The Annals of Statistics
2020, Vol. 48, No. 1, 397–412
https://doi.org/10.1214/19-AOS1812

THE NUMERICAL BOOTSTRAP

BY HAN HONG[1] AND JESSIE LI[2]

[1]*Department of Economics, Stanford University, hanhong@stanford.edu*
[2]*Department of Economics, University of California, Santa Cruz, jeqli@ucsc.edu*

This paper proposes a numerical bootstrap method that is consistent in many cases where the standard bootstrap is known to fail and where the m-out-of-n bootstrap and subsampling have been the most commonly used inference approaches. We provide asymptotic analysis under both fixed and drifting parameter sequences, and we compare the approximation error of the numerical bootstrap with that of the m-out-of-n bootstrap and subsampling. Finally, we discuss applications of the numerical bootstrap, such as constrained and unconstrained M-estimators converging at both regular and nonstandard rates, Laplace-type estimators, and test statistics for partially identified models.

1. Introduction. We propose a new type of bootstrap called the numerical bootstrap which offers an alternative to the m-out-of-n bootstrap [6, 30] and subsampling [25] in many cases where the standard bootstrap fails. Motivated by [16]'s work on inference for directionally differentiable functions, the numerical bootstrap is based on perturbing the sample by $\epsilon_n\sqrt{n}$ times the difference between the bootstrapped sample and the data. We show that when $\epsilon_n\sqrt{n} \to \infty$ and $\epsilon_n \downarrow 0$, the numerical bootstrap can be used to conduct pointwise asymptotically valid inference for a large class of M-estimators converging at possibly slower than $\sqrt{n}$ rates and subject to a set of known constraints which can be approximated in the limit by a cone centered at the true parameter value.

Section 2 provides an overview of the numerical bootstrap method. Section 2.1 contains some heuristic arguments comparing the error of the numerical bootstrap to that of the m-out-of-n bootstrap and subsampling. Section 3 studies the asymptotic coverage properties of confidence intervals constructed using the numerical bootstrap for drifting sequences of parameters. Section 4 validates the consistency of the numerical bootstrap for a class of M-estimators that includes the maximum score estimator developed by [23] and whose asymptotics are derived in [20] and [11]. In Section 4.1, we allow the true parameter to lie on the boundary of a constrained set, as in the setup of [17]. For the sample extremum counter example in Section 4.2, subsampling works, but the numerical bootstrap does not. Section 5 reports Monte Carlo simulation results comparing the numerical bootstrap to the standard bootstrap, the perturbation bootstrap [14, 24], and the m-out-of-n bootstrap [6, 30]. The Supplementary Material [18] contains more theoretical and simulation results on the differences between the numerical bootstrap, the m-out-of-n bootstrap, and subsampling. Local asymptotics and simulations results are presented for the LASSO estimator [33] in the one-dimensional mean model. We also demonstrate how to consistently estimate the asymptotic distribution of sample size dependent statistics such as the Laplace-type estimators of [13] and [19]. Additionally, we illustrate how the numerical bootstrap can be used to perform hypothesis testing in partially identified moment inequality models (see, e.g., [2, 3, 8, 9]). We also discuss the role of recentering in hypothesis testing and how to use the numerical bootstrap to estimate an unknown polynomial convergence rate. A list of commonly used symbols and proofs of the theorems are also included in the Supplementary Material.

Received June 2017; revised November 2018.
MSC2010 subject classifications. 62F40.
Key words and phrases. Bootstrap, numerical differentiation, directional differentiability.

2. A generalized numerical bootstrap method. To motivate, we note that many estimators and test statistics can be written as a functional of the empirical distribution $\theta(P_n)$ with a population analog $\theta(P)$. Typically, for an increasing function $a(n)$ of the sample size n, for a limiting distribution $\mathcal{J}$ (which can depend on P), and using weak convergence notation,[1] $\hat{\mathcal{J}}_n \equiv a(n)(\theta(P_n) - \theta(P)) \rightsquigarrow \mathcal{J}$. This can be rewritten as

$$\hat{\mathcal{J}}_n \equiv a(n)\left(\theta\left(P + \frac{1}{\sqrt{n}}\sqrt{n}(P_n - P)\right) - \theta(P)\right) \rightsquigarrow \mathcal{J}.$$

Since it is often times the case that $\hat{\mathcal{G}}_n = \sqrt{n}(P_n - P) \rightsquigarrow \mathcal{G}_0$ where $\mathcal{G}_0$ is a properly defined Brownian bridge, we also expect that

$$a(n)\left(\theta\left(P + \frac{1}{\sqrt{n}}\mathcal{G}_0\right) - \theta(P)\right) \rightsquigarrow \mathcal{J}.$$

If we take $\epsilon_n = \frac{1}{\sqrt{n}}$, so that $a(n) = a\left(\sqrt{n}^2\right)$ is replaced by $a\left(\frac{1}{\epsilon_n^2}\right)$, then we also anticipate that for other $\epsilon_n \downarrow 0$,

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\theta(P + \epsilon_n \mathcal{G}_0) - \theta(P)\right) \rightsquigarrow \mathcal{J}.$$

The goal is to provide a consistent estimate of $\mathcal{J}$, which approximates the left-hand side above. To obtain such a consistent estimate, we need to estimate the unknown P and $\mathcal{G}_0$. Intuitively, P can be estimated by P_n, and $\mathcal{G}_0$ can be consistently estimated by the bootstrapped empirical process $\hat{\mathcal{G}}_n^* = \sqrt{n}(P_n^* - P_n)$. A popular choice for $\hat{\mathcal{G}}_n^*$ is the multinomial bootstrap in which $\hat{\mathcal{G}}_n^* = \sqrt{n}(P_n^* - P_n)$ and $P_n^* = \frac{1}{n}\sum_{i=1}^n M_{ni}\delta_i$, where δ_i is the point mass on observation i, and $M_{ni}, i = 1, \dots, n$ is a multinomial distribution with parameters $(n^{-1}, n^{-1}, \dots, n^{-1})$. Other common choices for $\hat{\mathcal{G}}_n^*$ include the Wild bootstrap, where $\hat{\mathcal{G}}_n^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i - \bar{\xi})\delta_i$ for $\bar{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$ and ξ_i are i.i.d. variables with variance 1 and finite 3rd moment, and exchangeable bootstrap schemes in [34] (Chapter 3.6). Other forms of $\hat{\mathcal{G}}_n^*$ that consistently estimate $\mathcal{G}_0$ can also be used, such as $\hat{\mathcal{G}}_n^* = \sqrt{m_n}(P_{m_n}^* - P_n)$ where $m_n \to \infty$ as $n \to \infty$ and $P_{m_n}^*$ is a multinomial i.i.d. sample from P_n of size m_n. A choice of $m_n/n \to 0$ and $\epsilon_n = 1/\sqrt{m_n}$ corresponds to the m-out-of-n bootstrap. Convolved subsampling (e.g., [32]) can be used to handle time series data, but we focus on the i.i.d. case.

Under regularity conditions, $\hat{\mathcal{G}}_n^*$ converges in distribution to $\mathcal{G}_1$ both conditionally on the sample in probability, and unconditionally, where $\mathcal{G}_1$ is an independent and identical copy of $\mathcal{G}_0$. To offset the noise of estimating P with P_n, the step size parameter ϵ_n is chosen such that $\sqrt{n}\epsilon_n \to \infty$. Therefore, we propose a numerical bootstrap method that estimates $\mathcal{J}$ with

$$\hat{\mathcal{J}}_n^* = a\left(\frac{1}{\epsilon_n^2}\right)\left(\theta\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*\right) - \theta(P_n)\right).$$

To see why the numerical bootstrap might work, note that

$$\hat{\mathcal{J}}_n^* = a\left(\frac{1}{\epsilon_n^2}\right)\left(\theta\left(P + \epsilon_n\left(\hat{\mathcal{G}}_n^* + \frac{P_n - P}{\epsilon_n}\right)\right) - \theta(P)\right) - a\left(\frac{1}{\epsilon_n^2}\right)\left(\theta(P_n) - \theta(P)\right).$$

In the above, we rewrite the second term as

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\theta(P_n) - \theta(P)\right) = \frac{1}{a(n)} a\left(\frac{1}{\epsilon_n^2}\right) a(n)\left(\theta(P_n) - \theta(P)\right).$$

[1] $X_n \rightsquigarrow X$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in \mathrm{BL}_1} |E^* f(X_n) - E f(X)| \to 0$ where BL_1 is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1.

Since $a(n)(\theta(P_n)-\theta(P)) \rightsquigarrow \mathcal{J}$ and typically $\frac{1}{a(n)}a(\frac{1}{\epsilon_n^2}) \to 0$ (e.g., when $a(n)=n^\gamma$) as $n\epsilon_n^2 \to \infty$, the second term vanishes asymptotically:

$$a\left(\frac{1}{\epsilon_n^2}\right)\big(\theta(P_n)-\theta(P)\big)=o_P(1).$$

Using conditional weak convergence notation,[2] $\hat{\mathcal{G}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{G}_1$ in the first term of $\hat{\mathcal{J}}_n^*$. Additionally, since $\sqrt{n}\epsilon_n \to \infty$, heuristically we expect that

$$\frac{P_n-P}{\epsilon_n}=\frac{\sqrt{n}(P_n-P)}{\sqrt{n}\epsilon_n}\approx\frac{\mathcal{G}_0}{\sqrt{n}\epsilon_n}\xrightarrow{p}0.$$

Therefore, since $\mathcal{G}_1$ has the same distribution as $\mathcal{G}_0$, we also expect that

$$\hat{\mathcal{J}}_n^* \approx a\left(\frac{1}{\epsilon_n^2}\right)\big(\theta(P+\epsilon_n\mathcal{G}_1)-\theta(P)\big) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}.$$

Note that [15]'s rescaled bootstrap is a special case of the numerical bootstrap for estimators that satisfy $a(n)=\sqrt{n}$.

2.1. *Comparison of numerical bootstrap with m-out-of-n bootstrap and subsampling*. In situations where m-out-of-n bootstrap, subsampling, and the numerical bootstrap method can be used, the numerical bootstrap can potentially offer a more accurate approximation to the limiting distribution. Because the analysis is similar between subsampling and m-out-of-n bootstrap, for brevity we focus on subsampling. Recall that subsampling [25] approximates the limiting distribution $\mathcal{J}$ using the finite sample distribution of $a(b)(\theta(P_b)-\theta(P_n))$ which in large samples is close to $a(b)(\theta(P_b)-\theta(P))$ whenever $a(b)(\theta(P_n)-\theta(P))=o_P(1)$. In turn, as $b\to\infty$, $a(b)(\theta(P_b)-\theta(P)) \rightsquigarrow \mathcal{J}$. To compare subsampling to the numerical bootstrap, write the subsampling distribution as

$$a(b)\big(\theta(P_b)-\theta(P_n)\big)=a(b)\left(\theta\left(P_n+\frac{1}{\sqrt{b}}\sqrt{b}(P_b-P_n)\right)-\theta(P_n)\right).$$

In the numerical bootstrap setup, subsampling is essentially using $\epsilon_n=\frac{1}{\sqrt{b}}$ as the step size and using $\sqrt{b}(P_b-P_n)$ to estimate $\mathcal{G}_0$ based on subsamples of size b. The numerical bootstrap method is different and instead uses $\hat{\mathcal{G}}_n^* \equiv \sqrt{n}(P_n^*-P_n)$ to estimate $\mathcal{G}_0$ based on the entire sample of size n. In addition, $\mathcal{G}_0$ can also be approximated by a multivariate normal distribution in finite dimensional situations.

For $X_i \overset{i.i.d.}{\sim} (\mu(P),\sigma^2)$ and $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^n X_i$, consider the finite dimensional setup where $\theta(P)=\phi(\mu(P))$ for some finite dimensional Hadamard directionally differentiable mapping $\phi:\mathbb{R}^d\to\mathbb{R}$. Recall the following definition of first order Hadamard directional differentiability:

DEFINITION 2.1. ϕ is first order Hadamard directionally differentiable at $\mu_0\equiv\mu(P)\in\mathbb{R}^d$ tangentially to a set $\mathbb{D}_0\subseteq\mathbb{R}^d$ if there is a continuous map $\phi'_{\mu_0}:\mathbb{D}_0\to\mathbb{R}$ such that for all $h\in\mathbb{D}_0$,

$$\lim_{n\to\infty}\left|\frac{\phi(\mu_0+t_nh_n)-\phi(\mu_0)}{t_n}-\phi'_{\mu_0}(h)\right|=0$$

for all $\{h_n\}\subset\mathbb{D}$ and $\{t_n\}\in\mathbb{R}_+$ such that $t_n\downarrow 0$, $h_n\to h$ as $n\to\infty$ and $\mu_0+t_nh_n\in\mathbb{R}^d$.

[2] $X_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} X$ in the metric space $(\mathbb{D},d)$ if and only if $\sup_{f\in \mathrm{BL}_1}|E_{\mathbb{W}}f(X_n)-Ef(X)|\to 0$ and $E_{\mathbb{W}}f(X_n)^*-E_{\mathbb{W}}f(X_n)_*\xrightarrow{p}$ for all $f\in\mathrm{BL}_1$, where BL_1 is the space of functions $f:\mathbb{D}\mapsto\mathbb{R}$ with Lipschitz norm bounded by 1 and $E_{\mathbb{W}}$ denotes expectation with respect to the bootstrap weights $\mathbb{W}$ conditional on the data $\mathcal{X}_n$.

When the first order Hadamard directional derivative is degenerate, that is, $\phi'_{\mu_0}(h) = 0$ for all h, it will be necessary to assume second order Hadamard directional differentiability.

DEFINITION 2.2. ϕ is second order Hadamard directionally differentiable at $\mu_0 \in \mathbb{R}^d$ tangentially to $\mathbb{D}_0$ if it is first order Hadamard directionally differentiable and there is a continuous map $\phi''_{\mu_0} : \mathbb{D}_0 \to \mathbb{R}$ such that for all $h \in \mathbb{D}_0$,

$$\lim_{n\to\infty}\left|\frac{\phi(\mu_0 + t_n h_n) - \phi(\mu_0) - t_n \phi'_{\mu_0}(h_n)}{\frac{1}{2}t_n^2} - \phi''_{\mu_0}(h)\right| = 0$$

for all $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \in \mathbb{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$ and $\mu_0 + t_n h_n \in \mathbb{R}^d$.

Consider approximating the limiting distribution of $\sqrt{n}(\phi(\bar{X}_n) - \phi(\mu))$ for any twice Hadamard directionally differentiable function $\phi(\cdot)$. It is known that $\phi'_\mu(h)$ is positively homogeneous of degree 1. We demonstrate in the Supplementary Material that one dimensional positively homogeneous functions of degree 1 have a piecewise linear representation: there exists constants $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\phi'_\mu(h) = \lambda_1 h^+ + \lambda_2 h^-$. Using Taylor expansion arguments detailed in the Supplementary Material, for $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$,

$$P(\hat{\mathcal{J}}_n^* \leq x|\mathcal{X}_n) = \Phi\left(\frac{x}{\lambda_1}\right) + \Phi\left(\frac{x}{\lambda_2}\right) - 1 + O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p(\epsilon_n).$$

In particular, when the second order directional derivative is nonzero and $\phi'_\mu(\cdot)$ is not a linear function, then the error for the numerical bootstrap is $O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p(\epsilon_n) = O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right) + O_p(\epsilon_n)$. The optimal choice of ϵ_n that balances the two terms satisfies $\epsilon_n = O(n^{-1/4})$, leading to an error on the order of $n^{-1/4}$. The error for subsampling is $O_p\left(\sqrt{\frac{b}{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{b}}\right)$, so the optimal choice of b satisfies $b = O\left(n^{1/2}\right)$, which also leads to an error on the order of $n^{-1/4}$.

If however, $\phi'_\mu(\cdot)$ is a linear function that is not degenerate at μ, then the error for the numerical bootstrap is $O_p\left(\frac{1}{\sqrt{n}}\right) + O_p(\epsilon_n)$, and is minimized by $\epsilon_n = O\left(\frac{1}{\sqrt{n}}\right)$. In contrast, subsampling's error would still be $O_p(n^{-1/4})$ because of the additional error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ introduced by estimating the distribution of $\sqrt{b}(\mu(P_b) - \mu(P_n))$ using the empirical distribution of $\sqrt{b}(\mu(P_{b,i}) - \mu(P_n))$ over $i = 1, \ldots, \binom{n}{b}$ sub-blocks. The presence of the error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ is implied by Lemma A.2 in [27] and also demonstrated in Theorem 1 of [5] and Theorem 3 of [4]. Finally, if the second order derivative is zero, then the error for the numerical bootstrap is $O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$, and is smaller than $O_p\left(n^{-1/4}\right)$ for all values of ϵ_n satisfying $\sqrt{n}\epsilon_n \to \infty$ while subsampling's error would still be $O_p\left(n^{-1/4}\right)$ due to the error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ when estimating the distribution of $\sqrt{b}\,(\mu(P_b) - \mu(P))$. Therefore, the numerical bootstrap should not have an error that is of larger order than subsampling and it may outperform subsampling in some situations when the first derivative is linear and the second order derivative is nonzero, or when the second order derivative is zero.

3. Local analysis. Consider the finite dimensional setup where $\theta(P) = \phi(\mu(P))$ for some finite dimensional mapping $\phi : \mathbb{R}^d \to \mathbb{R}$ that is Hadamard directionally differentiable at $\mu(P)$ tangentially to $\mathbb{D}_0 \subseteq \mathbb{R}^d$. Suppose we consider perturbing the data generating process P so that we perform asymptotic analysis on drifting sequences of parameters given by

$$\mu(P^n) - \mu(P) = a_n c,$$

where $a_n \downarrow 0$ is the rate of drift and c is the slackness parameter. Let $\hat{\mu}_n$ be a $\sqrt{n}$-consistent estimator for $\mu_n = \mu(P^n)$ and $\hat{\mu}_n^*$ its bootstrapped version.

ASSUMPTION 3.1. For $r_n \uparrow \infty$ and some tight limiting distribution $\mathbb{G}_0$ supported on $\mathbb{D}_0$,

$$\sqrt{n}(\hat{\mu}_n - \mu_n) \rightsquigarrow \mathbb{G}_0, \quad \sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_0.$$

We first consider statistics $\hat{\mathcal{J}}_n \equiv \sqrt{n}(\phi(\hat{\mu}_n) - \phi(\mu_n))$ that have the same rate of convergence as $\hat{\mu}_n$. Define $\hat{c}_\alpha^*$ to be the α-th quantile of $\hat{\mathcal{J}}_n^* \equiv \frac{1}{\epsilon_n}(\phi(\hat{\mu}_n + \epsilon_n\sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n)) - \phi(\hat{\mu}_n))$. In the following theorem, we describe the coverage properties under drifting sequences for the following three kinds of confidence intervals: equal-tailed $\left[\phi(\hat{\mu}_n) - \frac{\hat{c}_{1-\alpha}^*}{\sqrt{n}}, \phi(\hat{\mu}_n) - \frac{\hat{c}_\alpha^*}{\sqrt{n}}\right]$, lower $\left[\phi(\hat{\mu}_n) - \frac{\hat{c}_{1-\alpha}^*}{\sqrt{n}}, \infty\right)$, and upper $\left(-\infty, \phi(\hat{\mu}_n) - \frac{\hat{c}_\alpha^*}{\sqrt{n}}\right]$.

THEOREM 3.1. *Let $\phi : \mathbb{D}_\phi \mapsto \mathbb{R}$ be a Hadamard directionally differentiable function at μ_0. Let $\hat{\mu}_n$ and $\hat{\mu}_n^*$ satisfy assumption 3.1. If ϕ'_{μ_0} is linear, then equal-tailed and one-sided confidence intervals are asymptotically exact for all $a_n \downarrow 0$. If ϕ'_{μ_0} is nonlinear and subadditive* (*superadditive*), *the lower* (*upper*) *confidence interval will be conservatively valid for the following types of sequences*: (i) $a_n\sqrt{n} = 1$, (ii) $a_n\sqrt{n} \to \infty$ *and* $a_n/\epsilon_n \to 0$. *Equal-tailed and one-sided intervals are asymptotically exact for* (i) $a_n\sqrt{n} \to 0$ (ii) $a_n\sqrt{n} \to \infty$ *and* $a_n/\epsilon_n \to \infty$.

The appendix in the Supplementary Material includes the proof of theorem 3.1 and a discussion of local asymptotics for the negative part of the mean example and for LASSO in the one-dimensional mean model.

It is not surprising that the numerical bootstrap consistently estimates the limiting distribution of $\sqrt{n}(\phi(\hat{\mu}_n) - \phi(\mu_n))$ when ϕ'_{μ_0} is linear because linearity of ϕ'_{μ_0} amounts to Hadamard differentiability (as opposed to directional differentiability) of ϕ. It is known that the standard bootstrap is consistent when ϕ is Hadamard differentiable (see Theorem 3.9.11 in [34]), so it should be the case that the numerical bootstrap is consistent as well. This property of sharing the same asymptotic distribution as the standard bootstrap when the standard bootstrap is consistent also applies to other bootstrap methods in the literature such as bootstrap bounding methods [12, 22] and adaptive projection intervals [26].

4. Consistency of numerical bootstrap for M-estimators. In this section, we demonstrate the asymptotic consistency of the numerical bootstrap for a class of M-estimators $\hat{\theta}_n$ that converge at rate n^γ for some $\gamma \in (\frac{1}{4}, 1)$. Our proofs in this section assume that the researcher knows γ, but in practice, we can estimate an unknown γ using methods described in the appendix in the Supplementary Material. Consider

$$\hat{\theta}_n \equiv \underset{\theta \in \Theta}{\arg\max} P_n \pi(\cdot, \theta) = \frac{1}{n}\sum_{i=1}^{n} \pi(z_i, \theta).$$

We approximate the limiting distribution of $n^\gamma\left(\hat{\theta}_n - \theta_0\right)$ using the finite sample distribution of $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, where $\hat{\theta}_n^* \equiv \arg\max_{\theta\in\Theta} \mathcal{Z}_n^* \pi(\cdot, \theta)$, and $\mathcal{Z}_n^* = P_n + \epsilon_n \hat{\mathcal{G}}_n^*$ is a linear combination between the empirical distribution and the bootstrapped empirical process. For example, when $\hat{\mathcal{G}}_n^*$ is the multinomial bootstrap, for each bootstrap sample $z_i^*, i = 1, \ldots, n$,

$$\hat{\theta}_n^* = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \pi(z_i, \theta) + \epsilon_n \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\left(\pi\left(z_i^*, \theta\right) - \pi(z_i, \theta)\right).$$

On the other hand, when $\hat{\mathcal{G}}_n^*$ is the Wild bootstrap,

$$\hat{\theta}_n^* = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \pi(z_i, \theta) + \epsilon_n \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)\pi(z_i, \theta).$$

In the following theorem, we show that for a suitable choice of the step size ϵ_n, $n^\gamma\left(\hat{\theta}_n - \theta_0\right)$ and $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$ converge to the same limiting distribution for a large class of estimators that includes the typical $\sqrt{n}$ consistent estimators like OLS and IV as well as $n^{1/3}$ consistent estimators like the maximum score estimator studied in [20, 23] and [1]. Other valid bootstrap methods for the maximum score estimator, such as [29], are available in the literature. Recently, [10] propose to bootstrap the Gaussian process and estimate the Hessian term in the quadratic limit separately in the context of M-estimation. Let $X_n^* = o_P^*(1)$ if the law of X_n^* is governed by P_n and if $P_n(|X_n^*| > \epsilon) = o_P(1)$ for all $\epsilon > 0$. Also define $M_n^* = O_p^*(1)$ (hence also $O_p(1)$) if $\lim_{m\to\infty}\limsup_{n\to\infty} P(P_n(M_n^* > m) > \epsilon) \to 0, \forall\epsilon > 0$.

THEOREM 4.1 (Consistency of Numerical Bootstrap for M-estimators). *Define* $g(\cdot, \theta) \equiv \pi(\cdot, \theta) - \pi(\cdot, \theta_0)$. *Suppose the following conditions are satisfied for some* $\rho \in (0, 3/2)$ *and for* $\gamma \equiv \frac{1}{2(2-\rho)}$:

(i) $P_n g\left(\cdot, \hat{\theta}_n\right) \geq \sup_{\theta\in\Theta} P_n g(\cdot, \theta) - o_P\left(n^{-2\gamma}\right)$ *and* $\mathcal{Z}_n^* g\left(\cdot, \hat{\theta}_n^*\right) \geq \sup_{\theta\in\Theta} \mathcal{Z}_n^* g(\cdot, \theta) - o_P^*\left(\epsilon_n^{4\gamma}\right)$.

(ii) $\hat{\theta}_n \xrightarrow{p} \theta_0$ *and* $\hat{\theta}_n^* - \hat{\theta}_n = o_P^*(1)$.

(iii) θ_0 *is an interior point of* $\Theta \in \mathbb{R}^d$.

(iv) *The class of functions* $\mathcal{G}_R = \{g(\cdot, \theta) : |\theta - \theta_0| \leq R\}$ *is uniformly manageable with envelope function* $G_R(\cdot) \equiv \sup_{g\in\mathcal{G}_R} |g(\cdot)|$.

(v) $\mathrm{Pg}(\cdot, \theta)$ *is twice differentiable at* θ_0 *with negative definite Hessian matrix* $-H$.

(vi) $\Sigma_\rho(s, t) = \lim_{\alpha\to\infty} \alpha^{2\rho}\,\mathrm{Pg}\left(\cdot, \theta_0 + \frac{s}{\alpha}\right) g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)$ *exists for each* s, t *in* $\mathbb{R}^d$.

(vii) $\lim_{\alpha\to\infty} \alpha^{2\rho}\,\mathrm{Pg}\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)^2 1\left(\left|g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)\right| > \epsilon\alpha^{2(1-\rho)}\right) = 0$ *for each* $\epsilon > 0$ *and* $t \in \mathbb{R}^d$.

(viii) *There exists a* $R_0 > 0$ *such that* $\mathrm{PG}_R^2 = O\left(R^{2\rho}\right)$ *for all* $R \leq R_0$.

(ix) $\sqrt{n}\epsilon_n \to \infty$ *and* $\epsilon_n \downarrow 0$.

(x) *For some* $\eta > 0$, *there exists a* K *such that* $\mathrm{PG}_R^2 1(G_R > K) < \eta R^{2\rho}$ *for* $R \to 0$.

(xi) $P|g(\cdot, \theta_1) - g(\cdot, \theta_2)| = O\left(|\theta_1 - \theta_2|^{2\rho}\right)$ *for* $|\theta_1 - \theta_2| \to 0$.

Then $\hat{\theta}_n - \theta_0 = O_p\left(n^{-\gamma}\right)$ *and* $\hat{\theta}_n^* - \theta_0 = O_p^*\left(\epsilon_n^{2\gamma}\right)$. *Furthermore, for* $\mathcal{Z}_0(h)$ *a mean zero Gaussian process with covariance kernel* Σ_ρ *and nondegenerate increments*,

$$\hat{\mathcal{J}}_n \equiv n^\gamma\left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow \mathcal{J} \equiv \arg\max_h \mathcal{Z}_0(h) - \frac{1}{2}h'Hh,$$

$$\hat{\mathcal{J}}_n^* \equiv \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J} \quad \textit{and} \quad \hat{\mathcal{J}}_n^* \rightsquigarrow \mathcal{J}.$$

The assumptions above are modeled after [20] but generalized so that results for both the $\sqrt{n}$ and $n^{1/3}$ cases can be stated concisely.

To explain the intuition for the above theorem, note that for $\hat{h}_n = n^{\gamma}\left(\hat{\theta}_n - \theta_0\right)$,

$$\begin{aligned}\hat{h}_n &= \underset{h\in n^{\gamma}(\Theta-\theta_0)}{\arg\max}\; n^{2\gamma} P_n g\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right)\\ &= n^{2\gamma-\frac{1}{2}}\sqrt{n}(P_n-P)g\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right)+n^{2\gamma}Pg\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right).\end{aligned} \tag{4.1}$$

Under the stated conditions, $n^{2\gamma}Pg\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right)\to -\frac{1}{2}h'Hh$, and

$$n^{2\gamma-\frac{1}{2}}\sqrt{n}(P_n-P)g\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right)=n^{\rho\gamma}\mathcal{G}_n g\left(\cdot;\theta_0+\frac{h}{n^{\gamma}}\right)\rightsquigarrow \mathcal{Z}_0(h).$$

The numerical bootstrap seeks to approximate the limiting distribution $\mathcal{J}$ with the distribution of

$$\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\hat{\theta}_n\right)=\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\theta_0\right)-\epsilon_n^{-2\gamma}\left(\hat{\theta}_n-\theta_0\right),$$

which will be valid if (1) $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n-\theta_0\right)=o_p(1)$ and (2) $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\theta_0\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\mathcal{J}$. Part (1) follows from $\sqrt{n}\epsilon_n\to\infty$ since $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n-\theta_0\right)=\frac{1}{(\sqrt{n}\epsilon_n)^{2\gamma}}n^{\gamma}\left(\hat{\theta}_n-\theta_0\right)=o_P(1)$. For part (2), write $\mathcal{Z}_n^*g(\cdot,\theta)=(\mathcal{Z}_n^*-P)g(\cdot,\theta)+Pg(\cdot,\theta)$, so that

$$\hat{\theta}_n^*=\underset{\theta\in\Theta}{\arg\max}\,\mathcal{Z}_n^*g(\cdot,\theta)=\left(\mathcal{Z}_n^*-P\right)g(\cdot,\theta)-\frac{1}{2}(\theta-\theta_0)'\left(H+o_p(1)\right)(\theta-\theta_0).$$

For the first term, note that $(\mathcal{Z}_n^*-P)=\frac{1}{\sqrt{n}}\sqrt{n}(P_n-P)+\epsilon_n\hat{\mathcal{G}}_n^*\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\frac{1}{\sqrt{n}}\mathcal{G}_0+\epsilon_n\mathcal{G}_1$ where $\mathcal{G}_0$ and $\mathcal{G}_1$ are independent copies of the same Gaussian process. Since $\epsilon_n >> \frac{1}{\sqrt{n}}$, the second term should dominate, so that $(\mathcal{Z}_n^*-P)\approx\epsilon_n\mathcal{G}_1$. Consequently, we expect

$$\begin{aligned}\hat{\theta}_n^* &\approx \underset{\theta\in\Theta}{\arg\max}\,\epsilon_n\mathcal{G}_1 g(\cdot,\theta)-\frac{1}{2}(\theta-\theta_0)'H(\theta-\theta_0)\\ &=\epsilon_n O_p\left(|\theta-\theta_0|^{\rho}\right)-\frac{1}{2}(\theta-\theta_0)'H(\theta-\theta_0).\end{aligned}$$

By the definition of $\hat{\theta}_n^*$, $\epsilon_n O_p\left(|\hat{\theta}_n^*-\theta_0|^{\rho}\right)+\left(\hat{\theta}_n^*-\theta_0\right)'H\left(\hat{\theta}_n^*-\theta_0\right)\geq 0$, implying that $|\hat{\theta}_n^*-\theta_0|^{2-\rho}\leq O_p(\epsilon_n)$ and therefore $|\hat{\theta}_n^*-\theta_0|\leq O_p\left(\epsilon_n^{\frac{1}{2-\rho}}\right)=O_p\left(\epsilon_n^{2\gamma}\right)$. To be more formal, let $\hat{h}_n^*=\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\theta_0\right)$. Then

$$\hat{h}_n^*=\underset{h\in\epsilon_n^{-2\gamma}(\Theta-\theta_0)}{\arg\max}\;\epsilon_n^{-4\gamma}\left(\left(\mathcal{Z}_n^*-P\right)g\left(\cdot;\theta_0+\epsilon_n^{2\gamma}h\right)+Pg\left(\cdot;\theta_0+\epsilon_n^{2\gamma}h\right)\right).$$

The second term $\epsilon_n^{-4\gamma}Pg\left(\cdot;\theta_0+\epsilon_n^{2\gamma}h\right)\to-\frac{1}{2}h'Hh$. It is shown in the Appendix that the first term satisfies

$$\epsilon_n^{-4\gamma}\left(\mathcal{Z}_n^*-P\right)g\left(\cdot;\theta_0+\epsilon_n^{2\gamma}h\right)\approx\epsilon_n^{-4\gamma}\left(\frac{1}{\sqrt{n}}\mathcal{G}_0+\epsilon_n\mathcal{G}_1\right)g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)$$

and that for a suitable Gaussian process $\mathcal{Z}_0$ (as in [20]),

$$\epsilon_n^{-4\gamma}\left(\frac{1}{\sqrt{n}}\mathcal{G}_0+\epsilon_n\mathcal{G}_1\right)g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\approx\epsilon_n^{1-4\gamma}\left(\mathcal{G}_1 g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\mathcal{Z}_0(h).$$

Combining the first and second terms implies that $\hat{h}_n^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J} = \arg\max_h \mathcal{Z}_0(h) - \frac{1}{2}h'Hh$. Altogether, parts (1) and (2) imply that $\hat{\mathcal{J}}_n^* \equiv \frac{1}{\epsilon^{2\gamma}}(\hat{\theta}_n^* - \hat{\theta}_n) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J}$, which validates the consistency of the numerical bootstrap method.

In a more conventional approach such as [19], $\mathcal{J}$ is approximated by $\bar{\mathcal{J}}^* = \arg\max_h \hat{\mathcal{Z}}_0(h) - \frac{1}{2}h'\hat{H}h$ where $\hat{H} \xrightarrow{p} H$ and $\hat{\mathcal{Z}}_0(h)$ is a Gaussian process with estimated covariance kernel $\hat{\Sigma}_\rho(s,t)$ for $\hat{\Sigma}_\rho(s,t) \xrightarrow{p} \Sigma_\rho(s,t)$. Instead, the numerical bootstrap essentially replaces

$$\hat{\mathcal{Z}}_0(h) - \frac{1}{2}h'\hat{H}h \quad \text{with } \epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot, \hat{\theta}_n + \epsilon_n h\right)$$

since $\hat{\mathcal{J}}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = \arg\max_{h \in \epsilon_n^{-2\gamma}(\Theta - \theta_0)} \epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot, \hat{\theta}_n + \epsilon_n h\right)$.

There are two leading cases for Theorem 4.1: the smooth case and the cubic root case. In the smooth case, $\rho = 1$ and $\gamma = \frac{1}{2}$, and the Gaussian process $\mathcal{G}_0 g(\cdot;\theta)$ is linearly separable in θ. Typically there exists a multivariate normal random vector $\mathcal{W}_0 \sim N(0,\Omega)$ such that $\mathcal{G}_0 g(\cdot;\theta) = \mathcal{W}_0'(\theta - \theta_0)$, and for an independent copy $\mathcal{W}_1$ of $\mathcal{W}_0$, $\mathcal{G}_1 g(\cdot;\theta) = \mathcal{W}_1'(\theta - \theta_0)$. The regular bootstrap is valid in this case due to linear separability, and corresponds to $\epsilon_n = 1/\sqrt{n}$. In particular,

$$\begin{aligned}\hat{\theta}_n^* &= \arg\max_{\theta\in\Theta} \mathcal{Z}_n^* g(\cdot;\theta) \equiv \left(\mathcal{Z}_n^* - P_n\right)g(\cdot;\theta) + (P_n - P)g(\cdot;\theta) + Pg(\cdot;\theta)\\ &\approx \frac{\mathcal{W}_0 + \mathcal{W}_1}{\sqrt{n}}(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)'H(\theta - \theta_0),\end{aligned}$$

since $(\mathcal{Z}_n^* - P_n)g(\cdot;\theta) \approx \mathcal{W}_1/\sqrt{n}$ and $(P_n - P)g(\cdot;\theta) \approx \mathcal{W}_0/\sqrt{n}$. Likewise the sample estimate satisfies

$$\begin{aligned}\hat{\theta}_n &= \arg\max_{\theta\in\Theta} P_n g(\cdot;\theta) = (P_n - P)g(\cdot;\theta) + Pg(\cdot;\theta)\\ &\approx \frac{\hat{\mathcal{G}}_n}{\sqrt{n}} g(\cdot;\theta) - \frac{1}{2}(\theta - \theta_0)'H(\theta - \theta_0)\\ &\approx \frac{\mathcal{W}_0}{\sqrt{n}}(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)'H(\theta - \theta_0).\end{aligned}$$

Hence, if we let $\hat{h}_n^* = \sqrt{n}(\hat{\theta}_n^* - \theta_0)$ and $\hat{h}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$, then $\hat{h}_n^* \xrightarrow{p} H^{-1}(\mathcal{W}_0 + \mathcal{W}_1)$ and $\hat{h}_n \xrightarrow{p} H^{-1}\mathcal{W}_0$, so that $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \hat{h}_n^* - \hat{h}_n \xrightarrow{p} H^{-1}\mathcal{W}_1 = N(0, H^{-1}\Omega H^{-1})$.

