Journal of Econometrics 217 (2020) 161–175

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Posterior distribution of nondifferentiable functions[☆]

Toru Kitagawa [a,*], José Luis Montiel Olea [b], Jonathan Payne [c], Amilcar Velez [d]

[a] Department of Economics, University College London, United Kingdom of Great Britain and Northern Ireland
[b] Department of Economics, Columbia University, United States of America
[c] Department of Economics, Princeton University, United States of America
[d] Department of Economics, Northwestern University, United States of America

ARTICLE INFO

Article history:
Received 3 April 2019
Received in revised form 6 September 2019
Accepted 25 October 2019
Available online 21 January 2020

Keywords:
Bootstrap
Bernstein–von Mises theorem
Directional differentiability
Posterior inference

ABSTRACT

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).

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 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]

☆ We would like to thank Gary Chamberlain, Tim Cogley, Han Hong, Hiroaki Kaido, Oliver Linton, Ulrich Müller, Mikkel Plagborg-Möller, Andres Santos, Quang Vuong, and three anonymous referees for detailed comments and suggestions on an earlier draft of this paper. We would also like to thank Ruby Steedle for excellent research assistance. 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. Velez thanks the Macro Modeling Division, Central Reserve Bank of Peru, where part of this work was written. First version: July 25th, 2016. This draft: October 31st, 2019.

* Corresponding author.
E-mail addresses: t.kitagawa@ucl.ac.uk (T. Kitagawa), montiel.olea@gmail.com (J.L. Montiel Olea), jepayne@princeton.edu (J. Payne), amilcare@u.northwestern.edu (A. Velez).

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 et al., 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).

https://doi.org/10.1016/j.jeconom.2019.10.009
0304-4076/

The potential nondifferentiability of $g(\cdot)$ poses challenges for 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 nondifferentiable.

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. 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.

It is important to note that the nondifferentiable functions studied in this paper depend only on θ. Our results do not apply to the bootstrap/posterior distributions of test statistics, which depend explicitly on both θ and the data. In fact, Chen et al. (2018a) have shown that the posterior and bootstrap distributions of profile likelihood ratio test statistics for set-identified models differ asymptotically.[3]

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.[6]

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., a *bagging* estimator reporting the mean of the bootstrap draws) can be interpreted as an approximately Bayes estimator for $g(\theta)$ (e.g., the posterior mean estimator).[7] 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 $g(\widehat{\theta}_n)$.

The rest of this paper is organized as follows. Section 2 presents a formal statement of the main results. Section 3 presents an illustrative example: the absolute value transformation. Section 4 concludes. The proofs of the main results are collected in Appendix A. Additional derivations are presented in Appendix B (see supplementary materials).

---

2 See, for example, DasGupta (2008), p. 291 for a Bernstein–von Mises theorem for regular parametric models where Z is Gaussian; Ghosal et al. (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 See Example 3 in Section 5.3.3. of Chen et al. (2018a) and Appendix E.2 of Chen et al. (2018b).

4 See Theorem 4.2 in Fiacco and Ishizuka (1990) and Proposition 6 in Morand et al. (2015).

5 We discuss this point in the framework of set-identified Structural Vector Autoregressions in Section 2 and in Appendix B.3.

6 The word *robust* refers to robustness with respect to the choice of a prior distribution for the set-identified structural parameters of the model; see Giacomini and Kitagawa (2018). The motivation for robustness in this context is that a prior imposed on unidentified parameters will not be updated by the data even in large samples. Therefore, if one cannot form a credible prior for unidentified parameters, it would be desirable to draw "robust" posterior statements that are true for any prior in a class of prior distributions.

7 See Section 8.7 of Friedman et al. (2017) for a definition of bootstrap aggregation or *bagging*.

2. Main results

Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of size n from the model $\mathbb{P}^n_\eta$, 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 $\mathbb{P}^n_{\eta_0}$. 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 exist 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.[8] 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.[9] Later, we will also present 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.[10] 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).

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

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

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

$$\beta(\phi^*_n, \psi^*_n;\ X^n) \equiv \sup_{f \in BL(1,\mathbb{R}^p)} \left| \mathbb{E}[f(\phi^*_n)|X^n] - \mathbb{E}[f(\psi^*_n)|X^n] \right|.$$

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 $n \to \infty$.[11]

Let P denote some prior for θ and let θ^{P*}_n denote the random variable with law equal to the posterior distribution of θ given a sample X^n of size n. Let θ^{B*}_n denote the random variable with law equal to the bootstrap distribution of $\widehat{\theta}_n$ given a sample X^n of size n.

