cemmap

centre for microdata methods and practice

Posterior Distribution of Nondifferentiable Functions

Toru Kitagawa
Jose-Luis Montiel-Olea
Jonathan Payne

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP20/16

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

POSTERIOR DISTRIBUTION OF NONDIFFERENTIABLE FUNCTIONS.[1]

TORU KITAGAWA[2], JOSÉ-LUIS MONTIEL-OLEA[3] AND JONATHAN PAYNE[4]

This paper examines the asymptotic behavior of the posterior distribution of a possibly nondifferentiable function $g(\theta)$, where θ is a finite dimensional parameter. The main assumption is that the distribution of the maximum likelihood estimator $\widehat{\theta}_n$, its bootstrap approximation, and the Bayesian posterior for θ all agree asymptotically.

It is shown that whenever g is Lipschitz, though not necessarily differentiable, the posterior distribution of $g(\theta)$ and the bootstrap distribution of $g(\widehat{\theta}_n)$ coincide asymptotically. One implication is that Bayesians can interpret bootstrap inference for $g(\theta)$ as approximately valid posterior inference in a large sample. Another implication—built on known results about bootstrap inconsistency—is that the posterior distribution of $g(\theta)$ does not coincide with the asymptotic distribution of $g(\widehat{\theta}_n)$ at points of nondifferentiability. Consequently, frequentists cannot presume that credible sets for a nondifferentiable parameter $g(\theta)$ can be interpreted as approximately valid confidence sets (even when this relation holds true for θ).

KEYWORDS: Bootstrap, Bernstein von-Mises Theorem, Directional Differentiability, Posterior Inference.

1. INTRODUCTION

This paper studies the posterior distribution of a real-valued function $g(\theta)$, where θ is a parameter of finite dimension. We focus on a class of models where the transformation $g(\theta)$ is Lipschitz continuous but possibly nondifferentiable. Some stylized examples are:

$$|\theta|, \max\{0, \theta\}, \max\{\theta_1, \theta_2\}.$$

Parameters of the type considered in this paper arise in a wide range of applications in economics and statistics. Some examples are the welfare level attained by an optimal treatment assignment rule in the treatment choice problem (Manski (2004)); a trading strategy in an asset market (Jha and Wolak (2015)); the regression function in a regression kink model with an unknown threshold (Hansen (2015)); the

[1]We would like to thank Gary Chamberlain, Tim Cogley and Quang Vuong for detailed comments and suggestions on a earlier draft of this paper. All errors remain our own. Financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001) is gratefully acknowledged. This draft: May 6, 2016.
[2]University College London, Department of Economics. E-mail: t.kitagawa@ucl.ac.uk.
[3]New York University, Department of Economics. E-mail: montiel.olea@nyu.edu.
[4]New York University, Department of Economics. E-mail: jep459@nyu.edu.

eigenvalues of a random symmetric matrix (Eaton and Tyler (1991)); and the value function of stochastic mathematical programs (Shapiro (1991)). The lower and upper bound of the identified set in a partially identified model are also examples of parameters that fall within the framework of this paper.[1]

The potential nondifferentiability of $g(\cdot)$ poses different challenges to frequentist inference. For example, different forms of the bootstrap lose their consistency whenever differentiability is compromised; see Dümbgen (1993), Andrews (2000) and the recent characterization of bootstrap failure in Fang and Santos (2015). To our knowledge, the literature has not yet explored how the Bayesian posterior of $g(\theta)$ relates to the distribution of the (plug-in) maximum likelihood (ML) estimator and its bootstrap distribution when g is allowed to be nondifferentible.

This paper studies these relations in large samples. The main assumptions are that: (i) the ML estimator for θ, denoted by $\widehat{\theta}_n$, is $\sqrt{n}$-asymptotically normal, (ii) the bootstrap consistently estimates the asymptotic distribution of $\widehat{\theta}_n$ and (iii) the Bernstein-von Mises Theorem holds for θ (DasGupta (2008), p. 291); i.e., the Bayesian posterior distribution of θ coincides with the asymptotic distribution of $\widehat{\theta}_n$.

This paper shows that—after appropriate centering and scaling—the posterior distribution of $g(\theta)$ and the bootstrap distribution of $g(\widehat{\theta}_n)$ are asymptotically equivalent. This means that the bootstrap distribution of $g(\widehat{\theta}_n)$ contains, in large samples, the same information as the posterior distribution for $g(\theta)$.[2]

This result provides two useful insights. First, Bayesians can interpret bootstrap-based inference for $g(\theta)$ as approximately valid posterior inference in a large sample. Thus, Bayesians can use bootstrap draws to conduct approximate posterior inference for $g(\theta)$ when computing $\widehat{\theta}_n$ is simpler than Markov Chain Monte Carlo (MCMC) sampling.

Second, combined with the known results on the failure of bootstrap inference, we show that the Bernstein-von Mises Theorem for $g(\theta)$ will not hold even under mild departures from differentiability. In particular, the posterior distribution of $g(\theta)$

[1] For example, treatment effect bounds (Manski (1990), Balke and Pearl (1997)); bounds in auction models (Haile and Tamer (2003)), and bounds for impulse-response functions (Giacomini and Kitagawa (2015), Gafarov, Meier, and Montiel Olea (2015)) and forecast-error variance decompositions (Faust (1998)) in structural vector autoregressions.

[2] Other results in the literature concerning the relations between bootstrap and posterior inference have focused on the Bayesian interpretation of the bootstrap in finite samples, for example Rubin (1981), or on how the parametric bootstrap output can be used for efficient computation of the posterior, for example Efron (2012).

will not coincide with the asymptotic distribution of $g(\widehat{\theta}_n)$ whenever $g(\cdot)$ only has directional derivatives as in the pioneering work of Hirano and Porter (2012). In fact, it is shown that whenever directional differentiability causes a bootstrap confidence set to cover $g(\theta)$ less often than desired, a credible set based on the quantiles of the posterior will have distorted frequentist coverage as well.

The rest of this paper is organized as follows. Section 2 presents a formal statement of the main results. Section 3 presents an illustrative example: the absolute value transformation. Section 4 concludes. All the proofs are collected in the Appendix.

2. MAIN RESULTS

Let $X^n = \{X_1, \dots X_n\}$ be a sample of i.i.d. data from the parametric model $f(x_i \mid \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^p$. Let $\widehat{\theta}_n$ denote the ML estimator of θ and let θ_0 denote the true parameter of the model. Consider the following assumptions:

Assumption 1 The function $g : \mathbb{R}^p \to \mathbb{R}$ is Lipschitz continuous with constant c. That is;

$$|g(x) - g(y)| \leq c||x - y|| \quad \forall\, x, y \in \mathbb{R}^p.$$

Assumption 1 implies—by means of the well-known Rademacher's Theorem (Evans and Gariepy (2015), p. 81)—that g is differentiable almost everywhere in $\mathbb{R}^p$. Thus, the functions considered in this paper allow only for mild departures from differentiability.[3]

Assumption 2 The sequence $Z_n \equiv \sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$, where $\mathcal{I}^{-1}(\theta_0)$ is the inverse of Fisher's Information matrix evaluated at θ_0.

Assumption 2 is high-level, but there are well-known conditions on the statistical model $f(x; \theta)$ under which Assumption 2 obtains (see, for example, Newey and McFadden (1994) p. 2146).

In order to state the next assumption, we introduce additional notation. Define the set:

$$\mathrm{BL}(1) \equiv \Big\{ f : \mathbb{R}^p \to \mathbb{R} \,|\, \sup_{a \in \mathbb{R}^k} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq ||a_1 - a_2|| \quad \forall a_1, a_2 \Big\}.$$

[3]Moreover, we assume that g is defined everywhere in $\mathbb{R}^p$ which rules out examples such as the ratio of means θ_1/θ_2, $\theta_2 \neq 0$ discussed in Fieller (1954) and weakly identified Instrumental Variables models.

Let ϕ_n^* and ψ_n^* be random variables whose distribution depends on the data X^n. The Bounded Lipschitz distance between the distributions induced by ϕ_n^* and ψ_n^* (conditional on the data X^n) is defined as:

$$\beta(\phi_n^*, \psi_n^*;\ X^n) \equiv \sup_{f \in BL(1)} \Big| \mathbb{E}[f(\phi_n^*)|X^n] - \mathbb{E}[f(\psi_n^*)|X^n] \Big|.$$

The random variables ϕ_n^* and ψ_n^* are said to converge *in Bounded Lipschitz distance in probability* if $\beta(\phi_n^*, \psi_n^*;\ X^n) \xrightarrow{p} 0$ as $n \to \infty$.[4]

Let θ_n^{P*} denote the random variable with law equal to the posterior distribution of θ in a sample of size n. Let θ_n^{B*} denote the random variable with law equal to the bootstrap distribution of the Maximum Likelihood estimator of θ in a sample of size n.

REMARK 1 In a parametric model for i.i.d. data there are different ways of bootstrapping the distribution of $\widehat{\theta}_n$. One possibility is a *parametric bootstrap*, which consists in generating draws $(x_1, \dots x_n)$ from the model $f(x_i; \widehat{\theta}_n)$ followed by an evaluation of the ML estimator for each draw (Van der Vaart (2000) p. 328). Another possibility is the standard *mutinomial bootstrap*, which generates draws $(x_1, \dots x_n)$ from its empirical distribution. We do not take a stand on the specific bootstrap procedure used by the researcher as long as it is consistent. This is formalized in the following assumption.

ASSUMPTION 3 The centered and scaled random variables:

$$Z_n^{P*} \equiv \sqrt{n}(\theta_n^{P*} - \hat{\theta}_n) \quad \text{and} \quad Z_n^{B*} \equiv \sqrt{n}(\theta_n^{B*} - \hat{\theta}_n),$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of the ML estimator $Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$, which is independent of the data. That is,

$$\beta(Z_n^{P*}, Z;\ X^n) \xrightarrow{p} 0$$

and

$$\beta(Z_n^{B*}, Z;\ X^n) \xrightarrow{p} 0.$$

Sufficient conditions for Assumption 3 to hold are the consistency of the boot-

[4]For a more detailed treatment of the bounded lipschitz metric over probability measures see the 'β' metric defined in p. 394 of Dudley (2002).

strap for the distribution of $\widehat{\theta}_n$ (Horowitz (2001), Van der Vaart and Wellner (1996) Chapter 3.6, Van der Vaart (2000) p. 340) and the Bernstein-von Mises Theorem for θ.[5]

The following theorem shows that under the first three assumptions, the Bayesian posterior for $g(\theta)$ and the frequentist bootstrap distribution of $g(\widehat{\theta}_n)$ converge (after appropriate centering and scaling). Note that for any measurable function $g(\cdot)$, be it differentiable or not, the posterior distribution of $g(\theta)$ can be defined as the *image measure* induced by the distribution of θ_n^{P*} under the mapping $g(\cdot)$.

Theorem 1 *Suppose that Assumptions 1, 2 and 3 hold. Then,*

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

That is, after centering and scaling, the posterior distribution $g(\theta)$ and the bootstrap distribution of $g(\widehat{\theta}_n)$ are asymptotically close to each other in terms of the Bounded Lipschitz metric in probability.

Proof: See Appendix A.1. *Q.E.D.*

The intuition behind Theorem 1 is the following. The centered and scaled posterior and bootstrap distributions can be written as:

$$\begin{aligned}\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) &= \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),\\ \sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) &= \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))\end{aligned}$$

Because Z_n^{P*} and Z_n^{B*} both converge, by assumption, to a common limit Z and g is Lipschitz, the centered and scaled posterior and bootstrap distributions (conditional on the data) can both be well approximated by:

$$\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

and so the desired convergence result obtains. Theorem 1 does not rely on the normality assumption, but we impose this condition for the sake of exposition.

