PERUVIAN ECONOMIC ASSOCIATION

Posterior Distribution of Nondifferentiable Functions

Toru Kitagawa

José Luis Montiel Olea

Jonathan Payne

Amilcar Velez

Working Paper No. 147, May 2019

Posterior Distribution of Nondifferentiable Functions*

Toru Kitagawa[†], José Luis Montiel Olea[‡],
Jonathan Payne[§] and Amilcar Velez[¶]

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

It is shown that whenever g is locally 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 credible intervals for a nondifferentiable parameter $g(\theta)$ cannot be presumed to be approximately valid confidence intervals (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 in either a parametric or semiparametric model. We focus

*We would like to thank Gary Chamberlain, Tim Cogley, Han Hong, Hiroaki Kaido, Ulrich Müller, Mikkel Plagborg-Möller, Andres Santos, Quang Vuong, and two anonymous referees for detailed comments and suggestions on an 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) and from the ERC through the ERC starting grant (grant number EPP-715940) is gratefully acknowledged. This draft: April 3, 2019.

[†]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`

[¶]Central Bank of Peru, Macro Modelling Division. E-mail: `velezs.ae@pucp.pe`

on transformations $g(\theta)$ that are locally Lipschitz continuous but possibly nondifferentiable. Some stylized examples are

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

More generally, our framework covers value functions of stochastic mathematical programs (Shapiro (1991)), which appear in the study of the bounds of the identified set in partially identified models.[1]

The potential nondifferentiability of $g(\cdot)$ challenges the frequentist inference. For example, different forms of the bootstrap lose their consistency whenever differentiability is compromised; see Dümbgen (1993), Beran (1997), Andrews (2000), Hong and Li (2018), and Fang and Santos (2019). To our knowledge, the literature has not yet explored how the Bayesian posterior of $g(\theta)$ relates to either the sampling or the bootstrap distribution of available plug-in estimators when g is allowed to be nondifferentible.

This paper studies these relations in large samples. The main assumptions are that: (i) there is an estimator for θ, denoted $\widehat{\theta}_n$, which is $\sqrt{n}$-asymptotically distributed according to some random vector Z (not necessarily Gaussian), (ii) the bootstrap consistently estimates the asymptotic distribution of $\widehat{\theta}_n$ and (iii) the Bayesian posterior distribution of θ coincides with the asymptotic distribution of $\widehat{\theta}_n$; i.e., the Bernstein–von Mises Theorem holds for θ.[2]

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, approximately the same information as the posterior distribution for $g(\theta)$.[3] Indisputably, these asymptotic relations are straightforward to deduce for (fully or directionally) differentiable functions. However, our main result shows that the asymptotic equivalence between the bootstrap and posterior distributions holds more broadly, highlighting that such a relation is better understood as a consequence of the continuous mapping theorem, as opposed to differentiability and the delta-method.

[1]For example, treatment effect bounds (Manski (1990), Balke and Pearl (1997)); bounds in auction models (Haile and Tamer (2003)); bounds for impulse-response functions (Giacomini and Kitagawa (2018), Gafarov, Meier, and Montiel Olea (2018)) and forecast-error variance decompositions (Faust (1998)) in Structural Vector Autoregressions. Other examples of value functions of stochastic programs that arise in different applications in economics and statistics are the welfare level attained by an optimal treatment assignment rule in the treatment choice problem (Manski (2004)) and the eigenvalues of a random symmetric matrix (Eaton and Tyler (1991)).

[2]See, for example, DasGupta (2008), p. 291 for a Bernstein–von Misses theorem for regular parametric models where Z is Gaussian; Ghosal, Ghosh, and Samanta (1995), p. 2147–2150 for a Bernstein–von Mises theorem for a class of parametric models whose likelihood ratio process is not Locally Asymptotically Normal; and Castillo and Rousseau (2015), p. 2357 for a Bernstein–von Mises theorem for semiparametric models where an efficiency theory at rate $\sqrt{n}$ is available.

[3]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).

The distinction between the local Lipschitz property and directional differentiability emphasized in our main result is not just a technical refinement. We believe that such a distinction is practically useful, for example, when conducting Bayesian estimation and inference of the bounds of the identified set in partially identified models, as recently suggested by Kline and Tamer (2016) and Giacomini and Kitagawa (2018). The bounds of the identified set are typically value functions of stochastic mathematical programs for which standard constraint qualifications suffice to verify the local Lipschitz property.[4] In contrast, directional differentiability requires additional conditions, which can be quite difficult to verify even in specific applications.[5] The local Lipschitz property allows us to relate robust Bayes procedures to bootstrap based approaches for estimation/inference.

Implications: The main results of this paper provide two useful and general insights. First, Bayesians can interpret bootstrap-based estimation/inference for $g(\theta)$ as approximately Bayesian in a large sample. For example, under regularity conditions ensuring convergence in distribution to imply convergence in mean, an estimator for $g(\theta)$ built upon the bootstrap distribution of $g(\widehat{\theta}_n)$ (e.g., the mean of the bootstrap draws) can be interpreted as an approximately Bayes estimator for $g(\theta)$ (e.g., the posterior mean estimator). Hence, decision-theoretic optimality of the Bayes estimator can be attached to the bootstrap-based estimator for $g(\theta)$ in large samples irrespective of $g(\theta)$ being differentiable or not. This means that Bayesians can use bootstrap draws to conduct approximate posterior estimation/inference for $g(\theta)$, if computing $\widehat{\theta}_n$ is simpler than Markov Chain Monte Carlo (MCMC) sampling.

Second, we show that whenever nondifferentiability causes the bootstrap confidence interval to cover $g(\theta)$ less often than desired—which is known to happen even under mild departures from differentiability—a credible interval based on the quantiles of the posterior will have distorted frequentist coverage as well. In the case where $g(\cdot)$ only has directional derivatives, as in the pioneering work of Hirano and Porter (2012), the unfortunate frequentist properties of credible intervals can be attributed to the fact that the posterior distribution of $g(\theta)$ does not coincide with the asymptotic distribution of $g(\widehat{\theta}_n)$.[6]

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.

[4] See Theorem 4.2 in Fiacco and Ishizuka (1990) and Proposition 6 in Morand, Reffett, and Tarafdar (2015)

[5] We discuss this point in the framework of set-identified Structural Vector Autoregressions in Section 2 and in the Appendix.

[6] Woutersen and Ham (2016) refer to the confidence interval for $g(\theta)$ based on draws from the bootstrap distribution of $g(\widehat{\theta}_n)$ as the ADR bootstrap and credit Runkle (1987). They show that a projection confidence set for $g(\theta)$ based on draws from the bootstrap distribution of $\widehat{\theta}_n$ will have correct frequentist coverage even if g is not differentiable. The assumptions in our paper imply that the Bayesian credible set for $g(\theta)$ formed in this way attains the frequentist coverage for $g(\theta)$ in large samples. However, the posterior probability on the projection credible set can be strictly larger than the nominal credibility attached to the credible set of θ.

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 size n from the model $f(X^n \mid \eta)$, where η is a possibly infinite dimensional parameter taking values in some space S. We assume there is a *finite-dimensional* parameter of interest, $\theta : S \to \Theta \subseteq \mathbb{R}^p$, and some estimator $\widehat{\theta}_n$ of θ. Let θ_0 denote the true parameter—that is, $\theta_0 \equiv \theta(\eta_0)$ with data generated according to $f(X^n|\eta_0)$. Consider the following assumptions:

Assumption 1. The function $g : \mathbb{R}^p \to \mathbb{R}$ is locally Lipschitz at (or near) θ_0. That is, there exists a neighborhood V_0 of θ_0 and a constant $c_0 > 0$ such that

$$|g(x) - g(y)| \leq c_0||x - y|| \quad \forall\, x, y \in V_0.$$

See Clarke (1990), Chapter 1, p. 9 for a textbook reference.

