\left\{\beta_{(i)}' x_{(i),t}\right\}\right]^2} x_{(i),t} \right| \leq \left|x_{(i,t}\right| \\
\sup_{\delta\in\mathcal{D}} \left|f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right| &= \sup_{\delta\in\mathcal{D}} \left| \frac{1}{1 + \exp\left\{\beta_0' x_t\right\}} - \frac{1}{1 + \exp\left\{\delta' x_{\delta,t}\right\}} \right| \leq 2 \\
\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} &\left|\left\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right\} g_{(i)}(x_t, \beta_{(i)})\right| \leq 2\left|x_{(i,t}\right| \\
\sup_{\beta_{(i)}\in\mathcal{B}_{(i)}} &\left|\left\{f(x_t, \beta_0) - f(x_t, [\delta, \mathbf{0}_{k_\theta}])\right\} h_{(i)}(x_t, \beta_{(i)}) \leq 2\right| x_{(i),t} x_{(i),t}'.
\end{aligned}$$

The required exponential tail bounds now follow from either boundedness, or arguments in the proof of Lemma 5.1 and sub-additivity.

d. The moment bounds hold under sub-exponentiality, and the Hessian eigenvalue and related bounds hold by supposition. Finally, (B.12) is trivial under homogeneity. $\mathcal{QED}$

E Complete Simulation Results

E.1 Linear Regression Model

E.1.1 Benchmark Results

Table 1: Linear Regression: Rejection Frequencies under $H_0 : \theta = 0$

	$k_\delta = 0$									$k_\delta = 10$								
	$n = 100$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.008	.051	.101	.006	.051	.107	.008	.041	.086	.013	.090	.118	.006	.057	.129	.007	.053	.118
Max-t-Test(b)	.010	.048	.110	.010	.055	.114	.003	.042	.089	.018	.072	.142	.019	.092	.180	.018	.107	.183
Wald(a)	.082	.327	1.00	.588	1.00	1.00	.939	1.00	1.00	.131	.420	1.00	.779	1.00	1.00	.986	1.00	1.00
Wald(b)	.003	.035	.076	.000	.001	.016	.000	.000	.005	.002	.057	.122	.000	.018	.081	.000	.003	.040
SWald(a)	.067	.134	.190	.448	.630	.726	.864	.942	.963	.094	.193	.262	.689	.810	.855	.973	.985	.988
SWald(b)	.003	.035	.076	.000	.001	.016	.000	.000	.005	.002	.057	.122	.000	.018	.081	.000	.003	.040
	$n = 250$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.003	.046	.094	.010	.046	.095	.007	.052	.111	.010	.064	116	.007	.053	.110	.006	.049	.101
Max-t-Test(b)	.005	.043	.090	.008	.052	.094	.006	.040	.082	.014	.064	.136	.009	.057	.101	.008	.069	.120
Wald(a)	.055	.282	1.00	.196	1.00	1.00	.790	1.00	1.00	.057	.334	1.00	.255	1.00	1.00	.870	1.00	1.00
Wald(b)	.005	.041	.091	.001	.019	.059	.001	.003	022	.010	.053	.113	.003	.035	.083	.000	.005	.041
SWald(a)	.032	.069	.114	.116	.237	.317	.611	.760	.824	.052	.109	.174	.145	.277	.368	.737	.860	.900
SWald(b)	.005	.041	.091	.001	.019	.059	.001	.003	022	.010	.053	.113	.003	.035	.083	.000	.005	.041
	$n = 500$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.015	.049	.099	.006	.054	.107	.007	.045	.098	.009	.041	.103	.009	.070	.117	.004	.027	.085
Max-t-Test(b)	.013	.041	.093	.008	.048	.104	.006	.042	.105	.008	.046	.103	.011	.061	.103	.005	.034	.094
Wald(a)	.028	.257	1.00	.120	1.00	1.00	.700	1.00	1.00	.033	.286	1.00	.139	1.00	1.00	.785	1.00	1.00
Wald(b)	.005	.034	.082	.003	.028	.070	.000	.004	.023	.005	.041	.099	.004	.029	.089	.002	.015	.063
SWald(a)	.035	.082	.121	.036	.114	.177	.413	.603	.707	.028	.083	.139	.059	.141	.216	.522	.701	.780
SWald(b)	.005	.034	.082	.003	.028	.070	.000	.004	.023	.005	.041	.099	.004	.029	.089	.002	.015	.063
	$n = 1000$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.006	.064	.115	.005	.,044	.094	.004	.005	.101	.015	.063	.108	.013	045	.096	.010	.051	.089
Max-t-Test(b)	012	.056	.116	.010	.045	.089	.006	.050	.102	.010	.061	.115	.014	.055	.107	.012	.042	.087
Wald(a)	.038	.276	1.00	.087	1.00	1.00	.637	1.00	1.00	.037	.279	1.00	.093	1.00	1.00	.681	1.00	1.00
Wald(b)	.015	.045	.099	.007	034	.080	.000	.017	.056	.010	.052	.116	.007	.034	.076	.002	.028	.052
SWald(a)	.026	.073	.111	.028	.090	.155	.345	.531	.628	.032	.089	.147	.036	.098	.147	.447	.580	.604
SWald(b)	.015	.045	.099	.007	034	.080	.000	.017	.056	.010	.052	.116	.007	.034	.076	.002	.028	.052