Remark 1. In a parametric model there are different ways of bootstrapping the distribution of $\hat{\theta}_n$. One possibility is a *parametric bootstrap*, which consists of generating draws $(x_1, \ldots, x_n)$ from the model $\mathbb{P}^n_{\hat{\eta}}$, where $\hat{\eta}$ is an estimator of η, followed by an evaluation of $\widehat{\theta}_n$ 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, as specified below.

---

8 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.

9 Directional differentiability is not a special case of local Lipschitzness. It is true that if g is locally Lipschitz at a point θ, then there is a neighborhood of θ over which g is almost everywhere differentiable (by virtue of Rademacher's theorem in Evans and Gariepy, 2015). However, even if g is almost everywhere differentiable in a neighborhood of θ, it does not follow that the directional derivative of g at θ exists.

10 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).

11 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).

The following assumption restricts the prior P for θ and the bootstrap procedure for $\widehat{\theta}_n$ (after appropriate centering and scaling).[12]

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,

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

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).[13,14]

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 1–3 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. □

To the best of our knowledge, Theorem 1 is new.[15] An analogous result for the case in which g is a directionally differentiable function can be derived from an application of the delta method for directionally differentiable functions in Shapiro (1991) or from Proposition 1 in Dümbgen (1993). We provide more details on p. 6 after Remark 2.

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). A crucial step is to show that Z_n^{P*} and Z_n^{B*} converge unconditionally and, therefore, are tight; see Lemma 1. This means that the asymptotic relation between the

[12] We center around $\widehat{\theta}_n$ because we assume that θ is a well-behaved parameter, for which neither the bootstrap distribution nor the posterior distribution need to be bias-adjusted.

[13] 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. Convergence in 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.

[14] 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.

[15] Theorem 1 does not follow directly from the existing literature. First, we are assuming that g is locally Lipschitz at θ_0. This requires us to restrict attention to the neighborhood of θ_0, which takes some additional work. 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 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

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

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 $g(Z)$.

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 Vector Autoregressions (SVARs): One illustration of the usefulness of Theorem 1 is in the "robust Bayes" analysis of SVARs. Consider an n-dimensional SVAR 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:

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

where $A \equiv (A_1, A_2, \ldots, A_p)$ are the autoregressive parameters of the model and $\Sigma \equiv BB'$ is the covariance matrix of reduced-form residuals, $B\varepsilon_t$. Define the impact of shock j in variable i at k periods in the future, referred to as 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 ith and jth columns 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: $\theta \equiv (\mathrm{vec}(A)', \mathrm{vech}(\Sigma)')'$.

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 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 (MFCQ) to hold at an optimal solution (Proposition 6 in Morand et al., 2015).[16] 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., a *bagging* estimator reporting 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 bagging estimator that reports the mean of the bootstrap draws.[17]

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 and may involve stronger assumptions than we impose in this paper. For example, Morand et al. (2015) argue that '*because the MFCQ is not sufficient to guarantee the uniqueness of KKT [Karush–Kuhn–Tucker] multipliers, it is very difficult to obtain directional derivatives and sharp characterizations of the generalized gradient*'.[18] Thus, this paper emphasizes local Lipschitz continuity because it is simpler to verify and so of greater potential use to practitioners.

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

16 In addition to the choice set being nonempty and uniformly compact at θ. We verify all these conditions in Lemma 6 of Appendix B.3.

17 Such *bagging* estimator might also be easier to compute, if bootstrap draws are computationally less demanding than posterior draws. This could happen, for example, in SVARs with conditional heteroskedasticity.

18 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). These are stronger assumptions than we impose in Lemma 6 of Appendix B.3, where we formally apply the results of our paper to SVAR example. To the best of our knowledge, directional differentiability of the value function is only straightforward to verify in well-behaved optimization problems; for example, convex programs with non-empty interiors (see Corollary 5 in Milgrom and Segal, 2002 and Theorem 4.2 in Fiacco and Ishizuka, 1990).

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 $g(\theta)$ for any $\epsilon > 0$.[19]

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).

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

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

where $\mathbb{P}_\eta^n$ refers to the distribution of $(X_1, \ldots, X_n)$ under the parameter η.

The next theorem relates the coverage probability of the Bayesian credible interval for $g(\theta)$ to the coverage probability of its bootstrap confidence interval. We state the theorem with a high level assumption on the closeness of quantiles and later provide a number of conditions under which the required closeness in quantiles can be established.