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 a neighborhood of θ_0. Thus, the functions considered in this paper allow only for mild departures from differentiability near θ_0.[7] We have made local Lipschitz continuity our starting point—as opposed to some form of directional differentiability—to emphasize that the asymptotic relation between Bootstrap and Bayes inference does not hinge on delta-method considerations. Later, we will also present economically relevant examples where the local Lipschitz property is easier to verify than directional differentiability.

Assumption 2. The sequence $Z_n \equiv \sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z$.

Despite being high-level, there are well-known conditions for parametric or semiparametric models under which Assumption 2 obtains (see, for example, Newey and McFadden (1994) p. 2146). The convergence at rate $\sqrt{n}$ is used for notational simplicity, but it is not relevant for deriving our main results.[8]

The asymptotic distribution of Z_n is typically normal, but our main theorems do not exploit this feature (and thus, we have decided to leave the distribution of Z unspecified).

[7] 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.

[8] An additional motivation for using $\sqrt{n}$ is that the Bernstein –von Mises theorem for θ, which will be invoked in Assumption 3, is usually verified at this rate. Examples of the Bernstein–von Mises theorem at rates different to $\sqrt{n}$ appear on Bochkina and Green (2014)).

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

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

Let ϕ_n^* and ψ_n^* be random vectors 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

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

The random vectors 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 $n \to \infty$.[9]

Let P denote some prior for θ and 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 $\widehat{\theta}_n$ in a sample of size n.

Remark 1. In a parametric model there are different ways of bootstrapping the distribution of $\widehat{\theta}_n$. One possibility is a *parametric bootstrap*, which consists in generating draws $(x_1, \ldots x_n)$ 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 *multinomial bootstrap*, which generates draws $(x_1, \ldots x_n)$ from its empirical distribution. Different options are also available in semiparametric models. We do not take a stand on the specific bootstrap procedure used by the researcher as long as it is consistent.

The following assumption restricts the prior P for θ and the bootstrap procedure for $\widehat{\theta}_n$:

Assumption 3. The centered and scaled random variables

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

converge (in the bounded Lipschitz distance in probability) to the asymptotic distribution of $\widehat{\theta}_n$, denoted Z, which is independent of the data. That is,

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

[9] For a more detailed treatment of the bounded Lipschitz distance over probability measures see the 'β' metric defined in p. 394 of Dudley (2002). It is well-known that convergence in distribution is equivalent to convergence in the bounded Lipschitz distance; for example, see Lemma 2.2 in Van der Vaart (2000).

Sufficient conditions for Assumption 3 to hold are the consistency of the bootstrap 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 (see DasGupta (2008) for parametric versions and Castillo and Rousseau (2015) for semiparametric ones).[10,11]

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 distance in probability.

Proof. See Appendix A.1. □

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

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\hat{\theta}_n + Z_n^{P*}/\sqrt{n}) - g(\hat{\theta}_n)),$$

$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\hat{\theta}_n + Z_n^{B*}/\sqrt{n}) - g(\hat{\theta}_n)).$$

Since Z_n^{P*} and Z_n^{B*} both converge to a common limit and $\widehat{\theta}_n$ is asymptotically close to θ_0, we can apply an argument analogous to the one used in the proof of the (Lipschitz) continuous mapping theorem to get the desired result, but focusing on a neighborhood around θ_0 (where we have Lipschitz continuity).[12] A crucial step is to show that Z_n^{P*} and

[10]The Bernstein–von Mises Theorem is oftentimes stated in terms of almost-sure convergence of the posterior to a Gaussian distribution (DasGupta (2008) p. 291) or possibly to a non-Gaussian limit (Ghosal et al. (1995)) in terms of the total variation distance. This mode of convergence (total variation metric) implies convergence in the bounded Lipschitz distance in probability. In this paper, all the results concerning the asymptotic behavior of the posterior are presented in terms of the bounded Lipschitz distance. This facilitates comparisons with the bootstrap.

[11]Z_n^{*P} and Z_n^{*B} could be defined as general conditional distributions that depend on X^n that have a limit Z. We focus on bootstrap and posterior draws because we are specifically interested in their behavior and also because there are well known conditions in parametric and semiparametric models to verify our assumption.

[12] To the best of our knowledge Theorem 1 does not follow directly from the existing literature. First, we are assuming the g is locally Lipschitz at θ_0. Second, even if we were willing to assume that g is globally, rather than locally, Lipschitz, the continuous mapping theorem in Proposition 10.7 in Kosorok (2008) does not imply our result. The reason is that we are interested in the limit

Z_n^{B*} converge unconditionally and, therefore, are tight; see Lemma 1. This means that the asymptotic relation between the bootstrap and Bayes distributions is a consequence of a (locally Lipschitz) continuous mapping theorem, and not of the delta-method.

Application to Set-Identified Structural VARs: One illustration of the usefulness of Theorem 1 is robust Bayes analysis of set-identified structural vector autoregression (VARs). Consider an n-dimensional structural VAR with p lags; i.i.d. structural innovations—denoted ε_t—distributed according to an independent multivariate normal; and unknown $n \times n$ structural matrix B:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \;\; \mathbb{E}[\varepsilon_t] = 0_{n\times 1}, \;\; \mathbb{E}[\varepsilon_t \varepsilon_t'] \equiv \mathbb{I}_n. \tag{2.1}$$

Define the (k,i,j)-*coefficient of the structural impulse-response function* to be the scalar parameter

$$\lambda_{k,i,j}(A,B) \equiv e_i' C_k(A) B e_j,$$

where e_i and e_j denote the i-th and j-th column of the identity matrix $\mathbb{I}_n$ and $C_k(A)$ are the reduced-form moving average coefficients. The structural parameters are set-identified by sign restrictions

$$S(\theta) B e_j \geq 0,$$

where $S(\theta)$ is an $m \times n$ matrix whose entries are allowed to depend on the *reduced-form parameters* of the structural vector autoregression: $\theta \equiv (\text{vec}(A)', \text{vech}(\Sigma)')'$, $A \equiv (A_1, A_2, \ldots, A_p), \Sigma \equiv BB'$.

Gafarov et al. (2018) show that, given reduced form parameter θ, the upper bound of the identified set for $\lambda_{k,i,j}$ is given by the solution of the program

$$g(\theta) \equiv \max_{x \in \mathbb{R}^n} e_i' C_k(A) x, \;\; \text{s.t. } x'\Sigma^{-1} x = 1, \; S(\theta) x \geq 0. \tag{2.2}$$

Giacomini and Kitagawa (2018) have recently suggested estimating the impulse-response identified set by reporting the posterior mean of $g(\theta)$ (starting from a prior over θ), and

---

of the transformation

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

which depends on n. Unfortunately, a direct application of Kosorok's theorem does not imply that the distribution of the random variable above is close to

$$\sqrt{n}(g(\widehat{\theta}_n + Z/n) - g(\widehat{\theta}_n)),$$

which would be sufficient to establish our theorem, as both Z_n^{*P} and Z_n^{*B} converge to Z. Kosorok's theorem only allows us to show that both

$$g(\sqrt{n}(\widehat{\theta}_n^{*P} - \widehat{\theta}_n)), g(\sqrt{n}(\widehat{\theta}_n^{*B} - \widehat{\theta}_n))$$

converge in bounded Lipschitz distance to $g(Z)$.

have shown that the posterior distribution of the upper and lower bounds can be used to construct a robust credible set for $\lambda_{k,i,j}$.

In this context, Theorem 1 is relevant for a number of reasons. Firstly, standard results in nonlinear optimization imply that a sufficient condition for the value function in (2.2) to be locally lipschitz at θ is for the Mangasarian-Fromowitz constraint qualification to hold at an optimal solution (Proposition 6 in Morand et al. (2015)).[13] Theorem 1 thus implies that for a large class of priors on θ, the bootstrap distribution of the plug-in estimator is asymptotically equivalent to the posterior distribution of $g(\theta)$.