[5]Note that the Berstein-von Mises Theorem is oftentimes stated in terms of almost-sure convergence of the posterior density to a normal density (DasGupta (2008) p. 291). This mode of convergence (total variation metric) implies convergence in the bounded Lipschitz metric in probability. In this paper, all the results concerning the asymptotic behavior of the posterior are presented in terms of the Bounded-Lipschitz metric. This facilitates comparisons with the boostrap.

If the additional assumption is made that the function g is directionally differentiable, a common approximation to the distribution of the bootstrap and the posterior can be characterized explictly.

Assumption 4 There exists a continuous function $g'_{\theta_0} : \mathbb{R}^p \to \mathbb{R}$ such that for any compact set $K \subseteq \mathbb{R}^p$ and any sequence of positive numbers $t_n \to 0$:

$$\sup_{h \in K} \left| t_n^{-1}(g(\theta_0 + t_n h) - g(\theta_0)) - g'_{\theta_0}(h) \right| \to 0,$$

The continuous, not necessarily linear, function $g'_\theta(\cdot)$ will be referred to as the (Hadamard) directional derivative of g at θ_0.[6]

Remark 2 Note that the notion of Hadamard directional derivative is, in principle, stronger than the notion of one-sided directional derivative (or Gateaux directional derivative) used in Hirano and Porter (2012). The former requires the approximation to be uniform over directions h that belong to a compact set, whereas the latter is pointwise in h. However, when g is Lipschitz it can be shown that one-sided directional differentiability implies Hadamard directional differentiability and so, for the environment described in this paper, the concepts are equivalent. We state Assumption 4 using the Hadamard formulation (as that is the property required in the proofs), but we remind the reader that in a Lipschitz environment it is sufficient to show one-sided directional differentiability to verify Assumption 4.

Corollary 1 *Let θ_0 denote the parameter that generated the data. Under Assumptions 1, 2, 3, and 4:*

$$\beta(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)),\ g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)\ ;\ X^n) \xrightarrow{p} 0,$$

where $Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$ and $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$.

Proof: See Appendix A.2. *Q.E.D.*

[6] Equivalently, one could say there is a continuous function $g'_\theta : \mathbb{R}^k \to \mathbb{R}$ such that for any converging sequence $h_n \to h$:

$$\left| \sqrt{n}\Big(g\left(\theta_0 + \frac{h_n}{\sqrt{n}} \right) - g(\theta_0) \Big) - g'_{\theta_0}(h_n) \right| \to 0.$$

See p. 479 in Shapiro (1990).

The distribution $g'_{\theta_0}(Z+Z_n) - g'_{\theta_0}(Z_n)$ (which still depends on the sample size) provides a large-sample approximation to the distribution of $g(\theta_n^{P*})$. Our result shows that, in large samples, after centering aroung $g(\hat{\theta}_n)$, the data will only affect the posterior distribution through $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$.

The approximating distribution has appeared in the literature before, see Proposition 1 in Dümbgen (1993) and equation A.41 in Theorem A.1 in Fang and Santos (2015). Thus, verifying the assumptions for any of these papers in combination with our Theorem 1 would suffice to establish our Corollary 1. In order to keep the exposition self-contained, we decided to present a simpler derivation of this law using our own specific assumptions.

The intuition behind our proof is as follows. When g is directionally differentiable the approximation used to establish Theorem 1:

$$\begin{aligned}\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)) &= \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) \\ &- \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0)),\end{aligned}$$

can be further refined to:

$$g'_{\theta_0}(Z+Z_n) - g'_{\theta_0}(Z_n).$$

This follows from the fact that

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

is a perturbation around θ_0 in the random direction $(Z + Z_n)$ and is well approximated by $g'_{\theta_0}(Z+Z_n)$. Likewise,

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

is a perturbation around θ_0 in direction Z_n and it is well approximated by $g'_{\theta_0}(Z_n)$.

Remark 3 Note that if g'_{θ_0} is linear (which is the case if g is fully differentiable), then $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ converges to:

$$g'_{\theta_0}(Z+Z_n) - g'_{\theta_0}(Z_n) = g'_{\theta_0}(Z) \sim \mathcal{N}(0, (g'_{\theta_0})^T \mathcal{I}^{-1}(\theta_0)(g'_{\theta_0})),$$

where $(g'_{\theta_0})^T$ denotes the transpose of the gradient vector g'_{θ_0}. This is the same limit as one would get from applying the delta method to $g(\widehat{\theta}_n)$. Thus, under full differentiability, the posterior distribution of $g(\theta)$ can be approximated as:

$$g(\theta_n^{P*}) \approx g(\widehat{\theta}_n) + \frac{1}{\sqrt{n}} g'_{\theta_0}(Z).$$

Moreover, this distribution coincides with the asymptotic distribution of the plug-in estimator $g(\widehat{\theta}_n)$. The obvious remark is that full differentiability is sufficient for a Bernstein-von Mises Theorem to hold for $g(\theta_n^{P*})$.

If g'_{θ_0} is nonlinear the limiting distribution of $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ becomes a nonlinear transformation of Z. This nonlinear transformation need not be Gaussian, and need not be centered at zero. Moreover, the nonlinear transformation $g'_{\theta_0}(Z+Z_n) - g'_{\theta_0}(Z_n)$ is different from the asymptotic distribution of the plug-in estimator $g(\widehat{\theta}_n)$ which is given by $g'_{\theta_0}(Z)$.[7] Informally, one can say that for directionally differentiable functions:

$$g(\theta_n^{P*}) \approx g(\widehat{\theta}_n) + \frac{1}{\sqrt{n}}(g'_{\theta_0}(Z+Z_n) - g'_{\theta_0}(Z_n)), \text{ where } Z_n = \sqrt{n}(\widehat{\theta}_n - \theta_0).$$

Failure of Bootstrap Inference: Theorem 1 established the large-sample equivalence between the bootstrap distribution of $g(\widehat{\theta}_n)$ and the posterior distribution of $g(\theta)$. We now use this Theorem to make a concrete connection between the coverage of bootstrap-based confidence sets and the coverage of Bayesian credible sets based on the quantiles of the posterior.

Neither the results of Dümbgen (1993) nor those of Fang and Santos (2015) offer a concrete chacterization of the asymptotic coverage of bootstrap-based confidence sets. Despite the bootstrap inconsistency established in these papers, it is still possible that the bootstrap confidence sets have correct asymptotic coverage.

In this section we will not insist in characterizing the asymptotic coverage of bootstrap and/or posterior inference. Instead, we start by assuming that a nominal $(1-\alpha)$ bootstrap confidence set fails to cover $g(\theta)$ at a point of directional differentiability. Then, we show that a $(1-\alpha-\epsilon)$ credible set based on the quantiles of the posterior distribution of $g(\theta)$ will also fail to cover $g(\theta)$ for any $\epsilon > 0$.[8]

[7] This follows from an application of the delta-method for directionally differentiable functions in Shapiro (1991)) or from Proposition 1 in Dümbgen (1993).

[8] The adjustment factor ϵ is introduced because the the quantiles of both the bootstrap and the posterior remain random even in large samples.

This result is not a direct corollary of Theorem 1 as there is some extra work needed to relate the quantiles of the bootstrap distribution of $g(\widehat{\theta}_n)$ and the quantiles of the posterior of $g(\theta)$.

Set-up: Let $q^B_\alpha(X^n)$ be defined as:

$$q^B_\alpha(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{B*}(g(\theta^{B*}_n) \leq c \mid X^n) \geq \alpha\}.$$

The quantile based on the posterior distribution $q^P_\alpha(X^n)$ is defined analogously.

A nominal $(1-\alpha)\%$ two-sided confidence set for $g(\theta)$ based on the bootstrap distribution $g(\theta^{B*}_n)$ can be defined as follows:

$$CS^B_n(1-\alpha) \equiv \Big[q^B_{\alpha/2}(X^n)\ ,\ q^B_{1-\alpha/2}(X^n)\Big]. \tag{2.1}$$

This is a typical confidence set based on the percentile method of Efron, p. 327 in Van der Vaart (2000).

Definition We say that the nominal $(1-\alpha)\%$ bootstrap confidence set fails to cover the parameter $g(\theta)$ at θ by at least $d_\alpha\%$ ($d_\alpha > 0$) if:

$$\limsup_{n\to\infty} \mathbb{P}_\theta\Big(g(\theta) \in CS^B_n(1-\alpha)\Big) \leq 1 - \alpha - d_\alpha, \tag{2.2}$$

where $\mathbb{P}_\theta$ refers to the distribution of X_i under parameter θ.

Let $F_\theta(y|Z_n)$ denote the c.d.f. of the random variable $Y \equiv g'_\theta(Z + Z_n)$ conditional on Z_n. In order to relate our weak convergence results to the behavior of the quantiles, we assume that at the point θ where the bootstrap fails the following assumption is satisfied:

Assumption 5 The c.d.f. $F_\theta(y|Z_n)$ is Lipschitz continuous with a constant k that does not depend on Z_n.[9]

Assumption 5 suffices to relate the coverage of a confidence set for $g(\theta)$ based on the quantiles of the posterior of $g(\theta)$ with the coverage of a bootstrap confidence set.[10] We establish this connection as another Corollary to Theorem 1.

---

[9] A sufficient condition for this result to hold is that the density $h_\theta(y|Z_n)$ admits an upper bound independent of Z_n. This will be the case in the illustrative example we consider.

[10] Assumption 5 could be relaxed. Appendix A.3 presents a high-level condition implied by Assumption 5 that requires the existence of a value ζ such that, for all $c \in \mathbb{R}$, the probability that $g'_\theta(Z + Z_n) - g'_\theta(Z_n)$ is in the interval $[c-\zeta, c+\zeta]$ can be made arbitrarily small (for most data realizations).

Corollary 2 *Suppose that the nominal $(1-\alpha)\%$ bootstrap confidence set fails to cover $g(\theta)$ at θ by at least $d_\alpha\%$. If Assumptions 1 to 5 hold then for any $\epsilon > 0$:*

$$\limsup_{n\to\infty} \mathbb{P}_\theta \left(g(\theta) \in \left[q^P_{(\alpha+\epsilon)/2}(X^n) \, , \, q^P_{1-(\alpha+\epsilon)/2}(X^n) \right] \right) \leq 1 - \alpha - d_\alpha .$$

Thus, for any $0 < \epsilon < d$, the nominal $(1-\alpha-\epsilon)\%$ credible set based on the quantiles of the posterior fails to cover $g(\theta)$ at θ by at least $(d_\alpha - \epsilon)\%$.

Proof: See Appendix A.3. *Q.E.D.*

The intuition behind the theorem is the following. For convenience, let θ^*_n denote either the bootstrap or posterior random variable and let $c^*_\beta(X^n)$ denote the β-critical value of $g(\theta^*_n)$ defined by:

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

We show that $c^*_\beta(X^n)$ is asymptotically close to the β-quantile of $g'_\theta(Z+Z_n) - g'_\theta(Z_n)$, denoted by $c_\beta(Z_n)$. More precisely, we show that for arbitrarily small $0 < \epsilon < \beta$ and $\delta > 0$, the probability that $c^*_\beta(X^n) \in [c_{\beta-\epsilon/2}(Z_n), c_{\beta+\epsilon/2}(Z_n)]$ is greater than $1-\delta$ for sufficiently large n.