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

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

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 $0 < \epsilon < \alpha$:

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

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 $100(d_\alpha-\epsilon)\%$.

Proof. See Appendix A.2.[20] □

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 $g(\theta)$.

Remark 2. The desired closeness of quantiles can be established under a few more regularity assumptions. For instance, if we assume that the posterior distribution $\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))$ admits a p.d.f that is uniformly bounded for all n, then we can also verify the conditions of Theorem 2 (see Theorem 3 in section B.1 of the Appendix for the details). This condition is not implied by (nor implies) directional differentiability. We verify our condition in the illustrative example presented in Section 3.

Posterior Distribution of $g(\theta^{P*})$ under Directional Differentiability: In order to give a more concrete characterization of the posterior distribution of $g(\theta^{P*})$ in large samples, we now assume that g is directionally differentiable. 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 $t_n \to 0$[21]:

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

[19] 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.

[20] We can observe that 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.

[21] One could simply say there is a continuous function $g'_\theta : \mathbb{R}^p \to \mathbb{R}$ such that for any converging sequence $h_n \to h$:

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

Proposition 1 in Dümbgen (1993) and equation A.41 in Theorem A.1 in Fang and Santos (2019) in combination with our Theorem 1 can be used to show that, under directional differentiability, the limiting distribution of $g(\theta)$ is $g'_{\theta_0}(Z+Z_n)-g'_{\theta_0}(Z_n)$.[22]

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.

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 (even if Z is), and need not be centered at zero. Moreover, the nonlinear transformation $g'_{\theta_0}(Z+Z_n)-g'_{\theta_0}(Z_n)$ is different from the asymptotic distribution of the plug-in estimator $g(\widehat{\theta}_n)$ which is simply $g'_{\theta_0}(Z)$.[23] Thus, one can say that for directionally differentiable functions

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

Even though we have a closed-form expression for the bootstrap and posterior large-sample approximation, we cannot directly apply the results in Romano and Shaikh (2010) to establish the closeness of quantiles. The reason is that the large-sample approximation still depends on the data. In the appendix, we show that a (Lipschitz) continuity restriction on the c.d.f of the limiting distribution can be used to verify the high-level assumption of Theorem 2.[24]

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 $X^n=(X_1,\ldots,X_n)$ 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 1–3. Then, the bootstrap distribution for $g(\widehat{\theta}_n)$ and posterior distributions of $g(\theta)$ are explicitly computed and compared.

Relation to main assumptions: The transformation g is Lipschitz continuous and differentiable everywhere, except at $\theta_0=0$. 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.

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.

[22] For the sake of completeness, Lemma 5 in Appendix B.2 shows that if Assumptions 1–3 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.

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

[24] See Assumption 4 in Appendix B.2 for the details. This additional assumption is enough to establish the closeness of bootstrap/posterior quantiles as required by Theorem 2 under the assumption that g is directionally differentiable.

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.[25]

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.

Asymptotic Behavior of Posterior/Bootstrap Inference for $g(\theta) = |\theta|$: Since Assumptions 1–3 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 $g(\hat{\theta}_n)$ (after appropriate centering and scaling).

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.1 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.

[25] 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

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

Therefore:

