# Theoretical Comparison of Bootstrap t Confidence Bounds

By

P.J. Bickel

Univ. of California, Berkeley

Technical Report 273
October 1990

Department of Statistics
University of California
Berkeley, California

# Theoretical Comparison of Bootstrap t Confidence Bounds

By

P.J. Bickel
Univ. of California, Berkeley

**Summary.** We compare in a formal way the behaviour to second order of bootstrap confidence bounds for a parameter $\theta$ based on t statistics. We investigate the effect of:

1) Varying the estimate of scale in the denominator
2) Varying the estimate $\hat{\theta}$ of $\theta$ used in the numerator
3) Varying the bootstrap method, parametric or nonparametric in terms of
   a) Equivalence of the resulting procedures
   b) Correctness of the probability of coverage
   c) Minimization of the amount of undershoot
   d) Robustness to failure of parametric assumptions

*Key Words:* (AMS 1980 Classification) Bootstrap, Confidence bounds, Second order properties.

---

Department of Statistics, University of California, Berkeley, California 94720

1. *Introduction.* Recent years have seen the development of a large number of approximate confidence bounds for a population parameter $\theta(F)$ on the basis of an i.i.d. sample $X_1, \ldots, X_n$ from a population F. These bounds, all based on resampling ideas, include the original Efron (1982) percentile method and BC (1985) and BCA (1987) modifications as well as the bootstrap t discussed extensively by Hall (1988), Beran's (1987) prepivoting approach as well as many others — see Hall (1988) and diCiccio and Romano (1988) for recent surveys.

Methods leading to bounds $\underline{\theta}_{n1}$, $\underline{\theta}_{n2}$ respectively have been considered equivalent to first order if,

$$\underline{\theta}_{n1} - \underline{\theta}_{n2} = O_p(n^{-1}) \tag{1.1}$$

and to second order if

$$\underline{\theta}_{n1} - \theta_{n2} = O_p(n^{-3/2}). \tag{1.2}$$

The principal criterion for a good method has been correctness of the coverage probability,

$$P[\underline{\theta}_n \leq \theta(F)] = 1 - \alpha + O(n^{-1/2}) \quad \text{(1st order)} \tag{1.3}$$

$$P[\underline{\theta}_n \leq \theta(F)] = 1 - \alpha + O(n^{-1}) \quad \text{(2nd order)}. \tag{1.4}$$

It is plausible and under assumptions explored in the literature true that bounds equivalent to a given order have coverage probabilities correct to the same order.

Hall (1988), (see also my discussion), diCiccio and Romano (1988), Beran (1987), diCiccio and Efron (1990) have shown that the various methods of construction of bootstrap confidence bounds are equivalent to first or second order to bootstrap t methods.

In this note we limit ourselves to different bootstrap t methods and investigate the effect of:

1) Varying the estimate of scale in the denominator

2) Varying the estimate $\hat{\theta}$ of $\theta$ used in the numerator

3) Varying the bootstrap method (nonparametric or parametric)

in terms of,

a) Equivalence of the resulting procedures

b) Correctness of probability of coverage and two aspects which have not been studied as extensively,

c) Minimization of the amount of undershoot

d) Robustness to failure of parametric assumption.

Our observations in c) are a simple application of the results of Pfanzagl (1981, 1985) — see also Bickel, Cibisov, van Zwet (1981) that first order efficiency implies second order efficiency.

The robustness remarks are trivial but we believe worth noting — see also Parr (1985).

2. *Second order correctness and equivalence.*

As we indicated our observations are $X_1, \ldots, X_n$ i.i.d. F, with empirical distribution which we denote by $F_n$. We are interested in a parameter $\theta$: $\mathbf{F} \to R$ where $\mathbf{F}$ is a nonparametric family of distributions containing all distributions with finite support, for example $\mathbf{F} = \{$all distributions$\}$ or $\{$all distributions with finite variance$\}$. We consider the natural estimate $\hat{\theta} \equiv \theta(F_n)$. We assume as a model $\mathbf{F}_0 \subset \mathbf{F}$ and we are interested in lower confidence bounds for $\theta$ restricted to $\mathbf{F}_0$ which are based on $\hat{\theta}$. The estimate $\hat{\theta}$ may or may not be efficient for estimating $\theta$ on $\mathbf{F}_0$ but we can essentially always think of efficient estimates in this way. For instance, if $\mathbf{F}_0$ is parametric with densities $p(\cdot, \theta)$ then (under regularity conditions) $\theta(F)$ formally solving