4.1. *Constrained M estimation.* A related application is to constrained M-estimators when the parameter (in a correctly specified model) can possibly lie on the boundary of the constrained set. In the following, we verify the consistency of the numerical bootstrap, under conditions given in [17, 21], and in Theorem 4.1. Alternative approaches to similar problems are provided in [28] and [7]. While the latter approach provides a closer tie between the numerical bootstrap and the numerical delta method, the former approach seems more in line with the convention in the statistics literature. To simplify notation when we make use of results from [17], we consider arg min instead of arg max.

Following the previous notation, replace the parameter space Θ by a constrained subset C such that for $\hat{\theta}_n \in C$ and $\hat{\theta}_n^* \in C$,

$$P_n\pi(\cdot, \hat{\theta}_n) \le \inf_{\theta\in C} P_n\pi(\cdot,\theta) + o_P\left(n^{-2\gamma}\right), \tag{4.2}$$

$$\mathcal{Z}_n^*\pi\left(\cdot, \hat{\theta}_n^*\right) \le \inf_{\theta\in C} \mathcal{Z}_n^*\pi(\cdot,\theta) + o_P^*\left(\epsilon_n^{4\gamma}\right). \tag{4.3}$$

Let C be approximated by a cone $T_C(\theta_0)$ at θ_0 in the sense of Theorem 2.1 in [17], which implies (p. 2002 [17]) that for $n \to \infty$,

$$+\infty 1\big(\delta \notin n^\gamma(C - \theta_0)\big) \xrightarrow{e} +\infty 1\big(\delta \notin T_C(\theta_0)\big). \tag{4.4}$$

Here, $\xrightarrow{e}$ denotes epigraphical convergence as defined in [17], p. 1997. The difficulty of practical inference lies in the challenge of estimating the approximating cone $T_C(\theta_0)$ [31], which is easily handled by the numerical bootstrap method.

The following theorem combines the results in [17, 21] and Theorem 4.1. A restricted version of Theorem 4.2 corresponding to $\rho = 1$ and $\gamma = 1/2$ can also be stated using only Assumptions A–D, Lemma 4.1, and Theorem 4.4 in [17]. It also includes Theorem 4.1 as a special case when $T_C(\theta_0) = R^d$.

THEOREM 4.2. *Assume θ_0 uniquely minimizes $P\pi(\cdot, \theta)$ over $\theta \in C$. Let* (4.4) *and the conditions except* (i) *and* (iii) *in Theorem* 4.1 *hold* (*and also replace* (v) *with a positive definite H*). *Also assume that*

$$\mathcal{J} \equiv \arg\min_{h \in T_C(\theta_0)} \mathcal{Z}_0(h) + \frac{1}{2} h' H h \tag{4.5}$$

is almost surely unique. Then $\hat{\mathcal{J}}_n \equiv n^\gamma(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{J}$, $\hat{\mathcal{J}}_n^* \equiv \epsilon_n^{-2\gamma}(\hat{\theta}_n^* - \hat{\theta}_n) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$, and $\hat{\mathcal{J}}_n^* \equiv \epsilon_n^{-2\gamma}(\hat{\theta}_n^* - \hat{\theta}_n) \rightsquigarrow \mathcal{J}$.

If θ_0 is in the interior of C, then $T_C(\theta_0) = R^d$ and the proof of Theorem 4.1 can be applied. In other special cases, the proof of Theorem 4.1 can also be applied without change to Theorem 4.2, without having to appeal to the notion of epi-convergence. For example, it applies when θ_0 is on the boundary of C and $C - \theta_0$ already contains a cone at the origin, meaning for any compact set K, $\exists \alpha > 0$ such that $T_C(\theta_0) \cap K \subset \alpha(C - \theta_0)$ where $C - \theta_0$ is the tensor product between a cone at the origin and an open set.

Theorem 4.2 is based on the M-estimation framework, but generalization to (correctly specified) GMM models is immediate. In GMM models, $\hat{\theta}_n = \arg\min_{\theta \in C} n\hat{Q}_n(\theta)$, where for a positive definite W and $\hat{W} = W + o_P(1)$

$$\hat{Q}_n(\theta) = \hat{\pi}(\theta)' \hat{W} \hat{\pi}(\theta) \quad \text{and} \quad \hat{\pi}(\theta) = \frac{1}{n} \sum_{i=1}^n \pi(z_i, \theta).$$

ASSUMPTION 4.1. 1. Θ is compact and $\pi(\theta) = E\pi(z_i, \theta)$. 2. $\pi(\theta)$ is four times continuously differentiable. 3. $\{\pi(\cdot, \theta) : \theta \in \Theta\}$ is a VC class of functions. 4. $\pi(\theta) = 0$ if and only $\theta = \theta_0$ and $\theta_0 \in C$.

Define $G_0 = \frac{\partial}{\partial\theta}\pi(\theta_0)$, and let $\frac{1}{\sqrt{n}} \sum_{i=1}^n \pi(z_i, \theta_0) \rightsquigarrow Z = N(0, \Omega)$. Also define $\Delta_n = G_0 W \frac{1}{\sqrt{n}} \sum_{i=1}^n \pi(z_i, \theta_0)$, $\Delta_0 = G_0 W Z$, and $H = G_0 W G_0'$. It is known (e.g., [13]) that Assumption 4.1 implies the following identification condition and quadratic expansion of the objective function $\hat{Q}_n(\theta)$:

$$\forall \delta > 0, \exists \epsilon > 0 \quad \text{s.t.} \quad \limsup P\Big(\inf_{|\theta - \theta_0| \geq \delta} \hat{Q}_n(\theta) - \hat{Q}_n(\theta_0) \geq \epsilon\Big) = 1 \tag{4.6}$$

and for $R_n(\theta) = n\hat{Q}_n(\theta) - n\hat{Q}_n(\theta_0) - \Delta_n'\sqrt{n}(\theta - \theta_0) - n(\theta - \theta_0)'\frac{H}{2}(\theta - \theta_0)$,

$$\forall \delta_n \downarrow 0, \qquad \sup_{|\theta - \theta_0| \leq \delta_n} \frac{|R_n(\theta)|}{1 + \sqrt{n}|\theta - \theta_0| + n|\theta - \theta_0|^2} = o_P(1). \tag{4.7}$$

Under (4.7), which also holds for most M-estimators, $n\hat{Q}_n(\theta)$ is locally approximated by a quadratic function:

$$n\tilde{Q}_n(\theta) = \frac{1}{2}\big(\sqrt{n}(\theta - \theta_0) + H^{-1}\Delta_n\big)' H\big(\sqrt{n}(\theta - \theta_0) + H^{-1}\Delta_n\big) - \frac{1}{2}\Delta_n' H^{-1}\Delta_n.$$

This leads to the asymptotic distribution

$$\begin{aligned} \hat{\mathcal{J}}_n &= \sqrt{n}(\hat{\theta}_n - \theta_0) \\ &\rightsquigarrow \mathcal{J} = \arg\min_{h \in T_C(\theta_0)} \big(h + H^{-1}\Delta_0\big)' H\big(h + H^{-1}\Delta_0\big). \end{aligned} \tag{4.8}$$

Each of the three unknown components can be consistently estimated. (1) Let $\hat{G}$ be either $\frac{\partial}{\partial\theta}\frac{1}{n}\sum_{i=1}^n \pi(z_i; \hat{\theta}_n)$ or a numerical derivative analog, and let $\hat{H} = \hat{G}\hat{W}\hat{G}'$. (2) Let $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \pi(z_i; \hat{\theta}_n)\pi(z_i; \hat{\theta}_n)'$. Then let $\hat{Z}_n^* = N(0, \hat{\Omega})$ be such that $\hat{Z}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} Z$, $\hat{\Delta}_n^* = \hat{G}W\hat{Z}_n^*$ so that $\hat{\Delta}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \Delta_0$. (3) Since $T_C(\theta_0)$ is the limit of $\sqrt{n}(C - \theta_0)$, we can also estimate $T_C(\theta_0)$ by $\epsilon_n^{-1}(C - \hat{\theta}_n)$. Therefore we define, with $\hat{\mathbb{G}}_n^* = -\hat{H}^{-1}\hat{\Delta}_n^*$,

$$\hat{\mathcal{J}}_n^* = \arg\min_{h \in \epsilon_n^{-1}(C - \hat{\theta}_n)} \big(h - \hat{\mathbb{G}}_n^*\big)' \hat{H}\big(h - \hat{\mathbb{G}}_n^*\big). \tag{4.9}$$

If $C = \{\theta : \theta \geq 0\}$, then $\{h \in \epsilon_n^{-1}(C - \hat{\theta}_n)\} = \{h \geq -\epsilon_n^{-1}\hat{\theta}_n\}$.

In the regular M-estimator problem where $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \pi(z_i, \theta)$, we typically have $\hat{H} = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta\partial\theta'}\pi(z_i; \hat{\theta}_n)$ or a numerical derivative analog, and $\hat{\Delta}_n^* \sim N(0, \hat{\Sigma})$, where $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\pi(z_i; \hat{\theta}_n)\frac{\partial}{\partial\theta}\pi(z_i; \hat{\theta}_n)'$, or a numerical derivative analog.

THEOREM 4.3. *Given* (4.4), *under* (4.6) (*implied by Assumption* 4.1) *and* (4.7), (4.8) *holds, and* $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$.

Theorem 4.2 allows for more general nonstandard asymptotics with $\gamma = 1/3$. Theorem 4.3 is only confined to the regular case of $\gamma = 1/2$, but can be easier to implement since the objective function $(h - \hat{\mathbb{G}}_n^*)'\hat{H}(h - \hat{\mathbb{G}}_n^*)$ is convex whenever $\hat{H}$ is positive semi-definite. In particular, if C is a polyhedron, then the problem can be solved by quadratic programming.

If an unconstrained estimate $\bar{\theta}_n = \arg\min_{\theta\in\Theta} \hat{Q}_n(\theta)$ with $\theta_0 \in \text{int}(\Theta)$ is available, it is well known that $\sqrt{n}(\bar{\theta}_n - \theta_0) = -H^{-1}\Delta_n + o_P(1) \rightsquigarrow -H^{-1}\Delta_0$, and that the bootstrap estimate $\bar{\theta}_n^* = \arg\min_{\theta\in\Theta} \hat{Q}_n^*(\theta)$ also satisfies $\sqrt{n}(\bar{\theta}_n^* - \bar{\theta}_n) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} -H^{-1}\Delta_0$. Therefore, we can replace $\hat{\mathbb{G}}_n^* = -\hat{H}^{-1}\hat{\Delta}_n^*$ with $\hat{\mathbb{G}}_n^* = \sqrt{n}(\bar{\theta}_n^* - \bar{\theta}_n)$. The proof of Theorem 4.3 goes through verbatim with this replacement. Furthermore, a direct application of the numerical bootstrap in the GMM setup approximates the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ by that of $\epsilon_n^{-1}(\hat{\theta}_n^* - \hat{\theta}_n)$, where

$$\hat{\theta}_n^* = \arg\min_{\theta\in C} \epsilon_n^{-2}\hat{Q}_n^*(\theta), \qquad \hat{Q}_n^*(\theta) = \hat{\pi}^*(\theta)'W\hat{\pi}^*(\theta), \tag{4.10}$$

$$\hat{\pi}^*(\theta) = \mathcal{Z}_n^*\pi(z_i, \theta) = \big(P_n + \epsilon_n\hat{\mathcal{G}}_n^*\big)\pi(z_i, \theta), \tag{4.11}$$

and where $\hat{\mathcal{G}}_n^*$ can be the multinomial bootstrap or the wild bootstrap or other schemes that consistently estimate the limiting Gaussian process $\mathcal{G}_0$.

THEOREM 4.4. *Under Assumption* 4.1, $\hat{\mathcal{J}}_n^* = \frac{(\hat{\theta}_n^* - \hat{\theta}_n)}{\epsilon_n} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$, *for* $\mathcal{J}$ *defined in* (4.8).

4.2. *Sample extremum*: *A counter example*. We now provide a counter example in which both the bootstrap and the numerical bootstrap fail, but subsampling and the m-out-of-n bootstrap are valid. Let $P \sim U(0,1)$,

$$\theta(P) = \inf\bigl(t : F(t) \geq 1\bigr) = 1,$$

$$\theta(P_n) = \inf\bigl(t : F_n(t) \geq 1\bigr) = \max(X_1, \dots, X_n).$$

It is well known that for $a(n) = n$ and $\mathcal{E}$ a standard exponential,

$$a(n)\bigl(\theta(P_n) - \theta(P)\bigr) \rightsquigarrow \mathcal{J} = -\mathcal{E}, \tag{4.12}$$

which is also the limit of the subsampling distribution. In this one-dimensional example, $\mathcal{G}_0$ is the standard Brownian bridge $\mathcal{B}(t)$ on $t \in (0,1)$ with covariance function $\min(s,t) - st$ for $0 \leq s, t \leq 1$. Consider now

$$a\left(\frac{1}{\epsilon_n^2}\right)\bigl(\theta(P + \epsilon_n \mathcal{G}_0) - \theta(P)\bigr) = \frac{1}{\epsilon_n^2}\bigl(\theta(P + \epsilon_n \mathcal{G}_0) - \theta(P)\bigr), \tag{4.13}$$

where, since $F(t) = t$, $\mathcal{G}_0 = \mathcal{B}$,

$$\mathcal{T}_n \equiv \theta(F + \epsilon_n \mathcal{G}_0) \equiv \inf\bigl(t : t + \epsilon_n \mathcal{B}(t) \geq 1\bigr) = \inf\left(t : \mathcal{B}(t) = \frac{1-t}{\epsilon_n}\right).$$

In other words, $\mathcal{T}_n$ is the first passage time of the standard Brownian bridge over the linear barrier $\frac{1-t}{\epsilon_n}$. It is known that $\mathcal{B}(t)$ has the same (joint) distribution as $(1-t)\mathcal{W}(\frac{t}{1-t})$ where $\mathcal{W}(\cdot)$ is a standard Brownian motion. Therefore, $\mathcal{T}_n$ is equivalent in distribution to

$$\mathcal{T}_n = \inf\left(t : (1-t)\mathcal{W}\left(\frac{t}{1-t}\right) = \frac{1-t}{\epsilon_n}\right) = \inf\left(t : \mathcal{W}\left(\frac{t}{1-t}\right) = \frac{1}{\epsilon_n}\right).$$

This can be rewritten as $\mathcal{T}_n = \frac{\tau_n}{1+\tau_n}$, where $\tau_n = \inf\left(t : \mathcal{W}(t) = \frac{1}{\epsilon_n}\right)$. It is a standard result that

$$P(\tau_n \leq t) = 2P\left(\mathcal{W}(t) \geq \frac{1}{\epsilon_n}\right) = 2 - 2\Phi\left(\frac{1}{\epsilon_n \sqrt{t}}\right).$$

Transforming τ_n monotonically to $\mathcal{T}_n$,

$$P(\mathcal{T}_n \leq t) = 2 - 2\Phi\left(\epsilon_n^{-1}\sqrt{\frac{1-t}{t}}\right).$$

Finally, consider $\mathcal{Y}_n = \frac{1}{\epsilon_n^2}(\mathcal{T}_n - 1) \in (-\infty, 0)$. For $y > 0$, as $\epsilon_n \downarrow 0$, we obtain a limit distribution different from the exponential limit distribution.

$$P(\mathcal{Y}_n \leq -y) = 2 - 2\Phi\left(\frac{1}{\epsilon_n\sqrt{\frac{-\epsilon_n^2 y + 1}{\epsilon_n^2 y}}}\right) = 2 - 2\Phi\left(\frac{1}{\sqrt{\frac{-\epsilon_n^2 y + 1}{y}}}\right) \longrightarrow 2 - 2\Phi(\sqrt{y}).$$

Intuitively, what makes the limit distributions in (4.12) and (4.13) differ seems to be too much dependence on the tail of $\mathcal{G}_0(t)$ in (4.13). In particular, for $\hat{\mathcal{G}}_n = \sqrt{n}(P_n - P)$, let

$$\theta(P_n) = \theta\left(P + \frac{\hat{\mathcal{G}}_n}{\sqrt{n}}\right) = \theta\left(P + \frac{\hat{\mathcal{G}}_n - \mathcal{G}_0}{\sqrt{n}} + \frac{\mathcal{G}_0}{\sqrt{n}}\right).$$

We would expect that $\hat{\mathcal{G}}_n - \mathcal{G}_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. In general, this should be smaller than $\mathcal{G}_0$ in order of magnitude. However, in the sample extremum example, $\theta(P + \epsilon_n \mathcal{G}_0)$ depends on a point t^* of $\mathcal{G}_0(t)$ such that $\mathcal{G}_0(t^*) = O_p\left(\frac{1}{\sqrt{n}}\right)$. This makes $\hat{\mathcal{G}}_n - \mathcal{G}_0$ and $\mathcal{G}_0$ similar in order of magnitude. The difference in the limit distributions of (4.12) and (4.13) results from the non-negligible error in $\hat{\mathcal{G}}_n - \mathcal{G}_0$. In other words, we expect the numerical bootstrap method to be valid whenever the error in $\hat{\mathcal{G}}_n - \mathcal{G}_0$ is small in comparison with $\mathcal{G}_0$.

TABLE 1
Standard and perturbation bootstrap equal-tailed coverage frequencies

	Standard bootstrap					Perturbation bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.488	0.494	0.500	0.514	0.504	0.490	0.497	0.500	0.509	0.504
	(1.487)	(1.495)	(1.511)	(1.533)	(1.502)	(1.548)	(1.546)	(1.548)	(1.582)	(1.552)
$n = 500$	0.552	0.585	0.572	0.543	0.589	0.606	0.621	0.620	0.604	0.635
	(1.129)	(1.096)	(1.125)	(1.114)	(1.127)	(1.339)	(1.293)	(1.317)	(1.331)	(1.331)
$n = 1000$	0.597	0.560	0.589	0.595	0.595	0.683	0.643	0.673	0.679	0.677
	(0.922)	(0.940)	(0.925)	(0.921)	(0.957)	(1.135)	(1.143)	(1.134)	(1.123)	(1.171)
$n = 5000$	0.638	0.627	0.625	0.636	0.674	0.751	0.738	0.752	0.755	0.780
	(0.562)	(0.566)	(0.565)	(0.576)	(0.570)	(0.721)	(0.728)	(0.727)	(0.734)	(0.729)
$n = 10{,}000$	0.644	0.664	0.645	0.638	0.665	0.763	0.763	0.770	0.763	0.782
	(0.453)	(0.459)	(0.450)	(0.459)	(0.453)	(0.578)	(0.581)	(0.584)	(0.588)	(0.579)

5. Monte Carlo simulations. We investigate the performance of the numerical bootstrap for a modal estimator that is similar to example 3.2.13 in [34]. Let $X_1, \ldots, X_n$ be i.i.d. random variables drawn from $N(\theta_0, 2)$. Define $\hat{\theta}_n = \arg\max_\theta \frac{1}{n}\sum_{i=1}^n \mathbf{1}(\theta - 5 \leq X_i \leq \theta + 5)$, the center of an interval of length 10 that contains the largest possible fraction of the observations. [34] shows that $n^{1/3}(\hat{\theta}_n - \theta_0)$ converges in distribution to the maximizer of a Gaussian process plus an additional quadratic term. We investigate the empirical coverage frequencies of nominal 95% confidence intervals constructed using the standard bootstrap, the perturbation bootstrap [14, 24], the numerical bootstrap with $\epsilon_n \in \{n^{-1/3}, n^{-1/4}, n^{-1/6}\}$, the m-out-of-n bootstrap [6, 30] with $m \in \{n^{2/3}, n^{1/2}, n^{1/3}\}$, and subsampling [25] with $b \in \{n^{2/3}, n^{1/2}, n^{1/3}\}$. We consider several values of $\theta_0 \in \{-n^{-1/4}, 0, n^{-1}, n^{-1/2}, 2\}$ and several values of $n \in \{100, 500, 1000, 5000, 10{,}000\}$. We use 1000 Monte Carlo simulations and 5000 bootstrap iterations. Tables 1 through 4 show the two-sided equal-tailed coverage frequencies along with the average widths of the confidence intervals (in parentheses).

We can see that the standard bootstrap confidence intervals severely undercover for all values of θ_0. The perturbation bootstrap improves upon the standard bootstrap but still undercovers for all θ_0. The m-out-of-n bootstrap performs better than the perturbation bootstrap

TABLE 2
m-out-of-n and numerical bootstrap equal-tailed coverage for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n bootstrap					Numerical bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.669	0.667	0.679	0.685	0.675	0.766	0.773	0.760	0.780	0.769
	(1.806)	(1.809)	(1.798)	(1.821)	(1.807)	(2.998)	(3.004)	(3.010)	(3.012)	(3.006)
$n = 500$	0.762	0.785	0.776	0.772	0.795	0.855	0.890	0.880	0.849	0.880
	(1.322)	(1.305)	(1.295)	(1.311)	(1.303)	(1.645)	(1.633)	(1.635)	(1.640)	(1.634)
$n = 1000$	0.791	0.814	0.806	0.817	0.803	0.895	0.866	0.872	0.886	0.872
	(1.064)	(1.063)	(1.059)	(1.067)	(1.069)	(1.254)	(1.255)	(1.254)	(1.254)	(1.253)
$n = 5000$	0.843	0.826	0.839	0.840	0.850	0.900	0.876	0.880	0.878	0.900
	(0.625)	(0.625)	(0.623)	(0.623)	(0.625)	(0.678)	(0.677)	(0.676)	(0.677)	(0.679)
$n = 10{,}000$	0.864	0.864	0.859	0.853	0.865	0.880	0.877	0.879	0.878	0.884
	(0.495)	(0.494)	(0.494)	(0.496)	(0.496)	(0.524)	(0.525)	(0.525)	(0.526)	(0.526)

TABLE 3
m-out-of-n and numerical bootstrap equal-tailed coverage for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-*n* bootstrap					Numerical bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.710	0.729	0.724	0.722	0.724	0.822	0.817	0.820	0.824	0.823
	(1.861)	(1.872)	(1.861)	(1.878)	(1.864)	(3.068)	(3.072)	(3.077)	(3.083)	(3.077)
$n = 500$	0.783	0.796	0.777	0.781	0.798	0.922	0.946	0.927	0.914	0.930
	(1.299)	(1.285)	(1.284)	(1.287)	(1.289)	(1.793)	(1.784)	(1.789)	(1.786)	(1.786)
$n = 1000$	0.783	0.815	0.792	0.812	0.808	0.950	0.935	0.938	0.947	0.941
	(1.048)	(1.049)	(1.046)	(1.053)	(1.048)	(1.397)	(1.396)	(1.397)	(1.396)	(1.396)
$n = 5000$	0.849	0.827	0.837	0.856	0.864	0.955	0.945	0.949	0.958	0.956
	(0.636)	(0.634)	(0.634)	(0.633)	(0.633)	(0.772)	(0.771)	(0.770)	(0.771)	(0.771)
$n = 10{,}000$	0.879	0.887	0.865	0.873	0.879	0.962	0.954	0.945	0.957	0.953
	(0.506)	(0.506)	(0.506)	(0.506)	(0.507)	(0.595)	(0.595)	(0.595)	(0.595)	(0.595)

but still gives coverage less than the nominal frequency for all values of m. For each ϵ_n, the numerical bootstrap outperforms the m-out-of-n bootstrap with $m = \epsilon_n^{-2}$.

We next use a version of the double bootstrap algorithm described in [12] and references therein to find the optimal choice of ϵ_n for $n = 1000$. Many other possibilities for choosing ϵ_n exist, and an extensive discussion of the theoretical properties of each method is beyond the scope of the paper. Starting from the smallest value in a grid of $\epsilon_n \in \{n^{-1/2}, n^{-1/3}, \dots, n^{-1/15}\}$, the algorithm draws $B_1 = 5000$ bootstrap samples and computes bootstrap estimates $\hat{\theta}_n^{(b_1)}$. Conditional on each of these bootstrap samples, the algorithm draws $B_2 = 5000$ bootstrap samples and computes bootstrap estimates $\hat{\theta}_n^{(b_1,b_2)}$. The algorithm then computes the empirical frequency with which equal tailed intervals centered at $\hat{\theta}_n^{(b_1)}$ cover $\hat{\theta}_n$. If the current value of ϵ_n achieves coverage at or above the nominal frequency, then it picks that value as the optimal ϵ_n. Otherwise, it increments ϵ_n to the next highest value in the grid and repeats the steps above.

Table 5 shows the double bootstrap coverage frequencies for $\epsilon_n \in \{n^{-1/2}, n^{-1/3}, \dots, n^{-1/11}\}$ and $\theta_0 \in \{-n^{-1/4}, 0, n^{-1}, n^{-1/2}, 2\}$. The coverage frequencies for the other values of ϵ_n are all less than the nominal frequency. We see that the smallest value of ϵ_n for which

TABLE 4
m-out-of-n and numerical bootstrap equal-tailed coverage for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-*n* bootstrap					Numerical bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.719	0.729	0.736	0.726	0.730	0.835	0.830	0.832	0.833	0.842
	(1.823)	(1.828)	(1.817)	(1.839)	(1.827)	(2.944)	(2.947)	(2.948)	(2.947)	(2.950)
$n = 500$	0.715	0.722	0.734	0.727	0.728	0.945	0.961	0.942	0.940	0.947
	(1.186)	(1.179)	(1.176)	(1.180)	(1.182)	(1.793)	(1.789)	(1.790)	(1.792)	(1.789)
$n = 1000$	0.694	0.721	0.705	0.715	0.733	0.969	0.963	0.964	0.967	0.963
	(0.954)	(0.958)	(0.955)	(0.957)	(0.959)	(1.433)	(1.433)	(1.432)	(1.433)	(1.434)
$n = 5000$	0.750	0.724	0.735	0.771	0.761	0.982	0.982	0.974	0.981	0.978
	(0.592)	(0.593)	(0.591)	(0.590)	(0.591)	(0.839)	(0.839)	(0.839)	(0.839)	(0.838)
$n = 10{,}000$	0.791	0.793	0.774	0.775	0.811	0.983	0.986	0.988	0.982	0.978
	(0.481)	(0.481)	(0.481)	(0.481)	(0.481)	(0.662)	(0.662)	(0.662)	(0.662)	(0.663)

TABLE 5
Double bootstrap equal-tailed coverage frequencies

θ_0/ϵ_n	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$	$n^{-1/7}$	$n^{-1/8}$	$n^{-1/9}$	$n^{-1/10}$	$n^{-1/11}$
0	0.8754	0.9838	0.9816	0.9556	0.9482	0.9502	0.9502	0.8692	0.8650	0.8732
$1/n$	0.8754	0.9838	0.9816	0.9572	0.9502	0.9510	0.9454	0.8668	0.8732	0.8528
$1/\sqrt{n}$	0.8796	0.9820	0.9830	0.9582	0.9510	0.9464	0.9470	0.8744	0.8528	0.8598
$n^{-1/4}$	0.8754	0.9838	0.9816	0.9556	0.9482	0.9502	0.9502	0.8668	0.8732	0.8528
2	0.8754	0.9838	0.9816	0.9556	0.9482	0.9502	0.9502	0.8692	0.8650	0.8732

the coverage exceeds the nominal frequency is $n^{-1/3}$. However, at this value, the coverage is around 0.98 for all θ_0, which is much higher than the nominal frequency of 0.95. It might make more sense to choose a value of ϵ_n for which the coverage is closer to the nominal frequency, for example $n^{-1/5}$.

Due to space constraints, results for subsampling and one-sided confidence intervals are in the Supplementary Material. Simulation results for the LASSO estimator in the one-dimensional mean model are also in the Supplementary Material.

Acknowledgments. We thank two anonymous referees, Andres Santos, Joe Romano, Xiaohong Chen, Zheng Fang, Bruce Hansen, David Kaplan, Adam McCloskey, Frank Wolak and participants at the Montreal Econometric Society World Congress and various conferences and seminars for helpful comments.

This work was supported by the National Science Foundation (SES 1164589), and both the IRiSS and the B. F. Haley and E. S. Shaw SIEPR dissertation fellowships.

SUPPLEMENTARY MATERIAL

Supplement to "The numerical bootstrap" (DOI: 10.1214/19-AOS1812SUPP; .pdf). The supplement contains a list of commonly used symbols, proofs of the theorems, further discussion of local asymptotics, and additional simulation results. Also included is a discussion of sample size dependent statistics, the role of recentering in hypothesis testing, estimating an unknown polynomial convergence rate, and inference for partially identified models.

REFERENCES

[1] ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245 https://doi.org/10.1111/j.1468-0262.2005.00613.x

[2] ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157. MR2642858 https://doi.org/10.3982/ECTA7502

[3] ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609–666. MR3043344 https://doi.org/10.3982/ECTA9370

[4] BABU, G. J. and SINGH, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17** 261–278. MR0813236 https://doi.org/10.1016/0047-259X(85)90084-3

[5] BERTAIL, P. (1997). Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli* **3** 149–179. MR1466305 https://doi.org/10.2307/3318585

[6] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

[7] BONNANS, J. F. and SHAPIRO, A. (2000). *Perturbation Analysis of Optimization Problems. Springer Series in Operations Research*. Springer, New York. MR1756264 https://doi.org/10.1007/978-1-4612-1394-9

[8] BUGNI, F. A., CANAY, I. A. and SHI, X. (2015). Specification tests for partially identified models defined by moment inequalities. *J. Econometrics* **185** 259–282. MR3300346 https://doi.org/10.1016/j.jeconom.2014.10.013

[9] BUGNI, F. A., CANAY, I. A. and SHI, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quant. Econ.* **8** 1–38. MR3638598 https://doi.org/10.3982/QE490

[10] CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2017). Bootstrap-Based Inference for Cube Root Consistent Estimators. Preprint. arXiv:1704.08066.

[11] CAVANAGH, C. L. (1987). Limiting behavior of estimators defined by optimization. Unpublished manuscript.

[12] CHAKRABORTY, B., LABER, E. B. and ZHAO, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69** 714–723. MR3106599 https://doi.org/10.1111/biom.12052

[13] CHERNOZHUKOV, V. and HONG, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115** 293–346. MR1984779 https://doi.org/10.1016/S0304-4076(03)00100-3

[14] DAS, D. and LAHIRI, S. N. (2019). Second order correctness of perturbation bootstrap M-estimator of multiple linear regression parameter. *Bernoulli* **25** 654–682. MR3892332 https://doi.org/10.3150/17-bej1001

[15] DÜMBGEN, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 https://doi.org/10.1007/BF01197342

[16] FANG, Z.and SANTOS, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies* **86** 377–412. https://doi.org/10.1093/restud/rdy049

[17] GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179 https://doi.org/10.1214/aos/1176325768

[18] HONG, H. and LI, J. (2019). Supplement to "The numerical bootstrap." https://doi.org/10.1214/19-AOS1812SUPP.

[19] JUN, S. J., PINKSE, J. and WAN, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *J. Econometrics* **187** 201–216. MR3347303 https://doi.org/10.1016/j.jeconom.2015.01.005

[20] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[21] KNIGHT, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Unpublished manuscript 37.

[22] LABER, E. B. and MURPHY, S. A. (2011). Adaptive confidence intervals for the test error in classification. *J. Amer. Statist. Assoc.* **106** 904–913. MR2894746 https://doi.org/10.1198/jasa.2010.tm10053

[23] MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905 https://doi.org/10.1016/0304-4076(75)90032-9

[24] MINNIER, J., TIAN, L. and CAI, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Amer. Statist. Assoc.* **106** 1371–1382. MR2896842 https://doi.org/10.1198/jasa.2011.tm10382

[25] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

[26] ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. *Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

[27] ROMANO, J. P. and SHAIKH, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *Ann. Statist.* **40** 2798–2822. MR3097960 https://doi.org/10.1214/12-AOS1051

[28] RÖMISCH, W. (2005). Delta method, infinite fimensional. In *Encyclopedia of Statistical Sciences* (S. Kotz et al., eds). Wiley, New York.

[29] SEIJO, E. and SEN, B. (2015). Bootstrapping Manski's Maximum Score Estimator. Preprint. arXiv:1105.1976.

[30] SHAO, J. (1994). Bootstrap sample size in nonregular cases. *Proc. Amer. Math. Soc.* **122** 1251–1262. MR1227529 https://doi.org/10.2307/2161196

[31] SHAPIRO, A. (1989). Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist.* **17** 841–858. MR0994271 https://doi.org/10.1214/aos/1176347146

[32] TEWES, J., POLITIS, D. N. and NORDMAN, D. J. (2019). Convolved subsampling estimation with applications to block bootstrap. *Ann. Statist.* **47** 468–496. MR3909939 https://doi.org/10.1214/18-AOS1695

[33] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

[34] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

1 List of Commonly Used Symbols

P_n	empirical measure
P_n^*	bootstrap empirical measure
$\mathcal{Z}_n^*$	$P_n + \epsilon_n \sqrt{n}\left(P_n^* - P_n\right)$
$\rightsquigarrow$	weak convergence
$\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}$	weak convergence conditional on the data
θ^-	$-min(\theta, 0)$
θ^+	$max(\theta, 0)$
$\rho_{BL_1}\left(F_1, F_2\right)$	$\sup_{f \in BL_1} \left\| Ef\left(F_1\right) - Ef\left(F_2\right)\right\|$
BL_1	the space of Lipschitz functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1

2 Comparison of Numerical Bootstrap with m-out-of-n bootstrap and Subsampling

In the first subsection, we discuss a simple example illustrating the differences between the numerical bootstrap, m-out-of-n bootstrap and subsampling and then in the second subsection we provide more detailed arguments for the more general case stated in the main text.

2.1 Negative Part of the Mean Example

Let $\phi\left(\mu\right) = \mu^-$ and $X_i \overset{i.i.d.}{\sim} \left(\mu, \sigma^2\right)$, where μ is a fixed parameter. Consider approximating the distribution of $\sqrt{n}\left(\theta\left(P_n\right) - \theta\left(P\right)\right) = \sqrt{n}\left(\bar{X}_n^- - \mu^-\right)$ using $\hat{\mathcal{J}}_n^* = \left(\frac{\bar{X}_n + \epsilon\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)}{\epsilon_n}\right)^- - \left(\frac{\bar{X}_n}{\epsilon_n}\right)^-$. First let $\mu = 0$ and $x > 0$. Then by standard Berry-Esseen arguments,

$$P\left(\sqrt{n}\bar{X}_n^- \leq x\right) = P\left(\sqrt{n}\bar{X}_n \geq -x\right) = \Phi\left(\frac{x}{\sigma}\right) + O\left(n^{-1/2}\right).$$

Then, since $\bar{X}_n/\epsilon_n = O_p\left(\frac{1}{\sqrt{n}\epsilon_n}\right)$, conditional on the data $\mathcal{X}_n = (X_1, \ldots, X_n)$:

$$
\begin{aligned}
&P\left(\hat{\mathcal{J}}_n^* \leq x | \mathcal{X}_n\right) \\
=&P\left(\left(\frac{\bar{X}_n}{\epsilon_n} + \sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)\right)^- - \left(\frac{\bar{X}_n}{\epsilon_n}\right)^- \leq x \middle| \mathcal{X}_n\right) \\
=&P\left(\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \geq -\left(x + \left(\frac{\bar{X}_n}{\epsilon_n}\right)^-\right) - \frac{\bar{X}_n}{\epsilon_n} \middle| \mathcal{X}_n\right) \\
=&\Phi\left(\frac{1}{\hat{\sigma}}\left(\left(x + \left(\frac{\bar{X}_n}{\epsilon_n}\right)^-\right) + \frac{\bar{X}_n}{\epsilon_n}\right)\right) + O_p\left(\frac{1}{\sqrt{n}}\right) = \Phi\left(\frac{x}{\hat{\sigma}}\right) + O_p\left(\frac{1}{\sqrt{n}\epsilon_n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Subsampling corresponds to $\epsilon_n = 1/\sqrt{b}$ and replacing $\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)$ with $\sqrt{b}\left(\bar{X}_b - \bar{X}_n\right)$:

$$
P\left(\left(\sqrt{b}\bar{X}_b\right)^- - \left(\sqrt{b}\bar{X}_n\right)^- \leq x\right) = \Phi\left(\frac{x}{\hat{\sigma}}\right) + O\left(\sqrt{\frac{b}{n}}\right) + O\left(b^{-1/2}\right).
$$

As shown in Lemma A.2 in Romano and Shaikh (2012), Theorem 1 in Bertail (1997), and Theorem 3 in Babu and Singh (1985), subsampling has an additional error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ from approximating the distribution of $\sqrt{b}\left(\bar{X}_b - \bar{X}_n\right)$ using the empirical distribution of $\binom{n}{b}$ sub-block estimates $\sqrt{b}\left(\bar{X}_{bi} - \bar{X}_n\right)$. For the m-out-of-n bootstrap, Bickel and Sakov (2008) suggest that

$$
P\left(\left(\sqrt{m}\bar{X}_m^*\right)^- - \left(\sqrt{m}\bar{X}_n\right)^- \leq x \middle| \mathcal{X}_n\right) = \Phi\left(\frac{x}{\hat{\sigma}}\right) + O_p\left(\sqrt{\frac{m}{n}}\right) + O_p\left(m^{-1/2}\right).
$$

An optimal choice of b (or m) will then be $b = O\left(n^{1/2}\right)$, resulting in an error of order $n^{-1/4}$. On the other hand, the error in numerical bootstrap can be made close to $O_p\left(n^{-1/2}\right)$ when $\epsilon_n \to 0$ slowly.