Moreover, under regularity conditions that ensure convergence in distribution implies convergence in mean, an estimator for $g(\theta)$ built upon the bootstrap distribution of $g(\widehat{\theta}_n)$ (e.g., the mean of the bootstrap draws) can be interpreted as an approximately Bayes estimator for $g(\theta)$ (e.g., the posterior mean estimator). Hence, decision-theoretic optimality of the Bayes estimator can be attached to the bootstrap-based estimator for $g(\theta)$ in large samples irrespective of $g(\theta)$ being differentiable or not. Consequently, if one is concerned with point estimation of the lower or upper bound of the impulse response identified set, then Theorem 1 suggests that the posterior mean estimator considered in Giacomini and Kitagawa (2018) can be well replicated by the 'bootstrap' mean estimator.

Although the asymptotic equivalence between the posterior and the bootstrap distributions can be shown under full or directional differentiability, verifying such properties for value functions is known to be complicated. For example, Morand et al. (2015) argue that 'because the MFCQ is not sufficient to guarantee the uniqueness of KKT multipliers, it is very difficult to obtain directional derivatives and sharp characterizations of the generalized gradient'.[14]

Finally, it is worth mentioning that the arguments above carry over to the more general framework concerning inference about the value function of a nonlinear program.

Failure of Bootstrap/Bayes 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 intervals and the coverage of Bayesian credible intervals based on the quantiles of the posterior.

We start by assuming that a nominal $100(1-\alpha)\%$ bootstrap confidence interval fails to cover $g(\theta)$ at a point of nondifferentiability. Then, we show that a $100(1-\alpha-\epsilon)\%$ credible interval based on the quantiles of the posterior distribution of $g(\theta)$ will also fail to cover

[13] In addition to the choice set being nonempty and uniformly compact at θ. We verify all these conditions in Lemma 6 of the Appendix.

[14] Indeed, standard theorems concerning the directional differentiability of the value function (Theorem 4.2 in Fiacco and Ishizuka (1990)) use the Mangasarian-Fromowitz constraint qualification to provide bounds on the directional derivatives. Establishing directional differentiability, however, requires verifying additional properties about the set of Lagrange multipliers or making additional assumptions on the sign restrictions under consideration (Gafarov et al. (2018)).

$g(\theta)$ for any $\epsilon > 0$.[15]

Set-up for Theorem 2: Let $q_\alpha^B(X^n)$ be defined as:

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

The quantile based on the posterior distribution $q_\alpha^P(X^n)$ is defined analogously. A nominal $100(1-\alpha)\%$ two-sided confidence interval for $g(\theta)$ based on the bootstrap distribution $g(\theta_n^{B*})$ can be defined as follows

$$CS_n^B(1-\alpha) \equiv \left[q_{\alpha/2}^B(X^n)\ ,\ q_{1-\alpha/2}^B(X^n)\right].$$

This is a typical confidence interval based on the percentile method of Efron, p. 327 in Van der Vaart (2000).[16]

Definition. We say that the nominal $100(1-\alpha)\%$ bootstrap confidence interval fails to cover the parameter $g(\theta)$ at θ by at least $100 d_\alpha\%$ $(0 < d_\alpha < 1-\alpha)$ if

$$\limsup_{n\to\infty} \mathbb{P}_\theta \left(g(\theta) \in CS_n^B(1-\alpha)\right) \leq 1 - \alpha - d_\alpha, \tag{2.3}$$

where $\mathbb{P}_\theta$ refers to the distribution of $(X_1, X_2, \ldots)$ under the parameter θ.

The next theorem shows the coverage probability of the Bayesian credible interval for $g(\theta)$ in relation to the coverage probability of its bootstrap confidence interval.

Theorem 2. *Suppose that the nominal $100(1-\alpha)\%$ bootstrap confidence interval fails to cover $g(\theta)$ at θ by at least $100 d_\alpha\%$. Suppose in addition that for any $0 < \epsilon < \alpha$*

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

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. Then for any $0 < \epsilon < \alpha$:

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

Thus, the nominal $100(1-\alpha-\epsilon)\%$ credible interval based on the quantiles of the posterior fails to cover $g(\theta)$ at θ by at least $100(d_\alpha - \epsilon)\%$.

[15] The adjustment factor ϵ is introduced because the quantiles of both the bootstrap and the posterior for nondifferentiable functions might remain random even in large samples.

[16] We focus on the percentile method rather than other bootstrap confidence intervals, such as the 'root method', for two reasons. First, to the best of our knowledge, there are no theoretical results showing that alternative forms of bootstrap confidence intervals can outperform the percentile method when conducting inference on nondifferentiable functions. Second, the Bayesian analogues of some of these procedures need not have correct Bayesian credibility.

Proof. See Appendix A.2. □

This result is not a direct corollary of Theorem 1 because convergence in distribution does not guarantee that the quantiles of the bootstrap distribution of $g(\widehat{\theta}_n)$ are close to the quantiles of the posterior of $g(\theta)$. Theorem 2 takes the closeness of the quantiles as given and establishes the frequentist coverage property of the Bayes credible interval based on the quantiles of the posterior.[17]

Remark 2. The desired closeness of quantiles can be established under a few more regularity assumptions. In particular, in the online Appendix B.2, we establish the closeness of bootstrap/posterior quantiles as required by Theorem 2 under the assumption that g is directionally differentiable.[18] That is, we assume there is 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 $t_n \to 0$:[19]

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

Proposition 1 in Dümbgen (1993) and equation A.41 in Theorem A.1 in Fang and Santos (2019) imply that, under directional differentiability, the limiting distribution of $g(\theta)$ is $g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$.[20] A (Lipschitz) continuity on the c.d.f of this limiting distribution then gives the required closeness in quantiles.[21]

Posterior Distribution of $g(\theta^{P*})$ under Directional Differentiability: The limiting distribution $g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$ allows us to characterize and compare large sample approximations of $g(\theta_n^{P*})$ with and without directional differentiability.

[17] It immediately follows that the reverse also applies. If the $100(1-\alpha)\%$–credible interval fails to cover the parameter $g(\theta)$ at θ, then so must the $100(1-\alpha-\epsilon)\%$–bootstrap confidence interval. Note that our approximation holds for any fixed ϵ, but we cannot guarantee that our approximation holds if we take the limit.

[18] If instead of assuming directional differentiability, we assume that posterior distribution $\sqrt{n}(g(\theta_n^{P*} - g(\hat{\theta}_n)$ admits a p.d.f that is uniformly bounded for all n, we can also verify the conditions of Theorem 2. See Theorem 3 in section B.4 of the Appendix.

[19] Equivalently, one could say there is a continuous function $g'_\theta : \mathbb{R}^p \to \mathbb{R}$ 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 continuous, not necessarily linear, function $g'_\theta(\cdot)$ will be referred to as the (Hadamard) directional derivative of g at θ_0.

[20] For the sake of completeness, Lemma 4 in Appendix B.2 shows that if Assumptions 1, 2 and 3 hold and g is directionally differentiable (in the sense defined in Remark 2), then,

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

holds, where $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$ and Z are as defined in Assumption 2.

[21] See Assumption 4 in Appendix B.2 for the details.

If $g'_{\theta_0}(\cdot)$ is linear (which is the case if g is fully differentiable), then the derivative can be characterized by a vector g'_{θ_0} and so $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$ converges to

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

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, $g(\widehat{\theta}_n)$, by a standard delta-method argument.

If g'_{θ_0} is nonlinear, then 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 (even if Z is). 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 $g'_{\theta_0}(Z)$.[22] Thus, one can say that for directionally differentiable functions

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

3 Illustration of main results for $|\theta|$

The main result of this paper, Theorem 1, can be illustrated in the following simple parametric 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:

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

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, 2 and 3. Then, the bootstrap distribution for $g(\widehat{\theta}_n)$ and posterior distributions of $g(\theta)$ are explicitly computed and compared.

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

Relation to main assumptions: The transformation 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, Assumption 1 is satisfied.

We consider the Maximum Likelihood estimator, which is $\widehat{\theta}_n = (1/n)\sum_{i=1}^n X_i$ 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 tractable so the distributions of θ_n^{P*} and θ_n^{B*} can be computed explicitly. The posterior distribution for θ is

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

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.[23]

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 bootstrap 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 bootstrap distribution of $\widehat{\theta}_n$ is

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

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.

