Online Supplementary Material for "Posterior Distribution of Nondifferentiable Functions"

Toru Kitagawa*, José Luis Montiel Olea†,
Jonathan Payne‡ and Amilcar Velez§

B Additional Results

B.1 Closeness of Bootstrap and Posterior Quantiles

This section establishes the closeness between bootstrap and posterior quantiles that we assumed in Theorem 2. We verify the high-level assumption of Theorem 2 assuming that the distribution $\sqrt{(}g(\theta^{*P}) - g(\widehat{\theta}_n))$ has a uniformly bounded p.d.f. As mentioned in Remark 2, this condition does not imply (nor is implied) by directional differentiability. We start with an intermediate lemma and corollary, then verify the required quantile closeness in Theorem 3.

Lemma 2. *Let W_n^*, Y_n^* be random variables dependent on the data $X^n = (X_1, X_2, \dots, X_n)$ inducing the probability measures P_W^n and P_Y^n respectively. Let $A \subset \mathbb{R}^k$ and let $A^\delta = \{y \in \mathbb{R}^k : \|x - y\| < \delta$ for some $x \in A\}$. Then,*

$$|P_W^n(A|X^n) - P_Y^n(A|X^n)| \leq \frac{1}{\delta} \sup_{f \in BL(1)} |\mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n]| + \max\{P_Y^n(A^\delta \backslash A|X^n),\ P_Y^n((A^c)^\delta \backslash A^c|X^n)\}$$

Proof. To show this lemma we use an argument analogous to that in Dudley (2002) p. 395. Define $f(x) \equiv \max(0, 1 - \|x - A\|/\delta)$. Then, $\delta f \in \mathrm{BL}(1)$ and:

*University College London, Department of Economics. E-mail: t.kitagawa@ucl.ac.uk
†Columbia University, Department of Economics. E-mail: montiel.olea@gmail.com
‡New York University, Department of Economics. E-mail:jep459@nyu.edu
§Northwestern University, Department of Economics. E-mail: amilcare@u.northwestern.edu

$$
\begin{aligned}
P^n_W(A|X^n) &= \int_A dP^n_W|X^n \\
&\leq \int f dP^n_W|X^n \\
&\quad (\text{ since } f \text{ is nonnegative and } f(x)=1 \text{ over } A \text{ }) \\
&= \int_{A^\delta} f dP^n_Y|X^n + \frac{1}{\delta}\left(\int_{A^\delta} \delta f dP^n_W|X^n - \int_{A^\delta} \delta f dP^n_Y|X^n\right) \\
&\leq \int_{A^\delta} dP^n_Y|X^n + \frac{1}{\delta}\sup_{f\in BL(1)}\left|\mathbb{E}[f(W^*_n)\mid X^n] - \mathbb{E}[f(Y^*_n)\mid X^n]\right| \\
&= P^n_Y(A^\delta|X^n) + \frac{1}{\delta}\sup_{f\in BL(1)}\left|\mathbb{E}[f(W^*_n)\mid X^n] - \mathbb{E}[f(Y^*_n)\mid X^n]\right|
\end{aligned}
$$

It follows that:

$$
P^n_W(A|X^n)-P^n_Y(A|X^n) \leq \frac{1}{\delta}|\mathbb{E}[f(W^*_n)|X^n]-\mathbb{E}[f(Y^*_n)|X^n]|+(P^n_Y(A^\delta|X^n)-P^n_Y(A|X^n))
$$

An analogous argument can be made for A^c. In this case we get:

$$
P^n_W(A^c|X^n)-P^n_Y(A^c|X^n) \leq \frac{1}{\delta}|\mathbb{E}[f(W^*_n)|X^n]-\mathbb{E}[f(Y^*_n)|X^n]|+(P^n_Y((A^c)^\delta|X^n)-P^n_Y(A^c|X^n)),
$$

which implies that:

$$
P^n_W(A|X^n)-P^n_Y(A|X^n) \geq -\frac{1}{\delta}|\mathbb{E}[f(W^*_n)|X^n]-\mathbb{E}[f(Y^*_n)|X^n]|-(P^n_Y((A^c)^\delta|X^n)-P^n_Y(A^c|X^n))
$$

The desired result follows. □

Corollary 1. *Suppose we have the same assumptions as Lemma 2. Then, for any $c \in \mathbb{R}$, we have*

$$
|P^n_W((-\infty,c)|X^n) - P^n_Y((-\infty,c)|X^n)| \leq \frac{1}{\zeta}\beta(W^*_n, Y^*_n; X^n) + P^n_Y([c-\zeta, c+\zeta]|X^n)
$$

Proof. Let us apply Lemma 2 to the set $A = (-\infty, c)$, it follows

$$
|P^n_W((-\infty,c)|X^n) - P^n_Y((-\infty,c)|X^n)|
$$

$$
\leq \frac{1}{\zeta}\beta(W^*_n, Y^*_n; X^n) + \max\{(P^n_Y(A^\zeta \setminus A|X^n), P^n_Y((A^c)^\zeta \setminus A^c|X^n)\}
$$

$$= \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \max\{P_Y^n(\, [c, c+\zeta)\, |X^n), P_Y^n(\, (c-\zeta, c)\, |X^n)\}$$

$$\leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + P_Y^n([c-\zeta, c+\zeta]|X^n),$$

which concludes the proof. □

Theorem 3. *Suppose that, for each n, the probability density function of $\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))$ is uniformly bounded. Suppose that Assumptions 1, 2 and 3 hold. Then for any $0 < \epsilon < \alpha$:*

$$\mathbb{P}_\eta^n[q_{\alpha-\epsilon}^P(X^n) \leq q_\alpha^B(X^n) \leq q_{\alpha+\epsilon}^P(X^n)] \to 1 \quad \text{as } n \to \infty.$$

That is, the α-quantile of the bootstrap is in between the $\alpha - \epsilon$ and $\alpha + \epsilon$ quantiles of the posterior of $g(\theta)$ with high probability.

Proof. Define, for any $0 < \beta < 1$, the critical values $c_\beta^{B*}(X^n)$ and $c_\beta^{P*}(X^n)$ as:

$$\begin{aligned} c_\beta^{B*}(X^n) &\equiv \inf_c\{c \in \mathbb{R} \mid \mathbb{P}^{B*}(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \beta\}, \\ c_\beta^{P*}(X^n) &\equiv \inf_c\{c \in \mathbb{R} \mid \mathbb{P}^{P*}(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \beta\}. \end{aligned}$$

Note that the critical values $c_\beta^{B*}(X^n)$, $c_\beta^{P*}(X^n)$ and the quantiles for $g(\theta_n^{B*})$ and $g(\theta_n^{P*})$ are related through the equation:

$$q_\beta^B(X^n) = g(\widehat{\theta}_n) + c_\beta^{B*}(X^n)/\sqrt{n},$$

$$q_\beta^P(X^n) = g(\widehat{\theta}_n) + c_\beta^{P*}(X^n)/\sqrt{n}.$$

This implies that:

$$\mathbb{P}_\eta^n[q_{\alpha-\epsilon}^P(X^n) \leq q_\alpha^B(X^n) \leq q_{\alpha+\epsilon}^P(X^n)] = \mathbb{P}_\eta^n[c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n)]$$

We start by deriving a convenient bound for the difference between the conditional distributions of $\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))$ and the distribution of $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$. Define the random variables:

$$W_n^* \equiv \sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)), \quad Y_n^* \equiv \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)).$$

