Inference on Directionally Differentiable Functions
Supplemental Appendix

Zheng Fang
Department of Economics
Texas A&M University
zfang@tamu.edu

Andres Santos*
Department of Economics
U.C. Los Angeles
andres@econ.ucla.edu

August, 2018

This Supplemental Appendix to "Inference on Directionally Differentiable Functions" is organized as follows. Section S.1 derives a generalization of the local analysis in Section 3.4 in the main text that allows the data to be dependent. In turn, Section S.2 contains the proofs of the main results in the text, while Section S.3 collects auxiliary results that are employed in the proofs of Section S.2. Finally, Section S.4 illustrates how to verify our assumptions in Examples 2.1 and 2.2, as well as two additional Examples based on Andrews and Shi (2013) and Linton et al. (2010).

S.1 Local Analysis Under Contiguity

In this Section, we present a set of results that generalize the analysis in Section 3.4 to allow for certain forms of dependence. To this end, we let $\mathbf{P}^\infty$ denote the set of possible distributions for $\{X_i\}_{i=1}^\infty$ and for any $P^\infty \in \mathbf{P}^\infty$ we set P^n to equal the distribution of $\{X_i\}_{i=1}^n$ induced by P^∞. As in the main text, we make the dependence of the parameter estimated by $\hat\theta_n$ on the unknown distribution of $\{X_i\}_{i=1}^\infty$ explicit by introducing a map $\theta : \mathbf{P}^\infty \to \mathbb{D}_\phi$ and letting $\theta(P^\infty)$ denote the value said estimand takes when $\{X_i\}_{i=1}^\infty$ is distributed according to P^∞. Given the introduced notation, we impose an Assumption that may be viewed as a generalization of Assumption 5 in the main text.

Assumption S.1. *There is a vector space Λ and $P_{n,\cdot}^\infty : \Lambda \to \mathbf{P}^\infty$, so that for any $\lambda \in \Lambda$*

(i) There is a $\theta'(\lambda) \in \mathbb{D}_0$ such that $\|r_n\{\theta(P_{n,\lambda}^\infty) - \theta(P_{n,0}^\infty)\} - \theta'(\lambda)\|_{\mathbb{D}} \to 0$.

*Research supported by NSF Grant SES-1426882.

(ii) $\theta(P^\infty_{n,\lambda}) \in \mathbb{D}_\phi$ *and* $r_n\{\hat{\theta}_n - \theta(P^\infty_{n,\lambda})\} \overset{L_\lambda}{\to} \mathbb{G}_0$*, where* $\overset{L_\lambda}{\to}$ *denotes convergence in distribution under* $\{X_i\}_{i=1}^n$ *distributed according to* $P^n_{n,\lambda}$*.*

(iii) $\mathbb{G}_0$ *is tight and its support is included in* $\mathbb{D}_0$*.*

(iv) The measures $\{P^\infty_{n,0}\}_{n=1}^\infty$ *are constant in* n*, and for any sequence of sets* $\{A_n\}_{n=1}^\infty$ *satisfying* $P^n_{n,0}(\{X_i\}_{i=1}^n \in A_n) \to 0$ *it follows that* $P^n_{n,\lambda}(\{X_i\}_{i=1}^n \in A_n) \to 0$*.*

Intuitively, Assumption S.1 introduces sequences of distributions $\{P^\infty_{n,\lambda}\}_{n=1}^\infty$, indexed by a parameter λ, that are "local" to a common fixed distribution. The value $\lambda = 0$ corresponds to the distribution to which all $\{P^\infty_{n,\lambda}\}_{n=1}^\infty$ are local to – notice that by Assumption S.1(iv) $P^\infty_{n,0}$ is constant in n. We note the construction in Section 3.4 maps into this setting with $P^\infty_{n,\lambda} = \bigotimes_{i=1}^\infty P_{\lambda/\sqrt{n}}$ and $P^\infty_{n,0} = \bigotimes_{i=1}^\infty P_0$. In parallel to Assumption 5, Assumption S.1(i) then demands that the parameter $\theta(P^\infty_{n,\lambda})$ be suitably differentiable (compare to Assumption 5(ii)), while Assumption S.1(ii) requires that the estimator $\hat{\theta}_n$ be regular (compare to Assumption 5(ii)). In turn, Assumption S.1(iii) is identical to Assumption 5(iii). Finally, Assumption S.1(iv) formalizes the sense in which $P^\infty_{n,\lambda}$ is local to $P^\infty_{n,0}$. Such a requirement is known as $\{P^n_{n,\lambda}\}_{n=1}^\infty$ being contiguous to $\{P^n_{n,0}\}_{n=1}^\infty$ and it may be understood as an asymptotic analogue to absolute continuity of a measure with respect to another; see Strasser (1985) for additional discussion. Assumption S.1(iv) is satisfied by multiple standard constructions, including the i.i.d. setting in the main text. However, we note that Assumption S.1 can also accommodate certain time series applications; see Bickel et al. (1998) and Garel and Hallin (1995).

The following result is a generalization of Lemma 3.1 in the main text.

Lemma S.1.1. *If Assumptions 1 and S.1 hold, then it follows that for any* $\lambda \in \Lambda$

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta(P^\infty_{n,\lambda}))\} \overset{L_\lambda}{\to} \phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)), \tag{S.1}$$

where $\theta_0 = \theta(P^\infty_{n,0})$ *and* L_λ *denotes convergence in distribution under* $\{X_i\}_{i=1}^n \sim P^n_{n,\lambda}$*.*

It is worth noting that the conclusion of Lemma S.1.1 implies that $\phi(\hat{\theta}_n)$ is a regular estimator of the parameters $\phi(\theta_0)$ if and only if the distribution of

$$\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)) \tag{S.2}$$

is constant in $\lambda \in \Lambda$. This "invariance" requirement is closely related to a necessary and sufficient condition for the consistency of the standard bootstrap that can be derived from results in Dümbgen (1993) (see (21)). In particular, if the closure of $\{\theta'(\lambda) : \lambda \in \Lambda\}$ in $\mathbb{D}$ equals the support of $\mathbb{G}_0$, then it can be shown that $\phi(\hat{\theta}_n)$ is a regular estimator if and only if the standard bootstrap is consistent. Such conclusion complements Beran (1997), who shows that in finite dimensional likelihood models the parametric bootstrap is consistent if and only if the estimator is regular. Thus, we can conclude that the failure of the standard bootstrap is an innate characteristic of irregular models.

We conclude this Section of the Supplemental Appendix by generalizing Theorem 3.3, Corollary 3.2, and providing the proofs for the stated results.

Theorem S.1.1. *Let* $\{X_i\}_{i=1}^n$ *be distributed according to* $P_{n,\lambda}^n$ *and set* $\theta_0 = \theta(P_{n,0}^\infty)$*. If Assumptions 1, 3, 4, and S.1 hold,* $\phi(\theta_0) = 0$*, and the cdf of* $\phi'_{\theta_0}(\mathbb{G}_0)$ *is continuous and increasing at its* $1-\alpha$ *quantile, denoted* $c_{1-\alpha}$*, then it follows that*

$$\limsup_{n\to\infty} P_{n,\lambda}^n(r_n\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \geq P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) > c_{1-\alpha}). \tag{S.3}$$

Moreover: (i) Result (S.3) holds with equality whenever $c_{1-\alpha}$ *is a continuity point of the cdf of* $\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda))$*, and (ii) The limiting rejection probability equals* α *when* $\lambda = 0$*.*

Corollary S.1.1. *Let the conditions of Theorem S.1.1 hold,* $\mathbb{G}_0$ *be Gaussian,* $\mathbb{D}_0$ *be a convex set, and* $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ *be a convex map. If* $\phi(\theta(P_{n,\lambda}^\infty)) \leq 0$ *for all* n*, then*

$$\limsup_{n\to\infty} P_{n,\lambda}^n(r_n\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) = P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) > c_{1-\alpha}) \leq \alpha. \tag{S.4}$$

Proof of Lemma S.1.1: First note that $\theta_0 = \theta(P_{n,0}^\infty)$ and Assumption S.1(i) imply

$$\lim_{n\to\infty} \|r_n\{\theta(P_{n,\lambda}^\infty) - \theta_0\} - \theta'(\lambda)\|_{\mathbb{D}} = 0. \tag{S.5}$$

Hence, letting $t_n \equiv r_n^{-1}$, $h_n \equiv r_n\{\theta(P_{n,\lambda}^\infty) - \theta_0\}$ we note $\theta_0 + t_nh_n = \theta(P_{n,\lambda}^\infty) \in \mathbb{D}_\phi$, and by (S.5) that $\|h_n - h\|_{\mathbb{D}} = o(1)$ for $h \equiv \theta'(\lambda)$. Therefore, Assumption 1(ii) yields

$$\begin{aligned}\lim_{n\to\infty} \|r_n\{\phi(\theta(P_{n,\lambda}^\infty)) - \phi(\theta_0)\} - \phi'_{\theta_0}(\theta'(\lambda))\|_{\mathbb{E}} \\ = \lim_{n\to\infty} \|\frac{\phi(\theta_0 + t_nh_n) - \phi(\theta_0)}{t_n} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = 0.\end{aligned} \tag{S.6}$$

Next, note that Theorem 2.1 and Assumption S.1(ii) applied with $\lambda = 0$ implies

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} = \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \tag{S.7}$$

under $P_{n,0}^n$. However, by Assumption S.1(iv) the sequence $\{P_{n,\lambda}^n\}_{n=1}^\infty$ is contiguous to $\{P_{n,0}^n\}_{n=1}^\infty$ and therefore results (S.6) and (S.7) allow us to conclude

$$\begin{aligned}r_n\{\phi(\hat{\theta}_n) - \phi(\theta(P_{n,\lambda}^n)\} &= r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} - r_n\{\phi(\theta(P_{n,\lambda}^\infty)) - \phi(\theta_0)\} \\ &= \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) - \phi'_{\theta_0}(\theta'(\lambda)) + o_p(1)\end{aligned} \tag{S.8}$$

under $P_{n,\lambda}^n$ as well. Furthermore, Assumption S.1(ii) and result (S.5) together imply

$$r_n\{\hat{\theta}_n - \theta_0\} = r_n\{\hat{\theta}_n - \theta(P_{n,\lambda}^\infty)\} + r_n\{\theta(P_{n,\lambda}^\infty) - \theta_0\} \overset{L_\lambda}{\to} \mathbb{G}_0 + \theta'(\lambda). \tag{S.9}$$

Finally, note that since (S.9) and $\{P_{n,\lambda}^n\}_{n=1}^\infty$ being contiguous to $\{P_{n,0}^n\}_{n=1}^\infty$ imply the support of $\mathbb{G}_0 + \theta'(\lambda)$ is a (weak) subset of the support of $\mathbb{G}_0$, Assumption S.1(iii) yields

$$P(\mathbb{G}_0 + \theta'(\lambda) \in \mathbb{D}_0) = 1. \tag{S.10}$$

Thus, the Lemma follows from (S.8)-(S.10) and the continuous mapping theorem. ■

Proof of Theorem S.1.1: We begin by establishing that $\hat{c}_{1-\alpha}$ is consistent for $c_{1-\alpha}$. To this end, let F denote the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ an similarly define $\hat{F}_n$ to equal

$$\hat{F}_n(c) \equiv P(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \leq c|\{X_i\}_{i=1}^n). \tag{S.11}$$

Next, observe that Theorem 3.2 and Lemma 10.11 in Kosorok (2008) imply that

$$\hat{F}_n(c) = F(c) + o_p(1), \tag{S.12}$$

for all $c \in \mathbf{R}$ that are continuity points of F. Fix $\epsilon > 0$, and note that since F is increasing at $c_{1-\alpha}$ and the set of continuity of points of F is dense in $\mathbf{R}$, it follows that there exist points $c_1, c_2 \in \mathbf{R}$ such that: (i) $c_1 < c_{1-\alpha} < c_2$, (ii) $|c_1 - c_{1-\alpha}| < \epsilon$ and $|c_2 - c_{1-\alpha}| < \epsilon$, (iii) c_1 and c_2 are continuity points of F, and (iv) $F(c_1) + \delta < 1 - \alpha < F(c_2) - \delta$ for some $\delta > 0$. Given these properties, we can then conclude that

$$\begin{aligned}\limsup_{n\to\infty} P^n_{n,\lambda}(|\hat{c}_{1-\alpha} - c_{1-\alpha}| > \epsilon) \\ \leq \limsup_{n\to\infty}\{P^n_{n,\lambda}(|\hat{F}_n(c_1) - F(c_1)| > \delta) + P^n_{n,\lambda}(|\hat{F}_n(c_2) - F(c_2)| > \delta)\} = 0,\end{aligned} \tag{S.13}$$

due to (S.12). In particular, since $\epsilon > 0$ was arbitrary, it follows that $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$.

Recall $\xrightarrow{L_\lambda}$ denotes convergence in distribution under the law induced by $\{X_i\}_{i=1}^n \sim P^n_{n,\lambda}$. From Lemma S.1.1 we are then able to conclude that

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta(P^\infty_{n,\lambda}))\} \xrightarrow{L_\lambda} \phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)). \tag{S.14}$$

Moreover, letting $t_n \equiv r_n^{-1}$ and $h_n \equiv r_n\{\theta(P^\infty_{n,\lambda}) - \theta_0\}$ we note $\theta_0 + t_n h_n = \theta(P^\infty_{n,\lambda}) \in \mathbb{D}_\phi$, and by Assumption S.1(i) that $\|h_n - \theta'(\lambda)\|_{\mathbb{D}} = o(1)$. Hence, Assumption 1(ii) yields

$$\lim_{n\to\infty} r_n\{\phi(\theta(P^\infty_{n,\lambda})) - \phi(\theta_0)\} = \lim_{n\to\infty} \frac{\phi(\theta_0 + t_n h_n) - \phi(\theta_0)}{t_n} = \phi'_{\theta_0}(\theta'(\lambda)). \tag{S.15}$$

Thus, since $\phi(\theta_0) = 0$, we may combine results (S.14) and (S.15) together with Slutksy's theorem and the continuous mapping theorem to obtain that

$$r_n\phi(\hat{\theta}_n) \xrightarrow{L_\lambda} \phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)). \tag{S.16}$$

Employing that $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$, Theorem 1.3.4(ii) in van der Vaart and Wellner (1996), and result (S.16) together with Slutsky's theorem then yields that

$$\liminf_{n\to\infty} P^n_{n,\lambda}(r_n\phi(\hat\theta_n) > \hat c_{1-\alpha}) \geq P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) > c_{1-\alpha}), \tag{S.17}$$

which implies result (S.3). Further notice that if $c_{1-\alpha}$ is a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda))$, then we may apply Theorem 1.3.4(vi) in van der Vaart and Wellner (1996) to obtain equality in (S.17). In particular, if $\lambda = 0$, then $\theta'(0) = 0$ by Assumption S.1(i) and the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ is continuous at $c_{1-\alpha}$ by assumption. Thus, we obtain

$$\lim_{n\to\infty} P^n_{n,0}(r_n\phi(\hat\theta_n) > \hat c_{1-\alpha}) = P(\phi'_{\theta_0}(\mathbb{G}_0) > c_{1-\alpha}) = \alpha, \tag{S.18}$$

where the final equality follows by definition of $c_{1-\alpha}$. ■

Proof of Corollary S.1.1: First note that $\phi(\theta_0) = 0$ and $\phi(\theta(P^\infty_{n,\lambda})) \leq 0$ implies

$$0 \geq \lim_{n\to\infty} r_n\{\phi(\theta(P^\infty_{n,\lambda})) - \phi(\theta_0)\} = \phi'_{\theta_0}(\theta(\lambda)), \tag{S.19}$$

where the equality was established in the proof of Theorem S.1.1 (see (S.15)). Further note that by Assumption 1(i) and Theorem 7.1.7 in Bogachev (2007), $\mathbb{G}_0$ is regular. Since in addition $\mathbb{G}_0$ is tight by Assumption S.1(iii), it follows $\mathbb{G}_0$ is Radon. In particular, $\mathbb{G}_0 + \theta'(\lambda)$ is also Radon, and since ϕ'_{θ_0} is continuous and convex, Theorem 11.1(i) in Davydov et al. (1998) implies that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda))$ is everywhere continuous except possibly at the point

$$r_0 \equiv \inf\{r : P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) \leq r) > 0\}. \tag{S.20}$$

However, we also note that since ϕ'_{θ_0} is convex and homogenous of degree one, we have

$$\phi'_{\theta_0}(h_1 + h_2) = 2\phi'_{\theta_0}(\frac{h_1}{2} + \frac{h_2}{2}) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2) \tag{S.21}$$

for any $h_1, h_2 \in \mathbb{D}_0$. Therefore, employing results (S.19) and (S.21) we can conclude

$$\begin{aligned} r_0 &\leq \inf\{r : P(\phi'_{\theta_0}(\mathbb{G}_0) + \phi'_{\theta_0}(\theta'(\lambda)) \leq r) > 0\} \\ &\leq \inf\{r : P(\phi'_{\theta_0}(\mathbb{G}_0) \leq r) > 0\} < c_{1-\alpha}, \end{aligned} \tag{S.22}$$

where the final inequality follows from the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ being increasing at $c_{1-\alpha}$. In particular, we conclude from the above discussion that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda))$ is continuous at $c_{1-\alpha}$. Hence, from Theorem S.1.1 we obtain that

$$\begin{aligned} \limsup_{n\to\infty} P^n_{n,\lambda}(r_n\phi(\hat\theta_n) > \hat c_{1-\alpha}) &= P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) > c_{1-\alpha}) \\ &\leq P(\phi'_{\theta_0}(\mathbb{G}_0) + \phi'_{\theta_0}(\theta'(\lambda)) > c_{1-\alpha}) \leq \alpha, \end{aligned} \tag{S.23}$$

where the second inequality is implied by (S.21), and the final inequality follows from (S.19) and the definition of $c_{1-\alpha}$. ■

S.2 Main Results

Below, we include the proofs for the main results in the text.

Proof of Proposition 2.1: One direction is clear since, by definition, ϕ being Hadamard differentiable implies that its Hadamard directional derivative exists, equals the Hadamard derivative of ϕ, and hence must be linear.

