BOOTSTRAP ADAPTIVE TRIMMED MEANS

BY

CHRISTIAN LÉGER AND JOSEPH P. ROMANO

TECHNICAL REPORT NO. 314
FEBRUARY 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
BOOTSTRAP ADAPTIVE TRIMMED MEANS

BY

CHRISTIAN LÉGER AND JOSEPH P. ROMANO

TECHNICAL REPORT NO. 314
FEBRUARY 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Bootstrap Adaptive Trimmed Means

Christian Léger
Département d'informatique et de recherche opérationnelle
Université de Montréal

Joseph P. Romano
Department of Statistics, Stanford University
Stanford, CA 94305, USA

February, 1989

Abstract. Jaeckel (1971) introduced an adaptive trimmed mean to estimate the location of a symmetric distribution whose trimming proportion is the one corresponding to the smallest estimate of the asymptotic variance of the $\alpha$-trimmed mean. Its asymptotic variance is equal to the smallest asymptotic variance among all trimmed means with trimming proportion in the range considered by the adaptive estimator. In this paper, we consider adaptive trimmed means whose trimming proportion corresponds to the smallest bootstrap estimate of a characteristic of the distribution of the $\alpha$-trimmed means. Jaeckel's estimator is one such estimator; it minimizes bootstrap estimates of the asymptotic variance. Two other bootstrap adaptive trimmed means are introduced which minimize bootstrap estimates of the finite sample variance and of the finite sample interquartile range. They all share the same first order asymptotic properties. Confidence intervals for the location parameter based on these bootstrap adaptive trimmed means are constructed. Intervals based on the asymptotic normality have a major flaw in this setting. Bootstrap intervals better incorporate the adaptive nature of the estimators, and are shown to be asymptotically valid. Simulations show that the bootstrap intervals have much better coverage than the classical intervals. Moreover, bootstrap adaptive trimmed means based on finite sample characteristics outperform Jaeckel's estimator. Finally, it is argued that the same ideas can be used in many other problems to use the bootstrap to select an adaptive procedure.

Key words: Adaptive estimation; Bootstrap; Confidence limits; Trimmed mean.
1. Introduction

Jaeckel (1971) introduced an adaptive trimmed mean to estimate the location of a symmetric distribution. The idea is simple: for each fixed trimming proportion in a given set compute an estimate of the asymptotic variance of the trimmed mean and let the adaptively chosen trimming proportion correspond to the smallest estimate. He showed that such an estimator is asymptotically normal with an asymptotic variance equal to the smallest among all trimmed means with a trimming proportion in the given set, in other words, this adaptive trimmed mean does asymptotically as well as the best trimmed mean for that particular distribution.

Jaeckel’s estimator was included in the Princeton Robustness Study (Andrews et. al. (1972)). Although it certainly showed signs of robustness, its performance, especially in small samples, was not very good. So its very good asymptotic properties must be contrasted with its rather modest small samples properties. One of the purposes of the paper is to present alternative adaptive trimmed means which would share the asymptotic properties of Jaeckel’s estimator and perform much better in small samples.

A careful look at Jaeckel’s estimate of the asymptotic variance of the trimmed mean shows that it is in fact a bootstrap estimate (Efron (1979)). It is then natural to try to modify Jaeckel’s adaptive trimmed mean by minimizing bootstrap estimates of characteristics of the finite sample distribution of the trimmed mean, such as the variance or the interquartile range, rather than characteristics of the asymptotic distribution such as the asymptotic variance. By minimizing estimates of the finite sample distribution, such adaptive trimmed means might have better small sample properties while retaining the asymptotic properties of Jaeckel’s estimator.

This paper is also concerned with the construction of approximate confidence intervals based on such adaptive estimators. In general, (at least) two alternative methods are available. The first one uses the asymptotic normality of the estimator and requires an estimate of its variance while the second one consists of bootstrapping the adaptive trimmed mean, a task which is computer intensive. Of course, when Jaeckel introduced his estimator, the second option was not yet available. Nevertheless, the first method has a major flaw in this particular problem. The only readily available estimate of variance for the adaptive trimmed mean is the estimate of asymptotic variance of the trimmed mean with trimming proportion fixed at the value of the adaptively chosen trimming proportion. But since this proportion is the one with the smallest estimate of asymptotic variance, it is clear that such an estimator of variance is biased downwards which leads to confidence intervals that are much too short. Hence the actual confidence level of an approximate
confidence interval based on that method will be smaller than the stated level.

The problem is caused by the fact that the estimator of variance pretends that the trimming proportion was chosen a priori rather than adaptively. This is not the case with a bootstrap confidence interval since the trimming proportion is adaptively selected for each bootstrap sample. Note that one might use a bootstrap estimate of variance in the standard approach, but one of the major advantages of that approach would be lost, namely, the relative ease of computation.