Applying Corollary 1 to W_n^* and Y_n^*, we obtain:

$$\Big|\mathbb{P}^{B*}\left(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n\right) - \mathbb{P}^{P*}\left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c \mid X^n\right)\Big|$$

$$\leq \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \, , \, \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)); X^n)$$

$$+ \sup_{c\in\mathbb{R}} \mathbb{P}^{P*}\left(c - \zeta \leq \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c + \zeta \mid X^n\right)$$

We use this relation between the conditional c.d.f. of $\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))$ and the conditional c.d.f. of $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ to show that the quantiles of these distributions should be close to each other. To simplify the notation, define the functions:

$$A_1(\zeta, X^n) \equiv \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \, , \, \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)); X^n),$$

$$A_2(\zeta, X^n) \equiv \sup_{c\in\mathbb{R}} \mathbb{P}^{P*}\left(c - \zeta \leq \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c + \zeta \mid X^n\right).$$

Observe that if the data X^n were such that $A_1(\zeta, X^n) < \epsilon/2$ and $A_2(\zeta, X^n) < \epsilon/2$, then for any $c \in \mathbb{R}$:

$$\Big|\mathbb{P}^{B*}\left(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n\right) - \mathbb{P}^{P*}\left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c \mid X^n\right)\Big|$$

$$\leq A_1(\zeta, X^n) + A_2(\zeta, X^n) < \epsilon. \tag{B.1}$$

This inequality implies that:

$$c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n),$$

whenever X^n is such that $A_1(\zeta, X^n) < \epsilon/2$ and $A_2(\zeta, X^n) < \epsilon/2$. To see this, evaluate equation (B.1) at $c = c_{\alpha+\epsilon}^{P*}(X^n)$. This implies that:

$$\begin{aligned}-\epsilon &< \mathbb{P}^{B*}\left(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}^{P*}(X^n) \mid X^n\right) - \mathbb{P}^{P*}\left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}^{P*}(X^n) \mid X^n\right) \\ &\leq \mathbb{P}^{B*}\left(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}^{P*}(X^n) \mid X^n\right) - (\alpha + \epsilon).\end{aligned}$$

Consequently:

$$c_{\alpha+\epsilon}^{P*}(X^n) \in \{c \in \mathbb{R} \mid \mathbb{P}^{B*}(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}.$$

Since:

$$c_\alpha^{B*}(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{B*}(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\},$$

it follows that

$$c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n).$$

To obtain the other inequality, evaluate equation (B.1) at $c = c_\alpha^{B*}(X^n)$. This implies that:

$$\begin{aligned} -\epsilon < \mathbb{P}^{P*}\left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c_\alpha^{B*}(X^n) \mid X^n\right) &- \mathbb{P}^{B*}(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c_\alpha^{B*}(X^n) \mid X^n) \\ \leq \mathbb{P}^{P*}\left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c_\alpha^{B*}(X^n) \mid X^n\right) &- \alpha, \end{aligned}$$

and so, by analogous reasoning, we get:

$$c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n).$$

Now we can finish the proof. Since the probability distribution function of $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ is uniformly bounded, there exists $K > 0$ such that:

$$\mathbb{P}^{P*}(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n) \in [a, b] \mid X^n) \leq K \cdot |a - b|, \ \forall a, b \in \mathbb{R}.$$

This implies that

$$A_2(\zeta^*, X^n) < 2\zeta^* \cdot K.$$

Given $\epsilon > 0$, we can choose $\zeta^* = \epsilon/(4K)$. Therefore,

$$\mathbb{P}(A_2(\zeta^*, X^n) < \epsilon/2) = 1.$$

Since assumptions 1, 2 and 3 hold, by Theorem 1, we have that there exists $\mathcal{N}(\zeta^*, \epsilon/2, \delta)$ such that for $n > N(\zeta^*, \epsilon/2, \delta)$:

$$\mathbb{P}_\theta^n(A_1(\zeta^*, X^n) > \epsilon/2\ ) < \delta.$$

It follows that for $n > \mathcal{N}(\epsilon/2, \delta)\} \equiv \mathcal{N}(\epsilon, \delta)$

$$
\begin{aligned}
\mathbb{P}_\eta^n(c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n))& \\
&\geq \mathbb{P}_\eta^n(A_1(\zeta^*, X^n) < \epsilon/2 \text{ and } A_2(\zeta^*, X^n) < \epsilon/2) \\
&= 1 - \mathbb{P}_\eta^n(A_1(\zeta^*, X^n) > \epsilon/2 \text{ or } A_2(\zeta^*, X^n) > \epsilon/2) \\
&\geq 1 - \mathbb{P}_\eta^n(A_1(\zeta^*, X^n) > \epsilon/2) - \mathbb{P}_\eta^n(A_2(\zeta^*, X^n) > \epsilon/2) \\
&\geq 1 - \delta,
\end{aligned}
$$

which concludes the proof. □

B.2 Posterior Distribution of $g(\theta^{P*})$ under directional differentiability

This section establishes the closeness between bootstrap and posterior quantiles that we assumed in Theorem 2 for the special case of directionally differentiable functions. For the sake of completeness, we provide a slightly more general result based on high-level assumptions that we then verify for the special case of directionally differentiable $g(\cdot)$. Our method of proof requires us to impose additional regularity conditions (besides directional differentiability) to establish closeness in quantiles. This means that directional differentiability is not a sufficient condition for establishing the required closeness in quantiles.

Assumption 4. There exists a function $h_{\theta_0}(Z, X^n)$ such that:

i) $\beta(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)), h_{\theta_0}(Z, X^n);\ X^n) \xrightarrow{p} 0$.

ii) The cumulative distribution function of $Y \equiv h_{\theta_0}(Z, X^n)$ conditional on X^n, denoted $F_{\theta_0}(y|X^n)$, is Lipschitz continuous in y—almost surely in X^n for every n—with a constant k that does not depend on X^n.

The first part of Assumption 4 simply requires the distribution of $\sqrt{n}(g(\theta_n^{B^*}) - g(\hat{\theta}_n))$, conditional on the data, to have a well-defined limit (which is neither assumed nor guaranteed by Theorem 1).

We now establish a Lemma based on a high-level assumption implied by the second part of Assumption 4. In what follows we use $\mathbb{P}^Z$ to denote the distribution of the random variable Z (which is independent of the data X^n for every n).

Assumption 5. The function $h_\theta(Z, X^n)$ is such that for all positive (ϵ, δ) there exists $\zeta(\epsilon, \delta) > 0$ and $\mathcal{N}(\epsilon, \delta)$ for which

$$\mathbb{P}_\eta^n\Big(\sup_{c \in \mathbb{R}} \mathbb{P}^Z\Big(c - \zeta(\epsilon, \delta) \le h_\theta(Z, X^n) \le c + \zeta(\epsilon, \delta) \Bigm| X^n\Big) > \epsilon\Big) < \delta,$$

provided $n \ge \mathcal{N}(\epsilon, \delta)$.

Assumption 5 is implied by the second part of Assumption 4:

$$\mathbb{P}^Z\Big(c - \zeta(\epsilon, \delta) \le h_\theta(Z, X^n) \le c + \zeta(\epsilon, \delta) \Bigm| X^n\Big),$$

equals:

$$F_\theta(c + \zeta(\epsilon, \delta)|X^n) - F_\theta(c - \zeta(\epsilon, \delta)|X^n) \le 2\zeta(\epsilon, \delta)k.$$

Last inequality holds since, by assumption, $F_\theta(y|X^n)$ is Lipschitz continuous—for almost every X_n for every n—with a constant k that does not depend on X^n. By choosing $\zeta(\epsilon,\delta)$ equal to $\epsilon/4k$, then

$$\mathbb{P}^Z\Big(c-\zeta(\epsilon,\delta)\leq h_\theta(Z,X^n)\leq c+\zeta(\epsilon,\delta)\;\Big|\;X^n\Big)\leq\frac{\epsilon}{2},$$

for every c, implying that Assumption 5 holds.