”b” = bootstrapped p-value; ”a” =asymptotic test. “SWald” is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples.

Table 2: Linear Regression: Rejection Frequencies under $H_1 : \theta_1 = .001$ and $\theta_i = 0,\ i \geq 2$]

	$k_\delta = 0$									$k_\delta = 10$								
	$n = 100$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.101	.307	.466	.451	.762	.870	.605	.862	.938	.011	.058	.125	.028	.119	.218	.039	.149	.266
Max-t-Test(b)	.142	.388	.536	.531	.795	.890	.673	.876	.937	.019	.083	.190	.053	.164	.275	.099	.267	.427
Wald(a)	.258	.570	1.00	.900	1.00	1.00	.988	1.00	1.00	.172	466	1.00	.832	1.00	1.00	.994	1.00	1.00
Wald(b)	.020	.133	.251	.002	.033	.131	.000	.001	.038	.004	.060	.156	.000	.031	.115	.000	.009	.060
SWald(a)	.245	.358	.440	.764	.861	.903	.980	.980	.993	.132	.250	.328	.756	.855	.895	.989	.992	.993
SWald(b)	.020	.133	.251	.002	.033	.131	.000	.001	.038	.004	.060	.156	.000	.031	.115	.000	.009	.060
	$n = 250$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.248	.460	.600	.924	.978	.992	1.00	1.00	1.00	.027	.127	.217	.148	.347	.485	.514	.742	.838
Max-t-Test(b)	.219	.462	.584	.914	.975	.990	1.00	1.00	1.00	.037	.149	.245	.281	.500	.616	.617	.799	.873
Wald(a)	.243	.589	1.00	.880	1.00	1.00	1.00	1.00	1.00	.149	.450	1.00	.590	1.00	1.00	.985	1.00	1.00
Wald(b)	.068	.224	.338	.212	.522	.685	.309	.850	.963	.031	.127	.217	.034	.190	.341	.008	.096	.260
SWald(a)	.282	.427	.500	.874	.935	.956	.998	.998	1.00	.105	.217	.294	.402	.573	.667	.959	.988	.994
SWald(b)	.068	.224	.338	.212	.522	.685	.309	.850	.963	.031	.127	.217	.034	.190	.341	.008	.096	.260
	$n = 500$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.833	.963	.987	1.00	1.00	1.00	1.00	1.00	1.00	.020	.122	.235	.490	.739	.836	1.00	1.00	1.00
Max-t-Test(b)	.931	.985	.994	1.00	1.00	1.00	1.00	1.00	1.00	.063	.199	.291	.680	.829	.893	1.00	1.00	1.00
Wald(a)	.863	.983	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.130	.448	1.00	.702	1.00	1.00	1.00	1.00	1.00
Wald(b)	.693	.876	.939	.971	.994	.98	1.00	1.00	1.00	.029	.142	.263	.199	.483	.631	.297	.680	.857
SWald(a)	.629	.776	.840	.992	.992	1.00	1.00	1.00	1.00	.290	.456	.560	.527	.691	.782	.992	1.00	1.00
SWald(b)	.693	.876	.939	.971	.994	.98	1.00	1.00	1.00	.029	.142	.263	.199	.483	.631	.297	.680	.857
	$n = 1000$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.463	.682	.768	.979	.997	.998	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.363	.595	.710	.997	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.355	.700	1.00	.965	1.00	1.00	1.00	1.00	1.00
Wald(b)	.987	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.1940	.395	.537	.821	.945	.968	.998	1.00	1.00
SWald(a)	.985	.989	.996	1.00	1.00	1.00	1.00	1.00	1.00	.557	..687	.767	.871	.946	.964	1.00	1.00	1.00
SWald(b)	.987	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.1940	.395	.537	.821	.945	.968	.998	1.00	1.00