Conversely suppose the Hadamard directional derivative $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ exists and is linear. Let $\{h_n\}$ and $\{t_n\}$ be sequences such that $h_n \to h \in \mathbb{D}_0$, $t_n \to 0$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n. Then note that from any subsequence $\{t_{n_k}\}$ we can extract a further subsequence $\{t_{n_{k_j}}\}$, such that either: (i) $t_{n_{k_j}} > 0$ for all j or (ii) $t_{n_{k_j}} < 0$ for all j. When (i) holds, ϕ being Hadamard directional differentiable, then immediately yields

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h). \tag{S.24}$$

On the other hand, if (ii) holds, then $h \in \mathbb{D}_0$ and $\mathbb{D}_0$ being a subspace implies $-h \in \mathbb{D}_0$. Therefore, Hadamard directional differentiability of ϕ and $-t_{n_{k_j}} > 0$ for all j imply

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} \\ = -\lim_{j\to\infty} \frac{\phi(\theta + (-t_{n_{k_j}})(-h_{n_{k_j}})) - \phi(\theta)}{-t_{n_{k_j}}} = -\phi'_\theta(-h) = \phi'_\theta(h), \tag{S.25}$$

where the final equality holds by the assumed linearity of ϕ'_θ. Thus, results (S.24) and (S.25) imply that every subsequence $\{t_{n_k}, h_{n_k}\}$ has a further subsequence along which

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h). \tag{S.26}$$

Since the subsequence $\{t_{n_k}, h_{n_k}\}$ is arbitrary, it follows that (S.26) must hold along the original sequence $\{t_n, h_n\}$ and hence ϕ is Hadamard differentiable tangentially to $\mathbb{D}_0$. ■

Proof of Theorem 2.1: The proof closely follows the proof of Theorem 3.9.4 in van der Vaart and Wellner (1996), and we include it here only for completeness. First, note that by Assumption 2(ii) we may assume without loss of generality that $\mathbb{D}_0$ is equal to the support of $\mathbb{G}_0$. Since the support of a random variable is closed and ϕ'_{θ_0} is continuous, Theorem 4.1 in Dugundji (1951) implies ϕ'_{θ_0} can be continuously extended to all of $\mathbb{D}$. Next, let $\mathbb{D}_n \equiv \{h \in \mathbb{D} : \theta_0 + h/r_n \in \mathbb{D}_\phi\}$ and define $g_n : \mathbb{D}_n \to \mathbb{E}$ by

$$g_n(h) \equiv r_n\{\phi(\theta_0 + \frac{h}{r_n}) - \phi(\theta_0)\} \tag{S.27}$$

for any $h \in \mathbb{D}_n$. Similarly, for any $(h_1, h_2) \in \mathbb{D}_n \times \mathbb{D}$, define $f_n : \mathbb{D}_n \times \mathbb{D} \to \mathbb{E} \times \mathbb{E}$ by

$$f_n(h_1, h_2) \equiv (g_n(h_1), \phi'_{\theta_0}(h_2)). \tag{S.28}$$

Then note that for any sequence $(h_{1n}, h_{2n}) \in \mathbb{D}_n \times \mathbb{D}$ satisfying $(h_{1n}, h_{2n}) \to (h_1, h_2)$ for some $(h_1, h_2) \in \mathbb{D}_0 \times \mathbb{D}_0$, it follows from Assumption 1(ii) and the continuity of ϕ'_{θ_0} that $f_n(h_{1n}, h_{2n}) \to (\phi'_{\theta_0}(h_1), \phi'_{\theta_0}(h_2))$. Therefore, Theorem 1.11.1 in van der Vaart and Wellner (1996) implies that as processes in $\mathbb{E} \times \mathbb{E}$ we have that

$$\begin{bmatrix} r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\ \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \phi'_{\theta_0}(\mathbb{G}_0) \\ \phi'_{\theta_0}(\mathbb{G}_0) \end{bmatrix}. \tag{S.29}$$

In particular, result (S.29) and the continuous mapping theorem allow us to conclude

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} - \phi'_{\theta_0}(r_n(\hat{\theta}_n - \theta_0)) \xrightarrow{L} 0. \tag{S.30}$$

Result (S.30) and Lemma 1.10.2(iii) in van der Vaart and Wellner (1996) then establishes claim (10), while $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$ follows directly from (S.29) and the continuous mapping theorem (or (10) and the continuous mapping theorem). ■

Proof of Theorem 3.1: Let P, $\mathbb{D}_L \subseteq \mathbb{D}$, and μ_0 respectively denote the distribution, support, and mean of $\mathbb{G}_0$. Since $\mathbb{D}_L$ is a vector space, it follows that $0 \in \mathbb{D}_L$ and $\mathbb{D}_L = \mathbb{D}_L + \mathbb{D}_L$, and hence Theorem S.3.1 implies that result (20) holds if and only if

$$E[f(\phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0))] = E[f(\phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + a_0) - \phi'_{\theta_0}(a_0))] \tag{S.31}$$

for $\bar{\mathbb{G}}_0 = \mathbb{G}_0 - \mu_0$, and all $a_0 \in \mathbb{D}_L$ and $f \in \mathrm{BL}_1(\mathbb{E})$. On the other hand, $\mathbb{D}_L$ is a subspace of $\mathbb{D}$, and therefore Proposition 2.1 implies that ϕ is Hadamard differentiable at $\theta_0 \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_L$ if and only if $\phi'_{\theta_0} : \mathbb{D}_L \to \mathbb{E}$ is linear. Thus, the claim of the Theorem will follow from establishing that (S.31) holds if and only if $\phi'_{\theta_0} : \mathbb{D}_L \to \mathbb{E}$ is linear. To this end, we note that one direction is trivial, since linearity of ϕ'_{θ_0} implies

$$\phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + a_0) - \phi'_{\theta_0}(a_0) = \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0) \tag{S.32}$$

P almost surely for all $a_0 \in \mathbb{D}_L$, and thus (S.31) must hold for any $f \in \mathrm{BL}_1(\mathbb{E})$.

The opposite direction is more challenging and requires us to introduce additional notation which closely follows Chapter 7 in Davydov et al. (1998). First, we note that by Lemma S.3.7 $\mathbb{D}_L$ is a separable Banach space under $\|\cdot\|_{\mathbb{D}}$. Next, let $\mathbb{D}_L^*$ denote the dual space of $\mathbb{D}_L$, and $\langle d^*, d\rangle_{\mathbb{D}} = d^*(d)$ for any $d \in \mathbb{D}_L$ and $d^* \in \mathbb{D}_L^*$. Similarly denote the dual space of $\mathbb{E}$ by $\mathbb{E}^*$ and corresponding bilinear form by $\langle\cdot,\cdot\rangle_{\mathbb{E}}$. In addition, we let $\bar{P}$ denote the distribution of $\bar{\mathbb{G}}_0$, and note that $\bar{P}$ is a centered Gaussian measure whose support also equals $\mathbb{D}_L$ since $\mu_0 \in \mathbb{D}_L$ by Lemma S.3.7. We further let

$$\mathbb{D}'_{\bar{P}} \equiv \Big\{d' : \mathbb{D}_L \to \mathbf{R} : d' \text{ is linear, Borel-measurable, and } \int_{\mathbb{D}} (d'(d))^2 d\bar{P}(d) < \infty\Big\}, \tag{S.33}$$

and with some abuse of notation also write $d'(d) = \langle d', d\rangle_{\mathbb{D}}$ for any $d' \in \mathbb{D}'_{\bar{P}}$ and $d \in \mathbb{D}_L$. Finally, for each $h \in \mathbb{D}_L$ we let $\bar{P}^h$ denote the law of $\bar{\mathbb{G}}_0 + h$, write $\bar{P}^h \ll \bar{P}$ whenever $\bar{P}^h$ is absolutely continuous with respect to $\bar{P}$, and define the set

$$\mathbb{H}_{\bar{P}} \equiv \{h \in \mathbb{D}_L : \bar{P}^{rh} \ll \bar{P} \text{ for all } r \in \mathbf{R}\}. \tag{S.34}$$

To proceed, note that since $\mathbb{D}_L$ is separable, the Borel σ-algebra, the σ-algebra generated by the weak topology, and the cylindrical σ-algebra all coincide (Ledoux and Talagrand, 1991, p. 38). Furthermore, by Theorem 7.1.7 in Bogachev (2007), $\bar{P}$ is Radon and thus by Theorem 7.1 in Davydov et al. (1998), it follows that there exists a linear map $I : \mathbb{H}_{\bar{P}} \to \mathbb{D}'_{\bar{P}}$ such that for every $h \in \mathbb{H}_{\bar{P}}$ we have

$$\frac{d\bar{P}^h}{d\bar{P}}(d) = \exp\Big\{\langle d, Ih\rangle_{\mathbb{D}} - \frac{1}{2}\sigma^2(h)\Big\} \qquad \sigma^2(h) \equiv \int_{\mathbb{D}} \langle d, Ih\rangle^2_{\mathbb{D}} d\bar{P}(d). \tag{S.35}$$

Next, fix an arbitrary $e^* \in \mathbb{E}^*$ and $h \in \mathbb{H}_{\bar{P}}$. Then note that if (S.31) holds, then Lemma 1.3.12 in van der Vaart and Wellner (1996) implies $\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + rh) - \phi'_{\theta_0}(rh)\rangle_{\mathbb{E}}$ and $\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0)\rangle_{\mathbb{E}}$ must be equal in distribution for all $r \in \mathbf{R}$. Thus, their characteristic functions must be equal, and hence for all $r \geq 0$ and $t \in \mathbf{R}$

$$\begin{aligned} E[\exp\{it\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0)\rangle_{\mathbb{E}}\}] &= E[\exp\{it\{\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + rh) - \phi'_{\theta_0}(rh)\rangle_{\mathbb{E}}\}\}] \\ &= \exp\{-itr\langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}}\}E[\exp\{it\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + rh)\rangle_{\mathbb{E}}\}], \end{aligned} \tag{S.36}$$

where in the second equality we have exploited that $\phi'_{\theta_0}(rh) = r\phi'_{\theta_0}(h)$ due to ϕ'_{θ_0} being positively homogenous of degree one. Setting $C(t) \equiv E[\exp\{it\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0)\rangle_{\mathbb{E}}\}]$ and exploiting result (S.36) we can then obtain by direct calculation that for all $t \in \mathbf{R}$

$$\begin{aligned} &itC(t) \times \langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}} = \lim_{r\downarrow 0} \frac{1}{r}\{E[\exp\{it\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0 + rh)\rangle_{\mathbb{E}}\}] - C(t)\} \\ &= \lim_{r\downarrow 0} \frac{1}{r}\int_{\mathbb{D}} \Big\{\exp\Big\{it\langle e^*, \phi'_{\theta_0}(d + \mu_0)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\Big\} - C(t)\Big\}d\bar{P}(d) \end{aligned} \tag{S.37}$$

where in the second equality we exploited result (S.35), linearity of $I : \mathbb{H}_{\bar{P}} \to \mathbb{D}'_{\bar{P}}$ and that $h \in \mathbb{H}_{\bar{P}}$ implies $rh \in \mathbb{H}_{\bar{P}}$ for all $r \in \mathbf{R}$. Furthermore, by the mean value theorem

$$\begin{aligned}
&\sup_{r\in(0,1]} \frac{1}{r}\Big| \exp\Big\{it\langle e^*, \phi'_{\theta_0}(d+\mu_0)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\Big\} - \exp\{it\langle e^*, \phi'_{\theta_0}(d+\mu_0)\rangle_{\mathbb{E}}\}\Big| \\
&\quad \le \sup_{r\in(0,1]} \Big| \exp\Big\{it\langle e^*, \phi'_{\theta_0}(d+\mu_0)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\Big\} \times \{\langle d, Ih\rangle_{\mathbb{D}} - r\sigma^2(h)\}\Big| \\
&\quad \le \exp\{|\langle d, Ih\rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih\rangle_{\mathbb{D}}| + \sigma^2(h)\}, \qquad \text{(S.38)}
\end{aligned}$$

where the final inequality follows from $\sigma^2(h) \ge 0$ and $|\exp\{it\langle e^*, \phi'_{\theta_0}(d+\mu_0)\rangle_{\mathbb{E}}\}| \le 1$. Moreover, by Proposition 2.10.3 in Bogachev (1998) and $Ih \in \mathbb{D}'_{\bar{P}}$, it follows that $\langle \bar{\mathbb{G}}_0, Ih\rangle_{\mathbb{D}} \sim N(0, \sigma^2(h))$. Thus, we can obtain by direct calculation:

$$\begin{aligned}
&\int_{\mathbb{D}} \exp\{|\langle d, Ih\rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih\rangle_{\mathbb{D}}| + \sigma^2(h)\} d\bar{P}(d) \\
&\qquad = \int_{\mathbf{R}} \frac{\{|u| + \sigma^2(h)\}}{\sigma(h)\sqrt{2\pi}} \times \exp\Big\{|u| - \frac{u^2}{2\sigma^2(h)}\Big\} du < \infty. \qquad \text{(S.39)}
\end{aligned}$$

Hence, results (S.38) and (S.39) justify the use of the dominated convergence theorem in (S.37). Also note that $t \mapsto C(t)$ is the characteristic function of $\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0)\rangle_{\mathbb{E}}$ and hence it is continuous. Thus, since $C(0) = 1$ there exists a $t_0 > 0$ such that $C(t_0)t_0 \neq 0$. For such t_0 we then finally obtain from the above results that

$$\langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}} = -\frac{iE[\exp\{it_0\langle e^*, \phi'_{\theta_0}(\bar{\mathbb{G}}_0 + \mu_0)\rangle_{\mathbb{E}}\}\langle \bar{\mathbb{G}}_0, Ih\rangle_{\mathbb{D}}]}{t_0 C(t_0)}. \qquad \text{(S.40)}$$

To conclude note that $\mathbb{H}_{\bar{P}}$ being a vector space (Davydov et al., 1998, p. 38) and $I : \mathbb{H}_{\bar{P}} \to \mathbb{D}'_{\bar{P}}$ being linear imply together with result (S.39) that $h \mapsto \langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}}$ is linear on $\mathbb{H}_P$. Moreover, note that $h \mapsto \langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}}$ is also continuous on $\mathbb{D}_L$ due to continuity of ϕ'_{θ_0} and having $e^* \in \mathbb{E}^*$. Hence, since $\mathbb{H}_{\bar{P}}$ is dense in $\mathbb{D}_L$ by Proposition 7.4(ii) in Davydov et al. (1998) we can conclude that $\langle e^*, \phi'_{\theta_0}(\cdot)\rangle_{\mathbb{E}} : \mathbb{D}_L \to \mathbf{R}$ is linear and continuous. Since this result holds for all $e^* \in \mathbb{E}^*$, Lemma A.2 in van der Vaart (1991) implies $\phi'_{\theta_0} : \mathbb{D}_L \to \mathbb{E}$ must be linear and continuous, which establishes the Theorem. ■

Proof of Corollary 3.1: By Theorem 3.1 and Proposition 2.1 the bootstrap is consistent if and only if ϕ'_{θ_0} is linear. However, since $\mathbb{G}_0$ is Gaussian and $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ is continuous, Lemma 2.2.2 in Bogachev (1998) implies $\phi'_{\theta_0}(\mathbb{G}_0)$ must be Gaussian (on $\mathbb{E}$) whenever ϕ'_{θ_0} is linear, and hence the claim of the Corollary follows. ■

Proof of Theorem 3.2: Fix arbitrary $\epsilon > 0$, $\eta > 0$, and for notational convenience let $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. By Assumption 2(ii) there is a compact set $K_0 \subseteq \mathbb{D}_0$ such that

$$P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2}. \qquad \text{(S.41)}$$

Thus, by Lemma S.3.1 and the Portmanteau Theorem, we conclude that for any $\delta > 0$

$$\limsup_{n\to\infty} P(\mathbb{G}_n^* \notin K_0^\delta) \le P(\mathbb{G}_0 \notin K_0^\delta) \le P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2}. \tag{S.42}$$

On the other hand, since K_0 is compact, Assumption 4 yields that for some $\delta_0 > 0$

$$\limsup_{n\to\infty} P(\sup_{h\in K_0^{\delta_0}} \|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\|_{\mathbb{E}} > \epsilon) < \eta. \tag{S.43}$$

Next, note that Lemma 1.2.2(iii) in van der Vaart and Wellner (1996), $h \in \mathrm{BL}_1(\mathbb{E})$ being bounded by one and satisfying $|h(e_1) - h(e_2)| \le \|e_1 - e_2\|_{\mathbb{E}}$ for all $e_1, e_2 \in \mathbb{E}$, imply

$$\begin{aligned}
\sup_{f\in \mathrm{BL}_1(\mathbb{E})} &|E[f(\hat{\phi}_n'(\mathbb{G}_n^*))|\{X_i\}] - E[f(\phi_{\theta_0}'(\mathbb{G}_n^*))|\{X_i\}]| \\
&\le \sup_{f\in \mathrm{BL}_1(\mathbb{E})} E[|f(\hat{\phi}_n'(\mathbb{G}_n^*)) - f(\phi_{\theta_0}'(\mathbb{G}_n^*))||\{X_i\}] \\
&\le E[2 \times 1\{\mathbb{G}_n^* \notin K_0^{\delta_0}\} + \sup_{h\in K_0^{\delta_0}} \|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\|_{\mathbb{E}}|\{X_i\}] \\
&\le 2P(\mathbb{G}_n^* \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) + \sup_{h\in K_0^{\delta_0}} \|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\|_{\mathbb{E}},
\end{aligned} \tag{S.44}$$

where in the final inequality we exploited Lemma 1.2.2(i) in van der Vaart and Wellner (1996) and $\hat{\phi}_n' : \mathbb{D} \to \mathbb{E}$ depending only on $\{X_i\}_{i=1}^n$. Furthermore, Markov's inequality, Lemma 1.2.7 in van der Vaart and Wellner (1996), and result (S.42) yield

$$\limsup_{n\to\infty} P(P(\mathbb{G}_n^* \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) > \epsilon) \le \limsup_{n\to\infty} \frac{1}{\epsilon} P(\mathbb{G}_n^* \notin K_0^{\delta_0}) < \eta. \tag{S.45}$$

Next, also note that Assumption 3 and Theorem 10.8 in Kosorok (2008) imply that

$$\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E[f(\phi_{\theta_0}'(\mathbb{G}_n^*))|\{X_i\}_{i=1}^n] - E[f(\phi_{\theta_0}'(\mathbb{G}_0))]| = o_p(1). \tag{S.46}$$

Thus, by combining results (S.43), (S.44), (S.45) and (S.46) we can finally conclude:

$$\limsup_{n\to\infty} P(\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E[f(\hat{\phi}_n'(\mathbb{G}_n^*))|\{X_i\}_{i=1}^n] - E[f(\phi_{\theta_0}'(\mathbb{G}_0))]| > 3\epsilon) < 3\eta. \tag{S.47}$$

Since ϵ and η were arbitrary, the claim of the Theorem then follows from (S.47). ■

PROOFS OF LEMMA 3.1, THEOREM 3.3, AND COROLLARY 3.2: Lemma 3.1, Theorem 3.3, and Corollary 3.2 are respectively special cases of Lemma S.1.1, Theorem S.1.1, and Corollary S.1.1 established in the Supplemental Appendix. To see this, set $\Lambda = \mathbf{R}$, $P_{n,\lambda}^n \equiv \bigotimes_{i=1}^n P_{\lambda/\sqrt{n}}$, and note Assumptions S.1(i), S.1(ii), and S.1(iii) are implied by Assumptions 5(i), 5(ii), and 5(iii) respectively. In turn, we note Assumption S.1(iv) is satisfied due to $t \mapsto P_t$ being a path and Theorem 12.2.3 and Corollary 12.3.1 in

Lehmann and Romano (2005). ■

Proof of Proposition 4.1: We proceed by verifying Assumptions 1 and 2, and then employing Theorem 2.1 to obtain (42). To this end, define the maps $\phi_1 : \mathbb{H} \to \mathbb{H}$ to be given by $\phi_1(\theta) = \theta - \Pi_\Lambda \theta$, and $\phi_2 : \mathbb{H} \to \mathbf{R}$ by $\phi_2(\theta) \equiv \|\theta\|_\mathbb{H}$. Letting $\phi \equiv \phi_2 \circ \phi_1$ and noting $\phi_1(\theta_0) = 0$ due to $\theta_0 \in \Lambda$, we then obtain the equality

$$r_n\|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} = r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}. \tag{S.48}$$

By Lemma 4.6 in Zarantonello (1971), ϕ_1 is then Hadamard directionally differentiable at θ_0 with derivative $\phi'_{1,\theta_0} : \mathbb{H} \to \mathbb{H}$ given by $\phi'_{1,\theta_0}(h) = h - \Pi_{T_{\theta_0}} h$; see also (Shapiro, 1994, p. 135). Moreover, since ϕ_2 is Hadamard directionally differentiable at $0 \in \mathbb{H}$ with derivative $\phi'_{2,0}(h) = \|h\|_\mathbb{H}$, Proposition 3.6 in Shapiro (1990) implies ϕ is Hadamard directionally differentiable at θ_0 with $\phi'_{\theta_0} = \phi'_{2,0} \circ \phi'_{1,\theta_0}$. In particular, we have

$$\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_\mathbb{H}, \tag{S.49}$$

for any $h \in \mathbb{H}$. Thus, (S.49) verifies Assumption 1. Since Assumption 2 was directly imposed, the Proposition then follows form Theorem 2.1. ■

Proof of Proposition 4.2: We first observe that Λ being convex implies T_{θ_0} is a closed convex cone. Hence, by Proposition 46.5(4) in Zeidler (1984), it follows that $\|\Pi_{T_{\theta_0}} h\|^2_\mathbb{H} = \langle h, \Pi_{T_{\theta_0}} h\rangle_\mathbb{H}$ for any $h \in \mathbb{H}$. In particular, for any $h_1, h_2 \in \mathbb{H}$ we must have

$$\|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|^2_\mathbb{H} = \langle h_1 + h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle_\mathbb{H}. \tag{S.50}$$

However, Proposition 46.5(4) in Zeidler (1984) further implies that $\langle c, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle \leq 0$ for any $h_1, h_2 \in \mathbb{H}$ and $c \in T_{\theta_0}$. Therefore, since $\Pi_{T_{\theta_0}} h_1, \Pi_{T_{\theta_0}} h_2 \in T_{\theta_0}$, we can conclude from result (S.50) and the Cauchy Schwarz inequality

$$\begin{aligned} \|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|^2_\mathbb{H} &\leq \langle h_1 - \Pi_{T_{\theta_0}} h_1 + h_2 - \Pi_{T_{\theta_0}} h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle_\mathbb{H} \\ &\leq \|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_\mathbb{H} \times \|(h_1 - \Pi_{T_{\theta_0}} h_1) + (h_2 - \Pi_{T_{\theta_0}} h_2)\|_\mathbb{H}\ . \end{aligned} \tag{S.51}$$

Thus, employing result (S.51), the triangle inequality, and the definition of ϕ'_{θ_0} yield

$$\phi'_{\theta_0}(\lambda h_1 + (1-\lambda)h_2) \leq \phi'_{\theta_0}(\lambda h_1) + \phi'_{\theta_0}((1-\lambda)h_2) = \lambda\phi'_{\theta_0}(h_1) + (1-\lambda)\phi'_{\theta_0}(h_2) \tag{S.52}$$

for any $h_1, h_2 \in \mathbb{H}$, $0 \leq \lambda \leq 1$, and where in the final equality we employed that ϕ'_{θ_0} is positively homogenous of degree one. ■