We now show that any random variable satisfying the weak convergence assumption in the first part of Assumption 4 has a conditional α-quantile that—with high probability—lies in between the conditional $(\alpha-\epsilon)$ and $(\alpha+\epsilon)$-quantiles of the limiting distribution.

Lemma 3. *Let θ^*_n denote a random variable whose distribution, P^*, depends on $X^n=(X_1,\ldots,X_n)$ and let Z be the limiting distribution of $Z_n\equiv\sqrt{n}(\widehat{\theta}_n-\theta)$ as defined in Assumption 2. Suppose that*

$$\beta(\sqrt{n}(g(\theta^*_n)-g(\widehat{\theta}_n)),h_\theta(Z,X^n);X^n)\xrightarrow{p}0.$$

Define $c^*_\alpha(X^n)$ and $c_\alpha(X^n)$ as the critical values such that:

$$c^*_\alpha(X^n)\equiv\inf_c\{c\in\mathbb{R}\mid\mathbb{P}^*(\sqrt{n}(g(\theta^*_n)-g(\widehat{\theta}_n))\leq c\mid X^n)\geq\alpha\}.$$

$$c_\alpha(X^n)\equiv\inf_c\{c\in\mathbb{R}\mid\mathbb{P}^*(h_\theta(Z,X^n)\leq c\mid X^n)\geq\alpha\}.$$

Suppose $h_\theta(Z,X^n)$ satisfies Assumption 5. Then for any $0<\epsilon<\alpha$ and $\delta>0$ there exists $\mathcal{N}(\epsilon,\delta)$ such that for $n>N(\epsilon,\delta)$:

$$\mathbb{P}^n_\eta(c_{\alpha-\epsilon}(X^n)\leq c^*_\alpha(X^n)\leq c_{\alpha+\epsilon}(X^n))\geq1-\delta.$$

Proof. We start by deriving a convenient bound for the difference between the conditional distributions of $\sqrt{n}(g(\theta^*_n)-g(\widehat{\theta}_n))$ and the distribution of $h_\theta(Z,X^n)$. Define the random variables:

$$W^*_n\equiv\sqrt{n}(g(\theta^*_n)-g(\widehat{\theta}_n)),\quad Y^*_n\equiv h_\theta(Z,X^n).$$

Denote by P^n_W and P^n_Y the probabilities that each of these random variables induce over the real line. Let $c \in \mathbb{R}$ be some constant. By applying Lemma 2 in Appendix B.1 to the set $A = (-\infty, c)$ it follows that for any $\zeta > 0$:

$$
\begin{aligned}
&|P_W^n((-\infty, c)|X^n) - P_Y^n((-\infty, c)|X^n)| \\
&\leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \max\{(P_Y^n(A^\zeta \setminus A|X^n), P_Y^n((A^c)^\zeta \setminus A^c|X^n)\} \\
&= \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \max\{P_Y^n(\,[c, c+\zeta)\,|X^n), P_Y^n(\,(c-\zeta, c)\,|X^n)\} \\
&\leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \mathbb{P}^Z\left(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta \mid X^n\right)
\end{aligned}
$$

where for any set A, we define $A^\delta \equiv \{y \in \mathbb{R}^k : \|x - y\| < \delta \text{ for some } x \in A\}$ (see Lemma 2). Therefore:

$$
\begin{aligned}
&|\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c \mid X^n\right)| \\
&\leq \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))\,,\ h_\theta(Z, X^n); X^n) \\
&+ \sup_{c \in \mathbb{R}} \mathbb{P}^Z\left(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta \mid X^n\right)
\end{aligned}
$$

We use this relation between the conditional c.d.f. of $\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))$ and the conditional c.d.f. of $h_\theta(Z, X^n)$ to show that quantiles of these distributions should be close to each other.

To simplify the notation, define the functions:

$$
\begin{aligned}
A_1(\zeta, X^n) &\equiv \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))\,,\ h_\theta(Z, X^n); X^n), \\
A_2(\zeta, X^n) &\equiv \sup_{c \in \mathbb{R}} \mathbb{P}^Z\left(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta \mid X^n\right)
\end{aligned}
$$

Observe that if the data X^n were such that $A_1(\zeta, X^n) < \epsilon/2$ and $A_2(\zeta, X^n) < \epsilon/2$ then for any $c \in \mathbb{R}$:

$$
\begin{aligned}
&|\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c \mid X^n\right)| \\
&\leq A_1(\zeta, X^n) + A_2(\zeta, X^n) \\
&< \epsilon.
\end{aligned}
$$

This would imply that for any $c \in \mathbb{R}$:

$$-\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c \mid X^n\right) < \epsilon. \tag{B.2}$$

We now show that this inequality implies that:

$$c_{\alpha-\epsilon}(X^n) \leq c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(X^n),$$

whenever X^n is such that $A_1(\zeta, X^n) < \epsilon/2$ and $A_2(\zeta, X^n) < \epsilon/2$. To see this, evaluate equation (B.2) at $c_{\alpha+\epsilon}(X^n)$. This implies that:

$$\begin{aligned}-\epsilon &< \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}(X^n) \mid X^n) - \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c_{\alpha+\epsilon}(X^n) \mid X^n\right) \\ &\leq \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}(X^n) \mid X^n) - (\alpha + \epsilon).\end{aligned}$$

Consequently:

$$c_{\alpha+\epsilon}(X^n) \in \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}.$$

Since:

$$c^*_\alpha(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\},$$

it follows that

$$c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(X^n).$$

To obtain the other inequality, evaluate equation (B.2) at $c^*_\alpha(X^n)$. This implies that:

$$\begin{aligned}-\epsilon &< \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c^*_\alpha(X^n) \mid X^n\right) - \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c^*_\alpha(X^n) \mid X^n) \\ &\leq \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c^*_\alpha(X^n) \mid X^n\right) - \alpha,\end{aligned}$$

it follows that

$$c_{\alpha-\epsilon}(X^n) \leq c^*_\alpha(X^n).$$

This shows that whenever the data X^n is such that $A_1(\zeta, X^n) < \epsilon/2$ and $A_2(\zeta, X^n) < \epsilon/2$

$$c_{\alpha-\epsilon}(X^n) \leq c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(X^n).$$

To finish the proof, note that by Assumption 5 there exists $\zeta^* \equiv \zeta(\epsilon/2, \delta/2)$ and