In this paper, several bootstrap adaptive trimmed means are introduced and in section 2, are shown to be asymptotically normal with the smallest asymptotic variance among the family of trimmed means. So these adaptive trimmed means are asymptotically as good as the best trimmed mean for the distribution at hand, as is Jaeckel's adaptive trimmed mean.

In section 3, the two methods for constructing approximate confidence intervals will be discussed and compared. The asymptotic consistency of the methods is obtained. Section 4 will be devoted to the results of a small sample simulation which will show that bootstrap adaptive trimmed means based on minimizing characteristics of the finite sample distribution rather than the asymptotic distribution are better. Moreover it will demonstrate that bootstrap confidence intervals are superior to the standard confidence intervals. Some concluding remarks are presented in section 5 and section 6 consists of an appendix containing a sketch of the proofs.

Finally, note that the ideas introduced here on using the bootstrap to construct adaptive statistical procedures can be used in many other similar problems where one has to choose among a family of estimators. See Léger and Romano (1989).

2. Bootstrap Adaptive Trimmed Means

In this section, we present three bootstrap adaptive trimmed means. By a bootstrap adaptive trimmed mean is meant a trimmed mean whose trimming proportion is chosen to minimize bootstrap estimates of a characteristic (of spread, most of the time) of the (finite or asymptotic) distribution of the trimmed mean. Let's begin with some notation.

Let $X_1, X_2, \ldots, X_n$ denote a sample of identically and independently distributed random variables according to the distribution function $F$. Let $\hat{F}_n$ be the empirical distribution function of the sample. The usual definition of the (symmetric) $\alpha$-trimmed mean $T_n^\alpha$ is given by:

$$T_n^\alpha = (1 - 2[n\alpha])^{-1} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)},$$
where \( \alpha \in [0, 1/2) \), \( \lfloor \cdot \rfloor \) represents the greatest integer function, and \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \) are the order statistics. We shall instead work with the asymptotically equivalent trimmed mean functional given by

\[
T(\alpha, F) = (1 - 2\alpha)^{-1} \int_{\alpha}^{1-\alpha} F^{-1}(t) \, dt.
\]

The \( \alpha \)-trimmed mean of the sample is then \( T(\alpha, \hat{F}_n) \). Note that the two definitions agree for all \( \alpha = i/n \) with \( i = 0, 1, \ldots, \lfloor (n+1)/2 \rfloor - 1 \).

Let’s now see how the trimming proportion will be selected in this paper. Consider a characteristic \( S_n(\alpha, F) \) of the distribution of \( T(\alpha, \hat{F}_n) \), usually a measure of spread. Here we either mean the finite sample distribution or the asymptotic distribution. In the latter case, the subscript \( n \) will later be omitted. As a concrete example, suppose \( S_n(\alpha, F) \) is the exact variance under \( F \) of \( T(\alpha, \hat{F}_n) \). The idea consists of estimating \( S_n(\alpha, F) \) for all values of \( \alpha \) in a given set and to choose \( \alpha_n(\hat{F}_n) \) corresponding to the smallest estimate. But we must first find an estimator of the functional \( S_n(\alpha, F) \).

The bootstrap methodology, introduced by Efron (1979), offers a natural estimate of functionals such as \( S_n(\alpha, F) \), namely \( S_n(\alpha, \hat{F}_n) \). Therefore, let

\[
\alpha_n(\hat{F}_n) = \arg \min_{\alpha \in A_n} S_n(\alpha, \hat{F}_n),
\]

where \( A_n \) is some set of trimming proportions in \([0, 1/2] \). For technical reasons, \( A_n \) will be of the form

\[
A_n = \{ \alpha : \alpha \in [\alpha_0, 1/2 - b_n] \text{ and } n\alpha \in \mathbb{N} \}, \tag{2.1}
\]

and \( \alpha_0 \) and \( b_n \) are constants between 0 and 1/2. Note that \( \alpha_0 \), the lower bound of \( A_n \), is fixed whereas the upper bound \( 1/2 - b_n \) is allowed to vary with the sample size. Typically, the sequence \( \{b_n\} \) will either be constant, or tend to 0 as \( n \to \infty \). The bootstrap adaptive trimmed mean estimate of location is \( T_n(\hat{F}_n) = T(\alpha_n(\hat{F}_n), \hat{F}_n) \). Also, let \( J_n(x, F) \) denote the distribution function of the appropriately normalized bootstrap adaptive trimmed mean \( T_n(\hat{F}_n) \), i.e.,

\[
J_n(x, F) = \Pr_F \{ n^{1/2} [T_n(\hat{F}_n) - T(\alpha_n(\hat{F}_n), F)] \leq x \} = \Pr_F \{ n^{1/2} [T(\alpha_n(\hat{F}_n), \hat{F}_n) - T(\alpha_n(\hat{F}_n), F)] \leq x \}.
\]

Since more than one such adaptive trimmed mean will be introduced, we will use superscripts to differentiate among the different \( S_n, \alpha_n, T_n, \) and \( J_n \).