$$\int \nabla \log p(x, \theta)\, dF(x) = 0$$

where $\nabla$ is the gradient, corresponds to $\hat{\theta}$ = M.L.E. We also assume that $\hat{\theta}$ is asymptotically linear with influence function $\psi(\cdot, F)$ see Hampel (1986). That is,

$$\hat{\theta} = \theta(F) + n^{-1} \Sigma_{i=1}^{n} \psi(X_i, F) + O_P(n^{-1}) \tag{2.1}$$

where

$$\int \psi(x, F)\, dF(x) = 0$$

and

$$\sigma^2(F) \equiv \int \psi^2(x, F)\, dF(x) < \infty,$$

for all $F \in \mathbf{F}$. Thus if $\mathbf{L}(\cdot | F)$ denotes the distribution of a function of $X_1, \ldots, X_n$ under F then

$$\mathbf{L}\left[\frac{\sqrt{n}(\hat{\theta} - \theta(F))}{\sigma(F)} \,|\, F\right] \to \mathbf{N}(0, 1).$$

We further suppose that we are given an asymptotically linear estimate $\hat{\sigma}$ of $\sigma(F)$ on $\mathbf{F}_0$ with influence function r,

$$\hat{\sigma} = \sigma(F) + n^{-1} \Sigma_{i=1}^{n} r(X_i, F) + O_P(n^{-1}). \tag{2.2}$$

for $F \in \mathbf{F}_0$. If,

$$T_n(F_n, F) \equiv \sqrt{n}(\hat{\theta} - \theta(F))/\hat{\sigma}$$

and we know the exact distribution $\mathbf{L}(T_n(F_n, F) | F)$ we are led to the exact $1 - \alpha$ LCB

$$\underline{\theta}_{EX} \equiv \hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}} c_n(F)$$

where

$$P_F[T_n(F_n, F) \leq c_n(F)] = 1 - \alpha. \tag{2.3}$$

The bootstrap bound(s) corresponding to $\underline{\theta}_{EX}$ are as usual obtained by estimating F by $\tilde{F}_n$ and replacing the, in general, unknown $c_n(F)$ by $c_n(\tilde{F}_n)$ where,

$$P_{\tilde{F}_n}[T_n(F_n^*, \tilde{F}_n) \leq e_n(\tilde{F}_n)] = 1 - \alpha \tag{2.4}$$

and $F_n^*$ is the empirical df. of a sample $X_1^*, \ldots, X_n^*$ i.i.d. $\tilde{F}_n$. That is

$$\underline{\theta}_{BOOT} = \hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}} c_n(\tilde{F}_n). \tag{2.5}$$

If $\tilde{F}_n = F_n$ we are dealing with the usual nonparametric bootstrap. If $\mathbf{F}_0 = \{F_\gamma\}$ where $\gamma = (\theta, \eta)$ is a Euclidean parameter, and $\hat{\gamma} = (\hat{\theta}, \hat{\eta})$ is the M.L.E. of $\gamma$, then $\tilde{F}_n \equiv F_{\hat{\gamma}}$ is the parametric bootstrap. How do different bootstrap bounds compare in terms of second order correctness and equivalence? The following "theorem", stated under unspecified regularity conditions gives the answer. Parts A and B of the theorem have been noted by Hall, Beran, and others. We give a heuristic proof and then indicate what kind of regularity conditions are needed.

We let $\underline{\theta}_{BOOT}$, $\underline{\theta}_{EXACT}$ denote the nonparametric bootstrap bounds corresponding to $\tilde{F}_n = F_n$ and the corresponding exact bound. Superscripts 1,2 will indicate different choices of numerator $\hat{\theta}^{(1)}$ or $\hat{\theta}^{(2)}$ while subscripts will similarly correspond to different choices of $\hat{\sigma}^{(1)}$ or $\hat{\sigma}^{(2)}$. $\underline{\theta}_{PBOOT}$ with or without indices will indicate $\tilde{F}_n = F_{(\hat{\theta}, \hat{\eta})}$ when $\mathbf{F} = \{F_\gamma\}$, $\gamma = (\theta, \eta)$.