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

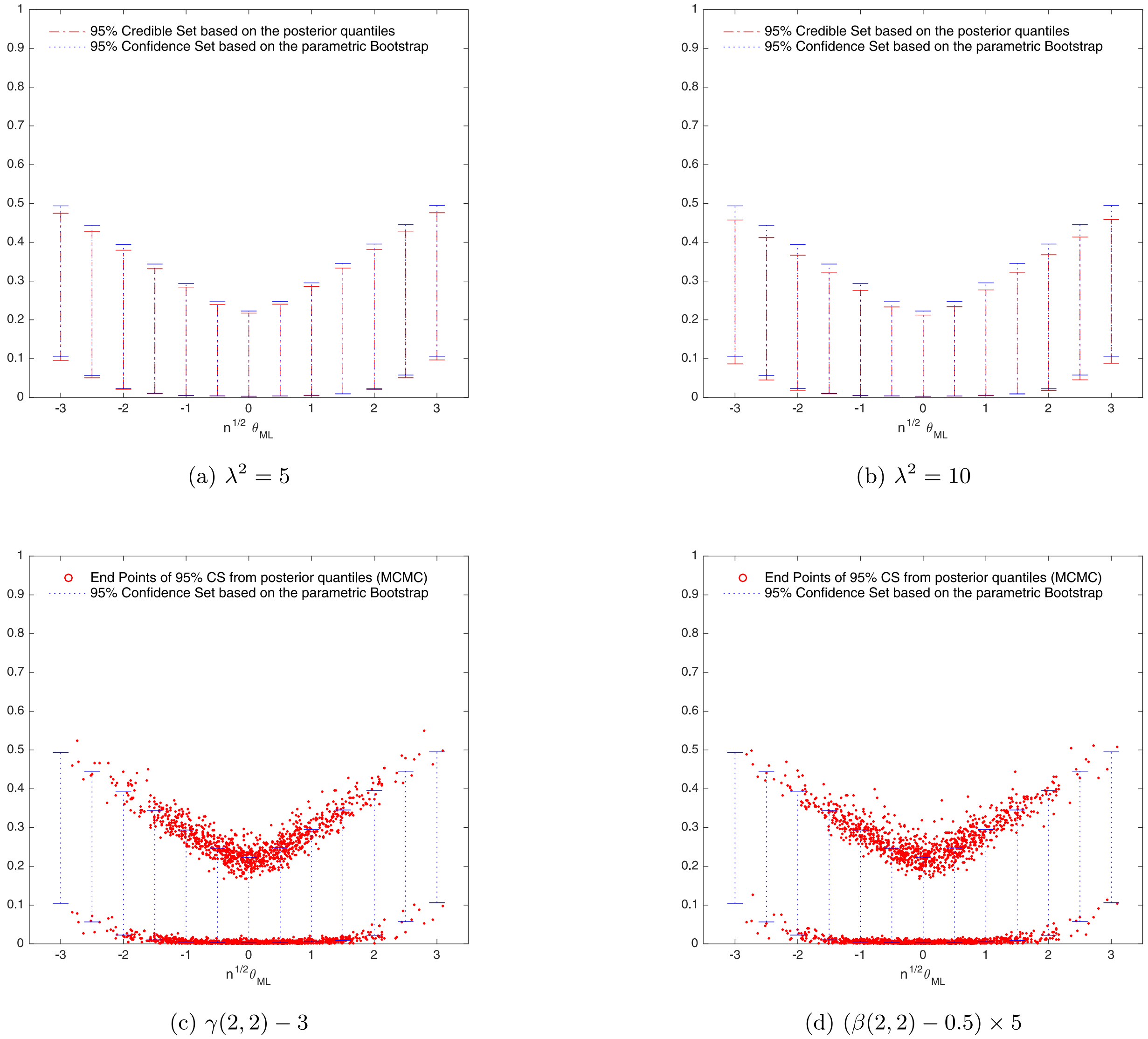
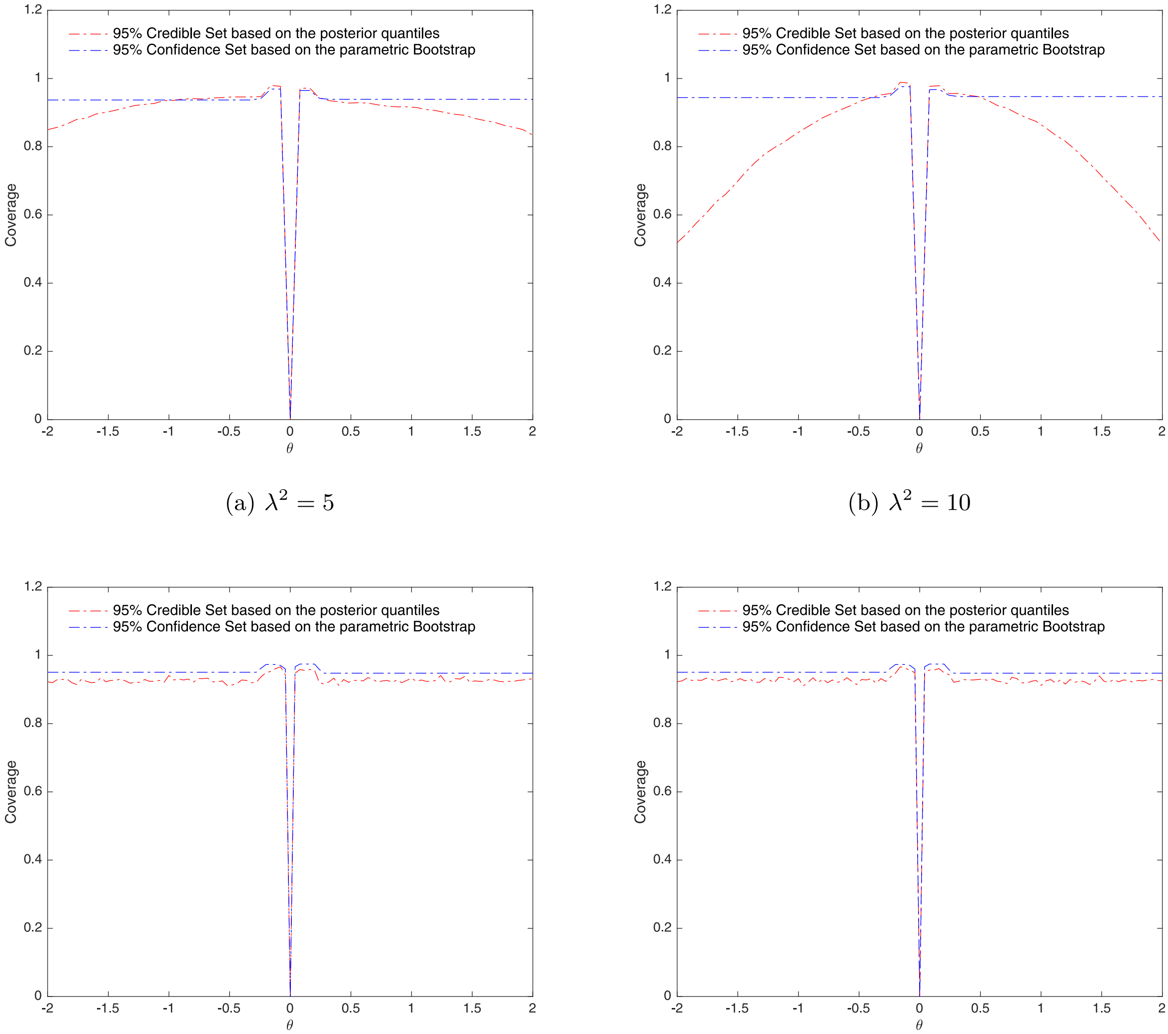