The asymptotic distribution of the \( \alpha \)-trimmed mean \( T(\alpha, \hat{F}_n) \) for any distribution \( F \) and fixed \( \alpha \) was obtained by Stigler (1973). It is normal with an asymptotic mean of
\( T(\alpha, F) \) and an asymptotic variance \( \sigma^2(\alpha, F) \) if and only if \( F^{-1} \) is continuous at \( \alpha \) and \( 1 - \alpha \). For the special case of a symmetric \( F \) we have

\[
\sigma^2(\alpha, F) = (1 - 2\alpha)^{-2} \left[ \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} (x - T(\alpha, F))^2 dF(x) \\
+ \alpha(F^{-1}(\alpha) - T(\alpha, F))^2 + \alpha(F^{-1}(1-\alpha) - T(\alpha, F))^2 \right]. (2.2)
\]

Note that \( T(\alpha, F) = \theta \) for all \( \alpha \)'s when \( F \) is symmetric about \( \theta \) so that the location of the distribution is unambiguously defined.

If we let \( A_0(F) = \arg \inf_{\alpha \in A} \sigma^2(\alpha, F) \) where \( A \) is a subset of \((0, 1/2)\), then the best trimmed mean with trimming proportion in the set \( A \) is, from an asymptotic point of view, \( T(A_0(F), \hat{F}_n) \). We are interested in functionals \( S_n(\alpha, F) \) such that \( T(\alpha_n(\hat{F}_n), \hat{F}_n) \) is asymptotically as good as \( T(A_0(F), \hat{F}_n) \).

The first candidate should now be clear. Let

\[
S^{\text{JaC}}(\alpha, F) = \sigma^2(\alpha, F),
\]

i.e., the asymptotic variance of the \( \alpha \)-trimmed mean of a sample of \( F \)-distributed random variables. This leads to

\[
\alpha_n^{\text{JaC}}(\hat{F}_n) = \arg \min_{\alpha \in A_n} S^{\text{JaC}}(\alpha, \hat{F}_n).
\]

The corresponding adaptive trimmed mean, \( T_n^{\text{JaC}}(\hat{F}_n) = T(\alpha_n^{\text{JaC}}(\hat{F}_n), \hat{F}_n) \), was introduced by Jaeckel (1971). Note that the asymptotic variance can be estimated by the bootstrap because \( \sigma^2(\alpha, F) \) is an explicit functional of \( F \). This is not the case for many estimators other than the trimmed mean, a fact which greatly limits the extension of this approach.

Instead of estimating characteristics of the asymptotic distribution, it would seem preferable to estimate characteristics of the finite sample distribution. De Jongh and de Wet (1985) were the first ones to suggest using the bootstrap to minimize bootstrap estimates of characteristics of the finite sample distribution of trimmed means, but in the context of trimmed mean regression. Their study of such estimators was limited to a simulation.

Thus, let’s introduce two such estimators. Let’s first define

\[
S_{n}^{\text{Var}}(\alpha, F) = \text{Variance of } n^{1/2}[T(\alpha, \hat{F}_n) - T(\alpha, F)] \text{ under } F;
\]

and,

\[
S_{n}^{\text{IQR}}(\alpha, F) = \text{IQR of } n^{1/2}[T(\alpha, \hat{F}_n) - T(\alpha, F)] \text{ under } F,
\]
where IQR stands for the interquartile range, i.e., the difference between the upper and lower quartiles. (In general define the $q^{th}$ quantile of a distribution $G$ to be $G^{-1}(q) = \inf\{x : G(x) \geq q\}$). Also, let

$$\alpha_n^{\text{Var}}(\hat{F}_n) = \arg\min_{\alpha \in A_n} S_n^{\text{Var}}(\alpha, \hat{F}_n),$$

and

$$\alpha_n^{\text{IQR}}(\hat{F}_n) = \arg\min_{\alpha \in A_n} S_n^{\text{IQR}}(\alpha, \hat{F}_n).$$

So, $\alpha_n^{\text{Var}}(F)$ chooses the trimming proportion that minimizes the variance of the finite sample distribution of $T(\alpha, \hat{F}_n)$, whereas $\alpha_n^{\text{IQR}}(F)$ minimizes its interquartile range. The corresponding adaptive trimmed means will be denoted $T_n^{\text{Var}}(\hat{F}_n)$ and $T_n^{\text{IQR}}(\hat{F}_n)$.

Before discussing the asymptotic behavior of these three bootstrap adaptive trimmed means, the following assumptions on $F$ will be made.

**Assumption C**: Consider a distribution function $F$ and let $0 < \alpha_0 < 1/2$ be fixed.

(C.1) The distribution function $F$ is symmetric about an unknown parameter $\theta$, and it has a density $f$.

