

Prepivoting by weighted bootstrap iteration

BY STEPHEN M. S. LEE

*Department of Statistics and Actuarial Science, The University of Hong Kong,
Pokfulam Road, Hong Kong
smslee@khusua.hku.hk*

AND G. ALASTAIR YOUNG

*Statistical Laboratory, University of Cambridge, Wilberforce Road,
Cambridge CB3 0WB, U.K.
g.a.young@statslab.cam.ac.uk*

SUMMARY

Prepivoting by conventional bootstrap iteration is known to yield a progressively more accurate pivot in certain problems, and has important application in the construction of confidence limits and estimation of null distributions. We investigate the theoretical effects of weighted bootstrap iteration on prepivoting and show that each weighted bootstrap iteration, with weights chosen carefully but empirically, is asymptotically equivalent to two consecutive conventional bootstrap iterations. In terms of reducing the order of error, prepivoting can therefore be carried out much more efficiently if based on weighted bootstrap iterations. This is shown for a variety of problem settings, including the smooth function model, M -estimation and the regression context. A numerical illustration is provided, demonstrating the potential practical usefulness of weighted prepivoting.

Some key words: Bootstrap iteration; Linear regression; M -estimation; Pivot; Prepivoting; Smooth function model; Weighted bootstrap iteration.

1. INTRODUCTION

The iterated bootstrap (Hall, 1986; Beran, 1987) provides a satisfactory theoretical solution to the problem of reducing the error in bootstrap methods of inference. A general theory of bootstrap iteration is given by Hall & Martin (1988), who show that successive iterations produce successive reductions in the order of error of the bootstrap procedure. In practice, however, iterated bootstrap procedures generally require a computationally expensive Monte Carlo simulation, involving nested levels of bootstrap sampling from data, which restricts the practical feasibility of higher levels of iteration, and demands analysis of whether or not the benefits of higher level iteration are obtainable without prohibitive computational expense. A number of authors have discussed issues relating to computationally efficient implementation of bootstrap iteration; see, among others, DiCiccio et al. (1992), Lee & Young (1995, 1999), Hall et al. (2000), Chan & Lee (2001) and references therein.

In this paper we establish that the theoretical effects of bootstrap iteration can be accelerated, using weighted bootstrap techniques in place of conventional bootstrap sam-

pling. To be specific, we reconsider a formulation of bootstrap iteration, due to Beran (1987, 1988), known as pre-pivoting, and show that pre-pivoting using weighted bootstrap sampling enhances the effect of iteration on the reduction of error of bootstrap techniques.

Pre-pivoting has been proposed as a unified strategy for transforming a function $R_n(\mathcal{X}, \theta)$ of a sample \mathcal{X} and an unknown parameter θ into an approximate pivot, or, more precisely, an approximately $\text{Un}(0, 1)$ random variable. The idea can be iterated and each iteration brings the transformed R_n closer to a pivot by an order of $O(n^{-1/2})$ in regular settings.

Pre-pivoting has found important applications in reducing the coverage error of bootstrap confidence sets when R_n takes the form of a confidence set root, and in reducing the error of the rejection probability when R_n is a test statistic. In the former application, we suppose a one-sided bootstrap confidence set for θ of nominal coverage α is $\{\gamma: R_n(\mathcal{X}, \gamma) \leq \alpha\}$ or $\{\gamma: R_n(\mathcal{X}, \gamma) \geq 1 - \alpha\}$, depending on the side for which the confidence set is intended; examples are given in §§ 3.2–3.5. Expressed in these terms, it is clear that if the sampling distribution of $R_n(\mathcal{X}, \theta)$ were exactly $\text{Un}(0, 1)$ the bootstrap confidence set would have coverage exactly equal to the nominal desired coverage. When we are testing the hypothesis $H_0: \theta = \gamma$, the hypothesis is rejected in a one-sided test of nominal size α if $R_n(\mathcal{X}, \gamma) \leq \alpha$, or if $R_n(\mathcal{X}, \gamma) \geq 1 - \alpha$. If $R_n(\mathcal{X}, \theta)$ were $\text{Un}(0, 1)$, the rejection probability would be exactly α , under H_0 .

The conventional pre-pivoting procedure may be described as follows. Let \mathcal{X} be a sample drawn from a distribution F , and let $\eta(F)$ be an \mathbb{R}^k -valued functional of F . The parameter of interest θ is assumed to be $\theta = g\{\eta(F)\}$, for some real function g on \mathbb{R}^k . For each possible value ξ of $\eta(\cdot)$, derive a sequence of n vector-valued functions $\mathcal{Z} = (z_1(\mathcal{X}, \xi), \dots, z_n(\mathcal{X}, \xi))$, where each z_i has values in \mathbb{R}^d . Bootstrap sampling is performed on this ‘derived sample’ \mathcal{Z} . Precise definitions of the z_i depend on the context of the problem, but in conventional bootstrapping ξ is typically taken as an estimate of $\eta(F)$. In the simplest case, where $\mathcal{X} = (X_1, \dots, X_n)$ represents a sample of independent and identically distributed data, we may define $z_i(\mathcal{X}, \xi) = X_i$ so that $\mathcal{X} = \mathcal{Z}$. The $z_i(\mathcal{X}, \xi)$ need not be independent in general. For example, in linear regression, $z_i(\mathcal{X}, \xi)$ may be defined as the i th centred residual, in which case ξ could be taken as the least squares estimates of the regression coefficients. In our weighted bootstrap pre-pivoting procedure to be introduced later, ξ could be taken as some hypothesised or pilot values of $\eta(F)$, and their explicit roles will be elucidated by examples in §§ 3–5. Let $\hat{\theta} = \hat{\theta}(\mathcal{X})$ be an estimator of θ . Denote by \mathcal{Z}^* a bootstrap sample formed by sampling with replacement from \mathcal{Z} , and by \mathcal{X}^* the bootstrap version of \mathcal{X} reconstructed from \mathcal{Z}^* . For each possible value γ of θ , write $R_{n,0}(\mathcal{X}, \gamma) = R_n(\mathcal{X}, \gamma)$. Denote by $R_{n,j}(\mathcal{X}, \gamma)$ the transformation of $R_n(\mathcal{X}, \gamma)$ after j iterations of pre-pivoting. Let \hat{G}_j be the bootstrap cumulative distribution function of $R_{n,j}(\mathcal{X}, \theta)$, so that

$$\hat{G}_j(x) = \mathbb{P}^*\{R_{n,j}(\mathcal{X}^*, \hat{\theta}) \leq x\},$$

where \mathbb{P}^* denotes the probability, given \mathcal{X} , under the drawing of bootstrap samples \mathcal{X}^* . Then pre-pivoting $R_{n,j}$ yields $R_{n,j+1}(\mathcal{X}, \gamma) = \hat{G}_j\{R_{n,j}(\mathcal{X}, \gamma)\}$, for $j = 0, 1, 2, \dots$.

We remark in passing that, if it were possible to use the true distribution function of $R_{n,j}(\mathcal{X}, \theta)$ in place of its bootstrap estimator \hat{G}_j , then $R_{n,j+1}(\mathcal{X}, \theta)$ would be exactly distributed as $\text{Un}(0, 1)$ under appropriate continuity conditions. Note also that construction of $R_{n,j+1}$ will generally require nested levels of bootstrap sampling from \mathcal{X} ; illustration is provided in §§ 3–5.

It can be shown that in regular situations, if $R_n(\mathcal{X}, \theta)$ is asymptotically pivotal to order $O(n^{-j_0/2})$, then $R_{n,j}(\mathcal{X}, \theta)$ is asymptotically pivotal to $O(n^{-(j_0+j)/2})$. In other words,

$R_{n,j}(\mathcal{X}, \theta)$ differs from $\text{Un}(0, 1)$ by an order of $O(n^{-(j_0+j)/2})$. Each iteration step has the effect of reducing the error by $O(n^{-1/2})$; see Hall & Martin (1988) for a detailed discussion.

Recently, there has been much interest in procedures of bootstrap inference which replace uniform resampling from \mathcal{Z} by weighted sampling, in which unequal weights are placed on the elements $z_i(\mathcal{X}, \xi)$ of \mathcal{Z} ; see, for example, Hall & Presnell (1999a, b) and Chuang & Lai (2000). These techniques, in which bootstrap samples are drawn under non-uniform, but fixed, weights on the observed data points, may be contrasted with weighted bootstrap techniques discussed by Barbe & Bertail (1995). The latter are procedures based on the heuristic of randomly reweighting each observation in the sample.

Our primary contribution in this paper is to investigate the effect of using weighted resampling instead of uniform resampling on the properties of bootstrap prepivoting. We provide a general account of techniques by which the effectiveness of bootstrap prepivoting can be enhanced. In particular, we provide a detailed analysis of the effects of prepivoting in general settings of practical importance, such as those arising in regression problems and M -estimation.