Fig. 1. 95% Credible Intervals for $|\theta|$ and 95% Parametric Bootstrap Confidence Intervals. Description of Fig. 1: 95% Credible intervals for $|\theta|$ obtained from four different priors and evaluated at different realizations of the data ($n = 100$). (Blue, Dotted Line) 95% confidence intervals 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 1000 possible data sets from a standard normal model). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Observe from (3.2) that the approximation to the centered and scaled posterior and bootstrap distributions depends on the data via $\sqrt{n}\widehat{\theta}_n$. Thus, in Fig. 1 we report the 95% credible and confidence intervals for data realizations $\sqrt{n}\widehat{\theta}_n \in [-3, 3]$. 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 Fig. 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 et al. (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



(a) $\lambda^2 = 5$ (b) $\lambda^2 = 10$

(c) $\gamma(2,2) - 3$ (d) $(\beta(2,2) - 0.5) \times 5$

Fig. 2. Coverage Probability of 95% Credible Intervals and Parametric Bootstrap Confidence Intervals for $|\theta|$. Description of Fig. 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to consider one-sided credible intervals or Highest Posterior Density intervals. In Section C of the online supplementary materials 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.

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" (Friedman et al., 2017, 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.

Appendix 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.*

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

Proof. Define

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

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

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

Taking expectation on both sides

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

Consequently,

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

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 1–3 imply that the assumptions of Lemma 1 are verified for both θ_n^{P*} and θ_n^{B*}. Define

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

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 Lipschitz 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,

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

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:

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

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

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

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

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

where

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

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

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

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

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

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

From the definition of F_n (as extension of Δ_n)

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

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

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 converges 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:

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

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

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:

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

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

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

This implies

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

and

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

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

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

Consequently:

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

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

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

which implies that

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

This implies that if the bootstrap fails at θ by at least $100d_{\alpha}\%$ 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 $(1-\alpha-\epsilon)$.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.10.009.

References

Andrews, D.W., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68 (2), 399–405.
Balke, A., Pearl, J., 1997. Bounds on treatment effects from studies with imperfect compliance. J. Amer. Statist. Assoc. 92 (439), 1171–1176.