(C.2) For some $0 < \epsilon_0 < \alpha_0$, and $f_0 > 0$, $f(x) \geq f_0$ on $\{x : \alpha_0 - \epsilon_0 \leq F(x) \leq 1 - \alpha_0 + \epsilon_0\}$.

(C.3) $\sigma^2(\alpha, F)$ has a unique global infimum over the set $A$ where $A = [\alpha_0, 1/2)$ if $b_n \to 0$ in (2.1), or $A = [\alpha_0, \alpha_1]$ if $b_n$ is constant and $\alpha_1 = 1/2 - b_n$. Then let $A_0(F) = \arg\inf_{\alpha \in A} \sigma^2(\alpha, F)$.

(C.4) $f(\theta)$ is finite.

(C.5) $A_0(F) \neq 1/2$.

The asymptotic distribution of the estimators is given in the next theorem.

**THEOREM 2.1**: Let $F$ satisfy assumptions (C.1) through (C.3).

1. Let $n^{1/4}b_n \to \infty$ as $n \to \infty$. If $b_n \to 0$, also assume (C.5). Then

$$\sup_x |J_n^{\text{Jac}}(x, F) - \Phi(x/\sigma(A_0(F), F))| \to 0 \quad \text{as } n \to \infty.$$

2. Let $n^{1/2}b_n \to \infty$ as $n \to \infty$. If $b_n \to 0$, also assume (C.5). Then

$$\sup_x |J_n^{\text{Var}}(x, F) - \Phi(x/\sigma(A_0(F), F))| \to 0 \quad \text{as } n \to \infty.$$
(3) Let \( n^{1/2}b_n \to \infty \) as \( n \to \infty \). If \( b_n \to 0 \), also assume (C.4) and (C.5). Then
\[
\sup_x |J_{n \to \infty}^F(x, F) - \Phi(x/\sigma_0(F), F)| \to 0 \quad \text{as } n \to \infty.
\]
where \( \Phi(x) \) is the distribution function of the standard normal distribution, \( \sigma^2(\alpha, F) \) is given by (2.2), and \( A_0(F) \) is defined in assumption (C.3).

This theorem states that, under different conditions, the three bootstrap adaptive trimmed means are asymptotically normal with the smallest possible asymptotic variance among trimmed means with a trimming proportion in the set \( A \) defined in assumption (C.3). In other words, by using one of these adaptive trimmed means, one does asymptotically as well as if one knew in advance the best trimming proportion for the given distribution \( F \). The first part of the theorem with \( b_n \) equal to a positive constant is theorem 1 of Jaeckel (1971).

Minimizing the finite sample variance is a natural alternative to minimizing the asymptotic variance as Jaeckel's estimator does. Minimizing the interquartile range is yet another alternative. To obtain results like theorem 2.1, the latter approach involves the convergence of the quantile function of the estimator whereas the former involves the convergence of moments. Consequently, it is expected that an extension of this adaptive method to estimators other than the trimmed mean would succeed more often when the adaptation is done through minimizing bootstrap estimates of the interquartile range rather than bootstrap estimates of variance. Indeed, whenever \( G_n \) is a sequence of distribution functions converging weakly to a continuous and strictly increasing distribution function \( G \), then the quantiles and the interquartile range of \( G_n \) converge to those of \( G \).

A careful look at the proofs of the results in this section and of the next shows that the assumption of symmetry (the first part of (C.1)) is not required. On the other hand, the interpretation of the results is more complex when the underlying distribution is not symmetric. The reason is that the location is no longer unambiguously defined. Each trimmed mean \( T(\alpha, \hat{F}_n) \) estimates a trimmed mean "parameter" \( T(\alpha, F) \) which varies with \( \alpha \). For instance, theorem 2.1 says that the adaptive trimmed means have an asymptotic mean of \( T(A_0(F), F) \) and an asymptotic variance which is as small as possible among trimmed means and is equal to
\[
\sigma^2(\alpha, F) = (1 - 2\alpha)^{-2} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(\min(s, t)) - F(s)F(t) \, ds \, dt.
\]
Note that this formula is consistent with (2.2) for symmetric distributions. The disadvantage is obviously that the functional or parameter being estimated is unknown. So not
only does the adaptive trimmed mean chooses the trimming proportion, but in so doing, it also chooses the functional being estimated!

This completes the asymptotic treatment of the bootstrap adaptive trimmed means as point estimators of the location of a symmetric distribution. The next section is devoted to the construction of (approximate) confidence intervals for the location parameter based on these point estimators.

3. Confidence Intervals

In order to construct an exact confidence interval for the location parameter \( \theta \) based on a bootstrap adaptive trimmed mean \( T_n(\hat{F}_n) \), one needs its distribution \( J_n(x, F) \). Unfortunately, this distribution usually depends on the unknown \( F \), but at least two approaches are possible.