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples.

Table 3: Linear Regression: Rejection Frequencies under $H_1 : \theta_i = i/2$ for $i = 1, ..., 10$ and $\theta_i = 0$ $i \geq 11$]

	$k_\delta = 0$									$k_\delta = 10$								
$n = 100$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	.869	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.868	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	.869	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.868	1.00	1.00
$n = 250$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$n = 500$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$n = 1000$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples.

Table 4: Linear Regression: Rejection Frequencies under $H_1 : \theta_i = .001$ for each i

	$k_\delta = 0$									$k_\delta = 10$								
	$n = 100$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.729	.917	.965	.995	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.706	.898	.951	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.920	.985	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	.996	1.00	1.00	.885	1.00	1.00	.499	.812	.909	.991	1.00	1.00	.747	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.907	.952	.971	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	.996	1.00	1.00	.885	1.00	1.00	.499	.812	.909	.991	1.00	1.00	.747	1.00	1.00
	$n = 250$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.583	.879	.947	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.883	.964	.991	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.932	.989	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.756	.926	.965	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.926	.963	.978	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.756	.926	.965	1.00	1.00	1.00	1.00	1.00	1.00
	$n = 500$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$n = 1000$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max-t-Test(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(a)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SWald(b)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples.

E.1.2 Robustness Checks

Table 5: Linear Regression: Robustness Check a.i: $V[x_{\theta,i,t}] \in [1, 100]$ increasing over i

	$k_\delta = 0$									$k_\delta = 10$								
$n = 100$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.010	.051	.098	.007	.046	.089	.006	.039	.087	.012	.059	.116	.011	.057	.112	.010	.059	.115
Max-t-Test(b)	.006	053	.106	.011	.055	.107	.014	.051	.116	.012	.062	.121	.016	.063	.122	.016	.064	.124
Wald(a)	.071	.324	1.00	.609	1.00	1.00	.930	1.00	1.00	.117	.422	1.00	.756	1.00	1.00	.979	1.00	1.0
Wald(b)	.001	.032	.070	.000	.002	.017	.000	.001	.005	.006	.050	.117	.01	.014	.077	.000	.001	.030
SWald(a)	.070	.130	.194	.459	.631	.694	.855	.927	.949	.106	.182	.253	.660	.777	.839	.965	.983	.993
SWald(b)	.001	.032	.070	.000	.002	.017	.000	.001	.005	.006	.050	.117	.01	.014	.077	.000	.001	.030
$n = 250$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.009	.054	.114	.010	.053	.104	.008	.047	.104	.014	.062	.112	.010	.053	.104	.015	.063	.113
Max-t-Test(b)	.007	.041	.091	.009	.047	.097	.010	.043	.098	.016	.066	.122	.007	.059	.116	.011	.060	.120
Wald(a)	.045	.273	1.00	.187	1.00	1.00	.802	1.00	1.00	.081	.308	1.00	.280	1.00	1.00	.876	1.00	1.00
Wald(b)	.004	.033	.092	.002	.022	.053	.000	.004	.021	.015	.064	.127	.001	.027	.071	.000	.011	.052
SWald(a)	.035	.088	.121	.103	.231	.310	.615	.771	.837	.047	.115	.176	.143	.287	.369	.735	.859	.911
SWald(b)	.004	.033	.092	.002	.022	.053	.000	.004	.021	.015	.064	.127	.001	.027	.071	.000	.011	.052
$n = 500$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.009	.057	.105	.011	.042	.095	.010	.049	.099	.006	.052	.111	.009	.049	.092	.011	.055	.099
Max-t-Test(b)	.014	.052	093	.009	.043	.095	.012	.045	..092	.009	.050	.102	.013	.047	.103	.007	.046	.110
Wald(a)	.041	.280	1.00	.127	1.00	1.00	.733	1.00	1.00	.040	.281	1.00	.119	1.00	1.00	.785	1.00	1.00
Wald(b)	.006	.045	.090	.007	.036	.079	.000	.012	.041	.008	.042	.097	.05	.023	.078	.003	.026	.062
SWald(a)	.030	.065	.110	.057	.126	.188	.420	.611	.707	.037	.075	.116	.070	.157	.218	.507	.688	.774
SWald(b)	.006	.045	.090	.007	.036	.079	.000	.012	.041	.008	.042	.097	.05	.023	.078	.003	.026	.062
$n = 1000$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.007	.055	.107	.011	.058	.114	.008	.042	.090	.005	.045	.100	.012	.049	.093	.011	.050	.100
Max-t-Test(b)	.008	.052	.099	.013	.052	.091	.012	.055	.102	.011	.052	.111	.009	.046	.095	.007	.060	.115
Wald(a)	.033	.253	1.00	.104	1.00	1.00	.624	1.00	1.00	.052	.274	1.00	.105	1.00	1.00	.658	1.00	1.00
Wald(b)	.009	.043	090	.008	.047	.100	.000	.014	.590	.007	.063	.116	.003	.050	.100	.001	.024	.062
SWald(a)	.027	.072	.119	.027	.091	.159	.302	.487	.590	.021	.070	.115	.042	.098	.149	.424	.502	.669
SWald(b)	.009	.043	090	.008	.047	.100	.000	.014	.590	.007	.063	.116	.003	.050	.100	.001	.024	.062