$\mathcal{N}(\epsilon/2, \delta/2)$ that guarantees that if $n > \mathcal{N}(\epsilon/2, \delta/2)$:

$$\mathbb{P}^n_\theta(A_2(\zeta^*, X^n) > \epsilon/2) < \delta/2.$$

Also, by the convergence assumption of this Lemma, there is $\mathcal{N}(\zeta^*, \epsilon/2, \delta/2)$ such that for $n > N(\zeta^*, \epsilon/2\delta/2)$:

$$\mathbb{P}^n_\theta(A_1(\zeta^*, X^n) > \epsilon/2\ ) < \delta/2.$$

It follows that for $n > \max\{\mathcal{N}(\zeta^*, \epsilon/2, \delta/2), \mathcal{N}(\epsilon/2, \delta/2)\} \equiv \mathcal{N}(\epsilon, \delta)$

$$\begin{aligned}
&\mathbb{P}^n_\eta(c_{\alpha-\epsilon}(X^n) \leq c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(X^n)) \\
&\qquad \geq \mathbb{P}^n_\eta(A_1(\zeta^*, X^n) < \epsilon/2 \text{ and } A_2(\zeta^*, X^n) < \epsilon/2) \\
&\qquad = 1 - \mathbb{P}^n_\eta(A_1(\zeta^*, X^n) > \epsilon/2 \text{ or } A_2(\zeta^*, X^n) > \epsilon/2) \\
&\qquad \geq 1 - \mathbb{P}^n_\eta(A_1(\zeta^*, X^n) > \epsilon/2) - \mathbb{P}^n_\eta(A_2(\zeta^*, X^n) > \epsilon/2) \\
&\qquad \geq 1 - \delta, \qquad \square
\end{aligned}$$

We have shown that if $\sqrt{n}(g(\theta^*_n) - g(\widehat{\theta}_n))$ is any random variable satisfying the assumptions of Lemma 3, its conditional α-quantile lies—with high probability—between the conditional $(\alpha - \epsilon)$ and $(\alpha + \epsilon)$ quantiles of the limiting distribution $h_\theta(Z, X^n)$. The next Lemma considers the case in which θ^*_n is either θ^{B*}_n or θ^{P*}_n and characterizes the asymptotic behavior of the c.d.f. of $\sqrt{n}(g(\widehat{\theta}_n) - g(\theta))$ evaluated at bootstrap and posterior quantiles. The main result is that the c.d.f. evaluated at the bootstrap α-quantile is—in large samples—close to same c.d.f. evaluated at the $(\alpha - \epsilon)$ and $(\alpha + \epsilon)$ posterior quantiles. We note that this result could not be obtained directly from the fact that the bootstrap and posterior quantiles converge in probability to each other, as some additional regularity in the limiting distribution is needed. This is why it was important to establish Lemma 3 before the following Lemma.

Lemma 4. *Suppose that Assumptions 1, 2, 3 and 4 hold. Fix $\alpha \in (0,1)$. Let $c^{B*}_\alpha(X^n)$ and $c^{P*}_\alpha(X^n)$ denote critical values satisfying:*

$$c^{B*}_\alpha(X^n) \equiv \inf_c\{c \in \mathbb{R} \mid \mathbb{P}^{B*}(\sqrt{n}(g(\theta^{B*}_n) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\},$$

$$c^{P*}_\alpha(X^n) \equiv \inf_c\{c \in \mathbb{R} \mid \mathbb{P}^{P*}(\sqrt{n}(g(\theta^{P*}_n) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}.$$

Then, for any $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $N(\epsilon, \delta)$ such that for all $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \leq \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \tag{B.3}$$

$$\mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \geq \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta. \tag{B.4}$$

Proof. Let θ^* denote either θ_n^{P*} or θ_n^{B*}. Let $c_\alpha(X^n)$ and $c_\alpha^*(X^n)$ be defined as in Lemma 3. Under Assumptions 1, 2, 3 and 4, the conditions for Lemma 3 are satisfied. It follows that for any $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $\mathcal{N}(\epsilon, \delta)$ such that for all $n > \mathcal{N}(\epsilon, \delta)$:

$$\mathbb{P}_\eta^n(c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \leq \delta/2 \quad \text{and} \quad \mathbb{P}_\eta^n(c_\alpha^*(X^n) < c_{\alpha-\epsilon/2}(X^n)) \leq \delta/2.$$

Therefore:

$$\begin{aligned}
&\mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n)) \\
&\qquad = \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) \geq c_\alpha^*(X^n)) \\
&\qquad + \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\
&\qquad \text{(by the additivity of probability measures)} \\
&\qquad \leq \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\eta^n(c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\
&\qquad \text{(by the monotonicity of probability measures)} \\
&\qquad \leq \mathbb{P}_\eta^n(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \delta/2.
\end{aligned} \tag{B.5}$$

Also, we have that:

$$
\begin{aligned}
&\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n) \text{ and } c^*_\alpha(X_n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n) \text{ and } c^*_\alpha(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad = \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n)) + \mathbb{P}^n_\eta(c^*_\alpha(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad - \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n) \text{ or } c^*_\alpha(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \text{(using } P(A \cap B) = P(A) + P(B) - P(A \cup B)) \\
&\qquad \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n)) - (1 - \mathbb{P}^n_\eta(c^*_\alpha(X^n) \geq c_{\alpha-\epsilon/2}(X^n))) \\
&\qquad \text{(since } \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n) \text{ or } c^*_\alpha(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \leq 1) \\
&\qquad = \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n)) - \mathbb{P}^n_\eta(c^*_\alpha(X^n) < c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^*_\alpha(X^n)) - \delta/2. \qquad \text{(B.6)}
\end{aligned}
$$

Replacing c^*_α by c^{B*}_α in (B.6) and c^*_α by c^{P*}_α and α by $\alpha - \epsilon$ in (B.5) implies that for $n > N(\epsilon, \delta)$:

$$
\begin{aligned}
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_\alpha(X^n)) - \delta/2 \\
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) \leq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{\alpha-\epsilon}(X^n)) + \delta/2.
\end{aligned}
$$

Combining the previous two equations gives that for $n > N(\epsilon, \delta)$:

$$
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_\alpha(X^n)) \leq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{\alpha-\epsilon}(X^n)) + \delta.
$$

This establishes equation (B.3). Replacing θ^*_n by θ^{B*}_n in (B.5) and replacing θ^*_n by θ^{P*}_n, α by $\alpha + \epsilon$ (B.6) implies that for $n > N(\epsilon, \delta)$:

$$
\begin{aligned}
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X_n)) \leq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_\alpha(X^n)) + \delta/2 \\
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X_n)) \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{\alpha+\epsilon}(X^n)) - \delta/2
\end{aligned}
$$

and combining the previous two equations gives that for $n > N(\epsilon, \delta)$:

$$
\mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_\alpha(X^n)) \geq \mathbb{P}^n_\eta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{\alpha+\epsilon}(X^n)) - \delta,
$$

which establishes equation (B.4). □

Lemma 5. *let Z be the limiting distribution of $Z_n \equiv \sqrt{n}(\widehat{\theta}_n - \theta)$ as defined in Assumption 2. Let Z^* be a random variable independent of both $X^n = (X_1, \ldots, X_n)$ and Z and let θ_0 denote the parameter that generated the data. Suppose that g is directionally differentiable in the sense defined in Remark 2 of the main text. Then, Assumption 4 (i) holds with $h_{\theta_0}(Z, Z_n) = g'_{\theta_0}(Z^* + Z_n) - g'_{\theta_0}(Z_n)$.*

Proof. We start by analyzing the limiting distribution of both:

$$\sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\theta_0))$$

and

$$\sqrt{n}(g(\theta_0 + \frac{Z_n}{\sqrt{n}}) - g(\theta_0))$$

as a function of (Z^*, Z_n). Note that the delta method for directionally differentiable functions (e.g., Theorem 2.1 in Fang and Santos (2019)) and the continuity of the directional derivative implies that jointly:

$$\begin{aligned}
\sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\theta_0)) &\overset{d}{\to} g'_{\theta_0}(Z^* + Z) \\
g'_{\theta_0}(Z^* + Z_n) &\overset{d}{\to} g'_{\theta_0}(Z^* + Z) \\
\sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0)) &\overset{d}{\to} g'_{\theta_0}(Z) \\
g'_{\theta_0}(Z_n) &\overset{d}{\to} g'_{\theta_0}(Z)
\end{aligned}$$

where Z is independent of Z^*. Note that the joint (and unconditional) convergence in distribution above implies that:

$$A_n \equiv \sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\widehat{\theta}_n))$$

and

$$B_n \equiv g'_{\theta_0}(Z^* + Z_n) - g'_{\theta_0}(Z_n)$$

are such that $|A_n - B_n| = o_p(1)$, where the $o_p(1)$ term refers to convergence in probability unconditional on the data as a function of Z^* and Z_n.