Proof of Proposition 4.3: We first observe that for any $h_1, h_2 \in \mathbb{H}$ we must have

$$\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2) \le \sup_{\theta \in \Lambda: r_n\|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n} \{\|h_1 - \Pi_{T_\theta} h_1\|_{\mathbb{H}} - \|h_2 - \Pi_{T_\theta} h_2\|_{\mathbb{H}}\}$$
$$\le \sup_{\theta \in \Lambda: r_n\|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n} \{\|h_1 - \Pi_{T_\theta} h_2\|_{\mathbb{H}} - \|h_2 - \Pi_{T_\theta} h_2\|_{\mathbb{H}}\} \le \|h_1 - h_2\|_{\mathbb{H}}, \quad \text{(S.53)}$$

where the first inequality follows from the definition of $\hat{\phi}_n'(h)$, the second inequality is implied by $\|h_1 - \Pi_{T_\theta} h_1\|_{\mathbb{H}} \le \|h_1 - \Pi_{T_\theta} h_2\|_{\mathbb{H}}$ for all $\theta \in \Lambda$, and the third inequality holds by the triangle inequality. Result (S.53) further implies $\hat{\phi}_n'(h_2) - \hat{\phi}_n'(h_1) \le \|h_1 - h_2\|_{\mathbb{H}}$, and hence we can conclude $\hat{\phi}_n' : \mathbb{H} \to \mathbf{R}$ is Lipschitz – i.e. for any $h_1, h_2 \in \mathbb{H}$

$$|\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2)| \le \|h_1 - h_2\|_{\mathbb{H}}. \quad \text{(S.54)}$$

Thus, by Lemma S.3.6, in verifying $\hat{\phi}_n'$ satisfies Assumption 4 it suffices to show

$$|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)| = o_p(1) \quad \text{(S.55)}$$

for all $h \in \mathbb{H}$. To this end, note that convexity of Λ and Proposition 46.5(2) in Zeidler (1984) imply $\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \theta\|_{\mathbb{H}} \le \|\theta_0 - \theta\|_{\mathbb{H}}$ for any $\theta \in \mathbb{H}$. Thus, since $r_n\{\hat{\theta}_n - \theta_0\}$ is asymptotically tight by Assumption 2 and $\kappa_n \uparrow \infty$ by hypothesis, we conclude

$$\liminf_{n\to\infty} P(r_n\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n) \ge \liminf_{n\to\infty} P(r_n\|\theta_0 - \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n) = 1. \quad \text{(S.56)}$$

Moreover, the same arguments as in (S.56) and the triangle inequality further yield

$$\liminf_{n\to\infty} P(r_n\|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \le 2\kappa_n \text{ for all } \theta \in \Lambda \text{ s.t. } r_n\|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n)$$
$$\ge \liminf_{n\to\infty} P(r_n\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n) = 1. \quad \text{(S.57)}$$

Hence, from the definition of $\hat{\phi}_n'$ and results (S.56) and (S.57) we obtain for any $h \in \mathbb{H}$

$$\liminf_{n\to\infty} P(\|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} \le \hat{\phi}_n'(h) \le \sup_{\theta \in \Lambda: r_n\|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \le 2\kappa_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}) = 1. \quad \text{(S.58)}$$

Next, select a sequence $\{\theta_n\}_{n=1}^\infty \subseteq \Lambda$ such that $r_n\|\theta_n - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \le 2\kappa_n$ for all n and

$$\limsup_{n\to\infty} \{\sup_{\theta \in \Lambda: r_n\|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \le 2\kappa_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}\} = \lim_{n\to\infty} \|h - \Pi_{T_{\theta_n}} h\|_{\mathbb{H}}. \quad \text{(S.59)}$$

By Theorem 4.2.2 in Aubin and Frankowska (1990), the cone valued map $\theta \mapsto T_\theta$ is lower semicontinuous on Λ and hence since $\|\theta_n - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \le 2\kappa_n / r_n = o(1)$, there is a sequence $\{\tilde{h}_n\}_{n=1}^{\infty}$ such that $\tilde{h}_n \in T_{\theta_n}$ for all n and $\|\Pi_{T_{\theta_0}} h - \tilde{h}_n\|_{\mathbb{H}} = o(1)$. Thus,

$$\limsup_{n\to\infty}\{\sup_{\theta\in\Lambda : r_n\|\theta-\Pi_\Lambda\theta_0\|_{\mathbb{H}}\le 2\kappa_n} \|h-\Pi_{T_\theta}h\|_{\mathbb{H}}\} \\ = \lim_{n\to\infty}\|h-\Pi_{T_{\theta_n}}h\|_{\mathbb{H}} \le \lim_{n\to\infty}\|h-\tilde{h}_n\|_{\mathbb{H}} = \|h-\Pi_{T_{\theta_0}}h\|_{\mathbb{H}}, \quad \text{(S.60)}$$

where the first equality follows from (S.59), the inequality by $\tilde{h}_n \in T_{\theta_n}$, and the second equality by $\|\tilde{h}_n - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} = o(1)$. Hence, combining (S.58) and (S.60) we conclude that (S.55) holds, and by Lemma S.3.6 and (S.54) that $\hat{\phi}'_n$ satisfies Assumption 4. ■

S.3 Auxiliary Results

This Section contains auxiliary results employed in the Appendix to the main text.

Theorem S.3.1. *Let Assumptions 1, 2, and 3 hold,* $\mathbb{D}_L$ *denote the support of* $\mathbb{G}_0$, $0 \in \mathbb{D}_L$, *and* $\mathbb{D}_0 = \mathbb{D}_0 + \mathbb{D}_0$. *Then, the following statements are equivalent*

(i) $E[f(\phi'_{\theta_0}(\mathbb{G}_0))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))]$ *for all* $a_0 \in \mathbb{D}_L$ *and* $f \in BL_1(\mathbb{E})$.

(ii) $\sup_{f\in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1)$.

Proof: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. In addition, for notational convenience we let $\mathbb{G}_n \equiv r_n\{\hat{\theta}_n - \theta_0\}$ and $\mathbb{G}^*_n \equiv r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$. To begin, note that Lemma S.3.2 and the continuous mapping theorem imply that

$$(r_n\{\hat{\theta}^*_n - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\}) \\ = (r_n\{\hat{\theta}^*_n - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\}) \xrightarrow{L} (\mathbb{G}_1 + \mathbb{G}_2, \mathbb{G}_2) \quad \text{(S.61)}$$

on $\mathbb{D}\times\mathbb{D}$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent copies of $\mathbb{G}_0$. Further let $\Phi : \mathbb{D}_\phi \times \mathbb{D}_\phi \to \mathbb{E}$ be given by $\Phi(\theta_1, \theta_2) = \phi(\theta_1) - \phi(\theta_2)$ for any $\theta_1, \theta_2 \in \mathbb{D}_\phi \times \mathbb{D}_\phi$. Then observe that Assumption 1(ii) implies Φ is Hadamard directionally differentiable at (θ_0, θ_0) tangentially to $\mathbb{D}_0 \times \mathbb{D}_0$ with derivative $\Phi'_{\theta_0} : \mathbb{D}_0 \times \mathbb{D}_0 \to \mathbb{E}$ given by

$$\Phi'_{\theta_0}(h_1, h_2) = \phi'_{\theta_0}(h_1) - \phi'_{\theta_0}(h_2) \quad \text{(S.62)}$$

for any $(h_1, h_2) \in \mathbb{D}_0 \times \mathbb{D}_0$. Thus, by Assumptions 2(ii) and $\mathbb{D}_0 = \mathbb{D}_0 + \mathbb{D}_0$, Theorem 2.1, result (S.61), and $r_n\{\hat{\theta}^*_n - \theta_0\} = \mathbb{G}^*_n + \mathbb{G}_n$ we can conclude that

$$r_n\{\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)\} = r_n\{\Phi(\hat{\theta}^*_n, \hat{\theta}_n) - \Phi(\theta_0, \theta_0)\} \\ = \Phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n, \mathbb{G}_n) + o_p(1) = \phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n) + o_p(1). \quad \text{(S.63)}$$

Further observe that for any $\epsilon > 0$, it follows from the definition of $\text{BL}_1(\mathbb{E})$ that

$$\sup_{h\in\text{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat\theta_n^*) - \phi(\hat\theta_n)\}) - h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]| \\ \leq \epsilon + 2P^*(\|r_n\{\phi(\hat\theta_n^*) - \phi(\hat\theta_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n) \quad \text{(S.64)}$$

Moreover, Lemma 1.2.6 in van der Vaart and Wellner (1996) and result (S.63) also yield

$$E^*[P^*(\|r_n\{\phi(\hat\theta_n^*) - \phi(\hat\theta_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n)] \\ \leq P^*(\|r_n\{\phi(\hat\theta_n^*) - \phi(\hat\theta_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon) = o(1). \quad \text{(S.65)}$$

Therefore, since $\epsilon > 0$ was arbitrary, we obtain from results (S.64) and (S.65) that

$$\sup_{h\in\text{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat\theta_n^*) - \phi(\hat\theta_n)\})|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| \\ = \sup_{h\in\text{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| + o_p(1). \quad \text{(S.66)}$$

Thus, in establishing the Theorem, it suffices to study the right hand side of (S.66).

First Claim: We aim to show that (ii) implies (i). To this end, note by Lemma S.3.2

$$(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n), \mathbb{G}_n) \xrightarrow{L} (\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2), \mathbb{G}_2) \quad \text{(S.67)}$$

on $\mathbb{E} \times \mathbb{D}$ by the continuous mapping theorem. Let $f \in \text{BL}_1(\mathbb{E})$ and $g \in \text{BL}_1(\mathbb{D})$ satisfy $f(h_1) \geq 0$ and $g(h_2) \geq 0$ for any $h_1 \in \mathbb{E}$ and $h_2 \in \mathbb{D}$. By (S.67) we then have

$$\lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^*+\mathbb{G}_n)-\phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_1+\mathbb{G}_2)-\phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)]. \quad \text{(S.68)}$$

On the other hand, also note that if the bootstrap is consistent, then result (S.66) yields

$$\sup_{h\in\text{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1). \quad \text{(S.69)}$$

Moreover, since $\|g\|_\infty \leq 1$ and $\|f\|_\infty \leq 1$, it also follows that for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|g(\mathbb{G}_n)] \\ \leq \lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|] \\ \leq \lim_{n\to\infty} 2P^*(|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| > \epsilon) + \epsilon. \quad \text{(S.70)}$$

Thus, result (S.69), ϵ being arbitrary in (S.70), Lemma S.3.5(v), $g(h) \geq 0$ for all $h \in \mathbb{D}$, and $\mathbb{G}_n \xrightarrow{L} \mathbb{G}_2$ by result (S.67) allow us to conclude that

$$\lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)]$$
$$= \lim_{n\to\infty} E^*[E[f(\phi'_{\theta_0}(\mathbb{G}_0))]g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)]. \quad \text{(S.71)}$$

In addition, we also note that by Lemma 1.2.6 in van der Vaart and Wellner (1996)

$$\lim_{n\to\infty} E_*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)]$$
$$\leq \lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)]$$
$$\leq \lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] \quad \text{(S.72)}$$

since $\mathbb{G}_n$ is a function of $\{X_i\}_{i=1}^n$ only and $g(\mathbb{G}_n) \geq 0$. However, by (S.67) and Lemma 1.3.8 in van der Vaart and Wellner (1996), $(\phi'_{\theta_0}(\mathbb{G}_n^*+\mathbb{G}_n)-\phi'_{\theta_0}(\mathbb{G}_n), \mathbb{G}_n)$ is asymptotically measurable, and thus combining results (S.71) and (S.72) we can conclude:

$$\lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)]. \quad \text{(S.73)}$$

Hence, comparing (S.68) and (S.73) with $g \in \text{BL}_1(\mathbb{D})$ given by $g(a) = 1$ for all $a \in \mathbb{D}$,

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)] = E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))]E[g(\mathbb{G}_2)]$$
$$= E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)], \quad \text{(S.74)}$$

where the second equality follows again by (S.68) and (S.73). Since (S.74) must hold for any $f \in \text{BL}_1(\mathbb{E})$ and $g \in \text{BL}_1(\mathbb{D})$ with $f(h_1) \geq 0$ and $g(h_2) \geq 0$ for any $h_1 \in \mathbb{E}$ and $h_2 \in \mathbb{D}$, Lemma 1.4.2 in van der Vaart and Wellner (1996) implies $\phi'_{\theta_0}(\mathbb{G}_1+\mathbb{G}_2)-\phi'_{\theta_0}(\mathbb{G}_2)$ must be independent of $\mathbb{G}_2$, and hence (i) must hold by Lemma S.3.3.

<u>Second Claim:</u> To conclude, we show (i) implies (ii). Fix $\epsilon > 0$ and note that by Assumption 2, Lemma S.3.1, and Lemma 1.3.8 in van der Vaart and Wellner (1996), $\mathbb{G}_n$ and $\mathbb{G}_n^*$ are asymptotically tight. Hence, there is a compact set $K \subset \mathbb{D}$ such that

$$\liminf_{n\to\infty} P_*(\mathbb{G}_n^* \in K^\delta) \geq 1 - \epsilon \qquad \liminf_{n\to\infty} P_*(\mathbb{G}_n \in K^\delta) \geq 1 - \epsilon, \quad \text{(S.75)}$$

for any $\delta > 0$ and $K^\delta \equiv \{a \in \mathbb{D} : \inf_{b\in K} \|a-b\|_{\mathbb{D}} < \delta\}$. Furthermore, by the Portmanteau Theorem we may assume without loss of generality that K is a subset of the support of $\mathbb{G}_0$ and that $0 \in K$. Next, let $K + K \equiv \{a \in \mathbb{D} : a = b + c \text{ for some } b, c \in K\}$ and note that the compactness of K implies $K + K$ is also compact. Thus, by Lemma S.3.4 and continuity of $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$, there exist scalars $\delta_0 > 0$ and $\eta_0 > 0$ such that:

$$\sup_{a,b\in(K+K)^{\delta_0}:\|a-b\|_{\mathbb{D}}<\eta_0} \|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < \epsilon. \quad \text{(S.76)}$$

Next, for each $a \in K$, let $B_{\eta_0/2}(a) \equiv \{b \in \mathbb{D} : \|a - b\|_{\mathbb{D}} < \eta_0/2\}$. Since $\{B_{\eta_0/2}(a)\}_{a\in K}$ is an open cover of K, there exists a finite collection $\{B_{\eta_0/2}(a_j)\}_{j=1}^J$ also covering K. Therefore, since for any $b \in K^{\frac{\eta_0}{2}}$ there is a $\Pi b \in K$ such that $\|b - \Pi b\|_{\mathbb{D}} < \eta_0/2$, it follows that for every $b \in K^{\frac{\eta_0}{2}}$ there is a $1 \leq j \leq J$ such that $\|b - a_j\|_{\mathbb{D}} < \eta_0$. Setting $\delta_1 \equiv \min\{\delta_0, \eta_0\}/2$, we obtain that if $a \in K^{\delta_1}$ and $b \in K^{\delta_1}$, then: (i) $a + b \in (K + K)^{\delta_0}$ since $K^{\frac{\delta_0}{2}} + K^{\frac{\delta_0}{2}} \subseteq (K + K)^{\delta_0}$, (ii) there is a $1 \leq j \leq J$ such that $\|b - a_j\|_{\mathbb{D}} < \eta_0$, and (iii) $(a + a_j) \in (K + K)^{\delta_0}$ since $a_j \in K$ and $a \in K^{\frac{\delta_0}{2}}$. Therefore, since $0 \in K$, we can conclude from (S.76) that for every $b \in K^{\delta_1}$ there exists a $1 \leq j(b) \leq J$ such that

$$\sup_{a\in K^{\delta_1}} \|\{\phi'_{\theta_0}(a + b) - \phi'_{\theta_0}(b)\} - \{\phi'_{\theta_0}(a + a_{j(b)}) - \phi'_{\theta_0}(a_{j(b)})\}\|_{\mathbb{E}}$$
$$\leq \sup_{a,b\in(K+K)^{\delta_0}:\|a-b\|_{\mathbb{D}}<\eta_0} 2\|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < 2\epsilon. \quad \text{(S.77)}$$

In particular, if we define the set $\Delta_n \equiv \{\mathbb{G}_n^* \in K^{\delta_1}, \mathbb{G}_n \in K^{\delta_1}\}$, then (S.77) implies that for every realization of $\mathbb{G}_n$ there is an a_j independent of $\mathbb{G}_n^*$ such that

$$\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |(f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)) - f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j)))1\{\Delta_n\}| < 2\epsilon. \quad \text{(S.78)}$$

Letting Δ_n^c denote the complement of Δ_n, result (S.78) then allows us to conclude

$$\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^*+\mathbb{G}_n)-\phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]-E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| \leq 2P^*(\Delta_n^c|\{X_i\}_{i=1}^n)$$
$$+ \max_{1\leq j\leq J} \sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| + 2\epsilon \quad \text{(S.79)}$$

since $\|f\|_\infty \leq 1$ for all $f \in \mathrm{BL}_1(\mathbb{E})$. However, by Assumptions 3(i)-(ii) and 3(iv), and Theorem 10.8 in Kosorok (2008) it follows that for any $1 \leq j \leq J$

$$\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_j) - \phi'_{\theta_0}(a_j))]| = o_p(1). \quad \text{(S.80)}$$

Thus, since K is a subset of the support of $\mathbb{G}_0$ and property (i) holds by hypothesis, result (S.80), the continuous mapping theorem, and $J < \infty$ allow us to conclude that

$$\max_{1\leq j\leq J} \sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1). \quad \text{(S.81)}$$

Moreover, for any $\epsilon \in (0, 1)$ we also have by Markov's inequality, Lemma 1.2.6 in van der Vaart and Wellner (1996), $1\{\Delta_n^c\} \leq 1\{\mathbb{G}_n^* \notin K^{\delta_1}\} + 1\{\mathbb{G}_n \notin K^{\delta_1}\}$, and (S.75) that

$$\limsup_{n\to\infty} P^*(2P^*(\Delta_n^c|\{X_i\}_{i=1}^n) + 2\epsilon > 6\sqrt{\epsilon}) \leq \limsup_{n\to\infty} P^*(P^*(\Delta_n^c|\{X_i\}_{i=1}^n) > 2\sqrt{\epsilon})$$
$$\leq \frac{1}{2\sqrt{\epsilon}} \times \limsup_{n\to\infty}\{P^*(\mathbb{G}_n \notin K^{\delta_1}) + P^*(\mathbb{G}_n^* \notin K^{\delta_1})\} \leq \sqrt{\epsilon}. \quad \text{(S.82)}$$

Since $\epsilon > 0$ was arbitrary, combining (S.66), (S.79), (S.81), and (S.82) imply (ii) holds, thus establishing the claim of the Theorem. ■

Lemma S.3.1. *If Assumptions 1(i), 2(ii), and 3(i)-(iii) hold, then* $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \stackrel{L}{\to} \mathbb{G}_0$.