Our resampling frame permits the functions z_i to depend on \mathcal{X} as well as on a hypothesised or pilot value $\gamma = g(\xi)$ through some k -vector ξ . Typically γ is set to the hypothesised value of θ in testing problems or taken from a number of trial values of θ when constructing confidence sets for θ . A weighted bootstrap sample \mathcal{X}^\dagger is reconstructed from $\mathcal{Z}^\dagger = (Z_1^\dagger, \dots, Z_n^\dagger)$, where each Z_i^\dagger is independently drawn from $\mathcal{Z} = (z_1(\mathcal{X}, \xi), \dots, z_n(\mathcal{X}, \xi))$ such that $Z_i^\dagger = z_j(\mathcal{X}, \xi)$ with probability p_j . Here $p = (p_1, \dots, p_n)$ is a vector of sampling weights such that $\sum_{i=1}^n p_i = 1$. We reserve the more common notation \mathcal{X}^* for bootstrap samples generated by uniform resampling. Denote by $F_{\xi,p}^\dagger$ the sampling distribution of \mathcal{X}^\dagger conditional on \mathcal{X} . For each fixed γ , the functional form of the sampling weights $p = p(\gamma)$, as a function of γ , is determined by minimising a 'distance' between $p(\gamma)$ and the vector of uniform weights (n^{-1}, \dots, n^{-1}) subject to conditions that $\eta(F_{\xi,p(\gamma)}^\dagger) = \xi$ and $g(\xi) = \gamma$. Hall & Presnell (1999a) propose a general class of distance measures suitable for this purpose. Let $\mathbb{P}_\gamma^\dagger$ denote the probability, given \mathcal{X} , under the drawing of weighted bootstrap samples \mathcal{X}^\dagger using weights $p(\gamma)$.

Weighted prepivoting replaces \hat{G}_j by $\tilde{G}_j(\cdot|\gamma)$, the weighted bootstrap distribution function of $R_{n,j}(\mathcal{X}, \gamma)$:

$$\tilde{G}_j(x|\gamma) = \mathbb{P}_\gamma^\dagger\{R_{n,j}(\mathcal{X}^\dagger, \gamma) \leq x\}.$$

Now $R_{n,j+1}(\mathcal{X}, \gamma) = \tilde{G}_j\{R_{n,j}(\mathcal{X}, \gamma)|\gamma\}$, for $j = 0, 1, 2, \dots$.

We show below that $R_{n,j}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to order $O(n^{-(j_0+2j)/2})$ if $R_n(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to order $O(n^{-j_0/2})$. Prepivoting by weighted bootstrap iterations thus improves upon the same number of conventional bootstrap iterations in terms of the order of the difference from the exact pivot $\text{Un}(0, 1)$. In fact, each weighted bootstrap iteration reduces the error by $O(n^{-1})$ and is asymptotically equivalent to two conventional bootstrap iterations.

Section 2 gives a theoretical account of the effects of weighted bootstrap iterations on prepivoting, and suggests a general device for selecting the appropriate sampling weights. Sections 3–5 illustrate our theory in three general settings, namely smooth function models, M -estimation and linear regression. Examples are provided of common bootstrap applications. A numerical illustration is reported in § 6. All technical details are given in the Appendix.

2. THEORY

Let σ^2 be the asymptotic variance of $n^{1/2}\hat{\theta}$ and let $\hat{\sigma}^2$ be an appropriate sample estimator. The forms of estimator are described for particular cases later, where σ^2 assumes the general expression in (A7). Studentisation of $\hat{\theta}$ yields an asymptotically standard normal variable $T = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$.

The initial function $R_n(\mathcal{X}, \gamma)$ may take on a variety of forms, depending on the application in hand. For example, it may be a test statistic for testing the null parameter value γ or an approximate pivot for constructing confidence sets. Sections 3.2–3.5 present examples corresponding to different choices of R_n . Recall that $\theta = g\{\eta(F)\}$. Let $\bar{Z} = n^{-1}\sum_{i=1}^n z_i\{\mathcal{X}, \eta(F)\}$. We assume that, for some $j \geq 0$, $R_{n,j}(\mathcal{X}, \theta)$ admits an asymptotic expansion

$$R_{n,j}(\mathcal{X}, \theta) = \Phi(T) + \phi(T)\{n^{-1/2}r_1(\bar{Z}, T) + n^{-1}r_2(\bar{Z}, T) + \dots\}, \quad (1)$$

where Φ and ϕ denote the standard normal distribution and density functions respectively, and each $r_i(z, t)$ is a polynomial in t with coefficients being smooth functions of the vector z . It is clear from (1) that $R_{n,j}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$. In subsequent sections of the paper, we shall give the forms of (1) under three important general settings and briefly describe the regularity conditions required in each case.

Consider an arbitrary smooth function f defined on \mathbb{R}^d . Define $\mu_Z = \lim_{n \rightarrow \infty} E(\bar{Z})$, which will exist under mild regularity conditions. We assume that, given the sample \mathcal{X} , the family of resampling weights $p(\gamma)$ satisfies, at $\gamma = \theta$,

$$f\left[\sum_{i=1}^n p_i(\theta)z_i\{\mathcal{X}, \eta(F)\}\right] = f(\bar{Z}) + n^{-1/2}H(\mu_Z, T, f) + O_p(n^{-1}), \quad (2)$$

where $H(z, t, f)$ is a smooth functional of $z \in \mathbb{R}^d$, with $t \in \mathbb{R}$ and f in the class of smooth functions from \mathbb{R}^d to \mathbb{R} . We remark that (1) and (2) hold in all the standard cases looked at in the remainder of the paper. Define $D(z, f)$ to be a smooth functional which satisfies

$$E[n^{1/2}T\{f(\bar{Z}) - f(\mu_Z)\}] = \sigma^{-1}D(\mu_Z, f) + O(n^{-1}).$$

Recall that prepivoting using weights $p(\gamma)$ yields $R_{n,j+1}(\mathcal{X}, \gamma) = \tilde{G}_j\{R_{n,j}(\mathcal{X}, \gamma)|\gamma\}$. Define $z_x = \Phi^{-1}(x)$. Our main result is given in the following proposition.

PROPOSITION 1. *Assume that (1) and (2) hold, that*

$$\text{pr}\{R_{n,j}(\mathcal{X}, \theta) \leq x\} = x - n^{-j_0/2}\phi(z_x)d(\mu_Z, z_x) + O(n^{-(j_0+1)/2}) \quad (3)$$

for some polynomial $d(\mu_Z, z_x)$ in z_x with coefficients depending smoothly on μ_Z and for some integers $j \geq 0$ and $j_0 \geq 1$, and that the expansion (3) is uniform over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$. Then

$$\begin{aligned} \text{pr}\{R_{n,j+1}(\mathcal{X}, \theta) \leq x\} &= x + n^{-(j_0+1)/2}\phi(z_x)[H\{\mu_Z, z_x, d(\cdot, z_x)\} + \sigma^{-1}z_x D\{\mu_Z, d(\cdot, z_x)\}] \\ &\quad + O(n^{-(j_0+2)/2}), \end{aligned} \quad (4)$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$.

The proof of Proposition 1 is outlined in the Appendix. The conditions under which the proposition applies differ in different problem settings. Loosely speaking, conditions which permit Edgeworth expansions in powers of $n^{-1/2}$ to order $O(n^{-(j_0+1)/2})$ suffice. Note that the form of the functional $H(z, t, f)$ depends on the choice of sampling weights p . In particular, if $p_i(\gamma) \equiv n^{-1}$ for all i and γ as in conventional bootstrap prepivoting, then

$H(z, t, f) \equiv 0$. The next proposition follows immediately by iterating the conclusion of Proposition 1.

PROPOSITION 2. *Under the conditions of Proposition 1 and for $m = 1, 2, \dots$, we have the following:*

(i) *if $p(\gamma) \equiv (n^{-1}, \dots, n^{-1})$, then*

$$\text{pr}\{R_{n,j+m}(\mathcal{X}, \theta) \leq x\} = x + O(n^{-(j_0+m)/2}),$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$;

(ii) *if the sampling weights $p(\gamma)$ used in the weighted bootstrapping are such that $p(\theta)$ satisfies (2) with*

$$H(\mu_Z, T, f) = -\sigma^{-1}TD(\mu_Z, f) + O_p(n^{-1/2}) \tag{5}$$

for smooth functions f on \mathbb{R}^d , then

$$\text{pr}\{R_{n,j+m}(\mathcal{X}, \theta) \leq x\} = x + O(n^{-(j_0+2m)/2}),$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$.

Proposition 2(i) reiterates the well-known result that each conventional bootstrap iteration succeeds in reducing the error by an order $O(n^{-1/2})$. Proposition 2(ii) establishes that weighted bootstrap iterations based on appropriate sampling weights can be more effective in prepivoting, with each iteration reducing the error by $O(n^{-1})$.

One level of uniform bootstrapping generally eliminates the leading-order discrepancy, exemplified by the term $-n^{-j_0/2}\phi(z_x)d(\mu_Z, z_x)$ in (3), between the distributions of $R_{n,j}(\mathcal{X}, \theta)$ and $\text{Un}(0, 1)$. This results in a smaller discrepancy in the form of

$$n^{-(j_0+1)/2}\phi(z_x)\sigma^{-1}z_xD\{\mu_Z, d(\cdot, z_x)\},$$

contributed by the correlation between $R_{n,j}(\mathcal{X}, \theta)$ and the eliminated leading-order discrepancy. Fine tuning of the uniform weights within $O(n^{-1/2})$ gives rise to a perturbation of order $O(n^{-1/2})$, see (2), which translates to the term $n^{-(j_0+1)/2}\phi(z_x)H\{\mu_Z, z_x, d(\cdot, z_x)\}$ in the coverage expansion (4). Such perturbation may be deployed constructively to eliminate the $O(n^{-(j_0+1)/2})$ error term in (4), provided that condition (5) holds, which may be achieved by careful tuning of the sampling weights.