Note that for any two random variables A_n and B_n we have that for any ϵ

$$\sup_{BL(1)} \Big| \mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n] \Big|$$

is bounded above by:

$$\epsilon + 2\mathbb{P}^{Z^*}[\,|A_n - B_n| > \epsilon\,|X^n],$$

where the probability is taken over the distribution of Z^*, denoted $\mathbb{P}^{Z^*}$.[1] Note that the unconditional convergence in probability result for $|A_n - B_n|$ implies that:

$$\mathbb{E}_\theta[\mathbb{P}^{Z^*}[\,|A_n - B_n| > \epsilon\,|X^n]] \to 0,$$

as the expectation is taken over different data realizations. Note that in light of the inequalities above we have that:

$$\mathbb{P}^n_\eta\left(\sup_{BL(1)}\Big|\mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n]\Big| > 2\epsilon\right) \tag{B.7}$$

is bounded above by:

$$\mathbb{P}^n_\eta\left(\epsilon + 2\mathbb{P}^{Z^*}[\,|A_n - B_n| > \epsilon\,|X^n] > 2\epsilon\right),$$

which equals

$$\mathbb{P}^n_\eta\left(\mathbb{P}^{Z^*}[\,|A_n - B_n| > \epsilon\,|X^n] > \epsilon/2\right).$$

Thus, by Markov's inequality:

$$\mathbb{P}^n_\eta\left(\sup_{BL(1)}\Big|\mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n]\Big| > 2\epsilon\right) \leq 2\mathbb{E}_\theta[\mathbb{P}^{Z^*}[\,|A_n - B_n| > \epsilon\,|X^n]]/\epsilon.$$

Implying that:

$$\sup_{BL(1)}\Big|\mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n]\Big| \xrightarrow{p} 0,$$

as desired.[2] □

B.3 Additional Lemmas

Lemma 6. *Let $S(\theta)$ be an $m \times n$ matrix of sign restrictions whose entries depend on the finite dimensional parameter $\theta \equiv (vec(A)', vech(\Sigma)')'$. Given Σ is invertible,*

[1] This is a common bound used in bootstrap analysis; see for example, Theorem 23.9 p. 333 in Van der Vaart (2000).

[2] We are extremely thankful to an anonymous referee who suggested major simplifications to the previous version of the proof of this Lemma.

consider the program

$$g(\theta) \equiv \max_{x \in \mathbb{R}^n} e_i' C_k(A)x, \;\; s.t. \;\; x'\Sigma^{-1}x = 1, \; S(\theta)x \geq 0, \tag{B.8}$$

where e_i *denotes the i-th column of the identity matrix of dimension* n. *Suppose*

1. $\{x \in \mathbb{R}^n \mid S(\theta)x \geq 0\}$ *is nonempty in a neighborhood of* θ.

2. $S(\theta)$ *is a continuously differentiable function of* θ,

3. *There exists an optimal solution* $x^*(\theta)$ *to (B.8) for which its corresponding active constraints* $S^*(\theta) \in \mathbb{R}^{m^* \times n}$ $(m^* \leq m)$ *can be written as positive linear combination of a full row-rank matrix* $\tilde{S}^*(\theta) \in \mathbb{R}^{r \times n}$, $r \leq m^*$ *and* $\tilde{S}^*(\theta)x^*(\theta) = \boldsymbol{0}_{r \times 1}$. *That is, there exists* $\alpha \in \mathbb{R}_+^{r \times m^*}$ *s.t.*

$$\tilde{S}^*(\theta)'\alpha = S^*(\theta)',$$

and

$$\tilde{S}^*(\theta)x^*(\theta) = \boldsymbol{0}_{r \times 1}.$$

Then $g(\theta)$ *is locally Lipschitz.*

Proof. Define $D(\theta) \equiv \{x \in \mathbb{R}^n | x'\Sigma^{-1}x = 1, S(\theta)x \geq 0\}$. Assumption 1 of the current lemma implies that $D(\theta)$ is nonempty in a neighborhood of θ. We use Proposition 6 from Morand, Reffett, and Tarafdar (2015) to prove that $g(\theta)$ is a locally Lipschitz function. Thus, we need to verify that (i) $D(\theta)$ is uniformly compact near θ and (ii) the Mangasarian-Fromowitz constraint qualification (MFCQ) holds at some optimal solution $x^*(\theta)$. This second requirement is equivalent to verifying:

1. The gradient of the equality constraints ($\nabla_x h^j(x^*(\theta), \theta)$ for $j = 1, .., q$) are linear independent vectors. In our problem we only have one equality constraint that is defined by $h(x, \theta) \equiv x'\Sigma^{-1}x - 1$. Since $\nabla_x h(x, \theta) = 2\Sigma^{-1}x$ and $x^*(\theta)'\Sigma^{-1}x^*(\theta) = 1$, it follows that $\nabla_x h(x^*(\theta), \theta) \neq 0$ verifies this linear independent condition.

2. There exists $y \in \mathbb{R}^n$ such that, $\nabla_x g^i(x^*(\theta), \theta) \cdot y < 0$ for all $i \in I \equiv \{i | g^i(x^*(\theta), \theta) = 0\}$ and $\nabla_x h^j(x^*(\theta), \theta) \cdot y = 0$ for all $j = 1, \ldots, q$. In our problem we have m-inequality constraints $g^i(x, \theta) \equiv -e_i' S(\theta)x$ for $i = 1, ..., m$ and only one equality constraint $h(x, \theta) = x'\Sigma^{-1}x - 1$. Under the assumption of this lemma, we

have that at $x^*(\theta)$ the set I has m^* elements that are defined by the active constraints (the rows of $S^*(\theta)$). Then, the verification of this condition is equivalent to $-S^*(\theta)y < 0$ and $\Sigma^{-1}x^*(\theta) \cdot y = 0$. We will verify this condition in **step 2**.

Step 1: Define

$$D(\theta,\delta) \equiv \bigcup_{\{\tilde{\theta}:||\tilde{\theta}-\theta||<\delta\}} D(\tilde{\theta}) \subseteq E(\theta,\delta) \equiv \bigcup_{\{\tilde{\theta}:||\tilde{\theta}-\theta||<\delta\}} E(\tilde{\theta})$$

where $E(\tilde{\theta}) \equiv \{x \in \mathbb{R}^n \,|\, x'\tilde{\Sigma}^{-1}x = 1\}$. It is sufficient to show that for δ small enough, there exists $M_\theta(\delta) > 0$ such that $E(\tilde{\theta}) \subseteq B_0(M_\theta(\delta))$ for all $\tilde{\theta}$ such that $||\tilde{\theta} - \theta|| < \delta$; where $B_0(M_\theta(\delta))$ is an open ball centered at 0 with radius $M(\delta)$. This is sufficient since

$$\text{Closure}(D(\theta,\delta)) \subseteq \text{Closure}(E(\theta,\delta)) \subseteq \text{Closure}(B_0(M_\theta(\delta))) = \{x \mid ||x'x|| \leq M_\theta(\delta)\},$$

implies that the closure of $D(\theta,\delta)$ is a subset of a compact subset, which implies the uniform compactness of $D(\theta)$.