**Theorem:** Under suitable regularity conditions, for $F \in \mathbf{F}_0$,

A. $\underline{\theta}_{BOOT}$, $\underline{\theta}_{PBOOT}$ are second order correct.

B. $\underline{\theta}_{BOOT} = \underline{\theta}_{PBOOT} + O_p(n^{-3/2}) = \underline{\theta}_{EXACT} + O_p(n^{-3/2})$

C. If $\hat{\sigma}_1 = \hat{\sigma}_2$ but $\hat{\theta}^{(2)} = \hat{\theta}^{(n)} + \frac{\Delta_n}{n}$ where $\mathbf{L}_F(\Delta_n) \to \mathbf{L}_F(\Delta)$ then

$$\mathbf{L}_F\{n(\underline{\theta}^{(2)}_{BOOT} - \underline{\theta}^{(1)}_{BOOT}\} \;\to\; \mathbf{L}_F(\Delta - d(F)) \tag{2.6}$$

where $d(F)$ is a constant. If $\Delta$ is constant, $\Delta = d(F)$. Thus, unless $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ differ to second order only in bias i.e. $\hat{\theta}^{(2)} = \hat{\theta}^{(1)} + \frac{\Delta(F)}{n} + O_p(n^{-3/2})$ for $\Delta$ a constant, then $\underline{\theta}^{(2)}_{BOOT}$ and $\underline{\theta}^{(1)}_{BOOT}$ are not equivalent to second order.

D. Suppose $\hat{\theta}^{(1)} = \hat{\theta}^{(2)} = \hat{\theta}_{eff}$, an efficient estimate for $\theta$ on $\mathbf{F}_0$, but $r_1 \neq r_2$ where $r_j$ correspond to $\hat{\sigma}_j$ via (2.2). Then,

$$\mathbf{L}_F\{n(\underline{\theta}_{BOOT1} - \underline{\theta}_{BOOT2})\} \;\to\; \mathbf{L}_F(\Delta) \;\neq\; 0$$

**Comments:** A,B: If $F \in \mathbf{F}_0$ there is nothing to choose between the parametric and nonparametric bootstrap t bounds to second order

(C) If $\hat{\theta} \equiv$ MLE, $b_n \equiv E_F(\hat{\theta}) - \theta$ then typically $b_n = \frac{b(F)}{n} + O(n^{-2})$ so that if $\hat{\theta}^{(2)} = \hat{\theta} - \frac{b(F_n)}{n}$, the debiased MLE, we expect (under regularity conditions!) bootstrap t bounds based on $\hat{\theta}$, $\hat{\theta}^{(2)}$ both to be second order correct and second order equivalent even though the parent estimates differ to second order. This continues to hold for estimates $\hat{\theta} + d(F_n)/n$ for d smooth. However, if $\hat{\theta}^{(2)}$ is an efficient estimate produced by an alternative method such as modified minimum $\chi^2$ where we can expect $\mathbf{L}_F(\Delta)$ to be nondegenerate then the bounds are no longer second order equivalent. Admittedly such estimates $\hat{\theta}^{(2)}$ can always be improved to second order by procedures of the form $\hat{\theta} + \frac{d(F_n)}{n}$. For a discussion see Berkson (1980), Pfanzagl (1981)

(D) If $\hat{\theta}$ is the MLE of $\theta$ for a parametric model $\mathbf{F}_0$ suppose

$$\hat{\sigma}_1^2 \;\equiv\; \int \psi^2(x, F_n)\, dF_n(x),$$

the nonparametric estimate of $\sigma^2(F_\gamma)$ while

$$\hat{\sigma}_2^2 \;=\; \sigma^2(F_{\hat{\gamma}})$$

where $\hat{\gamma}$ is the MLE of $\gamma$. Then $r_1 \neq r_2$ unless $\hat{\sigma}_1^2$ is also efficient for $\mathbf{F}_0$. Thus the bounds are not equivalent to second order. As a consequence we note the following phenomenon. Hall (1988) shows that the Efron parametric and non-parametric BCA bounds are second order equivalent to bootstrap t bounds based on $\frac{\hat{\theta}-\theta}{\hat{\sigma}_i}$, $i = 1, 2$, respectively. We conclude that, in general, the parametric and nonparametric BCA bounds are not second order equivalent despite the equivalence in A. Using the parametric or nonparametric bootstrap for the distribution of $\hat{\theta}$ which is a starting point in this method does make a difference.

