The Annals of Statistics
2020, Vol. 48, No. 1, 274–299
https://doi.org/10.1214/18-AOS1803

BOOTSTRAP CONFIDENCE REGIONS BASED ON M-ESTIMATORS UNDER NONSTANDARD CONDITIONS

BY STEPHEN M. S. LEE[1] AND PUYUDI YANG[2]

[1]*Department of Statistics and Actuarial Science, University of Hong Kong, smslee@hku.hk*
[2]*Department of Statistics, University of California, Davis, pydyang@ucdavis.edu*

Suppose that a confidence region is desired for a subvector θ of a multidimensional parameter $\xi = (\theta, \psi)$, based on an M-estimator $\hat{\xi}_n = (\hat{\theta}_n, \hat{\psi}_n)$ calculated from a random sample of size n. Under nonstandard conditions $\hat{\xi}_n$ often converges at a nonregular rate r_n, in which case consistent estimation of the distribution of $r_n(\hat{\theta}_n - \theta)$, a pivot commonly chosen for confidence region construction, is most conveniently effected by the m out of n bootstrap. The above choice of pivot has three drawbacks: (i) the shape of the region is either subjectively prescribed or controlled by a computationally intensive depth function; (ii) the region is not transformation equivariant; (iii) $\hat{\xi}_n$ may not be uniquely defined. To resolve the above difficulties, we propose a one-dimensional pivot derived from the criterion function, and prove that its distribution can be consistently estimated by the m out of n bootstrap, or by a modified version of the perturbation bootstrap. This leads to a new method for constructing confidence regions which are transformation equivariant and have shapes driven solely by the criterion function. A subsampling procedure is proposed for selecting m in practice. Empirical performance of the new method is illustrated with examples drawn from different nonstandard M-estimation settings. Extension of our theory to row-wise independent triangular arrays is also explored.

1. Introduction. Let $\mathcal{X}_n = (X_1, \ldots, X_n)$ be a random sample of size n drawn from an unknown distribution F. The parameter of interest $\xi_F \in \mathbb{R}^p$ is assumed to be the unique value of ξ that maximises $\mathbb{E}_F[b_{\xi,\eta_F}(X_1)]$, for some real-valued function $b_{\xi,\eta}$ indexed by $(\xi, \eta) \in \mathbb{R}^p \times \mathbb{R}^q$, where η_F denotes the true value of a nuisance parameter. Given a consistent estimator $\hat{\eta}_n$ of η_F, define an M-estimator $\hat{\xi}_n$ of ξ_F by the value of ξ which maximises, at least approximately, the criterion function $\hat{B}_n(\xi) = n^{-1}\sum_{i=1}^n b_{\xi,\hat{\eta}_n}(X_i)$. We are interested in constructing a level $1-\alpha$ confidence region for a d-dimensional subvector θ_F of ξ_F, for $d \le p$, without assuming the usual regularity conditions on $b_{\xi,\eta}$. Without loss of generality, we write $\xi_F = (\theta_F, \psi_F)$ and $\hat{\xi}_n = (\hat{\theta}_n, \hat{\psi}_n)$, for $(p-d)$-dimensional vectors ψ_F and $\hat{\psi}_n$.

Bootstrap confidence regions for θ_F have been studied mainly under regularity conditions which guarantee asymptotic normality of the pivot $n^{1/2}(\hat{\theta}_n - \theta_F)$ or its Studentized form. Consistency of the conventional, n out of n, bootstrap in this context has been established by [20] in the absence of nuisance parameters and by [9] in the presence of a Banach space-valued nuisance parameter. Multidimensionality of the bootstrapped pivot calls for special devices for fixing the final confidence region. One common approach is to impose on the region an ellipsoidal shape by calibrating the norm of the pivot [13], or a rectangular shape by referring to some properly chosen univariate quantiles along each of the d dimensions [3, 9]. Alternatively, the bootstrapped pivot can be calibrated by a data depth function to yield a confidence region more adaptive to the observed data; see, for example, [39]. Wei and Lee [38] investigate second-order properties of depth-based bootstrap regions under regularity

Received August 2017; revised December 2018.
MSC2010 subject classifications. 62G15, 62G09.
Key words and phrases. Confidence region, m out of n bootstrap, M-estimator, pivot, subsampling.

conditions. In a similar spirit, [30] introduce the notion of generalized spatial quantiles to handle multidimensionality of confidence region pivots. In the special case where $\hat{\theta}_n$ solves a multivariate estimating function that depends possibly on an infinite-dimensional nuisance parameter, [16] propose a bootstrap method for estimating the distribution of a univariate pivot formed by the empirical likelihood. Resampling methods other than the bootstrap have also been proposed; see, for example, [17] for a simulation procedure which estimates the distribution of $n^{1/2}(\hat{\xi}_n - \xi_F)$ by randomly perturbing the criterion function.

Without assuming regularity conditions on $b_{\xi,\eta}$, nonstandard convergence rates r_n and non-Gaussian weak limits of $r_n(\hat{\xi}_n - \xi_F)$ have been established using empirical process methods; see, for example, [37], Section 3.2, for the case where η is absent and [18] for the case where $r_n = n^{1/3}$ in the presence of η. Assuming convexity of $b_{\xi,\eta}$ in ξ, asymptotic Gaussianity of a special structure and absence of nuisance parameters, [5] prove that the distribution of $r_n(\hat{\xi}_n - \xi_F)$ can be consistently estimated by the m out of n bootstrap. Lee and Pun [23] establish m out of n bootstrap consistency under weaker conditions on $b_{\xi,\eta}$ and the assumption that $\hat{\eta}_n$ converges to η_F at a rate faster than r_n. To our knowledge, there has not been work done on m out of n bootstrap confidence regions constructed using pivots other than $r_n(\hat{\theta}_n - \theta_F)$ or its Studentized version under nonstandard conditions. The usual difficulties inherent in multidimensional pivots remain unresolved in this context.

We propose in Section 2 a new m out of n bootstrap procedure for constructing confidence regions for θ_F based on a univariate pivot expressed in terms of the criterion function $\hat{B}_n$. Section 3 establishes consistency of our proposed confidence region under nonstandard conditions on $b_{\xi,\eta}$, and shows that the same also holds for a modified perturbation bootstrap procedure. Extension of our theory to row-wise independent triangular arrays is explored in Section 4. Section 5 illustrates our results with a number of examples, and compares the finite-sample performance of the proposed procedure with that based on the usual d-variate pivot $r_n(\hat{\theta}_n - \theta_F)$ in each of the examples. A subsampling procedure is introduced in Section 6 for selecting m, supplemented with theoretical justification and empirical evidence. The issue of unknown convergence rate is briefly discussed in Section 7. Section 8 concludes our findings. All proofs of our theorems, corollaries and propositions are provided in the Supplementary Material [22].

2. Method. Suppose that $(\xi_F, \eta_F) \in \Xi \times \mathcal{H} \subset \mathbb{R}^p \times \mathbb{R}^q$. Let $\Xi_1 \subset \mathbb{R}^d$ be the projection of Ξ onto its first d dimensions and, for each $\theta \in \Xi_1$, $\Xi_2(\theta) = \{\psi \in \mathbb{R}^{p-d} : (\theta, \psi) \in \Xi\}$. Define $\Xi_F^\dagger = \{a(\xi - \xi_F) : a \geq 0, \xi \in \Xi\}$ and $\Xi_{F,2}^\dagger = \{u \in \mathbb{R}^{p-d} : (0, u) \in \Xi_F^\dagger\}$. Note that $\Xi_F^\dagger = \mathbb{R}^p$ if ξ_F is an interior point in Ξ, and often assumes the form of a convex cone if ξ_F lies on the boundary of an order-restricted parameter space Ξ.

For a given consistent estimator $\hat{\eta}_n$ of $\eta_F \in \mathcal{H}$, the M-estimator $\hat{\xi}_n = (\hat{\theta}_n, \hat{\psi}_n)$ of ξ_F is defined to be any value of $\xi \in \Xi$ which maximises the criterion function $\hat{B}_n(\xi) = n^{-1}\sum_{i=1}^n b_{\xi,\hat{\eta}_n}(X_i)$ approximately in the sense that $\hat{B}_n(\hat{\xi}_n) \geq \sup_{\xi\in\Xi} \hat{B}_n(\xi) - o_p(r_n^{-\nu})$, for some sequence $r_n^\nu \uparrow \infty$ to be specified below in (G1) and (G2). Note that we do not require that $n^{-1}\sum_{i=1}^n b_{\xi,\eta}(X_i)$ be maximised at $(\xi, \eta) = (\hat{\xi}_n, \hat{\eta}_n)$. In cases where $\hat{B}_n(\xi)$ admits multiple global maximisers, $\hat{\xi}_n$ may be defined to be an approximate value of any one of the maximisers. Similarly, for any fixed $\theta \in \Xi_1$, a profile M-estimator $\tilde{\psi}_n(\theta)$ of ψ is required to satisfy $\tilde{\psi}_n(\theta) \in \Xi_2(\theta)$ and $\hat{B}_n(\theta, \tilde{\psi}_n(\theta)) \geq \sup_{\psi\in\Xi_2(\theta)} \hat{B}_n(\theta, \psi) - o_p(r_n^{-\nu})$.

Define, for $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$, $g_{s,t} = b_{\xi_F+s,\eta_F+t} - b_{\xi_F,\eta_F+t}$. We assume, for some fixed $\nu > 0$ and some sequence of positive constants $r_n \uparrow \infty$, the following conditions on the random process $s \mapsto g_{s,0}(X_1)$:

(G1) there exists a nonzero function Λ on $\Xi_F^\dagger$ such that for every compact $\mathcal{K} \subset \Xi_F^\dagger$, Λ is uniformly continuous on $\mathcal{K}$ and $\lim_{\epsilon\to 0} \epsilon^{-\nu}\mathbb{E}_F[g_{\epsilon s,0}(X_1)] = \Lambda(s)$, uniformly in $s \in \mathcal{K}$;

(G2) for any $s_1, s_2 \in \Xi_F^\dagger$, $\Sigma(s_1, s_2) = \lim_{n\to\infty} n^{-1} r_n^{2\nu} \mathbb{E}_F[g_{s_1/r_n,0}(X_1) g_{s_2/r_n,0}(X_1)]$ exists, with $\sup_{s_1,s_2\in\Xi_F^\dagger} |\Sigma(s_1, s_2)| > 0$.

The power ν in (G1) characterises the index of regular variation of $\xi \mapsto \mathbb{E}_F[b_{\xi,\eta_F}]$ at its maximum when $\xi = \xi_F$ [29]. This condition, together with (G2), identifies the sharpest rate r_n that balances the pointwise mean and standard deviation of the empirical process $s \mapsto n^{-1} \sum_{i=1}^n g_{s/r_n,0}(X_i)$. To see this, suppose that for some $\varpi \in (0, 2\nu)$, $\mathbb{E}_F[g_{\epsilon s_1,0}(X_1) g_{\epsilon s_2,0}(X_1)]$ has an order ϵ^ϖ as $\epsilon \to 0$. It then follows by (G1) and (G2) that $n^{-1} \sum_{i=1}^n g_{s/r_n,0}(X_i)$ has pointwise mean and standard deviation both of order $n^{-\nu/(2\nu-\varpi)}$, with $r_n \propto n^{1/(2\nu-\varpi)}$. It can be shown, under further conditions to be stated in Section 3, that r_n indeed gives the precise convergence rate of $\hat{\xi}_n$. The conditions (G1) and (G2) also ensure that the mean and covariance functions of the scaled process $s \mapsto n^{-1} r_n^\nu \sum_{i=1}^n g_{s/r_n,0}(X_i)$ both converge to nontrivial limits.

Standard regularity assumptions often impose quadratic structures on Λ and Σ such that $\Lambda(s) = s^\top \Lambda_0 s$ and $\Sigma(s_1, s_2) = s_1^\top \Sigma_0 s_2$, for some nonsingular matrices Λ_0 and Σ_0. This amounts to the special case $\nu = \varpi = 2$, for which (G2) implies a standard convergence rate $r_n \propto n^{1/2}$. Our nonstandard setup removes the above strong assumptions, allows for greater flexibility in the forms of Λ and Σ, and requires s, s_1, s_2 to range only over a problem-specific region $\Xi_F^\dagger \subset \mathbb{R}^p$. Note that in some practical situations, r_n may be analytically intractable and needs to be estimated from the data $\mathcal{X}_n$, a point which we shall return to in Section 7.

Define, for $\theta \in \Xi_1$, a univariate confidence region pivot by

$$\Pi_n(\theta) = r_n^\nu \{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\theta, \tilde{\psi}_n(\theta))\}.$$

We describe below a bootstrap procedure for estimating the distribution of $\Pi_n(\theta_F)$ whereby a confidence region for θ_F can be constructed. Let $\mathcal{X}_m^* = (X_1^*, \ldots, X_m^*)$ denote a random sample drawn with replacement from $\mathcal{X}_n$, with $m = m_n \to \infty$ as $n \to \infty$. Given any nuisance parameter estimator $\hat{\eta}_n^*$, possibly but not necessarily depending on $\mathcal{X}_m^*$, define $\hat{B}_n^*(\xi) = m^{-1} \sum_{i=1}^m b_{\xi,\hat{\eta}_n^*}(X_i^*)$, for any $\xi \in \Xi$. Analogous to $\hat{\xi}_n$ and $\tilde{\psi}_n(\theta)$, we define $\hat{\xi}_n^* = (\hat{\theta}_n^*, \hat{\psi}_n^*)$ and $\tilde{\psi}_n^*(\theta)$ to be the M-estimator and profile M-estimator obtained by maximising approximately $\hat{B}_n^*(\xi)$ and, for a fixed $\theta \in \Xi_1$, $\hat{B}_n^*(\theta, \psi)$, respectively. The m out of n bootstrap analogue of $\Pi_n(\theta)$ is given by $\Pi_n^*(\theta) = r_m^\nu \{\hat{B}_n^*(\hat{\xi}_n^*) - \hat{B}_n^*(\theta, \tilde{\psi}_n^*(\theta))\}$. Denote by $\hat{G}_n$ the conditional distribution of $\Pi_n^*(\hat{\theta}_n)$ given $\mathcal{X}_n$. Our proposed m out of n bootstrap confidence region of nominal level $1 - \alpha$ is defined to be $\mathcal{R}_{n,1-\alpha} = \{\theta \in \Xi_1 : \Pi_n(\theta) \le \hat{G}_n^{-1}(1 - \alpha)\}$.

REMARK 2.1. Based on the d-dimensional pivot $r_n(\hat{\theta}_n - \theta)$, a more conventional approach constructs an m out of n bootstrap confidence region to be $\mathcal{R} = \{\theta \in \Xi_1 : \mathrm{pr}(\hat{\mathcal{D}}(r_m(\hat{\theta}_n^* - \hat{\theta}_n)) \le \hat{\mathcal{D}}(r_n(\hat{\theta}_n - \theta))|\mathcal{X}_n) \le 1 - \alpha\}$, where the real-valued function $\hat{\mathcal{D}}$ is either chosen to endow $\mathcal{R}$ with a pre-determined, typically elliptical or rectangular, shape or derived from a data-driven depth function or spatial quantile. Given any smooth injection $\varkappa : \Xi_1 \to \mathbb{R}^d$, the same procedure yields the confidence region $\{\varkappa_0 \in \varkappa(\Xi_1) : \mathrm{pr}(\hat{\mathcal{D}}(r_m(\varkappa(\hat{\theta}_n^*) - \varkappa(\hat{\theta}_n))) \le \hat{\mathcal{D}}(r_n(\varkappa(\hat{\theta}_n) - \varkappa_0))|\mathcal{X}_n) \le 1 - \alpha\}$ for the parameter $\varkappa(\theta)$, which is in general distinct from the $\varkappa$-transformed region $\varkappa(\mathcal{R})$. Thus, the conventional m out of n bootstrap confidence region is not transformation equivariant. Our proposed confidence region $\mathcal{R}_{n,1-\alpha}$, by contrast, has an entirely data-driven shape and is transformation equivariant.