Next let $\mu < 0$. Then with probability converging to 1, the numerical bootstrap is identical to bootstrap, which has an error of $O_p\left(n^{-1/2}\right)$ regardless of how $\epsilon_n \to 0$, while the subsampling error can still be $O_p\left(n^{-1/4}\right)$.

If instead $\phi(\mu) = (\mu^-)^2$, then we would approximate the distribution of $n\left(\theta\left(P_n\right) - \theta\left(P\right)\right) = n\left(\left(\bar{X}_n^-\right)^2 - \left(\mu^-\right)^2\right)$ using $\hat{\mathcal{J}}_n^* = \left(\left(\frac{\bar{X}_n + \epsilon\sqrt{n}\left(\bar{X}_n^* - \bar{X}\right)}{\epsilon_n}\right)^-\right)^2 - \left(\left(\frac{\bar{X}_n}{\epsilon_n}\right)^-\right)^2$ where now $a(n) = n$ instead of $\sqrt{n}$ as before. Under a null hypothesis of $\mu \geq 0$, the approximation error of the numerical bootstrap is the same as before, so the optimal choice of ϵ_n remains the same.

2.2 More general case

Suppose $\phi : \mathbb{R} \mapsto \mathbb{R}$ is twice Hadamard directionally differentiable at $\mu = \mu(P)$ with directional derivatives $\phi_{\mu}^{'}(\cdot)$ and $\phi_{\mu}^{''}(\cdot)$ that can be continuously extended to $\mathbb{R}$. It is known that $\phi'_{\mu}(h)$ is positively homogeneous of degree 1. First we demonstrate that one dimensional positively homogeneous functions of degree 1 have a piecewise linear representation:

Claim: There exists constants $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\phi'_{\mu}(h) = \lambda_1 h^+ + \lambda_2 h^-$$

For example, in the case of $\phi(\mu) = a\mu^+ + b\mu^-$, if $\mu = 0$, $\phi'_{\mu}(h) = ah^+ + bh^-$.

Proof: Suppose $f(h) = \phi'_{\mu}(h)$ is a homogeneous of degree one function of a scalar argument h. Then for all $\lambda \geq 0$:

$$f(\lambda h) = \lambda f(h) \quad \text{or} \quad \frac{f(\lambda h) - f(0)}{\lambda h} = \frac{1}{h} f(h)$$

If $h > 0$, then $f'_{+}(0) = \lim_{\lambda \to 0^+} \frac{f(\lambda h) - f(0)}{\lambda h} = \frac{1}{h} f(h)$ so that $f(h) = f'_{+}(0)h$. On the other hand, if $h < 0$, then

$$f'_{-}(0) = \lim_{\lambda \to 0^+} \frac{f(\lambda h) - f(0)}{\lambda h} = \frac{1}{h} f(h)$$

so that $f(h) = f'_{-}(0)h$. Therefore $f(h)$ must be of the form of

$$f(h) = f'_{+}(0)h^+ - f'_{-}(0)h^-.$$

■

We are interested in approximating the limiting distribution of $\sqrt{n}\left(\phi\left(\bar{X}_n\right) - \phi(\mu)\right)$ using the numerical bootstrap. Taylor expanding $\phi\left(\bar{X}_n + \epsilon_n \sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right)\right)$ around

$\phi(\mu)$ gives us

$$\begin{aligned}\phi\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)\right) &= \phi(\mu)+\epsilon_n\phi_\mu^{'}\left(\frac{\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu}{\epsilon_n}\right)\\ &+ \frac{1}{2}\epsilon_n^2\phi_\mu^{''}\left(\frac{\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu}{\epsilon_n}\right)+o_p^*\left(\epsilon_n^2\right)\end{aligned}$$

Taylor expanding $\phi\left(\bar{X}_n\right)$ around $\phi(\mu)$ gives

$$\phi\left(\bar{X}_n\right)=\phi(\mu)+\frac{1}{\sqrt{n}}\phi_\mu^{'}\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\right)+\frac{1}{2n}\phi_\mu^{''}\left(\sqrt{n}\left(\bar{X}_n-\mu\right)\right)+o_p\left(\frac{1}{n}\right).$$

If $\lambda_1\geq 0$ and $\lambda_2\geq 0$, we expect that the numerical bootstrap distribution satisfies

$$\begin{aligned}&P\left(\hat{\mathcal{J}}_n^*\leq x\middle|\mathcal{X}_n\right)\\ =\ &P\left(\frac{1}{\epsilon_n}\left(\phi\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)\right)-\phi\left(\bar{X}_n\right)\right)\leq x\middle|\mathcal{X}_n\right)\\ =\ &P\left(\frac{1}{\epsilon_n}\left(\phi_\mu^{'}\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right)-\phi_\mu^{'}\left(\bar{X}_n-\mu\right)\right)\right.\\ &\left.+O_p^*\left(\frac{1}{\epsilon_n}\epsilon_n^2\right)+O_p\left(\frac{1}{\epsilon_n n}\right)\leq x\middle|\mathcal{X}_n\right)\\ =\ &P\left(\frac{1}{\epsilon_n}\left(\lambda_1\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right)^++\lambda_2\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right)^-\right)\right.\\ &\left.-\frac{1}{\epsilon_n}\left(\lambda_1\left(\bar{X}_n-\mu\right)^++\lambda_2\left(\bar{X}_n-\mu\right)^-\right)+O_p^*\left(\frac{1}{\epsilon_n}\epsilon_n^2+\frac{1}{\epsilon_n n}\right)\leq x\middle|\mathcal{X}_n\right)\\ =\ &P\left(\lambda_1\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right)^++\lambda_2\left(\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right)^-\right.\\ &\left.\leq\epsilon_n x+\lambda_1\left(\bar{X}_n-\mu\right)^++\lambda_2\left(\bar{X}_n-\mu\right)^-+O_p^*\left(\epsilon_n^2\right)+O_p\left(\frac{1}{n}\right)\middle|\mathcal{X}_n\right)\end{aligned}$$

Using the fact that $\mu(P_n)=\bar{X}_n$ is a linear function of P_n, and the fact that for $a,b\in\mathbb{R}^+$,

$$\begin{aligned}P\left(aX^++bX^-\leq x\right) &= P(aX\leq x\cap X>0)+P(-bX\leq x\cap X<0)\\ &= P(0<X\leq x/a)+P(-x/b\leq X<0),\end{aligned}$$

we can write this as

$$
\begin{aligned}
& P\left(0<\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu\right.\\
& \quad\left.\le\epsilon_n\frac{x}{\lambda_1}+\left(\bar{X}_n-\mu\right)^++\frac{\lambda_2}{\lambda_1}\left(\bar{X}_n-\mu\right)^--O_p^*\left(\epsilon_n^2\right)+O_p\left(\frac{1}{n}\right)\right|\mathcal{X}_n\right)\\
+\ & P\left(-\epsilon_n\frac{x}{\lambda_2}-\frac{\lambda_1}{\lambda_2}\left(\bar{X}_n-\mu\right)^+-\left(\bar{X}_n-\mu\right)^-+O_p^*\left(\epsilon_n^2\right)\right.\\
& \quad\left.<\bar{X}_n+\epsilon_n\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)-\mu+O_p\left(\frac{1}{n}\right)\le 0\right|\mathcal{X}_n\right)\\
=\ & P\left(\frac{-\bar{X}_n+\mu}{\epsilon_n}<\sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)\right.\\
& \quad\left.\le\frac{x}{\lambda_1}+\frac{1}{\epsilon_n}\left(\left(\bar{X}_n-\mu\right)^++\frac{\lambda_2}{\lambda_1}\left(\bar{X}_n-\mu\right)^--\bar{X}_n+\mu\right)-O_p^*\left(\epsilon_n\right)+O_p\left(\frac{1}{n\epsilon_n}\right)\right|\mathcal{X}_n\right)\\
+\ & P\left(-\frac{x}{\lambda_2}-\frac{1}{\epsilon_n}\left(\frac{\lambda_1}{\lambda_2}\left(\bar{X}_n-\mu\right)^++\left(\bar{X}_n-\mu\right)^-+\bar{X}_n-\mu\right)+O_p^*\left(\epsilon_n\right)+O_p\left(\frac{1}{n\epsilon_n}\right)\right.\\
& \quad\left.<\ \sqrt{n}\left(\bar{X}_n^*-\bar{X}_n\right)\le\frac{-\bar{X}_n+\mu}{\epsilon_n}\right|\mathcal{X}_n\right)\\
=\ & \Phi\left(\frac{x}{\lambda_1}+\frac{1}{\epsilon_n}\left(\left(\bar{X}_n-\mu\right)^++\frac{\lambda_2}{\lambda_1}\left(\bar{X}_n-\mu\right)^--\left(\bar{X}_n-\mu\right)\right)-O_p\left(\epsilon_n+\frac{1}{n\epsilon_n}\right)\right)\\
-\ & \Phi\left(-\frac{x}{\lambda_2}-\frac{1}{\epsilon_n}\left(\frac{\lambda_1}{\lambda_2}\left(\bar{X}_n-\mu\right)^++\left(\bar{X}_n-\mu\right)^-+\bar{X}_n-\mu\right)+O_p\left(\epsilon_n+\frac{1}{n\epsilon_n}\right)\right)+O_p\left(\frac{1}{\sqrt{n}}\right)\\
=\ & \Phi\left(\frac{x}{\lambda_1}\right)+\Phi\left(\frac{x}{\lambda_2}\right)-1+O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\epsilon_n\right)
\end{aligned}
$$

The other sign combinations of λ_1 and λ_2 can be considered analogously. We can see that when the second order derivative is nonzero and $\phi_\mu^{'}(\cdot)$ is not a linear function, then the error for the numerical bootstrap is $O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\epsilon_n\right)=O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right)+O_p\left(\epsilon_n\right)$. The optimal choice of ϵ_n that balances the two terms satisfies $\epsilon_n=O\left(n^{-1/4}\right)$, leading to an error on the order of $n^{-1/4}$. If however, $\phi_\mu^{'}(\cdot)$ is a linear function, then $\lambda_1=1$ and $\lambda_2=-1$, which implies that $\left(\bar{X}_n-\mu\right)^++\frac{\lambda_2}{\lambda_1}\left(\bar{X}_n-\mu\right)^--\left(\bar{X}_n-\mu\right)=0$ and $\frac{\lambda_1}{\lambda_2}\left(\bar{X}_n-\mu\right)^++\left(\bar{X}_n-\mu\right)^-+\left(\bar{X}_n-\mu\right)=0$. Therefore, the error for the numerical bootstrap is $O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\epsilon_n\right)$, and is minimized by $\epsilon_n=O\left(\frac{1}{\sqrt{n}}\right)$. Finally, if the second order derivative is zero, then the error for the numerical bootstrap is $O_p\left(\frac{1}{\epsilon_n\sqrt{n}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)$, and is smaller for larger values of ϵ_n.

Subsampling targets the distribution of $\hat{\mathcal{J}}_b\equiv\sqrt{b}\left(\phi\left(\mu\left(P_b\right)\right)-\phi\left(\mu\left(P_n\right)\right)\right)$, and takes the form:

$$P\left(\hat{\mathcal{J}}_b \leq x\right) = \Phi\left(\frac{x}{\lambda_1}\right) + \Phi\left(\frac{x}{\lambda_2}\right) - 1 + O\left(\sqrt{\frac{b}{n}}\right) + O\left(\frac{1}{\sqrt{b}}\right)$$

The optimal choice of b satisfies $b = O\left(n^{1/2}\right)$, which leads to an error on the order of $n^{-1/4}$. If $\phi'_{\mu}(\cdot)$ were a linear function that is not degenerate at μ or $\phi(\cdot)$ had zero second order or higher derivatives, the error would still be on the order of $n^{-1/4}$ because of the additional error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ introduced by estimating the distribution of $\sqrt{b}\left(\mu\left(P_b\right) - \mu\left(P_n\right)\right)$ using the empirical distribution of $\sqrt{b}\left(\mu\left(P_{b,i}\right) - \mu\left(P_n\right)\right)$ over $i = 1, \ldots, \binom{n}{b}$ sub-blocks. Lemma A.2 in Romano and Shaikh (2012) provides a detailed proof for the presence of an error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ when estimating the distribution of $\sqrt{b}\left(\mu\left(P_b\right) - \mu\left(P\right)\right)$ using the empirical distribution of $\sqrt{b}\left(\mu\left(P_{b,i}\right) - \mu\left(P\right)\right)$ over $i = 1, \ldots, \binom{n}{b}$ sub-blocks. Since $\sqrt{b}\left(\mu\left(P_n\right) - \mu\left(P\right)\right) = \sqrt{\frac{b}{n}}\sqrt{n}\left(\mu\left(P_n\right) - \mu\left(P\right)\right) = O_p\left(\sqrt{\frac{b}{n}}\right)$, it follows that the error from estimating the distribution of $\sqrt{b}\left(\mu\left(P_b\right) - \mu\left(P_n\right)\right)$ using the empirical distribution of $\sqrt{b}\left(\mu\left(P_{b,i}\right) - \mu\left(P_n\right)\right)$ is also $O_p\left(\sqrt{\frac{b}{n}}\right)$. The presence of the error of $O_p\left(\sqrt{\frac{b}{n}}\right)$ is also demonstrated in Theorem 1 of Bertail (1997) and Theorem 3 of Babu and Singh (1985). Therefore, the numerical bootstrap should not have an error that is of larger order than subsampling and it may outperform subsampling in some situations when the first derivative is linear and the second order derivative is nonzero, or when the second order derivative is zero.

3 Local Asymptotics

3.1 Proof Theorem 3.1

The asymptotic distribution of $\hat{\mathcal{J}}_n$ can be derived as follows:

$$\begin{aligned}
\hat{\mathcal{J}}_n &= \sqrt{n}\left(\phi\left(\mu\left(P\right) + \frac{1}{\sqrt{n}}\left(\sqrt{n}\left(\hat{\mu}_n - \mu\left(P^n\right)\right) + \sqrt{n}\left(\mu\left(P^n\right) - \mu\left(P\right)\right)\right)\right) - \phi\left(\mu\left(P\right)\right)\right) \\
&\quad - \sqrt{n}a_n\left(\frac{\phi\left(\mu\left(P\right) + a_n c\right) - \phi\left(\mu\left(P\right)\right)}{a_n}\right) \\
&= \phi'_{\mu_0}\left(\mathbb{G}_0 + \lim_{n\to\infty}\sqrt{n}a_n c\right) - \lim_{n\to\infty}\sqrt{n}a_n\phi'_{\mu_0}\left(c\right) + o_p\left(1\right) \\
&= \lim_{n\to\infty}\phi'_{\mu_0}\left(\mathbb{G}_0 + \sqrt{n}a_n c\right) - \phi'_{\mu_0}\left(\sqrt{n}a_n c\right) + o_p\left(1\right)
\end{aligned}$$

where the last line follows from ϕ'_{μ_0} being continuous and positively homogeneous of degree 1 (Römisch (2005)). For $\hat{\mathcal{J}_n}^* \equiv \frac{1}{\epsilon_n}\left(\phi\left(\hat{\mu}_n + \epsilon_n\sqrt{n}\left(\hat{\mu}_n^* - \hat{\mu}_n\right)\right) - \phi\left(\hat{\mu}_n\right)\right)$,

$$
\begin{aligned}
\hat{\mathcal{J}_n}^* &= \frac{1}{\epsilon_n}\left(\phi\left(\mu\left(P\right) + \epsilon_n\left(\sqrt{n}\left(\hat{\mu}_n^* - \hat{\mu}_n\right) + \frac{\hat{\mu}_n - \mu\left(P\right)}{\epsilon_n}\right)\right) - \phi\left(\mu\left(P\right)\right)\right) \\
&- \frac{1}{\epsilon_n}\left(\phi\left(\hat{\mu}_n\right) - \phi\left(\mu\left(P^n\right)\right)\right) - \frac{1}{\epsilon_n}\left(\phi\left(\mu\left(P^n\right)\right) - \phi\left(\mu\left(P\right)\right)\right) \\
&= \frac{1}{\epsilon_n}\left(\phi\left(\mu\left(P\right) + \epsilon_n\left(\sqrt{n}\left(\hat{\mu}_n^* - \hat{\mu}_n\right) + \frac{\sqrt{n}\left(\hat{\mu}_n - \mu\left(P^n\right)\right) + \sqrt{n}\left(\mu\left(P^n\right) - \mu\left(P\right)\right)}{\epsilon_n\sqrt{n}}\right)\right) - \phi\left(\mu\left(P\right)\right)\right) \\
&- o_p^*\left(1\right) - \frac{1}{\epsilon_n}a_n\left(\frac{\phi\left(\mu\left(P\right) + a_n c\right) - \phi\left(\mu\left(P\right)\right)}{a_n}\right) \\
&= \phi'_{\mu_0}\left(\mathbb{G}_0 + \lim_{n\to\infty}\frac{a_n}{\epsilon_n}c\right) - \lim_{n\to\infty}\frac{a_n}{\epsilon_n}\phi'_{\mu_0}\left(c\right) + o_p^*\left(1\right) \\
&= \lim_{n\to\infty}\phi'_{\mu_0}\left(\mathbb{G}_0 + \frac{a_n}{\epsilon_n}c\right) - \phi'_{\mu_0}\left(\frac{a_n}{\epsilon_n}c\right) + o_p^*\left(1\right)
\end{aligned}
$$

First note that if ϕ'_{μ_0} is linear, then both $\hat{\mathcal{J}}_n$ and $\hat{\mathcal{J}_n}^*$ converge to $\phi'_{\mu_0}\left(\mathbb{G}_0\right)$, which implies that two-sided equal-tailed and one-sided confidence intervals are asymptotically exact for all values of a_n.

If ϕ'_{μ_0} is nonlinear, we consider the following cases: (i) $a_n\sqrt{n} \to 0$, (ii) $a_n\sqrt{n} = 1$, (iii) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to 0$, (iv) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to 1$, (v) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to \infty$.

If $a_n\sqrt{n} \to 0$, then $\hat{\mathcal{J}}_n \rightsquigarrow \phi'_{\mu_0}\left(\mathbb{G}_0\right)$ and $\hat{\mathcal{J}_n}^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mu_0}\left(\mathbb{G}_0\right)$ since $\sqrt{n}\epsilon_n \to \infty$. Two-sided equal-tailed and one-sided confidence intervals are asymptotically exact up to first order asymptotics.

If $a_n = \frac{1}{\sqrt{n}}$, then $\hat{\mathcal{J}}_n \rightsquigarrow \phi'_{\mu_0}\left(\mathbb{G}_0 + c\right) - \phi'_{\mu_0}\left(c\right)$ while $\hat{\mathcal{J}_n}^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mu_0}\left(\mathbb{G}_0\right)$. If ϕ'_{μ_0} is subadditive (meaning $\phi'_{\mu_0}\left(h_1 + h_2\right) \leq \phi'_{\mu_0}\left(h_1\right) + \phi'_{\mu_0}\left(h_2\right)$ for all $h_1, h_2 \in \mathbb{D}_0$), then the limiting distribution of $\hat{\mathcal{J}_n}^*$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n$. Let $\hat{c}_{1-\alpha}^*$ be the $1-\alpha$ quantile of $\hat{\mathcal{J}_n}^*$ and let $c_{1-\alpha}^*$ be the $1-\alpha$ quantile of $\phi'_{\mu_0}\left(\mathbb{G}_0\right)$. Since conditional weak convergence implies convergence in quantiles, $\hat{c}_{1-\alpha}^* = c_{1-\alpha}^* + o_p\left(1\right)$. It follows that

$$
\begin{aligned}
\limsup_{n\to\infty} P^n\left(\hat{\mathcal{J}}_n \geq \hat{c}_{1-\alpha}^*\right) &\leq \limsup_{n\to\infty} P^n\left(\phi'_{\mu_0}\left(\mathbb{G}_0 + c\right) - \phi'_{\mu_0}\left(c\right) \geq c_{1-\alpha}^*\right) \\
&\leq \limsup_{n\to\infty} P^n\left(\phi'_{\mu_0}\left(\mathbb{G}_0\right) \geq c_{1-\alpha}^*\right) \\
&= \alpha
\end{aligned}
$$

Therefore, the lower one-sided confidence interval $\left[\phi\left(\hat{\mu}_n\right) - \frac{\hat{c}^*_{1-\alpha}}{\sqrt{n}}, \infty\right)$ will be asymptotically conservatively valid. Similarly, if ϕ'_{μ_0} is superadditive (meaning $\phi'_{\mu_0}\left(h_1 + h_2\right) \geq \phi'_{\mu_0}\left(h_1\right) + \phi'_{\mu_0}\left(h_2\right)$ for all $h_1, h_2 \in \mathbb{D}_0$), then the limiting distribution of $\hat{\mathcal{J}}_n$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n^*$. Similar arguments show that $\underset{n\to\infty}{limsup} P^n\left(\hat{\mathcal{J}}_n \leq \hat{c}^*_\alpha\right) \leq \alpha$, which implies that the upper one-sided confidence interval $\left(-\infty, \phi\left(\hat{\mu}_n\right) - \frac{\hat{c}^*_\alpha}{\sqrt{n}}\right]$ is asymptotically conservatively valid.

If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to 0$, then $\hat{\mathcal{J}}_n \rightsquigarrow \underset{\kappa_n\to\infty}{lim} \phi'_{\mu_0}\left(\mathbb{G}_0 + sign(c)\kappa_n\right) - \phi'_{\mu_0}\left(sign(c)\kappa_n\right)$ while $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mu_0}\left(\mathbb{G}_0\right)$. The limiting distribution of $\hat{\mathcal{J}}_n^*$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n$ if ϕ'_{μ_0} is subadditive, in which case the lower one-sided confidence intervals are conservatively valid. If ϕ'_{μ_0} is superadditive, the upper one-sided confidence intervals are conservatively valid.

If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to 1$, then $\hat{\mathcal{J}}_n \rightsquigarrow \underset{\kappa_n\to\infty}{lim} \phi'_{\mu_0}\left(\mathbb{G}_0 + sign(c)\kappa_n\right) - \phi'_{\mu_0}\left(sign(c)\kappa_n\right)$ while $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mu_0}\left(\mathbb{G}_0 + c\right) - \phi'_{\mu_0}\left(c\right)$. Without additional assumptions, we cannot say which intervals will be asymptotically valid. In the negative of the mean example, it turns out that two-sided equal-tailed and one-sided intervals are asymptotically exact.

If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to \infty$, then the limiting distributions of $\hat{\mathcal{J}}_n$ and $\hat{\mathcal{J}}_n^*$ coincide, and two-sided equal-tailed and one-sided intervals are asymptotically exact. ■

3.2 Negative Part of the Mean Example

We now discuss local asymptotics for the example $\phi\left(\mu\right) = \mu^-$, which has directional derivative $\phi'_\mu\left(h\right) = -h1\left(\mu < 0\right) + h^- 1\left(\mu = 0\right)$. Note that if $\mu_0 < 0$, then $\phi'_{\mu_0}\left(h\right)$ is a linear function of h, and both two-sided equal-tailed and one-sided intervals are asymptotically exact for all a_n and ϵ_n. When $\mu_0 = 0$, note that ϕ'_{μ_0} is subadditive because $\left(h_1 + h_2\right)^- \leq h_1^- + h_2^-$. It follows that for $\sqrt{n}a_n = 1$ or $\sqrt{n}a_n \to \infty$ and $a_n/\epsilon_n \to 0$, the lower one-sided confidence interval $\left[\phi\left(\hat{\mu}_n\right) - \frac{\hat{c}^*_{1-\alpha}}{\sqrt{n}}, \infty\right)$ will be asymptotically conservative. An example of such a sequence is $a_n = \frac{1}{\sqrt{n}\epsilon_n}$ where $\frac{1}{\sqrt{n}\epsilon_n^2} \to 0$. If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to \infty$, then two-sided equal-tailed and one-sided intervals are asymptotically exact. An example of such a sequence is $a_n = \frac{1}{\sqrt{n}\epsilon_n}$ where $\frac{1}{\sqrt{n}\epsilon_n^2} \to \infty$. If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to 1$, for example when $a_n = \frac{1}{\sqrt{n}\epsilon_n}$ and $\frac{1}{\sqrt{n}\epsilon_n^2} \to 1$, then $\hat{\mathcal{J}}_n \rightsquigarrow \begin{cases} 0 & \text{if } c > 0 \\ \mathbb{G}_0 & \text{if } c < 0 \end{cases}$ and $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \begin{cases} \left(\mathbb{G}_0 + c\right)^- & \text{if } c > 0 \\ \left(\mathbb{G}_0 + c\right)^- + c & \text{if } c < 0 \end{cases}$. In the case of $c > 0$, the asymptotic distribution of $\hat{\mathcal{J}_n}^*$ lies strictly to the right of the asymptotic distribution of $\hat{\mathcal{J}}_n$, which implies that all types of intervals are conservative. In the case of $c < 0$, we have for $x \geq c$, $P\left((\mathbb{G}_0 + c)^- + c \leq x\right) = P\left(\mathbb{G}_0 + c \geq -(x - c)\right) = P\left(\mathbb{G}_0 \leq x\right)$. If $x < c$, $P\left((\mathbb{G}_0 + c)^- + c \leq x\right) = P\left(\mathbb{G}_0 \leq x\right) = 0$. It follows that two-sided equal-tailed and one-sided intervals will be asymptotically valid.

3.3 LASSO in the one-dimensional mean model

Consider the LASSO estimator in the one-dimensional mean model:

$$\hat{\beta} = \underset{\beta}{argmin} \frac{1}{2n} \|Y - \iota\beta\|_2^2 + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1$$

where ι is the $n \times 1$ vector of 1's and λ_n satisfies $\lambda_n \to \lambda_0 \in [0, \infty)$ and $\lambda_n/\sqrt{n} \to 0$. Define $\hat{\beta}^*$ as the numerical bootstrap LASSO estimator:

$$\hat{\beta}^* = \underset{\beta}{argmin} \frac{1}{2n} \|Y - \iota\beta\|_2^2 + \epsilon_n\sqrt{n}\left(\frac{1}{2n} \|Y^* - \iota\beta\|_2^2 - \frac{1}{2n} \|Y - \iota\beta\|_2^2\right) + \lambda_n\epsilon_n \|\beta\|_1$$

where we impose the additional requirement that $\lambda_n\epsilon_n \to 0$. It is easy to show that both $\hat{\beta}$ and $\hat{\beta}^*$ have closed form solutions in terms of the sample means of the data $(\bar{Y})$ and bootstrap samples $(\bar{Y}^*)$:

$$\hat{\beta} = \left(\bar{Y} - \lambda_n/\sqrt{n}\right)^+ - \left(\bar{Y} + \lambda_n/\sqrt{n}\right)^-$$
$$\hat{\beta}^* = \left(\bar{Y} + \epsilon_n\sqrt{n}\left(\bar{Y}^* - \bar{Y}\right) - \lambda_n\epsilon_n\right)^+ - \left(\bar{Y} + \epsilon_n\sqrt{n}\left(\bar{Y}^* - \bar{Y}\right) + \lambda_n\epsilon_n\right)^-$$

Define $\hat{\mathcal{J}}_n \equiv \sqrt{n}\left(\hat{\beta} - \beta_n\right)$ for a drifting sequence $\beta_n = \beta_0 + a_n c$ (where $a_n \downarrow 0$ and c is a constant) and let $\hat{c}_\alpha^*$ be the α-th quantile of $\hat{\mathcal{J}_n}^* \equiv \frac{\hat{\beta}^* - \hat{\beta}}{\epsilon_n}$. In the following theorem, we describe the coverage properties of the following three kinds of confidence intervals: two-sided equal-tailed $\left[\hat{\beta} - \frac{\hat{c}_{1-\alpha}^*}{\sqrt{n}}, \hat{\beta} - \frac{\hat{c}_\alpha^*}{\sqrt{n}}\right]$, lower one-sided $\left[\hat{\beta} - \frac{\hat{c}_{1-\alpha}^*}{\sqrt{n}}, \infty\right)$, and upper one-sided $\left(-\infty, \hat{\beta} - \frac{\hat{c}_\alpha^*}{\sqrt{n}}\right]$.

Theorem 3.1 *If $c > 0 (< 0)$, the lower (upper) one-sided confidence interval will be conservatively valid for the following types of sequences: (i) $a_n\sqrt{n} = 1$, (ii) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to 0$, (iii) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to 1$. Two sided equal-tailed and one-sided intervals are asymptotically exact for (i) $a_n\sqrt{n} \to 0$ (ii) $a_n\sqrt{n} \to \infty$ and $a_n/\epsilon_n \to \infty$.*

Proof: For $\sqrt{n}\left(\bar{Y}-\beta_n\right) \rightsquigarrow \mathbb{G}_0$ and $\lambda_n \to \lambda_0$,

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}-\beta_n\right) &= \sqrt{n}\left(\left(\bar{Y}-\lambda_n/\sqrt{n}\right)^+ - \left(\bar{Y}+\lambda_n/\sqrt{n}\right)^- - \beta_n\right) \\
&= \left(\sqrt{n}\beta_0 + \sqrt{n}\left(\bar{Y}-\beta_n\right) + \sqrt{n}\left(\beta_n-\beta_0\right) - \lambda_n\right)^+ - \sqrt{n}\beta_0^+ \\
&- \left(\sqrt{n}\beta_0 + \sqrt{n}\left(\bar{Y}-\beta_n\right) + \sqrt{n}\left(\beta_n-\beta_0\right) + \lambda_n\right)^- + \sqrt{n}\beta_0^- - \sqrt{n}\left(\beta_n-\beta_0\right) \\
&\rightsquigarrow 1\left(\beta_0=0\right)\left(\left(\mathbb{G}_0 + \lim_{n\to\infty}\sqrt{n}a_n c - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lim_{n\to\infty}\sqrt{n}a_n c + \lambda_0\right)^- - \lim_{n\to\infty}\sqrt{n}a_n c\right) \\
&+ 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)
\end{aligned}$$

For $\sqrt{n}\left(\bar{Y}^*-\bar{Y}\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_0$,

$$\begin{aligned}
\frac{\hat{\beta}^*-\hat{\beta}}{\epsilon_n} &= \frac{1}{\epsilon_n}\left(\left(\bar{Y}+\epsilon_n\sqrt{n}\left(\bar{Y}^*-\bar{Y}\right)-\lambda_n\epsilon_n\right)^+ - \left(\bar{Y}+\epsilon_n\sqrt{n}\left(\bar{Y}^*-\bar{Y}\right)+\lambda_n\epsilon_n\right)^- - \hat{\beta}\right) \\
&= \frac{1}{\epsilon_n}\left(\left(\bar{Y}+\epsilon_n\sqrt{n}\left(\bar{Y}^*-\bar{Y}\right)-\lambda_n\epsilon_n\right)^+ - \left(\bar{Y}+\epsilon_n\sqrt{n}\left(\bar{Y}^*-\bar{Y}\right)+\lambda_n\epsilon_n\right)^- - \beta_n\right) + \frac{\hat{\beta}-\beta_n}{\epsilon_n} \\
&= \left(\frac{\beta_0}{\epsilon_n} + \sqrt{n}\left(\bar{Y}^*-\bar{Y}\right) + \frac{\bar{Y}-\beta_n}{\epsilon_n} + \frac{\beta_n-\beta_0}{\epsilon_n} - \lambda_n\right)^+ \\
&- \left(\frac{\beta_0}{\epsilon_n} + \sqrt{n}\left(\bar{Y}^*-\bar{Y}\right) + \frac{\bar{Y}-\beta_n}{\epsilon_n} + \frac{\beta_n-\beta_0}{\epsilon_n} + \lambda_n\right)^- \\
&- \frac{\beta_n-\beta_0}{\epsilon_n} - \left(\frac{\beta_0}{\epsilon_n}\right)^+ + \left(\frac{\beta_0}{\epsilon_n}\right)^- + o_p(1) \\
&\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} 1\left(\beta_0=0\right)\left(\left(\mathbb{G}_0 + \lim_{n\to\infty}\frac{a_n}{\epsilon_n}c - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lim_{n\to\infty}\frac{a_n}{\epsilon_n}c + \lambda_0\right)^- - \lim_{n\to\infty}\frac{a_n}{\epsilon_n}c\right) \\
&+ 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)
\end{aligned}$$

where the $o_p(1)$ term refers to $\frac{\hat{\beta}-\beta_n}{\epsilon_n}$. If $a_n\sqrt{n} \to 0$, then the limiting distributions of $\hat{\mathcal{J}}_n$ and $\hat{\mathcal{J}}_n{}^*$ both equal

$$1\left(\beta_0=0\right)\left(\left(\mathbb{G}_0-\lambda_0\right)^+ - \left(\mathbb{G}_0+\lambda_0\right)^-\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

Two-sided equal-tailed and one-sided intervals are asymptotically exact up to first order asymptotics.

If $a_n = \frac{1}{\sqrt{n}}$, then

$$\hat{\mathcal{J}}_n \rightsquigarrow 1\left(\beta_0 = 0\right)\left(\left(\mathbb{G}_0 + c - \lambda_0\right)^+ - \left(\mathbb{G}_0 + c + \lambda_0\right)^- - c\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

while

$$\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} 1\left(\beta_0 = 0\right)\left(\left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

If $c > 0$, then $\left(\mathbb{G}_0 + c - \lambda_0\right)^+ - \left(\mathbb{G}_0 + c + \lambda_0\right)^- - c \leq \left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-$ for all values of $\mathbb{G}_0$, so the limiting distribution of $\hat{\mathcal{J}}_n^*$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n$. Therefore, the lower one-sided confidence interval $\left[\hat{\beta}_n - \frac{\hat{c}^*_{1-\alpha}}{\sqrt{n}}, \infty\right)$ will be asymptotically conservatively valid. If $c < 0$, then $\left(\mathbb{G}_0 + c - \lambda_0\right)^+ - \left(\mathbb{G}_0 + c + \lambda_0\right)^- - c \geq \left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-$ for all values of $\mathbb{G}_0$, so the limiting distribution of $\hat{\mathcal{J}}_n$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n^*$. Then the upper one-sided confidence interval $\left(-\infty, \hat{\beta}_n - \frac{\hat{c}^*_{\alpha}}{\sqrt{n}}\right]$ is asymptotically conservatively valid.