Proof: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. For notational simplicity also let $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. First, let $f \in \text{BL}_1(\mathbb{D})$, and then note that by Lemma S.3.5(i) and Lemma 1.2.6 in van der Vaart and Wellner (1996) we have that

$$\begin{aligned} E^*[f(\mathbb{G}_n^*)] - E[f(\mathbb{G}_0)] &\geq E^*[E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\ &\geq -E^*[|E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\geq -E^*[\sup_{f\in \text{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|]. \end{aligned} \tag{S.83}$$

Similarly, applying Lemma 1.2.6 in van der Vaart and Wellner (1996) once again together with Lemma S.3.5(ii), and exploiting that $f \in \text{BL}_1(\mathbb{D})$ we can conclude that

$$\begin{aligned} E_*[f(\mathbb{G}_n^*)] - E[f(\mathbb{G}_0)] &\leq E_*[E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\ &\leq E^*[|E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\leq E^*[\sup_{f\in \text{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|]. \end{aligned} \tag{S.84}$$

However, since $\|f\|_\infty \leq 1$ for all $f \in \text{BL}_1(\mathbb{D})$, it also follows that for any $\eta > 0$ we have

$$\begin{aligned} &E^*[\sup_{f\in \text{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\qquad \leq 2P^*(\sup_{f\in \text{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| > \eta) + \eta. \end{aligned} \tag{S.85}$$

Moreover, by Assumption 3(iii), $E^*[f(\mathbb{G}_n^*)] = E_*[f(\mathbb{G}_n^*)] + o(1)$. Thus, Assumption 3(ii), η being arbitrary, and results (S.83) and (S.84) together imply that

$$\lim_{n\to\infty} E^*[f(\mathbb{G}_n^*)] = E[f(\mathbb{G}_0)] \tag{S.86}$$

for any $f \in \text{BL}_1(\mathbb{D})$. Further note that since $\mathbb{G}_0$ is tight by Assumption 2(ii) and $\mathbb{D}$ is a Banach space by Assumption 1(i), Lemma 1.3.2 in van der Vaart and Wellner (1996) implies $\mathbb{G}_0$ is separable. Therefore, the claim of the Lemma follows from (S.86), Theorem 1.12.2 and Addendum 1.12.3 in van der Vaart and Wellner (1996). ■

Lemma S.3.2. *Let Assumptions 1(i), 2, 3(i)-(iii) hold, and* $\mathbb{G}_1, \mathbb{G}_2 \in \mathbb{D}$ *be independent random variables with the same law as* $\mathbb{G}_0$. *Then, it follows that on* $\mathbb{D} \times \mathbb{D}$

$$(r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}) \stackrel{L}{\to} (\mathbb{G}_1, \mathbb{G}_2). \tag{S.87}$$

Proof: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. For notational convenience we also let $\mathbb{G}_n \equiv r_n\{\hat{\theta}_n - \theta_0\}$ and $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. Then, note that Assumptions 2(i)-(ii), Lemma S.3.1, and Lemma 1.3.8 in van der Vaart and Wellner (1996) imply that both $\mathbb{G}_n$ and $\mathbb{G}_n^*$ are asymptotically measurable, and asymptotically tight in $\mathbb{D}$. Therefore, by Lemma 1.4.3 in van der Vaart and Wellner (1996) $(\mathbb{G}_n, \mathbb{G}_n^*)$ is asymptotically tight in $\mathbb{D} \times \mathbb{D}$ and asymptotically measurable as well. Thus, by Prohorov's theorem (Theorem 1.3.9 in van der Vaart and Wellner (1996)), each subsequence $\{(\mathbb{G}_{n_k}, \mathbb{G}_{n_k}^*)\}$ has an additional subsequence $\{(\mathbb{G}_{n_{k_j}}, \mathbb{G}_{n_{k_j}}^*)\}$ such that

$$(\mathbb{G}_{n_{k_j}}, \mathbb{G}_{n_{k_j}}^*) \xrightarrow{L} (\mathbb{Z}_1, \mathbb{Z}_2) \tag{S.88}$$

for a tight Borel random variable $\mathbb{Z} \equiv (\mathbb{Z}_1, \mathbb{Z}_2) \in \mathbb{D} \times \mathbb{D}$. Since the sequence $\{(\mathbb{G}_{n_k}, \mathbb{G}_{n_k}^*)\}$ was arbitrary, the Lemma follows if we show the law of $\mathbb{Z}$ equals that of $(\mathbb{G}_1, \mathbb{G}_2)$.

Towards this end, let $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ satisfy $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$. Then note that by result (S.88) it follows that:

$$\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}_{n_{k_j}}^*)] = E[f_1(\mathbb{Z}_1) f_2(\mathbb{Z}_2)]. \tag{S.89}$$

However, $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ satisfying $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$, Lemma 1.2.6 in van der Vaart and Wellner (1996), and Lemma S.3.5(iii) imply that

$$\begin{aligned}
&\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}_{n_{k_j}}^*)] - E^*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]] \\
&\qquad \geq \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) E^*[f_2(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n]] - E^*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]] \\
&\qquad \geq -\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) |E^*[f_2(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|] \\
&\qquad \geq -\lim_{j\to\infty} E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|],
\end{aligned} \tag{S.90}$$

where in the final inequality we exploited that $f_1 \in \mathrm{BL}_1(\mathbb{D})$. Similarly, Lemma 1.2.6 in van der Vaart and Wellner (1996), Lemma S.3.5(iv), and $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ also imply

$$\begin{aligned}
&\lim_{j\to\infty} E_*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}_{n_{k_j}}^*)] - E_*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]] \\
&\qquad \leq \lim_{j\to\infty} E_*[f_1(\mathbb{G}_{n_{k_j}}) E^*[f_2(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n]] - E_*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]] \\
&\qquad \leq \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) |E^*[f_2(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|] \\
&\qquad \leq \lim_{j\to\infty} E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_{n_{k_j}}^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|].
\end{aligned} \tag{S.91}$$

Thus, combining result (S.85) together with (S.90) and (S.91), and the fact that $(\mathbb{G}_n, \mathbb{G}_n^*)$

and $\mathbb{G}_n$ are asymptotically measurable, we can conclude that

$$\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})f_2(\mathbb{G}^*_{n_{k_j}})] = \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]]$$
$$= E[f_1(\mathbb{G}_0)]E[f_2(\mathbb{G}_0)], \tag{S.92}$$

where the final result follows from $\mathbb{G}_n \xrightarrow{L} \mathbb{G}_0$ in $\mathbb{D}$. Hence, (S.89) and (S.92) imply

$$E[f_1(\mathbb{Z}_1)f_2(\mathbb{Z}_2)] = E[f_1(\mathbb{G}_0)]E[f_2(\mathbb{G}_0)] \tag{S.93}$$

for all $f_1, f_2 \in \text{BL}_1(\mathbb{D})$ satisfying $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$. Since $\mathbb{Z}$ is tight on $\mathbb{D} \times \mathbb{D}$ it is also separable by Lemma 1.3.2 in van der Vaart and Wellner (1996) and Assumption 1(i), and hence result (S.93) and Lemma 1.4.2 in van der Vaart and Wellner (1996) imply the law of $\mathbb{Z}$ equals that of $(\mathbb{G}_1, \mathbb{G}_2)$. In view of (S.88), the claim of the Lemma then follows. ■

Lemma S.3.3. *Let Assumptions 1, 2(ii) hold, $\mathbb{D}_L$ denote the support of $\mathbb{G}_0$, $0 \in \mathbb{D}_L$, $\mathbb{D}_0 = \mathbb{D}_0 + \mathbb{D}_0$, and $\mathbb{G}_1$ be an independent copy of $\mathbb{G}_0$. If $\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_1)$ is independent of $\mathbb{G}_1$, then for any $a_0 \in \mathbb{D}_L$ and bounded continuous $f : \mathbb{E} \to \mathbf{R}$*

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))]. \tag{S.94}$$

Proof: We first note that since $\mathbb{D}_L \subseteq \mathbb{D}_0$ by Assumption 2(ii) and $\mathbb{G}_1$ is independent of $\mathbb{G}_0$, it follows that the support of $\mathbb{G}_0 + \mathbb{G}_1$ is included in $\mathbb{D}_0 + \mathbb{D}_0 = \mathbb{D}_0$, and hence $\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1)$ is well defined. Next, for any $a_0 \in \mathbb{D}$ and sequence $\{a_n\} \in \mathbb{D}$ with $\|a_0 - a_n\|_{\mathbb{D}} = o(1)$, we observe that continuity of ϕ'_{θ_0} and f, f being bounded, and the dominated convergence theorem allow us to conclude that

$$\lim_{n\to\infty} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_n) - \phi'_{\theta_0}(a_n))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))]. \tag{S.95}$$

Hence, letting $B_\epsilon(a_0) \equiv \{a \in \mathbb{D} : \|a_0 - a\|_{\mathbb{D}} < \epsilon\}$, we note that result (S.95) implies

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] = \liminf_{\epsilon\downarrow 0} \inf_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))]$$
$$\leq \limsup_{\epsilon\downarrow 0} \sup_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))]. \tag{S.96}$$

Letting L denote the law of $\mathbb{G}_0$, and for $\mathbb{G}_1$ and $\mathbb{G}_2$ independent copies of $\mathbb{G}_0$, we have

$$\inf_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))]P(\mathbb{G}_2 \in B_\epsilon(a_0))$$
$$\leq \int_{B_\epsilon(a_0)} \int_{\mathbb{D}_L} f(\phi'_{\theta_0}(z_1 + z_2) - \phi'_{\theta_0}(z_2))dL(z_1)dL(z_2)$$
$$\leq \sup_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))]P(\mathbb{G}_2 \in B_\epsilon(a_0)). \tag{S.97}$$

In particular, if $a_0 \in \mathbb{D}_L$, then $P(\mathbb{G}_2 \in B_\epsilon(a_0)) > 0$ for all $\epsilon > 0$, and thus we conclude

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] = \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))|\mathbb{G}_2 \in B_\epsilon(a_0)]$$
$$= \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))|\mathbb{G}_2 \in B_\epsilon(0)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))], \quad \text{(S.98)}$$

where the first equality follows from (S.96) and (S.97), the second by $\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2)$ being independent of $\mathbb{G}_2$ by hypothesis, and the final equality follows by results (S.96), (S.97), and $\phi'_{\theta_0}(0) = 0$ due to ϕ'_{θ_0} being homogenous of degree one. ■

Lemma S.3.4. *Let Assumption 1(i) hold, $\psi : \mathbb{D} \to \mathbb{E}$ be continuous, and $K \subset \mathbb{D}$ be compact. It then follows that for every $\epsilon > 0$ there exist $\delta > 0, \eta > 0$ such that*

$$\sup_{(a,b) \in K^\delta \times K^\delta : \|a-b\|_{\mathbb{D}} < \eta} \|\psi(a) - \psi(b)\|_{\mathbb{E}} < \epsilon. \quad \text{(S.99)}$$

PROOF: Fix $\epsilon > 0$ and note that since $\psi : \mathbb{D} \to \mathbb{E}$ is continuous, it follows that for every $a \in \mathbb{D}$ there exists a ζ_a such that $\|\psi(a) - \psi(b)\|_{\mathbb{E}} < \epsilon/2$ for all $b \in \mathbb{D}$ with $\|a - b\|_{\mathbb{D}} < \zeta_a$. Letting $B_{\zeta_a/4}(a) \equiv \{b \in \mathbb{D} : \|a - b\|_{\mathbb{D}} < \zeta_a/4\}$, then observe that $\{B_{\zeta_a/4}(a)\}_{a \in K}$ forms an open cover of K and hence, by compactness of K, there exists a finite subcover $\{B_{\zeta_{a_j}/4}(a_j)\}_{j=1}^J$ for some $J < \infty$. To establish the Lemma, we then let

$$\eta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4} \qquad \delta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4}. \quad \text{(S.100)}$$

For any $a \in K^\delta$, there then exists a $\Pi a \in K$ such that $\|a - \Pi a\|_{\mathbb{D}} < \delta$, and since $\{B_{\zeta_{a_j}/4}(a_j)\}_{j=1}^J$ covers K, there also is a $\bar{j}$ such that $\Pi a \in B_{\zeta_{a_{\bar{j}}}/4}(a_{\bar{j}})$. Thus, we have

$$\|a - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - \Pi a\|_{\mathbb{D}} + \|\Pi a - a_{\bar{j}}\|_{\mathbb{D}} < \delta + \frac{\zeta_{a_{\bar{j}}}}{4} \le \frac{\zeta_{a_{\bar{j}}}}{2}, \quad \text{(S.101)}$$

due to the choice of δ in (S.100). Moreover, if $b \in \mathbb{D}$ satisfies $\|a - b\|_{\mathbb{D}} < \eta$, then

$$\|b - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - b\|_{\mathbb{D}} + \|a - a_{\bar{j}}\|_{\mathbb{D}} < \eta + \frac{\zeta_{a_{\bar{j}}}}{2} \le \zeta_{a_{\bar{j}}}, \quad \text{(S.102)}$$

by the choice of η in (S.100). We conclude from (S.101), (S.102) that $a, b \in B_{\zeta_{a_{\bar{j}}}}(a_{\bar{j}})$, and

$$\|\psi(a) - \psi(b)\|_{\mathbb{E}} \le \|\psi(a) - \psi(a_{\bar{j}})\|_{\mathbb{E}} + \|\psi(b) - \psi(a_{\bar{j}})\|_{\mathbb{E}} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad \text{(S.103)}$$

by our choice of $\zeta_{a_{\bar{j}}}$. Thus, the Lemma follows from result (S.103). ■

Lemma S.3.5. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $c \in \mathbf{R}_+$, and $U : \Omega \to \mathbf{R}$ and $V : \Omega \to \mathbf{R}$ be arbitrary maps satisfying $U(\omega) \ge 0$ and $V(\omega) \ge 0$ for all $\omega \in \Omega$. If E^* and E_* denote outer and inner expectations respectively, then it follows that:*

(i) $E^*[U] - c \geq -E^*[|U - c|]$.

(ii) $E_*[U] - c \leq E^*[|U - c|]$.

(iii) $E^*[UV] - E^*[Uc] \geq -E^*[U|V - c|]$ *whenever* $\min\{E^*[UV], E^*[Uc]\} < \infty$.

(iv) $E_*[UV] - E_*[Uc] \leq E^*[U|V - c|]$ *whenever* $\min\{E_*[UV], E_*[Uc]\} < \infty$.

(v) $|E^*[UV] - E^*[Uc]| \leq E^*[U|V - c|]$ *whenever* $\min\{E_*[UV], E_*[Uc]\} < \infty$.

Proof: The arguments are simple and tedious, but unfortunately necessary to address the possible nonlinearity of inner and outer expectations. Throughout, for a map $T : \Omega \to \mathbf{R}$, we let T^* and T_* denote the minimal measurable majorant and the maximal measurable minorant of T respectively. We will also exploit the fact that:

$$E_*[T] = -E^*[-T], \tag{S.104}$$

and that $E^*[T] = E[T^*]$ whenever $E[T^*]$ exists, which in the context of this Lemma is always satisfied since all variables are positive.

To establish the first claim of the Lemma, note that Lemma 1.2.2(i) in van der Vaart and Wellner (1996) implies $U^* - c = (U - c)^*$. Therefore, (S.104) and $E_* \leq E^*$ yield

$$\begin{aligned} E^*[U] - c = E[U^* - c] = E[(U - c)^*] = E^*[U - c] \\ \geq E^*[-|U - c|] = -E_*[|U - c|] \geq -E^*[|U - c|]. \end{aligned} \tag{S.105}$$

Similarly, for the second claim of the Lemma, exploit that $E_* \leq E^*$, and once again employ Lemma 1.2.2(i) in van der Vaart and Wellner (1996) to conclude that

$$E_*[U] - c \leq E^*[U] - c = E[U^* - c] = E[(U - c)^*] \leq E^*[|U - c|]. \tag{S.106}$$

For the third claim, note that Lemma 1.2.2(iii) in van der Vaart and Wellner (1996) implies $|(UV)^* - (Uc)^*| \leq |UV - Uc|^*$. Thus, since $|U(V - c)| = U|V - c|$ as a result of $U(\omega) \geq 0$ for all $\omega \in \Omega$, we obtain from relationship (S.104) and $E_* \leq E^*$ that

$$\begin{aligned} E^*[UV] - E^*[Uc] = E[(UV)^* - (Uc)^*] \geq E[-|(UV)^* - (Uc)^*|] \\ \geq E[-|UV - Uc|^*] = -E_*[U|V - c|] \geq -E^*[U|V - c|]. \end{aligned} \tag{S.107}$$

Similarly, for the fourth claim of the Lemma, employ (S.104), that $|(-Uc)^* - (-UV)^*| \leq |(-Uc) - (-UV)|^*$ by Lemma 1.2.2(iii) in van der Vaart and Wellner (1996), and that $|UV - Uc| = U|V - c|$ due to $U(\omega) \geq 0$ for all $\omega \in \Omega$ to obtain that

$$\begin{aligned} E_*[UV] - E_*[Uc] = E[(-Uc)^* - (-UV)^*] \leq E[|(-Uc)^* - (-UV)^*|] \\ \leq E[|(-Uc) - (-UV)|^*] = E^*[U|V - c|]. \end{aligned} \tag{S.108}$$

Finally, for the fifth claim of the Lemma, note the same arguments as in (S.108) yield

$$E^*[UV] - E^*[Uc] = E[(Uc)^* - (UV)^*] \leq E[|(Uc)^* - (UV)^*|]$$
$$\leq E[|(Uc) - (UV)|^*] = E^*[U|V - c|]. \quad \text{(S.109)}$$

Thus, part (v) of the Lemma follows from part (iii) and (S.109). ■

Lemma S.3.6. *Let Assumption 1 hold, and suppose that for some $\kappa > 0$ and potentially random $C_n \in \mathbf{R}$ we have $\|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)\|_{\mathbb{E}} \leq C_n\|h_1 - h_2\|^{\kappa}_{\mathbb{D}}$ for all $h_1, h_2 \in \mathbb{D}$. If $C_n = O_p(1)$, then Assumption 4 holds provided that for all $h \in \mathbb{D}_0$ we have*

$$\|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1). \quad \text{(S.110)}$$

Proof: Fix $\epsilon > 0$ and note that since $C_n = O_p(1)$ by assumption, there exists some constant $0 < M < \infty$ such that for all n sufficiently large

$$P(C_n > M) < \epsilon. \quad \text{(S.111)}$$

Next, let $K_0 \subseteq \mathbb{D}_0$ be compact, and for any $h \in \mathbb{D}$ let $\Pi : \mathbb{D} \to K_0$ satisfy $\|h - \Pi h\|_{\mathbb{D}} = \inf_{a\in K_0} \|h - a\|_{\mathbb{D}}$ – here attainment is guaranteed by compactness. Since $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is continuous, Lemma S.3.4 then implies there exists a $\delta_1 > 0$ such that

$$\sup_{h\in K_0^{\delta_1}} \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} < \epsilon. \quad \text{(S.112)}$$

Next, set $\delta_2 < (\epsilon/M)^{1/\kappa}$ and note that by hypothesis we have outer almost surely that

$$\sup_{h\in K_0^{\delta_2}} \|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} \leq \sup_{h\in K_0^{\delta_2}} C_n\|h - \Pi h\|^{\kappa}_{\mathbb{E}} \leq C_n\delta_2^{\kappa}. \quad \text{(S.113)}$$

Defining $\delta_3 \equiv \min\{\delta_1, \delta_2\}$, exploiting (S.112), (S.113), and $\Pi h \in K_0$ we then conclude

$$\sup_{h\in K_0^{\delta_3}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}}$$
$$\leq \sup_{h\in K_0^{\delta_3}} \{\|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} + \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} + \|\hat{\phi}'_n(\Pi h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}}\}$$
$$\leq \sup_{h\in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} + \epsilon + C_n\delta_2^{\kappa} \quad \text{(S.114)}$$

outer almost surely. Thus, since $K_0^{\delta} \subseteq K_0^{\delta_3}$ for all $\delta \leq \delta_3$ we obtain from (S.114) that

$$P(\sup_{h\in K_0^{\delta}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 5\epsilon) \leq P(\sup_{h\in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) + P(C_n\delta_2^{\kappa} > \epsilon)$$
$$\leq P(\sup_{h\in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) + \epsilon, \quad \text{(S.115)}$$

where the second inequality is due to result (S.111) and the fact that $\delta_2 < (\epsilon/M)^{1/\kappa}$.

Next, note that since K_0 is compact, ϕ'_{θ_0} is uniformly continuous on K_0. Thus, we can find a finite collection $\{h_j\}_{j=1}^J$ with $J < \infty$ such that $h_j \in K_0$ for all j and

$$\sup_{h\in K_0} \min_{1\le j\le J} \max\{\|h - h_j\|_{\mathbb{D}}, \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}}\} < \min\{\delta_2, \epsilon\}. \tag{S.116}$$

In particular, since $\|\hat{\phi}'_n(h) - \hat{\phi}'_n(h_j)\|_{\mathbb{E}} \le C_n\|h - h_j\|_{\mathbb{D}}^{\kappa}$, we obtain from (S.116) that

$$\sup_{h\in K_0} \|\hat{\phi}'_{\theta_0}(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \le \max_{1\le j\le J} \|\hat{\phi}'_{\theta_0}(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} + C_n\delta_2^{\kappa} + \epsilon. \tag{S.117}$$

Thus, we can conclude from (S.117) and $\hat{\phi}'_n$ satisfying (S.110) for any $h \in \mathbb{D}_0$ that

$$\begin{aligned} P(\sup_{h\in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) &\le P(\max_{1\le j\le J} \|\hat{\phi}'_n(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} > \epsilon) + P(C_n\delta_2^{\kappa} > \epsilon) \\ &\le P(\max_{1\le j\le J} \|\hat{\phi}'_n(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} > \epsilon) + \epsilon, \end{aligned} \tag{S.118}$$

where in the final inequality we exploited (S.111) and $\delta_2 < (\epsilon/M)^{1/\kappa}$. The claim of the Lemma then follows from condition (S.110) and results (S.115) and (S.118). ■

Lemma S.3.7. *Let Assumptions 1(i), 2(ii) hold, and $\mathbb{G}_0$ be a Gaussian measure. If the support of $\mathbb{G}_0$ is a vector subspace of $\mathbb{D}$, then it is also a separable Banach space under $\|\cdot\|_{\mathbb{D}}$ and it includes the mean of $\mathbb{G}_0$.*

Proof: By Assumption 1 and Theorem 7.1.7 in Bogachev (2007) it follows that $\mathbb{G}_0$ is regular. Hence, since in addition $\mathbb{G}_0$ is tight by Assumption 2(ii), we can further conclude that $\mathbb{G}_0$ is Radon. Letting $\mathbb{D}_L$ and μ_0 respectively denote the support and the mean of $\mathbb{G}_0$, we then obtain by Theorem 3.6.1 in Bogachev (1998) that

$$\mathbb{D}_L = \mu_0 + \mathbb{D}_A, \tag{S.119}$$

where $\mathbb{D}_A$ is a closed separable subspace of $\mathbb{D}$. However, since $\mathbb{D}_L$ is also a vector subspace of $\mathbb{D}$ by hypothesis, it follows that $\mu_0 \in \mathbb{D}_A$ and hence $\mathbb{D}_L = \mathbb{D}_A$ and $\mu_0 \in \mathbb{D}_L$. Moreover, since $\mathbb{D}_A$ is a closed separable subspace of $\mathbb{D}$, we further conclude $\mathbb{D}_A$ is a separable Banach space under $\|\cdot\|_{\mathbb{D}}$ and the Lemma follows from $\mathbb{D}_L = \mathbb{D}_A$. ■

Lemma S.3.8. *Let Assumptions 1 and 2 hold and for some sequence $s_n \downarrow 0$ set*

$$\hat{\phi}'_n(h) \equiv \frac{1}{s_n}\{\phi(\hat{\theta}_n + s_n h) - \phi(\hat{\theta}_n)\} \tag{S.120}$$

for any $h \in \mathbb{D}$. It then follows that Assumption 4 holds provided that $r_n s_n \to \infty$.

