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

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

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

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

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 $\epsilon_n \downarrow 0$,

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

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

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

To see why the numerical bootstrap might work, note that

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

In the above, we rewrite the second term as

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

[1] $X_n \rightsquigarrow X$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in \mathrm{BL}_1} |E^* f(X_n) - Ef(X)| \to 0$ 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 $h \in \mathbb{D}_0$,

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

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 $\mu_0 + t_n h_n \in \mathbb{R}^d$.

[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 $z_i^*, i = 1, \dots, n$,

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

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*,

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

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

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 $\hat{h}_n = n^{\gamma}\left(\hat{\theta}_n - \theta_0\right)$,

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

Under the stated conditions, $n^{2\gamma} P g\left(\cdot;\theta_0 + \frac{h}{n^{\gamma}}\right) \to -\frac{1}{2}h' H h$, 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 $\mathcal{Z}_n^* g(\cdot,\theta) = (\mathcal{Z}_n^* - P) g(\cdot,\theta) + P g(\cdot,\theta)$, 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

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

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

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

The second term $\epsilon_n^{-4\gamma} P g\left(\cdot;\theta_0 + \epsilon_n^{2\gamma}h\right) \to -\frac{1}{2}h' H h$. 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,

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

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 $\hat{\theta}_n^* \in C$,

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

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

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 $\hat{Q}_n(\theta)$:

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

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

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

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 $\theta_0 \in \operatorname{int}(\Theta)$ 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 $X_1, \dots, X_n$ be i.i.d. random variables drawn from $N(\theta_0, 2)$. Define $\hat{\theta}_n = \arg\max_\theta \frac{1}{n}\sum_{i=1}^n \mathbf{1}(\theta - 5 \le X_i \le \theta + 5)$, 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 (1.861)	0.729 (1.872)	0.724 (1.861)	0.722 (1.878)	0.724 (1.864)	0.822 (3.068)	0.817 (3.072)	0.820 (3.077)	0.824 (3.083)	0.823 (3.077)
$n = 500$	0.783 (1.299)	0.796 (1.285)	0.777 (1.284)	0.781 (1.287)	0.798 (1.289)	0.922 (1.793)	0.946 (1.784)	0.927 (1.789)	0.914 (1.786)	0.930 (1.786)
$n = 1000$	0.783 (1.048)	0.815 (1.049)	0.792 (1.046)	0.812 (1.053)	0.808 (1.048)	0.950 (1.397)	0.935 (1.396)	0.938 (1.397)	0.947 (1.396)	0.941 (1.396)
$n = 5000$	0.849 (0.636)	0.827 (0.634)	0.837 (0.634)	0.856 (0.633)	0.864 (0.633)	0.955 (0.772)	0.945 (0.771)	0.949 (0.770)	0.958 (0.771)	0.956 (0.771)
$n = 10{,}000$	0.879 (0.506)	0.887 (0.506)	0.865 (0.506)	0.873 (0.506)	0.879 (0.507)	0.962 (0.595)	0.954 (0.595)	0.945 (0.595)	0.957 (0.595)	0.953 (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 (1.823)	0.729 (1.828)	0.736 (1.817)	0.726 (1.839)	0.730 (1.827)	0.835 (2.944)	0.830 (2.947)	0.832 (2.948)	0.833 (2.947)	0.842 (2.950)
$n = 500$	0.715 (1.186)	0.722 (1.179)	0.734 (1.176)	0.727 (1.180)	0.728 (1.182)	0.945 (1.793)	0.961 (1.789)	0.942 (1.790)	0.940 (1.792)	0.947 (1.789)
$n = 1000$	0.694 (0.954)	0.721 (0.958)	0.705 (0.955)	0.715 (0.957)	0.733 (0.959)	0.969 (1.433)	0.963 (1.433)	0.964 (1.432)	0.967 (1.433)	0.963 (1.434)
$n = 5000$	0.750 (0.592)	0.724 (0.593)	0.735 (0.591)	0.771 (0.590)	0.761 (0.591)	0.982 (0.839)	0.982 (0.839)	0.974 (0.839)	0.981 (0.839)	0.978 (0.838)
$n = 10{,}000$	0.791 (0.481)	0.793 (0.481)	0.774 (0.481)	0.775 (0.481)	0.811 (0.481)	0.983 (0.662)	0.986 (0.662)	0.988 (0.662)	0.982 (0.662)	0.978 (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