Because under Assumptions 1 to 5 the critical values of both the bootstrap and posterior distributions are asymptotically close to the quantiles of $g'_\theta(Z + Z_n) - g'_\theta(Z_n)$, we can show that for a fixed $\epsilon > 0$ and sufficiently large n:

$$\mathbb{P}_\theta \left(g(\theta) \in CS^B_n(1-\alpha) \right) \;=\; \mathbb{P}_\theta \left(g(\theta) \in \left[q^P_{(\alpha+\epsilon)/2}(X^n) \, , \, q^P_{1-(\alpha+\epsilon)/2}(X^n) \right] \right) - \delta.$$

It follows that when the $(1-\alpha)\%$–bootstrap confidence set fails to cover the parameter $g(\theta)$ at θ, then so must the $(1-\alpha-\epsilon)\%$–credible set.[11]

[11] It immediately follows that the reverse also applies. If the $(1-\alpha)\%$–credible set fails to cover the parameter $g(\theta)$ at θ, then so must the $(1-\alpha-\epsilon)\%$–bootstrap confidence set. Note that our approximation holds for any fixed ϵ, but we cannot guarantee that our approximation holds if we take the limit.

3. ILLUSTRATION OF MAIN RESULTS FOR $|\theta|$

The main result of this paper, Theorem 1, can be illustrated in the following simple environment. Let $X^n = (X_1, \dots X_n)$ be an i.i.d. sample of size n from the statistical model:

$$X_i \sim \mathcal{N}(\theta, 1).$$

Consider the following family of priors for θ:

$$\theta \sim N(0, (1/\lambda^2)),$$

where the precision parameter satisfies $\lambda^2 > 0$. The transformation of interest is the absolute value function:

$$g(\theta) = |\theta|.$$

It is first shown that when $\theta_0 = 0$ this environment satisfies Assumptions 1 to 5. Then, the bootstrap and posterior distributions for $g(\theta)$ are explicitly computed and compared.

Relation to main assumptions: The tranformation g is Lipschitz continuous and differentiable everywhere, except at $\theta_0 = 0$. At this particular point in the parameter space, g has directional derivative $g'_0(h) = |h|$. Thus, Assumptions 1 and Assumption 4 are both satisfied.

The Maximum Likelihood estimator is given by $\widehat{\theta}_n = (1/n)\sum_{i=1}^n X_i$ and so $\sqrt{n}(\widehat{\theta}_n - \theta) \sim Z \sim \mathcal{N}(0,1)$. This means that Assumption 2 is satisfied.

This environment is analytically tractible so the distributions of θ_n^{P*} and θ_n^{B*} can be computed explicitly. The posterior distribution for θ is given by:

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

which implies that:

$$\sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n)|X^n \sim \mathcal{N}\Big(\frac{\lambda^2}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n, \quad \frac{n}{n+\lambda^2}\Big).$$

Consequently,

$$\beta\Big(\sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n)\ ,\ \mathcal{N}(0,1); X^n\Big) \xrightarrow{p} 0.$$

This implies that under, $\theta_0 = 0$, the first part of Assumption 3 holds.[12]

Second, consider a parametric bootstrap for the sample mean, $\widehat{\theta}_n$. We decided to focus on the parametric bootstrap to keep the exposition as simple as possible. The parametric bootsrap is implemented by generating a large number of draws $(x_1^j, \ldots, x_n^j)$, $j = 1, \ldots, J$ from the model

$$x_i^j \sim \mathcal{N}(\widehat{\theta}_n, 1), \quad i = 1, \ldots n,$$

recomputing the ML estimator for each of the draws. This implies that the boostrap distribution of $\widehat{\theta}_n$ is given by:

$$\theta_n^{B*} \sim \mathcal{N}(\widehat{\theta}_n, 1/n),$$

and so, for the parametric bootstrap it is straightforward to see that:

$$\beta\Big(\sqrt{n}(\theta_n^{B*} - \widehat{\theta}_n)\ ,\ \mathcal{N}(0,1); X^n\Big) = 0.$$

This means that the second part of Assumption 3 holds.

Finally, in this example the p.d.f. of $Y \equiv g_0'(Z + Z_n) = |Z + Z_n|$ is that of a folded normal:

$$h_0(y|Z_n) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y - Z_n)^2\Big) + \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y + Z_n)^2\Big),$$

this expression follows by direct computation or by replacing $\cosh(x)$ in equation 29.41 in p. 453 in Johnson, Kotz, and Balakrishnan (1995) by $(1/2)(\exp(x) + \exp(-x))$. Note that:

$$h_0(y|Z_n) \leq \sqrt{\frac{2}{\pi}},$$

[12] The last equation follows from the fact that for two Gaussian real-valued random variables $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ we have that:

$$\Big|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]\Big| \leq \sqrt{\frac{2}{\pi}}\Big|\sigma_1^2 - \sigma_2^2\Big| + \Big|\mu_1 - \mu_2\Big|.$$

Therefore:

$$\beta\Big(\sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n)\ ,\ \mathcal{N}(0,1); X^n\Big) \leq \sqrt{\frac{2}{\pi}}\Big|\frac{n}{n+\lambda^2} - 1\Big| + \Big|\frac{\lambda^2}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n\Big|.$$

which implies that Assumption 5 holds. To see this, take $y_1 > y_2$. Note that:

$$F_\theta(y_1|Z_n) - F_\theta(y_2|Z_n) = \int_{y_2}^{y_1} h(y|Z_n)dy \leq (y_1 - y_2)\sqrt{\frac{2}{\pi}}.$$

An analogous argument for the case in which $y_1 \leq y_2$ implies that Assumption 5 is verified.

Asymptotic Behavior of Posterior Inference for $g(\theta) = |\theta|$: Since Assumptions 1 to 4 are satisfied, Theorem 1 and its Corollary hold.

In this example the posterior distribution of $g(\theta_P^*)|X^n$ is given by:

$$\Big|\frac{1}{\sqrt{n+\lambda^2}}Z^* + \frac{n}{n+\lambda^2}\widehat{\theta}_n\Big|, \quad Z^* \sim \mathcal{N}(0,1)$$

and therefore $\sqrt{n}(g(\theta_P^*) - g(\widehat{\theta}_n)$ can be written as :

$$\text{(3.1)} \qquad \Big|\frac{\sqrt{n}}{\sqrt{n+\lambda^2}}Z^* + \frac{n}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n\Big| - \Big|\sqrt{n}\widehat{\theta}_n\Big|, \quad Z^* \sim \mathcal{N}(0,1).$$

Theorem 1 and its Corollary show that when $\theta_0 = 0$ and n is large enough, this expression can be approximated in the Bounded Lipschitz metric in probability by:

$$\text{(3.2)} \qquad \Big|Z + Z_n\Big| - \Big|Z_n\Big| = \Big|Z + \sqrt{n}\widehat{\theta}_n\Big| - \Big|\sqrt{n}\widehat{\theta}_n\Big|, \quad Z \sim \mathcal{N}(0,1).$$

Observe that at $\theta_0 = 0$ the sampling distribution of the plug-in ML estimator for $|\theta|$ is given by:

$$\sqrt{n}(|\widehat{\theta}_n| - |\theta_0|) \sim |Z|.$$

Thus, the approximate distribution of the posterior differs from the asymptotic distribution of the plug-in ML estimator and the typical Guassian approximation for the posterior will not be appropriate.

Asymptotic Behavior of Parametric Bootstrap Inference for $g(\theta) = |\theta|$: The parametric bootstrap distribution of $|\widehat{\theta}_n|$, centered and scaled, is simply given by:

$$\Big|Z + \sqrt{n}\widehat{\theta}_n\Big| - |\sqrt{n}\widehat{\theta}_n|, \quad Z \sim N(0,1),$$

which implies that posterior distribution of $|\theta|$ and the bootstrap distribution of $|\widehat{\theta}_n|$

are asymptotically equivalent.

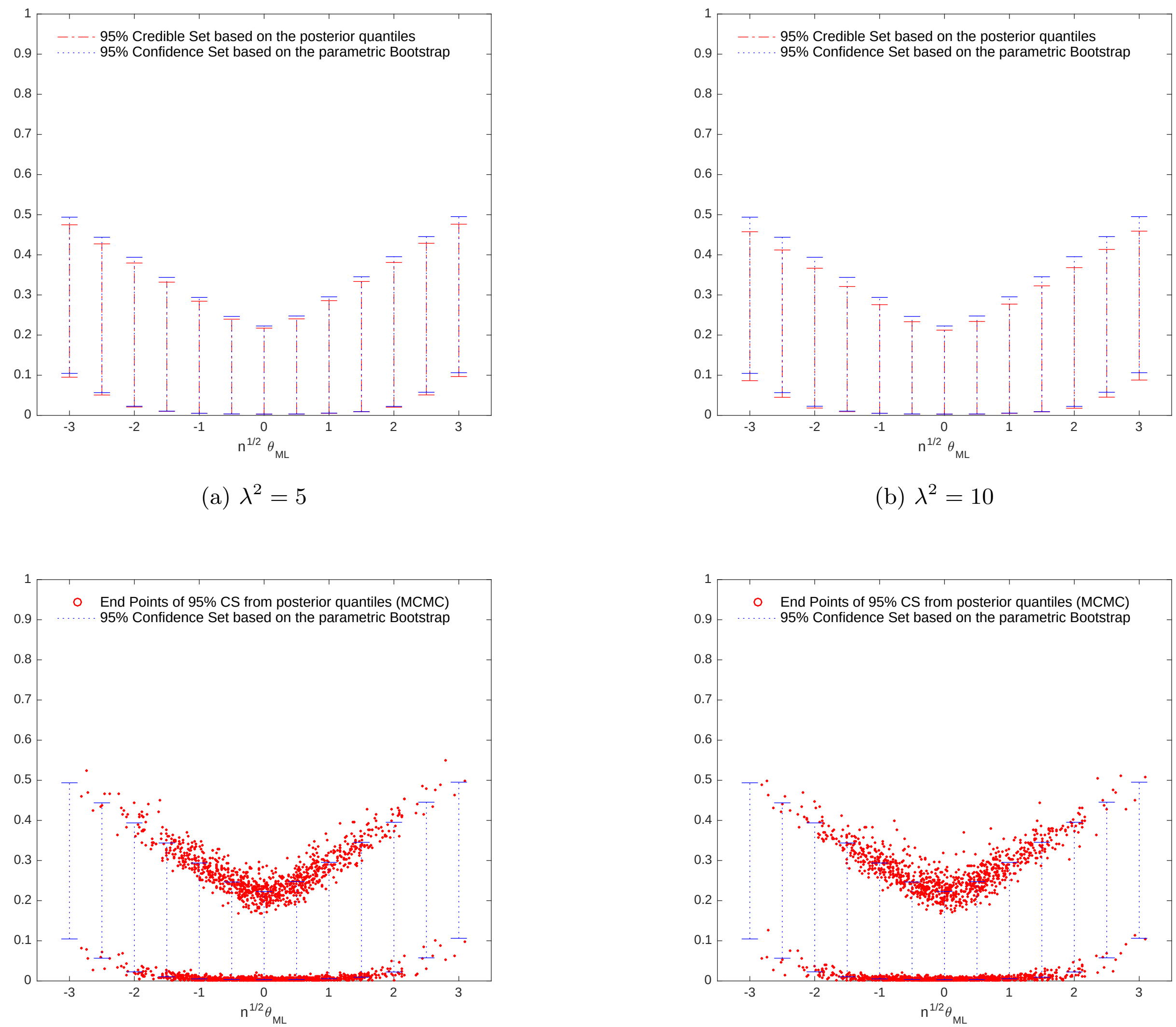
Graphical interpretation of Theorem 1: One way to illustrate Theorem 1 is to compute the 95% credible sets for $|\theta|$ when $\theta_0 = 0$ using the quantiles of the posterior. We can then compare the 95% credible sets to the 95% confidence sets from the bootstrap distribution. Note that in establishing Corollary 2 we have shown that this relation is indeed implied by Theorem 1.