REMARK 2.2. Under classical regularity conditions (e.g., [36], Section 5.6) and assuming absence of η_F, it can be shown that

$$n^{1/2}(\hat{\theta}_n - \theta_F) = -(A_{\theta\theta}, A_{\theta\psi}) n^{-1/2} \sum_{i=1}^n \dot{b}_{\xi_F}(X_i) + O_p(n^{-1/2}),$$

where $\dot{b}_{\xi_F} = (\partial/\partial\xi)b_\xi|_{\xi=\xi_F}$, $\ddot{b}_{\xi_F} = (\partial^2/\partial\xi\partial\xi^\top)b_\xi|_{\xi=\xi_F}$ and $\{\mathbb{E}_F[\ddot{b}_{\xi_F}(X_1)]\}^{-1}$ is partitioned as $((A_{\theta\theta}, A_{\theta\psi})^\top, (A_{\psi\theta}, A_{\psi\psi})^\top)$, with $A_{\theta\theta}$ having dimension $d \times d$. Similar arguments can be employed to establish that

$$\Pi_n(\theta_F) = -(n/2)(\hat{\theta}_n - \theta_F)^\top A_{\theta\theta}^{-1}(\hat{\theta}_n - \theta_F) + O_p\big(n^{-1/2}\big).$$

It follows that if we restrict $\mathcal{C}_{n,1-\alpha}$ to be an ellipsoid of the form $\{v \in \mathbb{R}^d : -v^\top \hat{A}_{\theta\theta}^{-1} v \le 2\hat{k}\}$, for some consistent estimator $\hat{A}_{\theta\theta}$ of $A_{\theta\theta}$ and some $\hat{k}$ calibrated to satisfy $\mathrm{pr}(m^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \in \mathcal{C}_{n,1-\alpha}|\mathcal{X}_n) = 1-\alpha$, then the two regions $\mathcal{R}_{n,1-\alpha}$ and $\{\theta \in \Xi_1 : n^{1/2}(\hat{\theta}_n - \theta) \in \mathcal{C}_{n,1-\alpha}\}$ are equivalent to first order. They are in general different from the conventional Wald-type region built upon the squared norm of the Studentized M-estimator, unless $\mathbb{E}_F[\dot{b}_{\xi_F}(X_1)\dot{b}_{\xi_F}(X_1)^\top] = -\mathbb{E}_F[\ddot{b}_{\xi_F}(X_1)]$, in which case $\mathrm{Var}(n^{1/2}\hat{\theta}_n)$ converges to $-A_{\theta\theta}$ and $2\Pi_n(\theta_F)$ becomes asymptotically χ_d^2.

3. Theory. Denote by $|\cdot|$ the Euclidean norm. For any $\delta, \epsilon > 0$, define function classes

$$\mathcal{G}_\delta = \big\{g_{s,t} : |(s,t)| \le \delta, (\xi_F + s, \eta_F + t) \in \Xi \times \mathcal{H}\big\},$$

$$\mathcal{G}_\delta(\epsilon) = \big\{g_{s_1,t_1} - g_{s_2,t_2} : |(s_1,t_1) - (s_2,t_2)| \le \epsilon, g_{s_1,t_1}, g_{s_2,t_2} \in \mathcal{G}_\delta\big\},$$

and G_δ to be an envelope function for $\mathcal{G}_\delta$. For any probability measure Q and any square-integrable function f with respect to Q, define $\|f\|_{Q,2} = (\int |f|^2 dQ)^{1/2}$. Denote by $N(\epsilon, \mathcal{F}, Q)$ the covering number of a function class $\mathcal{F}$ with respect to a $\|\cdot\|_{Q,2}$-radius $\epsilon > 0$. For any arbitrary set $\mathcal{T}$, define $\ell^\infty(\mathcal{T})$ to be the space of all bounded functions from $\mathcal{T}$ to $\mathbb{R}$, equipped with the sup-norm. Define, for $(\xi, \eta) \in \Xi \times \mathcal{H}$, $\Psi(\xi, \eta) = \mathbb{E}_F[b_{\xi,\eta}(X_1)]$. To establish the weak limit of $\Pi_n(\theta_F)$, we assume the following conditions:

(A1) $\hat{\eta}_n = \eta_F + o_p(r_n^{-1})$;

(A2) $\hat{\xi}_n = O_p(1)$ and $\tilde{\psi}_n(\theta_F) = O_p(1)$ satisfy

$$\hat{B}_n(\hat{\xi}_n) \ge \sup_{\xi\in\Xi} \hat{B}_n(\xi) - o_p\big(r_n^{-\nu}\big) \quad \text{and}$$

$$\hat{B}_n\big(\theta_F, \tilde{\psi}_n(\theta_F)\big) \ge \sup_{\psi\in\Xi_2(\theta_F)} \hat{B}_n(\theta_F, \psi) - o_p\big(r_n^{-\nu}\big);$$

(A3) $\Psi(\xi, \eta_F)$ is uniquely maximised at $\xi = \xi_F$, and for some $C_0 > 0$, $\Psi(\xi_F + s, \eta_F) - \Psi(\xi_F, \eta_F) \le -C_0|s|^\nu$ as $s \to 0$;

(A4) for every compact $\mathcal{K} \subset \Xi$, $|\Psi(\xi, \eta_F + t) - \Psi(\xi, \eta_F)| \to 0$ as $t \to 0$, uniformly in $\xi \in \mathcal{K}$, and there exist $K_0, \nu_1, \nu_2 > 0$ with $\nu_1 \le \nu \le \nu_1 + \nu_2$ such that

$$\big|\Psi(\xi_F + s, \eta_F + t) - \Psi(\xi_F, \eta_F + t) - \Psi(\xi_F + s, \eta_F) + \Psi(\xi_F, \eta_F)\big| \le K_0|s|^{\nu_1}|t|^{\nu_2}$$

as $s, t \to 0$;

(A5) for any $r > 0$, $\{b_{\xi,\eta} : |(\xi,\eta)| \le r,\ (\xi,\eta) \in \Xi \times \mathcal{H}\}$ is Glivenko–Cantelli;

(A6) there exist $\tau_0 < 3/4$ and $\alpha_0 < \nu$ such that for all $c, \epsilon > 0$,

$$r_n^{2\nu} n^{-1} \mathbb{E}\big[G_{c/r_n}(X_1)^2; G_{c/r_n}(X_1) > \epsilon n^{\tau_0} r_n^{-\nu}\big] \to 0,$$

and that $\varphi(\delta)^{-2}\mathbb{E}_F[G_\delta(X_1)^2]$ is bounded as $\delta \downarrow 0$, for some nonnegative function φ satisfying, as $\delta \downarrow 0$, $\varphi(a\delta) \le a^{\alpha_0}\varphi(\delta)$ for all $a \ge 1$ and $r_n^\nu n^{-1/2}\varphi(1/r_n) = O(1)$;

(A7) there exists $\delta_0 > 0$ such that

$$\int_0^1 \sup_{\delta<\delta_0} \sup_Q \sqrt{\log N\big(\epsilon\|G_\delta\|_{Q,2}, \mathcal{G}_\delta, Q\big)}\, d\epsilon < \infty,$$

where the second supremum is taken over all finitely discrete probability measures Q, and for any $c > 0$ and any sequence $\delta_n \downarrow 0$,

$$r_n^{2\nu} n^{-1} \sup\big\{\mathbb{E}_F\big[f(X_1)^2\big] : f \in \mathcal{G}_{c/r_n}(\delta_n/r_n)\big\} \to 0;$$

(A8) for any $\delta, \epsilon > 0$, the classes $\mathcal{G}_\delta$, $\mathcal{G}_\delta(\epsilon)$ and $\mathcal{G}_\delta(\epsilon)^2$ are measurable as defined in [37], p. 110.

The condition (A1) requires that η_F be consistently estimated by $\hat{\eta}_n$ at a rate faster than r_n. In a nonstandard setup one often encounters a convergence rate r_n slower than $n^{1/2}$, in which case (A1) typically holds if $\hat{\eta}_n$ can be obtained as solution to a system of q estimating equations subject to regularity conditions that guarantee $n^{1/2}$-consistency. Note that our setup allows the maximiser $\xi_F = (\theta_F, \psi_F)$ to accommodate another nuisance parameter subvector ψ_F, of which estimation is not required to be consistent at a rate faster than r_n. The condition (A2) requires that the M-estimator $\hat{\xi}_n$ and profile M-estimator $\tilde{\psi}_n(\theta_F)$ maximise the criterion function to order $o_p(r_n^{-\nu})$. The shape of the function $\xi \mapsto \Psi(\xi, \eta_F)$ near its maximum value is characterised by (A3). The condition (A4) requires that $\Psi(\xi, \eta)$ be smooth near $\eta = \eta_F$ in a certain sense, and is redundant if η_F is either known or absent. Conditions (A5)–(A8) concern properties of classes of functions $b_{\xi,\eta}$ and $g_{s,t}$: (A5) typically holds under an entropy condition on the function class and a finite mean of the envelope (e.g., [37], Theorem 2.4.3); (A6) requires the distribution of $r_n^{2\nu} n^{-1} G_{c/r_n}(X_1)^2$ to possess a bounded mean and a sufficiently light tail; (A7) imposes a uniform entropy condition on $\mathcal{G}_\delta$ and requires the $\|\cdot\|_{F,2}$ distance between $g_{(s_1,t_1)/r_n}, g_{(s_2,t_2)/r_n} \in \mathcal{G}_{c/r_n}$ to shrink at a uniform rate of order $o(n^{1/2} r_n^{-\nu})$ whenever $|(s_1, t_1) - (s_2, t_2)| \to 0$; (A8) facilitates application of Fubini's theorem to symmetrized processes and is satisfied, for example, by image admissible Suslin or pointwise measurable function classes. Note that (A5)–(A8) are essential to the proof of weak convergences of $\Pi_n(\theta_F)$ and $\Pi_n^*(\hat{\theta}_n)$. They are satisfied, in particular, by all VC-classes and VC-subgraph classes. We refer to [37], Chapter 1.3, for an extended definition of weak convergence to get around the problem of nonmeasurability.

Denote by $d_H(\cdot, \cdot)$ the Hausdorff distance between sets in Euclidean spaces. For in-probability weak convergence of $\Pi_n^*(\hat{\theta}_n)$, we assume further:

(B1) $m = o(n)$ and $m \to \infty$;
(B2) $\hat{\eta}_n^* = \hat{\eta}_n + o_p(r_m^{-1})$;
(B3) $\hat{\xi}_n^* = O_p(1)$ and $\tilde{\psi}_n^*(\hat{\theta}_n) = O_p(1)$ satisfy

$$\hat{B}_n^*(\hat{\xi}_n^*) \geq \sup_{\xi \in \Xi} \hat{B}_n^*(\xi) - o_p(r_m^{-\nu}) \quad \text{and}$$

$$\hat{B}_n^*(\hat{\theta}_n, \tilde{\psi}_n^*(\hat{\theta}_n)) \geq \sup_{\psi \in \Xi_2(\hat{\theta}_n)} \hat{B}_n^*(\hat{\theta}_n, \psi) - o_p(r_m^{-\nu});$$

(B4) there exists $C_H > 0$ such that $d_H(\Xi_2(\theta), \Xi_2(\theta_F)) \leq C_H |\theta - \theta_F|$ for all $\theta \in \Xi_1$ sufficiently close to θ_F.

Conditions (B2) and (B3) are m out of n bootstrap analogues of (A1) and (A2), respectively. Note that (B2) holds trivially if we set, for example, $\hat{\eta}_n^* = \hat{\eta}_n$. The condition (B4) requires Lipschitz continuity of $\Xi_2(\theta)$ at $\theta = \theta_F$. It holds if, for example, $\Xi = \mathbb{R}^p$ or an order-restricted subset of $\mathbb{R}^p$, and is redundant if $p = d$ so that $\theta_F = \xi_F$.

Our conditions generalise those assumed by [23] in their study of m out of n bootstrap consistency in a number of ways. The smoothness conditions (A3) and (A4) do not require that $\Psi(\xi, \eta)$ be second-order continuously differentiable with respect to (ξ, η). Assuming the latter, as has been required by [23], necessarily implies $\nu = 2$ and $\nu_1 = \nu_2 = 1$. The weaker conditions (A3) and (A4) extend our applications to cover, for example, L_1 regression problems under an error distribution with an infinite peak at the origin. Unlike [23], we do not assume that the function Σ defined in (G2) satisfies $\Sigma(s_1, s_1) + \Sigma(s_2, s_2) - 2\Sigma(s_1, s_2) \neq 0$ for any $s_1 \neq s_2$, a condition which is needed if the weak limit of $r_n(\hat{\xi}_n - \xi_F)$ is to be identified with the unique maximiser of a Gaussian process.

We state below our main theorem, the proof of which is given in the Supplementary Material [22].

THEOREM 3.1. *Assume that* (G1), (G2) *and* (A1)–(A8) *hold. Then, for any* $x \in \mathbb{R}$:

(i) $\lim_{n\to\infty} \mathrm{pr}_F(\Pi_n(\theta_F) \le x) = \mathrm{pr}(\sup_{s\in\Xi_F^\dagger} \inf_{u\in\Xi_{F,2}^\dagger} \mathbb{Z}(s,u) \le x)$, *for a Gaussian process* $\mathbb{Z}$ *in* $\ell^\infty(\Xi_F^\dagger \times \Xi_{F,2}^\dagger)$ *which satisfies, for any* $s, s_1, s_2 \in \Xi_F^\dagger$ *and* $u, u_1, u_2 \in \Xi_{F,2}^\dagger$, $\mathbb{E}[\mathbb{Z}(s,u)] = \Lambda(s) - \Lambda(0,u)$ *and*

$$\begin{aligned}&\mathrm{Cov}\big(\mathbb{Z}(s_1,u_1), \mathbb{Z}(s_2,u_2)\big)\\ &\quad = \Sigma(s_1,s_2) + \Sigma\big((0,u_1),(0,u_2)\big) - \Sigma\big(s_1,(0,u_2)\big) - \Sigma\big((0,u_1),s_2\big);\end{aligned}$$

(ii) *with* $\mathbb{Z}$ *specified as in* (i) *and assuming further* (B1)–(B4),

$$\hat{G}_n(x) = \mathrm{pr}\big(\Pi_n^*(\hat{\theta}_n) \le x | \mathcal{X}_n\big) \to \mathrm{pr}\Big(\sup_{s\in\Xi_F^\dagger} \inf_{u\in\Xi_{F,2}^\dagger} \mathbb{Z}(s,u) \le x\Big) \quad \textit{in probability}.$$

Theorem 3.1 shows that $\hat{G}_n$ provides a consistent estimator of the true distribution of $\Pi_n(\theta_F)$, the weak limit of which may not be analytically tractable. The above results lead immediately to the following corollary, which establishes asymptotic correctness of the confidence region $\mathcal{R}_{n,1-\alpha}$.

COROLLARY 3.1. *Under the conditions of Theorem* 3.1, *we have, for any* $\alpha \in (0,1)$,

$$\lim_{n\to\infty} \mathrm{pr}_F(\theta_F \in \mathcal{R}_{n,1-\alpha}) = 1 - \alpha.$$

REMARK 3.1. To show that the condition (B1) is indispensable in the nonstandard setup, we establish below inconsistency of the standard, n out of n, bootstrap under the simple scenario where $\theta_F = \xi_F \in \Xi$ and the nuisance parameter η_F is absent. The proof is given in the Supplementary Material [22].