**Formal proof of theorem:**

Our heuristic argument supposes that, for each studentized statistic we consider $T(F_n, F)$ both $\mathbf{L}(T(F_n, F) | F)$ and $\mathbf{L}(T(F_n^*, \tilde{F}_n) | \tilde{F}_n)$ admit (Edgeworth) expansions to order $n^{-3/2}$. That is,

$$P_F[T(F_n, F) \le x] = \Phi(x) - n^{-1/2}\phi(x) A(x, F) + O(n^{-1}) \qquad (2.7)$$

where $A(\cdot, F)$ is a polynomial of degree 2 and $O(n^{-1})$ is uniform in x. Similarly, we require,

$$P_{\tilde{F}_n}[T(F_n^*, \tilde{F}_n) \le x] = \Phi(x) - n^{-1/2}\phi(x) A(x, \tilde{F}_n) + O_p(n^{-1}). \qquad (2.8)$$

Agreement of $\underline{\theta}_{BOOT}$ and $\underline{\theta}_{EXACT}$ (part B) follows from asymptotic inversion of (2.7), (2.8) and

$$A(\cdot, \tilde{F}_n) = A(\cdot, F) + O_p(n^{-1/2}). \qquad (2.9)$$

If we further suppose that we can substitute the random $x = c_n(\tilde{F}_n)$ into (2.7) with $O(n^{-1})$ changing to $O_p(n^{-1})$ then evidently part A follows. We note in passing that the same type of heuristics indicate that all $\underline{\theta}_{BOOT}$ are first order equivalent to each other as well as first order correct.

The heuristics for C and D are based on the following lemma from Bai, Bickel, Olshen (1989).

**Lemma:** Suppose, for $j = 1, 2$, statistics $T_{nj}$,

1) $T_{nj}$ have Edgeworth expansions to order $n^{-1}$.

$$P[T_{nj} \le x] = \Phi(x) + n^{-1/2} A_j(x) + O(n^{-1}) \qquad (2.10)$$

with $O(n^{-1})$ uniform in x.

2) If $T_{n2} \equiv T_{n1} + \frac{\Delta_n}{\sqrt{n}}$

$$L(T_{n1}, \Delta_n) \rightarrow L(U, V) \text{ with } E|V| < \infty.$$

Then,

$$A_1(x) = A_2(x) \text{ for all } x \text{ iff } E(V|U) = 0.$$

The proof is obtained by considering the Edgeworth expansions of $Ee^{itT_{nj}}$ which are valid under (2.10).

To prove C, D we note first that by A we can equivalently consider $\underline{\theta}^{(1)}_{EXACT}$ and $\underline{\theta}^{(2)}_{EXACT}$.

By (2.7) we expect

$$c_n^{(1)}(F) = c_n^{(2)}(F) + \frac{p(F)}{\sqrt{n}} + O(n^{-1}).$$

Since

$$\hat{\sigma} = \sigma(F) + O_p(n^{-1/2})$$

we obtain that (2.6) holds with

$$d(F) = \sigma(F) + (F).$$

If $\Delta$ is constant we need only show that

$$r(F) = \frac{\Delta}{\sigma(F)} \tag{2.11}$$

since then,

$$\underline{\theta}^{(2)}_{EXACT} - \underline{\theta}^{(1)}_{EXACT} = \frac{\hat{\sigma}}{n}\left[\frac{\Delta_n}{\hat{\sigma}} - \frac{\Delta}{\sigma(F)}\right]$$

$$= o_p(n^{-1}).$$

But if,

$$T_{n1} \equiv \sqrt{n}\,\frac{(\hat{\theta}^{(1)} - \theta)}{\hat{\sigma}}, \qquad T_{n2} = \sqrt{n}\,\frac{(\hat{\theta}^{(2)} - \theta)}{\hat{\sigma}} - \frac{\Delta n^{-1/2}}{\sigma(F)}$$

evidently $n^{1/2}(T_{n1} - T_{n2}) = \frac{\Delta_n}{\hat{\sigma}} - \frac{\Delta}{\sigma(F)} = o_p(1)$. The lemma applies with $V = 0$ and (2.11) follows. For D we also need a result from the theory of efficient estimation, (see Pfanzagl (1981), Bickel, Klaassen, Ritov, and Wellner

(1991)) which we again state without explicit regularity conditions. These may however be found in the works cited above.