Beran, R., 1997. Diagnosing bootstrap success. Ann. Inst. Statist. Math. 49 (1), 1–24.
Bochkina, N.A., Green, P.J., 2014. The Bernstein–Von Mises theorem and nonregular models. Ann. Statist. 42, 1850–1878.
Castillo, I., Rousseau, J., 2015. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. Ann. Statist. 43 (6), 2353–2383.
Chen, X., Christensen, T.M., Tamer, E., 2018a. Monte Carlo confidence sets for identified sets. Econometrica 86 (6), 1965–2018.
Chen, X., Christensen, T.M., Tamer, E., 2018b. Supplement to Monte Carlo confidence sets for identified sets. Econometrica 86 (6), 1965–2018.
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. Probab. Theory Related Fields 95 (1), 125–140.
Eaton, M.L., Tyler, D.E., 1991. On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. Ann. Statist. 260–271.
Evans, L.C., Gariepy, R.F., 2015. Measure Theory and Fine Properties of Functions. CRC Press.
Fang, Z., Santos, A., 2019. Inference on directionally differentiable functions. Rev. Econom. Stud. 86, 377–412.
Faust, J., 1998. The robustness of identified VAR conclusions about money. In: Carnegie-Rochester Conference Series on Public Policy, vol. 49, Elsevier, pp. 207–244.
Fiacco, A.V., Ishizuka, Y., 1990. Sensitivity and stability analysis for nonlinear programming. Ann. Oper. Res. 27 (1), 215–235.
Fieller, E.C., 1954. Some problems in interval estimation. J. R. Stat. Soc. Ser. B Stat. Methodol. 175–185.
Friedman, J., Hastie, T., Tibshirani, R., 2017. The Elements of Statistical Learning: Data Mining, Inference and Prediction, second ed. In: Series in Statistics, vol. 1, Springer, New York.
Gafarov, B., Meier, M., Montiel Olea, J.L., 2018. Delta-method inference for a class of set-identified SVARs. J. Econometrics 203, 316–327.
Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics, Vol. 537. John Wiley & Sons.
Ghosal, S., Ghosh, J.K., Samanta, T., 1995. On convergence of posterior distributions. Ann. Statist. 2145–2152.
Giacomini, R., Kitagawa, T., 2018. Robust Bayesian inference for set-identified models. Cemmap Working Paper 61/18.
Haile, P.A., Tamer, E., 2003. Inference with an incomplete model of English auctions. J. Polit. Econ. 111 (1), 1–51.
Hirano, K., Porter, J.R., 2012. Impossibility results for nondifferentiable functionals. Econometrica 80 (4), 1769–1790.
Hong, H., Li, J., 2018. The numerical delta-method. J. Econometrics 206, 379–394.
Horowitz, J.L., 2001. The bootstrap. In: Handbook of Econometrics, Vol. 5. Elsevier, pp. 3159–3228.
Kline, B., Tamer, E., 2016. Bayesian Inference in a class of Partially Identified Models. Quant. Econ. 7.
Kosorok, M.R., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer Science & Business Media.
Liu, Y., Gelman, A., Zheng, T., 2015. Simulation-efficient shortest probability intervals. Stat. Comput. 25, 809–819.
Lo, A.Y., 1987. A large sample study of the Bayesian bootstrap. Ann. Statist. 360–375.
Manski, C.F., 1990. Nonparametric bounds on treatment effects. Am. Econ. Rev. 80 (2), 319–323.
Manski, C.F., 2004. Statistical treatment rules for heterogeneous populations. Econometrica 72 (4), 1221–1246.
McShane, E.J., 1934. Extension of range of functions. Bull. Amer. Math. Soc. 40.
Milgrom, P., Segal, I., 2002. Envelope theorems for arbitrary choice sets. Econometrica 70 (2), 583–601.
Morand, O., Reffett, K., Tarafdar, S., 2015. A nonsmooth approach to envelope theorems. J. Math. Econom. 61, 157–165.
Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Handbook of Econometrics. Elsevier, pp. 2111–2245.
Romano, J.P., Shaikh, A.M., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78 (1), 169–211.
Shapiro, A., 1990. On concepts of directional differentiability. J. Optim. Theory Appl. 66 (3), 477–487.
Shapiro, A., 1991. Asymptotic analysis of stochastic programs. Ann. Oper. Res. 30 (1), 169–186.
Van der Vaart, A., 2000. Asymptotic Statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
Van der Vaart, A., Wellner, J., 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.