A general strategy for ensuring (5) is to choose $p = p(\gamma)$ which is closest to (n^{-1}, \dots, n^{-1}) with respect to some appropriate distance measure, subject to constraints $\eta(F_{\xi,p}^\dagger) = \xi$ and $g(\xi) = \gamma$. Based on Read & Cressie's (1988, Ch. 7) notion of power divergence, Hall & Presnell (1999a) discuss a useful class of distance measures

$$D_\rho(p) = 2\{\rho(1 - \rho)\}^{-1} \left\{ n - \sum_{i=1}^n (np_i)^\rho \right\} \quad (-\infty < \rho < \infty)$$

which generalise the Kullback–Leibler distance between p and (n^{-1}, \dots, n^{-1}) . In particular, Owen's (1988) empirical likelihood ratio corresponds to D_ρ with $\rho \rightarrow 0$. In this case, p is chosen to minimise

$$D_0(p) = -2 \sum_{i=1}^n \log(np_i),$$

subject to the above constraints. Alternatively, we may follow DiCiccio & Romano's (1990) approach to determine p according to a least favourable family indexed by γ . The

sampling weights $p(\gamma)$ found by minimising D_ρ satisfy (5) for all ρ in all three settings considered in the rest of paper. Proofs are given in the Appendix.

3. SMOOTH FUNCTION MODEL

3.1. Preliminaries

Let $\mathcal{X} = (X_1, \dots, X_n)$ denote a sample of independent and identically distributed vector-valued observations in \mathbb{R}^k . Let $\eta(F)$ and \bar{X} be the population and sample means respectively. Assume Bhattacharya & Ghosh's (1978) smooth function model, so that $\theta = g\{\eta(F)\}$ for some smooth function g . A natural estimator of θ is $\hat{\theta} = g(\bar{X})$. Define $z_i = z_i(\mathcal{X}, \xi)$ to be a d -vector consisting of products of the components of X_i up to an order determined by the number of terms required in (1). Note that the z_i defined as such do not depend on ξ . The asymptotic variance σ^2 can be treated as a function of $\sigma(\mu_Z)^2$, for it depends smoothly on the first and second moments of X_1 . A plug-in estimator of σ^2 is $\hat{\sigma}^2 = \sigma(\bar{Z})^2$. Under the regularity conditions of Theorems 5.1 and 5.2 in Hall (1992) and applying the two theorems inductively, we can show that (1) is indeed satisfied by all common examples of $R_{n,j}(\mathcal{X}, \gamma)$.

The sampling weights $p = p(\gamma)$ are chosen to minimise $D_\rho(p)$ subject to the condition $g(\sum_{i=1}^n p_i X_i) = \gamma$. We prove in the Appendix that $p(\theta)$ satisfies (2) and (5) quite generally, so that all the assumptions hold. Our main conclusions are therefore available in wide generality.

We describe below a number of examples commonly found in problems of hypothesis testing and confidence limit construction. We assume that the sampling weights p are chosen such that (5) holds. As before, we denote by \mathbb{P}^* and $\mathbb{P}_\gamma^\dagger$ the conditional probabilities given \mathcal{X} under uniform and weighted bootstrap sampling respectively. Similarly, statistical functions superscripted by $*$ and \dagger are computed based on uniform and weighted bootstrap samples respectively.

Define $p_i(\mu_Z, x)$ and $q_i(\mu_Z, x)$ to be polynomials in x pertinent to Edgeworth expansions of the distributions of $T = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ and $S = n^{1/2}(\hat{\theta} - \theta)/\sigma$:

$$\text{pr}(T \leq x) = \Phi(x) + \phi(x)\{n^{-1/2}q_1(\mu_Z, x) + n^{-1}q_2(\mu_Z, x) + \dots\}, \quad (6)$$

$$\text{pr}(S \leq x) = \Phi(x) + \phi(x)\{n^{-1/2}p_1(\mu_Z, x) + n^{-1}p_2(\mu_Z, x) + \dots\}. \quad (7)$$

For details of the polynomials p_i and q_i , see Hall (1992, Ch. 2). It follows from Lemma A1 in the Appendix that

$$\begin{aligned} \text{pr}[\Phi(T) + \phi(T)\{n^{-1/2}r_1(\bar{Z}, T) + O_p(n^{-1})\} \leq x] \\ = x - \phi(z_x)[n^{-1/2}\{r_1(\mu_Z, z_x) - q_1(\mu_Z, z_x)\} + O(n^{-1})], \end{aligned} \quad (8)$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$.

3.2. Percentile method

Efron's (1979) percentile method takes

$$R_n(\mathcal{X}, \gamma) = \mathbb{P}^*(\hat{\theta}^* > \gamma)$$

as the root for constructing a bootstrap confidence interval $\{\gamma : R_n(\mathcal{X}, \gamma) \geq 1 - \alpha\}$. We can easily show by (7) that

$$R_n(\mathcal{X}, \theta) = \Phi(T) - \phi(T)\{n^{-1/2}p_1(\bar{Z}, T) - n^{-1}p_2(\bar{Z}, T) + \dots\}.$$

It follows from (8) and Proposition 2 that $R_n(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-1/2})$ and the weighted prepivoted $R_{n,m}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-(2m+1)/2})$.

A different version of the percentile method sets $R_n(\mathcal{X}, \gamma) = \mathbb{P}^*(\hat{\theta}^* - \hat{\theta} \leq \hat{\theta} - \gamma)$, so that

$$R_n(\mathcal{X}, \theta) = \Phi(T) + \phi(T)\{n^{-1/2}p_1(\bar{Z}, T) + n^{-1}p_2(\bar{Z}, T) + \dots\},$$

and the above conclusion still holds.

3.3 Inverse testing method

DiCiccio & Romano's (1990) nonparametric inverse testing method and a version of Carpenter's (1999) test inversion confidence interval both take $R_n(\mathcal{X}, \gamma) = g(\bar{X}) = \hat{\theta}$ and obtain the root by one iteration of weighted prepivoting, yielding

$$R_{n,1}(\mathcal{X}, \gamma) = \mathbb{P}_\gamma^\dagger(\hat{\theta}^\dagger \leq \hat{\theta}).$$

Note that

$$R_{n,1}(\mathcal{X}, \gamma) = \mathbb{P}_\gamma^\dagger\{n^{1/2}(\hat{\theta}^\dagger - \gamma)/\hat{\sigma}_\gamma \leq n^{1/2}(\hat{\theta} - \gamma)/\hat{\sigma}_\gamma\},$$

where $\hat{\sigma}_\gamma^2$ denotes the asymptotic variance of $n^{1/2}\hat{\theta}^\dagger$ conditional on \mathcal{X} . Substituting $\sum_i p_i(\theta)z_i$ for μ_Z and putting $x = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}_\theta$ in (7), we obtain an expansion for $R_{n,1}(\mathcal{X}, \theta)$. It then follows from (2) and (5) that

$$R_{n,1}(\mathcal{X}, \theta) = \Phi(T) + \phi(T)[n^{-1/2}\{p_1(\bar{Z}, T) - T^2D(\bar{Z}, \sigma^{-1})\} + O_p(n^{-1})].$$

We show in the Appendix that

$$p_1(z, t) - q_1(z, t) = t^2D(z, \sigma^{-1}) \quad (9)$$

and so $R_{n,1}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-1})$ by (8). Proposition 2 then implies that $R_{n,m+1}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-(m+1)})$.

3.4. Percentile-*t* method

Efron (1981) modifies the percentile method by studentising $\hat{\theta}$. The resulting percentile-*t* method chooses the root $R_n(\mathcal{X}, \gamma)$ to be

$$R_n(\mathcal{X}, \gamma) = \mathbb{P}^*\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq n^{1/2}(\hat{\theta} - \gamma)/\hat{\sigma}\},$$

so that $R_n(\mathcal{X}, \theta)$ admits the expansion

$$\Phi(T) + \phi(T)\{n^{-1/2}q_1(\bar{Z}, T) + n^{-1}q_2(\bar{Z}, T) + \dots\}$$

by (6). Applying Proposition 2 and (8), we have that $R_{n,m}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-(m+1)})$.

3.5. Studentised test inversion bootstrap method

Carpenter (1999) generalises Efron's (1981) percentile-*t* method by inverse testing. Application of this method to a nonparametric setting amounts to prepivoting the root $R_n(\mathcal{X}, \theta) = \Phi(T)$ by weighted bootstrapping once. It follows immediately from Proposition 2 that $R_{n,m}(\mathcal{X}, \theta)$ is asymptotically $\text{Un}(0, 1)$ to $O(n^{-(2m+1)/2})$. In particular, $R_{n,1}(\mathcal{X}, \theta)$ is $\text{Un}(0, 1)$ up to $O(n^{-3/2})$, which strengthens the result of Carpenter (1999) who obtains an order of $O(n^{-1})$ for the error.