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples. $H_0 : \theta_0 = 0$ is true. $x_{\theta,i,t}$ and $x_{\delta,i,t}$ are within and across block iid.

Table 6: Linear Regression: Robustness Check a.ii: $V[x_{\theta,1,t}] = 10$, and $\{\{V[x_{\theta,i,t}]\}_{i=2}^{\mathring{k}_{n,\theta}}, V[x_{\delta,i,t}]\} = 1$

	$k_\delta = 0$									$k_\delta = 10$								
$n = 100$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.005	.041	.095	.006	.044	.089	.004	.041	.088	.011	.058	.112	.009	.056	.116	.010	.049	.115
Max-t-Test(b)	.07	.050	.106	.007	.046	.095	.009	.057	.111	.015	.064	.113	.014	.067	.118	.013	.064	.120
Wald(a)	.085	.326	1.00	.618	1.00	1.00	.921	1.00	1.00	.117	.417	1.00	.758	1.00	1.00	.983	1.00	1.00
Wald(b)	.003	.037	.081	.000	.000	.015	.000	.001	.003	.006	.052	.111	.000	.017	.081	.000	.003	.044
SWald(a)	.058	.114	.187	.481	.627	.704	.865	928	.948	.119	.205	.271	.684	.793	.844	.963	.987	.990
SWald(b)	.003	.037	.081	.000	.000	.015	.000	.001	.003	.006	.052	.111	.000	.017	.081	.000	.003	.044
$n = 250$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.007	.049	.105	.008	.053	.106	.002	.045	.089	.008	.056	.112	.010	.044	.098	.011	.053	.112
Max-t-Test(b)	.008	.052	.115	.009	.060	.111	.007	.051	.100	.013	.058	.115	.013	.057	.115	.015	.064	.115
Wald(a)	.062	.324	1.00	.205	1.00	1.00	.792	1.00	1.00	.060	.299	1.00	.242	1.00	1.00	.869	1.00	1.00
Wald(b)	.005	.050	.104	.003	.026	.071	.000	.003	.108	,011	.050	.104	.003	.034	.088	.001	.014	.052
SWald(a)	.033	.085	.126	.118	.221	.316	.587	.743	.808	.040	.103	.152	.144	.252	.356	.732	.849	.893
SWald(b)	.005	.050	.104	.003	.026	.071	.000	.003	.108	,011	.050	.104	.003	.034	.088	.001	.014	.052
$n = 500$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.005	.044	.087	.007	.042	.088	.010	.055	.103	.014	.046	.090	.011	.049	.101	.008	.049	.101
Max-t-Test(b)	.005	.045	.093	.006	.042	.097	.016	.053	.113	.012	.046	.097	.012	.052	.102	.010	.053	.110
Wald(a)	.033	.200	1.00	.106	1.00	1.00	.694	1.00	1.00	.049	.274	1.00	.130	1.00	1.00	.747	1.00	1.00
Wald(b)	.004	.044	.100	.002	.021	.065	.000	.007	.030	.010	.048	.091	.007	.037	.086	.001	.018	.064
SWald(a)	.024	.060	.104	.042	.105	.172	.458	.633	.732	.031	.089	.128	.058	.148	.210	.496	.688	.772
SWald(b)	.004	.044	.100	.002	.021	.065	.000	.007	.030	.010	.048	.091	.007	.037	.086	.001	.018	.064
$n = 1000$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.004	.042	.091	.008	.047	.091	.006	.045	.092	.013	.050	.108	.006	.045	.090	.008	.043	.090
Max-t-Test(b)	.006	.045	.091	.007	.048	.095	.006	.046	.087	.014	.056	.111	.008	.046	.095	.010	.042	.092
Wald(a)	.038	.271	1.00	.084	1.00	1.00	.643	1.00	1.00	.043	.268	1.00	.085	1.00	1.00	.683	1.00	1.00
Wald(b)	.008	.041	.089	.009	.037	.080	.003	.024	.056	.009	.052	.105	.007	.031	.075	.001	.022	.064
SWald(a)	.032	.076	.117	.027	.085	.150	.308	.512	.622	.038	.084	.128	.035	.096	.163	.356	.560	.662
SWald(b)	.008	.041	.089	.009	.037	.080	.003	.024	.056	.009	.052	.105	.007	.031	.075	.001	.022	.064

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples. $H_0 : \theta_0 = 0$ is true. $x_{\theta,i,t}$ and $x_{\delta,i,t}$ are within and across block iid.

Table 7: Linear Regression: Robustness Check a.iii: $V[x_{\theta,1,t}] = 100$, and $\{\{V[x_{\theta,i,t}]\}_{i=2}^{\mathring{k}_{n,\theta}}, V[x_{\delta,i,t}]\} = 1$