Observe from (3.2) that the approximation to the centered and scaled posterior and bootstrap distributions depends on the data via $\sqrt{n}\hat{\theta}_n$. Thus, in Figure 2 we report the 95% credible and confidence sets for data realisations $\sqrt{n}\hat{\theta}_n \in [-3, 3]$. In all four plots the bootstrap confidence sets are computed using the parametric bootstrap. Posterior credible sets are presented for four different priors for θ: $\mathcal{N}(0, 1/5)$, $\mathcal{N}(0, 1/10)$, $\gamma(2, 2) - 3$ and $(\beta(2, 2) - 0.5) \times 5$. The posterior for the first two priors is obtained using the expression in (3.1), while the posterior for the last two priors is obtained using a the Metropolis-Hastings algorithm (Geweke (2005), p. 122).

Coverage of Credible Sets: In this example, the two-sided confidence set based on the quantiles of the bootstrap distribution of $|\widehat{\theta}_n|$ fails to cover $|\theta|$ when $\theta = 0$. Corollary 2 showed that the two-sided credible sets based on the quantiles of the posterior should exhibit the same problem. This is illustrated in Figure 2. Thus, a frequentist cannot presume that a credible set for $|\theta|$ based on the quantiles of the posterior will deliver a desired level of coverage.

As Liu, Gelman, and Zheng (2013) observe, although it is common to report credible sets based on the $\alpha/2$ and $1 - \alpha/2$ quantiles of the posterior, a Bayesian might find these credible sets unsatisfactory. In this problem, it is perhaps more natural to consider one-sided credible sets or Highest Posterior Density sets. In the online Appendix B we consider an alternative example, $g(\theta) = \max\{\theta_1, \theta_2\}$, where the decision between two-sided and one-sided credible sets is less obvious, but the two-sided credible set still experiences the same problem as the bootstrap.

Figure 1: 95% Credible Sets for $|\theta|$ and 95% Parametric Bootstrap Confidence Intervals



(a) $\lambda^2 = 5$

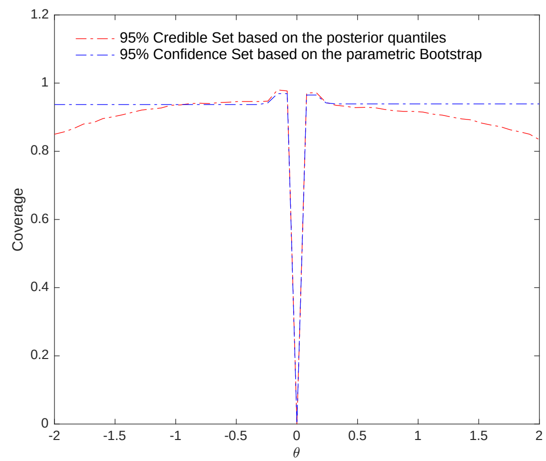
(b) $\lambda^2 = 10$

(c) $\gamma(2,2) - 3$

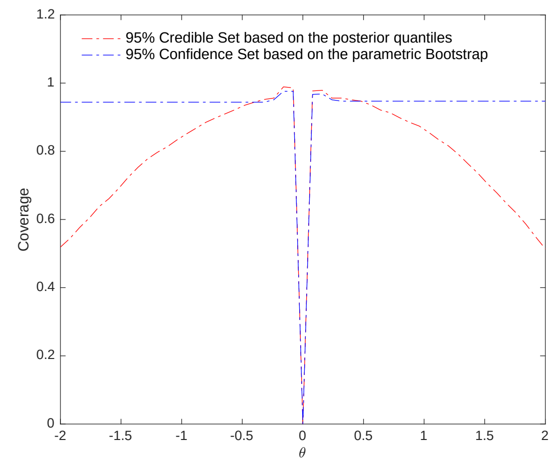
(d) $(\beta(2,2) - 0.5) \times 5$

Description of Figure 2: 95% Credible Sets for $|\theta|$ obtained from four different priors and evaluated at different realizations of the data ($n = 100$). (Blue, Dotted Line) 95% confidence intervals based on the quantiles of the bootstrap distribution $|N(\widehat{\theta}_n, 1/n)|$. The bootstrap distribution only depends on the data through $\widehat{\theta}_n$. (Red, Dotted Line) 95% credible sets based on the closed-form solution for the posterior. (Red, Circles) 95% credible sets based on Matlab's MCMC program (computed for a 1,000 possible data sets from a standard normal model).

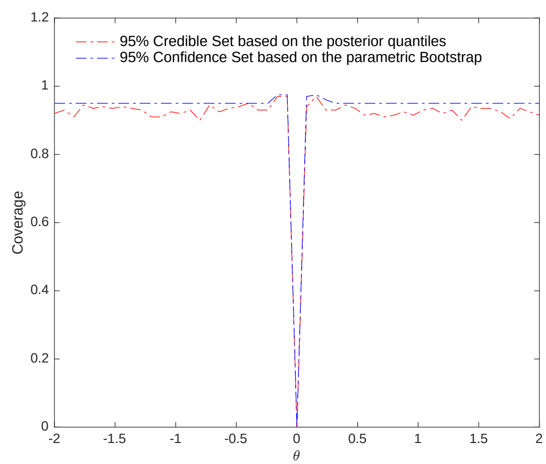
Figure 2: Coverage probability of 95% Credible Sets and Parametric Bootstrap Confidence Intervals for $|\theta|$



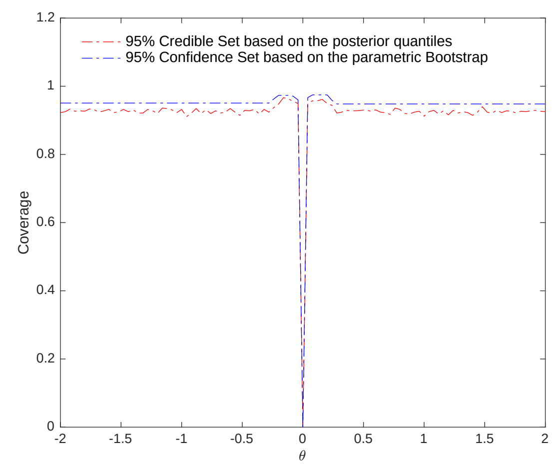
(a) $\lambda^2 = 5$



(b) $\lambda^2 = 10$



(c) $\gamma(2,2) - 3$



(d) $(\beta(2,2) - 0.5) \times 5$

Description of Figure 2: Coverage probability of 95% bootstrap confidence intervals and 95% Credible Sets for $|\theta|$ obtained from four different priors and evaluated at different realizations of the data ($n = 100$). (Blue, Dotted Line) Coverage probability of 95% confidence intervals based on the quantiles of the bootstrap distribution $|N(\widehat{\theta}_n, 1/n)|$. (Red, Dotted Line) 95% credible sets based on quantiles of the posterior. Cases (a) and (b) use the closed form expression for the posterior. Cases (c) and (d) use Matlab's MCMC program.

4. CONCLUSION

This paper studied the asymptotic behavior of the posterior distribution of parameters of the form $g(\theta)$, where $g(\cdot)$ is Lipschitz continuous but possibly nondifferentiable. We have shown that the bootstrap distribution of $g(\widehat{\theta}_n)$ and the posterior of $g(\theta)$ are asymptotically equivalent.

One implication from our results is that Bayesians can interpret bootstrap inference for $g(\theta)$ as approximately valid posterior inference in large samples. In fact, Bayesians can use bootstrap draws to conduct approximate posterior inference for $g(\theta)$ whenever bootstraping $g(\widehat{\theta}_n)$ is more convenient than MCMC sampling. This reinforces observations in the statistics literature noting that by "perturbing the data, the bootstrap approximates the Bayesian effect of perturbing the parameters" (Hastie, Tibshirani, Friedman, and Franklin (2005), p. 236).[13]

Another implication from our main result—combined with known results about bootstrap inconsistency—is that it takes only mild departures from differentiability (such as directional differentiability) to make the posterior distribution of $g(\theta)$ behave differently than the limit of $\sqrt{n}(g(\widehat{\theta}_n) - g(\theta))$. We showed that whenever directional differentiability causes a bootstrap confidence set to cover $g(\theta)$ less often than desired, a credible set based on the quantiles of the posterior will have distorted frequentist coverage as well.

For the sake of exposition, we restricted our analysis to parametric models. The main result of this paper should carry over to semiparametric models as far as the Bernstein-von Mises property and the bootstrap consistency for the finite-dimensional parameter θ hold. The Bernstein-von Mises theorem for smooth functionals in semiparametric models has been established recently in the work of Castillo and Rousseau (2015). The consistency of different forms of the bootstrap for semiparametric is well-known in the literature. The generalization of our main results to a semi-parametric environment considered in Castillo and Rousseau (2015) is left out for future work.

[13]Our results also provide a better understanding of what type of statistics could preserve, in large samples, the equivalence between bootstrap and posterior resampling methods, a question that have been explored by Lo (1987).

REFERENCES

Andrews, D. W. (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 68, 399–405.

Balke, A. and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

Castillo, I. and J. Rousseau (2015): "A Bernstein–von Mises theorem for smooth functionals in semiparametric models," *The Annals of Statistics*, 43, 2353–2383.

DasGupta, A. (2008): *Asymptotic theory of statistics and probability*, Springer Verlag.

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.

Eaton, M. L. and D. E. Tyler (1991): "On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix," *The Annals of Statistics*, 260–271.

Efron, B. (2012): "Bayesian inference and the parametric bootstrap," *The Annals of Applied Statistics*, 6, 1971.

Evans, L. C. and R. F. Gariepy (2015): *Measure theory and fine properties of functions*, CRC press.

Fang, Z. and A. Santos (2015): "Inference on Directionally Differentiable Functions," *Working paper, University of California at San Diego.*

Faust, J. (1998): "The Robustness of Identified VAR Conclusions about Money," in *Carnegie-Rochester Conference Series on Public Policy*, Elsevier, vol. 49, 207–244.

Fieller, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2015): "Delta-Method inference for a class of set-identified SVARs," *Working paper, New York University.*

Geweke, J. (2005): *Contemporary Bayesian econometrics and statistics*, vol. 537, John Wiley & Sons.

Giacomini, R. and T. Kitagawa (2015): "Robust Inference about partially identified SVARs," *Working Paper, University College London.*

Haile, P. A. and E. Tamer (2003): "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 111, 1–51.

Hansen, B. E. (2015): "Regression kink with an unknown threshold," *Journal of Business & Economic Statistics*.

Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin (2005): "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, 27, 83–85.

Hirano, K. and J. R. Porter (2012): "Impossibility results for nondifferentiable functionals," *Econometrica*, 80, 1769–1790.

Horowitz, J. L. (2001): "The bootstrap," *Handbook of econometrics*, 5, 3159–3228.

Jha, A. and F. A. Wolak (2015): "Testing for market efficiency with transactions costs: An application to convergence bidding in wholesale electricity markets," *Working paper, Stanford University*.

Johnson, N., S. Kotz, and N. Balakrishnan (1995): "Continuous Univariate Distributions, Vol. 2. 1995," .

Liu, Y., A. Gelman, and T. Zheng (2013): "Simulation-efficient shortest probability intervals," *arXiv preprint arXiv:1302.2142*.

Lo, A. Y. (1987): "A large sample study of the Bayesian bootstrap," *The Annals of Statistics*, 360–375.

Manski, C. F. (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 80, 319–323.

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246.

Newey, W. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 2111–2245.

Rubin, D. B. (1981): "The Bayesian Bootstrap," *The Annals of Statistics*, 9, 130–134.

Shapiro, A. (1990): "On concepts of directional differentiability," *Journal of optimization theory and applications*, 66, 477–487.

——— (1991): "Asymptotic analysis of stochastic programs," *Annals of Operations Research*, 30, 169–186.

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

Van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes.*, Springer, New York.

APPENDIX A: MAIN THEORETICAL RESULTS.