PROOF: Let $K \subset \mathbb{D}_0$ be compact and $\epsilon > 0$ be arbitrary. First, we note that

$$\hat{\phi}'_n(h) = \frac{1}{s_n}\{\phi(\theta_0 + s_n\{h + \frac{r_n\{\hat{\theta}_n - \theta_0\}}{r_n s_n}\}) - \phi(\theta_0)\} + \frac{r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}}{r_n s_n} \tag{S.121}$$

for any $h \in \mathbb{D}$. Moreover, we note that since $\mathbb{G}_0$ is tight by Assumption 2(ii) it follows that $\phi'_{\theta_0}(\mathbb{G}_0)$ is tight by continuity of ϕ'_{θ_0}. Thus, Lemma 1.3.8 in van der Vaart and Wellner (1996) together with Assumption 2(i) and Theorem 2.1 imply that $r_n\{\hat{\theta}_n - \theta_0\}$ and $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ are asymptotically tight in $\mathbb{D}$ and $\mathbb{E}$ respectively. Since in addition $s_n r_n \to \infty$ by hypothesis, we obtain from result (S.121) that

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P(\sup_{h\in K^\delta} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \epsilon) \\
&\quad \le \lim_{\delta\downarrow 0}\limsup_{n\to\infty} P(\sup_{h\in K^\delta} \|\frac{1}{s_n}\{\phi(\theta_0 + s_n\{h + \frac{r_n\{\hat{\theta}_n - \theta_0\}}{r_n s_n}\}) - \phi(\theta_0)\} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \frac{\epsilon}{2}) \\
&\quad \le \frac{2}{\epsilon}\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \sup_{h\in K^{2\delta}} \|\frac{1}{s_n}\{\phi(\theta_0 + s_n h) - \phi(\theta_0)\} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \qquad \text{(S.122)}
\end{aligned}$$

where in the final inequality we employed Markov's inequality, Lemma S.3.4, and that $\{\hat{\theta}_n - \theta_0\}/s_n = o_p(1)$. Next, fix an arbitrary sequence $\delta_n \downarrow 0$ and note there then exists a subsequence $\{n_k\}_{k=1}^\infty$ such that for some $h_{n_k} \in K^{2\delta_{n_k}}$ we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{h\in K^{2\delta_n}} \|\frac{1}{s_n}\{\phi(\theta_0 + s_n h) - \phi(\theta_0)\} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \\
&\qquad = \lim_{k\to\infty} \|\frac{1}{s_{n_k}}\{\phi(\theta_0 + s_{n_k} h_{n_k}) - \phi(\theta_0)\} - \phi'_{\theta_0}(h_{n_k})\|_{\mathbb{E}}. \qquad \text{(S.123)}
\end{aligned}$$

However, since K is compact and $\delta_{n_k} \downarrow 0$, there must be a further subsequence $\{n_{k_j}\}_{j=1}^\infty$ such that $h_{n_{k_j}} \to h^*$ for some $h^* \in K \subseteq \mathbb{D}_0$. In particular, since ϕ is Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}_0$ by Assumption 1(ii) we conclude

$$\begin{aligned}
&\lim_{k\to\infty} \|\frac{1}{s_{n_k}}\{\phi(\theta_0 + s_{n_k} h_{n_k}) - \phi(\theta_0)\} - \phi'_{\theta_0}(h_{n_k})\|_{\mathbb{E}} \\
&\qquad = \lim_{j\to\infty} \|\frac{1}{s_{n_{k_j}}}\{\phi(\theta_0 + s_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta_0)\} - \phi'_{\theta_0}(h_{n_{k_j}})\|_{\mathbb{E}} = 0, \qquad \text{(S.124)}
\end{aligned}$$

where we exploited that $\phi'_{\theta_0}(h_{n_{k_j}}) \to \phi'_{\theta_0}(h^*)$ by continuity of ϕ'_{θ_0}. We thus obtain

$$\limsup_{n\to\infty} \sup_{h\in K^{2\delta_n}} \|\frac{1}{s_n}\{\phi(\theta_0 + s_n h) - \phi(\theta_0)\} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = 0 \tag{S.125}$$

from results (S.123) and (S.124). Since $\delta_n \downarrow 0$ was arbitrary in (S.125), we can conclude

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \sup_{h\in K^{2\delta}} \|\frac{1}{s_n}\{\phi(\theta_0 + s_n h) - \phi(\theta_0)\} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = 0, \tag{S.126}$$

which together with result (S.122) establishes the claim of the Lemma. ■

Lemma S.3.9. *Let Assumption 2(ii), 3(i)-(ii), and 3(iv) hold and suppose that*

$$E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})^*|\{X_i\}_{i=1}^n] - E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})_*|\{X_i\}_{i=1}^n] \xrightarrow{p} 0 \tag{S.127}$$

for any $h \in BL_1(\mathbb{D})$ and $h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})^*$ and $h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})_*$ respectively the minimal measurable majorant and maximal measurable minorant of $h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})$. It then follows that Assumption 3(iii) holds.

Proof: Let $h \in \mathrm{BL}_1(\mathbb{D})$, fix an arbitrary $\epsilon > 0$, and define the set A_n to be given by

$$A_n \equiv \{\{X_i\}_{i=1}^n : E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})^*|\{X_i\}_{i=1}^n] - E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})_*|\{X_i\}_{i=1}^n] > \epsilon\}. \tag{S.128}$$

Note that since $h \in \mathrm{BL}_1(\mathbb{D})$ implies $\|h\|_\infty \leq 1$, we obtain from $h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})^*$ and $h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})_*$ being measurable, Fubini's theorem, and (S.127) that

$$\begin{aligned}\lim_{n\to\infty} E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})^*] - E[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})_*]& \\ \leq \lim_{n\to\infty} 2P(\{X_i\}_{i=1}^n \in A_n) + \epsilon P(\{X_i\}_{i=1}^n \notin A_n) = \epsilon.&\end{aligned} \tag{S.129}$$

In particular, since $\epsilon > 0$ and $h \in \mathrm{BL}_1(\mathbb{D})$ were arbitrary, Lemma 1.2.1 in van der Vaart and Wellner (1996) and result (S.129) allows us to conclude for any $h \in \mathrm{BL}_1(\mathbb{D})$ that

$$\lim_{n\to\infty} E^*[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})] - E_*[h(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})] = 0, \tag{S.130}$$

where E^* and E_* respectively denote outer an inner expectations.

We next aim to show that $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ is asymptotically tight. To this end, note that since $\mathbb{G}_0$ is tight by Assumption 2(ii), for any $\epsilon > 0$ we may find a compact K with

$$P(\mathbb{G}_0 \in K) \geq 1 - \epsilon. \tag{S.131}$$

Further let $h_{K,m}(d) \equiv \{1 - \min_{\tilde{d}\in K} m\|d - \tilde{d}\|_{\mathbb{D}}\} \vee 0$, and note that: (i) $h_{K,m}/m \in \mathrm{BL}_1(\mathbb{D})$, (ii) $h_{K,m}(d) \to 1\{d \in K\}$ as $m \to \infty$, (iii) $1\{d \in K^\delta\} \geq h_{K,\frac{1}{\delta}}(d)$ for any $\delta > 0$. Hence, (S.130), Assumption 3(iv), and Lemma 1.2.6 van der Vaart and Wellner (1996) imply

$$\begin{aligned}\liminf_{n\to\infty} E_*[1\{r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \in K^\delta\}] \geq \liminf_{n\to\infty} E_*[h_{K,\frac{1}{\delta}}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})]& \\ \geq \liminf_{n\to\infty}\{E[h_{K,\frac{1}{\delta}}(\mathbb{G}_0)] - E^*[|E[h_{K,\frac{1}{\delta}}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - h_{K,\frac{1}{\delta}}(\mathbb{G}_0)|]\}.&\end{aligned} \tag{S.132}$$

In particular, since $h_{K,\frac{1}{\delta}}(d) \geq 1\{d \in K\}$ for all $\delta > 0$ and $h_{K,m}/m \in \mathrm{BL}_1(\mathbb{D})$, we can obtain from Assumption 3(ii) and arguing as in (S.129) that

$$\liminf_{n\to\infty} E_*[1\{r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \in K^\delta\}] \geq P(\mathbb{G}_0 \in K) \geq 1 - \epsilon, \tag{S.133}$$

where the final inequality is due to (S.131). Thus, $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ is asymptotically tight, and since $h/M \in \text{BL}_1(\mathbb{D})$ for any $h \in \text{BL}_M(\mathbb{D})$ implies (S.129) holds for any bounded Lipschitz function $h : \mathbb{D} \to \mathbf{R}$, we can conclude from Lemma 1.3.13 in van der Vaart and Wellner (1996) that $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ is asymptotically measurable. ■

S.4 Analysis of Examples

We next include a more detailed analysis of Examples 2.1 and 2.2, as well as two additional applications based on Andrews and Shi (2013) and Linton et al. (2010). For illustrative purposes, we confine our discussion to i.i.d. settings. We stress, however, that our general results allow for dependent data as well. Extending our discussion to such settings solely requires applying central limit theorems and resampling methods that allow for dependence; see, e.g., Dehling and Philipp (2002), Bühlmann (1995), Radulović (1996), and Politis et al. (1999) among others.

S.4.1 Revisiting Example 2.1

Recall that for some random variable $X \in \mathbf{R}$, the parameter of interest is given by

$$\max\{E[X], 0\}.$$

We assume the availability of an i.i.d. sample $\{X_i\}_{i=1}^n$, and employ the sample analogue

$$\max\{\bar{X}_n, 0\}$$

as an estimator (here $\bar{X}_n$ denotes the sample mean of $\{X_i\}_{i=1}^n$). To map this problem into our general framework simply let $\theta_0 = E[X]$, $\hat{\theta}_n = \bar{X}_n$, $\mathbb{D}_\phi = \mathbb{D} = \mathbb{E} = \mathbf{R}$, and $\phi : \mathbf{R} \to \mathbf{R}$ be given by $\phi(\theta) = \max\{\theta, 0\}$ for any $\theta \in \mathbf{R}$. Under these choices, the requirements for the Delta method of Theorem 2.1 are easily verified.

Lemma S.4.1. *Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with $E[X^2] < \infty$. Then Assumptions 1 and 2 hold with $r_n = \sqrt{n}$, $\mathbb{G}_0 \sim N(0, Var\{X\})$, $\mathbb{D}_0 = \mathbf{R}$, and $\phi'_{\theta_0} : \mathbf{R} \to \mathbf{R}$ given by*

$$\phi'_{\theta_0}(h) = \begin{cases} h & \textit{if } \theta_0 > 0 \\ \max\{h, 0\} & \textit{if } \theta_0 = 0 \\ 0 & \textit{if } \theta_0 < 0 \end{cases}.$$

Turning to the analysis in Section 3, we focus for concreteness on the nonparametric bootstrap of Efron (1979). Specifically, for $\{W_{ni}\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$ and jointly distributed according to a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$, we let $\hat{\theta}_n^*$ be given by

$$\hat{\theta}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i W_{ni}. \tag{S.134}$$

Notice that conditional on $\{X_i\}_{i=1}^n$, the distribution of $\hat{\theta}_n^*$ equals the law of $\bar{X}_n^* \equiv \sum_{i=1}^n X_i^*/n$, where $\{X_i^*\}_{i=1}^n$ are drawn with replacement from $\{X_i\}_{i=1}^n$. Thus, the "non-parametric" bootstrap maps into our setting through the representation in (S.134).

Our next result verifies $\hat{\theta}_n^*$ satisfies Assumption 3, and proposes an estimator $\hat{\phi}_n' : \mathbf{R} \to \mathbf{R}$ for $\phi_{\theta_0}' : \mathbf{R} \to \mathbf{R}$ that meets the requirements of Assumption 4.

Lemma S.4.2. *Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with $0 < E[X^2] < \infty$, and $\{W_{ni}\}_{i=1}^n$ be independent of $\{X_i\}_{i=1}^n$ and jointly distributed according to a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$. Then Assumption 3 is satisfied. Moreover, Assumption 4 holds with $\hat{\phi}_n' : \mathbf{R} \to \mathbf{R}$ given by*

$$\hat{\phi}_n'(h) = \begin{cases} h & \text{if } \sqrt{n}\bar{X}_n/\hat{\sigma}_n > \kappa_n \\ \max\{h, 0\} & \text{if } |\sqrt{n}\bar{X}_n/\hat{\sigma}_n| \leq \kappa_n \\ 0 & \text{if } \sqrt{n}\bar{X}_n/\hat{\sigma}_n < -\kappa_n \end{cases},$$

where $\hat{\sigma}_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/n$ and κ_n is a sequence satisfying $\kappa_n \uparrow \infty$ and $\kappa_n/\sqrt{n} \downarrow 0$.

Lemmas S.4.1 and S.4.2 together with Proposition 2.1 and Corollary 3.1 therefore imply that the "standard" bootstrap is consistent for the asymptotic distribution of $\max\{\bar{X}_n, 0\}$ if and only if $E[X] \neq 0$. On the other hand, Theorem 3.2 and Lemma S.4.2 further establish that said limiting distribution may be consistently estimated by the conditional law of $\hat{\phi}_n'(\sqrt{n}\{\bar{X}_n^* - \bar{X}_n\})$ given the data. Finally, we note that ϕ_{θ_0}' is convex, as required for the tests discussed in Section 3.4 to provide local size control.

Below we include the proofs for the results stated in this subsection.

Proof of Lemma S.4.1: Since $\{X_i\}_{i=1}^n$ is i.i.d. with $E[X^2] < \infty$, Assumption 2 holds trivially with $\mathbb{D}_0 = \mathbf{R}$ by the central limit theorem. To verify Assumption 1, fix a sequence of scalars $t_n \downarrow 0$ and $\{h_n\}_{n=1}^\infty \subset \mathbf{R}$ satisfying $h_n \to h$ for some $h \in \mathbf{R}$. If $\theta_0 > 0$, then $\theta_0 + t_n h_n > 0$ for n sufficiently large, and we obtain that

$$\lim_{n\to\infty} \|\frac{\phi(\theta_0 + t_n h_n) - \phi(\theta_0)}{t_n} - \phi_{\theta_0}'(h)\|_{\mathbb{E}} = \lim_{n\to\infty} |\frac{(\theta_0 + t_n h_n) - \theta_0}{t_n} - h| = \lim_{n\to\infty} |h_n - h| = 0. \tag{S.135}$$

By identical arguments, it follows that if $\theta_0 < 0$, then $\phi_{\theta_0}'(h) = 0$ for all $h \in \mathbf{R}$. Finally, for the case $\theta_0 = 0$, simply note that $h_n \to 0$ and continuity of $u \mapsto \max\{u, 0\}$ implies

$$\lim_{n\to\infty} |\frac{\max\{\theta_0 + t_n h_n, 0\} - \max\{\theta_0, 0\}}{t_n} - \max\{h, 0\}| \\ = \lim_{n\to\infty} |\max\{h_n, 0\} - \max\{h, 0\}| = 0. \quad \text{(S.136)}$$

Thus, $\phi'_{\theta_0}(h) = \max\{h, 0\}$ when $\theta_0 = 0$ and the Lemma follows. ■

PROOF OF LEMMA S.4.2: We note Assumption 3(i) is satisfied by definition of $\hat{\theta}^*_n$, while Assumption 3(ii) holds by Theorem 3.6.13 and Example 3.6.10 in van der Vaart and Wellner (1996). In turn, Assumptions 3(iii)-(iv) hold since $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ is a measurable function of $\{X_i, W_{in}\}_{i=1}^n$. Further note that for any $h_1, h_2 \in \mathbf{R}$ we have

$$|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)| \leq |h_1 - h_2| \quad \text{(S.137)}$$

almost surely. Moreover, since $\hat{\sigma}^2_n \xrightarrow{p} \sigma^2 \equiv \text{Var}\{X\} > 0$, we additionally obtain for $\mathbb{Z}_0$ a standard normal random variable that $\sqrt{n}\{\bar{X}_n - E[X]\}/\hat{\sigma}_n \xrightarrow{L} \mathbb{Z}_0$. In particular, if $E[X] = 0$, then for any $h \in \mathbf{R}$ we are able to conclude that

$$\lim_{n\to\infty} P(\hat{\phi}'_n(h) = \phi'_{\theta_0}(h)) \geq \lim_{n\to\infty} P(|\frac{\sqrt{n}\{\bar{X}_n - E[X]\}}{\hat{\sigma}_n}| \leq \kappa_n) = 1, \quad \text{(S.138)}$$

since $\kappa_n \uparrow \infty$. Alternatively, if $E[X] \neq 0$, then $\bar{X}_n/\hat{\sigma}_n \xrightarrow{p} E[X]/\sigma \neq 0$ and therefore

$$\lim_{n\to\infty} P(\hat{\phi}'_n(h) = \phi_{\theta_0}(h)) = 1 \quad \text{(S.139)}$$

due to the definition of $\hat{\phi}'_n$ and $\kappa_n/\sqrt{n} \downarrow 0$. The fact that $\hat{\phi}'_n$ satisfies Assumption 4 then follows from Lemma S.3.6 and results (S.138) and (S.139). ■

S.4.2 Revisiting Example 2.2

We next return to Example 2.2 and illustrate how to apply our general results. Recall that in this application the parameter of interest is given by

$$E[p'(E[ZZ'])^{-1}ZY_l + \max\{p'(E[ZZ'])^{-1}Z, 0\}(Y_u - Y_l)], \quad \text{(S.140)}$$

where $Z \in \mathbf{R}^{d_z}$, $Y_l, Y_u \in \mathbf{R}$, and $Y_l \leq Y_u$ almost surely. To map (S.140) into our framework, we define $b_0 \equiv (E[ZZ'])^{-1}p$ and set $\psi_0 : \mathbf{R}^{d_z} \to \mathbf{R}$ to be given by

$$\psi_0(b) \equiv E[b'ZY_l + \max\{b'Z, 0\}(Y_u - Y_l)]. \quad \text{(S.141)}$$

For any compact set $\mathbf{B} \subset \mathbf{R}^{d_z}$ containing b_0 in its interior, we in turn let $\mathbb{D} \equiv \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ and define the map $\phi : \mathbb{D} \to \mathbf{R}$ to satisfy for any $\theta = (b, \psi) \in \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ the relation

$$\phi(\theta) \equiv \begin{cases} \psi(b) & \text{if } b \in \mathbf{B} \\ 0 & \text{if } b \notin \mathbf{B} \end{cases} . \tag{S.142}$$

For our analysis, the value of $\phi(\theta)$ when $b \notin \mathbf{B}$ is not relevant. Here, the introduction of $\mathbf{B}$ is just a technical device to simplify our arguments; see also Bontemps et al. (2012) who instead restrict the estimator for $(E[ZZ'])^{-1}$ to be norm bounded.

The following assumption imposes the conditions we require for our analysis.

Assumption S.2 (For Example 2.2)**.**

(i) $\{Y_{l,i}, Y_{u,i}, Z_i\}_{i=1}^n$ *is an i.i.d. sample with* $E[(Y_u^2 + Y_l^2)(1 + \|Z\|^2)] \vee E[\|Z\|^4] < \infty$.