THEOREM 3.2. *Assume the conditions* (G1), (G2), (A2), (A3), (A5)–(A8) *and that* $d = p$ *and* $q = 0$. *Then, for any* $x \in \mathbb{R}$:

(i) $\lim_{n\to\infty} \mathrm{pr}_F(\Pi_n(\xi_F) \le x) = \mathrm{pr}(\sup_{s\in\Xi_F^\dagger} \mathbb{Z}(s) \le x)$, *for a Gaussian process* $\mathbb{Z}$ *in* $\ell^\infty(\Xi_F^\dagger)$ *which satisfies, for any* $s, s_1, s_2 \in \Xi_F^\dagger$, $\mathbb{E}[\mathbb{Z}(s)] = \Lambda(s)$ *and* $\mathrm{Cov}(\mathbb{Z}(s_1), \mathbb{Z}(s_2)) = \Sigma(s_1,s_2)$;

(ii) *with* $m = n$ *and assuming further the condition* (B3), $\mathrm{pr}(\Pi_n^*(\hat{\xi}_n) \le x | \mathcal{X}_n)$ *converges in probability to*

$$\mathrm{pr}\Big(\sup_{s\in\Xi_F^\dagger} \mathbb{W}(s) - \mathbb{W}\Big(\operatorname*{argmax}_{t\in\Xi_F^\dagger} \tilde{\mathbb{Z}}(t)\Big) \le x | \tilde{\mathbb{Z}}\Big),$$

where $\mathbb{W} = \mathbb{Z} + \tilde{\mathbb{Z}} - \Lambda$ *and* $\mathbb{Z}, \tilde{\mathbb{Z}}$ *denote two independent replicates of the process* $\mathbb{Z}$ *specified in* (i).

We see from Theorem 3.2 that the n out of n bootstrap distribution converges in probability to a random limit in general, thus failing to estimate the distribution of $\Pi_n(\xi_F)$ consistently. An exception is provided by the special case commonly obtained under regularity conditions, where $\Xi_F^\dagger = \mathbb{R}^p$ and $\mathbb{Z}(s) = s^\top \Sigma_0 Z - (1/2) s^\top \Lambda_0 s$, for some positive-definite symmetric matrices Σ_0, Λ_0 and a standard normal p-variate vector Z. In this case, we have

$\sup_{s\in\mathbb{R}^p}\mathbb{Z}(s) = (1/2)Z^\top\Sigma_0\Lambda_0^{-1}\Sigma_0 Z$ and, writing $(Z,\tilde{Z})$ for two independent replicates of Z, that

$$\begin{aligned}
&\sup_{s\in\mathbb{R}^p}\mathbb{W}(s) - \mathbb{W}\bigl(\operatorname*{argmax}_{t\in\mathbb{R}^p}\tilde{\mathbb{Z}}(t)\bigr)\\
&\quad= (1/2)(Z+\tilde{Z})^\top\Sigma_0\Lambda_0^{-1}\Sigma_0(Z+\tilde{Z})\\
&\qquad-\{\tilde{Z}^\top\Sigma_0\Lambda_0^{-1}\Sigma_0 Z + (1/2)\tilde{Z}^\top\Sigma_0\Lambda_0^{-1}\Sigma_0\tilde{Z}\}\\
&\quad= (1/2)Z^\top\Sigma_0\Lambda_0^{-1}\Sigma_0 Z = \sup_{s\in\mathbb{R}^p}\mathbb{Z}(s).
\end{aligned}$$

Thus the two weak limits derived in parts (i) and (ii) of Theorem 3.2 coincide, so that the n out of n bootstrap is consistent.

REMARK 3.2. In the context of M-estimation, a resampling-based alternative to the bootstrap has been proposed with $\hat{B}_n^*(\xi)$ defined to be $(n\mathbb{E}[\omega_1^*])^{-1}\times\sum_{i=1}^n\omega_i^* b_{\xi,\hat{\eta}_n^*}(X_i)$, where $(\omega_1^*,\ldots,\omega_n^*)$ denotes a random sample of weights independent of $\mathcal{X}_n$ and the coefficient of variation $\phi = \sqrt{\mathrm{Var}(\omega_1^*)}/\mathbb{E}[\omega_1^*]$ is taken to be a fixed positive constant. The method has been variously termed the perturbation bootstrap [17], the weighted bootstrap [27] or the multiplier bootstrap [35]. Applying the perturbation bootstrap in our present setup, ϕ^2 plays the same role as n/m in the m out of n bootstrap, so that the standard choice of a fixed $\phi>0$ shares similar asymptotic properties with the n out of n bootstrap, which is inconsistent. For consistency, we may consider generating the random weights from a triangular array $(\omega_{n1}^*,\ldots,\omega_{nn}^*)$ satisfying the conditions:

(PB1) for each n, $(\omega_{n1}^*,\ldots,\omega_{nn}^*)$ are independent and identically distributed;

(PB2) $\phi_n = \sqrt{\mathrm{Var}(\omega_{n1}^*)}/\mathbb{E}[\omega_{n1}^*]\to\infty$ and $\phi_n = o(n^{1/2})$;

(PB3) for some constant $c>0$, $\mathrm{pr}(|\omega_{n1}^* - \mathbb{E}[\omega_{n1}^*]| > x\sqrt{\mathrm{Var}(\omega_{n1}^*)}) = O(e^{-cx})$ as $n, x\to\infty$.

The conditions (PB1)–(PB3) are readily satisfied by weights generated from the distribution of $\phi_n(W-\mathbb{E}[W])+\sqrt{\mathrm{Var}(W)}$, for any random variable W following a fixed, exponentially-tailed, distribution such as the normal, the gamma, the beta or the Bernoulli distributions. We prove in the Supplementary Material [22] the following theorem, which asserts consistency of the perturbation bootstrap based on weights satisfying (PB1)–(PB3).

THEOREM 3.3. *Let $(\omega_{n1}^*,\ldots,\omega_{nn}^*)$ be independent of $\mathcal{X}_n$ and satisfy* (PB1)–(PB3). *Then Theorem* 3.1(ii) *and Corollary* 3.1 *hold with* $\hat{B}_n^*(\xi) = (n\mathbb{E}[\omega_{n1}^*])^{-1}\times\sum_{i=1}^n\omega_{ni}^* b_{\xi,\hat{\eta}_n^*}(X_i)$, $\xi\in\Xi$ *and* $m = n\phi_n^{-2}$.

As with the choice of m in the m out of n bootstrap, the need for empirical tuning of ϕ_n does not make the perturbation bootstrap computationally more convenient in practice.

REMARK 3.3. Under nonstandard conditions, the smoothed bootstrap has occasionally been suggested to correct for inconsistency of the n out of n bootstrap. Compared to the m out of n bootstrap, the smoothed bootstrap enjoys the convenience of retaining the conventional bootstrap sample size n, but requires instead some problem-specific smoothing scheme and careful tuning of the accompanying bandwidth. Typically, smoothing is effective only when the distribution of the pivot depends on some local feature $T(F)$, such as density at a fixed point, of F and when the smoothed empirical distribution $\tilde{F}_n$, from which bootstrap samples are to be drawn, is constructed in a way such that $T(\tilde{F}_n)$ is sufficiently close to $T(F)$. In cases

where F is discrete or not everywhere differentiable, smoothing can hardly be expected to bring much benefit to the bootstrap. We shall see that smoothing may be possible in two of the examples studied in Section 5, examine its empirical performance and comment on its limitations when applied to those specific contexts.

REMARK 3.4. The nonstandard nature of our present setup forbids us to establish sharp bounds on the order of accuracy of $\hat{G}$ or the order of coverage error of $\mathcal{R}_{n,1-\alpha}$, though we expect both orders to converge more slowly than the regular $n^{1/2}$ rate. In cases where $\Pi_n(\theta_F)$ can be approximated by a measurable function of sample averages, Edgeworth expansion techniques may typically be applied to derive more explicit error orders. We illustrate results of this kind in Section 5.4 with the binomial example.

4. Extension to triangular arrays. Suppose now that $\mathcal{X}_n = (X_{n1}, \ldots, X_{nn})$ constitutes a row-wise independent triangular array such that the X_{ni}'s are defined on a common measure space but are not necessarily identically distributed. Although certain multiplier bootstrap procedures have been developed for empirical processes defined on triangular arrays [6, 19], application of the bootstrap to nonstandard M-estimation under this setting remains little studied. We shall show that our main results in Theorem 3.1 and Corollary 3.1 hold for triangular arrays, if the parameter of interest $\xi_n = (\theta_n, \psi_n)$ is assumed to uniquely maximise $\sum_{i=1}^n \mathbb{E}[b_{\xi,\eta_n}(X_{ni})]$ over $\xi \in \Xi$ and $\eta_n \in \mathcal{H}$ denotes the true value of a nuisance parameter. Note that both ξ_n and η_n are allowed to depend on n. For ease of exposition, we assume that each ξ_n is an interior point in Ξ. Following the notation introduced in Section 3, with (X_{ni}, ξ_n, η_n) replacing (X_i, ξ_F, η_F) and $(\mathbb{R}^p, \mathbb{R}^{p-d})$ replacing $(\Xi_F^\dagger, \Xi_{F,2}^\dagger)$, we define $\Psi_n(\xi, \eta) = n^{-1}\sum_{i=1}^n \mathbb{E}[b_{\xi,\eta}(X_{ni})]$ for $(\xi, \eta) \in \Xi \times \mathcal{H}$ and modify below the conditions of Theorem 3.1 to accommodate nonidentically distributed data. Note that the functions $g_{s,t}$, G_δ and the function classes $\mathcal{G}_\delta$, $\mathcal{G}_\delta(\epsilon)$ depend on n through (ξ_n, η_n), which we suppress in the notation for clarity.

(G1′) (G1) holds with $\Lambda(s) = \lim_{n\to\infty} k_n^\nu n^{-1}\sum_{i=1}^n \mathbb{E}[g_{s/k_n,0}(X_{ni})]$ for any sequence $\{k_n\}$ satisfying $n \geq k_n \to \infty$;

(G2′) (G2) holds with $\Sigma(s_1, s_2) = \lim_{n\to\infty}(nk_n)^{-1} r_{k_n}^{2\nu}\sum_{i=1}^n \mathbb{E}[g_{s_1/r_{k_n},0}(X_{ni}) \times g_{s_2/r_{k_n},0}(X_{ni})]$ for any sequence $\{k_n\}$ satisfying $n \geq k_n \to \infty$;

(A3′) (A3) holds with $\Psi = \Psi_n$ for some $C_0 > 0$ independent of n and for sufficiently large n;

(A4′) (A4) holds with $\Psi = \Psi_n$ for some $K_0, \nu_1, \nu_2 > 0$ independent of n, with $\nu_1 \leq \nu \leq \nu_1 + \nu_2$, and for sufficiently large n;

(A5′) for any $r > 0$, the function class $\mathcal{B}_r = \{b_{\xi,\eta} : |(\xi, \eta)| \leq r,\ (\xi, \eta) \in \Xi \times \mathcal{H}\}$ has an envelope function $\bar{B}_r$ and satisfies $\int_0^1 \sup_Q \sqrt{\log N(\epsilon \|\bar{B}_r\|_{Q,2}, \mathcal{B}_r, Q)}\, d\epsilon < \infty$, where the supremum is taken over all finitely discrete probability measures Q and $n^{-1}\sum_{i=1}^n \mathbb{E}[\bar{B}_r(X_{ni})^2] = O(1)$;

(A6′) (A6) holds with $r_n^{2\nu} n^{-1}\mathbb{E}[G_{c/r_n}(X_1)^2; G_{c/r_n}(X_1) > \epsilon n^{\tau_0} r_n^{-\nu}]$ and $\mathbb{E}_F[G_\delta(X_1)^2]$ replaced respectively by $r_{k_n}^{2\nu}(nk_n)^{-1}\sum_{i=1}^n \mathbb{E}[G_{c/r_{k_n}}(X_{ni})^2; G_{c/r_{k_n}}(X_{ni}) > \epsilon k_n^{\tau_0} r_{k_n}^{-\nu}]$ and $n^{-1}\sum_{i=1}^n \mathbb{E}[G_\delta(X_{ni})^2]$, for any sequence $\{k_n\}$ satisfying $n \geq k_n \to \infty$;

(A7′) (A7) holds with finiteness replaced by boundedness of the entropy integral and the last condition replaced by

$$r_{k_n}^{2\nu} k_n^{-1} \sup\left\{ n^{-1}\sum_{i=1}^n \mathbb{E}\big[f(X_{ni})^2\big] : f \in \mathcal{G}_{c/r_{k_n}}(\delta_n / r_{k_n}) \right\} \to 0,$$

for any sequence $\{k_n\}$ satisfying $n \geq k_n \to \infty$;

(B4′) (B4) holds for some $C_H > 0$ independent of n.

We prove in the Supplementary Material [22] the following theorem, which asserts consistency of the m out of n bootstrap under a triangular array setting.

THEOREM 4.1. *Let $\mathcal{X}_n = (X_{n1}, \ldots, X_{nn})$ be a row-wise independent triangular array. The conclusion of Theorem 3.1(i) holds for $\Pi_n(\theta_n)$ under the conditions* (G1′), (G2′), (A1), (A2), (A8) *and* (A3′)–(A7′). *If we assume further the conditions* (B1)–(B3) *and* (B4′), *then the conclusion of Theorem 3.1*(ii) *holds for $\hat{G}_n$.*

Similar to Corollary 3.1, asymptotic correctness of the confidence region $\mathcal{R}_{n,1-\alpha}$ for θ_n also follows under the conditions of Theorem 4.1.

5. Examples. We illustrate the applications of our proposed bootstrap confidence procedure with four examples. The first three are set in the context of a linear regression model $Y_i = Z_i^\top \beta_F + U_i$, $i = 1, \ldots, n$, for an unknown parameter vector $\beta_F \in \mathbb{R}^k$ and n independent and identically distributed observations $X_i = (Y_i, Z_i)$. We assume that the random errors U_i are independent of the covariates Z_i, and denote by F_U and F_Z the distribution functions of U_1 and Z_1, respectively. The nuisance parameter η is absent in the first two examples, and is implicitly defined in the third example. The fourth example considers the problem of estimating an ordered pair of binomial probabilities, for which standard inference procedures may fail when the two probabilities are equal. Each example is supplemented with a simulation study. Proofs of the theoretical results are given in the Supplementary Material [22].

5.1. *Maximum score estimation.* Suppose that $k \geq 2$, $\beta_F \neq 0$ and we are interested in estimating $\beta_F/|\beta_F|$, the normalised regression parameter vector. Introduced by [28], the maximum score estimator $\hat{\beta}_n$ is defined to be a value of $\beta \in \mathcal{S}_k \equiv \{x \in \mathbb{R}^k : |x| = 1\}$ which maximises $n^{-1}\sum_{i=1}^n m_\beta(Y_i, Z_i)$, where $m_\beta(y, z) = \operatorname{sgn}(y)\mathbf{1}\{z^\top \beta \geq 0\}$, $\mathbf{1}\{\cdot\}$ denotes the indicator function and $\operatorname{sgn}(y) = \mathbf{1}\{y \geq 0\} - \mathbf{1}\{y < 0\}$. Denote by $\hat{\beta}_n^*$ the bootstrap counterpart of $\hat{\beta}_n$ calculated from the m out of n bootstrap observations $X_1^*, \ldots, X_m^*$. To apply Theorem 3.1 we assume the following conditions:

(MS1) F_U has a continuous density $f_U = F_U'$ with $f_U(0) > 0$, $F_U(0) = 1/2$ and $F_U(u) \in (0, 1)$ for all $u \in \mathbb{R}$;

(MS2) F_Z has a positive, continuously differentiable, density f_Z on a compact support in $\mathbb{R}^k$.

Define a $k \times (k-1)$ matrix $\mathcal{O} = [e_1, \ldots, e_{k-1}]$, where $\{e_1, \ldots, e_{k-1}\}$ constitutes an orthonormal basis for the orthogonal complement of the space $\{c\beta_F : c \in \mathbb{R}\}$. We deduce from Theorem 3.1 the following corollary.