A.1. *Proof of Theorem 1*

Lemma 1 *Suppose that Assumption 1 holds. Suppose that θ_n^* is a random variable satisfying:*

$$\sup_{f \in BL(1)} \Big| \mathbb{E}[f(Z_n^*) \mid X^n] - \mathbb{E}[f(Z^*)] \Big| \overset{p}{\to} 0,$$

where $Z_n^* = \sqrt{n}(\theta_n^* - \hat{\theta}_n)$ and Z^* is a random variable independent of $X^n = (X_1, \ldots, X_n)$ for every n. Then,

$$\sup_{f \in BL(1)} \Big| \mathbb{E}[f(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))) \mid X^n] - \mathbb{E}[f(\sqrt{n}(g(\theta_0 + Z^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\widehat{\theta}_n))) \mid X^n] \Big| \overset{p}{\to} 0, \tag{A.1}$$

where θ_0 is the parameter that generated the data and $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$.

Proof: By Assumption 1, g is Lipschitz continuous. Define $\Delta_n(a) = \sqrt{n}(g(\theta_0 + a/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$. Observe that $\Delta_n(\cdot)$ is Lipschitz since:

$$\begin{aligned} |\Delta_n(a) - \Delta_n(b)| &= |\sqrt{n}(g(\theta_0 + a/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + b/\sqrt{n} + Z_n/\sqrt{n}))| \\ &\leq c\|a - b\|, \\ &\quad \text{(by Assumption 1).} \end{aligned}$$

Define $\tilde{c} = \max\{c, 1\}$. Then the function $(f \circ \Delta_n)/\tilde{c}$ is an element of BL(1). Consequently,

$$\begin{aligned} &\Big| \mathbb{E}[f(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))) \mid X^n] \\ &\qquad - \mathbb{E}[f(\sqrt{n}(g(\theta_0 + Z^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\widehat{\theta}_n))) \mid X^n] \Big| \\ &= \tilde{c} \Big| \mathbb{E}\Big[\frac{f \circ \Delta_n(Z_n^*)}{\tilde{c}} \Bigm| X^n \Big] - \mathbb{E}\Big[\frac{f \circ \Delta_n(Z^*)}{\tilde{c}} \Bigm| X^n \Big] \Big|, \\ &\quad (\text{since } \theta_n^* = \theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) \\ &\leq \tilde{c} \sup_{f \in BL(1)} \Big| \mathbb{E}[f(Z_n^*)|X^n] - \mathbb{E}[f(Z^*)|X^n] \Big|, \\ &\quad (\text{since } (f \circ \Delta_n)/\tilde{c} \in BL(1)). \end{aligned}$$

Q.E.D.

Proof of Theorem 1: Theorem 1 follows from Lemma 1. Note first that Assumptions 1, 2 and 3 imply that the assumptions of Lemma 1 are verified for both θ_n^{P*} and θ_n^{B*}. Note then that:

$$\begin{aligned} &\sup_{f \in BL(1)} \Big| \mathbb{E}[f(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))) \mid X^n] - \mathbb{E}[f(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))) \mid X^n] \Big| \\ &\quad \leq \sup_{f \in BL(1)} \Big| \mathbb{E}[f(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))) \mid X^n] \end{aligned}$$

$$
\begin{aligned}
&\qquad - \mathbb{E}\Big[f\Big(\sqrt{n}\Big(g\Big(\theta_0 + \frac{Z}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}} \Big) - g(\widehat{\theta}_n) \Big) \Big) \mid X^n \Big] \Big| \\
&+ \sup_{f \in BL(1)} \Big| \mathbb{E}[f(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))) \mid X^n] \\
&\qquad - \mathbb{E}\Big[f\Big(\sqrt{n}\Big(g\Big(\theta_0 + \frac{Z}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}} \Big) - g(\widehat{\theta}_n) \Big) \Big) \mid X^n \Big] \Big|.
\end{aligned}
$$

Lemma 1 implies that both terms converge to zero in probability. *Q.E.D.*

A.2. *Proof of the corollary to Theorem 1*

Lemma 2 *Let Z^* be a random variable independent of $X^n = (X_1, \ldots, X_n)$ and let θ_0 denote the parameter that generated the data. Suppose that Assumption 4 holds. Then,*

$$\sup_{f \in BL(1)} \left| \mathbb{E}\Big[f\Big(\sqrt{n}\Big(g\Big(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}\Big) - g\Big(\widehat{\theta}_n\Big)\Big)\Big) \Big| X^n \Big] - \mathbb{E}\Big[f\Big(g'_{\theta_0}\Big(Z^* + Z_n\Big) - g'_{\theta_0}\Big(Z_n\Big)\Big) \Big| X^n \Big] \right| \xrightarrow{p} 0.$$

Proof: Define the random variable:

$$W_n \equiv \sqrt{n}\Big(g\Big(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}\Big) - g\Big(\widehat{\theta}_n\Big)\Big) - \Big(g'_{\theta_0}\Big(Z^* + Z_n\Big) - g'_{\theta_0}\Big(Z_n\Big)\Big).$$

Let $\mathbb{P}^*$ denote the law of Z^* which, by assumption, is independent of X^n. We first show that for every $\epsilon > 0$:

$$\mathbb{P}^*(|W_n| > \epsilon | X^n) \xrightarrow{p} 0,$$

and then argue that this statement implies the desired result.

In order to prove $\mathbb{P}^*(|W_n| > \epsilon | X^n) \xrightarrow{p} 0$ we must show that for every $\epsilon, \eta, \delta > 0$ there exists $N(\epsilon, \eta, \delta)$ such that if $n > N(\epsilon, \eta, \delta)$:

$$\mathbb{P}^n\Big(\mathbb{P}^*(|W_n| > \epsilon | X^n) > \eta \Big) < \delta,$$

where $\mathbb{P}^n$ denotes the distribution of $(X_1, \ldots X_n)$. Observe that:

$$\Big| W_n \Big| \leq \Big| \sqrt{n}\Big(g\Big(\theta_0 + \frac{Z_n}{\sqrt{n}}\Big) - g(\theta_0)\Big) - g'_{\theta_0}\Big(Z_n\Big) \Big| + \Big| \sqrt{n}\Big(g\Big(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}\Big) - g(\theta_0)\Big) - g'_{\theta_0}\Big(Z^* + Z_n\Big) \Big|.$$

Moreover, since Z_n is (uniformly) tight, there exists an M_δ such that $\mathbb{P}^n(\|Z_n\| > M_\delta) < \delta$. This means that:

$$\mathbb{P}^n\Big(\mathbb{P}^*(|W_n| > \epsilon | X^n) > \eta \Big) \leq \mathbb{P}^n\Big(\mathbb{P}^*(|W_n| > \epsilon | X^n) > \eta \text{ and } \|Z_n\| \leq M_\delta \Big) + \delta,$$

and we can focus on the first term to right of the inequality. Define $\Gamma_1(\delta) = \{a \in \mathbb{R}^p : \|a\| \leq M_\delta\}$. Since g is (Hadamard) directionally differentiable and $\Gamma_1(\delta)$ is compact, $\forall \epsilon > 0$, $\exists\ N_1(\epsilon, \delta)$ such that $\forall\ n > N_1(\epsilon, \delta)$.

$$\sup_{a \in \Gamma_1(\delta)} \Big| \sqrt{n}(g(\theta_0 + a/\sqrt{n}) - g(\theta_0)) - g'_{\theta_0}(a) \Big| < \frac{\epsilon}{2}.$$

This means that $\forall\ n > N_1(\epsilon, \delta)$:

$$\mathbb{P}^n\Big(\mathbb{P}^*(|W_n| > \epsilon | X^n) > \eta \text{ and } \|Z_n\| \leq M_\delta \Big)$$

is bounded above by

$$\mathbb{P}^n\Big(\mathbb{P}^*\Big(\Big| \sqrt{n}\Big(g\Big(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}\Big) - g\Big(\theta_0\Big) - g'_{\theta_0}\Big(Z^* + Z_n\Big) \Big| > \frac{\epsilon}{2} \Big| X^n \Big) > \eta \text{ and } \|Z_n\| \leq M_\delta \Big).$$

Trivially, there exists $M^*_{\eta/2}$ such that $\mathbb{P}^*(\|Z^*\| > M^*_{\eta/2}) < \eta/2$. Consequently, we can further bound the probability above by:

$$\mathbb{P}^n\Big(\mathbb{P}^*\Big(\Big|\sqrt{n}\Big(g\Big(\theta_0+\frac{Z^*}{\sqrt{n}}+\frac{Z_n}{\sqrt{n}}\Big)-g\Big(\theta_0\Big)-g'_{\theta_0}\Big(Z^*+Z_n\Big)\Big|>\frac{\epsilon}{2}\text{ and }||Z^*||\leq M^*_{\eta/2}\,\Big|\,X^n\Big)>\eta/2 \text{ and } \|Z_n\|\leq M_\delta\Big).$$

Define the set $\Gamma_2(\delta,\eta)=\{b\in\mathbb{R}^p:\ \|b\|\leq M_\delta+M^*_{\eta/2}\}$. Then, since g is (Hadamard) directionally differentiable and $\Gamma_2(\eta,\delta)$ is compact, $\forall\epsilon>0$, $\exists\ N_2(\epsilon,\delta,\eta)$ such that $\forall n>N_2(\epsilon,\delta,\eta)$

$$\sup_{b\in\Gamma_2(\delta,\eta)}\Big|\sqrt{n}(g(\theta_0+b/\sqrt{n})-g(\theta_0))-g'_{\theta_0}(b)\Big|\leq\frac{\epsilon}{2}.$$

This means that for $n>N_2(\epsilon,\delta,\eta)$:

$$\mathbb{P}^n\Big(\mathbb{P}^*\Big(\Big|\sqrt{n}\Big(g\Big(\theta_0+\frac{Z^*}{\sqrt{n}}+\frac{Z_n}{\sqrt{n}}\Big)-g\Big(\theta_0\Big)-g'_{\theta_0}\Big(Z^*+Z_n\Big)\Big|>\frac{\epsilon}{2}\text{ and }||Z^*||\leq M^*_{\eta/2}\,\Big|\,X^n\Big)>\eta/2 \text{ and } \|Z_n\|\leq M_\delta\Big)=0,$$

and, consequently, for any $\epsilon,\eta,\delta>0$ there is n sufficiently large such—in particular, $n>\max\{N_1(\epsilon,\delta),N_2(\epsilon,\delta,\eta)\}$—such that:

$$\mathbb{P}^n\Big(\mathbb{P}^*(|W_n|>\epsilon|X^n)>\eta\Big)\leq\delta.$$

To see that this implies the desired result, let

$$\psi^*_n\equiv\sqrt{n}\Big(g\Big(\theta_0+\frac{Z^*}{\sqrt{n}}+\frac{Z_n}{\sqrt{n}}\Big)-g\Big(\widehat{\theta}_n\Big)\Big)\quad\text{and}\quad\gamma^*_n\equiv g'_{\theta_0}\Big(Z^*+Z_n\Big)-g'_{\theta_0}\Big(Z_n\Big),$$

so that $W_n=\psi^*_n-\gamma^*_n$. Fix $\eta>0$ and note that

$$\begin{aligned}
&\sup_{f\in BL(1)}\Big|\mathbb{E}[f(\psi^*_n)|X^n]-\mathbb{E}[f(\gamma^*_n)|X^n]\Big|\\
&\quad\leq\sup_{f\in BL(1)}\mathbb{E}[\,|f(\psi^*_n)-f(\gamma^*_n)|\,|X^n]\\
&\quad\leq\mathbb{E}[\min\{|W_n|,2\}|X^n],\ \text{(since } f\in BL(1))\\
&\quad\leq\mathbb{E}[\min\{|W_n|,2\}\mathbb{I}_{\{|W_n|\leq\eta/2\}}|X^n]+\mathbb{E}[\min\{|W_n|,2\}\mathbb{I}_{\{|W_n|>\eta/2\}}|X^n]\\
&\quad\leq\eta/2+2\mathbb{P}^*(|W_n|>\eta/2|X^n).
\end{aligned}$$

This inequality implies that:

$$\begin{aligned}
\mathbb{P}^n\Big(\sup_{f\in BL(1)}\Big|\mathbb{E}[f(\psi^*_n)|X^n]-\mathbb{E}[f(\gamma^*_n)|X^n]\Big|>\eta\Big)&\leq\mathbb{P}^n\Big(\eta/2+2\mathbb{P}^*(|W_n|>\eta/2|X^n)>\eta\Big)\\
&=\mathbb{P}^n\Big(\mathbb{P}^*(|W_n|>\eta/2|X^n)>\eta/4\Big)
\end{aligned}$$

This means that for any $\eta,\delta>0$ there exists n large enough—specifically, we can choose $n>\max\{N_1(\eta/2,\delta),N_2(\eta/2,\delta,\eta/4)\}$ with N_1 and N_2 as defined before—such that:

$$\mathbb{P}^n\Big(\sup_{f\in BL(1)}\Big|\mathbb{E}[f(\psi^*_n)|X^n]-\mathbb{E}[f(\gamma^*_n)|X^n]\Big|>\eta\Big)<\delta.$$

Q.E.D.