For each $\tilde{\theta} = (\text{vec}(\tilde{A})', \text{vech}(\tilde{\Sigma})')'$ consider the optimization problem

$$v(\tilde{\Sigma}) \equiv \max_{x \in \mathbb{R}^n} x'x, \text{ s.t. } x'\tilde{\Sigma}^{-1}x = 1.$$

The necessary first-order conditions for this problem are

$$(\mathbb{I}_n - \lambda\tilde{\Sigma}^{-1})x = 0,$$

where λ is a scalar lagrange multiplier. The first-order conditions are thus satisfied by pairs (λ^*, x^*) where λ^* is the eigenvalue of $\tilde{\Sigma}$ and x^* is its corresponding eigenvector. By the definition of the eigenvector

$$\tilde{\Sigma}^{-1}x^* = (1/\lambda^*)x^*,$$

Thus,

$$x^{*\prime}x^* = \lambda^*.$$

This means that value of the program above is given by

$$v(\tilde{\Sigma}) = \text{maxeig}(\tilde{\Sigma}).$$

Consequently,

$$x \in E(\widehat{\theta}) \implies ||x'x|| \leq \left(\text{maxeig}(\tilde{\Sigma})\right)^{1/2}.$$

Since Σ is invertible, there exists δ small enough and a constant c such that

$$1/\text{maxeig}(\tilde{\Sigma}) = \text{mineig}(\tilde{\Sigma}^{-1}) > c, \text{ for all } ||\tilde{\theta} - \theta|| \leq \delta.$$

Then, $E(\tilde{\theta}) \subset B_0(c^{-1/2})$ for all $\tilde{\theta}$ such that $||\tilde{\theta} - \theta|| < \delta$.

Step 2: We now show that the MFCQ holds at a solution $x^*(\theta)$ that satisfies our assumptions. Let $S^*(\theta)$ denote the matrix of active constraints at $x^*(\theta)$, that is

$$S^*(\theta)x^*(\theta) = \mathbf{0}_{m^* \times 1}.$$

We have assumed there exists a full-row rank matrix $\tilde{S}^*(\theta)$ of dimension $r \times n$, $r \leq m^*$, and a matrix α of dimension $r \times m^*$ with nonnegative entries such that

$$\tilde{S}^*(\theta)'\alpha = S^*(\theta)', \quad \tilde{S}^*(\theta)x^*(\theta) = \mathbf{0}_{r\times 1}$$

The full row-rank assumption about $\tilde{S}^*(\theta)$ implies $r \leq n - 1$ (if not $x^*(\theta) = 0$ and this contradicts $x^*(\theta)'\Sigma^{-1}x^*(\theta) = 1$).

We now argue that $\tilde{S}^*(\theta)' \in \mathbb{R}^{n\times r}$ and $\Sigma^{-1}x^*(\theta)$ are linearly independent. Suppose this is not the case. Since $\tilde{S}^*(\theta)'$ has full column rank and $x^*(\theta) \neq 0$ (as $x^*(\theta)'\Sigma^{-1}x^*(\theta) = 1$) then there must exist $\beta \in \mathbb{R}^r$ such that

$$\tilde{S}^*(\theta)'\beta = \Sigma^{-1}x^*(\theta).$$

This implies

$$(x^*)'\tilde{S}^*(\theta)'\beta = x^*(\theta)'\Sigma^{-1}x^*(\theta) = 1,$$

but the left-hand side in the equation is equal to $(\tilde{S}^*(\theta)x^*)'\beta$, which is zero by the definition of $\tilde{S}^*(\theta)$ and so we get the required contradiction.

Linear independence implies that

$$[\Sigma^{-1}x^*, \tilde{S}^*(\theta)'],$$

has column rank $(r+1) \leq n$. This means that for any vector $c \in \mathbb{R}^r$ with strictly positive entries there exists $y(c) \in \mathbb{R}^n$ such that

$$[\Sigma^{-1}x^*, \tilde{S}^*(\theta)']'y(c) = [0, c']'.$$

Consequently,

$$S^*(\theta)y(c) = (\alpha'\tilde{S}^*(c))y(c) = \alpha'c > 0.$$

and

$$(\Sigma^{-1}x^*)'y(c) = 0.$$

Thus, the MFCQ condition is satisfied.

□

C max$\{\theta_1, \theta_2\}$

In this Appendix we illustrate Theorem 2 with an alternative example. Let $(X_1, \ldots X_n)$ be an i.i.d sample of size n from the statistical model:

$$X_i \sim \mathcal{N}_2(\theta, \Sigma), \quad \theta = (\theta_1, \theta_2)' \in \mathbb{R}^2, \; \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

where Σ is assumed known. Consider the family of priors:

$$\theta \sim \mathcal{N}_2(\mu, (1/\lambda^2)\Sigma), \quad \mu = (\mu_1, \mu_2)' \in \mathbb{R}^2$$

indexed by the location parameter μ and the precision parameter $\lambda^2 > 0$. The object of interest is the transformation:

$$g(\theta) = \max\{\theta_1, \theta_2\}.$$

Relation to the main assumptions: The transformation g is differentiable everywhere except at $\theta_1 = \theta_2$. It can be proved with standard arguments that g is locally Lipschitz. This implies that Assumption 1 is satisfied.

Once again, we take $\widehat{\theta}_n$ to be the maximum likelihood estimator given by $\widehat{\theta}_{\rm n} = (1/n)\sum_{i=1}^{n} X_i$ and so $\sqrt{n}(\widehat{\theta}_{\rm n} - \theta) \sim Z \sim \mathcal{N}_2(0, \Sigma)$. Thus, Assumption 2 is satisfied.

The posterior distribution for θ is given by Gelman, Carlin, Stern, and Rubin (2009), p. 89:

$$\theta_n^{P*}|X^n \sim \mathcal{N}_2\Big(\frac{n}{n+\lambda^2}\widehat{\theta}_n + \frac{\lambda^2}{n+\lambda^2}\mu\,,\ \frac{1}{n+\lambda^2}\Sigma\Big).$$

and so by an analogous argument to the absolute value example we have that:

$$\beta(\sqrt{n}(\theta_n^{P*} - \hat{\theta}_n), \mathcal{N}_2(0, \Sigma)); X^n) \xrightarrow{p} 0,$$

which implies that Assumption 3 holds. Since $\theta_n^{P*} = \hat{\theta}_n + Z_n^{P*}/\sqrt{n}$, we have

$$Z_n^{P*}|X^n \sim \mathcal{N}_2\Big(A_n, \Sigma_n\Big), \text{where } A_n = \frac{\lambda^2\sqrt{n}}{n+\lambda^2}(\mu - \hat{\theta}_n),\ \Sigma_n = \frac{\sqrt{n}}{\sqrt{n+\lambda}}\Sigma$$

Define the random variable $Y \equiv \sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$. We have

$$Y = \max\{Z_{n,1}^{P*} + \sqrt{n}\cdot\hat{\theta}_{n,1}, Z_{n,2}^{P*} + \sqrt{n}\cdot\hat{\theta}_{n,2}\} - \max\{\sqrt{n}\cdot\hat{\theta}_{n,1}, \sqrt{n}\cdot\hat{\theta}_{n,2}\}$$