The first one is classic and uses the asymptotic distribution of \( T_n(\hat{F}_n) \). Since \( T_n(\hat{F}_n) \) is asymptotically normal, one can use the family of normal distributions with mean 0 indexed by a scale parameter \( \tau \) as an approximating family of distributions. To complete the approximation, one must estimate the (asymptotic) variance \( \tau^2 \).

In the case of bootstrap adaptive trimmed means this turns out to be a problem. To estimate the variance of \( T(\alpha_n(\hat{F}_n), \hat{F}_n) \), we could either use \( S_{n_{\text{Jae}}}^{\alpha_n(\hat{F}_n), \hat{F}_n} \) or \( S_n^{\var}(\alpha_n(\hat{F}_n), \hat{F}_n) \), i.e., the estimates of asymptotic/finite sample variance for the trimmed mean with trimming proportion equal to \( \alpha_n(\hat{F}_n) \). But had we decided \textit{a priori} to use the trimming proportion \( \alpha = \alpha_n(\hat{F}_n) \), we would use the same estimates of variance. Therefore these estimates do not take into account the adaptive nature of the choice of the trimming proportion. In particular, if we use \( S_{n_{\text{Var}}}^{\alpha_n(\hat{F}_n), \hat{F}_n} \) as an estimate of the variance of \( T_n^{\var}(\hat{F}_n) \), then by construction \( S_n^{\var}(\alpha_n(\hat{F}_n), \hat{F}_n) \) is the smallest bootstrap estimate of variance among all trimmed means. This estimate is biased downwards as it is chosen to be the smallest one. The same comment applies if we use \( S_{n_{\text{Jae}}}^{\alpha_n(\hat{F}_n), \hat{F}_n} \) as an estimate of the variance of \( T_n^{\text{Jae}}(\hat{F}_n) \).

The second approach consists of choosing a member among the family of distributions \( J_n(x, \cdot) \), a family which contains \( J_n(x, F) \) as a member. This amounts to estimating \( F \). The bootstrap choice is to approximate \( J_n(x, F) \) by \( J_n(x, \hat{F}_n) \). For this approach to work, it should be clear that the function \( J_n \) must be smooth in its second argument. In that case, this method has some clear advantages over the standard approximation.

First, the family of approximating functions changes with \( n \) and always includes the distribution being estimated as a member. Secondly, since \( J_n(x, G) \) is the distribution of a
bootstrap adaptive trimmed mean for any distribution function \( G \), this family of approximating distribution functions clearly incorporates the adaptive nature of the estimator. In the event that \( J_n \) is quite smooth in \( F \) and that the convergence of \( J_n \) to \( \Phi \) is slow, it is likely that for small \( n \), \( J_n(x, F) \) will be closer to \( J_n(x, \hat{F}_n) \) than \( \Phi(x, \hat{F}) \).

Let’s see how these two approaches lead to approximate confidence intervals based on \( T_n^{\text{Var}}(\hat{F}_n) \). Let \( Z_\tau \) be distributed according to a normal distribution with mean 0 and variance \( \tau^2 \). In order to construct a nominal level \( 1 - 2\beta \) two-sided confidence interval for the location parameter \( \theta \), consider the following equality:

\[
1 - 2\beta = \text{Prob}_F\{x_\beta^{\text{Var}}(F) \leq n^{1/2}[T_n^{\text{Var}}(\hat{F}_n) - \theta] \leq x_{1-\beta}^{\text{Var}}(F)\}; \tag{3.1}
\]

\[
= \text{Prob}\{\tau z_\beta \leq Z_\tau \leq \tau z_{1-\beta}\},
\]

where \( z_\beta = \Phi^{-1}(\beta) \) is the \( \beta \)th percentile of the standard normal distribution and \( x_\beta^{\text{Var}}(F) = J_n^{\text{Var}}(\beta, F) \) is the \( \beta \)th percentile of the distribution of \( n^{1/2}[T_n^{\text{Var}}(\hat{F}_n) - \theta] \). The approximation based on asymptotic normality consists of inverting the probability statement (3.1) and using \( \tau z_\beta \) instead of \( x_\beta^{\text{Var}}(F) \), and similarly for the upper bound giving the following approximate confidence interval for \( \theta \):

\[
\theta \in [T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2} \tau z_{1-\beta}, T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2} \tau z_\beta]. \tag{3.2}
\]

In order to use that approximation, the approximate variance \( \tau^2 \) must be chosen. A natural choice, given that \( F \) is unknown, is \( S_n^{\text{Var}}(\alpha_n^{\text{Var}}(\hat{F}_n), \hat{F}_n) \). The approximate confidence interval based on the asymptotic normality of \( T_n^{\text{Var}}(\hat{F}_n) \) is given by

\[
\theta \in [T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2} \sqrt{S_n^{\text{Var}}(\alpha_n^{\text{Var}}(\hat{F}_n), \hat{F}_n) z_{1-\beta}}, T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2} \sqrt{S_n^{\text{Var}}(\alpha_n^{\text{Var}}(\hat{F}_n), \hat{F}_n) z_\beta}]. \tag{3.3}
\]

Another estimate of \( \tau^2 \) is \( S_n^{\text{Jac}}(\alpha_n^{\text{Var}}(\hat{F}_n), \hat{F}_n) \). In both cases, the asymptotic coverage of the approximate confidence interval is \( 1 - 2\beta \) as stated in the next theorem.