[23] 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:

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

Asymptotic Behavior of Posterior/Bootstrap Inference for $g(\theta) = |\theta|$: Since Assumptions 1, 2 and 3 are satisfied, Theorem 1 holds.

In this example, the posterior distribution of $g(\theta_n^{P*})|X^n$ can be characterized explicitly as

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

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

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

Theorem 1 shows that when $\theta_0 = 0$ and n is large enough, this expression can be approximated in the bounded Lipschitz distance in probability by

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

which corresponds to the bootstrap distribution of $\widehat{\theta}_n$.

Moreover, conditional on the data, the distribution of (3.1) has density equal to a shift of a folded normal and can bounded above by a constant that does not depend on n. Theorem 3 in Section B.4 of the Appendix implies that the assumptions of Theorem 2 are verified; that is, the quantiles of (3.1) and (3.2) are close to each other.

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

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

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

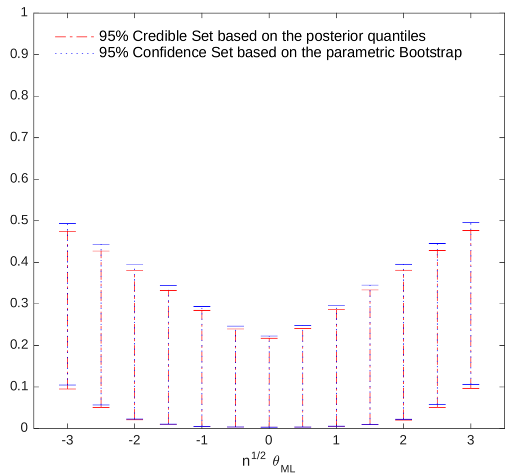
Graphical interpretation of Theorem 1: One way to illustrate Theorem 1 is to compute the 95% credible intervals for $|\theta|$ when $\theta_0 = 0$ using the quantiles of the posterior. We can then compare the 95% credible intervals to the 95% confidence intervals from the bootstrap distribution.

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 1 we report the 95% credible and confidence intervals for data realizations $\sqrt{n}\hat{\theta}_n \in [-3, 3]$. In all four plots the bootstrap confidence intervals are computed using the parametric bootstrap. Posterior credible intervals 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 the Metropolis–Hastings algorithm (Geweke (2005), p. 122).

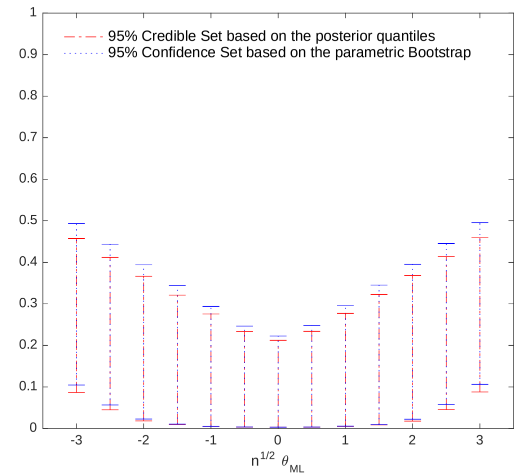
Coverage of Credible Intervals: In this example, the two-sided confidence interval based on the quantiles of the bootstrap distribution of $|\widehat{\theta}_n|$ fails to cover with the nominal probability $|\theta|$ when $\theta = 0$. Theorem 2 showed that the two-sided credible intervals based on the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

As Liu, Gelman, and Zheng (2015) observe, although it is common to report credible intervals based on the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior, a Bayesian might find these credible intervals unsatisfactory. In this problem, it is perhaps more natural to consider one-sided credible intervals or Highest Posterior Density intervals. In the online Appendix C we consider an alternative example, $g(\theta) = \max\{\theta_1, \theta_2\}$, where the decision between two-sided and one-sided credible intervals is less obvious, but the two-sided credible interval still experiences the same problem as the bootstrap.

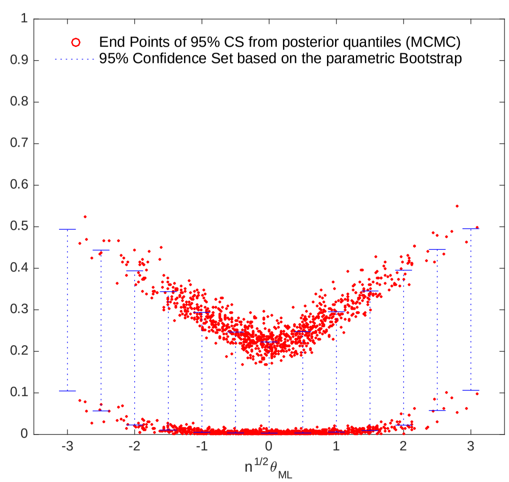
Figure 1: 95% Credible Intervals for $|\theta|$ and 95% Parametric Bootstrap Confidence Intervals



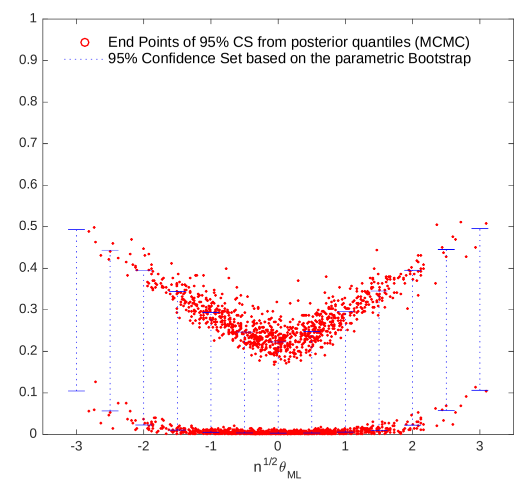
(a) $\lambda^2 = 5$



(b) $\lambda^2 = 10$



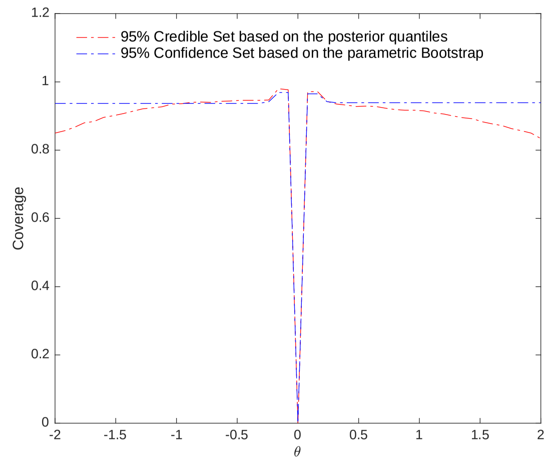
(c) $\gamma(2,2) - 3$



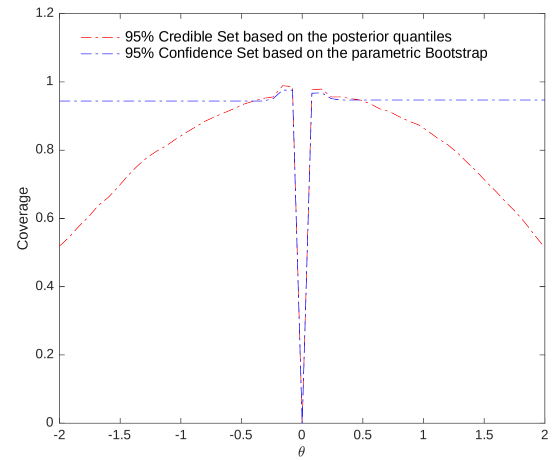
(d) $(\beta(2,2) - 0.5) \times 5$

Description of Figure 1: 95% Credible intervals 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 $|\mathcal{N}(\widehat{\theta}_n, 1/n)|$. The bootstrap distribution only depends on the data through $\widehat{\theta}_n$. (Red, Dotted Line) 95% credible intervals based on the closed-form solution for the posterior. (Red, Circles) 95% credible intervals based on Matlab's MCMC program (computed for 1,000 possible data sets from a standard normal model).