	$k_\delta = 0$									$k_\delta = 10$								
$n = 100$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.006	.043	.093	.005	.044	.090	.004	.041	.085	.014	.061	.121	.012	.058	.115	.011	.064	.122
Max-t-Test(b)	.007	.052	.116	.010	.050	.101	.008	.052	.106	.015	.064	.125	.016	.063	.125	.016	.066	.126
Wald(a)	.080	.346	1.00	.602	1.00	1.00	.919	1.00	1.00	.129	.431	1.00	.768	1.00	1.00	.976	1.00	1.00
Wald(b)	.005	.030	.073	.000	.004	.024	.000	.000	.004	.007	.056	.131	.000	.015	.072	.000	.006	.043
SWald(a)	.076	.353	1.00	.599	1.00	1.00	.930	1.00	1.00	.136	.423	1.00	.771	1.00	1.00	.979	1.00	1.00
SWald(b)	.005	.030	.073	.000	.004	.024	.000	.000	.004	.007	.056	.131	.000	.015	.072	.000	.006	.043
$n = 250$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.006	.042	.102	.005	.045	.091	.004	.045	.089	.004	.044	.096	.006	046	.101	.008	.055	.105
Max-t-Test(b)	.010	.048	.103	.007	.045	.100	.009	.046	.093	.010	.043	.099	.010	.055	.109	.014	.056	.116
Wald(a)	.046	.308	1.00	.207	1.00	1.00	.795	1.00	1.00	.061	.319	1.00	.266	1.00	1.00	.884	1.00	1.00
Wald(b)	.004	.036	.086	.002	.023	.064	.000	.004	.022	.011	.050	.108	.000	.028	.073	.000	.008	.044
SWald(a)	.046	.267	1.00	.202	1.00	1.00	.265	1.00	1.00	.057	.338	1.00	.814	1.00	1.00	.994	1.00	1.00
SWald(b)	.004	.036	.086	.002	.023	.064	.000	.004	.022	.011	.050	.108	.000	.028	.073	.000	.008	.044
$n = 500$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.005	.042	.087	.006	.045	.090	.005	.041	.085	.009	.061	.112	.012	.052	.105	.015	.057	.104
Max-t-Test(b)	.007	.046	.090	.007	.049	.091	.008	.041	.089	.013	.058	.116	.015	.055	.111	.015	.059	.116
Wald(a)	.038	.256	1.00	.138	1.00	1.00	.691	1.00	1.00	.047	.285	1.00	.145	1.00	1.00	.775	1.00	1.00
Wald(b)	.004	.043	.093	.003	.031	.086	.000	.013	.044	.009	.058	.119	.007	.044	.092	.000	.021	.054
SWald(a)	.043	.261	1.00	.114	1.00	1.00	.678	1.00	1.00	.045	.287	1.00	.128	1.00	1.00	.758	1.00	1.00
SWald(b)	.004	.043	.093	.003	.031	.086	.000	.013	.044	.009	.058	.119	.007	.044	.092	.000	.021	.054
$n = 1000$																		
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 158$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.006	,047	.103	.007	.045	.082	.008	.041	.089	.013	.062	.113	.011	.052	.099	.013	.054	.097
Max-t-Test(b)	.007	.038	.103	.008	.046	.092	.008	.044	.095	.014	.064	.114	.010	.054	.110	.013	.051	.104
Wald(a)	.036	.275	1.00	.080	1.00	1.00	.636	1.00	1.00	.039	.281	1.00	.101	1.00	1.00	.651	1.00	1.00
Wald(b)	.008	.052	.103	.006	.037	.083	.002	.019	.062	.012	.050	.095	.010	.041	.102	.000	.018	.063
SWald(a)	.039	.260	1.00	.101	1.00	1.00	.631	1.00	1.00	.035	.272	1.00	.100	1.00	1.00	.754	1.00	1.00
SWald(b)	.008	.052	.103	.006	.037	.083	.002	.019	.062	.012	.050	.095	.010	.041	.102	.000	.018	.063

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples. $H_0 : \theta_0 = 0$ is true. $x_{\theta,i,t}$ and $x_{\delta,i,t}$ are within and across block iid.

Table 8: Linear Regression: Robustness Check b: $k_\delta \in \{20, 40\}$, for $n \in \{100, 250, 500\}$