THEOREM 3.1: Let \( F \) satisfy assumptions (C.1) through (C.3).

(1) Let \( n^{1/4}b_n \to \infty \) as \( n \to \infty \). If \( b_n \to 0 \), also assume (C.5). Then the asymptotic coverage probability of the confidence interval (3.3) with the pair \( \{T_n^{\text{Var}}(\hat{F}_n), \alpha_n^{\text{Var}}(\hat{F}_n)\} \) replaced by \( \{T_n^{\text{Jac}}(\hat{F}_n), \alpha_n^{\text{Jac}}(\hat{F}_n)\} \) is \( 1 - 2\beta \).

(2) Let \( n^{1/2}b_n \to \infty \) as \( n \to \infty \). If \( b_n \to 0 \), also assume (C.5). Then the asymptotic coverage probability of the confidence interval (3.3) is \( 1 - 2\beta \).
(3) Let \( n^{1/2}b_n \to \infty \) as \( n \to \infty \). If \( b_n \to 0 \), also assume (C.4) and (C.5). Then the asymptotic coverage probability of the confidence interval (3.3) with \( \{ T_n^{\text{Var}}(\hat{F}_n), \alpha_n^{\text{Var}}(\hat{F}_n) \} \) replaced by \( \{ T_n^{\text{lar}}(\hat{F}_n), \alpha_n^{\text{lar}}(\hat{F}_n) \} \) is \( 1 - 2\beta \).

(4) Moreover, if \( n^{1/4}b_n \to \infty \) as \( n \to \infty \), then the results of (1) through (3) remain valid if \( S_n^{\text{Var}}(\cdot, \hat{F}_n) \) is replaced by \( S_n^{\text{lar}}(\cdot, \hat{F}_n) \).

Therefore, all variations of the approximate confidence interval (3.3) are asymptotically justified. Next, we will describe bootstrap confidence intervals based on our bootstrap adaptive trimmed means.

In order to describe the bootstrap approximation \( J_n^{\text{Var}}(x, \hat{F}_n) \) we need the following notation and conventions. Let \( X_1, X_2, \ldots, X_n \) be distributed according to \( F \), with corresponding empirical distribution function (e.d.f.) \( \hat{F}_n \). Conditionally on having observed \( \hat{F}_n \), we let \( Y_1, Y_2, \ldots, Y_n \) be distributed according to \( \hat{F}_n \), with corresponding e.d.f. \( \hat{G}_n \). Likewise, let \( Z_1, Z_2, \ldots, Z_n \) be distributed according to \( \hat{G}_n \) (conditionally), with e.d.f. \( \hat{H}_n \). So \( X, Y \), and \( Z \) stand for the random variables for the original, bootstrap, and double bootstrap distributions, respectively, and \( \hat{F}_n, \hat{G}_n, \) and \( \hat{H}_n \) for the corresponding e.d.f.'s.

Thus, we have

\[
J_n^{\text{lar}}(x, \hat{F}_n) = \text{Prob}_{\hat{F}_n}\{n^{1/2}[T(\alpha_n^{\text{lar}}(\hat{G}_n), \hat{G}_n) - T(\alpha_n^{\text{lar}}(\hat{F}_n), \hat{F}_n)] \leq x\}
= \text{Prob}_{\hat{F}_n}\{n^{1/2}[T_n^{\text{lar}}(\hat{G}_n) - T_n^{\text{lar}}(\hat{F}_n)] \leq x\}.
\]

So, as in equation (3.1), we can write

\[
1 - 2\beta = \text{Prob}_{\hat{F}}\{x_1^{\text{Var}}(F) \leq n^{1/2}[T_n^{\text{Var}}(\hat{F}_n) - \theta] \leq x_1^{-\beta}(F)\}
= \text{Prob}_{\hat{F}_n}\{x_1^{\text{Var}}(\hat{F}_n) \leq n^{1/2}[T_n^{\text{Var}}(\hat{G}_n) - T_n^{\text{Var}}(\hat{F}_n)] \leq x_1^{-\beta}(\hat{F}_n)\}.
\] (3.4)

Proceeding as before, we obtain the following approximate \( 1 - 2\beta \) level bootstrap two-sided confidence interval:

\[
\theta \in [T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2}x_1^{\text{Var}}(\hat{F}_n), T_n^{\text{Var}}(\hat{F}_n) - n^{-1/2}x_1^{-\beta}(\hat{F}_n)].
\] (3.5)