4. *M*-ESTIMATION

Let $\mathcal{X} = (X_1, \dots, X_n)$ denote an independent and identically distributed random sample. Consider a score function $\Psi = (\Psi_1, \dots, \Psi_k)'$. The parameter of interest is $\theta = g\{\eta(F)\}$, where g is a smooth function and $\eta(F)$ is the unique k -vector satisfying $E\Psi\{X_1, \eta(F)\} = 0$. The *M*-estimator of θ is defined as $\hat{\theta} = g(\hat{\eta})$, where $\hat{\eta}$ is the unique solution to

$$\sum_{i=1}^n \Psi(X_i, \hat{\eta}) = 0.$$

Under smoothness assumptions on Ψ , we may define, for each i and each $\xi \in \mathbb{R}^k$, a d -vector $z_i(\mathcal{X}, \xi)$ consisting of components of the form

$$\prod_{u=1}^v \frac{\partial^{m_u} \Psi_{l_u}(X_i, \xi)}{\partial \xi_{j_{u1}} \dots \partial \xi_{j_{um_u}}},$$

for $l_u = 1, \dots, k$, $m_u = 0, 1, \dots, M$, $v = 1, 2, \dots, N$, and for some sufficiently large positive integers M and N . The asymptotic variance σ^2 , as seen from the general formula (A7), depends on the expectations of the above components for $m_u = 0, 1$ and $v = 1, 2$, whence it can be regarded as a function of μ_Z : $\sigma^2 = \sigma(\mu_Z)^2$. If we assume Lahiri's (1992a) regularity conditions, (1) holds for a variety of choices of $R_{n,j}(\mathcal{X}, \gamma)$.

Note that the sampling weights $p = p(\gamma)$ should be chosen in this context to minimise $D_\rho(p)$ subject to the conditions $g(\xi) = \gamma$ and $\sum_{i=1}^n p_i \Psi(X_i, \xi) = 0$, for some k -vector ξ . We prove in the Appendix that $p(\theta)$ satisfies (2) and (5). Weighted sampling from \mathcal{X} using weights $p(\gamma)$ amounts to selecting X_i from \mathcal{X} with probability $p_i(\gamma)$.

If we argue as in Bhattacharya & Ghosh (1978), $T = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ and $S = n^{1/2}(\hat{\theta} - \theta)/\sigma$ are essentially smooth functions of means of independent and identically distributed terms and admit Edgeworth expansions of the types (6) and (7). Bootstrap versions of these expansions follow from the rigorous proof presented in Lahiri (1992a). Results specific to bootstrap applications discussed in §§ 3.2–3.5 also hold in the context of *M*-estimation.

5. LINEAR REGRESSION

Consider a linear regression set-up, where $\mathcal{X} = ((x_1, Y_1), \dots, (x_n, Y_n))$, Y_i is the i th observed response and x_i is a k -vector of covariates corresponding to Y_i . It is assumed that

$$Y_i = x_i' \eta(F) + \varepsilon_i \quad (i = 1, \dots, n),$$

where the ε_i are independent and identically distributed random errors independent of the x_i and $\eta(F)$ is a k -variate regression parameter.

Let $\theta = g\{\eta(F)\}$ be a real parameter of interest, for some smooth function g . Estimation of $\eta(F)$ typically amounts to solving the k equations

$$\sum_{i=1}^n x_i \psi(Y_i - x_i' \hat{\eta}) = 0$$

for $\hat{\eta}$, for some smooth function ψ satisfying $E\psi(\varepsilon_1) = 0$. Then θ is estimated by $\hat{\theta} = g(\hat{\eta})$.

There exist two major bootstrap strategies for estimating the sampling distribution of $\hat{\theta}$ or its studentised version. The first strategy, known as the paired bootstrap, treats the (x_i, Y_i) as independent and identically distributed pairs, rendering the problem a special

case of M -estimation with score function

$$\Psi(x_i, Y_i, \xi) = x_i \psi(Y_i - x_i' \xi).$$

Results of § 4 then carry over immediately.

The second strategy, known as the residual bootstrap and first introduced by Efron (1979), assumes fixed covariates x_i and requires more detailed discussion in our context of weighted prepivoting. Assuming smoothness of ψ , denote by $\psi^{(m)}$ the m th derivative of ψ . Define, for each $\xi \in \mathbb{R}^k$, $z_i(\mathcal{X}, \xi)$ to be a d -vector consisting of components of the form

$$\prod_{u=1}^v \psi^{(m_u)}(y_i - x_i' \xi) \quad (m_u = 0, 1, \dots, M, v = 1, \dots, N),$$

for some sufficiently large positive integers M and N . Again we have from (A7) that σ^2 depends on expectations of the above components for $m_u = 0, 1$ and $v = 1, 2$ and can be treated in general as $\sigma^2 = \sigma(\mu_Z)^2$. Weighted sampling from \mathcal{Z} amounts to drawing a weighted bootstrap sample of residuals $(\varepsilon_1^\dagger, \dots, \varepsilon_n^\dagger)$ from $(y_1 - x_1' \xi, \dots, y_n - x_n' \xi)$ for some hypothesised or pilot value ξ of the regression parameter. The corresponding weighted bootstrap sample $\mathcal{X}^\dagger = ((x_1, Y_1^\dagger), \dots, (x_n, Y_n^\dagger))$ is reconstructed by setting

$$Y_i^\dagger = x_i' \xi + \varepsilon_i^\dagger \quad (i = 1, \dots, n).$$

The bootstrap version of $\hat{\theta}$ is given by $\hat{\theta}^\dagger = g(\hat{\eta}^\dagger)$, where $\hat{\eta}^\dagger$ solves the equations

$$\sum_{i=1}^n x_i \psi(Y_i^\dagger - x_i' \hat{\eta}^\dagger) = 0.$$

Assume that the x_i are independent and identically distributed with sufficiently many finite moments. Then, modifying Lahiri's (1992b) regularity condition on ψ and the distribution of ε_i by strengthening the smoothness assumption, we can prove (1) for many variants of $R_{n,j}(\mathcal{X}, \theta)$.

As in M -estimation, the sampling weights $p = p(\gamma)$ are found by minimising $D_\rho(p)$ subject to the conditions $g(\xi) = \gamma$ and $\sum_{i=1}^n p_i x_i \psi(Y_i - x_i' \xi) = 0$, for some k -vector ξ ; see the Appendix for a proof of (2) and (5) for such sampling weights.

Hall (1992, § 5.4) outlines the key steps for proving the Edgeworth expansions (6) and (7) for T and S , where θ is specialised to the slope parameter in a simple linear regression set-up and $\psi(x) = x$. The proof can readily be adapted, following Lahiri's (1992b) more general approach, to yield similar expansions in our general setting. Note in particular that cumulants relevant to the expansions are calculated for weighted sums of independent random variables, which does not however affect the form of the expansions under the aforementioned regularity conditions. Our discussion in §§ 3.2–3.5 thus carries over to the present context.

6. AN ILLUSTRATION

As noted above, Efron's (1979) percentile method takes

$$R_n(\mathcal{X}, \gamma) = \mathbb{P}^*(\hat{\theta}^* > \gamma)$$

as the root for constructing a bootstrap confidence interval of nominal level α . The one-sided interval is $\mathcal{I}^* = (-\infty, \hat{\theta}_\alpha^*)$, where $\hat{\theta}_\alpha^*$ is the α quantile of the distribution of $\hat{\theta}$ under the drawing of bootstrap samples from \mathcal{X} . The coverage error of \mathcal{I}^* is of order $O(n^{-1/2})$.

For this root,

$$\hat{G}_0(x) = \mathbb{P}^* \{ \mathbb{P}^{**}(\hat{\theta}^{**} > \hat{\theta}) \leq x \},$$

where $\hat{\theta}^{**}$ is the version of $\hat{\theta}$ calculated on a second-level uniform bootstrap sample \mathcal{X}^{**} drawn from a first-level uniform bootstrap sample \mathcal{X}^* , and \mathbb{P}^{**} refers to the probability, given \mathcal{X}^* , under the drawing of \mathcal{X}^{**} . Construction of the confidence interval from the prepivoted root $R_{n,1}$ yields an upper nominal α level confidence limit which is the solution, in γ , of

$$R_{n,1}(\mathcal{X}, \gamma) = \hat{G}_0\{R_{n,0}(\mathcal{X}, \gamma)\} = 1 - \alpha.$$

This is easily seen to be the solution of

$$\mathbb{P}^*(\hat{\theta}^* \leq \gamma) = \tilde{\alpha},$$

where $\tilde{\alpha}$ is the α quantile of $\mathbb{P}^{**}(\hat{\theta}^{**} \leq \hat{\theta})$, under the drawing of bootstrap samples \mathcal{X}^* from \mathcal{X} .

The iterated interval is $\mathcal{I}^{**} = (-\infty, \hat{\theta}_{\tilde{\alpha}}^*)$. It is seen that the prepivoting amounts to making an additive adjustment to the nominal coverage of the confidence interval. The coverage error of \mathcal{I}^{**} is of order $O(n^{-1})$.