COROLLARY 5.1. *Assume* (B1), (MS1) *and* (MS2). *Then, for some Gaussian process $\mathbb{Z}$ in $\ell^\infty(\mathbb{R}^{k-1})$, $n^{-1/3}\sum_{i=1}^n\{m_{\hat{\beta}_n}(X_i) - m_{\beta_F}(X_i)\}$ converges weakly to $\sup_{s\in\mathbb{R}^{k-1}}\mathbb{Z}(s)$, and $m^{-1/3}\sum_{i=1}^m\{m_{\hat{\beta}_n^*}(X_i^*) - m_{\hat{\beta}_n}(X_i^*)\}$ converges weakly in probability to the same limit. The covariance and mean functions of $\mathbb{Z}$ are given, for $s, s_1, s_2 \in \mathbb{R}^{k-1}$, by*

$$\Sigma(s_1, s_2) = (1/2)\int\{|v^\top s_1| + |v^\top s_2| - |v^\top(s_1 - s_2)|\}f_Z(\mathcal{O}v)\,dv$$

and $\Lambda(s) = -f_U(0)\int(v^\top s)^2 f_Z(\mathcal{O}v)\,dv$, respectively.

Denoting by $\hat{G}_n$ the distribution function of $m^{-1/3}\sum_{i=1}^m\{m_{\hat{\beta}_n^*}(X_i^*) - m_{\hat{\beta}_n}(X_i^*)\}$ conditional on $\mathcal{X}_n$, it follows immediately from Corollary 5.1 that the confidence region $\mathcal{R}_{n,1-\alpha} =$

$\{\beta \in \mathcal{S}_k : n^{-1/3}\sum_{i=1}^n [m_{\hat\beta_n}(X_i) - m_\beta(X_i)] \le \hat G_n^{-1}(1-\alpha)\}$ for $\beta_F/|\beta_F|$ has asymptotically correct coverage probability $1-\alpha$, for $\alpha \in (0,1)$.

In their study of the same example, [18] show that the statistic $\Pi_n^C(\beta_F/|\beta_F|) = n^{1/3}\{\hat\beta_n - |\beta_F|^{-2}(\hat\beta_n^\top \beta_F)\beta_F\}$ converges weakly to $\mathcal{O}\cdot \operatorname{argmax}_{s\in\mathbb{R}^{k-1}} \mathbb{Z}(s)$. A conventional approach to constructing bootstrap confidence regions for $\beta_F/|\beta_F|$ would have taken $\Pi_n^C(\beta_F/|\beta_F|)$ as a pivot and estimated its distribution by that of $m^{1/3}\{\hat\beta_n^* - (\hat\beta_n^{*\top}\hat\beta_n)\hat\beta_n\}$, conditional on $\mathcal{X}_n$. It is clear that the latter distribution has support in the orthogonal complement $\hat{\mathcal{C}}$ of $\{c\hat\beta_n : c\in\mathbb{R}\}$. It then follows that the resulting bootstrap confidence region for $\beta_F/|\beta_F|$ is necessarily contained in the set $\{\beta\in\mathcal{S}_k : \Pi_n^C(\beta)\in\hat{\mathcal{C}}\} = \{\hat\beta_n, -\hat\beta_n\}$, a result which is practically useless. The problem persists even if $\Pi_n^C(\beta_F/|\beta_F|)$ is redefined to be $n^{1/3}(\hat\beta_n - \beta_F/|\beta_F|)$. Our proposed confidence region $\mathcal{R}_{n,1-\alpha}$ is free of such deficiencies.

Bootstrap inference based on the maximum score estimator has been studied in previous works. Abrevaya and Huang [1] show that the standard, n out of n, bootstrap distribution of $n^{1/3}(\hat\beta_n^* - \hat\beta_n)$ is inconsistent. Patra et al. [32] propose a smoothed bootstrap procedure for consistently estimating the distribution of $n^{1/3}(\hat\beta_n - \beta_F/|\beta_F|)$ and constructing one-dimensional confidence intervals for the first component of $\beta_F/|\beta_F|$ with accurate empirical coverages. When used for constructing multi-dimensional confidence regions for $\beta_F/|\beta_F|$, the above smoothed bootstrap procedure encounters the same difficulty intrinsic to any attempt to bootstrap the pivot $\Pi_n^C(\beta_F/|\beta_F|)$ directly, as has been discussed before. In their proof of consistency of the smoothed bootstrap, [32] require that f_Z be estimated with sup-norm error of order $o_p(n^{-1/3})$. It is well known that second-order kernel density estimators cannot achieve an error rate faster than $n^{-2/(k+4)}$. For the case $k\ge 2$, estimation of f_Z with $o_p(n^{-1/3})$ error necessarily calls for a higher-order kernel density estimator, which is not a proper density function suitable for simulation purposes. Patra et al. [32] ignore the above problem and use a Gaussian kernel, which is only second-order, in their simulation experiments.

Simulation study. Let $k=2$, $2U_1$ follow a standard Cauchy distribution and Z_1 be bivariate normal with mean $(1, 0.5)^\top$ and dispersion matrix $((0.623^2, 0)^\top, (0, 0.389^2)^\top)$. We set $\beta_F = \beta_F/|\beta_F| = (\cos 0.4, \sin 0.4)^\top$, for which 95% bootstrap confidence regions are to be constructed based on the random sample $(Y_1, Z_1), \ldots, (Y_n, Z_n)$. Note that we cannot invoke asymptotic normality to construct the confidence region based on the least squares estimator $\hat\beta_{\text{ols}}$ of β_F, as U_1 does not have a finite variance. In contrast, the maximum score estimator provides a robust solution, noting that (MS1) holds for the Cauchy distribution.

Because of the anomaly of the conventional bootstrap confidence region built upon the pivot $\Pi_n^C(\beta_F/|\beta_F|)$, we consider instead an alternative pivot defined by $\tilde\Pi_n^C(\beta_F/|\beta_F|) = n^{1/3}\{\tan^{-1}(\hat\beta_n^{[2]}/\hat\beta_n^{[1]}) - \tan^{-1}(\beta_F^{[2]}/\beta_F^{[1]})\}$, where $\beta^{[j]}$ denotes the jth component of the vector $\beta\in\mathbb{R}^2$. The corresponding $100(1-\alpha)\%$ equal-tailed m out of n bootstrap confidence region is then given by

$$\mathcal{R}_{n,1-\alpha}^C = \{(\cos\gamma, \sin\gamma)^\top : \gamma \in [\tan^{-1}(\hat\beta_n^{[2]}/\hat\beta_n^{[1]}) - n^{-1/3}\hat Q_{n,1-\alpha/2},$$
$$\tan^{-1}(\hat\beta_n^{[2]}/\hat\beta_n^{[1]}) - n^{-1/3}\hat Q_{n,\alpha/2}]\},$$

where $\hat Q_{n,\alpha}$ denotes the αth m out of n bootstrap quantile of $\tilde\Pi_n^C(\beta_F/|\beta_F|)$. In addition to the m out of n bootstrap, we apply the smoothed bootstrap of [32] to estimate the quantiles of $\Pi_n(\beta_F/|\beta_F|)$ and $\tilde\Pi_n^C(\beta_F/|\beta_F|)$, based on a Gaussian kernel and Scott's bandwidth selection rule, yielding smoothed bootstrap confidence regions denoted by $\mathcal{R}_{n,1-\alpha}^{\text{sb}}$ and $\mathcal{R}_{n,1-\alpha}^{C,\text{sb}}$, respectively. Although theoretically invalid, we calculate also the confidence region $\mathcal{R}_{n,1-\alpha}^{C,\text{ols}}$

TABLE 1
Maximum score example—coverage probabilities of $\mathcal{R}_{n,0.95}$, $\mathcal{R}^{C}_{n,0.95}$, $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$ *and* $\mathcal{R}^{C,\mathrm{ols}}_{n,0.95}$ *for* $n = 1000$, *each estimated by averaging over* 200 *random samples*

m	100	200	300	400	500	600	700	800	900	1000
$\mathcal{R}_{n,0.95}$	0.955	0.950	0.930	0.930	0.920	0.925	0.930	0.915	0.925	0.910
$\mathcal{R}^{C}_{n,0.95}$	0.480	0.505	0.520	0.510	0.520	0.555	0.530	0.540	0.515	0.500
$\mathcal{R}^{\mathrm{sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.980
$\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.700
$\mathcal{R}^{C,\mathrm{ols}}_{n,0.95}$	0.705									

derived from normal approximation to the pivot $\tan^{-1}(\hat{\beta}^{[2]}_{\mathrm{ols}}/\hat{\beta}^{[1]}_{\mathrm{ols}}) - \tan^{-1}(\beta^{[2]}_F/\beta^{[1]}_F)$, assuming, wrongly, a finite variance for U_1.

Setting $n = 1000$ and averaging over 200 independent replicates, we compare the coverage probabilities of $\mathcal{R}_{n,0.95}$, $\mathcal{R}^{C}_{n,0.95}$, $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$ and $\mathcal{R}^{C,\mathrm{ols}}_{n,0.95}$. Monte Carlo approximation to each bootstrap distribution is obtained from 100 bootstrap samples drawn from the parent sample. In each case, the criterion function is maximised numerically by simulated annealing. Table 1 shows the estimated coverages of the confidence regions, with m out of n bootstrap sample sizes set to be $m = 100j$, $j = 1, \ldots, 10$. We see that $\mathcal{R}^{C}_{n,0.95}$ yields very poor coverages, which in no cases exceed 55.5%. Applying the smoothed bootstrap to $\tilde{\Pi}^{C}_n(\beta_F/|\beta_F|)$ yields a slightly better coverage of 70.0%, which is on a par with the 70.5% given by the invalid normal approximation method. The coverages of our proposed bootstrap region $\mathcal{R}_{n,0.95}$ are considerably more accurate than those derived from the pivot $\tilde{\Pi}^{C}_n(\beta_F/|\beta_F|)$, and the results appear to vary little with the choice of m. Its smoothed bootstrap version $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$ exhibits slight over-coverage, making no apparent improvement over $\mathcal{R}_{n,0.95}$.

Figure 1 shows polar plots of the sample and population criterion functions against $\omega = \tan^{-1}(\beta^{[2]}/\beta^{[1]}) \in [0, 2\pi)$. The first random sample drawn in the above simulation study is used for plotting the sample criterion function. We see that although the population criterion function is maximised uniquely at the true value $\omega_F = 0.4$, the sample criterion function admits multiple global maximisers, among which simulated annealing settles on the particular solution $\hat{\omega}_n = 0.0692$. Nonuniqueness of maximisers as such accounts for the poor performance of methods based on the conventional pivot $\tilde{\Pi}^{C}_n(\beta_F/|\beta_F|) = n^{1/3}(\hat{\omega}_n - \omega_F)$. The problem recedes only when n becomes much bigger than 1000. Our proposed procedure depends on $\hat{\omega}_n$ only through the maximised criterion functions $\hat{B}_n(\hat{\beta}_n)$ and $\hat{B}^{*}_n(\hat{\beta}_n)$, which are less adversely affected by the problem of multiple maxima.

5.2. *L_1 regression estimation.* Following our general setting, we consider building confidence regions for θ_F, a d-variate subvector of β_F, for $1 \le d \le k$. Assume that F_U and F_Z satisfy, for some $\zeta, L, \Delta > 0$:

(LAD1) $F_U(u) - F_U(0) = \mathrm{sgn}(u)|u|^{\zeta} L/\zeta$ for $|u| \le \Delta$, and $\int |u|\, dF_U(u) < \infty$;
(LAD2) $\mathrm{pr}(Z_1^{\top}\xi = 0) < 1$ for any $\xi \ne 0$, and $\int |z|^{\max\{2,\zeta+1\}}\, dF_Z(z) < \infty$.

Note that (LAD1) covers a variety of shapes of F_U around 0. In particular, for $\zeta \ge 1$, F_U has a finite density f_U around 0 such that $f_U(0) = L$ or 0 according as $\zeta = 1$ or $\zeta > 1$. For $\zeta < 1$, F_U is not differentiable at 0. Under (LAD2), F_Z does not degenerate to a hyperplane passing through the origin in $\mathbb{R}^k$.

Define $b_\xi(y, z) = -|y - z^{\top}\xi|$ for $(y, z) \in \mathbb{R} \times \mathbb{R}^k$ and $\xi \in \mathbb{R}^k$. The L_1 regression estimator $\hat{\xi}_n = (\hat{\theta}_n, \hat{\psi}_n)$ is defined to be the maximiser of $\hat{B}_n(\xi) = n^{-1}\sum_{i=1}^{n} b_\xi(X_i)$. Similarly, for a

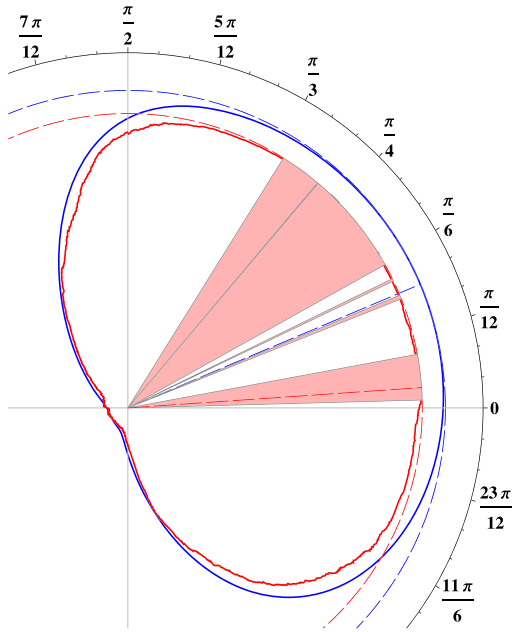

FIG. 1. *Maximum score example—polar plots of* $B_F(\beta) = \mathbb{E}_F[\operatorname{sgn}(Y_1)\mathbf{1}\{Z_1^\top \beta \geq 0\}]$ (*blue solid*) *and* $\hat{B}_n(\beta) = n^{-1}\sum_{i=1}^n \operatorname{sgn}(Y_i)\mathbf{1}\{Z_i^\top \beta \geq 0\}$ (*red solid*) *against* $\tan^{-1}(\beta^{[2]}/\beta^{[1]}) \in [0, 2\pi)$. *The blue dashed line and circle indicate the true direction* $\beta_F = (\cos(0.4), \sin(0.4))^\top$ *and* $B_F(\beta_F)$, *respectively. The red dashed line and circle indicate the estimated direction* $\hat{\beta}_n$ *found by simulated annealing and* $\hat{B}_n(\hat{\beta}_n)$, *respectively. The red shaded sectors indicate the set of maximisers of* $\hat{B}_n(\beta)$.

given $\theta \in \mathbb{R}^d$, the profile estimator $\tilde{\psi}_n(\theta)$ is the maximiser of $\hat{B}_n(\theta, \psi) = n^{-1}\sum_{i=1}^n b_{\theta,\psi}(X_i)$ over $\psi \in \mathbb{R}^{k-d}$. The m out of n bootstrap estimators $\hat{\xi}_n^*$ and $\tilde{\psi}_n^*(\hat{\theta}_n)$ are defined analogously, with the observations X_i replaced by X_i^*. Our general Theorem 3.1 then implies the following corollary, and hence asymptotic correctness of the confidence region $\mathcal{R}_{n,1-\alpha}$.

COROLLARY 5.2. *Assume* (B1), (LAD1) *and* (LAD2). *Then, for some Gaussian process* $\mathbb{Z}$ *in* $\ell^\infty(\mathbb{R}^k \times \mathbb{R}^{k-d})$, $\Pi_n(\theta_F) = n^{(1+\zeta)/(2\zeta)}\{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\theta_F, \tilde{\psi}_n(\theta_F))\}$ *converges weakly to* $\sup_{s\in\mathbb{R}^k}\inf_{u\in\mathbb{R}^{k-d}}\mathbb{Z}(s,u)$, *and* $m^{(1+\zeta)/(2\zeta)}\{\hat{B}_n^*(\hat{\xi}_n^*) - \hat{B}_n^*(\hat{\theta}_n, \tilde{\psi}_n^*(\hat{\theta}_n))\}$ *converges weakly in probability to the same limit. The covariance and mean functions of* $\mathbb{Z}$ *are as specified in Theorem* 3.1(i) *where, for* $s, s_1, s_2 \in \mathbb{R}^k$, $\Sigma(s_1, s_2) = \int s_1^\top z z^\top s_2\, dF_Z(z)$ *and* $\Lambda(s) = -2\zeta^{-1}(\zeta+1)^{-1}L\int |z^\top s|^{\zeta+1}\, dF_Z(z)$.