	$k_\delta = 20$									$k_\delta = 40$								
	$n = 100$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 50$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.015	.077	.159	.017	.108	.190	.016	.100	.192	.059	.177	.282	.076	.236	.413	.073	.265	.421
Max-t-Test(b)	.029	.118	.202	.034	.127	.233	.034	.141	.262	.091	.259	.388	.131	.361	.505	.136	.390	.552
Wald(a)	.222	.532	1.00	.915	1.00	1.00	1.00	1.00	1.00	.468	.766	1.00	.998	1.00	1.00	1.00	1.00	1.00
Wald(b)	.019	.112	.226	.002	.074	.241	.000	.045	.226	.098	.313	.467	.125	.595	.820	.122	.814	.952
SWald(a)	.198	.531	1.00	.907	1.00	1.00	.996	1.00	1.00	.332	.490	.775	.996	1.00	1.00	1.00	1.00	1.00
SWald(b)	.019	.112	.226	.002	.074	.241	.000	.045	.226	.098	.313	.467	.125	.595	.820	.122	.814	.952
	$n = 250$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 79$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(b)	.016	.069	.122	.009	.063	.121	.013	.070	.131	.016	.075	.137	.025	.090	.174	.026	.102	.185
Max-t-Test(b)	.016	.069	.133	.017	.066	.134	.015	.081	.150	.021	.084	.166	.027	.115	.205	.027	.127	.228
Wald(a)	.083	.345	1.00	.307	1.00	1.00	.928	1.00	1.00	.094	.369	1.00	.511	1.00	1.00	.968	1.00	1.00
Wald(b)	.016	.077	.132	.008	.060	.126	.000	.037	.112	.018	.078	.161	.018	.124	.252	.007	.127	.292
SWald(a)	.069	.299	1.00	.337	1.00	1.00	.926	1.00	1.00	.111	.408	1.00	.501	1.00	1.00	.974	1.00	1.00
SWald(b)	.016	.077	.132	.008	.060	.126	.000	.037	.112	.018	.078	.161	.018	.124	.252	.007	.127	.292
	$n = 500$																	
	$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$			$\mathring{k}_{\theta,n} = 10$			$\mathring{k}_{\theta,n} = 35$			$\mathring{k}_{\theta,n} = 112$		
Test / Size	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Max-Test(s)	.000	.027	.057	.004	.042	.096	.005	.050	.100	.000	.014	.073	.011	.075	.125	.017	.061	.112
Max-Test(b)	.009	.046	.098	.007	.053	.107	.007	.051	.107	.009	.056	.129	.017	.086	.135	.018	.067	.116
Max-t-Test(s)	.006	.039	.089	.009	.054	.109	.006	.057	.125	.003	.028	.090	.016	.068	.132	.016	.062	.127
Max-t-Test(b)	.013	.055	.106	.013	.059	.116	.007	.060	.132	.014	.067	.138	.020	.079	.144	.018	.064	.134
Wald(a)	.053	.296	1.00	.157	1.00	1.00	.817	1.00	1.00	.062	.304	1.00	.195	1.00	1.00	.881	1.00	1.00
Wald(b)	.012	.052	.098	.010	.056	.113	.002	.026	.073	.010	.067	.133	.019	.068	.139	.007	.073	.157
SWald(a)	.052	.293	1.00	.143	1.00	1.00	.808	1.00	1.00	.055	.336	1.00	.196	1.00	1.00	.882	1.00	1.00