Denote by $\hat{\theta}^{\dagger*}$ the version of $\hat{\theta}$ calculated on a second-level uniform bootstrap sample $\mathcal{X}^{\dagger*}$ drawn from a first-level weighted bootstrap sample \mathcal{X}^\dagger . Let $\mathbb{P}^{\dagger*}$ be the probability, given \mathcal{X}^\dagger , under the drawing of $\mathcal{X}^{\dagger*}$. Now

$$\tilde{G}_0(x|\gamma) = \mathbb{P}_\gamma^\dagger \{ \mathbb{P}^{\dagger*}(\hat{\theta}^{\dagger*} > \gamma) \leq x \},$$

so construction of the confidence interval from the weighted prepivoted root $R_{n,1}$ yields an upper nominal α -level confidence limit which is the solution, $L^{\dagger*}(\mathcal{X})$, in γ , of

$$R_{n,1}(\mathcal{X}, \gamma) = \tilde{G}_0\{R_{n,0}(\mathcal{X}, \gamma)|\gamma\} = 1 - \alpha.$$

The weighted bootstrap iterated interval is $\mathcal{I}^{\dagger*} = (-\infty, L^{\dagger*}(\mathcal{X}))$. The coverage error of $\mathcal{I}^{\dagger*}$ is of order $O(n^{-3/2})$.

An alternative to Efron's percentile method interval is the noniterated, weighted bootstrap interval \mathcal{I}^\dagger based on the root $R_{n,1}(\mathcal{X}, \gamma) = \mathbb{P}_\gamma^\dagger(\hat{\theta}^\dagger \leq \hat{\theta})$ described in §3.3. We may prepivot $R_{n,1}(\mathcal{X}, \gamma)$ by either uniform or weighted bootstrap sampling. Denote by $\mathcal{X}^{*\dagger}$ and $\mathcal{X}^{\dagger\dagger}$ the generic second-level weighted bootstrap samples drawn from a first-level uniform bootstrap sample \mathcal{X}^* and a first-level weighted bootstrap sample \mathcal{X}^\dagger respectively, and by $\mathbb{P}_\gamma^{*\dagger}$ and $\mathbb{P}_\gamma^{\dagger\dagger}$ the probability measures induced by the drawing of the second-level samples conditional on their respective first-level bootstrap samples. Similarly, let $\hat{\theta}^{*\dagger}$ and $\hat{\theta}^{\dagger\dagger}$ be the versions of $\hat{\theta}$ calculated on $\mathcal{X}^{*\dagger}$ and $\mathcal{X}^{\dagger\dagger}$ respectively.

Prepivoting $R_{n,1}(\mathcal{X}, \gamma)$ by uniform bootstrap sampling yields an interval $\mathcal{I}^{*\dagger} = (-\infty, L^{*\dagger}(\mathcal{X}))$, where the endpoint $L^{*\dagger}(\mathcal{X})$ solves

$$\hat{G}_1[R_{n,1}\{\mathcal{X}, L^{*\dagger}(\mathcal{X})\}] = \mathbb{P}^*[\mathbb{P}_{\hat{\theta}}^{*\dagger}(\hat{\theta}^{*\dagger} \leq \hat{\theta}^*) \leq R_{n,1}\{\mathcal{X}, L^{*\dagger}(\mathcal{X})\}] = 1 - \alpha.$$

On the other hand, prepivoting $R_{n,1}(\mathcal{X}, \gamma)$ by weighted bootstrap sampling yields an interval $\mathcal{I}^{\dagger\dagger} = (-\infty, L^{\dagger\dagger}(\mathcal{X}))$, where $L^{\dagger\dagger}(\mathcal{X})$ is the solution, in γ , to the equation

$$\tilde{G}_1\{R_{n,1}(\mathcal{X}, \gamma)|\gamma\} = \mathbb{P}_\gamma^\dagger \{ \mathbb{P}_\gamma^{\dagger\dagger}(\hat{\theta}^{\dagger\dagger} \leq \hat{\theta}^\dagger) \leq R_{n,1}(\mathcal{X}, \gamma) \} = 1 - \alpha.$$

The theoretical coverage errors of \mathcal{I}^\dagger , $\mathcal{I}^{*\dagger}$ and $\mathcal{I}^{\dagger\dagger}$ are of orders $O(n^{-1})$, $O(n^{-3/2})$ and $O(n^{-2})$ respectively.

Figures 1–3 present graphically the results of a small numerical study, in which the

practical effect on coverage accuracy of weighted bootstrap prepivoting, as opposed to conventional uniform bootstrap prepivoting, was studied for two cases, in each case for a standard Gaussian population. In the first case, Figs 1 and 2, the parameter of interest $\theta \equiv 1$ is the population variance, defined within the smooth function model of § 3, so that the estimator $\hat{\theta}$ is the usual sample variance. In the second, the estimator $\hat{\theta}$ is the M -estimator of scale described by Hampel et al. (1986, p. 107). In the notation of § 4, $\theta \equiv 1$ is the solution of $E\Psi(X_1, \theta) = 0$, where $\Psi(x, \xi) = \text{sgn}\{|x/\xi| - \Phi^{-1}(\frac{3}{4})\}$, so that $\hat{\theta}$ is the median of the absolute values of the observations X_1, \dots, X_n , multiplied by the factor $1/\Phi^{-1}(\frac{3}{4})$. The general theory described in § 4 holds for a slightly smoothed version of the score function Ψ , and can readily be extended to the present example by a limiting argument.

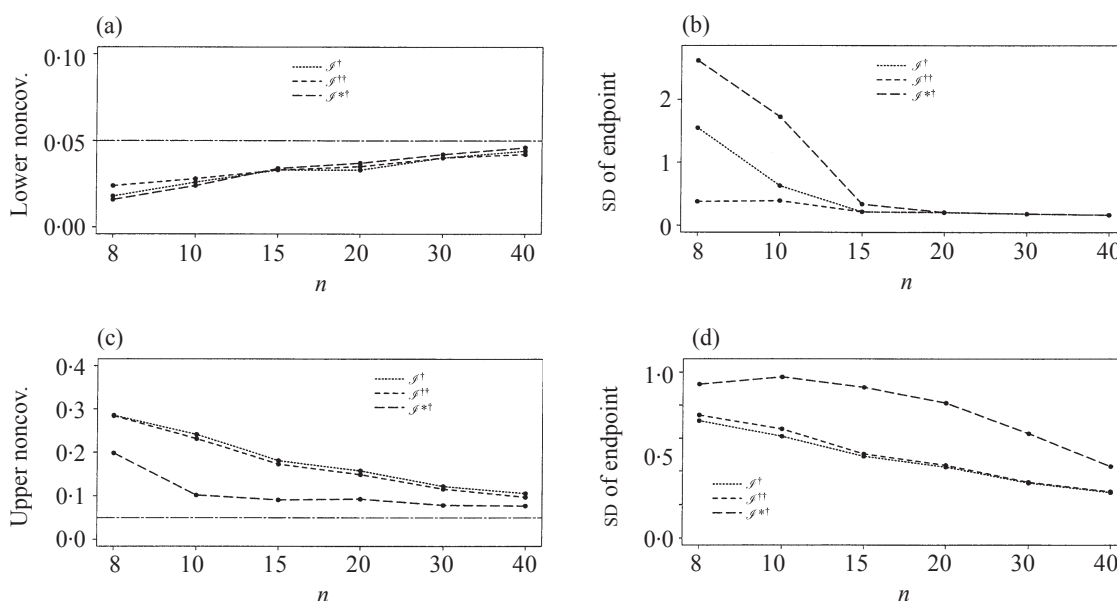


Fig. 1. Coverage accuracies of \mathcal{J}^\dagger , $\mathcal{J}^{*\dagger}$ and $\mathcal{J}^{\dagger\dagger}$, normal variance: (a) noncoverage of lower 95% confidence limit, (b) standard deviation of lower confidence limit, (c) noncoverage of upper 95% confidence limit, and (d) standard deviation of upper confidence limit.

In each case, for a range of sample sizes n , a series of 2000 samples were generated from the parent population $N(0, 1)$, and from each of these various intervals were constructed. Figure 1 compares the coverage accuracies of the intervals \mathcal{J}^\dagger , $\mathcal{J}^{*\dagger}$ and $\mathcal{J}^{\dagger\dagger}$ in the variance case. Figures 1(a), (b) refer to the case $\alpha = 0.05$, corresponding to construction of a lower confidence limit, while Figs 1(c), (d) refer to the case $\alpha = 0.95$, corresponding to an upper confidence limit. In Figs 1(a), (c) ‘lower noncoverage’ and ‘upper noncoverage’ refer to the estimated probabilities that θ lies below and above the interval respectively, each nominally 0.05. Figure 2 provides the corresponding comparison between \mathcal{J}^* and its iterated versions \mathcal{J}^{**} and $\mathcal{J}^{\dagger*}$. Figure 3 presents the same comparison as Fig. 1, but for the M -estimator of scale example. Also plotted are the standard deviations, over the 2000 replications, of the interval endpoints.