Note that this is not the percentile bootstrap confidence interval which, in this notation, can be written as:

\[
\theta \in [T_n^{\text{Var}}(\hat{F}_n) + n^{-1/2}x_1^{\text{Var}}(\hat{F}_n), T_n^{\text{Var}}(\hat{F}_n) + n^{-1/2}x_1^{-\beta}(\hat{F}_n)].
\] (3.6)

These intervals are also asymptotically justified based on the methods of the paper. For a review of bootstrap confidence intervals, see DiCiccio and Romano (1988c). Intervals
such as (3.5) have been referred to as the bootstrap pivotal method (Tibshirani (1984)), or the hybrid method (Hall (1988)).

From a computational point of view, here is how the bootstrap confidence interval is computed. In this problem, as in most problems where the bootstrap is used in practice, the bootstrap quantiles \( z^\text{Var}_\beta(\hat{F}_n) \) cannot be expressed in closed form. We must therefore resort to a Monte Carlo simulation to find an estimate \( \hat{z}^\text{Var}_\beta(\hat{F}_n) \). Note that \( \hat{z}^\text{Var}_\beta(\hat{F}_n) \) is an estimate of \( z^\text{Var}_\beta(F) \), which is itself an estimate for \( z^\text{Var}_\beta(F) \). The simulation is done as follows.

Let the e.d.f. of the sample of size \( n \) from \( F \) be \( \hat{F}_n \). From \( \hat{F}_n \), obtain \( B_1 \) bootstrap samples of size \( n \) with e.d.f. \( \hat{G}^i_n \). From each of these bootstrap samples (with e.d.f. \( \hat{G}^i_n \)), obtain \( B_2 \) further bootstrap samples with e.d.f. \( \hat{H}^{i,j}_n \). For fixed \( i \) and each \( \alpha \in A_n \), compute the variance of the \( B_2 \) estimates of location \( T(\alpha, \hat{H}^{i,j}_n) \). This variance is then \( S^\text{Var}(\alpha, \hat{G}^i_n) \), the bootstrap estimate of variance of \( T(\alpha, \hat{G}^i_n) \). Letting \( \alpha^\text{Var}(\hat{G}^i_n) \) be the trimming proportion corresponding to the smallest bootstrap estimate of the variance of \( T(\alpha, \hat{G}^i_n) \), we compute the \( B_1 \) differences \( T(\alpha^\text{Var}(\hat{G}^i_n), \hat{G}^i_n) - T(\alpha^\text{Var}(\hat{F}_n), \hat{F}_n) \). The estimates \( \hat{z}^\text{Var}_\beta(\hat{F}_n) \) and \( \hat{z}^\text{Var}_\beta(\hat{F}_n) \) are the \([1-\beta]B_1\)th and \([\beta B_1]\)th ordered such differences, respectively.

Note that this confidence interval involves a double bootstrap. A first level of bootstrapping (the inner loop) selects the trimming proportion while the second level (the outer loop) computes the adaptive trimmed mean of the bootstrap samples. It must be stressed that in practice \( B_2 \) need not be as large as \( B_1 \) (which is usually of the order of 1000) as we do not necessarily need a good estimate of the variance, but only a good estimate of the value of \( \alpha \) which minimizes these estimates. So we are only interested in the ordering of the different estimates. In particular, it wouldn't matter if those estimates were biased, as long as the bias is constant over \( \alpha \).

Let's now state the result establishing the asymptotic validity of the bootstrap approximations.

**THEOREM 3.2:** Let \( F \) satisfy assumptions (C.1) through (C.3).

(1) Let \( n^{1/4}b_n \to \infty \) as \( n \to \infty \). If \( b_n \to 0 \), also assume (C.5). Then

\[
\sup_x |J_n^\text{Jac}(x, F) - J_n^\text{Jac}(x, \hat{F}_n)| \to 0 \quad \text{in probability,}
\]

as \( n \to \infty \).
(2) Let $n^{1/2}b_n \to \infty$ as $n \to \infty$. If $b_n \to 0$, also assume (C.5). Then
\[
\sup_x |J_n^{\text{Var}}(x, F) - J_n^{\text{Var}}(x, \hat{F}_n)| \to 0 \quad \text{in probability},
\]
as $n \to \infty$.
(3) Let $n^{1/2}b_n \to \infty$ as $n \to \infty$. If $b_n \to 0$, also assume (C.4) and (C.5). Then
\[
\sup_x |J_n^{\text{lqr}}(x, F) - J_n^{\text{lqr}}(x, \hat{F}_n)| \to 0 \quad \text{in probability},
\]
as $n \to \infty$.