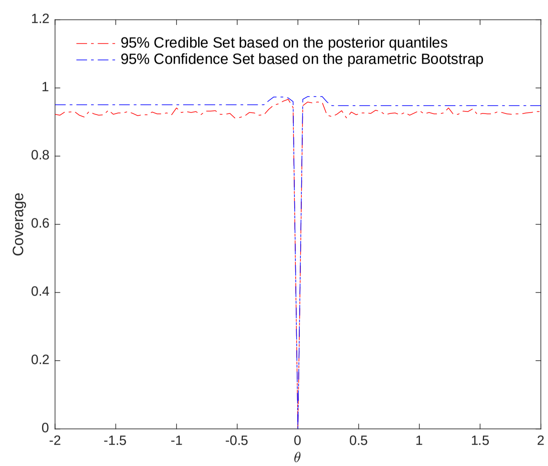
Figure 2: Coverage Probability of 95% Credible Intervals and Parametric Bootstrap Confidence Intervals for $|\theta|$



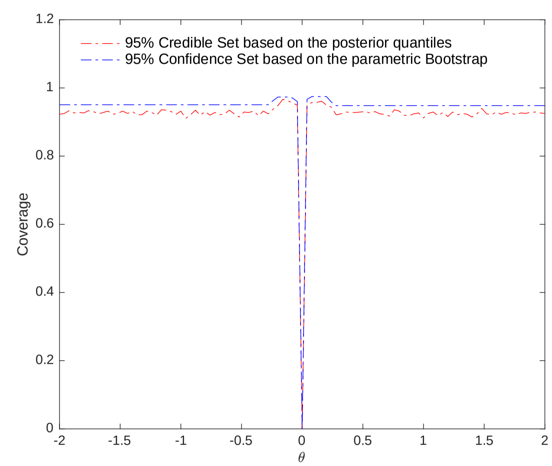
(a) $\lambda^2 = 5$



(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 intervals 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 $|\mathcal{N}(\widehat{\theta}_n, 1/n)|$. (Red, Dotted Line) 95% credible intervals 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 locally 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 bootstrapping $g(\widehat{\theta}_n)$ 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). Our results also provide a better understanding of what type of statistics could preserve the large-sample equivalence between bootstrap and posterior resampling methods, a question that has been explored by Lo (1987).

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 $\sqrt{n}(g(\theta) - g(\widehat{\theta}_n))$ behave differently than the limit of $\sqrt{n}(g(\widehat{\theta}_n) - g(\theta))$. We showed that whenever nondifferentiability causes a bootstrap confidence interval to cover $g(\theta)$ less often than desired, a credible interval based on the quantiles of the posterior will have distorted frequentist coverage as well.

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.

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

Bochkina, N. A. and P. J. Green (2014): "The Bernstein-Von Mises Theorem and Nonregular Models," *The Annals of Statistics*, 42, 1850–1878.

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

Chung, K. L. (2001): *A course in probability theory*, Academic Press.

Clarke, F. (1990): *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics.

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

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

Fiacco, A. V. and Y. Ishizuka (1990): "Sensitivity and stability analysis for nonlinear programming," *Annals of Operations Research*, 27, 215–235.

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 (2018): "Delta-Method inference for a class of set-identified SVARs," *Journal of Econometrics*, 203, 316–327.

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

Ghosal, S., J. K. Ghosh, and T. Samanta (1995): "On convergence of posterior distributions," *The Annals of Statistics*, 2145–2152.

Giacomini, R. and T. Kitagawa (2018): "Robust Bayesian inference for set-identified models," *Cemmap Working Paper 61/18.*

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

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.

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

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

Kline, B. and E. Tamer (2016): "Bayesian Inference in a class of Partially Identified Models," *Quantitative Economics*, 7.

Kosorok, M. R. (2008): *Introduction to empirical processes and semiparametric inference*, Springer Science & Business Media.

Liu, Y., A. Gelman, and T. Zheng (2015): "Simulation-efficient shortest probability intervals," *Statistics and Computing*, 25, 809–819.

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.

McShane, E. J. (1934): "Extension of range of functions," *Bull. Amer. Math. Soc.*, 40.

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

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.

Runkle, D. E. (1987): "Vector Autoregressions and Reality," *Journal of Business & Eco-

nomic Statistics, 5, 437–442.

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.

Whitney, H. (1934): "Analytic extensions of differentiable functions defined in closed sets," *Trans. Amer. Math. Soc.*, 36.

Woutersen, T. and J. Ham (2016): "Confidence Sets for Continuous and Discontinuous Functions of Parameters," *Working Paper.*

A Main Theoretical Results.

A.1 Proof of Theorem 1

Lemma 1. *If $\beta(Z_n^*, Z^*; X^n) \xrightarrow{p} 0$, then Z_n^* converge in distribution to Z^* unconditionally, i.e.*

$$\sup_{f \in BL(1,\mathbb{R}^p)} \Big| \mathbb{E}[f(Z_n^*)] - \mathbb{E}[f(Z^*)] \Big| \to 0, \textit{as} \;\; n \to \infty.$$

Proof. Define

$$A_n \equiv \sup_{f \in BL(1,\mathbb{R}^p)} \Big| \mathbb{E}[f(Z_n^*)|X^n] - \mathbb{E}[f(Z^*)|X^n] \Big|.$$

This random variable is bounded by 2 and converges to zero in probability by assumption. Theorem 4.1.4 of Chung (2001) (p. 71), implies that A_n converges in L^1-norm to zero; i.e., $\mathbb{E}[A_n] \to 0$ as $n \to \infty$.

For any $f \in BL(1, \mathbb{R}^p)$,

$$A_n \geq \Big| \mathbb{E}[f(Z_n^*)|X^n] - \mathbb{E}[f(Z^*)|X^n] \Big|.$$

Taking expectation on both sides

$$\begin{aligned} \mathbb{E}[A_n] &\geq \mathbb{E}\Big[\Big| \mathbb{E}[f(Z_n^*)|X^n] - \mathbb{E}[f(Z^*)|X^n] \Big| \Big], \\ &\geq \Big| \mathbb{E}\Big[\mathbb{E}[f(Z_n^*)|X^n] - \mathbb{E}[f(Z^*)|X^n] \Big] \Big|, \\ &= \Big| \mathbb{E}[f(Z_n^*)] - \mathbb{E}[f(Z^*)] \Big|. \end{aligned}$$

Consequently,

$$\sup_{f \in BL(1,\mathbb{R}^p)} \Big| \mathbb{E}[f(Z_n^*)] - \mathbb{E}[f(Z^*)] \Big| \leq \mathbb{E}[A_n] \to 0.$$

By part (iii) of Lemma 2.2 (Portmanteau) in Van der Vaart (2000) (p. 6), Z_n^* converges to Z^* in distribution (unconditionally). □

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*}. Define

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

where $Z_n^{*B} \equiv \sqrt{n}(\theta_n^{*B} - \widehat{\theta}_n)$ and $Z_n^{*P} = \sqrt{n}(\theta_n^{*P} - \widehat{\theta}_n)$. We break the proof of the Theorem into 8 steps.

Step 1: Fix $\epsilon > 0$. Lemma 1 implies that both Z_n^{*P} and Z_n^{*B} are tight, as they converge in distribution (unconditionally) to some random element Z^*. Then, there exists a compact subset $K_\epsilon \subseteq \mathbb{R}^p$ such that $\mathbb{P}[Z_n^{*P} \in K_\epsilon] \geq 1 - \epsilon$ and $\mathbb{P}[Z_n^{*B} \in K_\epsilon] \geq 1 - \epsilon$ for all n.

Step 2: By Assumption 1, g is locally Lipchitz at θ_0, then there exists $\delta_0 > 0$ such that:

$$|g(x) - g(y)| \leq c_0||x - y|| \quad x, y \in V_0 \equiv \{z : ||z - \theta_0|| < \delta_0\}.$$

Define $V_1 \equiv \{\theta : ||\theta - \theta_0|| < \delta_0/2\} \subset V_0$. By Assumption 2, there exists $N_1 \equiv N_1(\delta_0, \epsilon)$ such that $\mathbb{P}[\hat{\theta}_n \in V_1] \geq 1 - \epsilon$ for all $n \geq N_1$.