If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to 0$, then

$$\hat{\mathcal{J}}_n \rightsquigarrow 1\left(\beta_0 = 0\right)\left(\mathbb{G}_0 - sign(c)\lambda_0\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

while

$$\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} 1\left(\beta_0 = 0\right)\left(\left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

If $c > 0$, then $\mathbb{G}_0 - sign(c)\lambda_0 \leq \left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-$ for all values of $\mathbb{G}_0$, so the limiting distribution of $\hat{\mathcal{J}}_n^*$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n$ and the lower one-sided confidence intervals are conservatively valid. If $c < 0$, then $\mathbb{G}_0 - sign(c)\lambda_0 \geq \left(\mathbb{G}_0 - \lambda_0\right)^+ - \left(\mathbb{G}_0 + \lambda_0\right)^-$ for all values of $\mathbb{G}_0$ and the upper one-sided confidence intervals are conservatively valid.

If $\sqrt{n}a_n \to \infty$ and $\frac{a_n}{\epsilon_n} \to 1$, then

$$\hat{\mathcal{J}}_n \rightsquigarrow 1\left(\beta_0 = 0\right)\left(\mathbb{G}_0 - sign(c)\lambda_0\right) + 1\left(\beta_0 \neq 0\right)\left(\mathbb{G}_0 - sign\left(\beta_0\right)\lambda_0\right)$$

while

$$\hat{\mathcal{J}}_n^{\ *} \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} 1\left(\beta_0=0\right)\left(\left(\mathbb{G}_0+c-\lambda_0\right)^+-\left(\mathbb{G}_0+c+\lambda_0\right)^--c\right)+1\left(\beta_0\neq 0\right)\left(\mathbb{G}_0-sign\left(\beta_0\right)\lambda_0\right)$$

If $c>0$, then $\mathbb{G}_0-sign(c)\lambda_0\leq\left(\mathbb{G}_0+c-\lambda_0\right)^+-\left(\mathbb{G}_0+c+\lambda_0\right)^--c$ for all values of $\mathbb{G}_0$, so the limiting distribution of $\hat{\mathcal{J}}_n^{\ *}$ first order stochastically dominates the limiting distribution of $\hat{\mathcal{J}}_n$ and the lower one-sided confidence intervals are conservatively valid. If $c<0$, then $\mathbb{G}_0-sign(c)\lambda_0\geq\left(\mathbb{G}_0+c-\lambda_0\right)^+-\left(\mathbb{G}_0+c+\lambda_0\right)^--c$ for all values of $\mathbb{G}_0$ and the upper one-sided confidence intervals are conservatively valid.

If $\sqrt{n}a_n\to\infty$ and $\frac{a_n}{\epsilon_n}\to\infty$, then $\hat{\mathcal{J}}_n$ and $\hat{\mathcal{J}}_n^{\ *}$ have the same limiting distribution of $1\left(\beta_0=0\right)\left(\mathbb{G}_0-sign(c)\lambda_0\right)+1\left(\beta_0\neq 0\right)\left(\mathbb{G}_0-sign\left(\beta_0\right)\lambda_0\right)$, and two-sided equal-tailed and one-sided intervals are asymptotically exact.

4 Monte Carlo Simulation Results for LASSO Mean Example

In this section, we investigate the empirical coverage frequencies of confidence intervals formed using the following bootstrap estimators:

$$\begin{aligned}
\text{Standard: } \hat{\beta}^*_{\text{standard}} &= \left(\bar{Y}^*_n-\lambda_n/\sqrt{n}\right)^+-\left(\bar{Y}^*_n+\lambda_n/\sqrt{n}\right)^- \\
\text{Perturbation: } \hat{\beta}^*_{\text{perturb}} &= \left(\frac{\frac{1}{n}\sum_{i=1}^n Y_iG_i^*}{\frac{1}{n}\sum_{i=1}^n G_i^*}-\frac{\lambda_n}{\sqrt{n}}\right)^+-\left(\frac{\frac{1}{n}\sum_{i=1}^n Y_iG_i^*}{\frac{1}{n}\sum_{i=1}^n G_i^*}+\frac{\lambda_n}{\sqrt{n}}\right)^-, \quad G_i^*\sim exp(1) \\
m\text{-out-of-}n\text{: } \hat{\beta}^*_{\text{moutofn}} &= \left(\bar{Y}^*_m-\lambda_m/\sqrt{m}\right)^+-\left(\bar{Y}^*_m+\lambda_m/\sqrt{m}\right)^- \\
\text{Numerical: } \hat{\beta}^*_{\text{numerical}} &= \left(\bar{Y}_n+\epsilon_n\sqrt{n}\left(\bar{Y}^*_n-\bar{Y}_n\right)-\lambda_n\epsilon_n\right)^+-\left(\bar{Y}_n+\epsilon_n\sqrt{n}\left(\bar{Y}^*_n-\bar{Y}_n\right)+\lambda_n\epsilon_n\right)^-
\end{aligned}$$

Recall that $\hat{\beta}=\left(\bar{Y}_n-\lambda_n/\sqrt{n}\right)^+-\left(\bar{Y}_n+\lambda_n/\sqrt{n}\right)^-$ and let $\hat{c}^*_\alpha$ denote the α-th percentile of the empirical distributions of the above bootstrap estimators (properly centered and scaled). Tables 1 to 4 below show the empirical coverage frequencies of two-sided equal-tailed $\left[\hat{\beta}-\frac{\hat{c}^*_{1-\alpha}}{\sqrt{n}},\hat{\beta}-\frac{\hat{c}^*_\alpha}{\sqrt{n}}\right]$ intervals (with the average widths of the intervals in parentheses) for $\alpha=0.05$. Tables 5 to 12 show the empirical coverage frequencies of lower one-sided $\left[\hat{\beta}-\frac{\hat{c}^*_{1-\alpha}}{\sqrt{n}},\infty\right)$ intervals and upper one-sided $\left(-\infty,\hat{\beta}-\frac{\hat{c}^*_\alpha}{\sqrt{n}}\right]$ intervals.

The standard bootstrap generally behaves similarly to the perturbation bootstrap, while the numerical bootstrap generally outperforms the m-out-of-n bootstrap. Numerical bootstrap's improvement over the m-out-of-n bootstrap is more pronounced for larger values of ϵ_n and for larger values of λ_n. The better performance of the numerical bootstrap holds for both equal-tailed and one-sided intervals. When $\lambda_n = n^{-1/2}$, the numerical bootstrap behaves similarly to the perturbation bootstrap and standard bootstrap, achieving coverage close to the nominal level. For $\lambda_n = n^{-1/4}$ or $\lambda_n = n^{-1/6}$, the perturbation bootstrap and standard bootstrap intervals overcover for $\beta_0 \in \{0, n^{-1}, n^{-1/2}\}$ while the numerical bootstrap intervals still achieve coverage close to the nominal level for $n = 1000$ or $n = 10000$. For $n = 100$, the numerical bootstrap intervals have coverage that differ from the nominal level, although not as much as the perturbation bootstrap intervals.

Table 1: Standard and Perturbation Bootstrap Equal-Tailed Coverage Frequencies

	Standard Bootstrap				$\lambda_n = n^{-1/2}$	Perturbation Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.913	0.952	0.952	0.952	0.939	0.914	0.955	0.955	0.955	0.94
	(0.384)	(0.368)	(0.368)	(0.371)	(0.387)	(0.383)	(0.367)	(0.367)	(0.37)	(0.386)
$n = 1000$	0.944	0.948	0.948	0.948	0.944	0.943	0.948	0.948	0.948	0.945
	(0.124)	(0.122)	(0.122)	(0.122)	(0.124)	(0.124)	(0.122)	(0.122)	(0.122)	(0.124)
$n = 10000$	0.949	0.95	0.95	0.95	0.949	0.951	0.952	0.952	0.952	0.951
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
	Standard Bootstrap				$\lambda_n = n^{-1/4}$	Perturbation Bootstrap				
$n = 100$	0.876	0.966	0.966	0.964	0.919	0.88	0.972	0.972	0.97	0.918
	(0.378)	(0.327)	(0.327)	(0.335)	(0.386)	(0.378)	(0.328)	(0.328)	(0.335)	(0.386)
$n = 1000$	0.940	0.962	0.962	0.962	0.938	0.910	0.951	0.951	0.951	0.942
	(0.124)	(0.113)	(0.113)	(0.114)	(0.124)	(0.124)	(0.113)	(0.113)	(0.114)	(0.124)
$n = 10000$	0.949	0.96	0.96	0.96	0.949	0.949	0.961	0.961	0.961	0.949
	(0.039)	(0.037)	(0.037)	(0.037)	(0.039)	(0.039)	(0.037)	(0.037)	(0.038)	(0.039)
	Standard Bootstrap				$\lambda_n = n^{-1/6}$	Perturbation Bootstrap				
$n = 100$	0.857	0.979	0.979	0.931	0.914	0.856	0.978	0.978	0.928	0.906
	(0.376)	(0.300)	(0.300)	(0.312)	(0.387)	(0.375)	(0.299)	(0.299)	(0.310)	(0.385)
$n = 1000$	0.931	0.976	0.976	0.976	0.937	0.932	0.975	0.975	0.975	0.936
	(0.124)	(0.105)	(0.105)	(0.107)	(0.124)	(0.110)	(0.104)	(0.104)	(0.104)	(0.124)
$n = 10000$	0.944	0.97	0.97	0.97	0.944	0.944	0.97	0.97	0.97	0.944
	(0.039)	(0.035)	(0.035)	(0.036)	(0.039)	(0.039)	(0.035)	(0.035)	(0.036)	(0.039)

Table 2: m-out-of-n and Numerical Bootstrap Equal-Tailed Coverage for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$		Numerical Bootstrap			
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.894	0.932	0.932	0.932	0.944	0.916	0.947	0.947	0.947	0.939
	(0.352)	(0.344)	(0.344)	(0.344)	(0.387)	(0.371)	(0.367)	(0.367)	(0.367)	(0.387)
$n = 1000$	0.931	0.936	0.936	0.936	0.943	0.940	0.945	0.945	0.945	0.945
	(0.119)	(0.117)	(0.117)	(0.117)	(0.124)	(0.122)	(0.122)	(0.122)	(0.122)	(0.124)
$n = 10000$	0.944	0.946	0.946	0.946	0.95	0.949	0.951	0.951	0.951	0.951
	(0.039)	(0.038)	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$		Numerical Bootstrap			
$n = 100$	0.8	0.932	0.932	0.931	0.932	0.842	0.949	0.949	0.949	0.927
	(0.314)	(0.294)	(0.294)	(0.295)	(0.386)	(0.336)	(0.323)	(0.323)	(0.324)	(0.387)
$n = 1000$	0.884	0.932	0.932	0.932	0.943	0.910	0.951	0.951	0.951	0.942
	(0.110)	(0.104)	(0.104)	(0.104)	(0.124)	(0.116)	(0.112)	(0.112)	(0.112)	(0.124)
$n = 10000$	0.922	0.937	0.937	0.937	0.947	0.933	0.951	0.951	0.951	0.949
	(0.038)	(0.035)	(0.035)	(0.035)	(0.039)	(0.039)	(0.037)	(0.037)	(0.037)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$		Numerical Bootstrap			
$n = 100$	0.719	0.953	0.953	0.925	0.944	0.762	0.961	0.961	0.961	0.931
	(0.295)	(0.268)	(0.268)	(0.269)	(0.388)	(0.313)	(0.294)	(0.294)	(0.295)	(0.387)
$n = 1000$	0.824	0.939	0.939	0.939	0.945	0.863	0.955	0.955	0.955	0.942
	(0.104)	(0.094)	(0.094)	(0.094)	(0.124)	(0.110)	(0.104)	(0.104)	(0.104)	(0.124)
$n = 10000$	0.906	0.937	0.937	0.937	0.948	0.916	0.956	0.956	0.956	0.95
	(0.037)	(0.032)	(0.032)	(0.032)	(0.039)	(0.038)	(0.035)	(0.035)	(0.035)	(0.039)

Table 3: m-out-of-n and Numerical Bootstrap Equal-Tailed Coverage for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$		Numerical Bootstrap			
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.879	0.909	0.909	0.909	0.935	0.925	0.951	0.951	0.951	0.946
	(0.325)	(0.324)	(0.324)	(0.324)	(0.387)	(0.368)	(0.368)	(0.368)	(0.368)	(0.388)
$n = 1000$	0.915	0.923	0.923	0.923	0.941	0.939	0.943	0.943	0.943	0.943
	(0.112)	(0.112)	(0.112)	(0.112)	(0.124)	(0.122)	(0.122)	(0.122)	(0.122)	(0.124)
$n = 10000$	0.936	0.938	0.938	0.938	0.949	0.949	0.95	0.95	0.95	0.95
	(0.037)	(0.037)	(0.037)	(0.037)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$		Numerical Bootstrap			
$n = 100$	0.774	0.91	0.91	0.908	0.934	0.848	0.942	0.942	0.942	0.928
	(0.278)	(0.274)	(0.274)	(0.275)	(0.387)	(0.324)	(0.324)	(0.324)	(0.324)	(0.387)
$n = 1000$	0.86	0.912	0.912	0.912	0.939	0.912	0.948	0.948	0.948	0.943
	(0.097)	(0.097)	(0.097)	(0.097)	(0.124)	(0.112)	(0.112)	(0.112)	(0.112)	(0.124)
$n = 10000$	0.895	0.92	0.92	0.92	0.945	0.934	0.951	0.951	0.951	0.951
	(0.033)	(0.033)	(0.033)	(0.033)	(0.039)	(0.037)	(0.037)	(0.037)	(0.037)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$		Numerical Bootstrap			
$n = 100$	0.695	0.935	0.935	0.907	0.944	0.787	0.96	0.96	0.959	0.937
	(0.255)	(0.25)	(0.25)	(0.251)	(0.387)	(0.296)	(0.294)	(0.294)	(0.294)	(0.387)
$n = 1000$	0.797	0.915	0.915	0.915	0.94	0.867	0.952	0.952	0.952	0.945
	(0.088)	(0.088)	(0.088)	(0.088)	(0.124)	(0.104)	(0.104)	(0.104)	(0.104)	(0.124)
$n = 10000$	0.855	0.916	0.916	0.916	0.942	0.908	0.952	0.952	0.952	0.949
	(0.03)	(0.03)	(0.03)	(0.03)	(0.039)	(0.035)	(0.035)	(0.035)	(0.035)	(0.039)

Table 4: m-out-of-n and Numerical Bootstrap Equal-Tailed Coverage for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$		Numerical Bootstrap			
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.846	0.874	0.874	0.874	0.92	0.916	0.943	0.943	0.943	0.942
	(0.297)	(0.297)	(0.297)	(0.297)	(0.387)	(0.367)	(0.367)	(0.367)	(0.367)	(0.387)
$n = 1000$	0.892	0.899	0.899	0.899	0.932	0.94	0.944	0.944	0.944	0.943
	(0.104)	(0.104)	(0.104)	(0.104)	(0.124)	(0.122)	(0.122)	(0.122)	(0.122)	(0.124)
$n = 10000$	0.919	0.921	0.921	0.921	0.946	0.949	0.95	0.95	0.95	0.95
	(0.035)	(0.035)	(0.035)	(0.035)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$		Numerical Bootstrap			
$n = 100$	0.751	0.877	0.877	0.867	0.925	0.848	0.939	0.939	0.939	0.929
	(0.254)	(0.253)	(0.253)	(0.253)	(0.387)	(0.324)	(0.324)	(0.324)	(0.324)	(0.387)
$n = 1000$	0.823	0.881	0.881	0.881	0.927	0.913	0.945	0.945	0.945	0.943
	(0.088)	(0.088)	(0.088)	(0.088)	(0.124)	(0.112)	(0.112)	(0.112)	(0.112)	(0.124)
$n = 10000$	0.861	0.89	0.89	0.89	0.934	0.934	0.95	0.95	0.95	0.949
	(0.03)	(0.03)	(0.03)	(0.03)	(0.039)	(0.037)	(0.037)	(0.037)	(0.037)	(0.039)
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$		Numerical Bootstrap			
$n = 100$	0.668	0.909	0.909	0.856	0.935	0.796	0.954	0.954	0.954	0.942
	(0.235)	(0.234)	(0.234)	(0.234)	(0.387)	(0.294)	(0.294)	(0.294)	(0.294)	(0.387)
$n = 1000$	0.763	0.89	0.89	0.89	0.932	0.874	0.948	0.948	0.948	0.943
	(0.081)	(0.081)	(0.081)	(0.081)	(0.124)	(0.104)	(0.104)	(0.104)	(0.104)	(0.124)
$n = 10000$	0.811	0.886	0.886	0.886	0.932	0.909	0.952	0.952	0.952	0.952
	(0.027)	(0.027)	(0.027)	(0.027)	(0.039)	(0.035)	(0.035)	(0.035)	(0.035)	(0.039)

Table 5: Standard and Perturbation Bootstrap Upper Coverage Frequencies

	Standard Bootstrap				$\lambda_n = n^{-1/2}$	Perturbation Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.952	0.952	0.952	0.952	0.931	0.954	0.954	0.954	0.954	0.93
$n = 1000$	0.952	0.952	0.952	0.952	0.947	0.952	0.952	0.952	0.952	0.948
$n = 10000$	0.949	0.949	0.949	0.949	0.947	0.95	0.95	0.95	0.95	0.947
	Standard Bootstrap				$\lambda_n = n^{-1/4}$	Perturbation Bootstrap				
$n = 100$	0.969	0.969	0.969	0.933	0.889	0.965	0.965	0.965	0.932	0.887
$n = 1000$	0.963	0.963	0.963	0.963	0.932	0.963	0.963	0.963	0.963	0.932
$n = 10000$	0.959	0.959	0.959	0.959	0.940	0.958	0.958	0.958	0.958	0.940
	Standard Bootstrap				$\lambda_n = n^{-1/6}$	Perturbation Bootstrap				
$n = 100$	0.978	0.978	0.978	0.777	0.874	0.980	0.980	0.980	0.775	0.875
$n = 1000$	0.974	0.974	0.974	0.966	0.912	0.975	0.975	0.975	0.966	0.910
$n = 10000$	0.969	0.969	0.969	0.969	0.926	0.968	0.968	0.968	0.968	0.925

Table 6: Standard and Perturbation Bootstrap Lower Coverage Frequencies

	Standard Bootstrap				$\lambda_n = n^{-1/2}$	Perturbation Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.922	0.955	0.955	0.955	0.955	0.919	0.952	0.952	0.952	0.952
$n = 1000$	0.939	0.947	0.947	0.947	0.947	0.940	0.947	0.947	0.947	0.947
$n = 10000$	0.949	0.952	0.952	0.952	0.952	0.949	0.951	0.951	0.951	0.951
	Standard Bootstrap				$\lambda_n = n^{-1/4}$	Perturbation Bootstrap				
$n = 100$	0.885	0.964	0.964	0.964	0.964	0.883	0.964	0.964	0.964	0.964
$n = 1000$	0.923	0.961	0.961	0.961	0.961	0.924	0.961	0.961	0.961	0.961
$n = 10000$	0.937	0.960	0.960	0.960	0.960	0.936	0.960	0.960	0.960	0.960
	Standard Bootstrap				$\lambda_n = n^{-1/6}$	Perturbation Bootstrap				
$n = 100$	0.862	0.980	0.980	0.980	0.980	0.861	0.980	0.980	0.980	0.980
$n = 1000$	0.903	0.969	0.969	0.969	0.969	0.905	0.970	0.970	0.970	0.970
$n = 10000$	0.922	0.968	0.968	0.968	0.968	0.922	0.969	0.969	0.969	0.969

Table 7: m-out-of-n and Numerical Bootstrap Upper Coverage for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.928	0.928	0.928	0.928	0.948	0.946	0.946	0.946	0.946	0.933
$n = 1000$	0.945	0.945	0.945	0.945	0.953	0.950	0.950	0.950	0.950	0.949
$n = 10000$	0.946	0.946	0.946	0.946	0.952	0.949	0.949	0.949	0.949	0.948
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.937	0.937	0.937	0.895	0.937	0.952	0.952	0.952	0.947	0.916
$n = 1000$	0.942	0.942	0.942	0.942	0.955	0.954	0.954	0.954	0.954	0.945
$n = 10000$	0.940	0.940	0.940	0.940	0.958	0.951	0.951	0.951	0.951	0.946
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.949	0.949	0.949	0.764	0.927	0.963	0.963	0.963	0.883	0.912
$n = 1000$	0.945	0.945	0.945	0.943	0.953	0.956	0.956	0.956	0.956	0.940
$n = 10000$	0.940	0.940	0.940	0.940	0.958	0.953	0.953	0.953	0.953	0.945

Table 8: m-out-of-n and Numerical Bootstrap Lower Coverage for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.891	0.93	0.93	0.93	0.93	0.915	0.947	0.947	0.947	0.947
$n = 1000$	0.928	0.937	0.937	0.937	0.937	0.936	0.945	0.945	0.945	0.945
$n = 10000$	0.953	0.947	0.947	0.947	0.947	0.948	0.948	0.948	0.948	0.948
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.762	0.94	0.94	0.94	0.94	0.794	0.954	0.954	0.954	0.954
$n = 1000$	0.871	0.935	0.935	0.935	0.935	0.884	0.948	0.948	0.948	0.948
$n = 10000$	0.953	0.938	0.938	0.938	0.938	0.946	0.950	0.950	0.950	0.950
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.702	0.961	0.961	0.961	0.961	0.71	0.97	0.97	0.97	0.97
$n = 1000$	0.842	0.939	0.939	0.939	0.939	0.842	0.954	0.954	0.954	0.954
$n = 10000$	0.939	0.939	0.939	0.939	0.939	0.939	0.955	0.955	0.955	0.955

Table 9: m-out-of-n and Numerical Bootstrap Upper Coverage for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.916	0.916	0.916	0.914	0.961	0.946	0.946	0.946	0.946	0.938
$n = 1000$	0.935	0.935	0.935	0.935	0.96	0.95	0.95	0.95	0.95	0.949
$n = 10000$	0.941	0.941	0.941	0.941	0.958	0.949	0.949	0.949	0.949	0.949
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.918	0.918	0.918	0.844	0.952	0.951	0.951	0.951	0.945	0.923
$n = 1000$	0.928	0.928	0.928	0.928	0.966	0.952	0.952	0.952	0.952	0.947
$n = 10000$	0.928	0.928	0.928	0.928	0.968	0.949	0.949	0.949	0.949	0.947
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.933	0.933	0.933	0.715	0.95	0.955	0.955	0.955	0.899	0.921
$n = 1000$	0.931	0.931	0.931	0.913	0.964	0.954	0.954	0.954	0.954	0.944
$n = 10000$	0.926	0.926	0.926	0.926	0.97	0.95	0.95	0.95	0.95	0.946

Table 10: m-out-of-n and Numerical Bootstrap Lower Coverage for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.878	0.918	0.918	0.918	0.918	0.913	0.945	0.945	0.945	0.945
$n = 1000$	0.918	0.927	0.927	0.927	0.927	0.935	0.944	0.944	0.944	0.944
$n = 10000$	0.936	0.939	0.939	0.939	0.939	0.947	0.949	0.949	0.949	0.949
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.739	0.921	0.921	0.921	0.921	0.808	0.947	0.947	0.947	0.947
$n = 1000$	0.836	0.919	0.919	0.919	0.919	0.889	0.947	0.947	0.947	0.947
$n = 10000$	0.888	0.923	0.923	0.923	0.923	0.923	0.95	0.95	0.95	0.95
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.646	0.936	0.936	0.936	0.936	0.699	0.965	0.965	0.965	0.965
$n = 1000$	0.76	0.921	0.921	0.921	0.921	0.822	0.949	0.949	0.949	0.949
$n = 10000$	0.827	0.921	0.921	0.921	0.921	0.882	0.952	0.952	0.952	0.952

Table 11: m-out-of-n and Numerical Bootstrap Upper Coverage for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.889	0.889	0.889	0.88	0.967	0.943	0.943	0.943	0.943	0.94
$n = 1000$	0.918	0.918	0.918	0.918	0.973	0.951	0.951	0.951	0.951	0.951
$n = 10000$	0.929	0.929	0.929	0.929	0.968	0.948	0.948	0.948	0.948	0.948
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.894	0.894	0.894	0.762	0.964	0.942	0.942	0.942	0.938	0.929
$n = 1000$	0.902	0.902	0.902	0.896	0.978	0.951	0.951	0.951	0.951	0.948
$n = 10000$	0.901	0.901	0.901	0.901	0.977	0.949	0.949	0.949	0.949	0.948
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.918	0.918	0.918	0.67	0.962	0.955	0.955	0.955	0.916	0.925
$n = 1000$	0.906	0.906	0.906	0.76	0.976	0.953	0.953	0.953	0.953	0.946
$n = 10000$	0.898	0.898	0.898	0.889	0.978	0.949	0.949	0.949	0.949	0.947

Table 12: m-out-of-n and Numerical Bootstrap Lower Coverage for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-n Bootstrap				$\lambda_n = n^{-1/2}$	Numerical Bootstrap				
β_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.842	0.896	0.896	0.896	0.896	0.914	0.943	0.943	0.943	0.943
$n = 1000$	0.898	0.909	0.909	0.909	0.909	0.936	0.944	0.944	0.944	0.944
$n = 10000$	0.922	0.923	0.923	0.923	0.923	0.946	0.949	0.949	0.949	0.949
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/4}$	Numerical Bootstrap				
$n = 100$	0.715	0.897	0.897	0.897	0.897	0.822	0.947	0.947	0.947	0.947
$n = 1000$	0.807	0.894	0.894	0.894	0.894	0.893	0.945	0.945	0.945	0.945
$n = 10000$	0.857	0.9	0.9	0.9	0.9	0.923	0.949	0.949	0.949	0.949
	m-out-of-n Bootstrap				$\lambda_n = n^{-1/6}$	Numerical Bootstrap				
$n = 100$	0.622	0.92	0.92	0.92	0.92	0.723	0.959	0.959	0.959	0.959
$n = 1000$	0.727	0.9	0.9	0.9	0.9	0.828	0.946	0.946	0.946	0.946
$n = 10000$	0.791	0.899	0.899	0.899	0.899	0.885	0.951	0.951	0.951	0.951

5 Additional Monte Carlo Simulation Results for Modal Estimator

In this section, we report additional Monte Carlo simulation results for the modal estimator discussed in section 5 of the main text. The estimators we consider are the following:

$$\hat{\theta}^*_{\text{standard}} = \underset{\theta}{argmax} \frac{1}{n} \sum_{i=1}^{n} 1\left(\theta - 5 \leq X_i^* \leq \theta + 5\right)$$

$$\hat{\theta}^*_{\text{moutofn}} = \underset{\theta}{argmax} \frac{1}{m} \sum_{i=1}^{m} 1\left(\theta - 5 \leq X_i^* \leq \theta + 5\right)$$

$$\hat{\theta}^*_{\text{subsampling}} = \underset{\theta}{argmax} \frac{1}{b} \sum_{i=1}^{b} 1\left(\theta - 5 \leq X_i^* \leq \theta + 5\right)$$

$$\hat{\theta}^*_{\text{perturb}} = \underset{\theta}{argmax} \frac{1}{n} \sum_{i=1}^{n} 1\left(\theta - 5 \leq X_i^* \leq \theta + 5\right) G_i^*, \qquad G_i^* \sim exp(1)$$

$$\hat{\theta}^*_{\text{numerical}} = \underset{\theta}{argmax} \left\{ \left(1 - \epsilon_n \sqrt{n}\right) \frac{1}{n} \sum_{i=1}^{n} 1\left(\theta - 5 \leq X_i \leq \theta + 5\right) + \epsilon_n \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} 1\left(\theta - 5 \leq X_i^* \leq \theta + 5\right) \right\}$$

The tables below show that for equal-tailed intervals, subsampling performs similarly to the m-out-of-n bootstrap and hence performs worse than the numerical bootstrap. Subsampling and m-out-of-n bootstrap perform better than the numerical bootstrap for upper $\left(-\infty, \hat{\theta}_n - n^{-1/3}\hat{c}^*_{\alpha}\right]$ intervals, but the numerical bootstrap performs better than subsampling and m-out-of-n bootstrap for lower $\left[\hat{\theta}_n - n^{-1/3}\hat{c}^*_{1-\alpha}, \infty\right)$ intervals.

Table 13: Subsampling Two-Sided Equal-Tailed Coverage Frequencies

	$b = n^{2/3}$					$b = n^{1/2}$				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.644	0.647	0.651	0.636	0.649	0.700	0.711	0.721	0.688	0.706
	(1.708)	(1.694)	(1.714)	(1.712)	(1.738)	(1.815)	(1.813)	(1.824)	(1.823)	(1.854)
$n = 500$	0.758	0.759	0.750	0.783	0.768	0.795	0.787	0.779	0.790	0.780
	(1.277)	(1.276)	(1.264)	(1.266)	(1.272)	(1.287)	(1.285)	(1.273)	(1.273)	(1.283)
$n = 1000$	0.803	0.828	0.771	0.782	0.778	0.794	0.825	0.778	0.783	0.779
	(1.047)	(1.040)	(1.037)	(1.035)	(1.029)	(1.047)	(1.038)	(1.039)	(1.035)	(1.036)
$n = 5000$	0.819	0.850	0.843	0.842	0.799	0.794	0.825	0.778	0.783	0.779
	(0.614)	(0.614)	(0.616)	(0.616)	(0.617)	(0.632)	(0.633)	(0.634)	(0.634)	(0.632)
$n = 10000$	0.861	0.836	0.828	0.852	0.845	0.884	0.868	0.849	0.861	0.881
	(0.490)	(0.487)	(0.489)	(0.490)	(0.488)	(0.505)	(0.504)	(0.505)	(0.505)	(0.505)

Table 14: Subsampling Two-Sided Equal-Tailed Coverage Frequencies for $b = n^{1/3}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.720	0.726	0.726	0.700	0.725
	(1.809)	(1.810)	(1.816)	(1.813)	(1.839)
$n = 500$	0.736	0.719	0.717	0.722	0.718
	(1.178)	(1.184)	(1.175)	(1.176)	(1.180)
$n = 1000$	0.716	0.730	0.706	0.695	0.713
	(0.959)	(0.954)	(0.955)	(0.952)	(0.954)
$n = 5000$	0.734	0.780	0.753	0.765	0.743
	(0.592)	(0.593)	(0.592)	(0.593)	(0.591)
$n = 10000$	0.786	0.777	0.772	0.791	0.786
	(0.480)	(0.480)	(0.481)	(0.480)	(0.481)

Table 15: Standard Bootstrap and Perturbation Bootstrap Upper Coverage Frequencies

	Standard Bootstrap					Perturbation Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.613	0.617	0.594	0.635	0.622	0.618	0.619	0.595	0.633	0.626
$n = 500$	0.690	0.711	0.685	0.674	0.695	0.701	0.721	0.705	0.692	0.709
$n = 1000$	0.748	0.681	0.704	0.728	0.693	0.765	0.702	0.727	0.741	0.718
$n = 5000$	0.758	0.757	0.759	0.728	0.775	0.810	0.796	0.801	0.784	0.810
$n = 10000$	0.751	0.753	0.760	0.750	0.765	0.798	0.806	0.820	0.801	0.813

Table 16: m-out-of-n Bootstrap and Numerical Bootstrap Upper Coverage Frequencies for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.824	0.824	0.809	0.833	0.817	0.730	0.723	0.704	0.733	0.716
$n = 500$	0.895	0.926	0.918	0.892	0.914	0.816	0.854	0.834	0.810	0.837
$n = 1000$	0.937	0.910	0.916	0.920	0.910	0.874	0.830	0.839	0.850	0.837
$n = 5000$	0.946	0.921	0.923	0.919	0.938	0.889	0.870	0.870	0.869	0.885
$n = 10000$	0.921	0.920	0.932	0.920	0.919	0.873	0.874	0.884	0.881	0.880

Table 17: Subsampling Upper Coverage Frequencies for $b = n^{2/3}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.798	0.794	0.784	0.808	0.815
$n = 500$	0.868	0.893	0.894	0.902	0.886
$n = 1000$	0.884	0.919	0.908	0.904	0.898
$n = 5000$	0.917	0.889	0.925	0.913	0.910
$n = 10000$	0.911	0.925	0.917	0.910	0.910

Table 18: m-out-of-n Bootstrap and Numerical Bootstrap Upper Coverage Frequencies for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.882	0.882	0.876	0.884	0.879	0.776	0.777	0.753	0.791	0.770
$n = 500$	0.952	0.961	0.953	0.947	0.954	0.880	0.913	0.908	0.879	0.903
$n = 1000$	0.968	0.961	0.965	0.964	0.965	0.936	0.910	0.912	0.918	0.915
$n = 5000$	0.978	0.969	0.971	0.980	0.977	0.951	0.930	0.931	0.931	0.943
$n = 10000$	0.974	0.974	0.975	0.978	0.973	0.934	0.934	0.942	0.931	0.943

Table 19: Subsampling Upper Coverage Frequencies for $b = n^{1/2}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.876	0.874	0.855	0.874	0.881
$n = 500$	0.937	0.961	0.953	0.954	0.949
$n = 1000$	0.955	0.968	0.959	0.962	0.962
$n = 5000$	0.967	0.959	0.973	0.972	0.964
$n = 10000$	0.974	0.963	0.968	0.969	0.979

Table 20: m-out-of-n Bootstrap and Numerical Bootstrap Upper Coverage Frequencies for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.909	0.913	0.898	0.914	0.914	0.791	0.794	0.787	0.805	0.788
$n = 500$	0.971	0.984	0.974	0.977	0.976	0.918	0.935	0.935	0.905	0.925
$n = 1000$	0.984	0.982	0.985	0.985	0.980	0.956	0.936	0.944	0.948	0.942
$n = 5000$	0.993	0.993	0.990	0.992	0.992	0.974	0.959	0.970	0.974	0.975
$n = 10000$	0.991	0.996	0.996	0.995	0.990	0.974	0.971	0.971	0.977	0.972

Table 21: Subsampling Upper Coverage Frequencies for $b = n^{1/3}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.904	0.907	0.899	0.900	0.918
$n = 500$	0.966	0.978	0.977	0.976	0.969
$n = 1000$	0.981	0.986	0.982	0.986	0.985
$n = 5000$	0.992	0.983	0.997	0.994	0.993
$n = 10000$	0.993	0.994	0.988	0.990	0.996

Table 22: Standard Bootstrap and Perturbation Bootstrap Lower Coverage Frequencies

	Standard Bootstrap					Perturbation Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.805	0.803	0.820	0.785	0.794	0.811	0.802	0.829	0.810	0.816
$n = 500$	0.817	0.795	0.798	0.795	0.812	0.840	0.825	0.839	0.836	0.851
$n = 1000$	0.814	0.823	0.790	0.797	0.802	0.843	0.863	0.859	0.868	0.877
$n = 5000$	0.817	0.848	0.817	0.833	0.798	0.877	0.873	0.867	0.890	0.895
$n = 10000$	0.811	0.816	0.802	0.827	0.829	0.878	0.889	0.871	0.881	0.894

Table 23: m-out-of-n Bootstrap and Numerical Bootstrap Lower Coverage Frequencies for $m = n^{2/3}$ and $\epsilon_n = n^{-1/3}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.800	0.799	0.824	0.807	0.812	0.976	0.972	0.985	0.981	0.985
$n = 500$	0.803	0.807	0.814	0.802	0.828	0.969	0.969	0.958	0.962	0.973
$n = 1000$	0.789	0.834	0.817	0.810	0.821	0.941	0.950	0.954	0.960	0.956
$n = 5000$	0.836	0.818	0.832	0.855	0.850	0.930	0.925	0.931	0.930	0.936
$n = 10000$	0.866	0.864	0.856	0.868	0.872	0.935	0.928	0.925	0.926	0.934

Table 24: Subsampling Lower Coverage Frequencies for $b = n^{2/3}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.801	0.804	0.818	0.790	0.792
$n = 500$	0.820	0.803	0.798	0.809	0.811
$n = 1000$	0.822	0.832	0.793	0.797	0.806
$n = 5000$	0.833	0.865	0.852	0.849	0.828
$n = 10000$	0.861	0.851	0.843	0.854	0.861

Table 25: m-out-of-n Bootstrap and Numerical Bootstrap Lower Coverage Frequencies for $m = n^{1/2}$ and $\epsilon_n = n^{-1/4}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.798	0.799	0.820	0.802	0.809	0.982	0.979	0.987	0.985	0.989
$n = 500$	0.747	0.744	0.757	0.749	0.757	0.985	0.982	0.981	0.981	0.989
$n = 1000$	0.732	0.772	0.750	0.754	0.769	0.976	0.974	0.983	0.983	0.979
$n = 5000$	0.793	0.776	0.787	0.813	0.808	0.970	0.968	0.967	0.970	0.971
$n = 10000$	0.835	0.836	0.820	0.824	0.854	0.973	0.963	0.961	0.968	0.967

Table 26: Subsampling Lower Coverage Frequencies for $b = n^{1/2}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.798	0.800	0.817	0.788	0.790
$n = 500$	0.762	0.745	0.741	0.749	0.739
$n = 1000$	0.759	0.770	0.751	0.740	0.749
$n = 5000$	0.788	0.826	0.808	0.812	0.786
$n = 10000$	0.834	0.818	0.818	0.822	0.826

Table 27: m-out-of-n Bootstrap and Numerical Bootstrap Lower Coverage Frequencies for $m = n^{1/3}$ and $\epsilon_n = n^{-1/6}$

	m-out-of-n Bootstrap					Numerical Bootstrap				
θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.740	0.735	0.764	0.731	0.749	0.987	0.982	0.989	0.986	0.991
$n = 500$	0.654	0.650	0.658	0.661	0.662	0.988	0.991	0.986	0.990	0.992
$n = 1000$	0.615	0.658	0.623	0.627	0.660	0.987	0.988	0.990	0.990	0.990
$n = 5000$	0.638	0.646	0.635	0.683	0.662	0.986	0.989	0.985	0.989	0.984
$n = 10000$	0.696	0.695	0.665	0.671	0.715	0.993	0.989	0.983	0.985	0.986

Table 28: Subsampling Lower Coverage Frequencies for $b = n^{1/3}$

θ_0	$-n^{-1/4}$	0	n^{-1}	$n^{-1/2}$	2
$n = 100$	0.745	0.751	0.765	0.741	0.739
$n = 500$	0.677	0.646	0.656	0.647	0.664
$n = 1000$	0.653	0.658	0.613	0.640	0.641
$n = 5000$	0.649	0.673	0.653	0.672	0.655
$n = 10000$	0.688	0.681	0.689	0.687	0.685

6 Sample size dependent statistics

Estimators and test statistics may depend on both P_n and the sample size, e.g. $\theta(P_n, n, n)$. We break the dependence on n into two arguments, the first accounting for the variation of the Gaussian process and the second accounting for a bias consideration. How to handle the sample size dependence requires some knowledge of the model. In this case we write for $\hat{\mathcal{G}}_n = \sqrt{n}(P_n - P)$

$$\hat{\theta}_n = \theta(P_n, n, n) = \theta\left(P + \frac{1}{\sqrt{n}}\hat{\mathcal{G}}_n, \sqrt{n}^2, n\right). \tag{1}$$

To approximate the distribution of $\hat{\mathcal{J}}_n = a(n)(\theta(P_n, n, n) - \theta_0)$, the numerical bootstrap principle again replaces $\hat{\mathcal{G}}_n$ with a suitable bootstrap version $\hat{\mathcal{G}}_n^*$ and the first

$\sqrt{n}$ with $1/\epsilon_n$

$$\hat{\mathcal{J}}_n^* = a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \quad \text{where} \quad \hat{\theta}_n^* = \theta\left(\mathcal{Z}_n^*, \frac{1}{\epsilon_n^2}, n\right) = \theta\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*, \frac{1}{\epsilon_n^2}, n\right).$$

Examples in which only the first n is present in $\theta(P_n, n, n)$ are Chernozhukov and Hong (2003) and Jun et al. (2015). When θ_0 is in the interior of Θ, as in Chernozhukov and Hong (2003), the posterior distribution, in combination with the Sandwich formula, can be used to provide valid inference, or the bootstrap can also be used. In both cases,

$$\hat{\theta}_n \equiv \theta(P_n, n) = \frac{\int \theta\omega(\theta)\exp\left(n^{2\gamma}P_n\pi(\cdot;\theta)\right)d\theta}{\int \omega(\theta)\exp\left(n^{2\gamma}P_n\pi(\cdot;\theta)\right)d\theta}. \tag{2}$$

where $\omega(\cdot)$ is a prior distribution and $\pi(\cdot;\theta)$ and γ are defined as in Theorem 4.1. Recall that $\gamma = \frac{1}{2}$ in Chernozhukov and Hong (2003) but $\gamma = \frac{1}{3}$ in Jun et al. (2015). Both show that for $\mathcal{Z}_0$ given in Theorem 4.1,

$$\hat{\mathcal{J}}_n = n^{\gamma}\left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow \mathcal{J} = \frac{\int h \exp\left(\ell_\infty(h)\right)dh}{\int \exp\left(\ell_\infty(h)\right)dh}$$

where $\ell_\infty(h) = \mathcal{Z}_0(h) - \frac{1}{2}h'Hh$.