We see from Corollary 5.2 that the process $\mathbb{Z}$ has a quadratic covariance function in general. Lai and Lee [21] show under the same conditions that the conventional pivot $n^{1/(2\zeta)}(\hat{\xi}_n - \beta_F)$ converges weakly to the maximiser of a Gaussian process in $\ell^\infty(\mathbb{R}^k)$ with mean function Λ and covariance function Σ.

In the special case where $\zeta = 1$ so that F_U has a positive density L at 0, the mean function of $\mathbb{Z}$ is also quadratic. It follows that $\mathbb{Z}(s,u)$ has the representation $[s^\top - (0, u^\top)]W - Ls^\top\Sigma_Z s + L(0,u^\top)\Sigma_Z(0,u^\top)^\top$, where $\Sigma_Z = \int zz^\top\, dF_Z(z)$ and $W \sim N(0, \Sigma_Z)$. Denoting by W_2 the last $k-d$ components of W and by $[\Sigma_Z]_{22}$ its corresponding dispersion matrix, $\sup_{s\in\mathbb{R}^k}\inf_{u\in\mathbb{R}^{k-d}}\mathbb{Z}(s,u)$ has the closed-form expression $(1/4)L^{-1}(W^\top\Sigma_Z^{-1}W - W_2^\top[\Sigma_Z]_{22}^{-1}W_2)$, which is distributed as $(1/4)L^{-1}\chi_d^2$, a scaled chi-squared distribution on d degrees of freedom. We note that in this case the standard, n out of n, bootstrap region is asymptotically correct, while it generally fails to be consistent for $\zeta \neq 1$: see also Remark 3.1 for a general result.

TABLE 2
L_1 regression example—coverage probabilities of $\mathcal{R}^C_{n,0.95}$, and $\mathcal{R}_{n,0.95}$ for $n = 100$, each estimated by averaging over 100 random samples

m	10	20	30	40	50	60	70	80	90	100
$\mathcal{R}^C_{n,0.95}$	0.85	0.61	0.53	0.49	0.46	0.38	0.47	0.28	0.29	0.46
$\mathcal{R}_{n,0.95}$	0.96	0.96	0.95	0.91	0.92	0.87	0.89	0.92	0.86	0.96

Simulation study. Define, for $|u| \le 1$, $F_U(u) = (1/2)\{1 + \operatorname{sgn}(u)u^2\}$, which satisfies (LAD1) with $\zeta = 2$ and $\Delta = L = 1$. Setting $k = 3$, we take F_Z to be the distribution function of the three-dimensional standard normal distribution, which satisfies (LAD2). We are interested in constructing a two-dimensional 95% confidence region for $\theta_F = (\beta_F^{[1]}, \beta_F^{[2]})^\top$, the first two components of the three-dimensional regression parameter β_F. Two choices of pivot are considered, namely the conventional, two-dimensional pivot $n^{1/4}(\hat{\theta}_n - \theta)$ and our proposed univariate pivot $\Pi_n(\theta) = n^{3/4}\{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\theta, \tilde{\psi}_n(\theta))\}$. With the two-dimensional pivot, the Tukey depth is used for calibrating its bootstrap distribution. We set $n = 100$ and $\beta_F = (1, 0.5, 0.3)^\top$, so that $\theta_F = (1, 0.5)^\top$. From each parent sample, 100 bootstrap samples are generated for the construction of each bootstrap confidence region.

Table 2 shows the coverages of the bootstrap confidence regions $\mathcal{R}^C_{n,0.95}$ and $\mathcal{R}_{n,0.95}$, built respectively upon the conventional and our proposed pivots, with bootstrap sample sizes set to be $m = 10j$, $j = 1, \ldots, 10$. Each coverage is estimated by averaging over 100 random samples. As in the maximum score example, the coverages of $\mathcal{R}_{n,0.95}$ are much more accurate than those of $\mathcal{R}^C_{n,0.95}$ under all choices of m. Indeed, with the exception of the case $m = 10$, the coverages of $\mathcal{R}^C_{n,0.95}$ are consistently poor and never exceed 61%. The coverages of $\mathcal{R}_{n,0.95}$, by contrast, lie between 86% and 96% irrespective of the choice of m.

5.3. *Least quantile of squares estimation.* Introduced by [34], the least median of squares estimator of β_F is defined to be the minimiser of the median of $\{|Y_1 - Z_1^\top \xi|^2, \ldots, |Y_n - Z_n^\top \xi|^2\}$ with respect to ξ. In this example, we extend the notion of median and consider the least quantile of squares estimator $\hat{\xi}_n$, defined to be the value of ξ that minimises the qth quantile of $\{|Y_1 - Z_1^\top \xi|^2, \ldots, |Y_n - Z_n^\top \xi|^2\}$, for some fixed $q \in (0, 1)$.

Define, for $\xi \in \mathbb{R}^k$, $\nu > 0$ and $(y, z) \in \mathbb{R} \times \mathbb{R}^k$, $b_{\xi,\eta}(y, z) = \mathbf{1}\{|y - z^\top \xi| \le \eta\}$. Define the nuisance parameter $\eta_F = \inf\{\eta > 0 : \sup_\xi \mathbb{E}_F[b_{\xi,\eta}(X_1)] \ge q\}$ and its estimator $\hat{\eta}_n = \inf\{\eta > 0 : \sup_\xi n^{-1} \sum_{i=1}^n b_{\xi,\eta}(X_i) \ge q\}$. The least quantile of squares estimator $\hat{\xi}_n$ can then be identified with the value of ξ which maximises $\hat{B}_n(\xi) = n^{-1} \sum_{i=1}^n b_{\xi,\hat{\eta}_n}(X_i)$. We assume:

(LQ1) $\int zz^\top \, dF_Z(z)$ exists and is positive definite;

(LQ2) F_U has a density $f_U = F_U'$ which is bounded, symmetric about 0, continuously differentiable and nonincreasing on $[0, \infty)$, with $f_U'(\eta_F) < 0$.

Consider, as in Section 5.2, a partition $\beta_F = (\theta_F, \psi_F) \in \mathbb{R}^d \times \mathbb{R}^{k-d}$ and define, for a given $\theta \in \mathbb{R}^d$, $\tilde{\psi}_n(\theta)$ to be the maximiser of $\hat{B}_n(\theta, \psi)$ over $\psi \in \mathbb{R}^{k-d}$. The m out of n bootstrap estimators $\hat{\eta}_n^*$, $\hat{\xi}_n^*$ and $\tilde{\psi}_n^*(\hat{\theta}_n)$ are defined similarly as $\hat{\eta}_n$, $\hat{\xi}_n$ and $\tilde{\psi}_n(\theta)$, respectively, with the observations X_i replaced by X_i^*. The following corollary, which implies asymptotic correctness of $\mathcal{R}_{n,1-\alpha}$, follows from Theorem 3.1.

COROLLARY 5.3. *Assume* (B1), (LQ1) *and* (LQ2). *Then, for some Gaussian process $\mathbb{Z}$ in $\ell^\infty(\mathbb{R}^k \times \mathbb{R}^{k-d})$, $\Pi_n(\theta_F) = n^{2/3}\{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\theta_F, \tilde{\psi}_n(\theta_F))\}$ converges in distribution to $\sup_{s \in \mathbb{R}^k} \inf_{u \in \mathbb{R}^{k-d}} \mathbb{Z}(s, u)$, and $m^{2/3}\{\hat{B}_n^*(\hat{\xi}_n^*) - \hat{B}_n^*(\hat{\theta}_n, \tilde{\psi}_n^*(\hat{\theta}_n))\}$ converges weakly in*

probability to the same limit. The mean and covariance functions of $\mathbb{Z}$ *are as specified in Theorem* 3.1(i) *where, for* $s, s_1, s_2 \in \mathbb{R}^k$, $\Lambda(s) = f'_U(\eta_F) \int (z^\top s)^2 \, dF_Z(z)$ *and* $\Sigma(s_1, s_2) = f_U(\eta_F) \int \{|z^\top s_1| + |z^\top s_2| - |z^\top (s_1 - s_2)|\} \, dF_Z(z)$.

For the case $q = 1/2$, Kim and Pollard [18] show under the same conditions that $n^{1/3}(\hat{\xi}_n - \xi_F)$ converges weakly to the maximiser of a Gaussian process with mean function Λ and covariance function Σ as specified in Corollary 5.3 above.

In the special case where $k = d = 1$ and $Z_i \equiv 1$, the weak limit asserted in Corollary 5.3 has the expression $\{4 f_U(\eta_F)^2 / |f'_U(\eta_F)|\}^{1/3} \sup_{s \in \mathbb{R}} \{\mathbb{B}(s) - s^2\}$, where $\mathbb{B}$ denotes a standard two-sided Brownian motion on $(-\infty, \infty)$. Groeneboom [15] has derived the joint density function of the maximum and maximiser of the process $\mathbb{B}(s) - s^2$, which, combined with consistent estimation of $f_U(\eta_F)$ and $f'_U(\eta_F)$, provides a possible method for constructing an asymptotically correct confidence set for $\beta_F \in \mathbb{R}$. Assuming further that η_F is known, $\hat{\xi}_n$ then reduces to Chernoff's univariate modal estimator of β_F. In their motivation of a smoothed bootstrap confidence interval based on Chernoff's modal estimator, [24] stress the importance of drawing bootstrap samples from a kernel-smoothed empirical distribution which is symmetrized about a consistent estimator, such as the sample median, of β_F with a convergence rate faster than $n^{1/3}$. It is not clear how their procedure can be extended to higher-dimensional settings. Our proposed region $\mathcal{R}_{n,1-\alpha}$ requires neither explicit estimation of f_U and f'_U, nor any problem-specific modification of the bootstrapping scheme, and extends readily to high-dimensional settings with $k \geq d \geq 1$.

Simulation study. Consider first the case $k = d = 2$, which corresponds to a bivariate parameter of interest $\theta_F = \beta_F$. We compare coverage probabilities of 95% m out of n bootstrap confidence regions $\mathcal{R}^C_{n,0.95}$ and $\mathcal{R}_{n,0.95}$, built respectively upon the bivariate pivot $n^{1/3}(\hat{\theta}_n - \theta)$ and the univariate pivot $n^{2/3}\{\hat{B}_n(\hat{\theta}_n) - \hat{B}_n(\theta)\}$, where $\hat{\theta}_n$ is calculated by least median of squares. The Tukey depth is used for calibrating the bivariate bootstrap distribution in the construction of $\mathcal{R}^C_{n,0.95}$. Calculation of each bootstrap region is based on 100 bootstrap samples drawn from the parent sample. The first set of results, summarised in Table 3, compares the two bootstrap confidence regions under different choices of m, with n set to be

TABLE 3

Least median of squares example—coverage probabilities of $\mathcal{R}^C_{n,0.95}$ *and* $\mathcal{R}_{n,0.95}$ *for* $n = 100$, *each estimated by averaging over* 500 *random samples*

m	5	6	7	8	9
$\mathcal{R}^C_{n,0.95}$	0.86	0.81	0.85	0.78	0.80
$\mathcal{R}_{n,0.95}$	0.99	0.98	0.97	0.97	0.98
m	10	20	30	40	50
$\mathcal{R}^C_{n,0.95}$	0.76	0.76	0.66	0.61	0.61
$\mathcal{R}_{n,0.95}$	0.99	0.97	0.98	0.97	0.95
m	60	70	80	90	100
$\mathcal{R}^C_{n,0.95}$	0.60	0.64	0.57	0.55	0.54
$\mathcal{R}_{n,0.95}$	0.97	0.94	0.98	0.97	0.96

TABLE 4

Least median of squares example—coverage probabilities of $\mathcal{R}^C_{n,0.95}$ and $\mathcal{R}_{n,0.95}$. Each probability is estimated by averaging over 500 random samples for $n = 100$, and over 1000 random samples for $n = 50$

$n = 100, m = 10, Z_1 \sim N(\binom{10}{10}, \left(\begin{smallmatrix} 25 & 0 \\ 0 & 25 \end{smallmatrix}\right))$				
β_F	$(1,2)^\top$	$(1,10)^\top$	$(1,100)^\top$	$(10,20)^\top$
$\mathcal{R}^C_{n,0.95}$	0.78	0.88	0.89	0.89
$\mathcal{R}_{n,0.95}$	0.93	0.93	0.90	0.93
$n = 50, m = 10, \beta_F = (1,2)^\top, Z_1 \sim N(\binom{\mu}{0}, \left(\begin{smallmatrix} \sigma^2 & 0 \\ 0 & 25 \end{smallmatrix}\right))$				
(μ, σ)	$(0,5)$		$(10,1)$	$(10,10)$
$\mathcal{R}^C_{n,0.95}$	0.75		0.72	0.73
$\mathcal{R}_{n,0.95}$	0.97		0.96	0.96

100. Here, the covariate Z_1 is bivariate normally distributed with mean $(0,0)^\top$ and dispersion matrix $((25,0)^\top, (0,25)^\top)$, whereas U_1 has the standard normal distribution. The true value of β_F is set to be $(0,2)^\top$. Each coverage probability is estimated by averaging over 500 random samples. The coverage of $\mathcal{R}^C_{n,0.95}$ decreases in general from about 86% to 54% as m increases from 5 to 100, suggesting that a small choice of m may be desirable. On the other hand, $\mathcal{R}_{n,0.95}$ exhibits more accurate and stable coverages, which lie between 94% and 99% under all choices of m. Simulation results are also obtained under different settings of β_F and F_Z. The bootstrap sample size m is fixed at a small value compared to n, which finds favour with $\mathcal{R}^C_{n,0.95}$. The findings are summarised in Table 4. Each coverage probability is estimated by averaging over 500 random samples for $n = 100$, and over 1000 random samples for $n = 50$. We see again that $\mathcal{R}_{n,0.95}$ outperforms $\mathcal{R}^C_{n,0.95}$ in all cases. For a visual comparison of the two bootstrap regions, we calculate $\mathcal{R}_{n,0.95}$ and $\mathcal{R}^C_{n,0.95}$ from a random sample of size $n = 100$; see Figure 2. Here, we set $m = 10$, $\beta_F = (1,2)^\top$, $U_1 \sim N(0,1)$ and Z_1 to be bivariate normal with mean $(10,10)^\top$ and dispersion matrix $((25,0)^\top, (0,25)^\top)$. We see that

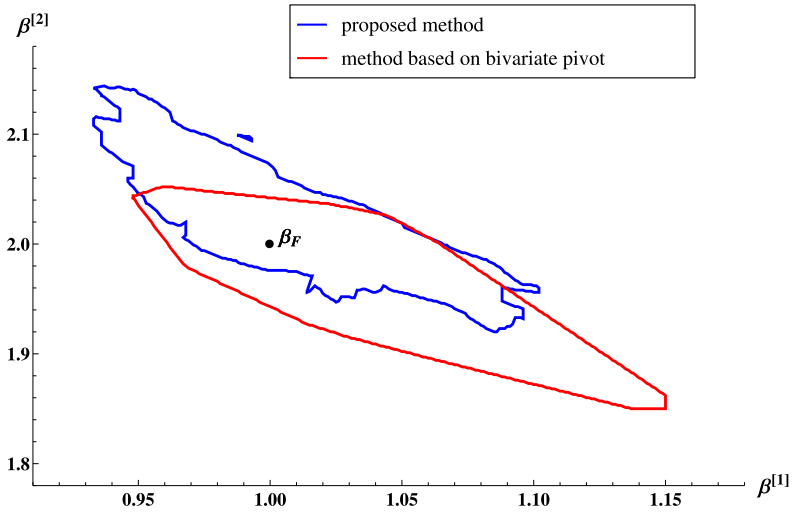

FIG. 2. *Least median of squares example—95% confidence regions $\mathcal{R}_{n,0.95}$ (blue) and $\mathcal{R}^C_{n,0.95}$ (red), based on a random sample of size 100 and $m = 10$. True value $\beta_F = (1,2)^\top$.*

TABLE 5

Chernoff's mode example—coverage probabilities of $\mathcal{R}_{n,0.95}$, $\mathcal{R}^{C}_{n,0.95}$, $\mathcal{R}^{\mathrm{sym}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sym}}_{n,0.95}$, $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{\mathrm{sym,sb}}_{n,0.95}$ *and* $\mathcal{R}^{C,\mathrm{sym,sb}}_{n,0.95}$ *for* $n = 100$, *each estimated by averaging over* 500 *random samples*

m	10	20	30	40	50	60	70	80	90	100
$\mathcal{R}_{n,0.95}$	0.786	0.800	0.790	0.826	0.810	0.808	0.818	0.796	0.814	0.826
$\mathcal{R}^{C}_{n,0.95}$	0.668	0.666	0.660	0.644	0.650	0.640	0.634	0.632	0.626	0.616
$\mathcal{R}^{\mathrm{sym}}_{n,0.95}$	0.808	0.896	0.908	0.936	0.942	0.948	0.956	0.944	0.960	0.966
$\mathcal{R}^{C,\mathrm{sym}}_{n,0.95}$	0.820	0.852	0.864	0.880	0.900	0.898	0.918	0.918	0.936	0.940
$\mathcal{R}^{\mathrm{sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.962
$\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.906
$\mathcal{R}^{\mathrm{sym,sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.962
$\mathcal{R}^{C,\mathrm{sym,sb}}_{n,0.95}$	–	–	–	–	–	–	–	–	–	0.970

the two regions have similar orientations, with $\mathcal{R}_{n,0.95}$ covering a slightly smaller area and having a more irregular boundary. The true value $\beta_F = (1, 2)^\top$ lies well within the interiors of both regions.