The results of the study show that the effectiveness of weighted bootstrap prepivoting depends both on the choice of initial pivot and on whether an upper or lower confidence limit is being constructed. Figure 1(a) shows that, for lower confidence limits, weighted prepivoting is, at least for small sample sizes, effective compared to conventional bootstrap

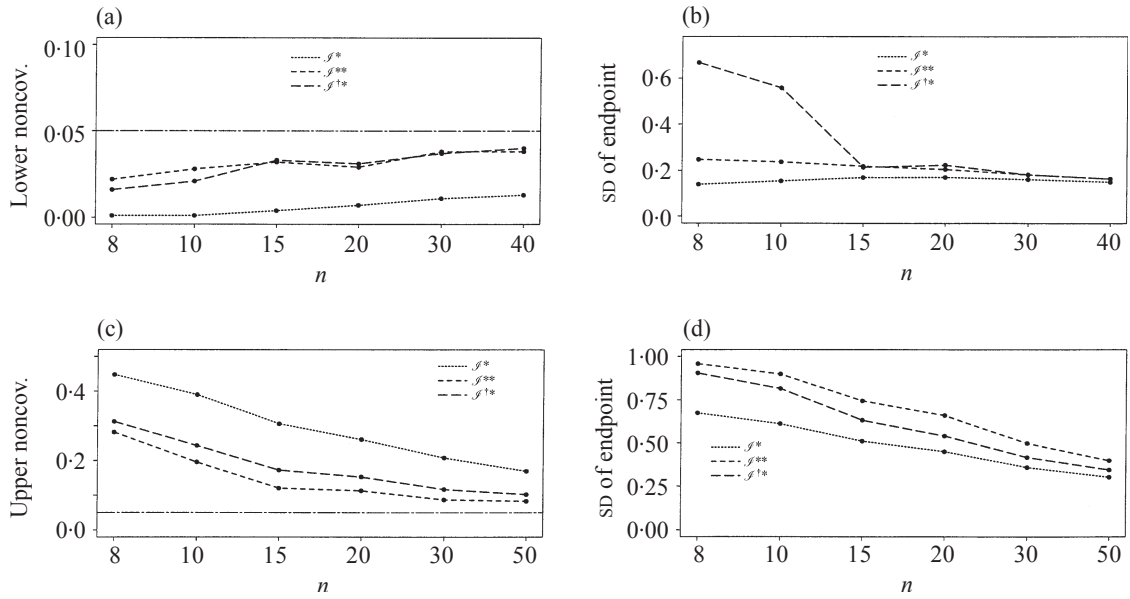


Fig. 2. Coverage accuracies of \mathcal{J}^* , \mathcal{J}^{**} and $\mathcal{J}^{\dagger*}$, normal variance: (a) noncoverage of lower 95% confidence limit, (b) standard deviation of lower confidence limit, (c) noncoverage of upper 95% confidence limit, and (d) standard deviation of upper confidence limit.

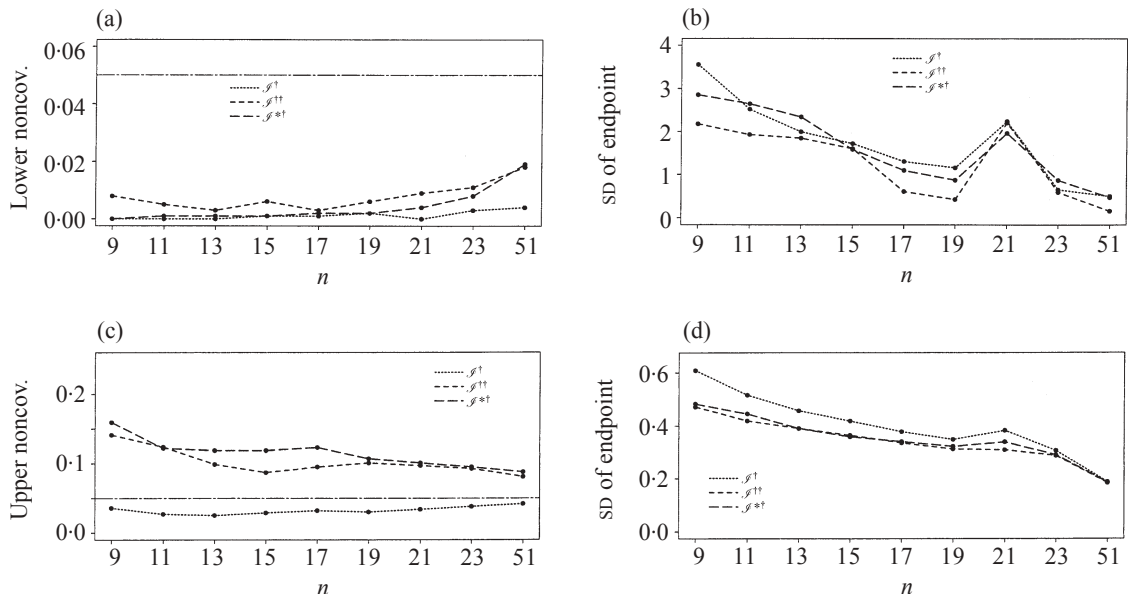


Fig. 3. Coverage accuracies of \mathcal{J}^\dagger , $\mathcal{J}^{*\dagger}$ and $\mathcal{J}^{\dagger\dagger}$, M-estimator of scale: (a) noncoverage of lower 95% confidence limit, (b) standard deviation of lower confidence limit, (c) noncoverage of upper 95% confidence limit, and (d) standard deviation of upper confidence limit.

prepivotting. We note also that weighted prepivotting produces more stable interval endpoints than does conventional prepivotting in this context. In the case of upper confidence limits, weighted prepivotting is less effective in terms of coverage error, but yields dramatic improvements over conventional prepivotting in terms of interval stability. This same broad picture is repeated in Fig. 2, and in other examples which we have studied but which are

not reported here. From Fig. 3 we see that the weighted bootstrap interval $\mathcal{I}^{\dagger\dagger}$ yields improvements in the M -estimator example over the conventional iterated interval $\mathcal{I}^{*\dagger}$, in terms of both upper and lower noncoverage. Again, relative stability of the weighted bootstrap interval is noted. It is worth remarking also that in this example the relatively good coverage accuracy of the noniterated interval \mathcal{I}^{\dagger} , which contrasts with what is seen in the variance example of Figs 1 and 2, is achieved at the expense of high variability.

We conclude this section with a brief discussion of computational issues relating to weighted bootstrap prepivoting, and with practical recommendations on its implementation and potential usefulness. Our primary remark is that the numerical requirements of weighted prepivoting are readily packaged in such a way as to provide algorithms which do not add significantly to the computational demands of conventional prepivoting algorithms.

Note that, in the context of the study reported above, prepivoting by weighted bootstrapping will generally require a search, over γ , for the confidence limit. An optimal set of sampling weights has to be determined specifically for each trial value of γ . These are, however, easily obtained using standard numerical techniques. In our empirical study, sampling weights were obtained using simple routines for sequential programming available in both the NAG and IMSL numerical packages. The search over values of γ required by the weighted bootstrap schemes is efficiently implemented by the Robbins–Monro stochastic search algorithm, as detailed by Garthwaite & Buckland (1992), and recommended also by Carpenter (1999).

In the problem of null distribution estimation, no search algorithm such as the Robbins–Monro algorithm is necessary, as a particular single value γ is specified.

We have investigated different approaches to choice of sampling weights with weighted bootstrap iteration, and recommend use of Owen's (1988) empirical likelihood weights. This approach was used in the numerical study, and seems both effective and reliable.

We contrast these requirements of weighted bootstrapping with construction of the conventional prepivoted percentile interval \mathcal{I}^{**} , described above, which requires no search over values of γ and operates under a fixed set of sampling weights.

Detailed comparison of the computational demands of weighted and conventional bootstrap iteration depends on the problem being studied, and on the particular implementation of the optimisation and stochastic search components of the weighted bootstrap prepivoting. However, expressed in general terms for the current problem of confidence set construction, conventional iteration will generally require a Monte Carlo simulation involving the drawing of, say, C second-level bootstrap samples from each of B first-level bootstrap samples drawn from \mathcal{X} , a total of $B(C + 1)$ bootstrap samples. By contrast, for weighted bootstrap iteration, for each of t values of γ used in the Robbins–Monro algorithm, it is necessary only to draw $B + 1$ first-level weighted bootstrap samples, with C second-level bootstrap samples being drawn from the $(B + 1)$ st of these, giving a total of $t(B + C + 1)$ bootstrap samples. In the study reported above, we set $t = B = C = 1000$, though it is our belief that it would be reasonable to reduce this value of t somewhat.

Specific timing comparisons between weighted and conventional prepivoting algorithms depend, among other factors, on the initial choice of root, the sample size n , the inference problem and the particular numerical implementation of the weighted bootstrap approach. In broad terms, however, it is our experience that, with the implementation as described above, weighted bootstrap prepivoting is in typical problems quite comparable with conventional prepivoting in terms of computational time. For the context described in Figs 1 and 2, for example, for sample size $n = 30$, computational loads in construction of the

intervals $\mathcal{J}^{\dagger\dagger}$ and $\mathcal{J}^{*\dagger}$ were observed to be very similar: those for \mathcal{J}^{**} are only marginally less than for $\mathcal{J}^{\dagger*}$. For smaller sample sizes n , the numerical optimisations required by weighted prepivoting have a larger impact on computational timings. For $n = 10$, for example, computational loads are about 20% less for $\mathcal{J}^{*\dagger}$ than for $\mathcal{J}^{\dagger\dagger}$, while \mathcal{J}^{**} requires only about 25% of the computation of the theoretically favoured $\mathcal{J}^{\dagger*}$.

We conclude from our numerical investigations that, while the theoretical reductions in the rate of convergence of the coverage error with sample size n obtained by weighted prepivoting are not particularly apparent in typical practical settings, weighted prepivoting can be easily implemented in such a way as to be computationally quite comparable to conventional prepivoting algorithms, and it can yield levels of coverage error which are quite acceptable compared to those obtained from conventional schemes, with improved stability. Its use in place of conventional bootstrap iteration procedures is strongly justified, on both theoretical and practical grounds.