PROOF OF THE COROLLARY TO THEOREM 1: The proof of the Corollary to Theorem 1 follows directly from Lemma 1 and Lemma 2. Remember that the goal is to show that:

$$\sup_{f\in BL(1)}\left|\mathbb{E}\left[f\left(\sqrt{n}\left(g(\theta_n^{P*})-g(\widehat{\theta}_n)\right)\right)-f\left(g'_{\theta_0}(Z+Z_n)-g'_{\theta_0}(Z_n)\right)\Big|X^n\right]\right|\xrightarrow{p}0,$$

where $Z\sim\mathcal{N}(\mathbf{0},\mathcal{I}^{-1}(\theta_0))$. The triangle inequality provides a natural upper bound for the probability above:

$$\begin{aligned}\text{(A.2)}\quad &\sup_{f\in BL(1)}\left|\mathbb{E}\left[f\left(\sqrt{n}\left(g(\theta_n^{P*})-g(\widehat{\theta}_n)\right)\right)-f\left(g'_{\theta_0}(Z+Z_n)-g'_{\theta_0}(Z_n)\right)\Big|X^n\right]\right|\\ &\leq\mathbb{E}\left[f\left(\sqrt{n}\left(g(\theta_n^{P*})-g(\widehat{\theta}_n)\right)\right)-f\left(\sqrt{n}\left(g\left(\theta_0+\frac{Z}{\sqrt{n}}+\frac{Z_n}{\sqrt{n}}\right)\right)-g\left(\widehat{\theta}_n\right)\right)\Big|X^n\right]\\ &+\mathbb{E}\left[f\left(\sqrt{n}\left(g\left(\theta_0+\frac{Z}{\sqrt{n}}+\frac{Z_n}{\sqrt{n}}\right)\right)-g\left(\widehat{\theta}_n\right)\right)-f\left(g'_{\theta_0}\left(Z+Z_n\right)-g'_{\theta_0}\left(Z_n\right)\right)\Big|X^n\right].\end{aligned}$$

Under Assumptions 1, 2 and 3, Lemma 1 applied to θ_n^{P*} implies that the term in the second line of (A.2) converges in probability to zero. Under Assumption 4, Lemma 2 applied to $Z\sim\mathcal{N}(\mathbf{0},\mathcal{I}^{-1}(\theta_0))$ implies that the term in the third line of (A.2) converges in probability to zero. The desired result then follows.

A.3. *Proof of Corollary 2*

We start by establishing a Lemma based on a high-level assumption implied by Assumption 5.

ASSUMPTION 6 The directional derivative g'_θ (at a point θ) is such that for all positive (M, ϵ, δ) there exists $\zeta(M, \epsilon, \delta) > 0$ and $N(M, \epsilon, \delta)$ for which:

$$\mathbb{P}_\theta\Big(\sup_{c\in\mathbb{R}} \mathbb{P}^Z_\theta\Big(c - \zeta(M,\epsilon,\delta) \leq g'_\theta(Z+Z_n) - g'_\theta(Z_n) \leq c + \zeta(M,\epsilon,\delta) \text{ and } ||Z|| \leq M \;\Big|\; X^n\Big) > \epsilon\Big) < \delta.$$

provided $n \geq N(M, \epsilon, \delta)$.

To see that Assumption 6 is implied by Assumption 5 simply note the following. Note that:

$$\mathbb{P}^Z\Big(c - \zeta(M,\epsilon,\delta) \leq g'_\theta(Z+Z_n) - g'_\theta(Z_n) \leq c + \zeta(M,\epsilon,\delta) \text{ and } ||Z|| \leq M \;\Big|\; X^n\Big),$$

is bounded above by

$$\mathbb{P}^Z\Big(c - \zeta(M,\epsilon,\delta) \leq g'_\theta(Z+Z_n) - g'_\theta(Z_n) \leq c + \zeta(M,\epsilon,\delta) \;\Big|\; X^n\Big),$$

which equals:

$$F_\theta(g'_\theta(Z_n) + c + \zeta(M,\epsilon,\delta)) - F_\theta(g'_\theta(Z_n) + c - \zeta(M,\epsilon,\delta)) \quad \leq \quad 2\zeta(M,\epsilon,\delta)k.$$

By choosing $\zeta(M, \epsilon, \delta)$ equal to $\epsilon/4k$, then

$$\mathbb{P}^Z\Big(c - \zeta(M,\epsilon,\delta) \leq g'_\theta(Z+Z_n) - g'_\theta(Z_n) \leq c + \zeta(M,\epsilon,\delta) \text{ and } ||Z|| \leq M \;\Big|\; X^n\Big) \leq \mathbb{P}^Z_\theta(\frac{\epsilon}{2} > \epsilon) = 0.$$

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 distributed as $\mathcal{N}(\boldsymbol{0}, \mathcal{I}^{-1}(\theta))$, where θ denote the parameter that generated the data. Let $Z_n \equiv \sqrt{n}(\widehat{\theta}_n - \theta)$. Suppose that*

$$\sup_{f\in BL(1)} \Big|\mathbb{E}\Big[f\Big(\sqrt{n}\Big(g\Big(\theta^*_n\Big) - g\Big(\widehat{\theta}_n\Big)\Big)\Big) \;\Big|\; X^n\Big] - \mathbb{E}\Big[f\Big(g'_{\theta_0}\Big(Z+Z_n\Big) - g'_{\theta_0}\Big(Z_n\Big)\Big)\Big|X^n\Big]\Big| \xrightarrow{p} 0.$$

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

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

Suppose that the distribution of $g'_\theta(Z+Z_n) - g'_\theta(Z_n)$ is continuous for every Z_n. Define $c_\alpha(Z_n)$ as:

$$\mathbb{P}^Z\left(g'_\theta(Z+Z_n) - g'_\theta(Z_n) \leq c_\alpha(Z_n) \mid X^n\right) = \alpha.$$

Under Assumption 6 for any $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $N(\epsilon, \delta)$ such that for $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\theta(c_{\alpha-\epsilon}(Z_n) \leq c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(Z_n)) \geq 1 - \delta.$$

PROOF: We start by deriving a convenient bound for the difference between the distribution of

$\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}))$ and the distribution of $g'_\theta(Z + Z_n) - g'_\theta(Z_n)$. Define the random variables:

$$W_n^* \equiv \sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)), \quad Y_n^* \equiv g'_\theta(Z + Z_n) - g'_\theta(Z_n).$$

Denote by P_W^n and P_Y^n the probabilities that each of these random variables induce over the real line. Let $c \in \mathbb{R}$ be some constant. Choose M_ϵ such that $\mathbb{P}^Z(||Z|| > M_\epsilon) \leq \epsilon/3$. By applying Lemma 5 in Appendix A.4 to the set $A = (-\infty, c)$ it follows that for any $\zeta > 0$:

$$\begin{aligned}
&|P_W^n((-\infty, c)|X^n) - P_Y^n((-\infty, c)|X^n)| \\
&\quad \leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \min\{(P_Y^n(A^\zeta \setminus A|X^n), P_Y^n((A^c)^\zeta \setminus A^c|X^n)\} \\
&\quad = \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \min\{P_Y^n(\, [c, c+\zeta]\, |X^n), P_Y^n(\, [c-\zeta, c]\, |X^n)\} \\
&\quad \leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \mid X^n\right) \\
&\quad = \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) \\
&\qquad\quad + \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \ \text{ and } ||Z|| \leq M_\epsilon | \, X^n\right) \\
&\qquad\quad + \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \ \text{ and } ||Z|| > M_\epsilon | \, X^n\right) \\
&\quad \leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) \\
&\qquad\quad + \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \ \text{ and } ||Z|| \leq M_\epsilon | \, X^n\right) \\
&\qquad\quad + \mathbb{P}^Z\left(\|Z\| > M_\epsilon\right) \\
&\qquad\quad \text{(since the random variable } Z \text{ is independent of } X^n\text{)}
\end{aligned}$$

That is:

$$\begin{aligned}
&|\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c \mid X^n\right)| \\
&\quad \leq \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))\, ,\, g'_\theta(Z + Z_n) - g'_\theta(Z_n); X^n) \\
&\qquad\quad + \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \text{ and } ||Z|| \leq M_\epsilon \mid X^n\right) \\
&\qquad\quad + \mathbb{P}^Z(||Z|| > M_\epsilon).
\end{aligned}$$

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

Note that for any $c \in \mathbb{R}$ the previous equation implies:

$$\begin{aligned}
&|\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c \mid X^n\right)| \\
&\quad \leq \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))\, ,\, g'_\theta(Z + Z_n) - g'_\theta(Z_n); X^n) \\
&\quad + \sup_{c \in \mathbb{R}} \mathbb{P}^Z\left(c - \zeta \leq g'_\theta(Z + Z_n) - g'_\theta(Z_n) \leq c + \zeta \text{ and } ||Z|| \leq M_\epsilon \mid X^n\right) \\
&\quad + \epsilon/3.
\end{aligned}$$

To simplify the notation, define the functions:

$$A_1(\zeta, X^n) \equiv \frac{1}{\zeta}\beta(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))\ ,\ g'_\theta(Z+Z_n) - g'_\theta(Z_n); X^n),$$
$$A_2(\zeta, X^n) \equiv \sup_{c\in\mathbb{R}} \mathbb{P}^Z\left(c - \zeta \le g'_\theta(Z+Z_n) - g'_\theta(Z_n) \le c + \zeta \text{ and } ||Z|| \le M_\epsilon \mid X^n\right).$$

Observe that if the data X^n were such that $A_1(\zeta, X^n) \le \epsilon/3$ and $A_2(\zeta, X^n) \le \epsilon/3$ then for any $c \in \mathbb{R}$:

$$\begin{aligned} &|\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c \mid X^n) - \mathbb{P}^Z\left(g'_\theta(Z+Z_n) - g'_\theta(Z_n) \le c \mid X^n\right)| \\ &\quad \le A_1(\zeta, X^n) + A_2(\zeta, X^n) + \epsilon/3 \\ &\quad < \epsilon. \end{aligned}$$

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

$$-\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c \mid X^n) - \mathbb{P}^Z\left(g'_\theta(Z+Z_n) - g'_\theta(Z_n) \le c \mid X^n\right) < \epsilon. \tag{A.3}$$

We now show that this inequality implies that:

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

whenever X^n is such that $A_1(\zeta, X^n) \le \epsilon/3$ and $A_2(\zeta, X^n) \le \epsilon/3$. To see this, evaluate equation (A.3) at $c_{\alpha+\epsilon}(Z_n)$. This implies that:

$$-\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c_{\alpha+\epsilon}(Z_n) \mid X^n) - (\alpha+\epsilon).$$

Consequently:

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

Since:

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

it follows that:

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

To obtain the other inequality, evaluate equation (A.3) at $c_{\alpha-\epsilon}(Z_n)$. This implies that:

$$\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c_{\alpha-\epsilon}(Z_n) \mid X^n) - (\alpha-\epsilon) < \epsilon.$$

Note that $c_{\alpha-\epsilon}(Z_n)$ is a lower bound of the set:

$$\{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c \mid X^n) \ge \alpha\}. \tag{A.4}$$

If this were not the case, there would exist c^* in the set above such that $c^* < c_{\alpha-\epsilon}(Z^n)$. As a consequence, the monotonicity of the c.d.f would then imply that:

$$\alpha \le \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c^* \mid X^n) \le \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \le c_{\alpha-\epsilon}(Z_n) \mid X^n) < \alpha,$$

which would imply that $\alpha < \alpha$; a contradiction. Therefore, $c_{\alpha-\epsilon}(Z_n)$ is indeed a lower bound for the set in (A.4) and, consequently:

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

This shows that whenever the data X^n is such that $A_1(\zeta, X^n) \leq \epsilon/3$ and $A_2(\zeta, X^n) \leq \epsilon/3$

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

To finish the proof, note that by Assumption 6 there exists $\zeta^* \equiv \zeta(M_\epsilon, \epsilon/3, \delta/2)$ and $N(M_\epsilon, \epsilon/3, \delta/2)$ that guarantees that if $n > N(M_\epsilon, \epsilon/3, \delta/2)$:

$$\mathbb{P}^n_\theta(A_2(\zeta^*, X^n) > \epsilon/3) < \delta/2.$$

Also, by the convergence assumption of this Lemma, there is $N(\zeta^*, \epsilon/3, \delta/2)$ such that for $n > N(\zeta^*, \epsilon/3\delta/2)$:

$$\mathbb{P}^n_\theta(A_1(\zeta^*, X^n) > \epsilon/3\ ) < \delta/2.$$

It follows that for $n > \max\{N(\zeta^*, \epsilon/3, \delta/2), N(M_\epsilon, \epsilon/3, \delta/2)\} \equiv N(\epsilon, \delta)$

$$\begin{aligned}
&\mathbb{P}_\theta(c_{\alpha-\epsilon}(Z^n) \leq c^*_\alpha(X^n) \leq c_{\alpha+\epsilon}(Z^n)) \\
&\qquad \geq \mathbb{P}_\theta(A_1(\zeta^*, X^n) < \epsilon/3 \text{ and } A_2(\zeta^*, X^n) < \epsilon/3) \\
&\qquad = 1 - \mathbb{P}_\theta(A_1(\zeta^*, X^n) > \epsilon/3 \text{ or } A_2(\zeta^*, X^n) > \epsilon/3) \\
&\qquad \geq 1 - \mathbb{P}_\theta(A_1(\zeta^*, X^n) > \epsilon/3) - \mathbb{P}_\theta(A_2(\zeta^*, X^n) > \epsilon/3) \\
&\qquad \geq 1 - \delta
\end{aligned}$$

Q.E.D.

Lemma 4 *Suppose that the Assumptions 1-5 hold. Fix $\alpha \in (0,1)$. Let $c_\alpha^B(X^n)$ and $c_\alpha^P(X^n)$ denote critical values satisfying:*

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

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

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \tag{A.5}$$

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta. \tag{A.6}$$

Proof: Let θ^* denote either θ_n^{P*} or θ_n^{B*}. Let $c_\alpha(Z^n)$ and $c_\alpha^*(X^n)$ be defined as in Lemma 3. Under Assumptions 1 to 5, the conditions for Lemma 3 are satisfied. It follows that for any $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $N(\epsilon, \delta)$ such for all $n > N(\epsilon, \delta)$:

$$\begin{aligned} &\mathbb{P}_\theta(c_{\alpha+\epsilon/2}(Z_n) < c_\alpha^*(X^n)) \leq \delta/2 \\ &\mathbb{P}_\theta(c_\alpha^*(X^n) < c_{\alpha-\epsilon/2}(Z_n)) \leq \delta/2 \end{aligned}$$

Therefore:

$$\begin{aligned} &\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(Z_n)) \\ &\quad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(Z_n) \text{ and } c_{\alpha+\epsilon/2}(Z_n) \geq c_\alpha^*(X^n)) \\ &\qquad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(Z_n) \text{ and } c_{\alpha+\epsilon/2}(Z_n) < c_\alpha^*(X^n)) \\ &\quad \text{(by the additivity of probability measures)} \\ &\quad \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\theta(c_{\alpha+\epsilon/2}(Z_n) < c_\alpha^*(X^n)) \\ &\quad \text{(by the monotonicity of probability measures)} \\ &\quad \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \delta/2. \end{aligned} \tag{A.7}$$

Also, we have that:

$$\begin{aligned} &\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(Z_n)) \\ &\quad \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(Z_n) \text{ and } c_\alpha^*(X_n) \geq c_{\alpha-\epsilon/2}(Z_n)) \\ &\quad \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ and } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(Z_n)) \\ &\quad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\theta(c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(Z_n)) \\ &\qquad - \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ or } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(Z_n)) \\ &\qquad \text{(using } P(A \cap B) = P(A) + P(B) - P(A \cup B)) \\ &\quad \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - (1 - \mathbb{P}_\theta(c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(Z_n))) \\ &\qquad \text{(since } \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ or } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(Z_n)) \leq 1) \\ &\quad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - \mathbb{P}_\theta(c_\alpha^*(X^n) < c_{\alpha-\epsilon/2}(Z_n)) \\ &\quad \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - \delta/2. \end{aligned} \tag{A.8}$$

Replacing θ_n^* by θ_n^{B*} in (A.8) and θ_n^* by θ_n^{P*} and α by $\alpha - \epsilon$ in (A.7) implies that for $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha-\epsilon/2}(Z_n)) \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_\alpha^{B*}(X^n)) - \delta/2$$
$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha-\epsilon/2}(Z_n)) \le \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta/2.$$

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

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

This establishes equation (A.5). Replacing θ_n^* by θ_n^{B*} in (A.7) and replacing θ_n^* by θ_n^{P*}, α by $\alpha + \epsilon$ (A.8) implies that for $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha+\epsilon/2}(Z_n)) \le \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_\alpha^{B*}(X^n)) + \delta/2$$
$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha+\epsilon/2}(Z_n)) \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta/2$$

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

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_\alpha^{B*}(X^n)) \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta,$$

which establishes equation (A.6).

Q.E.D.

Proof of Corollary 2: Define, for any $0<\beta<1$, the critical values $c^B_\beta(X^n)$ and $c^P_\beta(X^n)$ by the following:

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

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

$$\begin{aligned} q^B_\beta(X^n) &= g(\widehat{\theta}_n)+c^{B*}_\beta(X^n)/\sqrt{n}\\ q^P_\beta(X^n) &= g(\widehat{\theta}_n)+c^{P*}_\beta(X^n)/\sqrt{n}.\end{aligned}$$

This implies that:

$$\begin{aligned} CS^B(1-\alpha) &= \left[g(\widehat{\theta}_n)+c^{B*}_{\alpha/2}(X^n)/\sqrt{n}\;,\;g(\widehat{\theta}_n)+c^{B*}_{1-\alpha/2}(X^n)\right]\\ CS^P(1-\alpha-\epsilon) &= \left[g(\widehat{\theta}_n)+c^{P*}_{\alpha/2+\epsilon/2}(X^n)/\sqrt{n}\;,\;g(\widehat{\theta}_n)+c^{P*}_{1-\alpha/2-\epsilon/2}(X^n)\right].\end{aligned}$$

Under Assumptions 1 to 5 we can apply the previous lemma. This implies that for $n>N(\epsilon,\delta)$

$$\begin{aligned} \mathbb{P}_\theta\left(g(\theta)\in CS^B_n\right) &= \mathbb{P}_\theta\left(g(\theta)\in\left[g(\widehat{\theta}_n)+c^{B*}_{\alpha/2}(X^n)/\sqrt{n}\;,\;g(\widehat{\theta}_n)+c^{B}_{1-\alpha/2}/\sqrt{n}\right]\right)\\ &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n)-g(\theta))\le -c^{B*}_{\alpha/2}(X^n))\\ &\qquad -\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n)-g(\theta))\le -c^{B*}_{1-\alpha/2}(X^n))\\ &\ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n)-g(\theta))\le -c^{P*}_{\alpha/2+\epsilon/2}(X^n))\\ &\qquad -\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n)-g(\theta))\le -c^{P*}_{1-\alpha/2-\epsilon/2}(X^n))-\delta\\ &\qquad \text{(Replacing } \alpha \text{ by } \alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (A.6) and}\\ &\qquad \text{replacing } \alpha \text{ by } 1-\alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (A.5))}\\ &= \mathbb{P}_\theta\left(g(\theta)\in CS^P_n(1-\alpha-\epsilon)\right)-\delta\end{aligned}$$

This implies that for every $\epsilon>0$:

$$1-\alpha-d_\alpha\ge\limsup_{n\to\infty}\mathbb{P}_\theta\left(g(\theta)\in CS^B_n\right)\ge\limsup_{n\to\infty}\mathbb{P}_\theta\left(g(\theta)\in CS^P_n(1-\alpha-\epsilon)\right),$$

which implies that

$$1-\alpha-\epsilon-(d_\alpha-\epsilon)\ge\limsup_{n\to\infty}\mathbb{P}_\theta\left(g(\theta)\in CS^P_n(1-\alpha-\epsilon)\right).$$

This implies that if the bootstrap fails at θ by at least d_α% given the nominal confidence level $(1-\alpha)$%, then the confidence set based on the quantiles of the posterior will fail at θ—by at least $(d_\alpha-\epsilon)$%—given the nominal confidence level $(1-\alpha-\epsilon)$.

A.4. *Additional Lemmata*

Lemma 5 (Dudley (2002), p. 395) *Let W_n^*, Y_n^* be random variables dependent on the data $X^n = (X_1, X_2, \ldots X_n)$ 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,*

$$
\begin{aligned}
|P_W^n(A|X^n) - P_Y^n(A|X^n)| \leq &\frac{1}{\delta}\Big|\mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n]\Big| \\
&+ \min\{P_Y^n(A^\delta \backslash A|X^n),\ P_Y^n((A^c)^\delta \backslash A^c|X^n)\}
\end{aligned}
$$

Proof: First observe that:

$$
P_W^n(A|X^n) - P_Y^n(A|X^n) \leq P_W^n(A|X^n) - P_Y^n(A^\delta|X^n) + P_Y^n(A^\delta|X^n) - P_Y^n(A|X^n)
$$

Define $f(x) := \max(0, 1 - \|x - A\|/\delta)$. Then, $\delta f \in \mathrm{BL}(1)$ and:

$$
\begin{aligned}
P_W^n(A|X^n) &= \int_A dP_W^n|X^n \\
&\leq \int f dP_W^n|X^n \\
&\quad (\text{ since } f \text{ is nonnegative and } f(x) = 1 \text{ over } A\ ) \\
&= \int_A dP_Y^n|X^n + \frac{1}{\delta}\left(\int_A \delta f dP_W^n|X^n - \int_A \delta f dP_Y^n|X^n\right) \\
&\leq \int f dP_Y^n|X^n + \frac{1}{\delta}\sup_{f \in BL(1)}\Big|\mathbb{E}[f(W_n^*) \mid X^n] - \mathbb{E}[f(Y_n^*) \mid X^n]\Big| \\
&= \int_{A^\delta} f dP_Y^n|X^n + \frac{1}{\delta}\sup_{f \in BL(1)}\Big|\mathbb{E}[f(W_n^*) \mid X^n] - \mathbb{E}[f(Y_n^*) \mid X^n]\Big| \\
&\leq P_Y^n(A^\delta|X^n) + \frac{1}{\delta}\sup_{f \in BL(1)}\Big|\mathbb{E}[f(W_n^*) \mid X^n] - \mathbb{E}[f(Y_n^*) \mid X^n]\Big|
\end{aligned}
$$

It follows that:

$$
P_W^n(A|X^n) - P_Y^n(A|X^n) \leq \frac{1}{\delta}\Big|\mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n]\Big| + (P_Y^n(A^\delta|X^n) - P_Y^n(A|X^n))
$$

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

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

which implies that:

$$
P_W^n(A|X^n) - P_Y^n(A|X^n) \geq -\frac{1}{\delta}\Big|\mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n]\Big| - (P_Y^n((A^c)^\delta|X^n) - P_Y^n(A^c|X^n))
$$

The desired result follows. *Q.E.D.*

ONLINE APPENDIX B.