Define the random variable $M_n \equiv \max\{\sqrt{n}\cdot\hat{\theta}_{n,1}, \sqrt{n}\cdot\hat{\theta}_{n,2}\}$. Based on the results of Nadarajah and Kotz (2008), the (conditional) density of Y, denoted $f_{\theta_0}(y|X^n)$, is given by:

$$\begin{aligned} & \frac{1}{\sigma_{1,n}}\phi\left(\frac{C_{n,1}-y}{\sigma_{1,n}}\right)\Phi\left(\frac{1}{\sqrt{1-\rho_n^2}}\left(\frac{\rho_n(C_{n,2}-y)}{\sigma_{1,n}} + \frac{y-C_{n,2}}{\sigma_{2,n}}\right)\right) \\ + \quad & \frac{1}{\sigma_{2,n}}\phi\left(\frac{C_{n,2}-y}{\sigma_{2,n}}\right)\Phi\left(\frac{1}{\sqrt{1-\rho_n^2}}\left(\frac{\rho_n(C_{n,2}-y)}{\sigma_{2,n}} + \frac{y-C_{n,1}}{\sigma_{1,n}}\right)\right), \end{aligned}$$

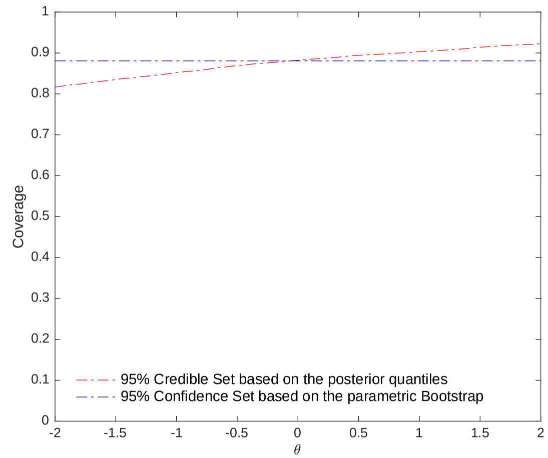
where $C_{n,1} = A_{n,1} + \sqrt{n}\cdot\hat{\theta}_{n,1} - M_n$ and $C_{n,2} = A_{n,2} + \sqrt{n}\cdot\hat{\theta}_{n,2} - M_n$. Also, the parameters $\sigma_{1,n}^2$, $\sigma_{2,n}^2$ and $\sigma_{12,n}$ define the entries of Σ_n. And, $\rho_n = \sigma_{12,n}/\sigma_{1,n}\sigma_{2,n}$ and ϕ, Φ are the p.d.f. and the c.d.f. of a standard normal. It follows that:

$$f_{\theta_0}(y|X^n) \leq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma_{1,n}} + \frac{1}{\sigma_{2,n}}\right) < \frac{2}{\sqrt{2\pi}}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$$

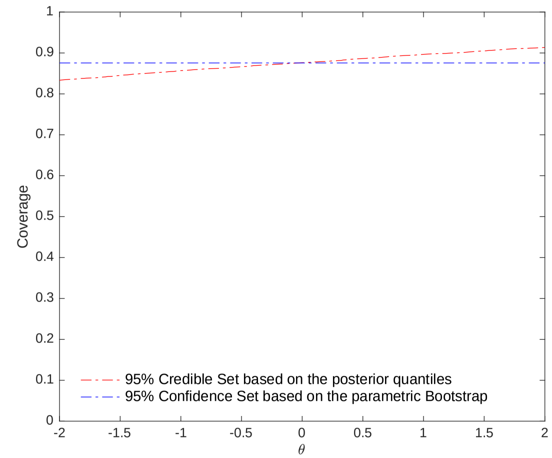
Last inequality follows since Σ_n converge to Σ. This implies that the probability distribution function of $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ is uniformly bounded. Theorem 3 in Section B.1 of the Appendix implies that assumptions of Theorem 2 are verified.

Graphical illustration of coverage failure: Theorem 2 implies that credible sets based on the quantiles of $g(\theta_n^{P*})$ will effectively have the same asymptotic coverage properties as confidence sets based on quantiles of the bootstrap. For the transformation $g(\theta) = \max\{\theta_1, \theta_2\}$, this means that both methods lead to deficient frequentist coverage at the points in the parameter space in which $\theta_1 = \theta_2$. This is illustrated in Figure 2, which depicts the coverage of a nominal 95% bootstrap confidence set and different 95% credible sets. The coverage is evaluated assuming $\theta_1 = \theta_2 = \theta \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$. The sample sizes considered are $n \in \{100, 200, 300, 500\}$. A prior characterized by $\mu = 0$ and $\lambda^2 = 1$ is used to calculate the credible sets. The credible sets and confidence sets have similar coverage as n becomes large and neither achieves 95% probability coverage for all $\theta \in [-2, 2]$.

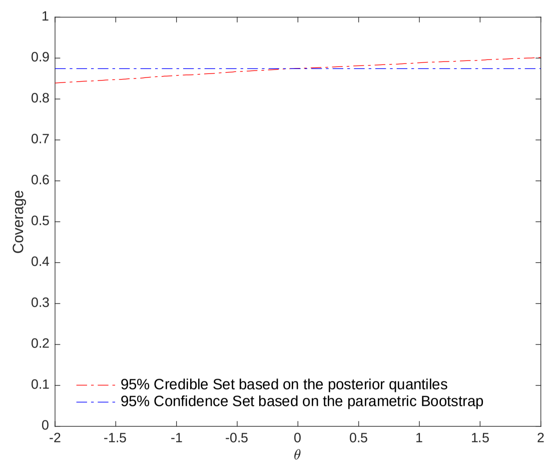
Figure 1: Coverage probability of 95% Credible Sets and Parametric Bootstrap Confidence Intervals.



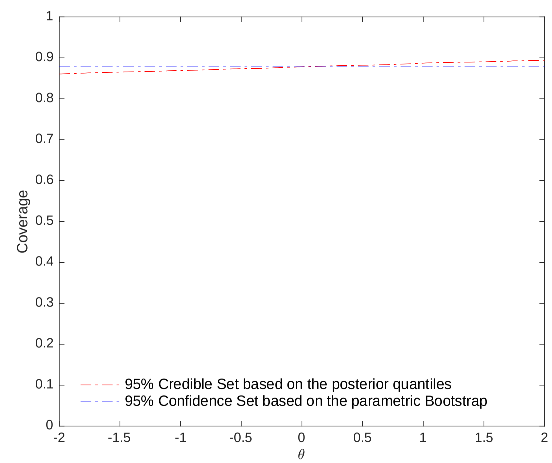
(a) $n = 100$



(b) $n = 200$



(c) $n = 300$



(d) $n = 500$

Description of Figure 2: Coverage probabilities of 95% bootstrap confidence intervals and 95% Credible Sets for $g(\theta) = \max\{\theta_1, \theta_2\}$ at $\theta_1 = \theta_2 = \theta \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$ based on data from samples of size $n \in \{100, 200, 300, 500\}$. (Blue, Dotted Line) Coverage probability of 95% confidence intervals based on the quantiles of the parametric bootstrap distribution of $g(\widehat{\theta}_n)$; that is, $g(N_2(\widehat{\theta}_n, \mathbb{I}_2/n))$. (Red, Dotted Line) 95% credible sets based on quantiles of the posterior distribution of $g(\theta)$; that is $g(\mathcal{N}_2(\frac{n}{n+\lambda^2}\widehat{\theta}_n + \frac{\lambda^2}{n+\lambda^2}\mu,\, \frac{1}{n+\lambda^2}\mathbb{I}_2))$ for a prior characterized by $\mu = 0$ and $\lambda^2 = 1$.