(ii) The matrix $E[ZZ']$ *is invertible.*

(iii) $\mathbf{B} \subset \mathbf{R}^{d_z}$ *is compact and* $b_0 \equiv (E[ZZ'])^{-1}p$ *belongs to its interior.*

We obtain estimators for $\theta_0 \equiv (b_0, \psi_0)$ by employing suitable sample analogues. In particular, we set $\hat{\theta}_n = (\hat{b}_n, \hat{\psi}_n)$ for $\hat{b}_n \equiv (\sum_{i=1}^n Z_i Z_i'/n)^{-1}p$ and $\hat{\psi}_n \in \ell^\infty(\mathbf{B})$ satisfying

$$\hat{\psi}_n(b) \equiv \frac{1}{n}\sum_{i=1}^n \{b'Z_i Y_{l,i} + \max\{b'Z_i, 0\}(Y_{u,i} - Y_{l,i})\}.$$

The estimator $\phi(\hat{\theta}_n)$ is then a direct sample analogue to the parameter of interest as defined in (S.140). Our next Lemma shows that Assumption S.2 implies Assumptions 1 and 2 are satisfied. Hence, the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ can be easily derived by the Delta method as stated in Theorem 2.1.

Lemma S.4.3. *If Assumption S.2 holds, then Assumptions 1 and 2 are satisfied with* $\mathbb{D}_\phi = \mathbb{D} = \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$, $\mathbb{E} = \mathbf{R}$, $r_n = \sqrt{n}$, $\mathbb{D}_0 = \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$, $\mathbb{G}_0 = (\mathbb{G}_b, \mathbb{G}_\psi)$ *centered Gaussian, and for any* $h = (h_b, h_\psi) \in \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$, *the map* $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ *satisfies*

$$\phi'_{\theta_0}(h) = h_\psi(b_0) + E[h_b' Z(Y_l + (Y_u - Y_l)1\{b_0'Z > 0\})] + E[\max\{h_b'Z, 0\}(Y_u - Y_l)1\{b_0'Z = 0\}].$$

Since the limiting distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ is centered Gaussian, Lemma S.4.3 together with Proposition 2.1 and Corollary 3.1 imply the "standard" bootstrap is consistent for the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ if and only if

$$P(b_0'Z = 0) = 0. \tag{S.143}$$

In previous work, Beresteanu and Molinari (2008) and Bontemps et al. (2012) show the consistency of the standard bootstrap under condition (S.143). Through a simple analytical calculation, our results enable to build on their analysis and show (S.143) is in fact necessary for the standard bootstrap to be consistent. In addition, Theorem 3.2 provides us with a framework for consistently estimating the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ when condition (S.143) fails to hold. To this end, we employ the analytic formula for ϕ'_{θ_0} derived in Lemma S.4.3 and define $\hat{\phi}'_n$ to be given by

$$\hat{\phi}'_n(h) = h_\psi(\hat{b}_n) + \frac{1}{n}\sum_{i=1}^n h'_b Z_i (Y_{l,i} + (Y_{u,i} - Y_{l,i}) 1\{\frac{\hat{b}'_n Z_i}{\|Z_i\|} > \frac{\kappa_n}{\sqrt{n}}\}) \\ + \frac{1}{n}\sum_{i=1}^n \max\{h'_b Z_i, 0\}(Y_{u,i} - Y_{l,i}) 1\{|\frac{\hat{b}'_n Z_i}{\|Z_i\|}| \leq \frac{\kappa_n}{\sqrt{n}}\},$$

where we understand $z/\|z\|$ to equal zero whenever $\|z\| = 0$, and κ_n is a sequence satisfying $\kappa_n \uparrow \infty$. In turn, we employ the nonparametric bootstrap to estimate the asymptotic distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$. For $\{W_{ni}\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$ and distributed according to a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$, we let $\hat{b}^*_n \equiv (\sum_{i=1}^n W_{ni} Z_i Z'_i / n)^{-1} p$ and define $\hat{\psi}^*_n \in \ell^\infty(\mathbf{B})$ by

$$\hat{\psi}^*_n(b) \equiv \frac{1}{n}\sum_{i=1}^n W_{ni}\{b' Z_i Y_{l,i} + \max\{b' Z_i, 0\}(Y_{u,i} - Y_{l,i})\}.$$

Our next result shows $\hat{\theta}^*_n = (\hat{b}^*_n, \hat{\psi}^*_n)$ and $\hat{\phi}'_n$ satisfy Assumptions 3 and 4.

Lemma S.4.4. *If Assumption S.3 holds and $\{W_{ni}\}_{i=1}^n$ is independent of $\{X_i\}_{i=1}^n$ and jointly distributed according to a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$, then Assumption 3 holds and $\hat{\phi}'_n$ satisfies Assumption 4 provided $\kappa_n \uparrow \infty$ and $\kappa_n/\sqrt{n} \to 0$.*

Theorem 3.2 and Lemma S.4.4 justify employing the conditional distribution of $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ given the data $\{X_i\}_{i=1}^n$ to estimate the asymptotic distribution of the estimator $\phi(\hat{\theta}_n)$. With regards to the analysis of Section 3.4, we note that ϕ'_{θ_0} is convex and $\phi'_{\theta_0}(\mathbb{G}_0)$ is continuously distributed on $\mathbf{R}$. Therefore, Corollary 3.2 implies the corresponding tests of the hypothesis that $\phi(\theta_0)$ is (weakly) negative are able to locally control size. With regards to the hypothesis testing problem

$$H_0 : \phi(\theta(P)) \geq 0 \qquad\qquad H_1 : \phi(\theta(P)) < 0, \tag{S.144}$$

we note that by Theorem 3.3 the limiting rejection probability of a test that rejects whenever $r_n \phi(\hat{\theta}_n) < c^\star$ for some critical value $c^\star$ is given by

$$\lim_{n\to\infty} P_{\lambda/\sqrt{n}}(r_n \phi(\hat{\theta}_n) < c^\star) = P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) < c^\star) \\ \leq P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)) < c^\star), \tag{S.145}$$

where the inequality holds whenever $\phi(\theta(P_{\lambda/\sqrt{n}})) \geq 0$ for all n, since in such a case

$\phi'_{\theta_0}(\theta'(\lambda)) \geq 0$. Letting $b(P) \equiv \{E_P[ZZ']\}^{-1}p$ and defining $\Delta(\lambda)$ to be given by

$$\Delta(\lambda) \equiv \lim_{n\to\infty} \sqrt{n}\{b(P_{\lambda/\sqrt{n}}) - b(P_0)\},$$

we may then obtain by direct calculation and Lemma S.4.3 the following lower bound

$$\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)) \geq \mathbb{G}_\psi(b_0) + E[\mathbb{G}_b' Z(Y_l + (Y_u - Y_l)1\{b_0'Z > 0\})] \\ + E[\mathbb{G}_b' Z(Y_u - Y_l)1\{b_0'Z = 0\}1\{\Delta(\lambda)'Z > 0\}]. \quad \text{(S.146)}$$

Combining results (S.145) and (S.146) we can then conclude that a test that rejects (S.144) whenever $r_n\phi(\hat{\theta}_n)$ is smaller than $c^\star$ will deliver local size control provided

$$\sup_{A\subseteq\{z:b_0'z=0\}} P(\mathbb{G}_\psi(b_0) + \mathbb{G}_b' E[Z(Y_l + (Y_u - Y_l)(1\{b_0'Z > 0\} + 1\{Z \in A\}))] < c^\star) \leq \alpha.$$

A consistent estimator for $c^\star$ may be obtained by appropriately modifying $\hat{\phi}'_n$. Moreover, we note that in most applications, the sets $A \subseteq \{z : b_0'z = 0\}$ satisfying $P(Z \in A) > 0$ consist only of points at which Z assigns positive probability. In such cases, when maximizing over sets A we need only examine sets consisting of points at which Z assigns positive probability. Finally, we also observe that since

$$\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) - \phi'_{\theta_0}(\theta'(\lambda)) \geq \mathbb{G}_\psi(b_0) + E[\mathbb{G}_b' Z(Y_l + (Y_u - Y_l)1\{b_0'Z > 0\})] \\ + E[\min\{\mathbb{G}_b' Z, 0\}(Y_u - Y_l)1\{b_0'Z = 0\}], \quad \text{(S.147)}$$

we may alternatively obtain (more conservative) critical values by employing the appropriate quantiles of the right hand side of (S.147).

Below, we include the proofs for Lemmas S.4.3 and S.4.4.

Proof of Lemma S.4.3: For notational simplicity we first define $\Omega_0 \equiv E[ZZ']$ and $\hat{\Omega}_n \equiv \sum_{i=1}^n Z_i Z_i'/n$. Since $E[\|Z\|^4] < \infty$, it follows $\sqrt{n}\{\hat{\Omega}_n - \Omega_0\} = O_p(1)$ and therefore

$$\hat{b} - b_0 = \{\hat{\Omega}_n^{-1} - \Omega_0^{-1}\}p = \Omega_0^{-1}\{\Omega_0 - \hat{\Omega}_n\}\hat{\Omega}_n^{-1}p = \Omega_0^{-1}\{\Omega_0 - \hat{\Omega}_n\}\Omega_0^{-1}p + o_p(n^{-1/2}). \quad \text{(S.148)}$$

Further defining $\mathcal{F} \equiv \{f : f(y_l, y_u, z) = b'zy_l + \max\{b'z(y_u - y_l), 0\}$ for some $b \in \mathbf{B}\}$, we note that Example 19.7 in van der Vaart (1998) and $\mathbf{B}$ being compact by Assumption S.2(iii) imply that the class $\mathcal{F}$ is Donsker. Thus, result (S.148) yields

$$(\sqrt{n}\{\hat{b}_n - b_0\}, \sqrt{n}\{\hat{\psi}_n - \psi_0\}) \xrightarrow{L} (\mathbb{G}_b, \mathbb{G}_\psi) \quad \text{(S.149)}$$

in $\mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$. Moreover, by Example 1.5.10 in van der Vaart and Wellner (1996), there is a version of $\mathbb{G}_\psi$ that is almost surely continuous with respect to

$$\rho_2^2(b_1, b_2) \equiv Var\{(b_1 - b_2)'ZY_l + (\max\{b_1'Z, 0\} - \max\{b_2'Z, 0\})(Y_u - Y_l)\}. \quad \text{(S.150)}$$

Since $\rho_2(b_1, b_2) \leq \|b_1 - b_2\|\{E[\|Z\|^2(Y_l^2 + (Y_u - Y_l)^2)]\}^{1/2}$ for any $b_1, b_2 \in \mathbf{R}^{d_z}$ by the Cauchy-Schwarz inequality, and $E[\|Z\|^2(Y_l^2 + Y_u^2)] < \infty$ by Assumption S.2(i) it follows that $\mathbb{G}_\psi$ is continuous with respect to the Euclidean norm, and hence $\mathbb{G}_\psi \in \mathcal{C}(\mathbf{B})$. We thus conclude Assumption 2 is satisfied with $\mathbb{D}_0 = \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$.

In order to verify Assumption 1, we let $t_n \downarrow 0$ and $h_n = (h_{bn}, h_{\psi n}) \in \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ form an arbitrary sequence satisfying $\|h_{bn} - h_b\| \vee \|h_{\psi n} - h_\psi\|_\infty = o(1)$ for some $h = (h_b, h_\psi) \in \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$. Since b_0 belongs to the interior of $\mathbf{B}$ by Assumption S.2(iii), it follows that $b_0 + t_n h_{bn} \in \mathbf{B}$ for n sufficiently large. Hence, definition (S.142) implies

$$\frac{1}{t_n}\{\phi(\theta_0 + t_n h_n) - \phi(\theta_0)\} = h_{\psi n}(b_0 + t_n h_{bn}) + \frac{1}{t_n}\{\psi_0(b_0 + t_n h_{bn}) - \psi_0(b_0)\} \quad \text{(S.151)}$$

for n sufficiently large. Further note that $h_\psi \in \mathcal{C}(\mathbf{B})$ and $\|h_{\psi n} - h_\psi\|_\infty = o(1)$ imply

$$\lim_{n\to\infty} |h_{\psi n}(b_0 + t_n h_{bn}) - h_\psi(b_0)| \leq \lim_{n\to\infty}\{\|h_{\psi n} - h_\psi\|_\infty + |h_\psi(b_0 + t_n h_{bn}) - h_\psi(b_0)|\} = 0. \quad \text{(S.152)}$$

In addition, Assumption S.2(i) and the dominated convergence theorem yield that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{t_n}\{\psi_0(b_0 + t_n h_{bn}) - \psi_0(b_0)\} \\
&= \lim_{n\to\infty}\{E[h_{bn}' Z Y_l] + \frac{1}{t_n} E[(\max\{(b_0 + t_n h_{bn})' Z, 0\} - \max\{b_0' Z, 0\})(Y_u - Y_l)]\} \\
&= E[h_b' Z(Y_l + (Y_u - Y_l)1\{b_0' Z > 0\})] + E[\max\{h_b' Z, 0\}(Y_u - Y_l)1\{b_0' Z = 0\}] \quad \text{(S.153)}
\end{aligned}$$

since $\|h_{bn} - h_b\| = o(1)$ by hypothesis. Thus, results (S.151) and (S.153) verify ϕ is indeed Hadamard directionally differentiable at θ_0 tangentially to $\mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$, which in turn implies Assumption 1 is satisfied. ■

Proof of Lemma S.4.4: We first let $\mathbf{M}^{d_z}$ denote the space of $d_z \times d_z$ matrices, and define $\nu : \mathbf{M}^{d_z} \to \mathbf{M}^{d_z}$ to be given by $\nu(\Omega) = \Omega^-$ where Ω^- denotes the Moore-Penrose pseudoinverse of Ω. Equipping $\mathbf{M}^{d_z}$ with the Frobenius norm $\|\cdot\|_F$, then note result (22) in Henderson and Searle (1981) implies ν is Hadamard differentiable at any invertible $\Omega \in \mathbf{M}^{d_z}$. Hence, setting $\pi : \mathbf{M}^{d_z} \times \ell^\infty(\mathbf{B}) \to \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ to be given by

$$\pi((\Omega, \psi)) = (\nu(\Omega)p, \psi) \quad \text{(S.154)}$$

for any $(\Omega, \psi) \in \mathbf{M}^{d_z} \times \ell^\infty(\mathbf{B})$, we can conclude that π is Hadamard differentiable at any (Ω, ψ) with Ω invertible. Moreover, we note that for $\Omega_0 \equiv E[ZZ']$, $\hat{\Omega}_n \equiv \sum_{i=1}^n Z_i Z_i'/n$, and $\hat{\Omega}_n^* \equiv \sum_{i=1}^n W_{ni} Z_i Z_i'/n$ we obtain that $\theta_0 = \pi((\Omega_0, \psi_0))$, $\hat{\theta}_n = \pi((\hat{\Omega}_n, \hat{\psi}_n))$, and $\hat{\theta}_n^* = \pi((\hat{\Omega}_n^*, \hat{\psi}_n^*))$. In particular, since $\mathcal{F} \equiv \{f : f(y_l, y_u, z) = b' z y_l + \max\{b' z(y_u - y_l), 0\}$ for some $b \in \mathbf{B}\}$ is Donsker by Example 19.7 in van der Vaart (1998), we can conclude from Example 2.10.8 and Theorems 3.6.13 and 3.9.11 in van der Vaart and

Wellner (1996) that Assumption 3(ii) is satisfied and in addition

$$E[h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})^*|\{X_i\}_{i=1}^n] - E[h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})_*|\{X_i\}_{i=1}^n] \xrightarrow{p} 0 \tag{S.155}$$

for any $h \in \text{BL}_1(\mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B}))$. Here, $h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})^*$ and $h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})_*$ denote the minimal measurable majorant and maximal measurable minorant of $h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})$ jointly in $\{X_i, W_{ni}\}_{i=1}^n$ respectively. We also note that Assumption 3(i) holds by construction, while Assumption 3(iv) is satisfied due to $f(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})$ being continuous in $\{W_{ni}\}_{i=1}^n$ for any continuous $f : \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B}) \to \mathbf{R}$. Therefore, Lemma S.4.3 implying Assumption 2(ii) holds together with (S.155) and Lemma S.3.9 establish Assumption 3(iii) is satisfied, which verifies Assumption 3.

In order to verify Assumption 4 holds, first fix an arbitrary $h = (h_b, h_\psi) \in \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$. Then note that since b_0 belongs to the interior of $\mathbf{B}$ by Assumption S.2(iii), it follows that $\hat{b}_n$ belongs to $\mathbf{B}$ with probability tending to one. As a result, we obtain that

$$h_\psi(\hat{b}_n) \xrightarrow{p} h_\psi(b_0) \tag{S.156}$$

due to $h_\psi \in \mathcal{C}(\mathbf{B})$, the continuous mapping theorem, and $\hat{b}_n \xrightarrow{p} b_0$ because Assumption 2(i) is satisfied by Lemma S.4.3. Next, we define the set of functions

$$\mathcal{G}_1 \equiv \{g : g(z) = 1\{\frac{z'\beta}{\|z\|} > \gamma\} \text{ for some } (\beta, \gamma) \in \mathbf{R}^{d_z+1}\}$$
$$\mathcal{G}_2 \equiv \{g : g(z) = 1\{\frac{|z'\beta|}{\|z\|} \leq \gamma\} \text{ for some } (\beta, \gamma) \in \mathbf{R}^{d_z+1}\}, \tag{S.157}$$

where $z/\|z\|$ is understood to equal zero whenever $\|z\| = 0$. Since the class $\mathcal{G}_0 \equiv \{g : g(z) = 1\{z'\beta \leq \gamma\}$ for some $(\beta, \gamma) \in \mathbf{R}^{d_z+1}\}$ is a VC subgraph class by Theorem B in Dudley (1979), Lemma 2.6.18(vii) and Theorem 2.10.6 in van der Vaart and Wellner (1996) establish $\mathcal{G}_1$ and $\mathcal{G}_2$ to be VC subgraph classes as well. Defining the classes

$$\mathcal{F}_1 \equiv \{f : f(y_l, y_u, z) = h_b'z(y_l + (y_u - y_l)g(z)) \text{ for some } g \in \mathcal{G}_1\}$$
$$\mathcal{F}_2 \equiv \{f : f(y_l, y_u, z) = \max\{h_b'z, 0\}(y_u - y_l)g(z) \text{ for some } g \in \mathcal{G}_2\}, \tag{S.158}$$

it then follows from Theorem 2.10.20 in van der Vaart and Wellner (1996) that $\mathcal{F}_1$ and $\mathcal{F}_2$ are Donsker as well. In particular, $\mathcal{F}_1$ being Donsker implies that

$$\frac{1}{n}\sum_{i=1}^n h_b'Z_i(Y_{l,i} + (Y_{u,i} - Y_{l,i})1\{\frac{\hat{b}_n'Z_i}{\|Z_i\|} > \frac{\kappa_n}{\sqrt{n}}\})$$
$$= E[h_b'Z(Y_l + (Y_u - Y_l)1\{\frac{\hat{b}_n'Z}{\|Z\|} > \frac{\kappa_n}{\sqrt{n}}\})] + o_p(1), \tag{S.159}$$

where the expectation is taken over $X = (Y_l, Y_u, Z)$ but not $\hat{b}_n$. We further note that since Lemma S.4.3 implies $\sqrt{n}\{\hat{b}_n - b_0\} = O_p(1)$ and $\kappa_n \uparrow \infty$ by assumption, we obtain

$$E[1\{\frac{\hat{b}_n' Z}{\|Z\|} > \frac{\kappa_n}{\sqrt{n}}\}1\{\frac{b_0' Z}{\|Z\|} \leq 0\}] \\ \leq E[1\{\frac{\sqrt{n}\{\hat{b}_n - b_0\}' Z}{\|Z\|} > \kappa_n\}] \leq 1\{\sqrt{n}\|\hat{b}_n - b_0\| > \kappa_n\} = o_p(1). \quad \text{(S.160)}$$