**Proposition:** Let $\mathbf{F}_0 = \{F_\gamma : \gamma \in \Gamma\}$, $\Gamma \subset R^k$ be a regular parametric model and $\mu : \mathbf{F}_0 \to R$ given by $\mu(F_\gamma) \equiv q(\gamma)$ where q is smooth. Let $\hat{\mu}_{eff}$ be an efficient (BAN) estimate of $\mu$ and $\hat{\mu}$ be another regular, asymptotically linear estimate of $\mu$. That is, for all $F \in \mathbf{F}_0$, in a uniform sense,

$$\hat{\mu} = \mu(F) + n^{-1} \Sigma_{i=1}^n w(X_i, F) + o_p(n^{-1/2}) \tag{2.12}$$

where $\int w(x,F)\, dF(x) = 0$, $\int w^2(x,F)\, dF(x) < \infty$ and the same holds for $w_{eff}$. Let $\frac{\partial l}{\partial \gamma_j}(X_1, \gamma)$, $j = 1, \ldots, k$ be the score function (derivatives of the loglikelihood of $X_1$). Then if $F = F_\gamma$, for $j = 1, \ldots, k$,

$$\mathrm{cov}_\gamma\left(w(X_1, F_\gamma) - w_{eff}(X_1, F_\gamma),\ \frac{\partial l}{\partial \gamma_j}(X_1, \gamma)\right) = 0. \tag{2.13}$$

That is, $w(X_1, F_\gamma) - w_{eff}(X_1, F_\gamma)$ is orthogonal to the tangent space of $\mathbf{F}_0$ at $F_0$. This is essentially a consequence of the Hájek-Le Cam convolution theorem. Alternatively it can be viewed as a consequence of the differential geometry of statistical models, see for example Efron (1975). If the linear span of $\{\frac{\partial l}{\partial \gamma_j}(X_1, \gamma)\}$ is interpreted as the tangent space of $\mathbf{F}_0$ at $F_\gamma$ claim (2.14) is true for semiparametric models as well see Bickel et al (1991). To prove D it is enough to establish that

$$c_{n1}(F) = c_{n2}(F) + O(n^{-1}). \tag{2.14}$$

For then, by A, the same applies to $c_{n1}(F_n)$ and $c_{n2}(F_n)$ and hence,

$$n(\underline{\theta}_{BOOT2} - \underline{\theta}_{BOOT1}) = n^{-1/2} c_{n1}(F) \Sigma_{i=1}^n (r_1(X_i, F) - r_2(X_i, F)) + o_p(1),$$

and D follows. But (2.15) follows from the lemma and proposition. Since $\hat{\theta}$ is efficient, $\psi(X_1, F_\gamma)$ is in the linear span of $\frac{\partial l}{\partial \gamma_j}(X_i, \gamma)$ $j = 1, \ldots, k$. If, without loss of generality we take, $\hat{\sigma}_1 = \hat{\sigma}_{eff}$ then, by the proposition, for all $F_\gamma \in \mathbf{F}_0$,

$$\mathrm{cov}_\gamma(\psi(X_1, F_\gamma),\ r_2(X_1, F) - r_1(X_1, F)) = 0. \tag{2.15}$$

Write,

$$U_n = \sqrt{n}\,\frac{(\hat\theta - \theta)}{\hat\sigma_1}, \qquad V_n = \sqrt{n}\,\frac{(\hat\theta - \theta)}{\hat\sigma_2}.$$