Jun et al. (2015) propose in their Theorem 4 to estimate $\mathcal{J}$ with

$$\bar{\mathcal{J}}_n^* = \frac{\int h \exp\left(\hat{\mathcal{Z}}_0(h) - \frac{1}{2}h'\hat{H}h\right)dh}{\int \exp\left(\hat{\mathcal{Z}}_0(h) - \frac{1}{2}h'\hat{H}h\right)dh}$$

where $\hat{H} = H + o_P(1)$ and $\hat{\mathcal{Z}}_0(h)$ is a Gaussian process with covariance kernel $\hat{\Sigma}_\rho(s,t)$ such that $\hat{\Sigma}_\rho(s,t) = \Sigma_\rho(s,t) + o_P(1)$. Alternatively, $\hat{\mathcal{Z}}_0(h)$ can be estimated by $n^{2\gamma-\frac{1}{2}}\hat{\mathcal{G}}_n^*\left(g\left(\cdot,\hat{\theta}+h/n^{\gamma}\right) - g\left(\cdot,\hat{\theta}\right)\right)$. In contrast, we propose a numerical bootstrap estimate:

$$\hat{\theta}_n^* \equiv \theta\left(P_n + \epsilon_n\hat{\mathcal{G}}_n^*, \frac{1}{\epsilon_n^2}\right) = \frac{\int \theta\exp\left(\pi_n^*(\theta)\right)d\theta}{\int \exp\left(\pi_n^*(\theta)\right)d\theta},$$

where $\pi_n^*(\theta) = \frac{1}{\epsilon_n^{4\gamma}}\left(P_n + \epsilon_n\hat{\mathcal{G}}_n^*\right)\pi(\cdot;\theta)$. In the following we first summarize the key heuristic steps for consistency of the numerical bootstrap estimator $\hat{\theta}_n^* \equiv \theta\left(P_n + \epsilon_n\hat{\mathcal{G}}_n^*, \frac{1}{\epsilon_n^2}\right) =$

$\frac{\int \theta \exp(\pi_n^*(\theta))d\theta}{\int \exp(\pi_n^*(\theta))d\theta}$ before presenting the formal theorem.

We want to show that $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) - \epsilon_n^{-2\gamma}\left(\hat{\theta}_n - \theta_0\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$. This will follow from (1) $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n - \theta_0\right) = o_p(1)$ and (2) $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$. Part (1) follows from $\sqrt{n}\epsilon_n \to \infty$. Since $n^\gamma\left(\hat{\theta} - \theta_0\right) = O_p(1)$, $\frac{1}{\epsilon_n^{2\gamma}}\left(\hat{\theta}_n - \theta_0\right) = \frac{1}{\epsilon_n^{2\gamma}} O_p\left(n^{-\gamma}\right) = o_p(1)$. To check (2) write

$$\pi_n^*(\theta) = \frac{\hat{\mathcal{G}}_n^* \pi(\cdot,\theta)}{\epsilon_n^{4\gamma-1}} + \frac{\hat{\mathcal{G}}_n \pi(\cdot,\theta)}{\epsilon_n^{4\gamma}\sqrt{n}} + \frac{P\pi(\cdot,\theta)}{\epsilon_n^{4\gamma}} \approx \frac{1}{\epsilon_n^{4\gamma-1}}\hat{\mathcal{G}}_n^*\pi(\cdot,\theta) + \frac{1}{\epsilon_n^{4\gamma}}P\pi(\cdot,\theta).$$

The second term is dominated by the other two terms since $\frac{1}{\epsilon_n^{4\gamma-1}} >> \frac{1}{\epsilon_n^{4\gamma}\sqrt{n}}$. Since $\pi_n^*(\theta)$ can be recentered at $\pi(\cdot,\theta_0)$ without changing $\hat{\theta}_n^*$, we redefine

$$\pi_n^*(\theta) = \frac{1}{\epsilon_n^{4\gamma-1}}\hat{\mathcal{G}}_n^* g(\cdot,\theta) + \frac{1}{\epsilon_n^{4\gamma}} P g(\cdot,\theta).$$

Next consider $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right)$. Then (ignoring the negligible prior)

$$\hat{h}_n^* \equiv \frac{\int h \exp\left(\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma} h\right)\right) dh}{\int \exp\left(\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma} h\right)\right) dh}, \quad \text{where}$$

$$\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma}h\right) = \frac{1}{\epsilon_n^{4\gamma-1}}\hat{\mathcal{G}}_n^* g\left(\cdot,\theta_0 + \epsilon_n^{2\gamma}h\right) + \frac{1}{\epsilon_n^{4\gamma}} P g\left(\cdot,\theta_0 + \epsilon_n^{2\gamma}h\right) = \epsilon_n^{-4\gamma}\mathcal{Z}_n^*\left(\theta_0 + \epsilon_n^{2\gamma}h\right)$$

By the arguments in Kim and Pollard (1990) and Theorem 4.1, $\hat{\mathcal{G}}_n^*(\cdot,\theta_0+\eta) = O_P^*(\eta^\rho)$, so that

$$\hat{\mathcal{G}}_n^* g\left(\cdot,\theta_0 + \epsilon_n^{2\gamma}h\right) = O_P^*\left(\left(\epsilon_n^{2\gamma}\right)^\rho\right) = O_P^*\left(\epsilon_n^{4\gamma-1}\right), \quad \frac{1}{\epsilon_n^{4\gamma-1}}\hat{\mathcal{G}}_n^* g\left(\cdot,\theta_0 + \epsilon_n^{2\gamma}h\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_0(h).$$

Together these imply that $\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma}h\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \ell_\infty(h) = \mathcal{Z}_0(h) - \frac{1}{2}h'Hh$.

The numerical bootstrap estimates $\mathcal{Z}_0(h) - \frac{1}{2}h'Hh$ using

$$\epsilon_n^{-4\gamma}\mathcal{Z}_n^*(h) = \epsilon_n^{-2\rho\gamma}\hat{\mathcal{G}}_n^*\left(\cdot,\hat{\theta}_n + \epsilon_n^{2\gamma}h\right) + \epsilon_n^{-4\gamma}P_n g\left(\cdot,\hat{\theta} + \epsilon_n^{2\gamma}h\right).$$

Essentially, $\epsilon_n^{-2\rho\gamma}\hat{\mathcal{G}}_n^*\left(\cdot,\hat{\theta}_n + \epsilon_n^{2\gamma}h\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_0(h)$ and $\epsilon_n^{-4\gamma}P_n g\left(\cdot,\hat{\theta}_n + \epsilon_n^{2\gamma}h\right) = -\frac{1}{2}h'Hh + o_P(1)$.

Theorem 6.1 *Let the conditions in Theorem 4.1 hold, with (iv) replaced by condition G(v) in Jun et al. (2015) where* $g_{n0}(\cdot,t) = \epsilon_n^{-1/3} g\left(\cdot, \theta_0 + t\epsilon_n^{2/3}\right)/(1+|t|)$. *Then for any* $\epsilon_n \downarrow 0$, $\hat{\mathcal{J}}_n \rightsquigarrow \mathcal{J}$ *and* $\hat{\mathcal{J}}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = \frac{\int h \exp\left(\epsilon_n^{-4\gamma} \mathcal{Z}_n^*(h)\right) dh}{\int \exp\left(\epsilon_n^{-4\gamma} \mathcal{Z}_n^*(h)\right) dh} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$.

Similar to Theorem 4.2, it is also straightforward to extend Jun et al. (2015) and Theorem 6.1 to allow for the case when the parameter can be at the boundary of a constrained set that is approximated by a cone, in the sense that for any $\alpha_n \to \infty$, and for any compact set $\mathcal{M}$ with radius $M < \infty$,

$$1\left(\delta \in \alpha_n (C - \theta_0) \cap \mathcal{M}\right) \to 1\left(\delta \in T_C(\theta_0) \cap \mathcal{M}\right) \tag{3}$$

almost surely in δ under the Lebesgue measure. We redefine, ignoring the negligible prior $\omega(\cdot)$,

$$\hat{\theta}_n \equiv \theta(P_n, n) = \int_C \theta \hat{p}_n(\theta)\, d\theta \quad \text{where} \quad \hat{p}_n(\theta) = \frac{\exp\left(n^{2\gamma} P_n \pi(\cdot;\theta)\right)}{\int_C \exp\left(n^{2\gamma} P_n \pi(\cdot;\theta)\right) d\theta} \tag{4}$$

and

$$\hat{\theta}_n^* = \int_C \theta \hat{p}_n^*(\theta)\, d\theta \quad \text{where} \quad \hat{p}_n^*(\theta) = \frac{\exp\left(\pi_n^*(\theta)\right)}{\int_C \exp\left(\pi_n^*(\theta)\right) d\theta}. \tag{5}$$

Also let $\hat{\mathcal{J}}_n = n^{\gamma}\left(\hat{\theta}_n - \theta_0\right) = \int_{\sqrt{n}(C-\theta_0)} h \hat{p}_n(h)\, dh$ where

$$\hat{p}_n(h) = \frac{\exp\left(n^{2\gamma} P_n \pi(\cdot;\theta_0 + n^{-\gamma} h)\right)}{\int_{\sqrt{n}(C-\theta_0)} \exp\left(n^{2\gamma} P_n \pi(\cdot;\theta_0 + n^{-\gamma} h)\right) dh} \tag{6}$$

and $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) = \int_{\epsilon_n^{-2\gamma}(C-\theta_0)} h \hat{p}_n^*(h)\, dh$ where

$$\hat{p}_n^*(h) = \frac{\exp\left(\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma} h\right)\right)}{\int_{\epsilon_n^{-2\gamma}(C-\theta_0)} \exp\left(\pi_n^*\left(\theta_0 + \epsilon_n^{2\gamma} h\right)\right) dh}, \tag{7}$$

and

$$\mathcal{J} = \int_{T_C(\theta_0)} h p_\infty(h)\, dh \quad \text{where} \quad p_\infty(h) = \frac{h \exp\left(\ell_\infty(h)\right)}{\int_{T_C(\theta_0)} \exp\left(\ell_\infty(h)\right) dh}. \tag{8}$$

Given these redefinitions, a constrained version of Theorem 6.1 is available.

Theorem 6.2 *Let the conditions in Theorem 6.1 and (3) hold. Then for $\hat{\mathcal{J}}_n$, $\hat{h}_n^*$ and $\mathcal{J}$ defined in (6), (7), and (8), $\hat{\mathcal{J}}_n \rightsquigarrow \mathcal{J}$, $\hat{h}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$, and $\hat{h}_n^* \rightsquigarrow \mathcal{J}$, which implies that for $\hat{\mathcal{J}}_n^* \equiv \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$ and $\hat{\mathcal{J}}_n^* \rightsquigarrow \mathcal{J}$.*

We remark that even in the "regular" case where $\gamma = 1/2$ and when the information matrix equality holds, the posterior quantiles in Chernozhukov and Hong (2003) only provide asymptotically valid confidence intervals when θ_0 is in the interior of Θ, so that $T_C(\theta_0) = R^d$. If θ_0 is on the boundary of Θ, neither posterior quantiles nor the bootstrap are asymptotically valid, but the numerical bootstrap will be.

Analogous to Theorem 4.3, in "regular" $\sqrt{n}$ convergent models (with $\gamma = 1/2$), $\ell_\infty(h)$ can be replaced by its local quadratic approximation. Under assumption 4.1, and more generally (4.6) and (4.7), $\ell_\infty(h)$ is determined by a finite dimensional sufficient statistic: $\ell_\infty(h) = \Delta_0' h - \frac{1}{2}h'Hh$ where consistent estimates $\hat{\Delta}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \Delta_0$ and $\hat{H} = H + o_P(1)$ are available. Then we can also define

$$\hat{\mathcal{J}}_n^* = \int_{\frac{C-\hat{\theta}}{\epsilon_n}} h\hat{p}_n^*(h)\,dh \text{ where } \hat{p}_n^*(h) = \frac{\exp\left(-\frac{1}{2}\left(h - \hat{\mathbb{G}}_n^*\right)'\hat{H}\left(h - \hat{\mathbb{G}}_n^*\right)\right)dh}{\int_{\frac{C-\hat{\theta}}{\epsilon_n}} \exp\left(-\frac{1}{2}\left(h - \hat{\mathbb{G}}_n^*\right)'\hat{H}\left(h - \hat{\mathbb{G}}_n^*\right)\right)dh} \tag{9}$$

and where $\hat{\mathbb{G}}_n^* = \hat{H}^{-1}\hat{\Delta}_n^*$. This can in fact be rewritten as $\hat{\mathcal{J}}_n^* = \epsilon_n^{-1}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, where

$$\hat{\theta}_n^* = \int_C \theta\hat{p}_n^*(\theta)\,d\theta \text{ and } \hat{p}_n^*(\theta) = \frac{\exp\left(\left(\frac{\theta-\hat{\theta}_n}{\epsilon_n} - \hat{\mathbb{G}}_n^*\right)'\hat{H}\left(\frac{\theta-\hat{\theta}_n}{\epsilon_n} - \hat{\mathbb{G}}_n^*\right)\right)}{\int_C \exp\left(\left(\frac{\theta-\hat{\theta}_n}{\epsilon_n} - \hat{\mathbb{G}}_n^*\right)'\hat{H}\left(\frac{\theta-\hat{\theta}_n}{\epsilon_n} - \hat{\mathbb{G}}_n^*\right)\right)d\theta} \tag{10}$$

Theorem 6.3 *Under the conditions of Theorem 4.3 and (3), $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$ and $\hat{\mathcal{J}}_n^* \rightsquigarrow \mathcal{J}$ for $\hat{\mathcal{J}}_n^*$ and $\mathcal{J}$ defined in (9) and (8).*

It is also immediate from the remark after Theorem 4.3 that $\hat{\mathbb{G}}_n^*$ can be replaced by other consistent estimates of $H^{-1}\Delta_0$, for example by taking $\hat{\mathbb{G}}_n^* = \sqrt{n}\left(\bar{\theta}_n^* - \bar{\theta}_n\right)$ where $\bar{\theta}_n$ is an unconstrained estimate and $\bar{\theta}_n^*$ is its bootstrapped version.

An analog for Theorem 4.4 for a direct application of the numerical bootstrap to GMM is also available. Recall from Theorem 4.4 that to approximate the distribution of

$$\hat{\theta}_n = \int_C \theta \hat{p}_n(\theta) \text{ where } \hat{p}_n(\theta) = \frac{\exp\left(n\hat{Q}_n(\theta)\right)}{\int_C \exp\left(n\hat{Q}_n(\theta)\right) d\theta}, \ \hat{Q}_n(\theta) = -\frac{1}{2}\hat{\pi}(\theta)' W \hat{\pi}(\theta), \tag{11}$$

$\pi_n^*(\theta)$ in (4) and (7) can be replaced by $\epsilon_n^{-2}\hat{Q}_n^*(\theta) = -\epsilon_n^{-2}\hat{\pi}^*(\theta)' W \hat{\pi}^*(\theta)$ for $\hat{\pi}^*(\theta) = \mathcal{Z}_n^*\pi(\cdot,\theta)$. Therefore $\hat{\theta}_n^* = \int_C \theta \hat{p}_n^*(\theta) d\theta$, and $\hat{\mathcal{J}}_n^* = \epsilon_n^{-1}\left(\hat{\theta}_n^* - \theta_0\right) = \int_{\epsilon_n^{-1}(C-\theta_0)} h\hat{p}_n^*(h) dh$, where

$$\hat{p}_n^*(\theta) = \frac{\exp\left(\epsilon_n^{-2}\hat{Q}_n^*(\theta)\right)}{\int_C \exp\left(\epsilon_n^{-2}\hat{Q}_n^*(\theta)\right) d\theta} \text{ and } \hat{p}_n^*(h) = \frac{\exp\left(\epsilon_n^{-2}\hat{Q}_n^*(\theta_0 + \epsilon_n h)\right)}{\int_{\frac{C-\theta_0}{\epsilon_n}} \exp\left(\epsilon_n^{-2}\hat{Q}_n^*(\theta_0 + \epsilon_n h)\right) dh}. \tag{12}$$

The following theorem follows directly from combining arguments in the proofs of Theorems 4.4, 6.2, and 6.3 and is thus stated without proof.

Theorem 6.4 *Under the conditions of Theorem 4.3,* $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$.

In addition to using plugin, estimation of (but not inference for except in "regular" cases) functions $\phi_0 = \phi(\theta_0)$ can also be based on functionals of the (quasi) posterior distribution, such as the posterior mean $\hat{\phi} = \int_C \phi(\theta) \hat{p}_n(\theta) d\theta$ or a posterior quantile, $\hat{\phi}_\tau = \inf\left\{x : \int_{C,\phi(\theta)\leq x} \hat{p}_n(\theta) d\theta \geq \tau\right\}$ where $\hat{p}_n(\theta)$ are defined in (4) and (11). The distributions of $\hat{h}_n = n^\gamma\left(\hat{\phi} - \phi_0\right)$ and $\hat{h}_\tau = n^\gamma\left(\hat{\phi}_\tau - \phi_0\right)$ will be approximated by

$$\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\phi}^* - \hat{\phi}\right) \quad \text{where} \quad \hat{\phi}^* = \int_C \phi(\theta) \hat{p}_n^*(\theta) d\theta$$

and

$$\hat{h}_\tau^* = \epsilon_n^{-2\gamma}\left(\hat{\phi}_\tau^* - \hat{\phi}_\tau\right) \quad \text{where} \quad \hat{\phi}_\tau^* = \inf\left\{x : \int_{C,\phi(\theta)\leq x} \hat{p}_n^*(\theta) d\theta \geq \tau\right\}$$

$\hat{p}_n^*(\theta)$ is defined in (5), or (10), or (12).

Theorem 6.5 *Let $\phi(\cdot)$ be continuous, majorized by a polynomial, and directionally differentiable at θ_0 in the sense that there exists a continuous map $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{R}$ such* *that for all* $M < \infty$,

$$\lim_{t_n \downarrow 0} \sup_{|h| \leq M} \left| \frac{\phi(\theta_0 + t_n h) - \phi(\theta_0)}{t_n} - \phi'_{\theta_0}(h) \right| = o(1)$$

Then (1) $\hat{h}_n \rightsquigarrow \mathcal{J}$, *(2)* $\hat{h}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$, *(3)* $\hat{h}_\tau \rightsquigarrow \mathcal{J}_\tau$, *(4)* $\hat{h}_\tau^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}_\tau$, *where*

$$\mathcal{J} = \int_{T_C(\theta_0)} \phi'_{\theta_0}(h) p_\infty(h) \, dh \quad \textit{and} \quad \mathcal{J}_\tau = \inf \left\{ x : \int_{h \in T_C(\theta_0), \phi'_{\theta_0}(h) \leq x} p_\infty(h) \, dh \geq \tau \right\}.$$

It is clear that the above results extend to the case when $\phi(\cdot)$ is a vector, $\phi = (\phi_1, \ldots, \phi_d)$. Consequently, Theorem 6.5 also includes other theorems in this section as special cases when $\phi(\theta) = \theta$.

6.1 Likelihood Ratio Statistics

The distribution of the optimized objective function for M-estimators is of interest for obtaining the distribution the likelihood ratio test statistic or for confidence interval construction for either the parameters or functions of the parameters based on inverting a likelihood ratio test.

In these situations the distributions of the rescaled objective function and of the parameter estimates both play first order roles. Let $Z_n^j(\cdot) \rightsquigarrow Z_\infty^j(\cdot), j = 1, \ldots, J$ for $Z_n^j(\cdot), Z_\infty^j(\cdot) \in \ell_\infty(R^d), \forall j$, and let $\rho_j(Z^j(\cdot))$ be a functional of $Z^j(\cdot)$ that is continuous almost surely on the support of $Z_\infty^j(\cdot)$. Examples are $\rho_j(Z^j(\cdot)) = \arg\max_{h \in H_j} Z^j(h)$ when $Z_\infty^j(h)$ is continuous with a unique maximizer on compact sets almost surely. Other functionals such as those defined through Bayesian posterior locations can replace the argmax functional. By the continuous mapping theorem (CMT), on the product topology of $\ell_\infty(R^d)^j$ and $||\cdot||_J$, for $\mathcal{J} = \{1, \ldots, J\}$,

$$\left(Z_n^j(\cdot), \rho_j\left(Z_n^j(\cdot)\right), j \in \mathcal{J}\right) \rightsquigarrow \left(Z_\infty^j(\cdot), \rho_j\left(Z_\infty^j(\cdot)\right), j \in \mathcal{J}\right). \tag{13}$$

Then for fixed constants $a_j, j = 1, \ldots, J$, by CMT again,

$$\sum_{j=1}^J a_j Z_n^j\left(\rho\left(Z_n^j(\cdot)\right)\right) \rightsquigarrow \sum_{j=1}^J a_j Z_\infty^j\left(\rho\left(Z_\infty^j(\cdot)\right)\right). \tag{14}$$

For example, $a_1 = 1$, $a_2 = -1$, $a_j = 0, \forall j \geq 3$. To apply (13) and (14) to sections 4 and 6 using the notations in Theorem 4.1 for the nested case, we first define

$$Z_n^j(h) = n^{2\gamma} P_n g\left(\cdot; \theta_0 + n^{-2\gamma} h\right) 1\left\{h \in n^\gamma\left(\Theta_j - \theta_0\right)\right\}$$

where Θ_j is the parameter space implied by the jth set of model constraints, and let

$$Z_\infty^j(h) = \left(\mathcal{Z}_\infty(h) - \frac{1}{2} h' H h\right) 1\left\{h \in C_j\right\},$$

where $C_j = \lim_{n\to\infty} n^\gamma\left(\Theta_j - \theta_0\right)$. $\phi_j\left(Z_n^j(\cdot)\right)$ can be defined through (4.1), (4.2), and (6). The numerical bootstrap replaces $Z_n^j(h)$ with

$$Z_{n,j}^*(h) = \epsilon_n^{-4\gamma}\left(P_n + \epsilon_n \mathcal{G}_n^*\right)\left(\pi\left(\cdot, \hat{\theta}_n + \epsilon_n^{2\gamma} h\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right),$$

so that $Z_{nj}^*(\cdot) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} Z_\infty^j(\cdot), j = 1, \ldots, J$. By the bootstrap continuous mapping theorem (theorem 10.8 of Kosorok (2007)),

$$\left(Z_{nj}^*(\cdot), \rho_j\left(Z_{nj}^*(\cdot)\right), j = 1, \ldots, J\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \left(Z_\infty^j(\cdot), \rho_j\left(Z_\infty^j(\cdot)\right), j = 1, \ldots, J\right),$$

and

$$\sum_{j=1}^J a_j Z_n^j\left(\rho\left(Z_n^j(\cdot)\right)\right) \rightsquigarrow \sum_{j=1}^J a_j Z_\infty^j\left(\rho\left(Z_\infty^j(\cdot)\right)\right).$$

Between two strictly nonnested constrained parameter spaces, as long as $\gamma > \frac{1}{4}$, the Likelihood Ratio test statistic converges at a $\sqrt{n}$ rate to a normal limit:

$$\begin{aligned}
&\sqrt{n}\left(P_n \pi\left(\cdot; \hat{\theta}_{n1}\right) - P_n \pi\left(\cdot; \hat{\theta}_{n2}\right)\right) \\
&= \sqrt{n}\left(P_n \pi\left(\cdot; \theta_{10}\right) - P_n \pi\left(\cdot; \theta_{20}\right)\right) + n^{\frac{1}{2} - 2\gamma}\left(Z_n^1\left(\phi_1\left(Z_n^1(\cdot)\right)\right) - Z_n^2\left(\phi_1\left(Z_n^2(\cdot)\right)\right)\right) \\
&= \sqrt{n}\left(P_n \pi\left(\cdot; \theta_{10}\right) - P_n \pi\left(\cdot; \theta_{20}\right)\right) + o_P(1) \\
&\rightsquigarrow N\left(0, Var\left(\pi\left(Z_i; \theta_{10}\right) - \pi\left(Z_i; \theta_{20}\right)\right)\right).
\end{aligned}$$

7 Recentering

In hypothesis testing or in confidence set construction based on test statistic inversion, subsampling does not require recentering (unlike the bootstrap) to achieve consistency and local power (Politis et al. (1999)). However, recentering does improve finite sample power (Chernozhukov and Fernández-Val (2005)). The same insight applies to the numerical bootstrap method, which this section illustrates.

Consider, for example, testing $H_0 : \theta(P) = \theta_*$ vs $H_1 : \theta(P) > \theta_*$. The difference between the centered and noncentered versions of the numerical bootstrap method is analogous to those in subsampling tests. For $\hat{\theta}_n^* \equiv \theta\left(P_n + \epsilon_n \sqrt{n}\left(P_n^* - P_n\right)\right)$, a numerical bootstrap test benchmarks the sample distribution of $a(n)\left(\hat{\theta}_n - \theta_*\right)$ to either (1) the noncentered numerical bootstrap distribution $a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_*\right)$, or (2) the centered numerical bootstrap distribution $a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$. Consistency and power analysis come from studying these three distributions. To illustrate, in the following we assume that (1) $a(n)\left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow \mathcal{J}$; (2) $a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_0\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$; and (3) $a\left(\frac{1}{\epsilon_n^2}\right)/a(n) \to 0$. Under the null of $\theta_0 = \theta_*$, they all have the same limiting distribution. Suppose first that the alternative is fixed, e.g. $\theta_0 = \theta_* + \mu$ for $\mu > 0$. Then the sample distribution diverges to ∞ at the rate of $a(n)$:

$$a(n)\left(\hat{\theta}_n - \theta_*\right) = a(n)\left(\hat{\theta}_n - \theta_0\right) + a(n)\mu \rightsquigarrow \mathcal{J} + a(n)\mu$$

where $X_n \rightsquigarrow Y_n$ is defined as $\rho_{BL_1}(X_n, Y_n) = o(1)$ as in Kosorok (2007). The noncentered numerical boostrap distribution diverges to ∞ at the rate $a\left(\frac{1}{\epsilon_n^2}\right)$, which is slower than $a(n)$, since $a\left(\frac{1}{\epsilon_n^2}\right)/a(n) \to 0$, and

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_*\right) = a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_0\right) + a\left(\frac{1}{\epsilon_n^2}\right)\mu \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J} + a\left(\frac{1}{\epsilon_n^2}\right)\mu,$$

where $X_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} Y_n \iff \rho_{BL_1}(X_n, Y_n) = o_P(1)$. Therefore the noncentered test is consistent. More formally, let $\hat{c}_{1-\alpha} = \inf\left\{x : P_n\left(a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_*\right) \le x\right) \ge 1 - \alpha\right\}$ and let $\mathcal{J}_{1-\alpha} = \inf\{x : P(\mathcal{J} \le x) \ge 1 - \alpha\}$. By arguments in Lemma 10.11 in Kosorok (2007) and the fact that the CDF of $\mathcal{J}$ is strictly increasing on its support,

$\hat{c}_{1-\alpha} - \mathcal{J}_{1-\alpha} - a\left(\frac{1}{\epsilon_n^2}\right)\mu = o_P(1)$. Then by Slutsky,

$$P\left(a(n)\left(\hat{\theta}_n - \theta_*\right) > \hat{c}_{1-\alpha}\right) = P\left(a(n)\left(\hat{\theta}_n - \theta_*\right) + o_P(1) > \mathcal{J}_{1-\alpha} + a\left(\frac{1}{\epsilon_n^2}\right)\mu\right)$$
$$= P\left(\mathcal{J} > \mathcal{J}_{1-\alpha} + \left(a\left(\frac{1}{\epsilon_n^2}\right) - a(n)\right)\mu\right) + o(1) \longrightarrow 1.$$

The noncentered test is consistent but can be less powerful in finite sample than the following recentered version:

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_0\right) - \frac{a\left(1/\epsilon_n^2\right)}{a(n)}a(n)\left(\hat{\theta}_n - \theta_0\right) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J}.$$

If we let $\bar{c}_{1-\alpha} = \inf\left\{x : P_n\left(a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \leq x\right) \geq 1-\alpha\right\}$, then we also have

$$P\left(a(n)\left(\hat{\theta}_n - \theta_*\right) > \bar{c}_{1-\alpha}\right) = P\left(a(n)\left(\hat{\theta}_n - \theta_*\right) + o_P(1) > \mathcal{J}_{1-\alpha}\right)$$
$$= P\left(\mathcal{J} > \mathcal{J}_{1-\alpha} - a(n)\mu\right) + o(1) \longrightarrow 1.$$

Observing that $\hat{c}_{1-\alpha} - \bar{c}_{1-\alpha} = a\left(\frac{1}{\epsilon_n^2}\right)\mu + \frac{a\left(1/\epsilon_n^2\right)}{a(n)}a(n)\left(\hat{\theta}_n - \theta_0\right)$, the power difference derives from

$$P\left(\hat{c}_{1-\alpha} - \bar{c}_{1-\alpha} > a\left(\frac{1}{\epsilon_n^2}\right)\mu - c\right) \to 1 \quad \text{for all} \quad c > 0.$$

Under the local alternative that $\theta_0 = \theta_* + \frac{c}{a(n)}$ for $c > 0$, the sample distribution satisfies

$$a(n)\left(\hat{\theta}_n - \theta_*\right) = a(n)\left(\hat{\theta}_n - \theta_0\right) + c \rightsquigarrow \mathcal{J} + c.$$

As $\frac{a\left(\frac{1}{\epsilon_n^2}\right)}{a(n)} \to 0$, the noncentered numerical bootstrap distribution converges to the null limit:

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_*\right) = a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_0\right) + \frac{a\left(\frac{1}{\epsilon_n^2}\right)}{a(n)}c \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J},$$

and has the correct asymptotic local power. The centered numerical bootstrap has a limit that also does not depend on the local drift c:

$$a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = a\left(\frac{1}{\epsilon_n^2}\right)\left(\hat{\theta}_n^* - \theta_0\right) - \frac{a\left(\frac{1}{\epsilon_n^2}\right)}{a\left(n\right)} a\left(n\right)\left(\hat{\theta}_n - \theta_0\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\leadsto}} \mathcal{J}.$$

The relation between the centered and non-centered numerical bootstrap distributions depends on $\mathcal{J}_n = a\left(n\right)\left(\hat{\theta}_n - \theta_0\right) \leadsto \mathcal{J}$. When $\mathcal{J}$ is a univariate centered normal distribution, the noncentered critical value $\hat{c}_{1-\alpha}$ is more likely than not larger than the centered critical value $\bar{c}_{1-\alpha}$, leading to less rejection. For $b\left(n\right) = \frac{a\left(\frac{1}{\epsilon_n^2}\right)}{a(n)}$ and $\hat{c}_{1-\alpha} - \bar{c}_{1-\alpha} = b\left(n\right)\left(c + \mathcal{J}_n\right)$,

$$P\left(\hat{c}_{1-\alpha} > \bar{c}_{1-\alpha}\right) = P\left(\mathcal{J}_n + c > 0\right) > \frac{1}{2} + \delta \quad \text{for some} \quad \delta > 0 \quad \text{and all large } n.$$

8 Unknown polynomial convergence rate

Similar to subsampling, the numerical bootstrap can be used to estimate the unknown rate of convergence when the convergence rate is a polynomial function $a\left(n\right) = n^\beta$ of the sample size and when the numerical bootstrap consistently estimates the limiting distribution. This is done by comparing the empirical distributions estimated by two (or more) sequences of step sizes ϵ_n. Let $\mathcal{L}_{n,\epsilon_n}\left(x\right)$ denote the distribution of $\theta\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*\right) - \theta\left(P_n\right)$, which is estimated by bootstrap simulations. Then

$$\epsilon_n^{-2\beta} \mathcal{L}_{n,\epsilon_n}^{-1}\left(t\right) = \hat{\mathcal{J}}_{\epsilon_n}^{-1}\left(t\right) = \mathcal{J}^{-1}\left(t, P\right) + o_P\left(1\right).$$

For $t_1 \in (0, 0.5)$, $t_2 \in (0.5, 1)$, $\epsilon_n^{-2\beta}\left(\mathcal{L}_{n,\epsilon_n}^{-1}\left(t_2\right) - \mathcal{L}_{n,\epsilon_n}^{-1}\left(t_1\right)\right) = \mathcal{J}^{-1}\left(t_2, P\right) - \mathcal{J}^{-1}\left(t_1, P\right) + o_P\left(1\right)$, or

$$-2\beta \log \epsilon_n + \log\left(\mathcal{L}_{n,\epsilon_n}^{-1}\left(t_2\right) - \mathcal{L}_{n,\epsilon_n}^{-1}\left(t_1\right)\right) = \mathcal{J}^{-1}\left(t_2, P\right) - \mathcal{J}^{-1}\left(t_1, P\right) + o_P\left(1\right).$$

Using two step size sequences, $\epsilon_{n,1}$ and $\epsilon_{n,2}$, it is then natural to estimate β by

$$\begin{aligned}\hat{\beta}_n &= \frac{\log\left(\mathcal{L}_{n,\epsilon_{n,2}}^{-1}\left(t_2\right) - \mathcal{L}_{n,\epsilon_{n,2}}^{-1}\left(t_1\right)\right) - \log\left(\mathcal{L}_{n,\epsilon_{n,1}}^{-1}\left(t_2\right) - \mathcal{L}_{n,\epsilon_{n,1}}^{-1}\left(t_1\right)\right)}{2\left(\log \epsilon_{n,2} - \log \epsilon_{n,1}\right)} \\ &= \beta + o_P\left(\left(\log \epsilon_{n,2} - \log \epsilon_{n,1}\right)^{-1}\right).\end{aligned}$$

For example, take $\epsilon_{n,1} = n^{-\gamma_1/2}$ and $\epsilon_{n,2} = n^{-\gamma_2/2}$ where $0 < \gamma_2 < \gamma_1 < 1$. Then

$$\hat{\beta}_n = \beta + o_P\left((\log n)^{-1}\right).$$

9 Application to Partially Identified Models

In this section, we relate the numerical bootstrap principle to applications in partially identified models studied by Andrews and Soares (2010),Bugni et al. (2015), and Bugni et al. (2017)). In particular, the numerical bootstrap method generates a special case of the generalized moment selection (GMS) function in Andrews and Soares (2010), which in one case can be adjusted by recentering, and also provides an alternative estimate of the asymptotic distribution developed in Bugni et al. (2015) and the recentering term in the second test statistic in Bugni et al. (2017). Moment selection is also an essential part of Hansen (2005) and of multiple inequality testing (Wolak (1989)).