SWald(b)	.012	.052	.098	.010	.056	.113	.002	.026	.073	.010	.067	.133	.019	.068	.139	.007	.073	.157

"b" = bootstrapped p-value; "a" =asymptotic test. "SWald" is the normalized Wald test. Bootstrapped p-vales are based on 1,000 independently drawn samples. $H_0 : \theta_0 = 0$ is true. $x_{\theta,i,t}$ and $x_{\delta,i,t}$ are within and across block iid.

References

Amemiya, T. (1973): "Regression Analysis when the Dependent Variable is a Truncated Normal," *Econometrica*, 41, 997–1016.

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Boyd, S., and L. Vanderberghe (2004): *Convex Optimization*. Cambridge Univ. Press.

Buldygin, V. V., and Y. V. Kozachenko (1980): "Sub-Gaussian Random Variables," *Ukrainian Mathematical Journal volume*, 32, 483–489.

Chang, J., X. Chen, and M. Wu (2021): "Central Limit Theorems for High Dimensional Dependent Data," Discussion paper, Dept. of Statistics, University of Illinios, Urbana-Champaign.

Chareka, P., O. Chareka, and S. Kennedy (2006): "Locally Sub-Gaussian Random Variables and the Strong Law of Large Numbers," *Altantic Electronic Journal of Mathematics*, 1, 75–81.

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

——— (2015): "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70.

Dudley, R. M. (1984): "A Course on Empirical Processes," in *Ecole Eté de Probabilités de St. Flour, Lecture Notes in Math*, vol. 1097, pp. 1–142. Springer, New York.

Hornik, K., M. Stinchcombe, and H. White (1989): "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366.

Jennrich, R. I. (1969): "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

Kahane, J. P. (1960): "Propriètès locales des fonctions á sèries de Fourier alèatoires," *Studia Mathematica*, 19, 1–25.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handnook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier Science.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

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

Stromberg, K. R. (1994): *Probability for Analysts*. Chapman & Hall/, New York.

Talagrand, M. (2011): *Mean Field Models for Spin Glasses*. Springer.

Vershynin, R. (2018): *High-Dimensional Probability*. Cambridge University Press, Cambridge, UK.

White, H. (1989): "An Additional Hidden Unit Test for Negliected NonNonlinear in Multilayer Feedforward Networks," in *Proceedings of the International Joint Conference on Neural Networks*, vol. II, pp. 451–455, Washington. IEEE Press, New York, NY.