Hence, results (S.159) and (S.160) together with the Cauchy-Schwarz inequality yield

$$\frac{1}{n}\sum_{i=1}^n h_b' Z_i (Y_{l,i} + (Y_{u,i} - Y_{l,i})1\{\frac{\hat{b}_n' Z_i}{\|Z_i\|} > \frac{\kappa_n}{\sqrt{n}}\}) \\ = E[h_b' Z(Y_l + (Y_u - Y_l)1\{\frac{\hat{b}_n' Z}{\|Z\|} > \frac{\kappa_n}{\sqrt{n}}\})1\{b_0' Z > 0\}] + o_p(1) \\ = E[h_b' Z(Y_l + (Y_u - Y_l)1\{b_0' Z > 0\})] + o_p(1), \quad \text{(S.161)}$$

where we employed the continuous mapping theorem applied to $(\hat{b}_n, \kappa_n/\sqrt{n}) \xrightarrow{p} (b_0, 0)$ and the function $(\beta, \gamma) \mapsto E[h_b' Z(Y_l + (Y_u - Y_l)1\{\beta' Z/\|Z\| > \gamma\})1\{b_0' Z > 0\}]$, which is continuous at $(\beta, \gamma) = (b_0, 0)$ by the dominated convergence theorem. Similarly, we may employ that $\mathcal{F}_2$ is Donsker and argue analogously to (S.160) and (S.161) to obtain

$$\frac{1}{n}\sum_{i=1}^n \max\{h_b' Z_i, 0\}(Y_{u,i} - Y_{l,i})1\{\frac{|Z_i'\hat{b}_n|}{\|Z_i\|} \leq \frac{\kappa_n}{\sqrt{n}}\} \\ = E[\max\{h_b' Z, 0\}(Y_u - Y_l)1\{\frac{|Z'\hat{b}_n|}{\|Z\|} \leq \frac{\kappa_n}{\sqrt{n}}\}] + o_p(1) \\ = E[\max\{h_b' Z, 0\}(Y_u - Y_l)1\{Z' b_0 = 0\}] + o_p(1). \quad \text{(S.162)}$$

In particular, since $h = (h_b, h_\psi) \in \mathbf{R}^{d_z} \times \mathcal{C}(\mathbf{B})$ was arbitrary, results (S.156), (S.161), and (S.162) establish $\hat{\phi}_n'(h) = \phi_{\theta_0}'(h) + o_p(1)$ for any $h \in \mathbb{D}_0$. Furthermore, if we equip $\mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ with the norm $\|h\|_{\mathbb{D}} = \|h_b\| + \|h_\psi\|_\infty$ for any $h = (h_b, h_\psi)$, then we obtain

$$|\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2)| \leq (\|h_{\psi 1} - h_{\psi 2}\|_\infty + \|h_{b1} - h_{b2}\|)\{1 + \frac{1}{n}\sum_{i=1}^n \|Z_i\|(|Y_{l,i}| + (Y_{u,i} - Y_{l,i}))\} \\ = \{\|h_{\psi 1} - h_{\psi 2}\|_\infty + \|h_{b1} - h_{b2}\|\} \times O_p(1), \quad \text{(S.163)}$$

for any $h_1, h_2 \in \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ with $h_j = (h_{bj}, h_{\psi j})$, and where $E[\|Z\|(|Y_l| + |Y_u - Y_l|)] < \infty$ by Assumption S.2(i). Since we have shown $\hat{\phi}_n'(h) = \phi_{\theta_0}'(h) + o_p(1)$ for any $h \in \mathbb{D}_0$, result (S.163) and Lemma S.3.6 imply $\hat{\phi}_n'$ satisfies Assumption 4. ■

S.4.3 Example: Stochastic Dominance

As a further example, we consider a test of first order stochastic dominance originally studied by Linton et al. (2010). Specifically, suppose $X = (X^{(1)}, X^{(2)}) \in \mathbf{R}^2$ is a bivariate random variable and let F_j denote the marginal distribution of $X^{(j)}$. For $w : \mathbf{R} \to \mathbf{R}_+$ a positive integrable weighting function, Linton et al. (2010) propose a test of first order stochastic dominance based on the observation that $F_1 \leq F_2$ if and only if

$$\int_{\mathbf{R}} \max\{F_1(u) - F_2(u), 0\} w(u) du$$

is equal to zero. In particular, for $\hat{F}_{jn}$ the empirical cdf of $\{X_i^{(j)}\}_{i=1}^n$, the preceding observation suggests constructing a test by employing the statistic

$$\sqrt{n} \int_{\mathbf{R}} \max\{\hat{F}_{1n}(u) - \hat{F}_{2n}(u), 0\} w(u) du.$$

In order to map this problem into our framework we let $\theta_0 = (F_1, F_2) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, set $\hat{\theta}_n = (\hat{F}_{1n}, \hat{F}_{2n}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, and define $\phi : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \mathbf{R}$ by

$$\phi(\theta) = \int_{\mathbf{R}} \max\{\theta^{(1)}(u) - \theta^{(2)}(u), 0\} w(u) du$$

for any $\theta = (\theta^{(1)}, \theta^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. The following Assumption imposes mild regularity conditions that we will employ to verify the requirements in the main text.

Assumption S.3 (Stochastic Dominance Example)**.**

(i) $\{X_i\}_{i=1}^n$ *is an i.i.d. sample with* $X_i = (X_i^{(1)}, X_i^{(2)}) \in \mathbf{R}^2$.

(ii) The weight function $w : \mathbf{R} \to \mathbf{R}_+$ *satisfies* $\int_{\mathbf{R}} w(u) du < \infty$.

Our first Lemma shows Assumption S.3 implies Assumptions 1 and 2 hold.

Lemma S.4.5. *If Assumption S.3 holds, then Assumptions 1 and 2 are satisfied with* $\mathbb{D}_\phi = \mathbb{D}_0 = \mathbb{D} = \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, $\mathbb{E} = \mathbf{R}$, $r_n = \sqrt{n}$, $\mathbb{G}_0$ *centered Gaussian, and*

$$\phi'_{\theta_0}(h) = \int_{B_+(\theta_0)} (h^{(1)}(u) - h^{(2)}(u)) w(u) du + \int_{B_0(\theta_0)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\} w(u) du$$

for any $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, *and where the sets* $B_0(\theta_0)$ *and* $B_+(\theta_0)$ *are given by* $B_0(\theta_0) \equiv \{u \in \mathbf{R} : \theta_0^{(1)}(u) = \theta_0^{(2)}(u)\}$ *and* $B_+(\theta_0) \equiv \{u \in \mathbf{R} : \theta_0^{(1)}(u) > \theta_0^{(2)}(u)\}$.

The formula for the directional derivative ϕ'_{θ_0} obtained by Lemma S.4.5 implies that ϕ'_{θ_0} is linear if and only if $\int_{B_0(\theta_0)} w(u) du = 0$. Hence, since $\mathbb{G}_0$ is Gaussian, Proposition 2.1 and Corollary 3.1 imply the "standard" bootstrap is consistent for the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ if and only if $\int_{B_0(\theta_0)} w(u) du = 0$. Nonetheless, Theorem 3.2 still provides us with a valid resampling procedure provided that we can construct a suitable estimator $\hat{\phi}'_n$ for ϕ'_{θ_0}. To this end, we note any $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ satisfying the null hypothesis will be such that $B_+(\theta_0) = \emptyset$. Thus, for the purposes of testing, it is without loss of generality to assume ϕ'_{θ_0} satisfies

$$\phi'_{\theta_0}(h) = \int_{B_0(\theta_0)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\} w(u) du$$

for any $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. A natural estimator for ϕ'_{θ_0} is then given by

$$\hat{\phi}'_n(h) = \int_{\hat{B}_0(\theta_0)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\} w(u) du,$$

where $\hat{B}_0(\theta_0)$ is a suitable estimator for $B_0(\theta_0)$. For concreteness, we follow Linton et al. (2010) and for some sequence $\kappa_n \uparrow \infty$ define $\hat{B}_0(\theta_0)$ to equal

$$\hat{B}_0(\theta_0) \equiv \{u \in \mathbf{R} : \sqrt{n}|\hat{F}_{1n}(u) - \hat{F}_{2n}(u)| \leq \kappa_n\}.$$

As a final ingredient for inference, we also require a consistent estimator of the law of $\mathbb{G}_0 \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. Fortunately, the nonparametric bootstrap is valid. Therefore, for $\{W_{ni}\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$ and following a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$, we set $\hat{\theta}_n^* = (\hat{\theta}_n^{(1)*}, \hat{\theta}_n^{(2)*}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ where for $j \in \{1, 2\}$ the function $\hat{\theta}_n^{(j)*} \in \ell^\infty(\mathbf{R})$ is given by

$$\hat{\theta}_n^{(j)*}(u) = \frac{1}{n} \sum_{i=1}^n W_{ni} 1\{X_i^{(j)} \leq u\}. \tag{S.164}$$

Our next result verifies these choices for $\hat{\theta}_n^*$ and $\hat{\phi}'_n$ satisfy Assumptions 3 and 4.

Lemma S.4.6. *If Assumption S.3 holds and $\{W_{ni}\}_{i=1}^n$ is independent of $\{X_i\}_{i=1}^n$ and jointly distributed according to a multinomial distribution over $\{1, \ldots, n\}$ with each element having probability $1/n$, then Assumption 3 is satisfied. Moreover, if in addition $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ is such that $\theta_0^{(1)} \leq \theta_0^{(2)}$, then $\hat{\phi}'_n$ satisfies Assumption 4 provided $\kappa_n \uparrow \infty$ and $\kappa_n/\sqrt{n} \to 0$.*

Thus, Lemma S.4.6, together with Theorem 3.2, implies that the asymptotic distribution of the test statistic $\sqrt{n}\phi(\hat{\theta}_n)$ may be consistently estimated by the conditional law of $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})$ given the data. With regards to inference, we notice that ϕ'_{θ_0} is convex, as required for the resulting test to provide local size control. However, a complication arises in that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ may fail to be continuous at its $1 - \alpha$ quantile. In particular, continuity of the cdf fails when $\int_{B_0(\theta_0)} w(u) du = 0$, in which case $\phi'_{\theta_0}(\mathbb{G}_0) = 0$ almost surely. To obtain a test that controls size in such settings the critical values must be slightly adjusted; see the discussion in Linton et al. (2010).

Below we include the proofs for Lemmas S.4.5 and S.4.6.

PROOF OF LEMMA S.4.5: First let $t_n \downarrow 0$ and $\{h_n\}_{n=1}^\infty = \{(h_n^{(1)}, h_n^{(2)})\}_{n=1}^\infty$ be any sequence in $\ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ such that $\|h_n^{(1)} - h^{(1)}\|_\infty \vee \|h_n^{(2)} - h^{(2)}\|_\infty = o(1)$ for some $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. We further define the set $B_-(\theta_0)$ to equal

$$B_-(\theta_0) \equiv \{u \in \mathbf{R} : \theta_0^{(1)}(u) < \theta_0^{(2)}(u)\}. \tag{S.165}$$

Since $\|h_n^{(1)} - h_n^{(2)}\|_\infty \to \|h^{(1)} - h^{(2)}\|_\infty < \infty$, it then follows from $\theta_0^{(1)}(u) - \theta_0^{(2)}(u) < 0$ for all $u \in B_-(\theta_0)$ and the dominated convergence theorem that

$$\begin{aligned} &\int_{B_-(\theta_0)} \max\{\theta_0^{(1)}(u) - \theta_0^{(2)}(u) + t_n(h_n^{(1)}(u) - h_n^{(2)}(u)), 0\}w(u)du \\ &\lesssim t_n \int_{B_-(\theta_0)} 1\{t_n(h_n^{(1)}(u) - h_n^{(2)}(u)) \geq -(\theta^{(1)}(u) - \theta^{(2)}(u))\}w(u)du = o(t_n). \end{aligned} \tag{S.166}$$

Hence, since $\mathbf{R} \setminus B_-(\theta_0) = B_+(\theta_0) \cup B_0(\theta_0)$, result (S.166) allows us to conclude that

$$\begin{aligned} &\frac{1}{t_n}\{\phi(\theta + t_n h_n) - \phi(\theta)\} \\ &= \int_{\mathbf{R}\setminus B_-(\theta)} \max\{h_n^{(1)}(u) - h_n^{(2)}(u), -\frac{\theta^{(1)}(u) - \theta^{(2)}(u)}{t_n}\}w(u)du + o(1) = \phi'_\theta(h) + o(1), \end{aligned}$$

where the final equality follows from the dominated convergence theorem. Thus, we conclude that Assumption 1 holds with $\mathbb{D}_0 = \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$.

In order to verify Assumption 2, define the class of sets $\mathcal{C} = \{C \subseteq \mathbf{R}^2 : C = (-\infty, c] \times \mathbf{R}$ or $C = \mathbf{R} \times (-\infty, c]$ for some $c \in \mathbf{R}\}$ and note $\mathcal{C}$ has VC index less than or equal to two; see, e.g., Example 2.6.1 in van der Vaart and Wellner (1996). Hence, the empirical process converges in $\ell^\infty(\mathcal{C})$ by Theorem 2.5.2 in van der Vaart and Wellner (1996), which in turn implies by the continuous mapping theorem that Assumption 2 holds with $r_n = \sqrt{n}$ and $\mathbb{G}_0$ a centered Gaussian process. ■

Proof of Lemma S.4.6: First note that Assumption 3(i) is satisfied by construction. Further letting $\mathcal{C} = \{C \subseteq \mathbf{R}^2 : C = (-\infty, c] \times \mathbf{R}$ or $C = \mathbf{R} \times (-\infty, c]$ for some $c \in \mathbf{R}\}$, observe that Lemma S.4.5 implying Assumption 2 holds with $\mathbb{D} = \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ allows us to conclude that $\mathcal{C}$ is a Donsker class. Hence, Theorem 3.6.13 in van der Vaart and Wellner (1996) establishes that Assumption 3(ii) holds and that

$$E[h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})^*|\{X_i\}_{i=1}^n] - E[h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})_*|\{X_i\}_{i=1}^n] = o_p(1) \tag{S.167}$$

for any $h \in \mathrm{BL}_1(\mathbb{D})$, and where $h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})^*$ and $h(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})_*$ respectively denote the minimal measurable majorant and maximal measurable minorant of $h(\sqrt{n}\{\hat{\theta}_n - \hat{\theta}_n^*\})$. Further note that Assumption 3(iv) is satisfied since $\hat{\theta}_n^*$ is continuous in $\{W_{in}\}_{i=1}^n$ by equation (S.164). Thus, since Assumption 2 is satisfied by Lemma S.4.5, we may apply result (S.167) and Lemma S.3.9 to conclude Assumption 3(iii) is satisfied as well.

In order to verify Assumption 4, next notice that for any $h_1 = (h_1^{(1)}, h_1^{(2)})$ and

$h_2 = (h_2^{(1)}, h_2^{(2)})$ with $h_1, h_2 \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ we have that

$$|\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2)| \leq \int_{\mathbf{R}} w(u)du \times \{\|h_1^{(1)} - h_2^{(1)}\|_\infty + \|h_1^{(2)} - h_2^{(2)}\|_\infty\}, \tag{S.168}$$

which implies $\hat{\phi}_n' : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \mathbf{R}$ is Lipschitz with Lipschitz constant one. Moreover, note that since Assumption 2(ii) holds, it follows that for $j \in \{1,2\}$ we can conclude $\sqrt{n}\|\hat{F}_{jn} - F_{jn}\|_\infty = O_p(1)$. Therefore, we obtain from $\kappa_n \uparrow \infty$ that

$$\begin{aligned}\lim_{n\to\infty} & P(B_0(\theta_0) \subseteq \hat{B}_0(\theta_0)) \\ & \geq \lim_{n\to\infty} P(\sup_{u\in\mathbf{R}} \sqrt{n}|\{\hat{F}_{1n}(u) - \hat{F}_{2n}(u)\} - \{F_1(u) - F_2(u)\}| \leq \kappa_n) = 1. \end{aligned}\tag{S.169}$$

Next define $B_\delta(\theta_0) \equiv \{u \in \mathbf{R} : |F_1(u) - F_2(u)| \leq \delta\}$ for any $\delta > 0$, and select $\delta_n \downarrow 0$ to satisfy $\sqrt{n}\delta_n - \kappa_n \uparrow \infty$ (which is possible since $\kappa_n/\sqrt{n} \downarrow 0$). We thus obtain

$$\begin{aligned}&\lim_{n\to\infty} P(\hat{B}_0(\theta_0) \subseteq B_{\delta_n}(\theta_0)) \\ &\geq \lim_{n\to\infty} P(\sup_{u\in\mathbf{R}} \sqrt{n}|\{\hat{F}_{1n}(u) - \hat{F}_{2n}(u)\} - \{F_1(u) - F_2(u)\}| \leq \sqrt{n}\delta_n - \kappa_n) = 1. \end{aligned}\tag{S.170}$$

For any $h = (h_1, h_2) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, we then obtain from (S.169) and (S.170) that

$$\begin{aligned}|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)| &\leq \int_{B_{\delta_n}(\theta_0)\setminus B_0(\theta_0)} \max\{h_1(u) - h_2(u), 0\}w(u)du + o_p(1) \\ &\leq \{\|h_1\|_\infty + \|h_2\|_\infty\} \int_{\mathbf{R}} 1\{0 < |F_1(u) - F_2(u)| \leq \delta_n\}w(u)du + o_p(1). \end{aligned}\tag{S.171}$$

Hence, the dominated convergence theorem implies $\hat{\phi}_n'(h) = \phi_{\theta_0}'(h) + o_p(1)$ for any $h \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. Result (S.168) and Lemma S.3.6 then verify Assumption 4. ■

S.4.4 Example: Moment Inequalities

As our final example, we examine models defined by conditional moment inequality restrictions. In such models, tests of whether a parameter belongs to the identified set often correspond to testing, for some $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^{d_z}$, the null hypothesis

$$E[Y|Z] \leq 0 \tag{S.172}$$

almost surely. Andrews and Shi (2013) propose testing whether (S.172) holds based on the observations that (S.173) is equivalent to the restriction

$$\sup_{f\in\mathcal{F}} E[Yf(Z)] \leq 0, \tag{S.173}$$

where $\mathcal{F}$ is a suitably chosen collection of functions $f : \mathbf{R}^{d_z} \to \mathbf{R}_+$. Relation (S.173) suggests testing the null hypothesis that $E[Y|Z] \leq 0$ by employing the test statistic

$$\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i f(Z_i). \tag{S.174}$$

In order to map this problem into our framework we let $\mathbb{D}_\phi = \mathbb{D} = \ell^\infty(\mathcal{F})$ and set $\theta_0, \hat{\theta}_n \in \ell^\infty(\mathcal{F})$ to be the functions satisfying for any $f \in \mathcal{F}$ the relations

$$\theta_0(f) \equiv E[Yf(Z)] \qquad \hat{\theta}_n(f) \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i f(Z_i).$$

Defining $\phi : \ell^\infty(\mathcal{F}) \to \mathbf{R}$ to be given by $\phi(\theta) = \sup_{f \in \mathcal{F}} \theta(f)$, we then note that (S.174) is equal to $\sqrt{n}\phi(\hat{\theta}_n)$. As a final piece of notation we introduce the intrinsic (semi)metric

$$\rho_2(f_1, f_2) \equiv (\text{Var}\{Y(f_1(Z) - f_2(Z))\})^{1/2}. \tag{S.175}$$

Endowing $\mathcal{F}$ with ρ_2, we then let $\mathcal{C}(\mathcal{F}) \equiv \{\psi : \mathcal{F} \to \mathbf{R} \text{ s.t. } \psi \text{ is continuous}\}$, where here continuity is understood with respect to the (semi)metric ρ_2.

We next introduce the Assumptions we require for this application.

Assumption S.4 (Moments Inequalities Example)**.**

(i) $\{Y_i, Z_i\}_{i=1}^n$ *is an i.i.d. sample with* $E[Y^2] < \infty$.

(ii) $\mathcal{F}$ *is a uniformly bounded* P*-measurable VC class that is closed under* ρ_2.

(iii) There is an $\epsilon > 0$ *such that* $Var\{Y|Z\} > \epsilon$ *almost surely.*

Assumption S.4(ii) imposes a mild measurability requirement on $\mathcal{F}$ by demanding that it be P-measurable; see pg. 110 in van der Vaart and Wellner (1996) for a definition and discussion. In turn, the condition that $\mathcal{F}$ be a VC class requires that $\mathcal{F}$ not be "too big" so that a functional central limit theorem applies. Such a requirement is usually satisfied, for example, when the class $\mathcal{F}$ is defined by a finite dimensional index; e.g. when $\mathcal{F}$ consists of cell indicators in $\mathbf{R}^{d_z}$. Assumption S.4(ii) additionally imposes that $\mathcal{F}$ be closed under ρ_2 for ease of exposition, though it is possible to adapt our arguments to work with the closure of $\mathcal{F}$ if $\mathcal{F}$ fails to be closed. Finally, Assumption S.4(iii) imposes a lower bound on the variance of Y conditional on Z.

Our first Lemma shows Assumption S.4 implies Assumptions 1 and 2 hold.