First we review their setup. For a finite dimensional parameter $\theta \in \Theta$ and a set of moment conditions $g(z,\theta) = \{g_k(z,\theta), k = 1,\ldots,K\}$, denote $g_k(\theta) = Pg_k(\cdot,\theta)$ and $g(\theta) = Pg(\cdot,\theta)$. Andrews and Soares (2010), Bugni et al. (2017) and Bugni et al. (2015) test

$$H_0 : \sup_{\theta\in\bar{\Theta}} \min_{k=1,\ldots,K} g_k(\theta) \geq 0 \quad \text{vs} \quad H_1 : \sup_{\theta\in\bar{\Theta}} \min_{k=1,\ldots,K} g_k(\theta) < 0.$$

On the one hand, in Andrews and Soares (2010), $\bar{\Theta} = \{\theta_0\}$ corresponds to a singleton parameter value for a pointwise testing procedure. On the other hand, in Bugni et al. (2015), $\bar{\Theta} = \Theta$ corresponds to the entire parameter space. Furthermore, in Bugni et al. (2017), $\bar{\Theta} = \Theta(\gamma) = \{\theta \in \Theta : f(\theta) = \gamma\}$ is the preimage of the functional parameter γ of interest.

All three papers employ a nonincreasing and continuous test function $S(\cdot)$ that satisfies, among other conditions: (1) $S(g(\theta)) \geq 0$ for all $g(\theta)$; $S(g(\theta)) = 0$ for all $g(\theta) \geq 0$; (2) $S(\cdot)$ is scale equivalent of degree ρ (e.g. 1 or 2), so that for $c > 0$, $S(cx) = c^\rho S(x)$. Examples are $S(x) = \sum_{k=1}^K x_k^-$, corresponding to $\rho = 1$, and $S(x) = \sum_{k=1}^K \left(x_k^-\right)^2$ corresponding to $\rho = 2$.

The null and alternative hypotheses are then converted using the test function

$S(\cdot)$ to

$$H_0 : T(P) = \inf_{\theta \in \bar{\Theta}} S(Pg(\cdot, \theta)) = 0, \quad \text{against} \quad H_1 : T(P) = \inf_{\theta \in \bar{\Theta}} S(Pg(\cdot, \theta)) > 0,$$

and are tested by benchmarking the sample test statistic

$$\sqrt{n}^{\rho} T(P_n) = \inf_{\theta \in \bar{\Theta}} S\left(\sqrt{n} P_n g(\cdot, \theta)\right)$$

against a consistent estimate of its limiting distribution $\mathcal{J}_{\bar{\Theta}}$ under H_0:

$$\hat{\mathcal{J}}_n \equiv \sqrt{n}^{\rho} \left(T(P_n) - T(P)\right) \overset{H_0}{=} \sqrt{n}^{\rho} T(P_n) \rightsquigarrow \mathcal{J}_{\bar{\Theta}}$$

Bugni et al. (2015) and Bugni et al. (2017) show that $\mathcal{J}_{\bar{\Theta}} = \inf_{\theta \in \bar{\Theta}} S(\ell(\theta) + \mathcal{G}_0 g(\cdot, \theta))$ where $\mathcal{G}_0$ is a Gaussian process with covariance function $Eg(\cdot, \theta) g(\cdot, \theta')$ and $\ell(\theta) = \lim_{t \downarrow 0} Pg(\cdot, \theta)/t$. The numerical bootstrap method approximates $\mathcal{J}_{\bar{\Theta}}$ using

$$\hat{\mathcal{J}}_n^* \equiv \frac{T\left(P_n + \epsilon_n \sqrt{n}\left(P_n^* - P_n\right)\right) - T(P_n)}{\epsilon_n^{\rho}} = T\left(\epsilon_n^{-1} P_n + \hat{\mathcal{G}}_n^*\right) - T\left(\epsilon_n^{-1} P_n\right). \tag{15}$$

In fact, the first term $T\left(\epsilon_n^{-1} P_n + \hat{\mathcal{G}}_n^*\right)$ in (15) corresponds to Type 4 GMS in Andrews and Soares (2010) (their $\phi_4(\cdot)$ GMS function) when $\bar{\Theta} = \{\theta_0\}$ and to the second test statistic (R2) in equation (2.12) of Bugni et al. (2017) when $\bar{\Theta} = \Theta(\gamma)$ for each γ. (15) differs by adding the recentering second term, which as discussed in section 7 usually does not alter asymptotic size and power properties, but might have finite sample implications. In light of this relation, asymptotic validity of the numerical bootstrap follows readily from Andrews and Soares (2010), Bugni et al. (2015), and Bugni et al. (2017). Intuitively, $\hat{\mathcal{J}}_n^* = \underset{\theta \in \bar{\Theta}}{inf} S\left(\frac{1}{\epsilon_n}\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*\right) g(\cdot, \theta)\right) - \underset{\theta \in \bar{\Theta}}{inf} S\left(\frac{1}{\epsilon_n} P_n g(\cdot, \theta)\right)$ is close to $\mathcal{J}_{\bar{\Theta}} = \inf_{\theta \in \bar{\Theta}} S(\ell(\theta) + \mathcal{G}_0 g(\cdot, \theta))$ because the bootstrapped empirical process $\hat{\mathcal{G}}_n^* g(\cdot, \theta) = \sqrt{n}\left(P_n^* - P_n\right) g(\cdot, \theta)$ converges weakly conditional on the data to $\mathcal{G}_0 g(\cdot, \theta)$, $\ell_n(\theta) \overset{p}{\to} \ell(\theta)$, where $\ell_n(\theta) = \frac{1}{\epsilon_n} P_n g(\cdot, \theta)$, and under the null, $\underset{\theta \in \bar{\Theta}}{inf} S(Pg(\cdot, \theta)) = 0$.

Confidence Set construction and Recentering the test statistic Andrews and Soares (2010) and Bugni et al. (2017) construct confidence sets by pointwise inversion of their test statistics, while Bugni et al. (2015) tests for model misspecification.

We remark that these are two different yet related issues. For example, Bugni et al. (2017) suggests a confidence set for γ as $C = \{\gamma : \inf_{\theta\in\Theta(\gamma)} S\left(\sqrt{n}P_n g\left(\cdot;\theta\right)\right) \leq \hat{c}_{1-\alpha}\}$ for $\hat{c}_{1-\alpha}$ being the αth quantile of the first term $T\left(\epsilon_n^{-1}P_n + \hat{\mathcal{G}}_n^*\right)$ in (15). The numerical bootstrap with recentering redefines $\hat{c}_\alpha$ to be the αth quantile of $\hat{\mathcal{J}}_n^*$ in equation (15). Note that the resulting confidence set in both cases can be empty with positive probability when the identified set Θ_0 has zero or small Lebesgue measure or when the model is misspecified.

This suggests another modification in the confidence set construction, where the left hand side statistic is also recentered as $\inf_{\theta\in\Theta(\gamma)} S\left(P_n g\left(\cdot;\theta\right)\right) - \inf_{\theta\in\Theta} S\left(P_n g\left(\cdot;\theta\right)\right)$. If the identified set Θ_0 has positive Lebesgue measure, $P\left(\inf_{\theta\in\Theta} S\left(P_n g\left(\cdot;\theta\right)\right) = 0\right) \to 1$. On the other hand, when Θ_0 has zero Lebesgue measure, for example in the case of point identification, it is possible that $P\left(\inf_{\theta\in\Theta} S\left(P_n g\left(\cdot;\theta\right)\right) > 0\right) = 1$. Recentering the left hand side can result in a confidence set that is non-empty with probability one, even when the model is misspecified, which is probably not desirable if one would like to use the size of the confidence set as a test for model misspecification. However, recentering the left hand side appears to be a widely acceptable practice in empirical research which dates back to point identified models, where model specification testing and confidence set construction are usually done separately (often without accounting for the sequential testing implications).

With the test statistic recentered, $\hat{\mathcal{J}}_n^*$ also can be further recentered in order to avoid a confidence set that is too large. In particular, redefine (for $\bar{\Theta} = \theta_0$ as in Andrews and Soares (2010) or $\bar{\Theta} = \Theta\left(\gamma_0\right)$ as in Bugni et al. (2017)),

$$\bar{T}\left(P\right) = \inf_{\theta\in\bar{\Theta}} S\left(Pg\left(\cdot,\theta\right)\right) - \inf_{\theta\in\Theta} S\left(Pg\left(\cdot,\theta\right)\right)$$
$$\bar{T}\left(P_n\right) = \inf_{\theta\in\bar{\Theta}} S\left(P_n g\left(\cdot,\theta\right)\right) - \inf_{\theta\in\Theta} S\left(P_n g\left(\cdot,\theta\right)\right)$$

so that $\bar{\mathcal{J}}_n = \hat{\mathcal{J}}_n - \hat{\mathcal{J}}_{n,\Theta}$ with $\hat{\mathcal{J}}_{n,\Theta} = \inf_{\theta\in\Theta} S\left(\sqrt{n}P_n g\left(\cdot,\theta\right)\right) - \inf_{\theta\in\Theta} S\left(\sqrt{n}Pg\left(\cdot,\theta\right)\right)$. The estimated distribution in (15) is defined analogously by replacing $T\left(\cdot\right)$ with $\bar{T}\left(\cdot\right)$:

$$\bar{\mathcal{J}}_n^* \equiv \hat{\mathcal{J}}_n^* - \hat{\mathcal{J}}_{n,\Theta}^* \quad \text{where} \quad \hat{\mathcal{J}}_{n,\Theta}^* = \underset{\theta\in\Theta}{inf} S\left(\frac{1}{\epsilon_n}\mathcal{Z}_n^* g(\cdot,\theta)\right) - \underset{\theta\in\Theta}{inf} S\left(\frac{1}{\epsilon_n}P_n g(\cdot,\theta)\right). \quad (16)$$

Theorem 9.1 below shows that under the null, both $\bar{\mathcal{J}}_n \equiv \sqrt{n}^\rho\left(\bar{T}\left(P_n\right) - \bar{T}\left(P\right)\right) \rightsquigarrow \bar{\mathcal{J}}$

and $\bar{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \bar{\mathcal{J}}$, where

$$\bar{\mathcal{J}} = \inf_{\theta \in \bar{\Theta}} S\left(\ell\left(\theta\right) + \mathcal{G}_0 g\left(\cdot, \theta\right)\right) - \inf_{\theta \in \Theta} S\left(\ell\left(\theta\right) + \mathcal{G}_0 g\left(\cdot, \theta\right)\right) \tag{17}$$

To illustrate the difference between $\hat{\mathcal{J}}_n{}^*$ and $\bar{\mathcal{J}}_n^*$, consider a conventional point identified, correctly specified, and possibly overidentified GMM model. On the one hand, $n\hat{g}\left(\theta_0\right)' W \hat{g}\left(\theta_0\right) \rightsquigarrow \chi^2_{\dim(g)}$ with the optimal weighting matrix W. On the other hand, it is standard to show that $n\hat{g}\left(\theta_0\right)' W \hat{g}\left(\theta_0\right) - n\hat{g}\left(\hat{\theta}_n\right)' W \hat{g}\left(\hat{\theta}_n\right) \rightsquigarrow \chi^2_{\dim(\theta)}$, where $\hat{\theta}_n = \arg\inf_{\theta \in \Theta} \hat{g}\left(\theta\right)' W \hat{g}\left(\theta\right)$. The first confidence set without left recentering is given by $C_1 = \{\theta : n\hat{g}\left(\theta\right)' W \hat{g}\left(\theta\right) \leq \chi^2_{\dim(g), 1-\alpha}\}$, while the second confidence set with left recentering is given by

$$C_2 = \left\{\theta : n\hat{g}\left(\theta\right)' W \hat{g}\left(\theta\right) \leq \chi^2_{\dim(\theta), 1-\alpha} + n\hat{g}\left(\hat{\theta}_n\right)' W \hat{g}\left(\hat{\theta}_n\right)\right\}.$$

Since $n\hat{g}\left(\hat{\theta}_n\right)' W \hat{g}\left(\hat{\theta}_n\right) \rightsquigarrow \chi^2_{\dim(g) - \dim(\theta)}$, the right hand sides in both C_1 and C_2 have unconditionally a $\chi^2_{\dim(g)}$ limiting distribution. However, they can behave very differently in every sample realization. In particular, C_2 is always nonempty even under misspecification.

When confidence sets are constructed by inverting a likelihood ratio or distance function test statistic and $\hat{\theta}$ is the maximum likelihood estimator, it is necessary to recenter the sample log likelihood at its optimum since the population objective function (the population entropy) is unknown even when the model is correctly specified. For GMM, the population objective function is zero under correct model specification, which makes recentering the sample objective function optional.

The following theorem demonstrates consistency of the numerical bootstrap for the statistic $\hat{\mathcal{J}}_n^*$ in (15) and the recentered statistic $\bar{\mathcal{J}}_n^*$ in (16). The first part of the theorem is a simplified version of Theorem 3.1 in Bugni et al. (2015).

Theorem 9.1 *Suppose the following assumptions are satisfied:*

(1) $\{g(\cdot, \theta) : \theta \in \Theta\}$ *is a measurable class of functions over a compact set* $\Theta \subseteq \mathbb{R}^{dim(\theta)}$.

(2) The empirical process $\hat{\mathcal{G}}_n g(\cdot, \theta) = \sqrt{n}(P_n - P) g(\cdot, \theta)$ *is stochastically equicontin-*

uous over Θ*: for any* $\epsilon > 0$*,*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P \left(\sup_{\|\theta - \theta'\| < \delta} \|\hat{\mathcal{G}}_n g(\cdot, \theta) - \hat{\mathcal{G}}_n g(\cdot, \theta')\| > \epsilon \right) = 0$$

(3) The bootstrapped empirical process $\hat{\mathcal{G}}_n^* g(\cdot, \theta) = \sqrt{n}(P_n^* - P_n) g(\cdot, \theta)$ *is stochastically equicontinuous conditional on the data: for any* $\epsilon > 0$*,*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P \left(\sup_{\|\theta - \theta'\| < \delta} \|\hat{\mathcal{G}}_n^* g(\cdot, \theta) - \hat{\mathcal{G}}_n^* g(\cdot, \theta')\| > \epsilon | \mathcal{X}_n \right) = 0$$

(4) $\sqrt{n}\epsilon_n \to \infty$ *and* $\epsilon_n \downarrow 0$*.*

Then under the null hypothesis of $T(P) = \underset{\theta \in \bar{\Theta}}{inf} S(Pg(\cdot, \theta)) = 0$*, for* $\mathcal{J}_{\bar{\Theta}} = \inf_{\theta \in \bar{\Theta}} S\left(\ell\left(\theta\right) + \mathcal{G}_0 g\left(\cdot, \theta\right)\right)$*,*

$$\hat{\mathcal{J}}_n \rightsquigarrow \mathcal{J}_{\bar{\Theta}} \text{ and } \hat{\mathcal{J}}_n{}^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}_{\bar{\Theta}}$$

where $\ell\left(\theta\right) = \lim_{t \searrow 0} Pg(\cdot, \theta)/t$*, and* $\mathcal{G}_0$ *is a tight mean zero Gaussian process with covariance kernel* $\Sigma(\theta, \theta') = Pg(\cdot, \theta) g(\cdot, \theta')$*. Likewise,* $\bar{\mathcal{J}}_n \rightsquigarrow \mathcal{J}_{\bar{\Theta}} - \mathcal{J}_{\Theta}$ *and* $\bar{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}_{\bar{\Theta}} - \mathcal{J}_{\Theta}$*, where* $\mathcal{J}_{\Theta} = \inf_{\theta \in \Theta} S\left(\ell\left(\theta\right) + \mathcal{G}_0 g\left(\cdot, \theta\right)\right)$*.*

Testing after estimation Another common empirical situation is when researchers already have enough moment conditions to point identify and estimate the parameter θ, but they want to test whether a second set of moment inequalities (denoted $g(\cdot)$ and different from the first set) are valid at the parameter point identified by the previous moment conditions. The insight of Andrews and Soares (2010) can be generalized to cover this case. Let $\hat{\theta}_n$ be the point estimate obtained from the first set of moment conditions, and $\hat{\theta}_n^*$ be its bootstrapped version. Also let $\hat{g}\left(\theta\right) = P_n g\left(\cdot; \theta\right)$ be the sample moment condition and let $\hat{g}^*\left(\theta\right) = P_n^* g\left(\cdot; \theta\right)$ be its bootstrap analog. To test $H_0 : T\left(P\right) = S\left(g\left(\theta\right)\right) = 0$ against $H_1 : T\left(P\right) = S\left(g\left(\theta\right)\right) > 0$, use the statistic $\sqrt{n}^{\rho} T\left(P_n\right) \equiv S\left(\sqrt{n}\hat{g}\left(\hat{\theta}_n\right)\right)$. The limiting distribution $\mathcal{J}$ of $\sqrt{n}^{\rho}\left(T\left(P_n\right) - T\left(P\right)\right) \equiv S\left(\sqrt{n}\hat{g}\left(\hat{\theta}_n\right)\right) - S\left(\sqrt{n} g\left(\theta\right)\right)$ can then be estimated consistently using the numerical bootstrap as

$$\hat{\mathcal{J}}_n^* = S\left(\frac{\hat{g}\left(\hat{\theta}_n\right)}{\epsilon_n} + \sqrt{n}\left(\hat{g}_n^*\left(\hat{\theta}_n^*\right) - \hat{g}\left(\hat{\theta}_n\right)\right)\right) - S\left(\frac{\hat{g}\left(\hat{\theta}_n\right)}{\epsilon_n}\right).$$

$\sqrt{n}\left(\hat{g}_n^*\left(\hat{\theta}_n^*\right) - \hat{g}\left(\hat{\theta}_n\right)\right)$ can be replaced by another consistent estimate $\hat{Z}$ of the limiting distribution of $\sqrt{n}\left(\hat{g}\left(\hat{\theta}_n\right) - g\left(\theta_0\right)\right)$. For example, if $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are difficult to compute, but are asymptotically linear (meaning $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi\left(z_i\right) + o_p\left(1\right)$ where the influence function can be uniformly consistently estimated by $\hat{\phi}\left(z_i\right)$), then we can use $\mathbb{Z}_n^* \sim N\left(0, \hat{\Omega}\right)$, where

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \left(g\left(Z_i, \hat{\theta}_n\right) + \frac{\partial \hat{g}\left(\hat{\theta}_n\right)}{\partial \theta}\hat{\phi}\left(Z_i\right)\right)\left(g\left(Z_i, \hat{\theta}_n\right) + \frac{\partial \hat{g}\left(\hat{\theta}_n\right)}{\partial \theta}\hat{\phi}\left(Z_i\right)\right)'.$$

and replace $\hat{\mathcal{J}}_n^*$ with $\hat{\mathcal{J}}_n^* = S\left(\frac{\hat{g}\left(\hat{\theta}_n\right)}{\epsilon_n} + \mathbb{Z}_n^*\right) - S\left(\frac{\hat{g}\left(\hat{\theta}_n\right)}{\epsilon_n}\right)$.

10 Proofs for the Theorems

Proof for Theorem 4.1 Our first step is to show that $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are respectively n^γ-consistent and $\epsilon_n^{-2\gamma}$-consistent for θ_0 . Assumptions (iv) and (viii) imply that the conditions of Lemma 4.1 of Kim and Pollard (1990) are satisfied. Therefore, for each $\eta > 0$, there exist random variables $\{M_n\} = O_p(1)$ such that $|P_n g(\cdot, \theta) - P g(\cdot, \theta)| \leq \eta|\theta - \theta_0|^2 + n^{-2\gamma}M_n^2$.[1] Here, $R_0 > 0$ is the constant such that $PG_R^2 = O(R^{2\rho})$ for all $R \leq R_0$. Assumptions (i) and (ii) imply that the conditions of Corollary 4.2 of Kim and Pollard (1990) are satisfied, which, in combination with Lemma 4.1, imply that $n^\gamma\left(\hat{\theta}_n - \theta_0\right) = O_p(1)$.

Next by Lemma 10.1, which generalizes Lemma 4.1 in Kim and Pollard (1990) to a bootstrap version with step size ϵ_n, $\exists M_n^* = O_P^*\left(1\right)$ such that $\epsilon_n|\hat{\mathcal{G}}_n^* g\left(\cdot, \hat{\theta}_n^*\right)| \leq$

[1] The main revisions to Lemma 4.1 of Kim and Pollard (1990) are redefining $A\left(n, j\right) = \left(j-1\right)n^{-\gamma} \leq |\theta| \leq jn^{-\gamma}$, bounding the jth summand in $P\left(M_n > m\right)$ by $n^{4\gamma} P \sup_{|\theta| < jn^{-\gamma}} |P_n g\left(\cdot, \theta\right) - P g\left(\cdot, \theta\right)|^2 \big/ \left[\eta\left(j-1\right)^2 + m^2\right]^2$, where the numerator is further bounded by $n^{4\gamma}\left(n^{-1}C' j n^{-\gamma(2\rho)}\right) = C' j$.

$\eta|\hat{\theta}_n^* - \theta_0|^2 + \epsilon_n^{4\gamma} M_n^{*2}$. Combine this with Kim and Pollard (1990) Lemma 4.1, and note that since $O_P^*\left(\epsilon_n^{4\gamma}\right) + O_P\left(n^{-2\gamma}\right) = O_P^*\left(\epsilon_n^{4\gamma}\right)$, we obtain $|\mathcal{Z}_n^* g(\cdot, \hat{\theta}_n^*) - Pg(\cdot, \hat{\theta}_n^*)| \leq \eta|\hat{\theta}_n^* - \theta_0|^2 + O_P^*\left(\epsilon_n^{4\gamma}\right)$. Then choose η so that $Pg\left(\cdot, \theta\right) - Pg\left(\cdot, \theta_0\right) \leq -2\eta|\theta - \theta_0|^2$, and since $g\left(\cdot, \theta_0\right) = 0$,

$$-O_P^*\left(\epsilon_n^{4\gamma}\right) = \mathcal{Z}_n^* g\left(\cdot, \theta_0\right) - O_P^*\left(\epsilon_n^{4\gamma}\right) \leq \mathcal{Z}_n^* g\left(\cdot, \hat{\theta}_n^*\right) \leq -\eta|\hat{\theta}_n^* - \theta_0|^2 + O_P^*\left(\epsilon_n^{4\gamma}\right),$$

from which we conclude that $|\hat{\theta}_n^* - \theta_0| = O_P^*\left(\epsilon_n^{2\gamma}\right)$ (and hence also $O_P\left(\epsilon_n^{2\gamma}\right)$).

Since $\hat{\theta}_n$ converges at rate n^γ, we can write $n^\gamma(\hat{\theta}_n - \theta_0) = \arg\max_h n^{\gamma\rho}\sqrt{n} P_n g(\cdot, \theta_0 + n^{-\gamma}h)$, and will derive its limiting distribution by verifying the Lindeberg Central Limit Theorem, stochastic equicontinuity and applying the Argmax Continuous Mapping Theorem. We first show that $W_n(h) \equiv n^{\gamma\rho}\sqrt{n}\left(P_n - P\right) g(\cdot; \theta_0 + n^{-\gamma}h)$ converges in finite dimensional distribution to a Gaussian process $\mathcal{Z}_0(h)$ and that $n^{\gamma\rho}\sqrt{n} Pg(\cdot; \theta_0 + n^{-\gamma}h) \to -\frac{1}{2}h'Hh$ for each h as $n \to \infty$. We then show that W_n is stochastically equicontinuous. Finally, we argue using the Argmax Continuous Mapping Theorem (Theorem 2.7 Kim and Pollard (1990)) that $n^\gamma(\hat{\theta}_n - \theta_0) \rightsquigarrow \arg\max_h \mathcal{Z}_0(h) - \frac{1}{2}h'Hh$. Consider the first part. Assumption (vii) implies the Lindeberg condition is satisfied and that $n^{\gamma\rho}\sqrt{n}\left(P_n - P\right) g(\cdot; \theta_0 + n^{-\gamma}h)$ converges in finite dimensional distribution to a mean zero Gaussian process with covariance kernel:

$$\begin{aligned}\Sigma_\rho(s,t) =& \lim_{n\to\infty} Cov\left(n^{\gamma\rho}\sqrt{n}\left(P_n - P\right) g(\cdot; \theta_0 + n^{-\gamma}s), n^{\gamma\rho}\sqrt{n}\left(P_n - P\right) g(\cdot; \theta_0 + n^{-\gamma}t)\right) \\ =& \lim_{n\to\infty} n^{2\gamma\rho} Pg(\cdot; \theta_0 + n^{-\gamma}s) g(\cdot; \theta_0 + n^{-\gamma}t) - n^{2\gamma\rho} Pg(\cdot; \theta_0 + n^{-\gamma}s) Pg(\cdot; \theta_0 + n^{-\gamma}t).\end{aligned}$$

Taking a second order Taylor expansion of $n^{\gamma\rho}\sqrt{n} Pg(\cdot; \theta_0 + n^{-\gamma}h)$ around θ_0 and using $g(\cdot; \theta_0) = 0$ and $\frac{\partial}{\partial\theta} Pg(\cdot; \theta_0) = 0$, $n^{\gamma\rho}\sqrt{n} Pg(\cdot; \theta_0 + n^{-\gamma}h) = -n^{\gamma\rho + \frac{1}{2}}\frac{1}{2}n^{-2\gamma}h'Hh + o(||h||^2) \to -\frac{1}{2}h'Hh$. To show that W_n is stochastically equicontinuous, it suffices to show that for every sequence of positive numbers $\{\delta_n\}$ converging to zero,

$$n^{\gamma\rho}\sqrt{n} E \sup_{\mathscr{D}(n)} |P_n d - Pd| = o(1) \tag{18}$$

where $\mathscr{D}(n) = \{d(\cdot, \theta_0, h_1, h_2) = g(\cdot; \theta_0 + n^{-\gamma}h_1) - g(\cdot; \theta_0 + n^{-\gamma}h_2)$ such that $max(|h_1|, |h_2|) \leq M$ and $|h_1 - h_2| \leq \delta_n\}$. Note that $\mathscr{D}(n)$ has envelope function $D_n = 2G_{R(n)}$ where $R(n) = Mn^{-\gamma}$.

Using the Maximal Inequality in section 3.1 of Kim and Pollard (1990), for sufficiently large n, splitting up the expectation according to whether $n^{2\gamma\rho}P_nD_n^2 \le \eta$ for each $\eta > 0$, and applying the Cauchy-Schwarz inequality,

$$n^{\gamma\rho}\sqrt{n}E\sup_{\mathscr{D}(n)}|P_nd - Pd| \le E\sqrt{n^{2\gamma\rho}P_nD_n^2}J\left(\frac{n^{2\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2}{n^{2\gamma\rho}P_nD_n^2}\right)$$
$$\le \sqrt{\eta}J(1) + \sqrt{En^{2\gamma\rho}P_nD_n^2}\sqrt{EJ^2\left(\min\left(1, \frac{1}{\eta}n^{2\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2\right)\right)}.$$

To show that this is $o(1)$ for each fixed $\eta > 0$, first, note that by assumption (viii), $En^{2\gamma\rho}P_nD_n^2 = 4n^{2\gamma\rho}EG_{R(n)}^2 = O(n^{2\gamma\rho}R(n)^{2\rho}) = O(1)$ since $R(n) = Mn^{-\gamma}$. The proof will then be complete if $n^{2\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2 = o(1)$. Next, for each $K > 0$ write $E\sup_{\mathscr{D}(n)}P_nd^2 \le EP_n\sup_{\mathscr{D}(n)}d^2\{D_n > K\} + KE\sup_{\mathscr{D}(n)}P_n|d| \le EP_nD_n^2\{D_n > K\} + K\sup_{\mathscr{D}(n)}P|d| + KE\sup_{\mathscr{D}(n)}|P_n|d| - P|d||$. By assumption (x), $EP_nD_n^2\{D_n > K\} < \eta n^{-2\gamma\rho}$ for large enough K. By assumption (xi) and the definition of $\mathscr{D}(n)$, $K\sup_{\mathscr{D}(n)}P|d| = O(n^{-2\gamma\rho}\delta_n) = o(n^{-2\gamma\rho})$. By assumption (viii) and the maximal inequality, $KE\sup_{\mathscr{D}(n)}|P_n|d| - P|d|| < Kn^{-\frac{1}{2}}J(1)\sqrt{PD_n^2} = O(n^{-(\gamma\rho+\frac{1}{2})}) = O(n^{-2\gamma}) = o(n^{-2\gamma\rho})$ whenever $\rho < 1$. Therefore, $En^{2\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2 = o(1)$ whenever $\rho < 1$. If $\rho = 1$, typically, $g(\cdot\,\theta)$ is Lipschitz in θ so that $D_n = O_P(n^{-\gamma}\delta_n)$, in which case one can argue directly that $n^{2\gamma}\sup_{\mathscr{D}(n)}P_nd^2 = o_P(1)$.

We have therefore shown that $n^{\gamma\rho}\sqrt{n}P_ng(\cdot, \theta_0 + n^{-\gamma}h) \rightsquigarrow \mathcal{Z}_0(h) - \frac{1}{2}h'Hh$. It follows from the Argmax Continuous Mapping Theorem and $n^{\gamma}\left(\hat{\theta}_n - \theta_0\right) = O_p(1)$ that

$$n^{\gamma}(\hat{\theta}_n - \theta_0) = \arg\max_h n^{\gamma\rho}\sqrt{n}P_ng(\cdot, \theta_0 + n^{-\gamma}h) \rightsquigarrow \mathbb{H} = \arg\max_h \mathcal{Z}_0(h) - \frac{1}{2}h'Hh \quad (19)$$

Similarly, since $|\hat{\theta}_n^* - \theta_0| = O_P^*(\epsilon_n^{2\gamma})$, we can write

$$\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) = \arg\max_h \epsilon_n^{-4\gamma}Z_n^*g(\cdot, \theta_0 + \epsilon_n^{2\gamma}h).$$

The goal is to show that $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{H}$ in (19). Note that

$$\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) - \epsilon_n^{-2\gamma}\left(\hat{\theta}_n - \theta_0\right).$$

By Assumption (ix) and n^γ consistency of $\hat{\theta}_n$: $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n-\theta_0\right)=\frac{1}{(\sqrt{n}\epsilon_n)^{2\gamma}}n^\gamma\left(\hat{\theta}_n-\theta_0\right)=o_P(1)$. It therefore suffices to show that $\mathbb{H}_n=\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\theta_0\right)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathbb{H}$, since $\mathbb{H}_n+o_P(1)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathbb{H}$ whenever $\mathbb{H}_n\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathbb{H}$.

To analyze $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^*-\theta_0\right)$, we use Lemma 10.2 which extends the Arg Max Theorem (Theorem 3.2.2 in van der Vaart and Wellner (1996)) to a bootstrap version. It therefore suffices to show that (i)

$$W_n^*(h)\equiv\epsilon_n^{-(1+2\gamma\rho)}\left(Z_n^*-P\right)g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{Z}_0\left(h\right)\quad\text{in}\quad\ell_\infty\left(K\right)\tag{20}$$

for any compact K and that (ii) $\epsilon_n^{-(1+2\gamma\rho)}Pg(\cdot;\theta_0+\epsilon_n^{2\gamma}h)\to-\frac{1}{2}h'Hh$ for each h.