Toru Kitagawa[1], José-Luis Montiel-Olea[2] and Jonathan Payne[3]

1. MAX$\{\theta_1, \theta_2\}$

In this Appendix we provide another illustration to Corollary 2. Let $(X_1, \dots X_n)$ 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 Lipschitz continuous everywhere and differentiable everywhere except at $\theta_1 = \theta_2$ where it has directional derivative $g'_\theta(h) = \max\{h_1, h_2\}$. Thus, Assumptions 1 and 4 are satisfied.

The maximum likelihood estimator is given by $\widehat{\theta}_{\text{ML}} = (1/n)\sum_{i=1}^n X_i$ and so $\sqrt{n}(\widehat{\theta}_{\text{ML}} - \theta) \sim Z \sim \mathcal{N}_2(0, \Sigma)$. 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:

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

[1]University College London, Department of Economics. E-mail: t.kitagawa@ucl.ac.uk.
[2]New York University, Department of Economics. E-mail: montiel.olea@nyu.edu.
[3]New York University, Department of Economics. E-mail: jep459@nyu.edu.

which implies that Assumption 3 holds.

Finally, we show that the cdf $F_\theta(y|Z_n)$ of the random variable $Y = g'_\theta(Z + Z_n) = \max\{Z_1 + Z_{n,1}, Z_2 + Z_{n,2}\}$ satisfies Assumption 5. Based on the results of Nadarajah and Kotz (2008), the density $f_\theta(y|Z_n)$ is given by:

$$
\begin{aligned}
& \frac{1}{\sigma_1}\phi\left(\frac{Z_{n,1}-y}{\sigma_1}\right)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\rho(Z_{n,1}-y)}{\sigma_1}+\frac{y-Z_{n,2}}{\sigma_2}\right)\right) \\
+ \quad & \frac{1}{\sigma_2}\phi\left(\frac{Z_{n,2}-y}{\sigma_2}\right)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\rho(Z_{n,2}-y)}{\sigma_2}+\frac{y-Z_{n,1}}{\sigma_1}\right)\right),
\end{aligned}
$$

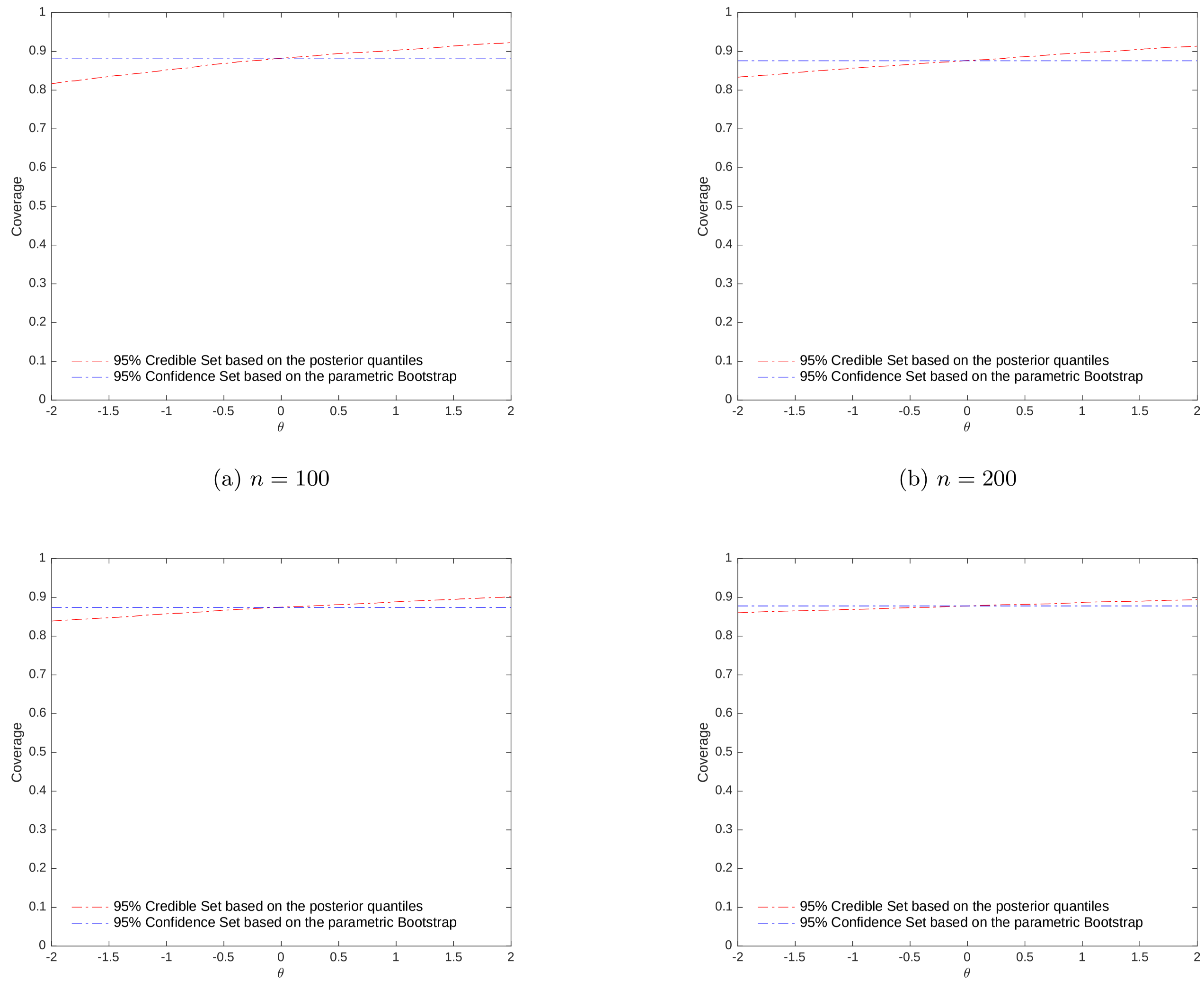
where $\rho = \sigma_{12}/\sigma_1\sigma_2$ and ϕ, Φ are the p.d.f. and the c.d.f. of a standard normal. It follows that:

$$
f(y|Z_n) \leq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right).
$$

and so, by an analogous argument to the absolute value case, $F(y|Z_n)$ is Lipschitz continuous with Lipschitz constant independent of Z_n and so Assumption 5 holds.

Graphical illustration of coverage failure: Corollary 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 (2015) 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 (2015) 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:

$$CS_n^H(1-\alpha) = \left[g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{1-\alpha/2},\ g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{\alpha/2}\right] \tag{1.1}$$

where c^*_β denote the β-quantile of y_n^*. Hong and Li (2015) have recently established the pointwise validity of the confidence interval above.

Whenever (1.1) 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 (1.1), assuming that the Hong and Li (2015) procedure is implemented using the posterior:

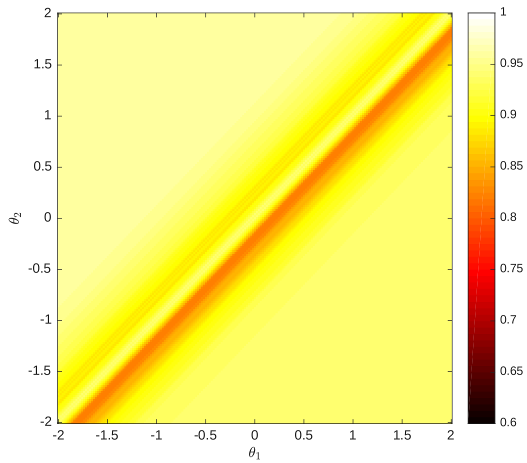
$$\theta_n^{P*}|X^n \sim \mathcal{N}_2\Big(\frac{n}{n+1}\widehat{\theta}_n\ ,\ \frac{1}{n+1}\mathbb{I}_2\Big).$$

The following figure shows that at least in this example fixing coverage comes at the expense of distorting Bayesian credibility.[1]

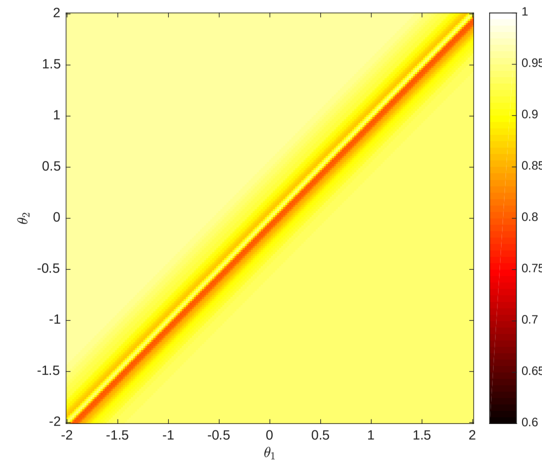
[1]The Bayesian credibility of $CS_n^H(1-\alpha)$ is given by:

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

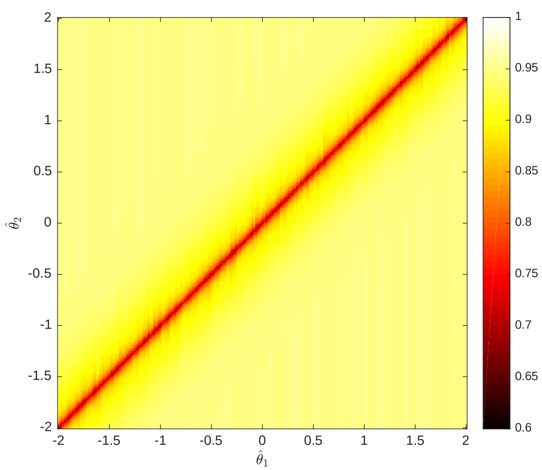
Figure 2: Coverage probability and Credibility of 95% Confidence Sets based on y_n^*



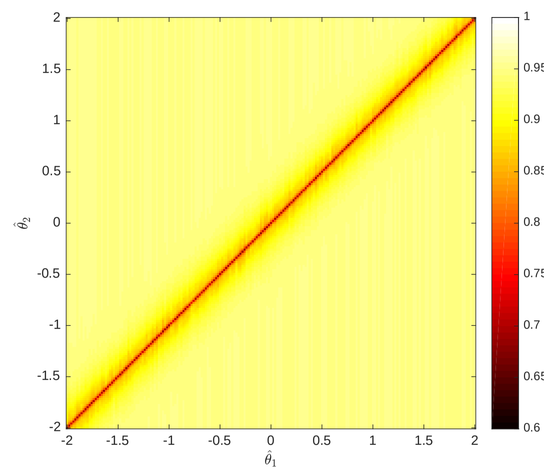
(a) Coverage Probability ($n = 100$)



(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

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

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 (2015): "The numerical delta-method," *Working Paper, Stanford University.*

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.