In the second study, we set $k = d = 1$, $Z_i \equiv 1$, assume $\eta_F = 1.7$ is known, and construct 95% m out of n bootstrap confidence intervals $\mathcal{R}^{C}_{n,0.95}$ and $\mathcal{R}_{n,0.95}$ for $\theta_F = \beta_F$ based on Chernoff's modal estimator $\hat{\theta}_n$, taken to be the mid-point of the interval $\mathrm{argmax}_\theta \hat{B}_n(\theta)$. The interval $\mathcal{R}^{C}_{n,0.95}$ is constructed to be equal-tailed. We include in the study six other intervals, namely $\mathcal{R}^{C,\mathrm{sym,sb}}_{n,0.95}$, $\mathcal{R}^{\mathrm{sym,sb}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$, $\mathcal{R}^{C,\mathrm{sym}}_{n,0.95}$ and $\mathcal{R}^{\mathrm{sym}}_{n,0.95}$, where $\mathcal{R}^{C,\mathrm{sym,sb}}_{n,0.95}$ and $\mathcal{R}^{\mathrm{sym,sb}}_{n,0.95}$ are constructed by the symmetrized smoothed bootstrap method [24], based respectively on the pivots $n^{1/3}(\hat{\theta}_n - \theta)$ and $n^{2/3}\{\hat{B}_n(\hat{\theta}_n) - \hat{B}_n(\theta)\}$, using a Gaussian kernel and a bandwidth set by the normal reference rule, $\mathcal{R}^{C,\mathrm{sb}}_{n,0.95}$ and $\mathcal{R}^{\mathrm{sb}}_{n,0.95}$ are constructed similarly by the smoothed bootstrap alone without symmetrization, and $\mathcal{R}^{C,\mathrm{sym}}_{n,0.95}$ and $\mathcal{R}^{\mathrm{sym}}_{n,0.95}$ are obtained by m out of n bootstrapping from the symmetrized empirical distribution without smoothing. Each interval is constructed using 500 bootstrap samples. The results are shown in Table 5 under different choices of m, with n set to be 100 and U_1 to be standard normal. Each coverage probability is estimated by averaging over 500 random samples. We see again that $\mathcal{R}_{n,0.95}$ is in general more accurate, and less sensitive to the choice of m, than $\mathcal{R}^{C}_{n,0.95}$. Coverage errors of both intervals are reduced if bootstrap samples are drawn from an empirical distribution modified by either symmetrization, kernel-smoothing or both symmetrization and kernel-smoothing. It is noteworthy that under either modification scheme, intervals constructed using our proposed pivot $n^{2/3}\{\hat{B}_n(\hat{\theta}_n) - \hat{B}_n(\theta)\}$ enjoy better coverages than those based on the conventional pivot $n^{1/3}(\hat{\theta}_n - \theta)$ in almost all cases.

5.4. *Ordered binomial probabilities*. It is well known that order restricted models often pose difficulties for standard inference procedures when the true parameter lies close to the boundary of the order restricted parameter space. In particular, the conventional bootstrap has been found to fail under the latter scenario; see, for example, [2]. Cohen and Sackrowitz [11] discuss several issues concerning order restricted inference in general. Construction of univariate confidence intervals for normal means under order restriction has been studied extensively; see [31] for a recent contribution. Multivariate confidence regions have, however, received relatively little attention. An exception is [25], who study a bootstrap method for constructing confidence regions for two ordered binomial probabilities. We take as our

last example the two-sample binomial model considered by [25], and compare our bootstrap regions with theirs.

Let $\mathcal{X}_n = \{X_i = (Y_i, Z_i) : i = 1, \ldots, n\}$ be independent and identically distributed observations in $\{0, 1\} \times \{1, 2\}$ such that $\mathrm{pr}(Z_i = 1) = \eta_F$ and $\mathrm{pr}(Y_i = 1 | Z_i = z) = \xi_F^{[z]}$, $z = 1, 2$. Thus $\mathcal{X}_n$ consists of two binomial samples of random sizes, with their underlying populations indexed by Z_i. Suppose that the two binomial probabilities $\xi_F = (\xi_F^{[1]}, \xi_F^{[2]})$ are order restricted with $\xi_F^{[1]} \le \xi_F^{[2]}$, so that $\Xi = \{(\xi^{[1]}, \xi^{[2]}) : 0 \le \xi^{[1]} \le \xi^{[2]} \le 1\}$. We consider three choices of θ_F: (i) $\theta_F = \xi_F$, (ii) $\theta_F = \xi_F^{[1]}$ and (iii) $\theta_F = \xi_F^{[2]}$; whereas $\eta_F \in \mathcal{H} = [0, 1]$ represents a nuisance parameter. Restricted maximum likelihood estimation of (ξ_F, η_F) amounts to setting $b_{\xi,\eta}(y, z) = z \log \eta + (1 - z) \log(1 - \eta) + y \log \xi^{[z]} + (1 - y) \log(1 - \xi^{[z]})$, for $(\xi, \eta) \in \Xi \times \mathcal{H}$. The corresponding M-estimator of ξ_F has an explicit expression given by

$$\begin{aligned}\hat{\xi}_n &= \big(\hat{\xi}_n^{[1]}, \hat{\xi}_n^{[2]}\big)^\top \\ &= \left(\min\left\{\frac{N_{11}}{N_{01} + N_{11}}, \frac{N_{11} + N_{12}}{n}\right\}, \max\left\{\frac{N_{12}}{N_{02} + N_{12}}, \frac{N_{11} + N_{12}}{n}\right\}\right)^\top,\end{aligned}$$

where $N_{ab} = \sum_{i=1}^n \mathbf{1}\{Y_i = a, Z_i = b\}$ for $a = 0, 1$ and $b = 1, 2$. For any $\theta \in [0, 1]$, the profile estimator $\tilde{\psi}_n(\theta)$ is given by $\tilde{\psi}_{n,j}(\theta) \equiv \max\{\theta, N_{1j}/(N_{0j} + N_{1j})\}$, where $j = 2$ and 1 under cases (ii) and (iii), respectively. The m out of n bootstrap estimators $\hat{\xi}_n^*$ and $\tilde{\psi}_n^*(\hat{\theta}_n)$ are defined similarly, with n replaced by m and N_{ab} replaced by $N_{ab}^* = \sum_{i=1}^m \mathbf{1}\{Y_i^* = a, Z_i^* = b\}$, where $\mathrm{pr}(Y_i^* = a, Z_i^* = b | \mathcal{X}_n) = N_{ab}/n$, for $a = 0, 1$ and $b = 1, 2$. The nuisance estimators $\hat{\eta}_n$ and $\hat{\eta}_n^*$ can be arbitrarily set, as they would have been eliminated during calculations of $\Pi_n(\theta)$ and $\Pi_n^*(\hat{\theta}_n)$. Asymptotic correctness of the confidence region $\mathcal{R}_{n,1-\alpha}$ then follows from Theorem 3.1, as elucidated in the following corollary.

COROLLARY 5.4. *Assume that $\eta_F \in (0, 1)$ and $\xi_F \in [\zeta, 1 - \zeta] \times [\zeta, 1 - \zeta]$ for some $\zeta \in (0, 1/2)$. Let W_1, W_2 denote independent standard normal random variables.*

For $\theta_F = \xi_F$, $2\Pi_n(\xi_F) = 2n\{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\xi_F)\}$ converges weakly to $W_1^2 + W_2^2$ if $\xi_F^{[1]} < \xi_F^{[2]}$, and to

$$\begin{aligned}&\big(\eta_F^{1/2} W_1 + (1 - \eta_F)^{1/2} W_2\big)^2 \mathbf{1}\big\{\eta_F^{-1/2} W_1 > (1 - \eta_F)^{-1/2} W_2\big\} \\ &\quad + \big(W_1^2 + W_2^2\big) \mathbf{1}\big\{\eta_F^{-1/2} W_1 \le (1 - \eta_F)^{-1/2} W_2\big\}\end{aligned}$$

if $\xi_F^{[1]} = \xi_F^{[2]}$.

For $\theta_F = \xi_F^{[b]}$, $b = 1, 2$, $2\Pi_n(\xi_F^{[b]}) = 2n\{\hat{B}_n(\hat{\xi}_n) - \hat{B}_n(\xi_F^{[b]}, \tilde{\psi}_{n,3-b}(\xi_F^{[b]}))\}$ converges weakly to W_b^2 if $\xi_F^{[1]} < \xi_F^{[2]}$, and to

$$\begin{aligned}&\big(\eta_F^{1/2} W_1 + (1 - \eta_F)^{1/2} W_2\big)^2 \mathbf{1}\big\{\eta_F^{-1/2} W_1 > (1 - \eta_F)^{-1/2} W_2\big\} \\ &\quad + \big(W_1^2 + W_2^2\big) \mathbf{1}\big\{\eta_F^{-1/2} W_1 \le (1 - \eta_F)^{-1/2} W_2\big\} - W_{3-b}^2 \mathbf{1}\big\{(-1)^b W_{3-b} < 0\big\}\end{aligned}$$

if $\xi_F^{[1]} = \xi_F^{[2]}$.

In all cases, $\Pi_n^*(\hat{\theta}_n)$ converges weakly in probability to the same limit as does $\Pi_n(\theta_F)$, provided that m satisfies (B1).

Note that the statistic $2\Pi_n(\theta_F)$ is nothing but the generalised likelihood ratio based on the restricted maximum likelihood estimator. The results given in Corollary 5.4 for the case $\xi_F^{[1]} < \xi_F^{[2]}$ reflect the classical large-sample likelihood theory under regularity conditions, and hold also with $m = n$. The asymptotic chi-squared limit changes to a very different form

when $\xi_F^{[1]} = \xi_F^{[2]}$, which poses problems for standard inference procedures. The m out of n bootstrap successfully circumvents such difficulties.

That $\Pi_n(\theta_F)$ can be expressed in terms of the N_{ab}'s renders it feasible to investigate into the error rate of the m out of n bootstrap. Consider for simplicity case (i) $\theta_F = \xi_F$. We prove in the Supplementary Material [22] the following result.

PROPOSITION 5.1. *Assume the conditions of Corollary 5.4. Then, for any fixed* $x \in \mathbb{R}$,

$$\mathrm{pr}\big(\Pi_n^*(\hat{\xi}_n) \le x | \mathcal{X}_n\big) = \mathrm{pr}_F\big(\Pi_n(\xi_F) \le x\big) + O_p\big(m^{-1/2} + (m/n)^{1/2}\mathbf{1}\{\xi_F^{[1]} = \xi_F^{[2]}\}\big).$$

Without prior knowledge of ξ_F, the error given in Proposition 5.1 attains a minimax order of $n^{-1/4}$ with the choice $m \propto n^{1/2}$. In an m out of n bootstrap application to construct Stein confidence sets, [10] propose a special iterative scheme to reduce coverage error from $O(n^{-1/4})$ to $O(n^{-1/3})$. Given the similarity in asymptotic properties between the two problems, we conjecture that a similar scheme can be adopted to improve the error rate obtained in Proposition 5.1, but refrain from pursuing it further in the present work.

Simulation study. As the nuisance parameter η_F does not play any role in inference about ξ_F, we consider $N_b = N_{0b} + N_{1b}$, $b = 1, 2$, fixed in the simulation study and apply the bootstrap separately to samples arising from the two different populations. Note that Corollary 5.4 holds here with $\eta_F = N_1/(N_1 + N_2) = N_1/n$.

Setting bootstrap sample sizes to be $M_1 = \eta_F m$ and $M_2 = (1 - \eta_F)m$ for the first and second populations, respectively, construction of $\mathcal{R}_{n,1-\alpha}$ is based on bootstrap observations $N_{1b}^* = M_b - N_{0b}^*$, which follow the binomial $(M_b, N_{1b}/N_b)$ distributions, for $b = 1, 2$. Table 6 shows the coverage probabilities of 95% confidence regions $\mathcal{R}_{n,0.95}$ for the parameter $\xi_F = (0.2, 0.2)^\top$ under different choices of m, with (n, η_F) set to be $(300, 1/3)$. Each region is constructed with 1000 bootstrap samples and its coverage estimated by averaging over 1000 independent replications. The coverages are in general accurate and rather insensitive to the choice of m, except for the cases of small m where $\mathcal{R}_{n,0.95}$ exhibits a slight degree of over-coverage.