We show (ii) by a second order Taylor expansion around θ_0 using $g(\cdot;\theta_0)=0$ and $\frac{\partial}{\partial\theta}Pg(\cdot;\theta_0)=0$.

$$\epsilon_n^{-(1+2\gamma\rho)}Pg(\cdot;\theta_0+\epsilon_n^{2\gamma}h)=-\,\epsilon_n^{-(1+2\gamma\rho)}\frac{1}{2}\epsilon_n^{4\gamma}h'Hh+o(||h||^2)\to-\frac{1}{2}h'Hh.$$

To show (20), we first show its unconditional version $W_n^*\left(h\right)\rightsquigarrow\mathcal{Z}_0\left(h\right)$, and then use an almost sure conditional finite dimensional CLT to convert it to $W_n^*\left(h\right)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{Z}_0\left(h\right)$ using arguments analogous to Theorem 2.9.6 in van der Vaart and Wellner (1996). Unconditional convergence $W_n^*\left(h\right)\rightsquigarrow\mathcal{Z}_0\left(h\right)$ is in turn shown by invoking the Lindeberg finite dimensional CLT and verifying stochastic equicontinuity. Note that

$$\begin{aligned}W_n^*(h)=&\epsilon_n^{-(1+2\gamma\rho)}\left(P_n-P\right)g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)+\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_n^*g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)\\=&\frac{1}{\sqrt{n}\epsilon_n}\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)+\hat{\mathcal{G}}_n^*\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h).\end{aligned}$$

For the first part, by assumption (vi), the covariance kernel of $\hat{\mathcal{G}}_ng(\cdot;\theta_0+\epsilon_n^{2\gamma}h)$ converges to the limit

$$\begin{aligned}\Sigma_\rho(s,t)=&\lim_{n\to\infty}Cov\left(\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_ng(\cdot;\theta_0+\epsilon_n^{2\gamma}s),\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_ng(\cdot;\theta_0+\epsilon_n^{2\gamma}t)\right)\\=&\lim_{n\to\infty}\epsilon_n^{-4\gamma\rho}Pg(\cdot;\theta_0+\epsilon_n^{2\gamma}s)g(\cdot;\theta_0+\epsilon_n^{2\gamma}t)-\epsilon_n^{-4\gamma\rho}Pg(\cdot;\theta_0+\epsilon_n^{2\gamma}s)Pg(\cdot;\theta_0+\epsilon_n^{2\gamma}t)\end{aligned}$$

The Lindeberg condition also holds for $\hat{\mathcal{G}}_n g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)$ by assumptions (vii) and (ix):

$$\begin{aligned}&\lim_{n\to\infty}\epsilon_n^{-4\gamma\rho}Pg\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)^2 1\left(\frac{1}{\sqrt{n}}\epsilon_n^{2\gamma\rho}g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\geq\epsilon\Sigma_\rho\left(h,h\right)\right)\\&=\lim_{n\to\infty}\epsilon_n^{-4\gamma\rho}Pg\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)^2 1\left(\epsilon_n^{1-2\gamma\rho}g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\geq\epsilon\Sigma_\rho\left(h,h\right)\sqrt{n}\epsilon_n\right)\to 0.\end{aligned}$$

Therefore by (xi), $\frac{1}{\sqrt{n}\epsilon_n}\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)=o_P\left(1\right)$. Alternatively, we can also impose a strong condition, such as $\sqrt{n}\epsilon_n/\log n\to\infty$, and invoke a version of the law of iterated logarithm, so that we can replace $o_P\left(1\right)$ with $o_{a.s.}\left(1\right)$. Using similar arguments as those after (18), it can also be shown that for $\mathscr{D}(n)=\{d(\cdot,\theta_0,h_1,h_2)=g(\cdot;\theta_0+\epsilon^{2\gamma}h_1)-g(\cdot;\theta_0+\epsilon_n^{2\gamma}h_2)$ such that $max(|h_1|,|h_2|)\leq M$ and $|h_1-h_2|\leq\delta_n\}$, with $D(n)=M\epsilon_n^{2\gamma}$ being an envelope of $\mathscr{D}(n)$,

$$\epsilon_n^{-2\gamma\rho}E\sup_{\mathscr{D}(n)}\left|\hat{\mathcal{G}}_n d\right|=o(1). \tag{21}$$

To see (21), as before bound, with $E\epsilon_n^{-4\gamma\rho}P_nD_n^2=O\left(1\right)$, and split according to η,

$$\begin{aligned}E\epsilon_n^{-2\gamma\rho}E\sup_{\mathscr{D}(n)}\left|\hat{\mathcal{G}}_n d\right|\leq&E\sqrt{\epsilon_n^{-4\gamma\rho}P_nD_n^2}J\left(\frac{\epsilon_n^{-4\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2}{\epsilon_n^{-4\gamma\rho}P_nD_n^2}\right)\\&\leq\sqrt{\eta}J\left(1\right)+\sqrt{E\epsilon_n^{-4\gamma\rho}P_nD_n^2}\sqrt{EJ^2\left(min\left(1,\frac{1}{\eta}\epsilon_n^{-4\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2\right)\right)}\end{aligned}$$

Finally to show that $E\epsilon_n^{-4\gamma\rho}\sup_{\mathscr{D}(n)}P_nd^2=o\left(1\right)$, split using large K,

$$E\sup_{\mathscr{D}(n)}P_nd^2\leq EP_nD_n^2 1\left(D_n>K\right)+K\sup_{\mathscr{D}(n)}P|d|+KE\sup_{\mathscr{D}(n)}\left|P_n|d|-P|d|\right|$$

By (x), $EP_nD_n^2 1\left(D_n>K\right)<\eta\epsilon_n^{4\gamma\rho}$. By (xi), $K\sup_{\mathscr{D}(n)}P|d|=O\left(\epsilon_n^{4\gamma\rho}\delta_n^\rho\right)=o\left(\epsilon_n^{4\gamma\rho}\right)$. By (viii) and the maximal inequality in Kim and Pollard (1990) result 3.1, $KE\sup_{\mathscr{D}(n)}\left|P_n|d|-P|d|\right|\leq Kn^{-1/2}J\left(1\right)\sqrt{ED_n^2}=O\left(n^{-1/2}\epsilon_n^{2\gamma\rho}\right)=O\left(\frac{1}{\sqrt{n}\epsilon_n}\epsilon_n^{1+2\gamma\rho}\right)=o\left(\epsilon_n^{4\gamma\rho}\right)$.

Hence we verified (21) to conclude that $\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)\rightsquigarrow\mathcal{Z}_0\left(h\right)$, and $\frac{1}{\sqrt{n}\epsilon_n}\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)=o_P\left(1\right)$, both as a process indexed by h in $\ell_\infty\left(K\right)$ for any compact K.

We remark that while condition (vii) is modeled after (iv) in Lemma 4.5 of Kim and Pollard (1990), neither seems to be needed for the Lindeberg condition. In fact they should all be implied by (vi). Under the integrability condition (vi), for any $\kappa_n \to \infty$,

$$\lim_{\alpha\to\infty} \alpha^{2\rho} P g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)^2 1\left(\alpha^\rho \left| g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)\right| \geq \kappa_n \epsilon\right) \to 0.$$

Then (vii) corresponds to $\kappa_n = \alpha^{2\rho}$, and the Lindeberg condition corresponds to $\alpha = \epsilon_n^{-2\gamma}$ and $\kappa_n = \sqrt{n}$.

Next, the conditional (given the sample) covariance kernel of $\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_n^* g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)$ satisfies

$$\hat{\Sigma}_\rho(s,t) = \frac{1}{n}\sum_{i=1}^n \epsilon_n^{-4\gamma\rho} g\left(\cdot, \theta_0 + \epsilon^{2\gamma} t\right) g\left(\cdot, \theta_0 + \epsilon^{2\gamma} s\right) \to \Sigma_\rho(s,t), \tag{22}$$

almost surely by a strong law of large numbers for both the Wild and multinomial bootstrap. The conditional (in ξ) Lindeberg condition is satisfied if almost surely,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \epsilon_n^{-4\gamma\rho} g^2\left(z_i, \theta_0 + \epsilon_n^{2\gamma}\right) E\xi^2 1\left(|\xi|\epsilon_n^{-2\gamma\rho}|g\left(z_i, \theta_0 + \epsilon_n^{2\gamma}\right)| \geq \sqrt{n}\epsilon\right) \\
&\leq E\xi^2 1\left(|\xi|\frac{\max_i |g\left(z_i, \theta_0 + \epsilon_n^{2\gamma}\right)|}{\sqrt{n}\epsilon_n^{2\gamma\rho}} \geq \epsilon\right) \frac{1}{n}\sum_{i=1}^n \epsilon_n^{-4\gamma\rho} g^2\left(z_i, \theta_0 + \epsilon_n^{2\gamma}\right) \to 0.
\end{aligned} \tag{23}$$

This holds by the strong LLN and that almost surely, $\frac{\max_i |g\left(z_i,\theta_0+\epsilon_n^{2\gamma}\right)|}{\sqrt{n}\epsilon_n^{2\gamma\rho}} \to 0$ using (vi). Therefore almost surely in finite dimension, the following conditional (in ξ) weak convergence holds:

$$\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_n^* g(\cdot;\theta_0+\epsilon_n^{2\gamma}h) \overset{\xi}{\rightsquigarrow} \mathcal{Z}_0(h), \quad h = \{h_1,\ldots,h_J\}.$$

With the multinomial bootstrap, (23) is replaced by

$$\begin{aligned}
&E^*\epsilon_n^{-4\gamma\rho} g^2\left(z_i^*, \theta_0 + \epsilon_n^{2\gamma}h\right) 1\left(\frac{1}{\sqrt{n}}\epsilon_n^{-2\gamma\rho}|g\left(z_i^*, \theta_0 + \epsilon_n^{2\gamma}h\right)| \geq \epsilon\right) \\
&= \frac{1}{n}\sum_{i=1}^n \epsilon_n^{-4\gamma\rho} g^2\left(z_i, \theta_0 + \epsilon_n^{2\gamma}h\right) 1\left(\frac{1}{\sqrt{n}}\epsilon_n^{-2\gamma\rho}|g\left(z_i, \theta_0 + \epsilon_n^{2\gamma}h\right)| \geq \epsilon\right) \to 0
\end{aligned}$$

almost surely by Strong LLN and conditions (vi) and (vii).

Finally we show (unconditional stochastic equicontinuity) of the $\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_n^* g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)$ part of W_n^*, separately for the wild and multinomial bootstrap. For the wild bootstrap,

$$\epsilon_n^{-2\gamma\rho}\hat{\mathcal{G}}_n^* g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)=\epsilon_n^{-2\gamma\rho}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i\left(\delta_i-P\right)g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)+\bar{\xi}_i\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h).$$

Since $\bar{\xi}\overset{a.s.}{\to}0$, the second term is $o_{a.s}\left(1\right)$ in $\ell_\infty\left(K\right)$. The first term is handled in van der Vaart and Wellner (1996) Lemmas 2.3.6 and 2.9.1, and is stochastically equicontinuous whenever $\hat{\mathcal{G}}_n\epsilon_n^{-2\gamma\rho}g(\cdot;\theta_0+\epsilon_n^{2\gamma}h)$ is. Combined with the unconditional versions of (22) and (23),

$$\epsilon_n^{-2\gamma\rho}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i\left(\delta_i-P\right)g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\rightsquigarrow\mathcal{Z}_0\left(h\right)\quad\text{in}\quad\ell_\infty\left(K\right)\tag{24}$$

Next using the approximation scheme in van der Vaart and Wellner (1996) Theorem 2.9.6, (24), (22) and (23),

$$\epsilon_n^{-2\gamma\rho}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i\left(\delta_i-P\right)g\left(\cdot,\theta_0+\epsilon_n^{2\gamma}h\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\mathcal{Z}_0\left(h\right)\quad\text{in}\quad\ell_\infty\left(K\right)\tag{25}$$

Finally, as the sum between (25) and $o_P\left(1\right)$ terms, $W_n^*\left(h\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\mathcal{Z}_0\left(h\right)$ in $\ell_\infty\left(K\right)$.

In the multinomial boostrap case, for $\mathscr{D}\left(n\right)$ in (21), we use a bootstrap version of Kim and Pollard (1990) to show $\epsilon_n^{-2\gamma\rho}E\underset{\mathscr{D}(n)}{sup}\left|\hat{\mathcal{G}}_n^* d\right|=o(1)$. As before bound, with $EP_n\epsilon_n^{-4\gamma\rho}P_n^*D_n^2=O\left(1\right)$,

$$\begin{aligned}\epsilon_n^{-2\gamma\rho}EP_n\underset{\mathscr{D}(n)}{sup}\left|\hat{\mathcal{G}}_n^* d\right|\leq&EP_n\sqrt{\epsilon_n^{-4\gamma\rho}P_n^*D_n^2}J\left(\frac{\epsilon_n^{-4\gamma\rho}\underset{\mathscr{D}(n)}{sup}P_n^*d^2}{\epsilon_n^{-4\gamma\rho}P_n^*D_n^2}\right)\\ \leq&\sqrt{\eta}J\left(1\right)+\sqrt{E\epsilon_n^{-4\gamma\rho}P_n^*D_n^2}\sqrt{EJ^2\left(min\left(1,\frac{1}{\eta}\epsilon_n^{-4\gamma\rho}\underset{\mathscr{D}(n)}{sup}P_n^*d^2\right)\right)}\end{aligned}$$

Finally to show that $E\epsilon_n^{-4\gamma\rho}\underset{\mathscr{D}(n)}{sup}P_n^* d^2 = o(1)$, split using large K,

$$
\begin{aligned}
E\underset{\mathscr{D}(n)}{sup}P_n^* d^2 \leq EP_n^* D_n^2 1 (D_n > K) + K\underset{\mathscr{D}(n)}{sup}P|d| \\
+ KE\underset{\mathscr{D}(n)}{sup}\left|P_n|d| - P|d|\right| + KE\underset{\mathscr{D}(n)}{sup}\left|P_n^*|d| - P_n|d|\right|.
\end{aligned}
$$

The first three terms are handled after equation (21). The last term is handled by a bootstrap version of Kim and Pollard (1990) maximal inequality 3.1, with $P_n D_n^2 = PD_n^2 + O_P\left(n^{-1/2}\right)$,

$$
EP_n\underset{\mathscr{D}(n)}{sup}\left|P_n^*|d| - P_n|d|\right| \leq J(1)\, En^{-1/2}\sqrt{P_n D_n^2} = O\left(n^{-1/2}\left(\epsilon_n^{2\gamma\rho} + n^{-1/4}\right)\right).
$$

This is $o\left(\epsilon_n^{4\gamma\rho}\right)$ when $\gamma = 1/3, \rho = 1/2$. For larger values of $\rho > 2/3$ and γ we will impose the additional condition that $n^{3/4}\epsilon_n^{4\gamma\rho} \to \infty$ to achieve $o\left(\epsilon_n^{4\gamma\rho}\right)$. Alternatively, we can impose Holder continuity to argue directly that $EP_n\underset{\mathscr{D}(n)}{sup}\left|P_n^*|d| - P_n|d|\right| = O(\delta_n) = o(1)$.

It is also clear from the above arguments that the bootstrapped statistic converges unconditionally to the limiting distribution of the root: $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \rightsquigarrow \mathcal{J}$. This is because unconditionally, $\epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) = O_P(1)$,

$$
\epsilon_n^{-(1+2\rho\gamma)}\mathcal{Z}_n^* g\left(\cdot, \theta_0 + \epsilon_n^{2\gamma}h\right) \rightsquigarrow \mathcal{Z}_0(h) - \frac{1}{2}h'Hh \quad \text{in} \quad \ell_\infty(K),
$$

so that the unconditional arg max theorem van der Vaart and Wellner (1996) 3.2.2 can be applied. ■

Lemma 10.1 *Under conditions (viii) and (ix) of Theorem 4.1, for each η there exist random variables $M_n^* = O_P^*(1)$ such that for all θ close to θ_0,*

$$
\epsilon_n|\hat{\mathcal{G}}_n^* g(\cdot, \theta)| \leq \eta|\theta - \theta_0|^2 + \epsilon_n^{4\gamma}M_n^{*2}.
$$

Proof: We first consider the Wild Bootstrap where $\hat{\mathcal{G}}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\xi_i - \bar{\xi}\right) \delta_i$. Also WLOG let $\theta_0 = 0$. Define $A(n,j) = \{\theta : (j-1)\epsilon_n^{2\gamma} \le |\theta| \le j\epsilon_n^{2\gamma}\}$. Then

$$\begin{aligned} P_\xi \left(M_n^* > m\right) \le & P_\xi \left(\exists \theta : \epsilon_n |\hat{\mathcal{G}}_n^* g(\cdot,\theta)| > \eta|\theta|^2 + \epsilon_n^{4\gamma} m^2\right) \\ & \le \sum_{j=1}^{\infty} P_n \{\exists \theta \in A(n,j) : \epsilon_n^{-4\gamma} \epsilon_n |\hat{\mathcal{G}}_n^* g(\cdot,\theta)| > \eta (j-1)^2 + m^2\}. \end{aligned}$$

The jth summand is then bounded by

$$\epsilon_n^{-8\gamma} \epsilon_n^2 P_\xi \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_n^* g(\cdot,\theta)|^2 / \left[(j-1)^2 + m^2\right]^2 \tag{26}$$

Note that $\hat{\mathcal{G}}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (\delta_i - P) - \bar{\xi}\hat{\mathcal{G}}_n$. We bound the expectation in the numerator in (26) by

$$P_\xi \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_n^* g(\cdot,\theta)|^2 \le P_\xi \sup_{|\theta|<j\epsilon_n^{2\gamma}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (\delta_i - P) g(\cdot,\theta) \right|^2 + \bar{\xi}^2 \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_n g(\cdot,\theta)|^2.$$

By Maximal Inequality 3.1 in Kim and Pollard (1990) and also $\bar{\xi} = o_P(1)$, $\bar{\xi} \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_n g(\cdot,\theta)|^2 = o_p\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right)$. Next by Lemmas 2.3.6 and (a square version of) 2.9.1 of van der Vaart and Wellner (1996), for large n_0 and n,

$$P \sup_{|\theta|<j\epsilon_n^{2\gamma}} \frac{|\sum_{i=1}^n \xi_i (\delta_i - P) g(\cdot,\theta)|^2}{\sqrt{n}} \le O\left(P \sup_{|\theta|<j\epsilon_n^{2\gamma}} |g(\cdot,\theta)|^2 + \max_{n_0 \le k \ge n} P \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_k g(\cdot,\theta)|^2\right).$$

Both terms are $O\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right)$ by (viii) and again Kim and Pollard (1990) Maximal Inequality 3.1. Therefore

$$P_\xi \sup_{|\theta|<j\epsilon_n^{2\gamma}} |\hat{\mathcal{G}}_n^* g(\cdot,\theta)|^2 = O_p\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right).$$

The numerator of (26) is thus further bounded by, for $C_n' = O_p(1)$, $\epsilon_n^{-8\gamma}\epsilon_n^2 C_n' \left(\epsilon_n^{2\gamma}\right)^{2\rho} = C_n'$. Then $M_n^* = O_P^*(1)$ since by choosing m, the following can be made asymptotically small. $\forall \epsilon > 0$,

$$P\left(P_n\left(M_n^* > m\right) > \epsilon\right) \leq P\left(C_n' \sum_{j=1}^{\infty} 1/\left[\left(j-1\right)^2 + m^2\right]^2 > \epsilon\right).$$

Next consider the multinomial bootstrap $\hat{\mathcal{G}}_n^* = \sqrt{n}\left(P_n^* - P_n\right)$. Note a bootstrap version of part (ii) of Maximal Inequality 3.1 in Kim and Pollard (1990) holds with:

$$nP_n \sup_{\mathcal{F}} \left|P_n^* f - P_n f\right|^2 \leq J\left(1\right)^2 P_n F^2 = J\left(1\right)^2 PF^2 + O_P\left(\frac{1}{\sqrt{n}}\right). \tag{27}$$

This can be used to similarly bound the numerator in (26) by $\epsilon_n^{-8\gamma}\epsilon_n^2 C_n' \left(\epsilon_n^{2\gamma}\right)^{2\rho} = C_n'$ for $C_n' = O_p\left(1\right)$. Strictly speaking, (27) is $O_P\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right)$ only when $n^{-1/2} = O\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right)$, which holds when $\rho = 1/2, \gamma = 1/3$ but not when $\rho = 1, \gamma = 1/2$. However, under the additional assumption that $g\left(\cdot, \theta\right)$ is Holder with index ρ, it can be directly verified that $P_n F^2 = O_P\left(\left(\epsilon_n^{2\gamma}\right)^{2\rho}\right)$. ■

Lemma 10.2 *Let* $\mathbb{M}_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{M}$ *in* $\ell_\infty\left(K\right)$ *for every compact* K*. If there exists a tight* $\hat{h}$ *such that for every open* G *containing* $\hat{h}$*,* $\mathbb{M}\left(\hat{h}\right) > \sup_{h \notin G, h \in K} \mathbb{M}\left(h\right)$*, and* $\mathbb{M}_n\left(\hat{h}_n\right) \geq \sup_h \mathbb{M}_n\left(h\right) - o_P^*\left(1\right)$*, where* $\hat{h}_n = O_P^*\left(1\right)$*, then* $\hat{h}_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \hat{h}$*.*

Proof: First a bootstrap version of the Portmanteau theorem can be shown. The following are equivalent (TFAE): (i) $X_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} X$; (ii) For every open G and $\forall \epsilon > 0$, $P\left(P_n\left(X_n \in G\right) \geq P\left(X \in G\right) - \epsilon\right) \to 1$; (iii) For every closed F and $\forall \epsilon > 0$, $P\left(P_n\left(X_n \in F\right) \leq P\left(X \in F\right) + \epsilon\right) \to 1$. By the bootstrap CMT (Theorem 10.8 Kosorok (2007)), $\sup_{h \in F \cap K} \mathbb{M}_n\left(h\right) - \sup_{h \in K} \mathbb{M}_n\left(h\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \sup_{h \in F \cap K} \mathbb{M}\left(h\right) - \sup_{h \in K} \mathbb{M}\left(h\right)$. Then by bootstrap Portmanteau, with probability converging to 1 (w.p.c.1), $\forall \epsilon > 0$,

$$\begin{aligned} P_n\left(\hat{h}_n \in F \cap K\right) &\leq P_n\left(\sup_{h \in F \cap K} \mathbb{M}_n\left(h\right) \geq \sup_{h \in K} \mathbb{M}_n\left(h\right) - o_P^*\left(1\right)\right) \\ &\leq P\left(\sup_{h \in F \cap K} \mathbb{M}\left(h\right) \geq \sup_{h \in K} \mathbb{M}\left(h\right)\right) + \epsilon \leq P\left(\hat{h} \in F\right) + P\left(\hat{h} \notin K\right) + \epsilon. \end{aligned}$$

Next split up $P_n\left(\hat{h}_n \in F\right) \leq P_n\left(\hat{h}_n \in F \cap K\right) + P_n\left(\hat{h} \notin K\right)$. Choose K large so that $P_n\left(\hat{h} \notin K\right) \leq \epsilon$ has probability larger than $1 - \delta$ for large n. Conclude that with probability larger than $1-2\delta$ for large n, $P_n\left(\hat{h}_n \in F\right) \leq P\left(\hat{h} \in F\right) + 2\epsilon$. ■

Proof for Theorem 4.2 Part 1: Define $w_n(h) = +\infty 1\left(h \notin n^{\gamma}(C-\theta_0)\right)$ and $w(h) = +\infty 1\left(h \notin T_C(\theta_0)\right)$. By (4.4), $w_n(\cdot) \stackrel{e}{\rightarrow} w(\cdot)$ as a sequence of nonrandom functions. Next define

$$\mathcal{H}_n(h) = n^{\gamma\rho}\sqrt{n}P_n g\left(\cdot, \theta_0 + n^{-\gamma}h\right) + w_n(h).$$

Similarly define $\mathcal{H}(h) = \mathcal{Z}_0(h) + \frac{1}{2}h'Hh + w(h)$. Then for $\hat{h}_n = n^{\gamma}\left(\hat{\theta}_n - \theta_0\right)$, $\mathcal{H}_n\left(\hat{h}_n\right) = \inf_h \mathcal{H}_n(h) + o_P(1)$. Almost the same arguments as in the proof of Theorem 4.1 can be applied to show that $\hat{h}_n = O_P(1)$. As in the proof of Theorem 4.3 of Geyer (1994), we confine our attention to compact sets.

In Theorem 4.1 it has been shown that $n^{\gamma\rho}\sqrt{n}P_n g\left(\cdot, \theta_0 + n^{-\gamma}h\right) \rightsquigarrow \mathcal{Z}_0(h) + \frac{1}{2}h'Hh$ in $\ell_\infty(K)$ in the sense of finite dimensional convergence and stochastic equicontinuity. Furthermore, $\mathcal{Z}_0(h) + \frac{1}{2}h'Hh$ has a continuous sample path. Then according to page 5 in Knight (1999), $n^{\gamma\rho}\sqrt{n}P_n g\left(\cdot, \theta_0 + n^{-\gamma}h\right) \rightarrow_{u-d} \mathcal{Z}_0(h) + \frac{1}{2}h'Hh$. Next by Theorem 4 in Knight (1999), $\mathcal{H}_n(\cdot) \rightarrow_{e-d} \mathcal{H}(\cdot)$. The remaining arguments are the same as in the second part of the proof of Theorem 4.4 in Geyer (1994). In other words, the arg min functional is continuous with respect to the metric of epi-convergence on the space of functions embedding $\mathcal{H}_n(\cdot)$ and $\mathcal{H}(\cdot)$, which allows for the application of a continuous mapping theorem as in Theorem 1 of Knight (1999).

Part 2: As before, for $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right)$, it suffices to show that $\hat{h}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$ and $\hat{h}_n^* \rightsquigarrow \mathcal{J}$. Define $w_n^*(h) = +\infty 1\left(h \notin \epsilon_n^{-2\gamma}(C-\theta_0)\right)$, so that by (4.4), $w_n^*(\cdot) \stackrel{e}{\rightarrow} w(\cdot)$ as a nonrandom sequence. Next let $\mathcal{H}_n^*(h) = \epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot, \theta_0 + \epsilon_n^{\gamma}h\right) + w_n(h)$. Then for $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right)$, by assumption, $\mathcal{H}_n^*\left(\hat{h}_n^*\right) = \inf_h \mathcal{H}_n^*(h) + o_P(1)$. Almost the same arguments as in the proof of Theorem 4.1 can be applied to show that $\hat{h}_n^*$ is $O_P^*(1)$ and $O_P(1)$, so we can confine our attention to compact sets. Theorem 4.1 has shown that $\epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot, \theta_0 + \epsilon_n^{\gamma}h\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} (\rightsquigarrow)\mathcal{Z}_0(h)$ in $\ell_\infty(K)$. A bootstrap in probability version of Theorem 4 Knight (1999) can be stated to show that $\mathcal{H}_n^*(\cdot) \rightarrow_{e-d} \mathcal{H}(\cdot)$ conditionally in probability, which can be equivalently stated as $\rho_{BL_1}\left(\mathcal{H}_n^*(\cdot), \mathcal{H}(\cdot)\right) = o_p(1)$ where BL_1 now represents the class of Lipschitz norm 1 functions with respect to the metric of epi-convergence (see last equation on page 4 of Knight (1999)), and $\mathcal{H}_n^*(\cdot)$ is understood to be the conditional law given the data. Finally, by revising the bootstrap argmax continuous mapping lemma 10.2 to replace weak convergence by epi-convergence after incorporating Theorem 1 in Knight (1999), we can show that $\hat{h}_n^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \hat{h}$.

Proof for Theorem 4.3 The same arguments as in the proof of Theorem 4.2 (see also Sherman (1993) and Newey and McFadden (1994)) show that in $\ell_\infty(K)$,

$$\begin{aligned}&n\left(\hat{Q}_n\left(\theta_0+h/\sqrt{n}\right)-\hat{Q}_n\left(\theta_0\right)\right)\\&\rightsquigarrow \Delta_0'h+\frac{1}{2}h'Hh=\left(h+H^{-1}\Delta_0\right)'H\left(h+H^{-1}\Delta_0\right)-\frac{1}{2}\Delta_0'H^{-1}\Delta_0.\end{aligned}$$

Then argue as in the previous proof that $\hat{h}_n=\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)=O_P(1)$ and that

$$\begin{aligned}&n\left(\hat{Q}_n\left(\theta_0+h/\sqrt{n}\right)-\hat{Q}_n\left(\theta_0\right)\right)+w_n(h)\\&\rightarrow_{e-d}\left(h+H^{-1}\Delta_0\right)'H\left(h+H^{-1}\Delta_0\right)-\frac{1}{2}\Delta_0'H^{-1}\Delta_0+w(h).\end{aligned}$$

for $w_n(h)=\infty 1\left(h\notin\sqrt{n}\left(C-\theta_0\right)\right)$ and $w(h)=\infty 1\left(h\notin T_C\left(\theta_0\right)\right)$. Therefore $\hat{\mathcal{J}}_n\rightsquigarrow\mathcal{J}$. Next note that $\hat{\mathcal{J}}_n^*=\bar{\mathcal{J}}_n^*-\epsilon_n^{-1}\left(\hat{\theta}_n-\theta_0\right)$ where

$$\bar{\mathcal{J}}_n^*=\arg\min_{\bar{h}\in\epsilon_n^{-1}(C-\theta_0)}\bar{Q}_n^*\left(\bar{h}\right)=\left(\bar{h}+\hat{H}^{-1}\hat{\Delta}_n^*-\frac{\hat{\theta}_n-\theta_0}{\epsilon_n}\right)'\hat{H}\left(\bar{h}+\hat{H}^{-1}\hat{\Delta}_n^*-\frac{\hat{\theta}_n-\theta_0}{\epsilon_n}\right).$$

Note that $\frac{\hat{\theta}_n-\theta_0}{\epsilon_n}=o_P(1)$, $\hat{H}^{-1}\hat{\Delta}_n^*\rightsquigarrow H^{-1}\Delta_0$. Therefore $\hat{H}^{-1}\hat{\Delta}_n^*-\frac{\hat{\theta}_n-\theta_0}{\epsilon_n}\rightsquigarrow H^{-1}\Delta_0$. Therefore by the bootstrap CMT (Proposition 10.7 in Kosorok (2007)), in $\ell_\infty(K)$,

$$\bar{Q}_n^*\left(\bar{h}\right)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\bar{Q}_\infty\left(\bar{h}\right)=\left(\bar{h}+H^{-1}\Delta_0\right)'H\left(\bar{h}+H^{-1}\Delta_0\right)$$

Therefore the same arguments as in part 2 of the proof of Theorem 4.2 apply. $\bar{Q}_n^*\left(\bar{h}\right)+w_n\left(\bar{h}\right)\rightarrow_{e-d}\bar{Q}_\infty\left(\bar{h}\right)+w\left(\bar{h}\right)$ conditionally in probability, $\bar{\mathcal{J}}_n^*\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{J}$, and $\hat{\mathcal{J}}_n^*\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{J}$.

Proof for Theorem 6.1 We follow the proof logic in Theorem 1(i) of Jun et al. (2015), who already show that $\hat{\mathcal{J}}_n\rightsquigarrow\mathcal{J}$. So we will focus on $\hat{\mathcal{J}}_n^*\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{J}$. It suffices to show that $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$. We write

$$\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\theta}_n^* - \theta_0\right) = \frac{\int hw\left(\theta_0 + h\epsilon_n^{2\gamma}\right)\exp\left(\epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right)dh}{\int w\left(\theta_0 + h\epsilon_n^{2\gamma}\right)\exp\left(\epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right)dh}$$

The goal is to show for any $\alpha \geq 0$, $\phi(H) = \exp\left(-\frac{1}{2}h'Hh\right)$, and $C_H = \sqrt{2\pi}^{d_\theta}\det(H)^{d_\theta/2}$,

$$\int h^\alpha w\left(\theta_0 + h\epsilon_n^{2\gamma}\right) e^{\frac{\mathcal{Z}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)}{\epsilon_n^{4\gamma}}} dh \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} C_H w(\theta_0)\int h^\alpha e^{\mathcal{Z}_0(h) - \frac{1}{2}h'Hh} dh.$$

Since Theorem 4.1 already shows that $\frac{(\mathcal{Z}_n^* - P)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)}{\epsilon_n^{4\gamma}} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_0(h)$ in $\ell_\infty(K)$, this will follow from CMT if it can be shown that, for $\omega_0(h) = \omega\left(\theta_0 + h\epsilon_n^{2\gamma}\right)$,

$$\int |h|^\alpha e^{\frac{(\mathcal{Z}_n^* - P)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)}{\epsilon_n^{4\gamma}}} \left| w_0(h) e^{\frac{Pg\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)}{\epsilon_n^{4\gamma}}} - w_0(0)\phi_H(h)\right| dh = o_P^*(1), \tag{28}$$

and if for any $\eta > 0$, $\exists K$ compact, such that w.p.c.1,

$$\int_{K^c} |h|^\alpha w\left(\theta_0 + h\epsilon_n^{2\gamma}\right)\exp\left(\epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right)dh < \eta. \tag{29}$$

Take $\gamma_n = o\left(\epsilon_n^{-2\gamma/3}\right)$, and let $\Gamma_n = \{h : |h| \leq \gamma_n\}$,

$$\begin{aligned}
\kappa_1(h) &= |h|^\alpha w_0(h)\exp\left(\epsilon_n^{-4\gamma}\mathcal{Z}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right), \\
\kappa_2(h) &= |h|^\alpha e^{\epsilon_n^{-4\gamma}(\mathcal{Z}_n^* - P)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)} w_0(0)\phi_H(h).
\end{aligned}$$

Bound (28) by bounding each of the terms in

$$\left|\int_{\Gamma_n}(\kappa_1 - \kappa_3)\right| + \left|\int_{\Gamma_n}(\kappa_3 - \kappa_2)\right| + \left|\int_{\Gamma_n^c}\kappa_2\right| + \left|\int_{\Gamma_n^c \cap \{\theta_0 + h\epsilon_n^{2\gamma} \in \Theta\}}\kappa_1\right| \tag{30}$$

where $\kappa_3(h) = |h|^\alpha w_0(h)\exp\left(\epsilon_n^{-4\gamma}(\mathcal{Z}_n^* - P)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right)\phi_H(h)$. Consider the first term in (30), bounded by $|\int_{\Gamma_n}|h|^\alpha w_0(h)\exp\left(\epsilon_n^{-4\gamma}(\mathcal{Z}_n^* - P)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)\right)|e^{R_n(h)} - 1|\phi_H(h)\,dh|$ for $R_n(h) = \epsilon_n^{-4\gamma}Pg\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right) + 1/2h'Hh$. Since Γ_n shrinks to θ_0, $\sup_{\Gamma_n}|R_n(h)| = O\left(\epsilon_n^{-4\gamma}\gamma_n^3\epsilon_n^{6\gamma}\right) = o(1)$. Next consider a bootstrap version of Lemma

B.5 in Jun et al. (2015). For any $\gamma^* \to \infty$, $\forall \epsilon > 0$,

$$P_n\left(\sup_{|h|\geq\gamma^*}\left|\epsilon_n^{-4\gamma}\left(\mathcal{Z}_n^* - P\right)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)/\left(1+|h|\right)\right| > \epsilon\right) = o_P\left(1\right) \tag{31}$$

The key to (31) is using condition G in Jun et al. (2015) to argue that

$$\epsilon_n^{1-4\gamma}\hat{\mathcal{G}}_n g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)/\left(1+|h|\right) \rightsquigarrow \mathcal{Z}_0\left(h\right)/\left(1+|h|\right)$$

and $\epsilon_n^{1-4\gamma}\hat{\mathcal{G}}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)/\left(1+|h|\right) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{Z}_0\left(h\right)/\left(1+|h|\right)$, both in $\ell_\infty\left(R^d\right)$ (and not only in $\ell_\infty\left(K\right)$). Jun et al. (2015) already handled $\epsilon_n^{1-4\gamma}\hat{\mathcal{G}}_n g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)/\left(1+|h|\right)$ so that

$$\epsilon_n^{-4\gamma}\left(P_n - P\right)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)/\left(1+|h|\right) = o_P\left(1\right),$$

in $\ell_\infty\left(R^d\right)$. By similar arguments, (31) follows from taking $j \to \infty$ in

$$\begin{aligned} P_n\left(\sup_{|h|\geq\gamma^*}\left|\frac{\epsilon_n^{-4\gamma}\hat{\mathcal{G}}_n^* g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)}{\left(1+|h|\right)}\right| > \epsilon\right) &\leq P_n\left(\sup_{|h|\geq j}\left|\frac{\mathcal{Z}_0\left(h\right)}{1+|h|}\right| > \epsilon\right) + o_P\left(1\right) \\ &\leq \frac{C}{\epsilon^4}\sum_{s=j+1}^{\infty} 1/s^2 + o_P\left(1\right). \end{aligned}$$

Then Lemma B.6 in Jun et al. (2015) shows that

$$\sup_{h\in R^d}|\epsilon_n^{-4\gamma}\left(\mathcal{Z}_n^* - P\right)g\left(\cdot,\theta_0 + h\epsilon_n^{2\gamma}\right)| - c|h| = O_P^*\left(1\right). \tag{32}$$

Then up to multiplicative $O_P^*\left(1\right)$, the first term in (30) is bounded by

$$\sup_{\Gamma_n}|R_n\left(h\right)|\left|\int_{\Gamma_n}|h|^\alpha \exp\left(c|h|\right)\phi_H\left(h\right)dh\right| = o\left(1\right)O_P^*\left(1\right) = o_P^*\left(1\right).$$

Similarly, (32) also bounds the second and third terms in (30) up to multiplicative $O_P^*\left(1\right)$ by,

$$\sup_{\Gamma_n}|\omega_0\left(h\right) - \omega_0\left(0\right)|\left|\int_{\Gamma_n}|h|^\alpha \exp\left(c|h|\right)\phi_H\left(h\right)dh\right| = o\left(1\right)O_P^*\left(1\right) = o_P^*\left(1\right).$$

and $\left|\int_{\Gamma_n^c} |h|^\alpha \exp\left(c|h|\right) \phi_H\left(h\right) dh\right| = o_P^*\left(1\right)$. Finally, via Jun et al. (2015) Lemma B.4,

$$\epsilon_n^{-4\gamma} Pg\left(\cdot, \theta_0 + \epsilon_n^{2\gamma} h\right) < -\min\left(h'Hh/4, c_\alpha \epsilon_n^{-4\gamma}\right).$$

Then with bounded Θ, using also (32), bound the 4th term in (30) up to multiplicative $O_P^*\left(1\right)$ by, for some large $M > 0$, small $\delta > 0$,

$$\int_{\Gamma_n^c \cap \{|h|\epsilon_n^{2\gamma} \leq \delta\}} |h|^\alpha e^{c|h|} e^{-\frac{1}{2}h'Hh} dh + e^{-c_\alpha \epsilon_n^{-4\gamma}} \int_{\{\delta < |h|\epsilon_n^{2\gamma} \leq M\}} |h|^\alpha e^{c|h|} dh.$$

Both terms are $o\left(1\right)$ as $\gamma_n \to \infty$ and $\epsilon_n \to 0$, where the 2nd term is $O\left(e^{-c_\alpha \epsilon_n^{-4\gamma}} e^{M\epsilon_n^{2\gamma}}\right)$. Finally, (29) indeed follows from identical arguments to the 4th term in (30).