Remark 1. Dümbgen (1993) and Hong and Li (2018) have proposed re-scaling the bootstrap to conduct inference about a directionally differentiable parameter. More specifically, the re-scaled bootstrap in Dümbgen (1993) and the numerical delta-method in Hong and Li (2018) can be implemented by constructing a new random variable:

$$y_n^* \equiv n^{1/2-\delta}\left(g\left(\frac{1}{n^{1/2-\delta}}Z_n^* + \widehat{\theta}_n\right) - g(\widehat{\theta}_n)\right),$$

where $0 \leq \delta \leq 1/2$ is a fixed parameter and Z_n^* could be either Z_n^{P*} or Z_n^{B*}. The suggested confidence interval is of the form:

$$CS_n^H(1-\alpha) = \left[g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{1-\alpha/2},\ g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{\alpha/2}\right] \tag{C.1}$$

where c^*_β denote the β-quantile of y_n^*. Hong and Li (2018) have recently established the pointwise validity of the confidence interval above.

Whenever (C.1) is implemented using posterior draws; i.e., by relying on draws from:

$$Z_n^{P*} \equiv \sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n),$$

it seems natural to use the same posterior distribution to evaluate the credibility of the proposed confidence set. Figure 2 reports both the frequentist coverage and the Bayesian credibility of (C.1), assuming that the Hong and Li (2018) procedure is implemented using the posterior:

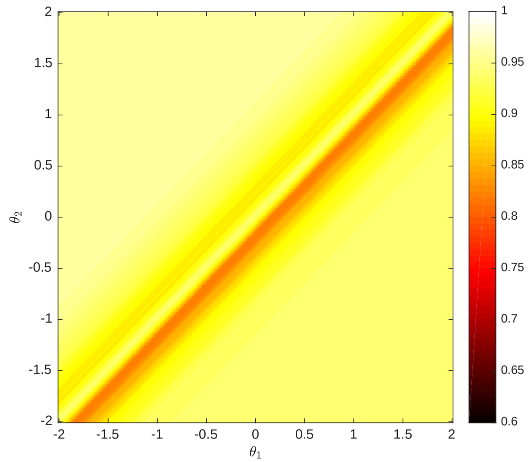
$$\theta_n^{P*}|X^n \sim \mathcal{N}_2\Big(\frac{n}{n+1}\widehat{\theta}_n\ ,\ \frac{1}{n+1}\mathbb{I}_2\Big).$$

The following figure shows that at least in this example fixing coverage comes at the expense of distorting Bayesian credibility.[3]

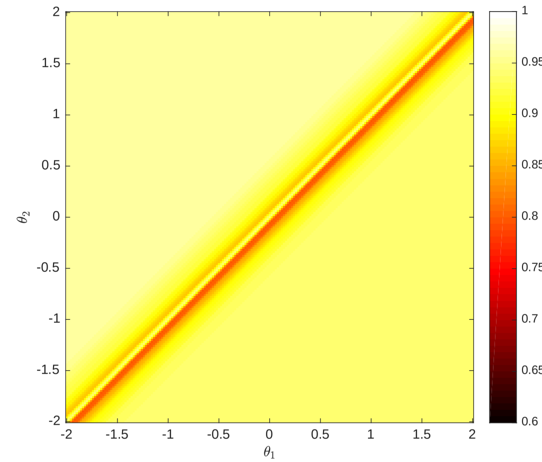
[3]The Bayesian credibility of $CS_n^H(1-\alpha)$ is given by:

$$\begin{aligned}&\mathbb{P}^*(g(\theta_n^{P*}) \in CS_n^H(1-\alpha)|X^n)\\ &\quad = \mathbb{P}^*\left(g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{1-\alpha/2}(X^n) \leq g(\theta_n^{P*}) \leq g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{\alpha/2}(X^n)\ \Big|\ X^n\right)\end{aligned}$$

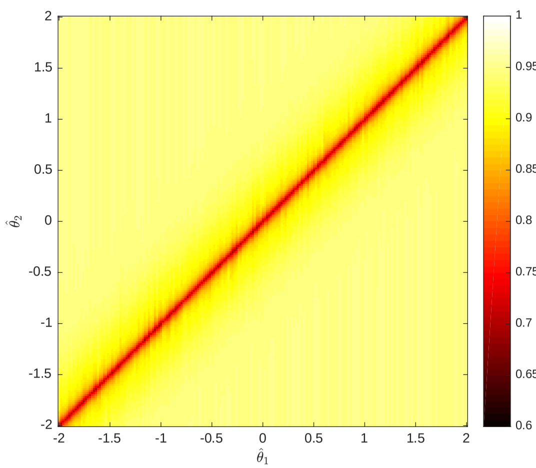
Figure 2: Coverage probability and Credibility of 95% Confidence Sets based on y_n^*



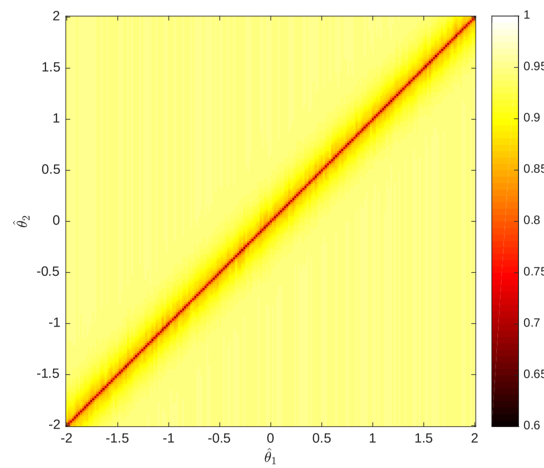
(a) Coverage Probability ($n = 100$)



(b) Coverage Probability ($n = 1000$)



(c) Credibility ($n = 100$)



(d) Credibility ($n = 1000$)

Description of Figure 2: Plots (a) and (b) show heat maps depicting the coverage probability of confidence sets based on the scaled random variable y_n^* for sample sizes $n \in \{100, 1000\}$ when $\theta_1, \theta_2 \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$. Plots (c) and (d) show heat maps depicting the credibility of confidence sets based on the scaled random variable y_n^* for sample sizes $n \in \{100, 1000\}$ when $\theta = 0$, $\Sigma = \mathbb{I}_2$, Z_n^* is approximated by $N_2(0, \Sigma)$ for computing the quantiles of y_n^* and $\widehat{\theta}_{n,1}, \widehat{\theta}_{n,2} \in [-2, 2]$.

References

Dudley, R. (2002): *Real Analysis and Probability*, vol. 74, Cambridge University Press.

Dümbgen, L. (1993): "On nondifferentiable functions and the bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Fang, Z. and A. Santos (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86, 377–412.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2009): *Bayesian data analysis*, vol. 2 of *Texts in Statistical Science*, Taylor & Francis.

Hong, H. and J. Li (2018): "The numerical delta-method," *Journal of Econometrics*, 206, 379–394.

Morand, O., K. Reffett, and S. Tarafdar (2015): "A nonsmooth approach to envelope theorems," *Journal of Mathematical Economics*, 61, 157–165.

Nadarajah, S. and S. Kotz (2008): "Exact distribution of the max/min of two Gaussian random variables," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 16, 210–212.

Van der Vaart, A. (2000): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.