Step 3: Consider $M \equiv \sup\{||a|| : a \in K_\epsilon\}$ and define $\Delta_n : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$

$$\Delta_n(a, b) \equiv \sqrt{n}(g(b + a/\sqrt{n}) - g(b)).$$

For each $b \in V_1$ and $n > N_2 \equiv (2M/\delta_0)^2$, we claim that $\Delta_n(\cdot, b)$ is a c_0-Lipschitz function in K_ϵ. This means that for each $b \in V_1$ and n sufficiently large

$$|\Delta_n(a_1, b) - \Delta_n(a_2, b)| \leq c||a_1 - a_2||, \quad \forall a_1, a_2 \in K_\epsilon.$$

It is then sufficient to show that

$$b + a/\sqrt{n} \in V_0$$

for all $a \in K_\epsilon$ and $n > N_2$. This is true because if any two points

$$b + a_1/\sqrt{n} \text{ and } b + a_2/\sqrt{n}$$

both belong to V_0 (which is a neighborhood of θ_0), the locally Lipschitz property of g at θ_0 readily gives

$$|\sqrt{n}(g(b + a_1/\sqrt{n}) - g(b)) - \sqrt{n}(g(b + a_2/\sqrt{n}) - g(b))| \leq c_0||a_1 - a_2||.$$

If $b \in V_1$, $n > 2$, and $a \in K_\epsilon$

$$\begin{aligned}
||b + a/\sqrt{n} - \theta_0|| &\leq ||b - \theta_0|| + ||a/\sqrt{n}|| \\
&\leq \delta_0/2 + ||a/\sqrt{n}||, \\
&(\text{as } b \in V_1) \\
&\leq \delta_0/2 + M/\sqrt{n}, \\
&(\text{as } a \in K_\epsilon) \\
&\leq \delta_0, \\
&(\text{as } n > N_2).
\end{aligned}$$

Step 4: For each $n > N_2$ and $b \in V_1$, there exists $F_n(\cdot, b) : \mathbb{R}^p \to \mathbb{R}$ such that $F_n(\cdot, b)$ is a c_0-Lipschitz function and $F_n(a, b) = \Delta_n(a, b)$ for all $a \in K_\epsilon$ (See McShane (1934), Whitney (1934)), i.e. $F_n(\cdot, b)$ is a c_0-Lipschitz extension of $\Delta_n(\cdot, b)|_{K_\epsilon}$.

Step 5: For each $n > N_2$, $b \in V_1$ and $f \in BL(1, \mathbb{R})$, we have that $\frac{1}{\tilde{c}} f \circ F_n(\cdot, b) \in BL(1, \mathbb{R}^p)$, where $\tilde{c} \equiv \max\{c_0, 1\}$. Therefore,

$$\left| \mathbb{E}\left[\frac{f \circ F_n(Z_n^{*P}, b)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*B}, b)}{\tilde{c}} \;\Big|\; X^n \right] \right|$$

is smaller than or equal to

$$\sup_{f \in BL(1, \mathbb{R}^p)} \left| \mathbb{E}[f(Z_n^{*P})|X^n] - \mathbb{E}[f(Z_n^{*B})|X^n] \right| \equiv A_n.$$

Consequently:

$$1\{\widehat{\theta}_n \in V_1\} \left| \mathbb{E}\left[\left(\frac{f \circ F_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} \right) \;\Big|\; X^n \right] \right| \leq A_n.$$

Since $1\{\widehat{\theta}_n \in V_1\}$ is X^n-measurable

$$\left| \mathbb{E}\Big[1\{\widehat{\theta}_n \in V_1\} \left(\frac{f \circ F_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} \right) \;\Big|\; X^n \Big] \right| \leq A_n.$$

Step 6: Let $N_3 \equiv \max\{N_1, N_2\}$. For each $f \in BL(1, \mathbb{R})$

$$
\begin{aligned}
&\left|\mathbb{E}[f(\sqrt{n}(g(\theta_n^{*P}) - g(\widehat{\theta}_n))) \mid X^n] - \mathbb{E}[f(\sqrt{n}(g(\theta_n^{*B}) - g(\widehat{\theta}_n))) \mid X^n]\right| \\
&= \tilde{c} \cdot \left|\mathbb{E}\left[\frac{f(\sqrt{n}(g(\widehat{\theta}_n + Z_n^{*P}/\sqrt{n}) - g(\widehat{\theta}_n)))}{\tilde{c}} \;\Big|\; X^n\right]\right. \\
&\quad \left. - \mathbb{E}\left[\frac{f(\sqrt{n}(g(\widehat{\theta}_n + Z_n^{*B}/\sqrt{n}) - g(\widehat{\theta}_n)))}{\tilde{c}} \mid X^n\right]\right| \\
&= \tilde{c} \cdot \left|\mathbb{E}\left[\frac{f \circ \Delta_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\right] - \mathbb{E}\Big[\frac{f \circ \Delta_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\Big]\right| \\
&\leq |I_1| + |I_2|,
\end{aligned}
$$

where

$$
I_1 \equiv \tilde{c} \cdot 1\{\widehat{\theta}_n \in V_1\} \cdot \left(\mathbb{E}\left[\frac{f \circ \Delta_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\right] - \mathbb{E}\left[\frac{f \circ \Delta_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\right]\right),
$$

$$
I_2 \equiv \tilde{c} \cdot 1\{\widehat{\theta}_n \notin V_1\} \cdot \left(\mathbb{E}\left[\frac{f \circ \Delta_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\right] - \mathbb{E}\left[\frac{f \circ \Delta_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} \;\Big|\; X^n\right]\right).
$$

For the second term, I_2, $|I_2| \leq 1\{\widehat{\theta}_n \notin V_1\} \cdot 2$ since $f \in BL(1, \mathbb{R})$. For the first term, I_1, $F_n(\cdot, \widehat{\theta}_n)$ is a well-defined c_0-Lipschitz function since $\widehat{\theta}_n \in V_1$. Since $1\{\widehat{\theta}_n \in V_1\}$ is X^n-measurable, we can further decompose I_1 as the sum of $I_3 + I_4 + I_5$ where

$$
I_3 \equiv \tilde{c} \cdot \mathbb{E}\Big[1\{\hat{\theta}_n \in V_1\} \cdot \left(\frac{f \circ \Delta_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}}\right) \;\Big|\; X^n\Big],
$$

$$
I_4 \equiv \tilde{c} \cdot \mathbb{E}\Big[1\{\hat{\theta}_n \in V_1\} \cdot \left(\frac{f \circ F_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ \Delta_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}}\right) \;\Big|\; X^n\Big],
$$

$$
I_5 \equiv \tilde{c} \cdot \mathbb{E}\Big[1\{\hat{\theta}_n \in V_1\} \cdot \left(\frac{f \circ F_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}}\right) \;\Big|\; X^n\Big].
$$

From The definition of F_n (as extension of Δ_n)

$$
I_3 = \tilde{c} \cdot \mathbb{E}\Big[1\{Z_n^{*P} \notin K_\epsilon\} \cdot 1\{\hat{\theta}_n \in V_1\} \cdot \left(\frac{f \circ \Delta_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ F_n(Z_n^{*P}, \widehat{\theta}_n)}{\tilde{c}}\right) \;\Big|\; X^n\Big],
$$

$$
I_4 = \tilde{c} \cdot \mathbb{E}\Big[\{Z_n^{*B} \notin K_\epsilon\} \cdot 1\{\hat{\theta}_n \in V_1\} \cdot \left(\frac{f \circ F_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}} - \frac{f \circ \Delta_n(Z_n^{*B}, \widehat{\theta}_n)}{\tilde{c}}\right) \;\Big|\; X^n\Big].
$$

and so

$$
|I_3| + |I_4| \leq \mathbb{E}[1\{Z_n^{*P} \notin K_\epsilon\} \cdot 2 \mid X^n] + \mathbb{E}[1\{Z_n^{*B} \notin K_\epsilon\} \cdot 2 \mid X^n].
$$