Then,

$$\begin{aligned} V_n &= U_n + U_n \frac{(\hat\sigma_1 - \hat\sigma_2)}{\hat\sigma_2} \qquad (2.16)\\ &= U_n + \frac{\Delta_n}{\sqrt{n}} \end{aligned}$$

where

$$\begin{aligned} \Delta_n = U_n (n^{-1/2} \textstyle\sum_{i=1}^n (r_1 (X_i, F) - r_2 (X_i, F))\sigma^{-1} (F)) \\ + o_p(1). \end{aligned}$$

Evidently, $\mathbf{L}_F (U_n, \Delta_n)$ tends to a $\mathbf{L}(Z, WZ)$ where $Z \sim N(0, \sigma^{-2}(F) \int \psi^2 (x, F)\, dF(x))$, $W$ is independent of $Z$ and $N(0, \sigma^{-2}(F) \int (r_1 - r_2)^2 (x, F)\, dF(x))$. Evidently $E(WZ \mid Z) = 0$ and we can apply the lemma. □

To make these results rigorous we need to justify (2.7), (2.8), (2.9) and substitution of the random $c_n(F_n)$ into (2.7). When the estimates are smooth functions of vector means $n^{-1} \sum_{i=1}^n M(X_i)$, $M_{p\times 1}$, the argument is due to Cibisov (1973) and Pfanzagl (1981), see also Hall (1986). In general the idea is:

a) To expand $T(F_n, F)$ in a von Mises or Hoeffding expansion and show that the remainder after three terms can be neglected in the Edgeworth expansion. That is, write for suitable $a_j$

$$\begin{aligned} T(F_n, F) = \sqrt{n}\{&\textstyle\int a_1(x)\, d(F_n - F)(x) \qquad (2.17)\\ &+ \textstyle\int a_2(x,y)\, d(F_n - F)(x)\, d(F_n - F)(y)\\ &+ \textstyle\int a_3(x,y,z)\, d(F_n - F)(x)\, d(F_n - F)(y)\, d(F_n - F)(z)\\ &+ r_n\} \end{aligned}$$

where $P[|r_n| \geq n^{-3/2-\delta}] = O(n^{-1})$ for some $\delta > 0$. This is to be expected since we expect $r_n = O_p(n^{-2})$ Conditions such that the sum of the first three terms has an Edgeworth expansion of the from (2.7) may be gleaned from Bickel, Götze, van Zwet (1989) for example. Of course (2.9) can, in principle, be justified in the same way save that techniques such as those of Singh (1981) and Bickel and Freedman (1980) have to be employed to get by the failure of Cramer's condition due to the discreteness of $F_n$. Substitution of $c_n(F_n)$ in (2.7) can be justified once we express $T(F_n, F) - c_n(F_n)$ in the form (2.17) by using the inversion of (2.8) for $c_n(F_n)$.

3. *Second order optimality and robustness.*

It is natural to define second order efficiency in terms of undershoot for a lower confidence bound $\underline{\theta}^*$ by: For all $F \in \mathbf{F}_0$,

i) $\underline{\theta}^*$ is second order correct

ii) If $\underline{\theta}$ is second order correct and $\delta_n > 0$, $n^{1/2}\delta_n = O(1)$ then

$$P_F[\underline{\theta}^* \le \theta(F) - \delta_n] \le P_F[\underline{\theta} \le \theta(F) - \delta_n] + o(n^{-1/2}). \tag{3.1}$$

In fact to avoid superefficiency phenomena we essentially have to require second order correctness to hold uniformly on shrinking neighbourhoods of every fixed F and then require (3.1) to similarly hold uniformly. If $o(n^{-1/2})$ is replaced in (3.1) by $o(1)$ then $\underline{\theta}^*$ is first order efficient. It is shown in Pfanzagl (1981) and Bickel, Cibisov, and van Zwet (1981) that first order efficiency implies second order efficiency. But $\underline{\theta}_{BOOT} = \hat{\theta} - \frac{\sigma(F)}{\sqrt{n}} z_{1-\alpha} + O_p(n^{-1})$. If $\hat{\theta}$ is efficient as an estimate it follows that $\underline{\theta}_{BOOT}$ is first order efficient.