We also compare $\mathcal{R}_{n,1-\alpha}$ with two different approaches to constructing parametric bootstrap confidence regions of the form $\{\theta : |\hat{\theta}_n - \theta| \le \hat{q}_{n,1-\alpha}\}$. The first approach calculates $\hat{q}_{n,1-\alpha}$ to be the $(1-\alpha)$th quantile of $|\hat{\theta}_n^* - \hat{\theta}_n^U|$, where $\hat{\theta}_n^*$ is derived from bootstrap observations $N_{1b}^* = N_b - N_{0b}^*$ generated from the binomial $(N_b, N_{1b}/N_b)$ distribution, $b = 1, 2$ and $\hat{\theta}_n^U$ denotes the unrestricted maximum likelihood estimator of θ_F calculated from $\mathcal{X}_n$. Li et al. [25] consider a similar approach with $\hat{\theta}_n^U$ replaced by $\hat{\theta}_n$, which is theoretically unsupported. The second approach calculates $\hat{q}_{n,1-\alpha}$ to be the $(1-\alpha)$th quantile of $|\hat{\theta}_n^* - \hat{\theta}_n^A|$, where $\hat{\theta}_n^*$

TABLE 6

Ordered binomial probability example—coverage probabilities of $\mathcal{R}_{n,0.95}$ *for* $n = 300$ *and* $\eta_F = 1/3$, *each estimated by averaging over* 1000 *random samples*

m	30	60	90	120	150
(M_1, M_2)	(10, 20)	(20, 40)	(30, 60)	(40, 80)	(50, 100)
Coverage	0.963	0.963	0.960	0.956	0.956
m	180	210	240	270	300
(M_1, M_2)	(60, 120)	(70, 140)	(80, 160)	(90, 180)	(100, 200)
Coverage	0.957	0.956	0.955	0.954	0.956

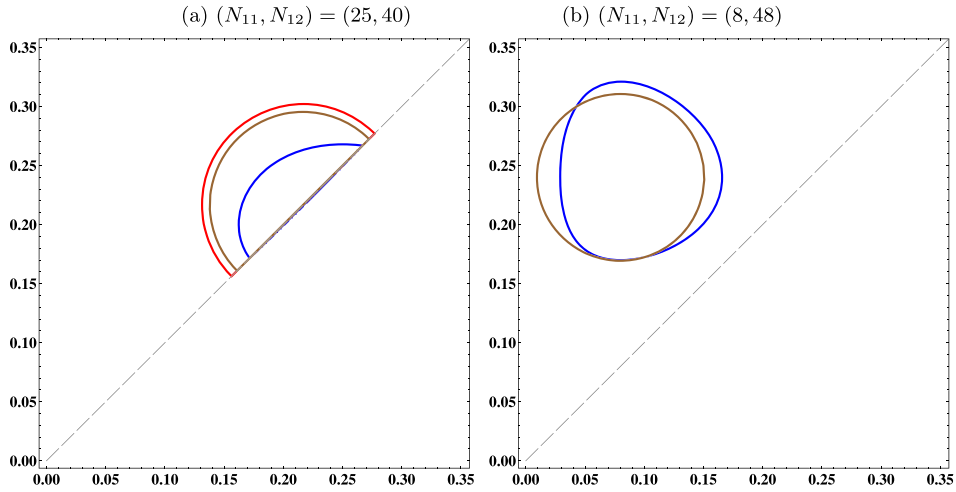

FIG. 3. *Ordered binomial probability example—95% confidence regions* $\mathcal{R}_{n,0.95}$ *(blue),* $\mathcal{R}^{E1}_{n,0.95}$ *(red) and* $\mathcal{R}^{E2}_{n,0.95}$ *(brown), based on two different random samples (a) and (b), with* $n = 300$ $(N_1 = 100, N_2 = 200)$, $\eta_F = 1/3$ *and* $m = 150$. *True value* $\xi_F = (0.2, 0.2)$.

is based on N^*_{1b} drawn from the binomial $(N_b, \hat{p}^A_b)$ distribution, and $\hat{\theta}^A_n = (\hat{p}^A_1, \hat{p}^A_2)^\top$ or $\hat{p}^A_b$ according as $\theta_F = \xi_F$ or $\xi^{[b]}_F$. Following [2], we define, for $b = 1, 2$, $\hat{p}^A_b$ to be N_{1b}/N_b if $N_{12}/N_2 - N_{11}/N_1 > 2\sqrt{\log\log n/n}$ and $(N_{11} + N_{12})/n$ otherwise, so that the two approaches coincide whenever $N_{12}/N_2 - N_{11}/N_1 > 2\sqrt{\log\log n/n}$. We denote by $\mathcal{R}^{E1}_{n,1-\alpha}$ and $\mathcal{R}^{E2}_{n,1-\alpha}$ the level $1-\alpha$ confidence regions constructed using the first and second approaches, respectively. It can be shown that $\mathcal{R}^{E1}_{n,1-\alpha}$ is asymptotically correct if $\xi^{[1]}_F < \xi^{[2]}_F$ but not if $\xi^{[1]}_F = \xi^{[2]}_F$, whereas $\mathcal{R}^{E2}_{n,1-\alpha}$ is asymptotically correct under either condition.

Consider first the case (i) $\theta_F = \xi_F$. Figure 3 displays the three bootstrap regions obtained from two different samples: (a) $(N_{11}, N_{12}) = (25, 40)$, and (b) $(N_{11}, N_{12}) = (8, 48)$, with $n = 300$, $m = 150$, $\eta_F = 1/3$ and $\xi_F = (0.2, 0.2)^\top$. Each region is constructed using 5000 bootstrap samples. The two regions $\mathcal{R}^{E1}_{n,0.95}$ and $\mathcal{R}^{E2}_{n,0.95}$ take on a circular shape due to their use of a Euclidean distance-based pivot. Sample (a) is typical of the setting $\xi_F = (0.2, 0.2)$, where $\mathcal{R}_{n,0.95}$ has a size far smaller than $\mathcal{R}^{E1}_{n,0.95}$ and $\mathcal{R}^{E2}_{n,0.95}$. Sample (b) represents a rare case under $\xi_F = (0.2, 0.2)$, in which $\mathcal{R}^{E1}_{n,0.95}$ and $\mathcal{R}^{E2}_{n,0.95}$ coincide with each other, and have a size comparable to that of $\mathcal{R}_{n,0.95}$. For a comparison in coverage accuracy, we construct $\mathcal{R}_{n,0.95}$, $\mathcal{R}^{E1}_{n,0.95}$ and $\mathcal{R}^{E2}_{n,0.95}$ under different settings of n and ξ_F, with η_F always set to be $1/3$ and $m = n/2$. Each region is constructed using 1000 bootstrap samples and each coverage probability estimated by averaging over 1000 replications. Table 7 summarises the coverage results, which suggest that $\mathcal{R}_{n,0.95}$ and $\mathcal{R}^{E2}_{n,0.95}$ have comparable coverage accuracies, while $\mathcal{R}^{E1}_{n,0.95}$ yields somewhat inferior results, especially when ξ_F assumes more extreme values. The above comparison is repeated for cases (ii) $\theta_F = \xi^{[1]}_F$ and (iii) $\theta_F = \xi^{[2]}_F$, with each coverage probability estimated by averaging over 100 replications. As before, we set $\eta_F = 1/3$, $m = n/2$ and construct each interval using 1000 bootstrap samples. The results, shown in Table 8, suggest that $\mathcal{R}_{n,0.95}$ yields the best, or almost the best, coverage accuracy among the three methods in all cases.

6. A subsampling method for selecting *m*. In their treatise on the subsampling method, [33] suggest several empirical approaches to fixing the subsample size. One of the approaches, which applies specifically to confidence interval construction, consists of min-

TABLE 7

Ordered binomial probability example—coverage probabilities of $\mathcal{R}_{n,0.95}$ $(m = n/2)$, $\mathcal{R}^{E1}_{n,0.95}$ *and* $\mathcal{R}^{E2}_{n,0.95}$, *each estimated by averaging over* 1000 *random samples, with* $\eta_F = 1/3$ *and parameter of interest* $\theta_F = \xi_F$

	$n = 150$			$n = 300$		
(i) $\theta_F = \xi_F$	(0.2, 0.2)	(0.5, 0.5)	(0.8, 0.8)	(0.2, 0.2)	(0.5, 0.5)	(0.8, 0.8)
$\mathcal{R}_{n,0.95}$ $(m = n/2)$	0.949	0.944	0.949	0.947	0.952	0.952
$\mathcal{R}^{E1}_{n,0.95}$	0.933	0.954	0.960	0.939	0.955	0.969
$\mathcal{R}^{E2}_{n,0.95}$	0.944	0.950	0.956	0.936	0.949	0.945

imising the volatility of the interval end points. This method extends naturally to our d-dimensional m out of n bootstrap confidence region $\mathcal{R}_{n,1-\alpha}$ as follows. For each possible value of m, the volatility of $\mathcal{R}_{n,1-\alpha}$ is calculated to be the running standard deviation of the $2k+1$ values of $\hat{G}_n^{-1}(1-\alpha)$, obtained using bootstrap sample sizes $m-k, m-k+1, \ldots, m+k$ respectively. The value of m which minimises this volatility is then chosen. Assuming validity of Edgeworth expansions of a certain type, [14] suggest estimating m by minimising some distance measure between a pair of m out of n bootstrap distribution estimators based on bootstrap sample sizes m and $m/2$, for even m. Bickel and Sakov [4] propose a similar procedure in which the pair of estimators are obtained from bootstrap sample sizes $q^j n$ and $q^{j+1}n$, for a fixed $q \in (0, 1)$ and $j = 0, 1, \ldots$.

We propose below an alternative, theoretically better justified, subsampling approach to selecting m. First, fix $T \geq 2$ distinct subsample sizes $n_1 < \cdots < n_T < n$. Generate a number B_1 of subsamples, each of size n_T drawn without replacement from $\mathcal{X}_n$. For each $t = 1, \ldots, T$ and $b_1 = 1, \ldots, B_1$, denote by $\mathcal{X}_t^{\dagger(b_1)}$ the set of the first n_t observations in the b_1th subsample. Similarly, generate from each $\mathcal{X}_t^{\dagger(b_1)}$ B_2 bootstrap samples, each of size $\lceil n_t/2 \rceil$, and denote by $\mathcal{X}_{t,m}^{\dagger *(b_1,b_2)}$ the set of the first m observations in the b_2th bootstrap sample, $b_2 = 1, \ldots, B_2$. Write $\Pi_t^{\dagger(b_1)}(\theta)$ and $\Pi_{t,m}^{\dagger *(b_1,b_2)}(\theta)$ for the analogues of $\Pi_n(\theta)$ and $\Pi_n^*(\theta)$, respectively, with $(\mathcal{X}_n, \mathcal{X}_m^*)$ replaced by $(\mathcal{X}_t^{\dagger(b_1)}, \mathcal{X}_{t,m}^{\dagger *(b_1,b_2)})$. Denote by $\hat{\xi}_t^{\dagger(b_1)} = (\hat{\theta}_t^{\dagger(b_1)}, \hat{\psi}_t^{\dagger(b_1)})$ the M-estimator based on $\mathcal{X}_t^{\dagger(b_1)}$. We then estimate the coverage error of the m out of n_t bootstrap

TABLE 8

Ordered binomial probability example—coverage probabilities of $\mathcal{R}_{n,0.95}$ $(m = n/2)$, $\mathcal{R}^{E1}_{n,0.95}$ *and* $\mathcal{R}^{E2}_{n,0.95}$, *each estimated by averaging over* 100 *random samples, with* $\eta_F = 1/3$ *and parameter of interest* θ_F

	$n = 60$		$n = 150$	
ξ_F	(0.5, 0.5)	(0.8, 0.8)	(0.5, 0.5)	(0.8, 0.8)
(ii) $\theta_F = \xi_F^{[1]}$:				
$\mathcal{R}_{n,0.95}$ $(m = n/2)$	0.95	0.97	0.96	0.94
$\mathcal{R}^{E1}_{n,0.95}$	0.98	0.98	0.98	0.96
$\mathcal{R}^{E2}_{n,0.95}$	0.98	0.98	0.99	0.96
(iii) $\theta_F = \xi_F^{[2]}$:				
$\mathcal{R}_{n,0.95}$ $(m = n/2)$	0.96	0.93	0.95	0.96
$\mathcal{R}^{E1}_{n,0.95}$	0.95	0.92	0.96	0.90
$\mathcal{R}^{E2}_{n,0.95}$	0.95	0.92	0.96	0.91

confidence region by

$$CE_t(m) = \Bigg| B_1^{-1} \sum_{b_1=1}^{B_1} \mathbf{1}\Bigg\{ B_2^{-1} \sum_{b_2=1}^{B_2} \mathbf{1}\big\{ \Pi_{t,m}^{\dagger *(b_1,b_2)}\big(\hat{\theta}_t^{\dagger(b_1)}\big) \le \Pi_t^{\dagger(b_1)}(\hat{\theta}_n) \big\} \le 1-\alpha \Bigg\} - (1-\alpha) \Bigg|,$$

for $m = 1, \dots, n_t$ and $t = 1, \dots, T$. Select $m = m_t$ which minimises $CE_t(m)$ over $m \in \{1, \dots, n_t\}$. Fitting the optimal choices m_t by a parametric form Ωn_t^{ω}, for $\Omega > 0$ and $\omega \in (0,1)$ independent of n, we obtain by standard least squares approximation that $\omega \approx \max\{0, \min\{1, (TL_2 - M_1L_1)/(TM_2 - M_1^2)\}\}$ and $\Omega \approx \exp\{(L_1 - \omega M_1)/T\}$, where $M_1 = \sum_{t=1}^T \log n_t$, $M_2 = \sum_{t=1}^T (\log n_t)^2$, $L_1 = \sum_{t=1}^T \log m_t$ and $L_2 = \sum_{t=1}^T (\log n_t)(\log m_t)$. Extrapolating the above results to the full sample, we calculate the optimal m to be $\min\{\max\{\Omega n^{\omega}, m^{\circ}\}, n\}$, rounded to the nearest integer, for some reasonably small lower bound $m^{\circ} < n$. The following proposition states conditions under which $CE_t(m)$ consistently estimates the coverage error of the m out of n_t bootstrap confidence region, thus justifying theoretically the above selection procedure.

PROPOSITION 6.1. *Assume the conditions of Theorem 3.1 and further that*

$$\text{(G3)} \quad \begin{cases} \tilde{\psi}_n(\vartheta) - \tilde{\psi}_n(\theta_F) = O_p\big(|\vartheta - \theta_F|\big), \\ \mathbb{E}_F\big[b_{(\theta,\psi),\eta}(X_1) - b_{(\theta_F,\psi'),\eta}(X_1)\big] = O\big(r_n^{-\nu}\big), \\ \mathbb{E}_F\big\{b_{(\theta,\psi),\eta}(X_1) - b_{(\theta_F,\psi'),\eta}(X_1)\big\}\big\{b_{(\theta',\psi''),\eta}(X_1) - b_{(\theta_F,\psi'''),\eta}(X_1)\big\} \\ \qquad = O\big(nr_n^{-2\nu}\big), \end{cases}$$

uniformly over ϑ in a neighbourhood of θ_F, (θ, ψ), (θ_F, ψ'), (θ', ψ''), (θ_F, ψ''') in a neighbourhood of (θ_F, ψ_F) and η in a neighbourhood of η_F, with $|\theta - \theta_F| + |\theta' - \theta_F| + |\psi - \psi'| + |\psi'' - \psi'''| = O(r_n^{-1})$. Assume that $B_1, B_2 \to \infty$ and $n_1^{-1} + n_T/n = o(1)$. Then we have, for $t = 1, \dots, T$,

$$CE_t(m) = \big|\mathrm{pr}_F(\theta_F \in \mathcal{R}_{n_t,1-\alpha}) - (1-\alpha)\big| + O_p\big\{(n/n_t)^{1/2}(r_{n_t}/r_n)^{\nu} + n^{-1/2} + B_1^{-1/2} + B_2^{-1/2}\big\}.$$

Condition (G3) requires that the profile M-estimator $\tilde{\psi}_n(\vartheta)$ depend sufficiently smoothly on ϑ in a neighbourhood of θ_F, and extends validity of the asymptotic orders prescribed in (G1) and (G2) from the fixed points $\psi' = \psi''' = \psi_F$ and $\eta = \eta_F$ to small neighbourhoods around them. Assuming that $\mathcal{R}_{n,1-\alpha}$ has an optimal coverage error of order not smaller than $n^{-1/2}$, as is typically the case under nonstandard conditions, Proposition 6.1 suggests that $CE_t(m)$ consistently estimates the coverage error of $\mathcal{R}_{n_t,1-\alpha}$, that is,

$$CE_t(m) = \mathrm{pr}_F(\theta_F \in \mathcal{R}_{n_t,1-\alpha}) - (1-\alpha) + o_p\big(n_t^{-1/2}\big), \qquad t = 1, \dots, T,$$

if we set $n_t = o(\min\{B_1, B_2\})$ and $(n/n_t)^{1/2}(r_{n_t}/r_n)^{\nu} = o(n_t^{-1/2})$ for each $t = 1, \dots, T$. The latter conditions are equivalent to setting

$$n_T = o\big(\min\{B_1, B_2\}\big) \qquad \text{and} \qquad r_{n_T} = o\big(n^{-1/(2\nu)} r_n\big). \tag{6.1}$$

Applications of (6.1) to the examples discussed in Section 5 suggest the choices $n_T = o(n^{1/4})$ for maximum score estimation (Section 5.1) and least quantile of squares estimation (Section 5.3), $n_T = o(n^{1/(\zeta+1)})$ for L_1 regression (Section 5.2) and $n_T = o(n^{1/2})$ for ordered binomial probability estimation (Section 5.4), provided that B_1 and B_2 have orders exceeding n_T in each case.