ACKNOWLEDGEMENT

S. M. S. Lee was supported on this research by a grant from the Hong Kong Research Grants Council.

APPENDIX

Proofs

Proof of Proposition 1. Let the polynomials q_i be as defined in (6). Our proof uses the following result.

LEMMA A1. Consider

$$R = \Phi(T) + \phi(T)\{n^{-1/2}r_1(\bar{Z}, T) + n^{-1}r_2(\bar{Z}, T) + \dots\},$$

where the $r_i(z, t)$ are polynomials in t with coefficients depending smoothly on z . Then we have

$$\text{pr}(R \leq x) = x - \phi(z_x)[n^{-1/2}\{r_1(\mu_Z, z_x) - s_1(\mu_Z, z_x)\} + n^{-1}\{r_2(\mu_Z, z_x) - s_2(\mu_Z, z_x)\} + \dots] \quad (\text{A1})$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$, where each $s_{i+1}(\mu_Z, z_x)$ is a polynomial in z_x depending on the r_k only for $k \leq i$, for $i = 1, 2, \dots$. In particular, we have

$$s_1(\mu_Z, z_x) = q_1(\mu_Z, z_x),$$

$$s_2(\mu_Z, z_x) = q_2(\mu_Z, z_x) + r_1(\mu_Z, z_x)[r'_1(\mu_Z, z_x) - q'_1(\mu_Z, z_x) - z_x\{r_1(\mu_Z, z_x) - q_1(\mu_Z, z_x)\}] - \sigma(\mu_Z)^{-1}z_x D\{\mu_Z, r_1(\cdot, z_x)\},$$

where r'_1, q'_1 denote the first derivatives with respect to the second arguments of r_1, q_1 respectively.

Proof of Lemma A1. For $x \in [\varepsilon, 1 - \varepsilon]$, we can show by standard algebraic inversion that $R \leq x$ if and only if $\tilde{T} \equiv T + n^{-1/2}w_1(\bar{Z}, z_x) + n^{-1}w_2(\bar{Z}, z_x) + \dots \leq z_x$ for some smooth functions w_i , where in particular $w_1(z, t) = r_1(z, t)$, $w_2(z, t) = r_2(z, t) - r_1(z, t)r'_1(z, t) + tr_1(z, t)^2/2$, and so on. Taylor-expanding \tilde{T} in the order of successive powers of $n^{-1/2}$ and hence obtaining asymptotic expansions for the cumulants of \tilde{T} , we derive the Edgeworth expansion for the distribution function of \tilde{T} and (A1) follows immediately using arguments similar to the proof of Proposition 3.1 in Hall (1992). \square

Lemma A1 enables us to define recursively functions $e_1 = s_1, e_2 = s_2, \dots$ such that

$$\mathcal{E} = \Phi(T) + \phi(T)\{n^{-1/2}e_1(\bar{Z}, T) + n^{-1}e_2(\bar{Z}, T) + \dots\}$$

is exactly $\text{Un}(0, 1)$. In particular, $e_1 = q_1$ and

$$e_2(\mu_Z, z_x) = q_2(\mu_Z, z_x) - \sigma(\mu_Z)^{-1} z_x D\{\mu_Z, q_1(\cdot, z_x)\},$$

and so on.

To prove Proposition 1, we first apply Lemma A1 to $R_{n,j}(\mathcal{X}, \theta)$ and extend its distribution expansion (3) to include an explicit $n^{-(j_0+1)/2}$ term, in terms of the difference between $R_{n,j}(\mathcal{X}, \theta)$ and the exactly uniform \mathcal{E} . A weighted bootstrap version of this extended distribution expansion will be employed to obtain the difference between $R_{n,j+1}(\mathcal{X}, \theta)$ and \mathcal{E} . This difference can then be used to update the distribution expansion of $R_{n,j}(\mathcal{X}, \theta)$ to that of $R_{n,j+1}(\mathcal{X}, \theta)$, up to and including the $n^{-(j_0+1)/2}$ term, which will complete the proof.

To be specific, letting $R = R_{n,j}(\mathcal{X}, \theta)$ in Lemma A1 and comparing (3) with (A1), we see that necessarily $r_i(\mu_Z, z_x) = e_i(\mu_Z, z_x)$ for $i < j_0$, and $r_{j_0}(\mu_Z, z_x) = e_{j_0}(\mu_Z, z_x) + d(\mu_Z, z_x)$. For $z \in \mathbb{R}^d$ and $t \in \mathbb{R}$, define $\tilde{d}(z, t) = r_{j_0+1}(z, t) - e_{j_0+1}(z, t)$. Then we have

$$R_{n,j}(\mathcal{X}, \theta) = \Phi(T) + \phi(T)[n^{-1/2}\{e_1(\bar{Z}, T) + \Delta_n(\bar{Z}, T)\} + n^{-1}e_2(\bar{Z}, T) + n^{-3/2}e_3(\bar{Z}, T) + \dots + O_p(n^{-(j_0+2)/2})], \tag{A2}$$

where $\Delta_n(z, t) = n^{-(j_0-1)/2}d(z, t) + n^{-j_0/2}\tilde{d}(z, t)$. Noting that $e_1 = q_1$ and applying Lemma A1 with r_i replaced by $q_1 + \Delta_n$ and r_i by e_i for $i > 1$, we have

$$\mathbb{P}\{R_{n,j}(\mathcal{X}, \theta) \leq x\} = x - \phi(z_x)\{n^{-j_0/2}d(\mu_Z, z_x) + n^{-(j_0+1)/2}b(\mu_Z, z_x) + O(n^{-(j_0+2)/2})\} \tag{A3}$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$, where

$$b(z, t) = \tilde{d}(z, t) + \sigma(z)^{-1}tD\{z, d(\cdot, t)\} - d'(z, t) - td(z, t)\{q_1(z, t) + \delta_{j_0,1}d(z, t)\}$$

and δ_{kl} denotes the Kronecker delta function. The Edgeworth expansion for the weighted bootstrap distribution of $R_{n,j}(\mathcal{X}, \theta)$ can be obtained by replacing μ_Z with $\sum_i p_i(\theta)z_i\{\mathcal{X}, \eta(F)\}$ in the expansion (A3) and applying (2) to the coefficients of polynomials d and b , which yields

$$\tilde{G}_j(x|\theta) = x - \phi(z_x)(n^{-j_0/2}d(\bar{Z}, z_x) + n^{-(j_0+1)/2}[H\{\mu_Z, T, d(\cdot, z_x)\} + b(\mu_Z, z_x)]) + O_p(n^{-(j_0+2)/2}) \tag{A4}$$

uniformly over $x \in [\varepsilon, 1 - \varepsilon]$ for all $\varepsilon \in (0, \frac{1}{2}]$. Without loss of generality we restrict consideration to samples \mathcal{X} for which $R_{n,j}(\mathcal{X}, \theta) \in [\varepsilon, 1 - \varepsilon]$. Substituting (A2) for x in (A4), we obtain an asymptotic expansion for the prepivoted function $R_{n,j+1}(\mathcal{X}, \theta) = \tilde{G}_j\{R_{n,j}(\mathcal{X}, \theta)|\theta\}$:

$$R_{n,j+1}(\mathcal{X}, \theta) = \mathcal{E} - n^{-(j_0+1)/2}\phi(T)[H\{\mu_Z, T, d(\cdot, T)\} + \sigma^{-1}TD\{\mu_Z, d(\cdot, T)\}] + O_p(n^{-(j_0+2)/2}). \tag{A5}$$

Note that (A5) is similar to (A2) with d and \tilde{d} set to 0 and $-H\{\mu_Z, T, d(\cdot, T)\} - \sigma^{-1}TD\{\mu_Z, d(\cdot, T)\}$ respectively. The expansion (4) then follows from (A3), which completes the proof of Proposition 1. \square

Proof of (5) for D_p -based sampling weights. We consider a general set-up encompassing the three settings discussed in §§ 3–5. Let $\mathcal{X} = (X_1, \dots, X_n)$ be independent random vectors having a joint distribution F . The parameter of interest is $\theta = g\{\eta(F)\}$ for some smooth function g , where $\eta(F)$ is a k -vector satisfying $E\kappa_i\{X_i, \eta(F)\} = 0$, for $i = 1, \dots, n$, and each κ_i is smooth in its second argument, with values in \mathbb{R}^k . Define an estimator $\hat{\eta}$ of $\eta(F)$ implicitly by the equations $\sum_{i=1}^n \kappa_i(X_i, \hat{\eta}) = 0$. Set $\hat{\theta} = g(\hat{\eta})$. Hu & Kalbfleisch (2000) propose an estimating function bootstrap procedure for this general set-up by resampling the summands $\kappa_i(X_i, \hat{\eta})$. Hu & Zidek (1995) specialise this method to a linear model set-up.

For each $i = 1, \dots, n$, let $z_i\{\mathcal{X}, \eta(F)\}$ be a d -vector which depends on \mathcal{X} only through X_i . The sampling weights $p(\gamma)$ are chosen to minimise $D_p(p)$ subject to constraints that

$$\sum_{i=1}^n p_i \kappa_i(X_i, \xi) = 0, \quad g(\xi) = \gamma, \tag{A6}$$

for some k -vector ξ .