We conclude that second order efficiency of bootstrap t confidence bounds depends only on the first order efficiency of the estimate $\hat{\theta}$ defining them and not on the choice of $\hat{\sigma}$ beyond its consistency. If $\hat{\theta}$ is not first order efficient then $\underline{\theta}_{BOOT}$ is not first order and a fortiori not second order efficient.

**Robustness:** Suppose $\theta(F)$ is the parameter we wish to estimate for all $F \in \mathbf{F}$. If $\mathbf{F}_0 = \{F_\gamma : \gamma \in \Gamma\}$, $\Gamma \subset R^k$ we may wish to estimate $\sigma(F_\gamma)$ by $\sigma(F_{\hat{\gamma}})$ in forming $\underline{\theta}_{BOOT}$. If $F \notin \mathbf{F}_0$ and $\hat{\gamma} = \gamma(F) + o_p(1)$ we expect $\sigma(F_{\hat{\gamma}}) = \sigma(F_{\gamma(F)}) + o_p(1)$. Unless $F_{\gamma(F)} = F$, $\underline{\theta}_{BOOT}$ will not even be first order correct. This does not happen if we use $\hat{\sigma}^2 = \int \psi^2(x, F_n)\, dF_n(x)$ and the nonparametric bootstrap.

On the other hand if we use $\hat{\sigma}^2 = \int \psi^2(x, F_n)\, dF_n(x)$ but use the parametric bootstrap, when $F \notin \mathbf{F}_0$, we are, in general, first order correct. The reason is that,

$$1 - \alpha = P_{\tilde{F}_n}\left[\frac{\sqrt{n}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*} \le c_n(\tilde{F}_n)\right] = P_{F_{\gamma(F)}}\left[\frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(F_{\gamma(F)})} \le z_{1-\alpha}\right] + o_p(1).$$

But $\underline{\theta}_{PBOOT}$ is not second order correct since that depends on $A(\cdot, F_{\gamma(F)}) = A(\cdot, F)$, which, in general, is false. Thus, robustness considerations strongly mandate a robust estimate of variance and more weakly mandate use of the nonparametric bootstrap.

## References

Bai, C., Bickel, P.J. and Olshen, R. (1989). The bootstrap for prediction. Proceedings of an Oberwolfach Conference, Springer-Verlag.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457-468.

Berkson, J. (1980). Minimum chi-square not maximum likelihood (with discussion). *Ann. Statist.* **8**, 457-487.

Bickel, P.J. (1974). Edgeworth expansion in nonparametric statistics. *Ann. Statist.* **2**, 1-20.

Bickel, P.J. and Freedman, D.A. (1981). Some asymptotics on the bootstrap. *Ann. Statist.* **9**, 1196-1217.

Bickel, P.J., Chibisov, D.M. and van Zwet W.R. (1981). On efficiency of first and second order. *International Statistical Review* **49**, 169-175.

Bickel, P.J., Götze, F., and van Zwet, W.R. (1983). A simple analysis of third-order efficiency of estimates. Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer. Wadsworth. Belmont.

Bickel, P.J., Götze, F. and van Zwet, W.R. (1989). The Edgeworth expansion for U statistics of degree two. *Ann. Statist.* **14**, 1463-1484.

Chibisov, D.M. (1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theory Probab. Appl.* **18**, 295-303.

diCiccio, T. and Romano, J. (1988). A review of bootstrap confidence intervals. *JRSS* **B 50**, 338-355.

diCiccio, T. and Efron, B. (1990). Better approxiamte confidence intervals in exponential families. *Tech Report, Stanford University.*

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3**, 1189-1242.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM. Philadelphia.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**, 45-58.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82**, 171-200.

Hall, P.J. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431-1452.

Hall, P.J. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16**, 927-985.

Hampel, F., Renchotti, E., Rousseuw, P., Stahel, W. (1986). Robust statistics: the approach based on influence functions. J. Wiley. New York.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187-1195.

Parr, W. (1985). The bootstrap: Some large sample theory and connections with robustness. *Stat. Prob. Letters* **3**, 97-100.

Pfanzagl, J. (1981). Contributions to a general asymptotic statistical theory. *Lecture Notes in Statistics* **13**, Springer Verlag.

Pfanzagl, J. (1985). Asymptotic expansions for general statistical models. *Lecture Notes in Statistics* **31**, Springer Verlag.