In addition, by Step 5,

$$|I_5| \leq \tilde{c} \cdot A_n.$$

Collecting the inequalities

$$\begin{aligned}
&\left|\mathbb{E}[f(\sqrt{n}(g(\theta_n^{*P}) - g(\widehat{\theta}_n))) \mid X^n] - \mathbb{E}[f(\sqrt{n}(g(\theta_n^{*B}) - g(\widehat{\theta}_n))) \mid X^n]\right| \\
&\quad \leq |I_1| + |I_2|, \\
&\quad \leq |I_3| + |I_4| + |I_5| + 1\{\widehat{\theta}_n \notin V_1\} \cdot 2, \\
&\quad \leq \mathbb{E}[1\{Z_n^{*P} \notin K_\epsilon\} \cdot 2 \mid X^n] + \mathbb{E}[1\{Z_n^{*B} \notin K_\epsilon\} \cdot 2 \mid X^n] + \tilde{c} \cdot A_n + 1\{\widehat{\theta}_n \notin V_1\} \cdot 2.
\end{aligned}$$

Taking the supremum over $f \in BL(1, \mathbb{R})$

$$B_n \leq \mathbb{E}[1\{Z_n^{*P} \notin K_\epsilon\} \cdot 2 \mid X^n] + \mathbb{E}[1\{Z_n^{*B} \notin K_\epsilon\} \cdot 2 \mid X^n] + \tilde{c} \cdot A_n + 2 \cdot 1\{\hat{\theta}_n \notin V_1\}.$$

Applying expectation:

$$\mathbb{E}[B_n] \leq 2 \cdot \mathbb{P}[Z_n^{*P} \notin K_\epsilon] + 2 \cdot \mathbb{P}[Z_n^{*B} \notin K_\epsilon] + \tilde{c} \cdot \mathbb{E}[A_n] + 2 \cdot \mathbb{P}[\hat{\theta}_n \notin V_1].$$

Step 7: For any $n > N_3$ (which was previously defined as the maximum of N_1 and N_2), using Step 1 and Step 2 in Step 6

$$\mathbb{E}[B_n] \leq 4 \cdot \epsilon + \tilde{c} \cdot \mathbb{E}[A_n] + 2 \cdot \epsilon.$$

Assumption 3 and the triangle inequality imply that A_n converges to zero in probability. Since A_n is also bounded by 2, once again applying Theorem 4.1.4 of Chung (2001), we conclude that A_n converges in L^1-norm to zero: $\mathbb{E}[A_n] \to 0$. Then,

$$\limsup \mathbb{E}[B_n] \leq 6 \cdot \epsilon + \tilde{c} \cdot \limsup \mathbb{E}[A_n] = 6 \cdot \epsilon$$

Thus, $\limsup \mathbb{E}[B_n] \leq 0$, which implies that $\lim \mathbb{E}[B_n] = 0$ since $B_n \geq 0$.

Step 8: Since B_n is a random variable bounded by 2, and it converge to zero in L^1-norm, we conclude, by a standard application of Markov's inequality, B_n converges to zero in probability.

These steps prove Theorem 1. □

A.2 Proof of Theorem 2

Proof of Theorem 2: 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:

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

By assumption of the theorem for every $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $N(\epsilon, \delta)$ such that

$$\mathbb{P}_\theta[c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n)] \geq 1 - \delta,\ \forall n \geq N(\epsilon, \delta).$$

This implies

$$\begin{aligned} \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n), c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n)) \\ &\quad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n), c_{\alpha-\epsilon}^{P*}(X^n) > c_\alpha^{B*}(X^n)), \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n), c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n)) \\ &\quad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n), c_\alpha^{B*}(X^n) > c_{\alpha+\epsilon}^{P*}(X^n)), \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) + \delta. \end{aligned}$$

Thus, for $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \tag{A.1}$$
$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta. \tag{A.2}$$

Consequently:

$$\begin{aligned}
\mathbb{P}_\theta\left(g(\theta) \in CS_n^B(1-\alpha)\right) &= \mathbb{P}_\theta\left(g(\theta) \in \left[g(\widehat{\theta}_n) + c_{\alpha/2}^{B*}(X^n)/\sqrt{n}\,,\ g(\widehat{\theta}_n) + c_{1-\alpha/2}^{B*}/\sqrt{n}\right]\right) \\
&= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha/2}^{B*}(X^n)) \\
&- \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{1-\alpha/2}^{B*}(X^n)) \\
&\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha/2+\epsilon/2}^{P*}(X^n)) \\
&- \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{1-\alpha/2-\epsilon/2}^{P*}(X^n)) - \delta \\
&\text{(Replacing } \alpha \text{ by } \alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (B.10) and} \\
&\text{replacing } \alpha \text{ by } 1-\alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (B.9))} \\
&= \mathbb{P}_\theta\left(g(\theta) \in CS_n^P(1-\alpha-\epsilon)\right) - \delta.
\end{aligned}$$

Therefore, for every $0 < \epsilon < \alpha$:

$$1 - \alpha - d_\alpha \geq \limsup_{n\to\infty} \mathbb{P}_\theta\left(g(\theta) \in CS_n^B\right) \geq \limsup_{n\to\infty} \mathbb{P}_\theta\left(g(\theta) \in CS_n^P(1-\alpha-\epsilon)\right),$$

which implies that

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

This implies that if the bootstrap fails at θ by at least $100 d_\alpha\%$ given the nominal confidence level $100(1-\alpha)\%$, then the confidence interval based on the quantiles of the posterior will fail at θ—by at least $100(d_\alpha - \epsilon)\%$—given the nominal confidence level $(1 - \alpha - \epsilon)$.

B Additional Results (intended for online publication only)

B.1 Bootstrap and Posterior quantiles

This section establishes the closeness between bootstrap/posterior quantiles as assumed in Theorem 2 for directional differentiable functions. For the sake of generality, we provide a slightly more general result based on high-level assumptions that we then verify for directionally differentiable $g(\cdot)$.

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

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

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

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

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

equals:

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

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 2. *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 $n > N(\epsilon, \delta)$:

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

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_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. By applying Lemma 5 in Appendix B.3 to the set $A = (-\infty, c)$ it follows that for any $\zeta > 0$:

$$|P_W^n((-\infty, c)|X^n) - P_Y^n((-\infty, c)|X^n)|$$

$$\leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \max\{(P_Y^n(A^\zeta \setminus A|X^n), P_Y^n((A^c)^\zeta \setminus A^c|X^n)\}$$

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

$$\leq \frac{1}{\zeta}\beta(W_n^*, Y_n^*; X^n) + \mathbb{P}^Z\left(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta \mid X^n\right)$$

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 5). Therefore:

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

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

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:

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

$$A_2(\zeta, X^n) \equiv \sup_{c \in \mathbb{R}} \mathbb{P}^Z\left(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta \mid X^n\right)$$

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 $c \in \mathbb{R}$:

$$-\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c \mid X^n) - \mathbb{P}^Z\left(h_\theta(Z, X^n) \leq c \mid X^n\right) < \epsilon. \tag{B.1}$$

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.1) 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.1) 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 $n > \mathcal{N}(\epsilon/2, \delta/2)$:

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

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

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

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

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

□

We have shown that if $\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))$ is any random variable satisfying the assumptions of Lemma 2, 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_n^{B*} or θ_n^{P*} 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 2 before the following Lemma.