This set-up covers the three cases considered in §§ 3–5: for the smooth function model of § 3, the X_i are independent and identically distributed random k -vectors, $\kappa_i(X_i, \xi) = X_i - \xi$ and the constraints (A6) reduce to $g(\sum_i p_i X_i) = \gamma$; for the M -estimation scenario of § 4, the X_i are independent and identically distributed random vectors and $\kappa_i(X_i, \xi) = \Psi(X_i, \xi)$; and, for the linear regression scenario of § 5, for the paired bootstrap the X_i are independent and identically distributed, $X_i = (x_i, Y_i)$ and $\kappa_i(x_i, Y_i, \xi) = x_i\psi(Y_i - x_i'\xi)$, whereas for the residual bootstrap the X_i are independent, $X_i = Y_i$ and $\kappa_i(Y_i, \xi) = x_i\psi(Y_i - x_i'\xi)$.

It follows from the regularity conditions described in §§ 3–5 that

$$n^{-1} \sum_{i=1}^n E \left. \frac{\partial \kappa_i}{\partial \xi'}(X_i, \xi) \right|_{\xi=\eta(F)} = M + O(n^{-1}),$$

$$n^{-1} \sum_{i=1}^n E[\kappa_i\{X_i, \eta(F)\}\kappa_i\{X_i, \eta(F)\}'] = V + O(n^{-1}),$$

for some nonsingular $k \times k$ matrices M and V . Denote by ∇g the gradient of g . Define

$$\zeta_i = \nabla g\{\eta(F)\}'M^{-1}\kappa_i\{X_i, \eta(F)\}, \quad \bar{\zeta} = n^{-1} \sum_{i=1}^n \zeta_i.$$

Then

$$E\zeta_i = 0, \quad n^{-1} \sum_{i=1}^n E\zeta_i^2 = \nabla g\{\eta(F)\}'M^{-1}V(M')^{-1}\nabla g\{\eta(F)\} + O(n^{-1}).$$

Taylor expansion yields $\hat{\theta} - \theta = -\bar{\zeta} + O_p(n^{-1})$. We may therefore set

$$\sigma^2 = \nabla g\{\eta(F)\}'M^{-1}V(M')^{-1}\nabla g\{\eta(F)\}. \tag{A7}$$

It follows that, for any smooth function f on \mathbb{R}^d ,

$$D(\mu_Z, f) = -n^{-1}\nabla f(\mu_Z)' \sum_{i=1}^n E[\zeta_i z_i\{\mathcal{X}, \eta(F)\}].$$

Using standard Lagrangian arguments, we can show that the optimal p satisfies

$$p_i = \begin{cases} \left(\left(\sum_j p_j^\rho \right)^{1/(\rho-1)} \left[1 - \lambda \nabla g(\xi)' \left\{ \sum_j p_j \frac{\partial \kappa_j}{\partial \xi'}(X_j, \xi) \right\}^{-1} \kappa_i(X_i, \xi) \right] \right)^{1/(\rho-1)} & (\rho \neq 1), \\ \exp \left[\sum_j p_j \log p_j - \lambda \nabla g(\xi)' \left\{ \sum_j p_j \frac{\partial \kappa_j}{\partial \xi'}(X_j, \xi) \right\}^{-1} \kappa_i(X_i, \xi) \right] & (\rho = 1), \end{cases}$$

where λ is the Lagrangian multiplier chosen such that $g(\xi) = \theta$ and $\sum_j p_j \kappa_j(X_j, \xi) = 0$. In either case, we see that $\lambda = O_p(n^{-1/2})$, $p_i = n^{-1}\{1 + O_p(n^{-1/2})\}$ and

$$\sum_j p_j \frac{\partial \kappa_j}{\partial \xi'}(X_j, \xi) = M + O_p(n^{-1/2}).$$

It follows by expanding each p_i around n^{-1} that

$$p_i(\theta) = n^{-1}\{1 + n^{-1/2}\sigma^{-1}T\zeta_i + O_p(n^{-1})\}.$$

Expanding $f[\sum_i p_i(\theta)z_i\{\mathcal{X}, \eta(F)\}]$ around $f(\bar{Z})$ yields

$$f \left[\sum_i p_i(\theta)z_i\{\mathcal{X}, \eta(F)\} \right] - f(\bar{Z}) = -n^{-1/2}\sigma^{-1}TD(\mu_Z, f) + O_p(n^{-1}),$$

which confirms the relationship (5). □

Proof of (9). Assume without loss of generality that g is defined on \mathbb{R}^d . Theorem 2.2 of Hall (1992) gives explicit expressions for p_1 and q_1 if we let $A(\bar{Z}) = \{g(\bar{Z}) - g(\mu_Z)\}/\sigma(\mu_Z)$ and $A(\bar{Z}) =$

$\{g(\bar{Z}) - g(\mu_Z)\}/\sigma(\bar{Z})$ respectively. With Hall's notation and by straightforward differentiation, a_{ij} is the (i, j) th entry of $\sigma^{-1}(\partial^2 g/\partial z \partial z')(\mu_Z)$ in the former case and that of

$$\sigma^{-1}(\partial^2 g/\partial z \partial z')(\mu_Z) + \nabla g(\mu_Z)\nabla\sigma^{-1}(\mu_Z)' + \nabla\sigma^{-1}(\mu_Z)\nabla g(\mu_Z)'$$

in the latter case, and a_i is the i th component of $\sigma^{-1}\nabla g(\mu_Z)$ in both cases. Using (2.32) and (2.33) of Hall (1992) and noting that

$$\sigma^2 = \nabla g(\mu_Z)' \text{var}[z_1\{\mathcal{X}, \eta(F)\}] \nabla g(\mu_Z),$$

we see that the values of A_1 for p_1 and q_1 differ by $-D(\mu_Z, \sigma^{-1})$, and those of A_2 differ by $-6D(\mu_Z, \sigma^{-1})$. It then follows from (2.36) of Hall (1992) that

$$p_1(\mu_Z, t) - q_1(\mu_Z, t) = D(\mu_Z, \sigma^{-1}) + D(\mu_Z, \sigma^{-1})(t^2 - 1) = t^2 D(\mu_Z, \sigma^{-1}). \quad \square$$

REFERENCES

- BARBE, P. & BERTAIL, P. (1995). *The Weighted Bootstrap*. Berlin: Springer.
- BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–68.
- BERAN, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J. Am. Statist. Assoc.* **83**, 687–97.
- BHATTACHARYA, R. N. & GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434–51.
- CARPENTER, J. (1999). Test inversion bootstrap confidence intervals. *J. R. Statist. Soc. B* **61**, 159–72.
- CHAN, K. Y. F. & LEE, S. M. S. (2001). An exact iterated bootstrap algorithm for small-sample bias reduction. *Comp. Statist. Data Anal.* **36**, 1–13.
- CHUANG, C.-S. & LAI, T. L. (2000). Hybrid resampling methods for confidence intervals (with Discussion). *Statist. Sinica* **10**, 1–50.
- DI CICCIO, T. J. & ROMANO, J. P. (1990). Nonparametric confidence limits by resampling methods and least favourable families. *Int. Statist. Rev.* **58**, 59–76.
- DI CICCIO, T. J., MARTIN, M. A. & YOUNG, G. A. (1992). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika* **79**, 285–95.
- EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.
- EFRON, B. (1981). Nonparametric standard errors and confidence intervals (with Discussion). *Can. J. Statist.* **9**, 139–72.
- GARTHWAITE, P. H. & BUCKLAND, S. T. (1992). Generating Monte Carlo confidence intervals by the Robbins–Monro process. *Appl. Statist.* **41**, 159–71.
- HALL, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431–52.
- HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer.
- HALL, P. & MARTIN, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661–71.
- HALL, P. & PRESNELL, B. (1999a). Intentionally biased bootstrap methods. *J. R. Statist. Soc. B* **61**, 143–58.
- HALL, P. & PRESNELL, B. (1999b). Biased bootstrap methods for reducing the effects of contamination. *J. R. Statist. Soc. B* **61**, 661–80.
- HALL, P., LEE, S. M. S. & YOUNG, G. A. (2000). Importance of interpolation when constructing double bootstrap confidence intervals. *J. R. Statist. Soc. B* **62**, 479–91.
- HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. A. (1986). *Robust Statistics: The Approach based on Influence Functions*. New York: Wiley.
- HU, F. & KALBFLEISCH, J. D. (2000). The estimating function bootstrap (with Discussion). *Can. J. Statist.* **28**, 449–99.
- HU, F. & ZIDEK, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82**, 263–75.
- LAHIRI, S. N. (1992a). On bootstrapping M -estimators. *Sankhyā A* **54**, 157–70.
- LAHIRI, S. N. (1992b). Bootstrapping M -estimators of a multiple linear regression parameter. *Ann. Statist.* **20**, 1548–70.
- LEE, S. M. S. & YOUNG, G. A. (1995). Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.* **23**, 1301–30.

- LEE, S. M. S. & YOUNG, G. A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. R. Statist. Soc. B* **61**, 353–66.
- OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.
- READ, T. R. C. & CRESSIE, N. A. C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*. New York: Springer.

[Received November 2001. Revised May 2002]