Lemma S.4.7. *If Assumptions S.4 hold, then Assumptions 1 and 2 are satisfied with* $\mathbb{D}_\phi = \mathbb{D} = \ell^\infty(\mathcal{F})$, $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$, $\mathbb{E} = \mathbf{R}$, $r_n = \sqrt{n}$, $\mathbb{G}_0$ *centered Gaussian, and*

$$\phi'_{\theta_0}(h) = \sup_{f \in \Psi_{\mathcal{F}}(\theta_0)} h(f) \qquad \Psi_{\mathcal{F}}(\theta_0) = \arg\max_{f \in \mathcal{F}} \theta_0(f).$$

Note that Lemma S.4.7 implies that $\phi'_{\theta_0} : \ell^\infty(\mathcal{F}) \to \mathbf{R}$ is linear if and only if $\Psi_{\mathcal{F}}(\theta_0)$ is a singleton. Therefore, Proposition 2.1 and Corollary 3.1 imply the "standard" bootstrap is consistent for the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ if and only if $\Psi_{\mathcal{F}}(\theta_0)$ is a singleton. Theorem 3.2, however, enables us to construct an estimator for the desired asymptotic distribution that is consistent even if $\Psi_{\mathcal{F}}(\theta_0)$ fails to be a singleton. To this end, we first introduce as an estimator for the asymptotic distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ the process $\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\} \in \ell^\infty(\mathcal{F})$ to be given by

$$\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\}(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i \{Y_i f(Z_i) - \frac{1}{n} \sum_{i=1}^n Y_i f(Z_i)\},$$

where $\{W_i\}_{i=1}^n$ is an i.i.d. sample of standard normal random variables. Here, we employ a multiplier bootstrap for the sake of variety in exposition, though we note the nonparametric bootstrap is also consistent. As an estimator for ϕ'_{θ_0}, we then let

$$\hat{\Psi}_{\mathcal{F}}(\theta_0) \equiv \{f \in \mathcal{F} : \sup_{\tilde{f} \in \mathcal{F}} \hat{\theta}_n(\tilde{f}) \leq \hat{\theta}_n(f) + \frac{\kappa_n}{\sqrt{n}}\},$$

and following the analytical expression for ϕ'_{θ_0} in Lemma S.4.7 we set $\hat{\phi}'_n$ to be given by

$$\hat{\phi}'_n(h) \equiv \sup_{f \in \hat{\Psi}_{\mathcal{F}}(\theta_0)} h(f).$$

The estimator $\hat{\phi}'_n$ may be viewed as a generalized moment selection procedure as in Andrews and Soares (2010) and Andrews and Shi (2013).

Our second Lemma verifies Assumption 3 and 4.

Lemma S.4.8. *If Assumption S.4 holds and $\{W_i\}_{i=1}^n$ is an i.i.d. sequence of standard normal random variables independent of $\{X_i\}_{i=1}^n$, then Assumption 3 is satisfied. Moreover, if in addition $\kappa_n \uparrow \infty$ and $\kappa_n/\sqrt{n} \to 0$, then $\hat{\phi}'_n$ satisfies Assumption 4.*

Lemma S.4.6 verifies the conditions of Theorem 3.2, which justifies estimating the asymptotic distribution of $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ by employing the conditional law of $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ given the data. Moreover, since ϕ'_{θ_0} is convex, Corollary 3.2 implies the resulting test is able to locally control size at any θ_0 such that $\phi(\theta_0) = 0$ – here the requirement that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ be continuous and increasing is satisfied at any $\alpha < 0.5$ since $\mathbb{G}_0$ is non-degenerate and $\Psi_{\mathcal{F}}(\theta_0)$ is not empty.

Below we include the proofs for Lemmas S.4.7 and S.4.8.

Proof of Lemma S.4.7: Define the class $\mathcal{G} \equiv \{g : g(y, z) = yf(z)$ for some $f \in \mathcal{F}\}$. Then note that since $\mathcal{F}$ is uniformly bounded, $E[Y^2] < \infty$, and $\mathcal{F}$ is a VC class by Assumptions S.4(i)-(ii), Theorem 2.10.20 in van der Vaart and Wellner (1996) implies $\mathcal{G}$ is a Donsker class and therefore Assumption 2(i) holds with $\mathbb{G}_0$ tight. Moreover, by

Example 1.5.10 in van der Vaart and Wellner (1996), there exists a version of $\mathbb{G}_0$ that has continuous paths with respect to the intrinsic (semi)metric ρ_2 defined in (S.175), and hence Assumption 2(ii) holds with $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$ as well.

In order to verify Assumption 1, we note Example 1.5.10 in van der Vaart and Wellner (1996) additionally implies $\mathcal{F}$ is totally bounded under ρ_2. Since $\mathcal{F}$ is also closed under ρ_2 by Assumption S.4(ii), Corollary 3.29 in Aliprantis and Border (2006) implies $\mathcal{F}$ is compact under ρ_2. Next fix any sequence $\{f_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ satisfying $\rho_2(f_n, f) \to 0$ for some $f \in \mathcal{F}$. By Assumption S.4(iii) there then exists an $\epsilon > 0$ such that

$$E[(f_n(Z) - f(Z))^2] \leq \frac{1}{\epsilon} E[\mathrm{Var}\{Y|Z\}(f_n(Z) - f(Z))^2] \leq \frac{1}{\epsilon}\rho_2^2(f_n, f). \qquad \text{(S.176)}$$

Hence, we obtain from the Cauchy-Schwarz inequality and result (S.176) that

$$\lim_{n\to\infty} |E[Y(f_n(Z) - f(Z))]| \leq \lim_{n\to\infty} \{E[Y^2]\}^{1/2}\{E[(f_n(Z) - f(Z))^2]\}^{1/2} = 0, \qquad \text{(S.177)}$$

which implies $\theta_0 \in \mathcal{C}(\mathcal{F})$. Applying Lemma S.4.9 we may then conclude that Assumption 1 is satisfied, which concludes the proof of the Lemma. ∎

PROOF OF LEMMA S.4.8: For notational simplicity, set $\mathbb{G}_n^* \equiv \sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$. Then note that Assumption 3(i) is satisfied by construction, while Assumption 3(iv) holds since $f(\mathbb{G}_n^*)$ is continuous in $\{W_i\}_{i=1}^n$. Further define $\mathbb{G}_n' \in \ell^\infty(\mathcal{F})$ to be given by

$$\mathbb{G}_n'(f) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\{Y_i f(Z_i) - E[Y_i f(Z_i)]\}, \qquad \text{(S.178)}$$

and note that since $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ satisfies Assumption 2 by Lemma S.4.7, it follows that $\|\hat{\theta}_n - \theta_0\|_\infty = o_p(1)$. Therefore, by direct manipulation we can conclude that

$$\sup_{h\in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))} |E[h(\mathbb{G}_n')|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n]|$$
$$\leq E[|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i|] \times \|\hat{\theta}_n - \theta_0\|_\infty = o_p(1), \qquad \text{(S.179)}$$

where we exploited that $\{W_i\}_{i=1}^n$ is independent of $\{X_i\}_{i=1}^n$ and that $\sum_{i=1}^n W_i/\sqrt{n}$ follows a standard normal distribution. Hence, Lemma S.4.7 implying Assumption 2 holds together with Theorem 2.9.6 in van der Vaart and Wellner (1996) yields

$$\sup_{h\in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))} |E[h(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}_0)]|$$
$$\leq \sup_{h\in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))} |E[h(\mathbb{G}_n')|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}_0)]| + o_p(1) = o_p(1), \qquad \text{(S.180)}$$

which verifies Assumption 3(ii). Next, for any $h \in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))$ let $h(\mathbb{G}_n')^*$ and $h(\mathbb{G}_n^*)^*$

denote the minimal measurable majorants with respect to $\{X_i, W_i\}_{i=1}^n$ of $h(\mathbb{G}'_n)$ and $h(\mathbb{G}^*_n)$ respectively. Lemma 1.2.2(iii) in van der Vaart and Wellner (1996) then yields

$$|E[h(\mathbb{G}'_n)^*|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}^*_n)^*|\{X_i\}]| \\ \leq E[|h(\mathbb{G}'_n) - h(\mathbb{G}^*_n)|^*|\{X_i\}_{i=1}^n] \leq E[|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i|] \times \|\hat{\theta}_n - \theta_0\|^*_\infty = o_p(1) \quad \text{(S.181)}$$

because $\|\hat{\theta}_n - \theta_0\|_\infty = o_p(1)$. Similarly, let $h(\mathbb{G}'_n)_*$ and $h(\mathbb{G}^*_n)_*$ denote the maximal measurable minorants of $h(\mathbb{G}'_n)$ and $h(\mathbb{G}^*_n)$. Since for any $T : \{X_i, W_i\}_{i=1}^n \to \ell^\infty(\mathcal{F})$ we have $T_* = -(-T^*)$, Lemma 1.2.2(iii) in van der Vaart and Wellner (1996) yields

$$|E[h(\mathbb{G}'_n)_*|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}^*_n)_*|\{X_i\}]| = |E[-(-h(\mathbb{G}'_n))^* + (-h(\mathbb{G}^*_n))^*|\{X_i\}_{i=1}^n]| \\ \leq E[|h(\mathbb{G}'_n) - h(\mathbb{G}^*_n)|^*|\{X_i\}_{i=1}] = o_p(1), \quad \text{(S.182)}$$

where the final equality follows from (S.181). Hence, for any $h \in \mathrm{BL}_1(\ell^\infty(\mathcal{F}))$ we obtain

$$E[h(\mathbb{G}^*_n)^*|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}^*_n)_*|\{X_i\}_{i=1}^n] \\ = E[h(\mathbb{G}'_n)^*|\{X_i\}_{i=1}^n] - E[h(\mathbb{G}'_n)_*|\{X_i\}_{i=1}^n] + o_p(1) = o_p(1), \quad \text{(S.183)}$$

where the first equality follows from (S.181) and (S.182), while the second is implied by Theorem 2.9.7 in van der Vaart and Wellner (1996). Since Assumptions 3(i)-(ii), and 3(iv) have been shown to hold, Lemma S.3.9 implies Assumption 3(iii) holds as well.

In order to verify Assumption 4, we first observe that since Assumption 1 holds by Lemma S.4.7 it follows $\sqrt{n}\|\hat{\theta}_n - \theta_0\|_\infty = O_p(1)$. Therefore, $\kappa_n \uparrow \infty$ yields

$$\lim_{n\to\infty} P(\Psi_{\mathcal{F}}(\theta_0) \subseteq \hat{\Psi}_{\mathcal{F}}(\theta_0)) \geq \lim_{n\to\infty} P(2\sqrt{n}\|\hat{\theta}_n - \theta_0\|_\infty \leq \kappa_n) = 1. \quad \text{(S.184)}$$

Next, fix an arbitrary $\delta > 0$ and let $(\Psi_{\mathcal{F}}(\theta_0))^\delta$ denote a δ enlargement of $\Psi_{\mathcal{F}}(\theta_0)$ under ρ_2. Since $\mathcal{F}$ is compact and θ_0 is continuous under ρ_2 (see (S.177)), we obtain that

$$\eta \equiv \{\sup_{f\in\mathcal{F}} \theta_0(f) - \sup_{f\in\mathcal{F}\setminus(\Psi_{\mathcal{F}}(\theta_0))^\delta} \theta_0(f)\} > 0. \quad \text{(S.185)}$$

Therefore, the definition of $\hat{\Psi}_{\mathcal{F}}(\theta_0)$ together with result (S.185) allows us to conclude

$$P(\hat{\Psi}_{\mathcal{F}}(\theta_0) \nsubseteq (\Psi_{\mathcal{F}}(\theta_0))^\delta) = P(\sup_{\tilde{f}\in\mathcal{F}} \hat{\theta}_n(\tilde{f}) \leq \hat{\theta}_n(f) + \frac{\kappa_n}{\sqrt{n}} \text{ for some } f \in \mathcal{F} \setminus (\Psi_{\mathcal{F}}(\theta_0))^\delta) \\ \leq P(\eta \leq 2\|\hat{\theta}_n - \theta_0\|_\infty + \frac{\kappa_n}{\sqrt{n}}) = o(1), \quad \text{(S.186)}$$

where in the final result we employed that $\eta > 0$, $\kappa_n/\sqrt{n} \to 0$, and $\|\hat{\theta}_n - \theta_0\|_\infty = o_p(1)$ since Assumption 2 holds by Lemma S.4.7. Next, fix an arbitrary $h \in \mathcal{C}(\mathcal{F})$ and note that by results (S.184) and (S.186), there exists a sequence $\delta_n \downarrow 0$ such that

$$|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)| \leq \sup_{f \in (\Psi_{\mathcal{F}}(\theta_0))^{\delta_n} \cap \mathcal{F}} h(f) - \sup_{f \in \Psi_{\mathcal{F}}(\theta_0)} h(f) + o_p(1)$$
$$\leq \sup_{f_1, f_2 \in \mathcal{F}: \rho_2(f_1, f_2) \leq \delta_n} \{h(f_1) - h(f_2)\} + o_p(1) = o_p(1), \quad \text{(S.187)}$$

where in the final equality we employed that $h \in \mathcal{C}(\mathcal{F})$ and $\mathcal{F}$ being compact imply h must in fact be uniformly continuous. Finally, we note that since $\hat{\phi}_n'$ is Lipschitz continuous with Lipschitz constant equal to one, Lemma S.3.6 and result (S.187) establish that Assumption 4 holds, and the Lemma therefore follows. ■

Lemma S.4.9. *Let* $\mathbf{A}$ *be compact under a metric* $d_{\mathbf{A}}$*,* $\phi : \ell^\infty(\mathbf{A}) \to \mathbf{R}$ *be given by* $\phi(\theta) = \sup_{a \in \mathbf{A}} \theta(a)$*, and set* $\Psi_{\mathbf{A}}(\theta) \equiv \arg\max_{a \in \mathbf{A}} \theta(a)$ *for any* $\theta \in \mathcal{C}(\mathbf{A})$*. Then,* ϕ *is Hadamard directionally differentiable tangentially to* $\mathcal{C}(\mathbf{A})$ *at any* $\theta \in \mathcal{C}(\mathbf{A})$*, and*

$$\phi_\theta'(h) = \sup_{a \in \Psi_{\mathbf{A}}(\theta)} h(a) \qquad h \in \mathcal{C}(\mathbf{A}).$$

PROOF: Let $\{t_n\}_{n=1}^\infty$ and $\{h_n\}_{n=1}^\infty$ be sequence with $t_n \in \mathbf{R}$, $h_n \in \ell^\infty(\mathbf{A})$ for all n, $t_n \downarrow 0$ and $\|h_n - h\|_\infty = o(1)$ for some $h \in \mathcal{C}(\mathbf{A})$. Then note that for any $\theta \in \mathcal{C}(\mathbf{A})$ we have

$$|\sup_{a \in \mathbf{A}}\{\theta(a) + t_n h_n(a)\} - \sup_{a \in \mathbf{A}}\{\theta(a) + t_n h(a)\}| \leq t_n \|h_n - h\|_\infty = o(t_n). \quad \text{(S.188)}$$

Further note that since $\mathbf{A}$ is compact, $\Psi_{\mathbf{A}}(\theta)$ is well defined for any $\theta \in \mathcal{C}(\mathbf{A})$. Setting $\Gamma_\theta : \mathcal{C}(\mathbf{A}) \to \mathcal{C}(\mathbf{A})$ to be given by $\Gamma_\theta(g) = \theta + g$, then note that Γ_θ is trivially continuous. Therefore, Theorem 17.31 in Aliprantis and Border (2006) and the relation

$$\Psi_{\mathbf{A}}(\theta + g) = \arg\max_{a \in \mathbf{A}} \Gamma_\theta(g)(a) \quad \text{(S.189)}$$

imply that $\Psi_{\mathbf{A}}(\theta + g)$ is upper hemicontinuous in g. In particular, for $\Psi_{\mathbf{A}}(\theta)^\epsilon \equiv \{a \in \mathbf{A} : \inf_{\tilde{a} \in \Psi_{\mathbf{A}}(\theta)} d_{\mathbf{A}}(a, \tilde{a}) \leq \epsilon\}$, it follows from $\|t_n h\|_\infty = o(1)$ that $\Psi_{\mathbf{A}}(\theta + t_n h) \subseteq \Psi_{\mathbf{A}}(\theta)^{\delta_n}$ for some sequence $\delta_n \downarrow 0$. Thus, since $\Psi_{\mathbf{A}}(\theta) \subseteq \Psi_{\mathbf{A}}(\theta)^{\delta_n}$ we can conclude that

$$|\sup_{a \in \mathbf{A}}\{\theta(a) + t_n h(a)\} - \sup_{a \in \Psi_{\mathbf{A}}(\theta)}\{\theta(a) + t_n h(a)\}|$$
$$= \sup_{a \in \Psi_{\mathbf{A}}(\theta)^{\delta_n}}\{\theta(a) + t_n h(a)\} - \sup_{a \in \Psi_{\mathbf{A}}(\theta)}\{\theta(a) + t_n h(a)\}$$
$$\leq \sup_{a_0, a_1 \in \mathbf{A}: d_{\mathbf{A}}(a_0, a_1) \leq \delta_n} t_n |h(a_0) - h(a_1)|$$
$$= o(t_n), \quad \text{(S.190)}$$

where the final result follows from h being uniformly continuous by compactness of $\mathbf{A}$.

Therefore, exploiting (S.188), (S.190) and θ being constant on $\Psi_{\mathbf{A}}(\theta)$ yields

$$
\begin{aligned}
&|\sup_{a\in\mathbf{A}}\{\theta(a)+t_n h_n(a)\}-\sup_{a\in\mathbf{A}}\theta(a)-t_n\sup_{a\in\Psi_{\mathbf{A}}(\theta)}h(a)| \\
&\leq |\sup_{a\in\Psi_{\mathbf{A}}(\theta)}\{\theta(a)+t_n h(a)\}-\sup_{a\in\Psi_{\mathbf{A}}(\theta)}\theta(a)-t_n\sup_{a\in\Psi_{\mathbf{A}}(\theta)}h(a)|+o(t_n)=o(t_n), \quad \text{(S.191)}
\end{aligned}
$$

which verifies the claim of the Lemma. ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Andrews, D. W. K. and Shi, X. (2013). Inference based on conditional moment inequatlies. *Econometrica*, **81** 609–666.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

Aubin, J.-P. and Frankowska, H. (1990). *Set-Valued Analysis*. Birkhauser, Boston.

Beran, R. (1997). Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics*, **49** 1–24.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.

Bickel, P. J., Ritov, Y. and Rydn, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. *Ann. Statist.*, **26** 1614–1635. URL http://dx.doi.org/10.1214/aos/1024691255.

Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society, Providence.

Bogachev, V. I. (2007). *Measure Theory: Volume II*. Springer-Verlag, Berlin.

Bontemps, C., Magnac, T. and Maurin, E. (2012). Set Identified Linear Models. *Econometrica*, **80** 1129–1155.

Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their applications*, **58** 247–265.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society, Providence.

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, eds.). Birkhäuser.

Dudley, R. M. (1979). Balls in rk do not cut all subsets of k+ 2 points. *Advances in Mathematics*, **31** 306–308.

Dugundji, J. (1951). An extension of tietze's theorem. *Pacific Journal of Mathematics*, **1** 353–367.

DÜMBGEN, L. (1993). On Nondifferentiable Functions and the Bootstrap. *Probability Theory and Related Fields*, **95** 125–140.

EFRON, B. (1979). Bootstrap methods: Another look at the jacknife. *Annals of Statistics*, **7** 1–26.

GAREL, B. and HALLIN, M. (1995). Local asymptotic normality of multivariate arma processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, **47** 551–579.

HENDERSON, H. V. and SEARLE, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review*, **23** 53–60.

KOSOROK, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer-Verlag, Berlin.

LEHMANN, E. and ROMANO, J. (2005). *Testing Statistical Hypotheses.* Springer Verlag.

LINTON, O., SONG, E., K. and WHANG, Y.-J. (2010). An Improved Bootstrap Test of Stochastic Dominance. *Journal of Econometrics*, **154** 186 – 202.

POLITIS, D., ROMANO, J. P. and WOLF, M. (1999). Weak convergence of dependent empirical measures with application to subsampling in function spaces. *Journal of statistical planning and inference*, **79** 179–190.

RADULOVIĆ, D. (1996). The bootstrap for empirical processes based on stationary observations. *Stochastic processes and their applications*, **65** 259–279.

SHAPIRO, A. (1990). On Concepts of Directional Differentiability. *Journal of Optimization Theory and Applications*, **66** 477–487.

SHAPIRO, A. (1994). Existence and differentiability of metric projections in hilbert spaces. *Siam Journal of Optimization*, **4** 130–141.

STRASSER, H. (1985). *Mathematical theory of statistics: statistical experiments and asymptotic decision theory*, vol. 7. Walter de Gruyter.

VAN DER VAART, A. (1991). On Differentiable Functionals. *The Annals of Statistics*, **19** pp. 178–204.

VAN DER VAART, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag.

ZARANTONELLO, E. H. (1971). Projections on convex sets and hilbert spaces and spectral theory. In *Contributions to Nonlinear Functional Analysis* (E. H. Zaranotello, ed.). Academic Press.

ZEIDLER, E. (1984). *Nonlinear Functional Analysis and its Applications III.* Springer-Verlag, New York.