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

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

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

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

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \tag{B.2}$$

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

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 2. Under Assumptions 1, 2, 3 and 4, the conditions for Lemma 2 are satisfied. It follows that for any $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $\mathcal{N}(\epsilon,\delta)$ such that for all $n > \mathcal{N}(\epsilon,\delta)$:

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

Therefore:

$$\begin{aligned}
&\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n)) \\
&\qquad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) \geq c_\alpha^*(X^n)) \\
&\qquad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\
&\qquad \text{(by the additivity of probability measures)} \\
&\qquad \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\theta(c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\
&\qquad \text{(by the monotonicity of probability measures)} \\
&\qquad \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \delta/2.
\end{aligned} \tag{B.4}$$

Also, we have that:

$$
\begin{aligned}
&\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c_{\alpha-\epsilon/2}(X^n) \text{ and } c^*_\alpha(X_n) \ge c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n) \text{ and } c^*_\alpha(X^n) \ge c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n)) + \mathbb{P}_\theta(c^*_\alpha(X^n) \ge c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad - \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n) \text{ or } c^*_\alpha(X^n) \ge c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad (\text{using } P(A \cap B) = P(A) + P(B) - P(A \cup B)) \\
&\qquad \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n)) - (1 - \mathbb{P}_\theta(c^*_\alpha(X^n) \ge c_{\alpha-\epsilon/2}(X^n))) \\
&\qquad (\text{since } \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n) \text{ or } c^*_\alpha(X^n) \ge c_{\alpha-\epsilon/2}(X^n)) \le 1) \\
&\qquad = \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n)) - \mathbb{P}_\theta(c^*_\alpha(X^n) < c_{\alpha-\epsilon/2}(X^n)) \\
&\qquad \ge \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \le -c^*_\alpha(X^n)) - \delta/2. \qquad \text{(B.5)}
\end{aligned}
$$

Replacing c^*_α by c^{B*}_α in (B.5) and c^*_α by c^{P*}_α and α by $\alpha - \epsilon$ in (B.4) implies that for $n > N(\epsilon, \delta)$:

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

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

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

This establishes equation (B.2). Replacing θ^*_n by θ^{B*}_n in (B.4) and replacing θ^*_n by θ^{P*}_n, α by $\alpha + \epsilon$ (B.5) implies that for $n > N(\epsilon, \delta)$:

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

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

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

which establishes equation (B.3). $\square$

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

Lemma 4. *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 $\mathbb{P}^{Z^*}$.[24] 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:

$$\mathbb{P}_\theta\left(\sup_{BL(1)}\Big|\mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n]\Big| > 2\epsilon\right) \tag{B.6}$$

is bounded above by:

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

which equals

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

Thus, by Markov's inequality:

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

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.[25] □

B.3 Additional Lemmas

Lemma 5. *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,*

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

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

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

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:

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

It follows that:

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

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

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

which implies that:

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

The desired result follows. □

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, consider the program*

$$g(\theta) \equiv \max_{x \in \mathbb{R}^n} e_i' C_k(A) x, \ \ s.t. \ \ x'\Sigma^{-1}x = 1, \ S(\theta)x \geq 0, \tag{B.7}$$

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.7) 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 $\tilde{S}^*(\theta)x^*(\theta) = \mathbf{0}_{r\times 1}$. That is, there exists $\alpha \in \mathbb{R}_+^{r \times m^*}$ s.t.*

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

and

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

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 et al. (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 $E(\tilde{\theta}) \equiv \{x \in \mathbb{R}^n \mid x'\tilde{\Sigma}^{-1}x = 1\}$. 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.

$\square$

B.4 Alternative Statement for Theorem 2

Theorem 3. *Suppose that the nominal* $100(1-\alpha)\%$ *bootstrap confidence interval fails to cover* $g(\theta)$ *at* θ *by at least* $100 d_\alpha\%$. *Suppose in addition that for each* n *the probability density function of* $\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))$ *is uniformly bounded. If the Assumptions 1, 2 and 3 hold. Then for any* $0 < \epsilon < \alpha$:

$$\mathbb{P}_\theta[q^P_{\alpha-\epsilon}(X^n) \leq q^B_\alpha(X^n) \leq q^P_{\alpha+\epsilon}(X^n)] \to 1 \quad \textit{as } n \to \infty.$$

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. And*

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

Thus, the nominal $100(1-\alpha-\epsilon)\%$ *credible interval based on the quantiles of the posterior fails to cover* $g(\theta)$ *at* θ *by at least* $100(d_\alpha - \epsilon)\%$.

Proof. **Step 1:** We will first prove the closeness in quantiles.

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:

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

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. By applying Lemma 5 in Appendix B.3 to the set $A = (-\infty, c)$ it follows that for any $\zeta > 0$:

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

where for any set A, we have defined $A^\zeta \equiv \{y \in \mathbb{R}^k : \|x - y\| < \zeta \text{ for some } x \in A\}$ (as in Lemma 5). Therefore:

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

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

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:

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

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

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 $c \in \mathbb{R}$:

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

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

$$-\epsilon < \mathbb{P}^* \left(\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \leq c \mid X^n \right) - \mathbb{P}^Z \left(\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \leq c \mid X^n \right) < \epsilon. \tag{B.8}$$

We now show that 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.8) at $c_{\alpha+\epsilon}^{P*}(X^n)$. This implies that:

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

Consequently:

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

Since:

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

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 $c_{\alpha}^{B*}(X^n)$. This implies that:

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

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:

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

This implies that

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

Given $\epsilon > 0$, we can choose $\zeta^* = \epsilon/(2K)$ that satisfies

$$\mathbb{P}(A_2(\zeta^*, X^n) > \epsilon) = 0.$$

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

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

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

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

Step 2: Now, we prove that when the bootstrap confidence interval fails to cover $g(\theta)$, then the posterior credible set also fails to cover.

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:

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

By step 1 we have that every $0 < \epsilon < \alpha$ and $\delta > 0$ there exists $N(\epsilon, \delta)$ such that

$$\mathbb{P}_\theta[c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n)] \geq 1 - \delta, \; \forall n \geq N(\epsilon, \delta).$$

This implies

$$\begin{aligned} \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n), c_{\alpha-\epsilon}^{P*}(X^n) \leq c_\alpha^{B*}(X^n)) \\ &\quad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n), c_{\alpha-\epsilon}^{P*}(X^n) > c_\alpha^{B*}(X^n)), \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n), c_\alpha^{B*}(X^n) \leq c_{\alpha+\epsilon}^{P*}(X^n)) \\ &\quad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n), c_\alpha^{B*}(X^n) > c_{\alpha+\epsilon}^{P*}(X^n)), \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) + \delta. \end{aligned}$$

Thus, for $n > N(\epsilon, \delta)$:

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P*}(X^n)) + \delta, \tag{B.9}$$

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B*}(X^n)) \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P*}(X^n)) - \delta. \tag{B.10}$$

Consequently:

$$\begin{aligned}
\mathbb{P}_\theta\left(g(\theta) \in CS_n^B(1-\alpha)\right) &= \mathbb{P}_\theta\left(g(\theta) \in \left[g(\widehat{\theta}_n) + c^{B*}_{\alpha/2}(X^n)/\sqrt{n}\,,\ g(\widehat{\theta}_n) + c^{B*}_{1-\alpha/2}/\sqrt{n}\right]\right) \\
&= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_{\alpha/2}(X^n)) \\
&- \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{B*}_{1-\alpha/2}(X^n)) \\
&\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{\alpha/2+\epsilon/2}(X^n)) \\
&- \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c^{P*}_{1-\alpha/2-\epsilon/2}(X^n)) - \delta \\
&\text{(Replacing } \alpha \text{ by } \alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (B.10) and} \\
&\text{replacing } \alpha \text{ by } 1-\alpha/2,\ \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (B.9))} \\
&= \mathbb{P}_\theta\left(g(\theta) \in CS_n^P(1-\alpha-\epsilon)\right) - \delta.
\end{aligned}$$

Therefore, for every $0 < \epsilon < \alpha$:

$$1 - \alpha - d_\alpha \geq \limsup_{n\to\infty} \mathbb{P}_\theta\left(g(\theta) \in CS_n^B\right) \geq \limsup_{n\to\infty} \mathbb{P}_\theta\left(g(\theta) \in CS_n^P(1-\alpha-\epsilon)\right),$$

which implies that

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

This implies that if the bootstrap fails at θ by at least $100 d_\alpha\%$ given the nominal confidence level $100(1-\alpha)\%$, then the confidence interval based on the quantiles of the posterior will fail at θ—by at least $100(d_\alpha - \epsilon)\%$—given the nominal confidence level $(1 - \alpha - \epsilon)$. □