Proof for Theorem 6.2 Since this is a straightforward extension of Theorem 6.1, it suffices to describe the key steps that differ. Recall that $\pi\left(\cdot\right)$ can always be replaced by $g\left(\cdot\right)$. First note that the convergence of integrals in Jun et al. (2015) and in Theorem 6.1 (such as (28) and (30)) all remain valid when the domain of integration is confined to $h \in n^\gamma\left(C - \theta_0\right)$. First consider (6). By the same arguments in Jun et al. (2015),

$$\int_{n^\gamma(C-\theta_0)} |h|^\alpha \left| e^{n^{2\gamma} P_n g\left(\cdot;\theta_0 + n^{-\gamma}h\right)} - e^{n^{2\gamma}(P_n - P) g\left(\cdot;\theta_0 + n^{-\gamma}h\right)} e^{-\frac{1}{2}h'Hh} \right| dh = o_P\left(1\right)$$

Hence we can always replace $\int_{n^\gamma(C-\theta_0)} h^\alpha \exp\left(n^{2\gamma} P_n g\left(\cdot;\theta_0 + n^{-\gamma}h\right)\right) dh$ with

$$\int_{n^\gamma(C-\theta_0)} h^\alpha \exp\left(n^{2\gamma}\left(P_n - P\right) g\left(\cdot;\theta_0 + n^{-\gamma}h\right)\right) e^{-\frac{1}{2}h'Hh} dh$$

in showing distributional convergence. Using the Skorohod representation, let $\mathcal{Z}_n\left(h\right) \equiv n^{2\gamma}\left(P_n - P\right) g\left(\cdot;\theta_0 + n^{-\gamma}h\right) \xrightarrow{a.s.} \mathcal{Z}_\infty\left(h\right)$. Then for all $\mathcal{M}$ in (3),

$$\int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_n\left(h\right)\right) e^{-\frac{1}{2}h'Hh} dh - \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_\infty\left(h\right)\right) e^{-\frac{1}{2}h'Hh} dh \xrightarrow{a.s.} 0.$$

Alternatively, we may note that $\int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} h^\alpha e^{\mathcal{Z}_n(h)} e^{-\frac{1}{2}h'Hh} dh$ is BL_1 in $\mathcal{Z}_n(\omega, h)$. Hence because the BL_1 property is closed under composition, for

$$X_{n,M} \equiv \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_n(h)\right) e^{-\frac{1}{2}h'Hh} dh \in R$$

and

$$Y_{n,M} \equiv \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_\infty(h)\right) e^{-\frac{1}{2}h'Hh} dh \in R$$

we have $\rho_{BL_1}(X_{n,M}, Y_{n,M}) \to 0$. Next by applying the dominated convergence theorem (DOM) almost surely,

$$Y_{n,M} \xrightarrow{a.s.} Y_M \equiv \int_{T_C(\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_\infty(h)\right) e^{-\frac{1}{2}h'Hh} dh. \tag{33}$$

Therefore $\rho_{BL_1}(X_{n,M}, Y_M) \to 0 \implies X_{n,M} \rightsquigarrow Y_M$ for all $\mathcal{M}$. Next by Lemmas B.2, B.5 and B.6 in Jun et al. (2015), since $\sup_{h\in R^d} |\mathcal{Z}_n(h)| - c|h| = O_P(1)$, for all $\delta, \epsilon > 0$ we can find $\mathcal{M}$ large enough so that

$$P\left(\int_{h\notin\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_\infty(h)\right) e^{-\frac{1}{2}h'Hh} dh > \epsilon\right) < \delta \tag{34}$$

and

$$\limsup_{n\to\infty} P\left(\int_{h\notin\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_n(h)\right) e^{-\frac{1}{2}h'Hh} dh > \epsilon\right) < \delta \tag{35}$$

Then we combine (35), (34) and $X_{n,M} \rightsquigarrow Y_M$ for all $\mathcal{M}$ to conclude that

$$X_n \equiv \int_{n^\gamma(C-\theta_0)} h^\alpha \exp\left(\mathcal{Z}_n(h)\right) e^{-\frac{1}{2}h'Hh} dh \rightsquigarrow Y \equiv \int_{T_C(\theta_0)} h^\alpha \exp\left(\mathcal{Z}_\infty(h)\right) e^{-\frac{1}{2}h'Hh} dh.$$

Hence also (jointly in finite set of α):

$$\int_{n^\gamma(C-\theta_0)} h^\alpha \exp\left(n^{2\gamma} P_n g\left(\cdot; \theta_0 + n^{-\gamma} h\right)\right) dh \rightsquigarrow \int_{T_C(\theta_0)} h^\alpha \exp\left(\mathcal{Z}_n(h)\right) e^{-\frac{1}{2}h'Hh} dh.$$

Then (8) follows from CMT.

Next consider (7). By the same arguments leading to (28),

$$\int_{\epsilon_n^{-2\gamma}(C-\theta_0)} |h|^\alpha e^{\epsilon_n^{-4\gamma}(\mathcal{Z}_n^*-P)g\left(\cdot,\theta_0+h\epsilon_n^{2\gamma}\right)} \left| e^{\epsilon_n^{-4\gamma}Pg\left(\cdot,\theta_0+h\epsilon_n^{2\gamma}\right)} - \phi_H(h) \right| dh = o_P^*(1),$$

so we can focus on analyzing the convergence of

$$X_n^* = \int_{\epsilon_n^{-2\gamma}(C-\theta_0)} h^\alpha e^{\epsilon_n^{-4\gamma}(\mathcal{Z}_n^*-P)g\left(\cdot,\theta_0+h\epsilon_n^{2\gamma}\right)} \phi_H(h)\, dh.$$

As in the proof of Theorem 6.1,

$$\epsilon_n^{-4\gamma}(\mathcal{Z}_n^*-P)\, g\left(\cdot,\theta_0+h\epsilon_n^{2\gamma}\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_\infty(h) \quad \text{in} \quad \ell_\infty(K)$$

Then using the same BL_1 arguments as above, for

$$X_{n,M}^* \equiv \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\epsilon_n^{-4\gamma}(\mathcal{Z}_n^*-P)\, g\left(\cdot,\theta_0+h\epsilon_n^{2\gamma}\right)\right) e^{-\frac{1}{2}h'Hh} dh$$

and

$$Y_{n,M}^* \equiv \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} h^\alpha \exp\left(\mathcal{Z}_\infty(h)\right) e^{-\frac{1}{2}h'Hh} dh,$$

there is $\rho_{BL_1}\left(X_{n,M}^*, Y_{n,M}^*\right) = o_P(1)$ for all $\mathcal{M}$ with radius M. Also as in (33) $\rho_{BL_1}\left(Y_{n,M}^*, Y_M\right) = o(1)$ for all M. Therefore $\rho_{BL_1}\left(X_{n,M}^*, Y_M\right) = o_P(1)$ for all M. Hence we can also find a sequence of $M_n \to \infty$ sufficiently slowly such that $\rho_{BL_1}\left(X_{n,M_n}^*, Y_{M_n}\right) = o_P(1)$.

Next use (32) to bound

$$\int_{\mathcal{M}_n^c} |h|^\alpha e^{\mathcal{Z}_n^*(h)} e^{-\frac{1}{2}h'Hh} dh = O_P^*(1) \int_{\mathcal{M}_n^c} |h|^\alpha e^{c|h|} e^{-\frac{1}{2}h'Hh} dh = O_P^*(1)\, o_P(1) = o_P^*(1).$$

Furthermore by (27) in Lemma B.5 of Jun et al. (2015), $\sup_{|h|>M_n} \frac{\mathcal{Z}_\infty(h)}{1+|h|} = O_P(1)$:

$$\int_{\mathcal{M}_n^c} |h|^\alpha e^{\mathcal{Z}_\infty(h)} e^{-\frac{1}{2}h'Hh} dh = O_P(1) \int_{\mathcal{M}_n^c} |h|^\alpha e^{c|h|} e^{-\frac{1}{2}h'Hh} dh = O_P(1)\, o_P(1) = o_P(1).$$

It then follows from $X_n^* = X_{n,M_n}^* + o_P^*(1)$, $Y = Y_{M_n} + o_p(1)$, and $\rho_{BL_1}\left(X_{n,M_n}^*, Y_{M_n}\right) = o_P(1)$, that $\rho_{BL_1}(X_n^*, Y) = o_P(1)$. Finally note that the above convergence rates all hold jointly in a finite set of $\alpha \geq 0$. Then $\hat{\mathcal{J}}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$ follows from the bootstrap CMT (e.g. Proposition 10.7) in Kosorok (2007)).

Proof for Theorem 6.3 For $\bar{h} = h + \epsilon_n^{-1}\left(\hat{\theta}_n - \theta_0\right)$ and for $\epsilon_n^{-1}\left(\hat{\theta}_n - \theta_0\right) = o_P(1)$, we can write

$$\begin{aligned}\hat{\mathcal{J}}_n^* &= \frac{\int_{\epsilon_n^{-1}(C-\theta_0)} \bar{h} e^{(\bar{h}+o_P(1))'\hat{\Delta}_n^* - \frac{1}{2}(\bar{h}+o_P(1))'H(\bar{h}+o_P(1))} d\bar{h}}{\int_{\epsilon_n^{-1}(C-\theta_0)} e^{(\bar{h}+o_P(1))'\hat{\Delta}_n^* - \frac{1}{2}(\bar{h}+o_P(1))'H(\bar{h}+o_P(1))} d\bar{h}} \\ &= \frac{\int_{\epsilon_n^{-1}(C-\theta_0)} \bar{h} e^{\bar{h}'(\hat{\Delta}_n^*+o_P(1)) - \frac{1}{2}\bar{h}'\hat{H}\bar{h}+o_P(1)} d\bar{h}}{\int_{\epsilon_n^{-1}(C-\theta_0)} e^{(\bar{h}+o_P(1))'\hat{\Delta}_n^* - \frac{1}{2}(\bar{h}+o_P(1))'H(\bar{h}+o_P(1))} d\bar{h}}\end{aligned}$$

It then suffices to show that jointly in a finite collection of $\alpha \geq 0$, for

$$X_n^* = \int_{\epsilon_n^{-1}(C-\theta_0)} \bar{h}^\alpha e^{\left(\bar{h}+o_P(1)\right)'\hat{\Delta}_n^* - \frac{1}{2}\left(\bar{h}+o_P(1)\right)'H\left(\bar{h}+o_P(1)\right)} d\bar{h}$$

and Y in the proof of Theorem 6.2, $\rho_{BL_1}(X_n^*, Y) = o_P(1)$. Define

$$X_{n,M}^* = \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} \bar{h}^\alpha e^{\left(\bar{h}+o_P(1)\right)'\hat{\Delta}_n^* - \frac{1}{2}\left(\bar{h}+o_P(1)\right)'H\left(\bar{h}+o_P(1)\right)} d\bar{h}.$$

By the same BL_1 argument in the previous proof, and the previously defined $Y_{n,M}^*$, $\rho_{BL_1}\left(X_{n,M}^*, Y_{n,M}^*\right) = o_P(1)$. Furthermore, for any $M_n \to \infty$

$$\int_{\bar{h}\notin\mathcal{M}_n} \bar{h}^\alpha e^{\bar{h}'\left(\hat{\Delta}_n^*+o_P(1)\right) - \frac{1}{2}\bar{h}'\hat{H}\bar{h}+o_P(1)} d\bar{h} = o_P^*(1).$$

The same arguments in the previous proof then apply.

Proof for Theorem 6.5 Consider first (1):

$$\hat{h}_n = \int_{n^\gamma(C-\theta_0)} n^\gamma\left(\phi\left(\theta_0 + n^{-\gamma}h\right) - \phi_0\right)\hat{p}_n(h)\,dh.$$

Let $\mathcal{W}_n(h) = n^{2\gamma}P_n\pi\left(z_i;\theta_0 + n^{-\gamma}h\right)$ in (4) or $\mathcal{W}_n(h) = n\hat{Q}_n\left(\theta_0 + h/\sqrt{n}\right)$ in (11), so that $\mathcal{W}_n(h) \rightsquigarrow \mathcal{W}_\infty(h) = \mathcal{Z}_\infty(h) - \frac{1}{2}h'Hh$ in $\ell_\infty(K)$. For each M, let

$$X_n^M = \left\{\int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} n^\gamma\left(\phi\left(\theta_0 + n^{-\gamma}h\right) - \phi_0\right)e^{\mathcal{W}_n(h)}dh, \quad \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_n(h)}dh\right\}$$

and

$$Y_n^M = \left\{ \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} n^\gamma \left(\phi\left(\theta_0 + n^{-\gamma}h\right) - \phi_0\right) e^{\mathcal{W}_\infty(h)} dh, \quad \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_\infty(h)} dh \right\}$$

By the same BL_1 embedding argument in the proof of Theorem 6.2, $\rho_{BL_1}\left(X_n^M, Y_n^M\right) = o(1)$. Next apply DOM almost surely:

$$Y_n^M \xrightarrow{a.s.} Y^M = \left\{ \int_{T_C(\theta_0)\cap\mathcal{M}} \phi'_{\theta_0}(h) e^{\mathcal{W}_\infty(h)} dh, \quad \int_{T_C(\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_\infty(h)} dh \right\}$$

Next by the polynomial growth condition on $\phi(\cdot)$ and the exponentially small tail of both $\mathcal{W}_n(h)$ and $\mathcal{W}_\infty(h)$, for any $M_n \to \infty$, we have

$$\left\{ \int_{\mathcal{M}_n^c} n^\gamma \left(\phi\left(\theta_0 + n^{-\gamma}h\right) - \phi_0\right) e^{\mathcal{W}_n(h)} dh, \quad \int_{\mathcal{M}_n^c} e^{\mathcal{W}_n(h)} dh \right\} = o_P(1),$$

and $\left\{ \int_{\mathcal{M}_n^c} \phi'_{\theta_0}(h) e^{\mathcal{W}_\infty(h)} dh, \quad \int_{\mathcal{M}_n^c} e^{\mathcal{W}_\infty(h)} dh \right\} = o_P(1)$. Therefore $X_n \equiv X_n^{M=\infty} \rightsquigarrow Y \equiv Y^{M=\infty}$ follows from taking a sequence $M_n \to \infty$ slowly enough so that

$$\rho_{BL_1}\left(X_n^{M_n}, Y_n^{M_n}\right) = o(1), \ \rho_{BL_1}\left(Y_n^{M_n}, Y^{M_n}\right) = o(1), \ X_n = X_n^{M_n} + o_P(1)$$

and $Y = Y^{M_n} + o_P(1)$. Then $\hat{h}_n = X_{n,1}/X_{n,2} \rightsquigarrow \mathcal{J} = Y_1/Y_2$ follows from CMT.

Next consider (2). As before since $\epsilon_n^{-2\gamma}\left(\hat{\phi} - \phi_0\right) = o_P(1)$, it suffices to consider (using redefined notation) $\hat{h}_n^* = \epsilon_n^{-2\gamma}\left(\hat{\phi}^* - \phi_0\right)$ which can be rewritten as

$$\hat{h}_n^* = \int_{h\in\epsilon_n^{-2\gamma}(C-\theta_0)} \epsilon_n^{-2\gamma}\left(\phi\left(\theta_0 + \epsilon_n^{2\gamma}h\right) - \phi_0\right) \hat{p}_n^*(h)\, dh,$$

where $\hat{p}_n^*(h)$ are defined in (7), (12), and (9) (with $\hat{\theta}_n$ replaced by θ_0 and $\hat{\mathbb{G}}_n^*$ by $\hat{\mathbb{G}}_n^* + \epsilon_n\left(\hat{\theta}_n - \theta_0\right)$). Note $\hat{p}_{n,h}^*(h) = \epsilon_n^{2\gamma}\hat{p}_{n,\theta}^*\left(\hat{\theta}_n + \epsilon_n^{2\gamma}h\right)$ and $\hat{p}_{n,\theta}^*(\theta) = \epsilon_n^{-1}\hat{p}_{n,h}^*\left(\epsilon_n^{-1}\left(\theta - \hat{\theta}_n\right)\right)$. Let $\mathcal{W}_n^*(h)$ be $\pi_n^*\left(\theta_0 + \epsilon_n^{-2\gamma}h\right)$ in (7), $-\frac{\left(h-\hat{\mathbb{G}}_n^*\right)'\hat{H}\left(h-\hat{\mathbb{G}}_n^*\right)}{2}$ in (9), and $\frac{\hat{Q}_n^*(\theta_0+\epsilon_n h)}{\epsilon_n}$ in (12). Then $\mathcal{W}_n^*(h) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{W}_\infty(h)$ in $\ell_\infty(K)$. For each M, let

$$X_{n,M}^* = \left\{ \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} \epsilon_n^{-2\gamma}\left(\phi\left(\theta_0 + \epsilon_n^{-2\gamma}h\right) - \phi_0\right) e^{\mathcal{W}_n^*(h)} dh, \quad \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_n^*(h)} dh \right\}$$

and

$$Y^*_{n,M} = \left\{ \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} \epsilon_n^{-2\gamma} \left(\phi\left(\theta_0 + \epsilon_n^{-2\gamma} h\right) - \phi_0\right) e^{\mathcal{W}_\infty(h)} dh, \quad \int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_\infty(h)} dh \right\}$$

First by BL_1 embedding, $\rho_{BL_1}\left(X^*_{n,M}, Y^*_{n,M}\right) = o_P(1)$. Next by almost sure DOM, $\rho_{BL_1}\left(Y^*_{n,M}, Y^M\right) = o(1)$. Furthermore, $\forall M_n \to \infty$, both

$$\left\{ \int_{\mathcal{M}_n^c} \epsilon_n^{-2\gamma} \left(\phi\left(\theta_0 + \epsilon_n^{-2\gamma} h\right) - \phi_0\right) e^{\mathcal{W}_n^*(h)} dh, \quad \int_{\mathcal{M}_n^c} e^{\mathcal{W}_n^*(h)} dh \right\} = o_P^*(1),$$

and $\left\{ \int_{\mathcal{M}_n^c} \epsilon_n^{-2\gamma} \left(\phi\left(\theta_0 + \epsilon_n^{-2\gamma} h\right) - \phi_0\right) e^{\mathcal{W}_\infty(h)} dh, \quad \int_{\mathcal{M}_n^c} e^{\mathcal{W}_\infty(h)} dh \right\} = o_P(1)$. Then find some $M_n \to \infty$ such that $\rho_{BL_1}\left(X^*_{n,M_n}, Y^*_{n,M_n}\right) = o_P(1)$, $\rho_{BL_1}\left(Y^*_{n,M_n}, Y^{M_n}\right) = o(1)$. Then use $X_n^* = X^*_{n,M_n} + o_P^*(1)$, $Y = Y^{M_n} + o_P(1)$ to conclude that $X_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} Y$. Apply CMT for $\hat{h}_n^* \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{J}$.

Next consider (3). Note that we can equivalently write

$$\hat{\phi}_\tau = \inf\left\{ x : \int_{n^\gamma(C-\theta_0), \phi(\theta_0 + n^{-\gamma} h) \le x} \hat{p}_n(h)\, dh \ge \tau \right\},$$

and $\hat{h}_\tau = n^\gamma \left(\hat{\phi}_\tau - \phi_0\right) = \inf\{x : \int_{n^\gamma(C-\theta_0), n^\gamma(\phi(\theta_0+n^{-\gamma}h)-\phi_0) \le x} \hat{p}_n(h)\, dh \ge \tau\}$. Given M, define random processes on $\ell_\infty\left(R^d\right) \times R$ as

$$X_n^M(\cdot) = \left\{ \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}, n^\gamma(\phi(\theta_0+n^{-\gamma}h)-\phi_0) \le \cdot} e^{\mathcal{W}_n(h)} dh, \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_n(h)} dh \right\}$$

and

$$Y_n^M(\cdot) = \left\{ \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}, n^\gamma(\phi(\theta_0+n^{-\gamma}h)-\phi_0) \le \cdot} e^{\mathcal{W}_\infty(h)} dh, \int_{n^\gamma(C-\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_\infty(h)} dh \right\}$$

Then by embedding the BL_1 norm, $\rho_{BL_1}\left(X_n^M(\cdot), Y_n^M(\cdot)\right) = o(1)$. Next by directional differentiability, uniformly in x,

$$\left|1\left(n^\gamma\left(\phi\left(\theta_0 + n^{-\gamma} h\right) - \phi_0\right) \le x, |h| \le M\right) - 1\left(\phi'_{\theta_0}(h) \le x, |h| \le M\right)\right| \le 1\left(\left|\phi'_{\theta_0}(h)\right| \le o(1)\right).$$

Therefore for $Y^M(\cdot) = \left\{\int_{T_C(\theta_0)\cap\mathcal{M},\phi'_{\theta_0}(h)\le\cdot} e^{\mathcal{W}_\infty(h)}dh, \int_{T_C(\theta_0)\cap\mathcal{M}} e^{\mathcal{W}_\infty(h)}dh\right\}$,

$$\sup_x |Y_n^M(x) - Y^M(x)| \to 0 \quad \text{almost surely} \Longrightarrow \rho_{BL_1}\left(Y_n^M(\cdot), Y^M(\cdot)\right) = o(1).$$

So $\exists M_n \to 0$ slowly enough that $\rho_{BL_1}\left(X_n^{M_n}(\cdot), Y^M(\cdot)\right) = o(1)$. Next for any $M_n \to \infty$,

$$\left\{\sup_x \int_{\mathcal{M}_n^c, n^\gamma(\phi(\theta_0+n^{-\gamma}h)-\phi_0)\le x} e^{\mathcal{W}_n(h)}dh, \int_{\mathcal{M}_n^c} e^{\mathcal{W}_n(h)}dh\right\} = o_P(1),$$

and

$$\left\{\sup_x \int_{\mathcal{M}_n^c,\phi'_{\theta_0}(h)\le x} e^{\mathcal{W}_\infty(h)}dh, \int_{\mathcal{M}_n^c} e^{\mathcal{W}_\infty(h)}dh\right\} = o_P(1).$$

Namely, $X_n(\cdot) \equiv X_n^{M=\infty}(\cdot) = X_n^{M_n}(\cdot) + o_P(1)$ and $Y(\cdot) \equiv Y^{M=\infty}(\cdot) = Y^{M_n}(\cdot) + o_P(1)$, leading to $\rho_{BL_1}(X_n(\cdot), Y(\cdot)) = o(1)$. Let

$$\hat{\mathcal{L}}_n(\cdot) = \int_{n^\gamma(C-\theta_0), n^\gamma(\phi(\theta_0+n^{-\gamma}h)-\phi_0)\le\cdot} \hat{p}_n(h)\, dh = X_{n,1}(\cdot)/X_{n,2}$$

and

$$\mathcal{L}_\infty(\cdot) = \int_{T_C(\theta_0),\phi'_{\theta_0}(h)\le\cdot} p_\infty(h)\, dh = Y_1(\cdot)/Y_2.$$

Then by CMT, $\hat{\mathcal{L}}_n(\cdot) \rightsquigarrow \mathcal{L}_\infty(\cdot)$ in $\ell_\infty\left(R^d\right)$. Finally, note that $F^{-1}(\tau)$ is a continuous functional of $F(\cdot)$ whenever $F(\cdot)$ is strictly increasing at $F^{-1}(\tau)$. Therefore, as long as $\mathcal{L}_\infty(\cdot)$ is almost surely strictly increasing at its τ percentile, the CMT implies that $\hat{h}_\tau \rightsquigarrow \mathcal{J}$.

Finally consider (4), again replace $\hat{h}_\tau^* = \epsilon_n^{-2\gamma}\left(\hat{\phi}_\tau^* - \hat{\phi}_\tau\right)$ with $\hat{h}_\tau^* = \epsilon_n^{-2\gamma}\left(\hat{\phi}_\tau^* - \phi_0\right)$ and write

$$\hat{h}_\tau^* = \inf\left\{x : \int_{h\in\epsilon_n^{-2\gamma}(C-\theta_0), \epsilon_n^{-2\gamma}\left(\phi\left(\theta_0+\epsilon_n^{2\gamma}h\right)-\phi_0\right)\le x} \hat{p}_n^*(h)\, dh \ge \tau\right\}$$

Given M, define the following bootstrap random processes on $\ell_\infty\left(R^d\right)\times R$:

$$X^*_{n,M}(\cdot)=\left\{\int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M},\epsilon_n^{-2\gamma}\left(\phi\left(\theta_0+\epsilon_n^{2\gamma}h\right)-\phi_0\right)\leq\cdot}e^{\mathcal{W}^*_n(h)}dh,\int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}}e^{\mathcal{W}^*_n(h)}dh\right\}$$

and

$$Y^*_{n,M}(\cdot)=\left\{\int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M},\epsilon_n^{-2\gamma}\left(\phi\left(\theta_0+\epsilon_n^{2\gamma}h\right)-\phi_0\right)\leq\cdot}e^{\mathcal{W}_\infty(h)}dh,\int_{\epsilon_n^{-2\gamma}(C-\theta_0)\cap\mathcal{M}}e^{\mathcal{W}_\infty(h)}dh\right\}$$

As before, $\rho_{BL_1}\left(X^*_{n,M}(\cdot),Y^*_{n,M}(\cdot)\right)=o_P(1)$, and

$$\sup_x|Y^*_{n,M}(x)-Y^M(x)|\to 0\quad\text{almost surely}\Longrightarrow\rho_{BL_1}\left(Y^*_{n,M}(\cdot),Y^M(\cdot)\right)=o(1)\,.$$

Hence for $M_n\to\infty$ slowly enough, $\rho_{BL_1}\left(X^*_{M_n}(\cdot),Y^M(\cdot)\right)=o_P(1)$. For any $M_n\to\infty$,

$$\left\{\sup_x\int_{\mathcal{M}_n^c,\epsilon_n^{-2\gamma}\left(\phi\left(\theta_0+\epsilon_n^{2\gamma}h\right)-\phi_0\right)\leq x}e^{\mathcal{W}_n(h)}dh,\int_{\mathcal{M}_n^c}e^{\mathcal{W}_n(h)}dh\right\}=o^*_P(1)\,,$$

Therefore, $X^*_n(\cdot)\equiv X^*_{n,M=\infty}(\cdot)=X^*_{n,M_n}(\cdot)+o^*_P(1)$, leading to $\rho_{BL_1}(X^*_n(\cdot),Y(\cdot))=o_P(1)$. Let $\hat{\mathcal{L}}^*_n(\cdot)=\int_{\epsilon_n^{-2\gamma}(C-\theta_0),\epsilon_n^{-2\gamma}\left(\phi\left(\theta_0+\epsilon_n^{-2\gamma}h\right)-\phi_0\right)\leq\cdot}\hat{p}^*_n(h)\,dh=X^*_{n,1}(\cdot)/X^*_{n,2}$. Then by the bootstrap CMT, $\hat{\mathcal{L}}^*_n(\cdot)\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{L}_\infty(\cdot)$ in $\ell_\infty\left(R^d\right)$. Therefore, as long as $\mathcal{L}_\infty(\cdot)$ is almost surely strictly increasing at its τ percentile, the bootstrap CMT again implies that $\hat{h}^*_\tau\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}\mathcal{J}$.

Proof for Theorem 9.1 The proof for first part of the theorem is a simplified version of the proof of Theorem 3.1 in Bugni et al. (2015) and is therefore omitted. We will show consistency of the numerical bootstrap. Note that

$$\begin{aligned}\hat{\mathcal{J}}^*_n&=\inf_{\theta\in\bar\Theta}S\left(\frac{1}{\epsilon_n}\mathcal{Z}^*_n g(\cdot,\theta)\right)-\inf_{\theta\in\bar\Theta}S\left(\frac{1}{\epsilon_n}P_n g(\cdot,\theta)\right)\\&=\inf_{\theta\in\bar\Theta}S\left(\frac{1}{\epsilon_n}\left(P_n+\epsilon_n\hat{\mathcal{G}}^*_n\right)g(\cdot,\theta)\right)-\inf_{\theta\in\bar\Theta}S\left(\frac{1}{\epsilon_n}P_n g(\cdot,\theta)\right)\\\hat{\mathcal{J}}^*_{n,\Theta}&=\inf_{\theta\in\Theta}S\left(\frac{1}{\epsilon_n}\mathcal{Z}^*_n g(\cdot,\theta)\right)-\inf_{\theta\in\Theta}S\left(\frac{1}{\epsilon_n}P_n g(\cdot,\theta)\right)\\&=\inf_{\theta\in\Theta}S\left(\frac{1}{\epsilon_n}\left(P_n+\epsilon_n\hat{\mathcal{G}}^*_n\right)g(\cdot,\theta)\right)-\inf_{\theta\in\Theta}S\left(\frac{1}{\epsilon_n}P_n g(\cdot,\theta)\right)\end{aligned}$$

Using conditions (1),(2), and (3) in combination with Theorem 2.6 in Kosorok (2007), we can show that $\hat{\mathcal{G}}_n^* g(\cdot,\theta) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{G}_0 g(\cdot,\theta)$ uniformly over $\theta \in \Theta$ and thereby also over $\theta \in \bar{\Theta}$ since $\bar{\Theta} \subseteq \Theta$. Let $\ell_n(\theta) = \frac{1}{\epsilon_n} P_n g(\cdot,\theta)$. Note that because $\sqrt{n}\epsilon_n \to \infty$, $\ell_n(\theta) = \frac{1}{\sqrt{n}\epsilon_n}\sqrt{n}\left(P_n - P\right) g(\cdot,\theta) + \frac{1}{\epsilon_n} P g(\cdot,\theta) = o_p(1) + \frac{1}{\epsilon_n} P g(\cdot,\theta)$, which implies that $\ell_n(\theta) \xrightarrow{p} \ell(\theta) \equiv \underset{t\searrow 0}{lim} P g(\cdot,\theta)/t$. By the bootstrap continuous mapping theorem (Theorem 10.8 in Kosorok (2007)) applied to the functional $\phi\left(\ell(\theta), \mathcal{G}_0 g(\cdot,\theta)\right) = \underset{\theta\in\bar{\Theta}}{inf} S\left(\ell(\theta) + \mathcal{G}_0 g(\cdot,\theta)\right) - \underset{\theta\in\Theta}{inf} S\left(\ell(\theta) + \mathcal{G}_0 g(\cdot,\theta)\right)$,

$$\begin{aligned}
&\underset{\theta\in\bar{\Theta}}{inf} S\left(\frac{1}{\epsilon_n}\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*\right) g(\cdot,\theta)\right) - \underset{\theta\in\Theta}{inf} S\left(\frac{1}{\epsilon_n}\left(P_n + \epsilon_n \hat{\mathcal{G}}_n^*\right) g(\cdot,\theta)\right) \\
&= \underset{\theta\in\bar{\Theta}}{inf} S\left(\ell_n(\theta) + \hat{\mathcal{G}}_n^* g(\cdot,\theta)\right) - \underset{\theta\in\Theta}{inf} S\left(\ell_n(\theta) + \hat{\mathcal{G}}_n^* g(\cdot,\theta)\right) \\
&\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \underset{\theta\in\bar{\Theta}}{inf} S\left(\ell(\theta) + \mathcal{G}_0 g(\cdot,\theta)\right) - \underset{\theta\in\Theta}{inf} S\left(\ell(\theta) + \mathcal{G}_0 g(\cdot,\theta)\right) \equiv \mathcal{J}_{\bar{\Theta}} - \mathcal{J}_{\Theta}
\end{aligned}$$

Under the null, $\underset{\theta\in\bar{\Theta}}{inf} S\left(P g(\cdot,\theta)\right) = 0$, which implies $\underset{\theta\in\bar{\Theta}}{inf} S\left(\frac{1}{\epsilon_n} P_n g(\cdot,\theta)\right) \xrightarrow{p} \underset{\theta\in\bar{\Theta}}{inf} S\left(\ell(\theta)\right) = 0$. Additionally, since $\underset{\theta\in\Theta}{inf} S\left(P g(\cdot,\theta)\right) \le \underset{\theta\in\bar{\Theta}}{inf} S\left(P g(\cdot,\theta)\right)$, it follows that $\underset{\theta\in\Theta}{inf} S\left(\frac{1}{\epsilon_n} P_n g(\cdot,\theta)\right) \xrightarrow{p} \underset{\theta\in\Theta}{inf} S\left(\ell(\theta)\right) = 0$. By Slutsky's Theorem, $\bar{\mathcal{J}}_n^* \equiv \hat{\mathcal{J}}_n^* - \hat{\mathcal{J}}_{n,\Theta}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \mathcal{J}_{\bar{\Theta}} - \mathcal{J}_{\Theta}$. ■

References

Andrews, Donald WK and Gustavo Soares, "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 2010, *78* (1), 119–157.

Babu, G Jogesh and Kesar Singh, "Edgeworth expansions for sampling without replacement from finite populations," *Journal of Multivariate Analysis*, 1985, *17* (3), 261–278.

Bertail, Patrice, "Second-order properties of an extrapolated bootstrap without replacement under weak assumptions," *Bernoulli*, 1997, *3* (2), 149–179.

Bickel, Peter J and Anat Sakov, "On the choice of m in the m out of n bootstrap and confidence bounds for extrema," *Statistica Sinica*, 2008, pp. 967–985.

Bugni, Federico A, Ivan A Canay, and Xiaoxia Shi, "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 2015, *185* (1), 259–282.

_ , _ , and _ , "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 2017, *8* (1), 1–38.

Chernozhukov, Victor and Han Hong, "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

_ and Iván Fernández-Val, "Subsampling inference on quantile regression processes," *Sankhyā: The Indian Journal of Statistics*, 2005, pp. 253–276.

Geyer, Charles J, "On the asymptotics of constrained M-estimation," *The Annals of Statistics*, 1994, pp. 1993–2010.

Hansen, Peter Reinhard, "A test for superior predictive ability," *Journal of Business & Economic Statistics*, 2005, *23* (4).

Jun, Sung Jae, Joris Pinkse, and Yuanyuan Wan, "Classical Laplace estimation for-consistent estimators: Improved convergence rates and rate-adaptive inference," *Journal of Econometrics*, 2015, *187* (1), 201–216.

Kim, Jeankyung and David Pollard, "Cube root asymptotics," *The Annals of Statistics*, 1990, pp. 191–219.

Knight, Keith, "Epi-convergence in distribution and stochastic equisemicontinuity," *Unpublished manuscript*, 1999, *37.*

Kosorok, Michael R, *Introduction to empirical processes and semiparametric inference*, Springer, 2007.

Newey, W. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, North Holland, 1994, pp. 2113–2241.

Politis, D., J. Romano, and M. Wolf, *Subsampling*, Springer Series in Statistics, 1999.

Romano, Joseph P and Azeem M Shaikh, "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 2012, *40* (6), 2798–2822.

Römisch, Werner, "Delta method, infinite dimensional," *Encyclopedia of Statistical Sciences*, 2005.

Sherman, Robert P., "The limiting distribution of the maximum rank correlation estimator," *Econometrica*, 1993, *61* (1), 123–137.

van der Vaart, AW and Jon Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996.

Wolak, F., "Testing Inequality Constraints in Linear Econometric Models," *Journal of Econometrics*, 1989, *41*, 205–235.