REMARK 6.1. Writing $m_{\mathrm{opt}}(n)$ for the theoretically optimal choice of m that minimises the coverage error, we have proposed approximating $m_{\mathrm{opt}}(n)$ by a parametric fit Ωn^{ω}, estimation of which requires at least $T = 2$. A rationale behind the form Ωn^{ω} lies in the notion that the coverage error of an m out of n bootstrap confidence region is typically dominated by terms of orders m^{-c_1} and $m^{c_2}n^{-c_3}$, for some $c_1, c_2, c_3 > 0$, leading to an optimal m of order $n^{c_3/(c_1+c_2)}$; see Proposition 5.1 for an example. At extra computational cost we may consider increasing T and approximating $m_{\mathrm{opt}}(n)$ by a nonparametric fit instead, an approach which we do not pursue in the present work.

REMARK 6.2. Since n_T satisfying (6.1) is typically small relative to n, our proposed subsampling method requires only moderate simulation sizes B_1 and B_2.

REMARK 6.3. In the rare event of $\mathcal{R}_{n,1-\alpha}$ having an optimal coverage error of order $o(n^{-1/2})$, which is atypically small, the optimal m selected by the subsampling method under (6.1) may not yield the best error rate but still guarantees a coverage error of order $O(n^{-1/2})$.

REMARK 6.4. Minimisation of $CE_t(m)$ over $m \in \{1, \ldots, n_t\}$, at each $t = 1, \ldots, T$, requires $\sum_{t=1}^{T} n_t$ evaluations of $CE_t(m)$, which may be computationally very costly. A more efficient approach is to evaluate $CE_t(m)$ for m selected from a small subset $\mathcal{M}_t \subset \{1, \ldots, n_t\}$ likely to contain the optimal m_t, and approximate m_t by minimising a quadratic interpolation of the points $\{(m, CE_t(m)) : m \in \mathcal{M}_t\}$ over the real interval $[1, \max\{m' : m' \in \mathcal{M}_t\}]$.

Simulation study. For an empirical illustration, we repeat the simulation study in Section 5.1, with m selected by our proposed subsampling method. We set $T = 2$, $n_1 = 16$, $n_2 = 48$, $B_1 = 100$, $B_2 = 50$ and $m^{\circ} = 5$. For $t = 1, 2$, $CE_t(m)$ is minimised approximately by quadratic interpolation as described in Remark 6.4, with $\mathcal{M}_1 = \{4, 5, 6, 7, 8, 9\}$ and $\mathcal{M}_2 = \{4, 6, 8, 10, 16, 25\}$. The coverage probability, estimated by averaging over 200 replications, is found to be 0.955, which is comparable to the best result achieved by fixing m at different values shown in Table 1. Figure 4(a) shows the histogram of the 200 empirically selected values of m, which have a mean value of 48.275 and a standard deviation of 87.096. Among the 200 replications, m is selected to be the lower bound m° in 135 cases. Generally speaking, our subsampling method favours use of a small m, whose values accord well with the optimal range suggested by fixed-m results shown in Table 1.

We also apply the subsampling method to select m in the simulation study reported in Table 6 for the ordered binomial probability example. Here, we set $T = 2$, $n_1 = 10$, $n_2 = 30$, $B_1 = B_2 = 100$, $m^{\circ} = 6$, $\mathcal{M}_1 = \{3, 4, 5, 6, 7, 8\}$ and $\mathcal{M}_2 = \{4, 6, 8, 11, 15, 20\}$. The coverage probability is estimated to be 0.950 based on 1000 replications, which outperforms all the fixed-m results shown in Table 6. The 1000 selected values of m have mean 27.642 and standard deviation 41.135, and take on the lower bound m° 535 times. Their histogram, plotted in Figure 4(b), shows a trend similar to that observed in Figure 4(a) for the maximum score example.

7. Unknown convergence rate. Practical implementation of our proposed confidence procedure presupposes some working knowledge of the normalising factor r_n^{ν} as a function of n, typically derived from the conditions (G1) and (G2). This has been shown to be analytically available in the examples considered in Section 5. However, situations may arise where the criterion function has a structure too complex to permit tractable analysis of r_n^{ν}, or where r_n and ν involve some unknown features of the underlying distribution F. In such cases, one may resort to empirical methods for estimating the values of r_n and ν. One possible approach, suggested by [33], Section 8.2, is as follows. For distinct ρ_1, ρ_2 in $(0, 1)$, calculate the $(1 - \alpha)$th

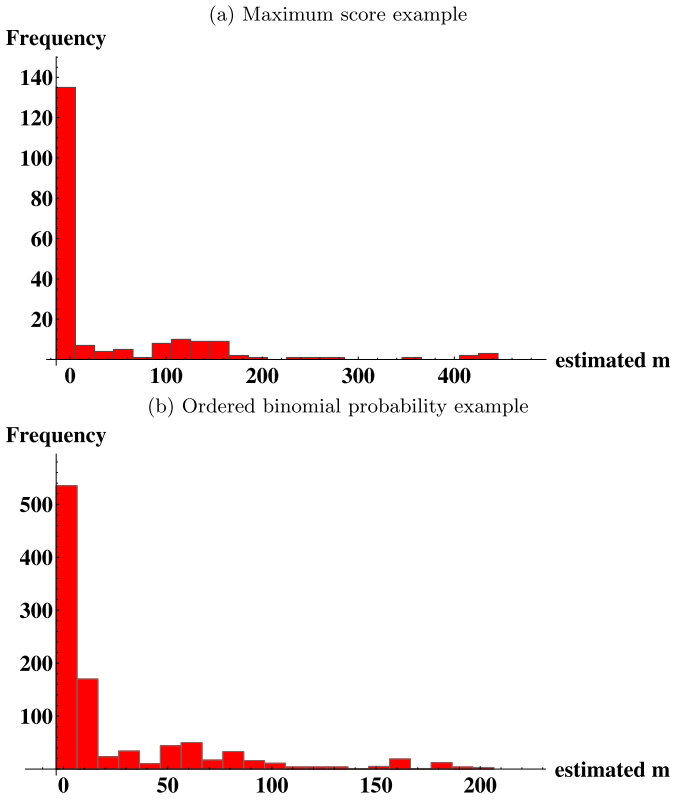

FIG. 4. *Histograms of bootstrap sample size m selected by subsampling method in simulation studies conducted for examples* (*a*) *maximum score estimation* ($n = 1000$), *and* (*b*) *ordered binomial probability estimation* ($n = 300$).

bootstrap quantiles of $\hat{B}_n^*(\hat{\xi}_n^*) - \hat{B}_n^*(\hat{\theta}_n, \tilde{\psi}_n^*(\theta))$ based on bootstrap sample sizes $m_1 = n^{\rho_1}$ and $m_2 = n^{\rho_2}$, respectively, and denote their values by Q_1 and Q_2. Then estimate the normalising factors r_n^ν and r_m^ν by $n^{\hat{\lambda}}$ and $m^{\hat{\lambda}}$, respectively, where $\hat{\lambda} = \log(Q_1/Q_2)/\log(m_2/m_1)$. It can be shown, under the assumption that $r_n^\nu = n^\lambda$ for some $\lambda > 0$, the above procedure yields a consistent estimator $\hat{\lambda} = \lambda + o_p(1/\log n)$. It follows that $n^{\hat{\lambda}} = n^\lambda\{1 + o_p(1)\}$ and $m^{\hat{\lambda}} = m^\lambda\{1 + o_p(1)\}$, so that Theorem 3.1 remains valid with λ replaced by $\hat{\lambda}$. Similarly, noting that $n_t^{\hat{\lambda}} = n_t^\lambda\{1 + o_p(1)\}$, the subsampling procedure described in Section 6 for selecting m retains its theoretical properties with λ estimated by $\hat{\lambda}$, while the second condition in (6.1) suggests the choice $n_T = o(n^{1-1/(2\hat{\lambda})})$.

8. Conclusion. We have proposed a novel m out of n bootstrap procedure for constructing confidence regions under nonstandard M-estimation settings. By using a univariate pivot defined in terms of the maximised criterion function, our procedure is proved to be consistent, applies readily to multidimensional parameters, and is rid of anomalies typical of conventional pivots based directly on multivariate M-estimators. Simulation results show that our procedure improves upon conventional methods in coverage accuracy. Interestingly, the improvement remains noticeable even after the bootstrap scheme has been modified by smoothing or symmetrization, techniques which have been proposed in the literature in selected contexts. The need for selecting m motivates us to develop a subsampling procedure for consistently estimating the coverage error under different subsample sizes, based on which some kind of extrapolation can be applied to yield an empirical choice of m. The procedure has been tested via simulations, generating very encouraging results.

When developing our procedure, we have made a critical assumption (A1) that the nuisance estimator $\hat{\eta}_n$ converges at a faster rate than does the M-estimator. However, it is not uncommon to find in a semiparametric M-estimation context an infinite-dimensional η_F which can only be estimated at a rate slower than the Euclidean parameter of interest ξ_F. In the latter case and under standard regularity conditions, $\hat{\xi}_n$ can be shown to be $n^{1/2}$-consistent and asymptotically normal, thus allowing for applications of the standard bootstrap [9] or the perturbation bootstrap [27]. Advances have also been made in problems concerning high-dimensional ξ_F. Spokoiny and Zhilova [35] propose a multiplier bootstrap confidence region for ξ_F of growing dimension $p = o(n^{1/3})$, based on a log-likelihood ratio pivot akin to ours but still subject to standard regularity conditions. Chatterjee and Bose [8] consider asymptotically normal M-estimators, defined as solutions to differentiable estimating equations, and establish consistency for a general class of bootstrap methods with d fixed and p subject to a growing rate which amounts to $o(n)$ under regularity conditions. For high-dimensional regression problems under sparsity assumptions, a popular strategy is provided by penalised M-estimation of the lasso type, for which a variety of bootstrap methods have been developed to draw inference for the regression parameter subvector θ_F when either d or p increases with n [7, 12, 26, 40, 41]. It would be of interest to extend our m out of n bootstrap procedure to the above contexts under nonstandard conditions, when asymptotic normality no longer holds for the M-estimator $\hat{\xi}_n$ and when p or q increases with n.

Acknowledgments. The first author was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 17303715).

SUPPLEMENTARY MATERIAL

Supplement to "Bootstrap confidence regions based on M-estimators under nonstandard conditions" (DOI: 10.1214/18-AOS1803SUPP; .pdf). The supplement contains proofs of Theorems 3.1–3.3, 4.1, Corollaries 5.1–5.4, Propositions 5.1 and 6.1.

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. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009 https://doi.org/10.1111/1468-0262.00114

[3] BERAN, R. (1988). Balanced simultaneous confidence sets. *J. Amer. Statist. Assoc.* **83** 679–686. MR0963795

[4] BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist. Sinica* **18** 967–985. MR2440400

[5] BOSE, A. and CHATTERJEE, S. (2001). Generalised bootstrap in non-regular M-estimation problems. *Statist. Probab. Lett.* **55** 319–328. MR1867535 https://doi.org/10.1016/S0167-7152(01)00161-4

[6] CHANG, C. and TODD OGDEN, R. (2009). Bootstrapping sums of independent but not identically distributed continuous processes with applications to functional data. *J. Multivariate Anal.* **100** 1291–1303. MR2508388 https://doi.org/10.1016/j.jmva.2008.11.008

[7] CHATTERJEE, A. and LAHIRI, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41** 1232–1259. MR3113809 https://doi.org/10.1214/13-AOS1106

[8] CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808 https://doi.org/10.1214/009053604000000904

[9] CHENG, G. and HUANG, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38** 2884–2915. MR2722459 https://doi.org/10.1214/10-AOS809

[10] CHEUNG, K. Y., LEE, S. M. S. and YOUNG, G. A. (2006). Stein confidence sets based on non-iterated and iterated parametric bootstraps. *Statist. Sinica* **16** 45–75. MR2256079

[11] COHEN, A. and SACKROWITZ, H. B. (2004). A discussion of some inference issues in order restricted models. *Canad. J. Statist.* **32** 199–205. MR2064402 https://doi.org/10.2307/3315943

[12] DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2

[13] GHOSH, S. and POLANSKY, A. M. (2014). Smoothed and iterated bootstrap confidence regions for parameter vectors. *J. Multivariate Anal.* **132** 171–182. MR3266268 https://doi.org/10.1016/j.jmva.2014.08.003

[14] GÖTZE, F. and RAČKAUSKAS, A. (2001). Adaptive choice of bootstrap sample sizes. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 286–309. IMS, Beachwood, OH. MR1836566 https://doi.org/10.1214/lnms/1215090074

[15] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[16] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068 https://doi.org/10.1214/07-AOS555

[17] JIN, Z., YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. MR1844838 https://doi.org/10.1093/biomet/88.2.381

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

[19] KOSOROK, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84** 299–318. MR1965224 https://doi.org/10.1016/S0047-259X(02)00040-4

[20] LAHIRI, S. N. (1992). On bootstrapping M-estimators. *Sankhyā Ser. A* **54** 157–170. MR1192092

[21] LAI, P. Y. and LEE, S. M. S. (2005). An overview of asymptotic properties of L_p regression under general classes of error distributions. *J. Amer. Statist. Assoc.* **100** 446–458. MR2160549 https://doi.org/10.1198/016214504000001385

[22] LEE, S. M. S. and YANG, P. (2019). Supplement to "Bootstrap confidence regions based on M-estimators under nonstandard conditions." https://doi.org/10.1214/18-AOS1803SUPP.

[23] LEE, S. M. S. and PUN, M. C. (2006). On m out of n bootstrapping for nonstandard M-estimation with nuisance parameters. *J. Amer. Statist. Assoc.* **101** 1185–1197. MR2328306 https://doi.org/10.1198/016214506000000014

[24] LÉGER, C. and MACGIBBON, B. (2006). On the bootstrap in cube root asymptotics. *Canad. J. Statist.* **34** 29–44. MR2267708 https://doi.org/10.1002/cjs.5550340104

[25] LI, Z., TAYLOR, J. M. G. and NAN, B. (2010). Construction of confidence intervals and regions for ordered binomial probabilities. *Amer. Statist.* **64** 291–298. MR2758560 https://doi.org/10.1198/tast.2010.09096

[26] LIU, H. and YU, B. (2013). Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression. *Electron. J. Stat.* **7** 3124–3169. MR3151764 https://doi.org/10.1214/14-EJS875

[27] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96** 190–217. MR2202406 https://doi.org/10.1016/j.jmva.2004.09.008

[28] 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

[29] MEERSCHAERT, M. M. (1988). Regular variation in $\mathbf{R}^k$. *Proc. Amer. Math. Soc.* **102** 341–348. MR0920997 https://doi.org/10.2307/2045886

[30] MUKHOPADHYAY, N. D. and CHATTERJEE, S. (2011). High dimensional data analysis using multivariate generalized spatial quantiles. *J. Multivariate Anal.* **102** 768–780. MR2772334 https://doi.org/10.1016/j.jmva.2010.12.002

[31] PARK, Y., KALBFLEISCH, J. D. and TAYLOR, J. M. G. (2014). Confidence intervals under order restrictions. *Statist. Sinica* **24** 429–445. MR3183692

[32] PATRA, R. K., SEIJO, E. and SEN, B. (2018). A consistent bootstrap procedure for the maximum score estimator. *J. Econometrics* **205** 488–507. MR3813528 https://doi.org/10.1016/j.jeconom.2018.04.001

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

[34] ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. MR0770281

[35] SPOKOINY, V. and ZHILOVA, M. (2015). Bootstrap confidence sets under model misspecification. *Ann. Statist.* **43** 2653–2675. MR3405607 https://doi.org/10.1214/15-AOS1355

[36] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

[38] WEI, B. and LEE, S. M. S. (2012). Second-order accuracy of depth-based bootstrap confidence regions. *J. Multivariate Anal.* **105** 112–123. MR2877506 https://doi.org/10.1016/j.jmva.2011.08.016

[39] YEH, A. B. and SINGH, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B* **59** 639–652. MR1452031 https://doi.org/10.1111/1467-9868.00088

[40] ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

[41] ZHOU, Q. and MIN, S. (2017). Uncertainty quantification under group sparsity. *Biometrika* **104** 613–632. MR3694586 https://doi.org/10.1093/biomet/asx037