Basically, this result says that with high probability $J_n(x, F)$ and $J_n(x, \hat{F}_n)$ will be uniformly close when $n$ is large. This result is about the distribution functions. It is easy to see that this implies that for fixed $\beta$, the two corresponding quantiles, $x_\beta(F)$ and $x_\beta(\hat{F}_n)$, are also asymptotically close which is sufficient to get the following result:

**Corollary 3.1:** Under the conditions of Theorem 3.2, the bootstrap confidence interval (3.5) and the corresponding intervals based on $T_n^{\text{Jae}}(\hat{F}_n)$ and $T_n^{\text{lqr}}(\hat{F}_n)$ have an asymptotic coverage probability of $1 - 2\beta$.

Therefore, both standard confidence intervals such as (3.3) and bootstrap intervals such as (3.5) have asymptotically the claimed coverage probability. So they are both said to be consistent. But this is not sufficient to conclude that one method is superior to the other, either asymptotically or for finite samples. Second order results would be rather difficult to obtain because of the adaptive choice of the trimming proportion. These results usually require the existence of an Edgeworth expansion for $J_n(x, F)$. Bjerve (1974) showed the existence of such an Edgeworth expansion for the distribution of the $\alpha$-trimmed mean. It is unclear whether his lengthy argument could be generalized to an adaptive trimmed mean.

Second order correct methods usually involve more work. For instance, the BC$_\alpha$ interval of Efron (1987) requires the calculation of a so-called acceleration constant which would undoubtedly be difficult to compute in this adaptive setting. On the other hand, the percentile-t as discussed, for instance, in Hall (1988) could, in principle, be computed. Consider the confidence interval based on $T_n^{\text{Jae}}(\hat{F}_n)$. Instead of bootstrapping the root $T_n^{\text{Jae}}(\hat{F}_n) - T_n^{\text{Jae}}(F)$, the percentile-t consists of bootstrapping a “studentized” root $(T_n^{\text{Jae}}(\hat{F}_n) - T_n^{\text{Jae}}(F))/\sigma(\alpha_n^{\text{Jae}}(\hat{F}_n), \hat{F}_n)$. Therefore, for each bootstrap sample, one computes an estimate of the asymptotic variance of Jaeckel’s estimator. Note that this
estimate is readily available, having been computed in order to determine \( \alpha_n^{Iae} \). Letting \( K_n^{Iae}(\cdot,F) \) be the distribution function of \( n^{1/2}[T_n^{Iae}(\hat{F}_n) - T_n^{Iae}(F)]/\sigma(\alpha_n^{Iae}(\hat{F}_n), \hat{F}_n) \), and \( y_\beta^{Iae}(F) = K_n^{Iae}^{-1}(\beta,F) \), the percentile-t interval is given by

\[
\theta \in [T_n^{Iae}(\hat{F}_n) - n^{-1/2}y_\beta^{Iae}(\hat{F}_n)\sigma(\alpha_n^{Iae}(\hat{F}_n), \hat{F}_n), T_n^{Iae}(\hat{F}_n) - n^{-1/2}y_{1-\beta}^{Iae}(\hat{F}_n)\sigma(\alpha_n^{Iae}(\hat{F}_n), \hat{F}_n)].
\]

But although this interval is consistent based on the methods of this paper, it need not be second order correct as the sufficient conditions of Hall (1988) are not satisfied. Another second order correct method, the automatic percentile, due to DiCiccio and Romano (1989, 1988a,b) also requires the same type of conditions.

In the next section we present the results of a small sample simulation which compares the variance of the different bootstrap adaptive trimmed means and also compares the two methods of constructing approximate confidence intervals based on these estimators.

4. Small Sample Simulations

In this section we will study the small sample behavior of the three adaptive trimmed means introduced in section 2. We will simulate the distribution of the bootstrap adaptive trimmed means and of the trimmed means with fixed trimming proportion when the observations come from one of four sampling plans. First, the observations will be normally distributed. As is well known, the mean has the smallest variance among all trimmed means for every sample sizes. Note that according to the asymptotic theory developed in the previous section, the mean should not be one of the possible candidate of our adaptive estimators. But by letting \( \alpha_0 \) be arbitrarily close to 0, one could get an asymptotic variance arbitrarily close to that of the mean.

The second sampling plan is known as the one-wild. It consists of \( n - 1 \) independent standard normals and of an independent normal random variable with variance equal to 100. It is supposed to be a good approximation to real life samples which often include an outlier. It does not fit the scheme of i.i.d. samples, but it is often included in robustness studies.

The next distribution, known as the slash, is the ratio of a normal to an independent uniform on (0,1). It has been included in numerous simulation studies of robustness as a very long-tailed distribution without moments.

The other distribution under consideration is the double exponential. With tails between that of the Gaussian and the slash, it has moments of all orders. As is well known, its maximum likelihood estimator is the median and so is asymptotically the
best trimmed mean for that distribution. This contradicts assumption (C.5). On the other hand, it does not minimize the finite sampling variance among the trimmed means with fixed trimming proportion. These four rather different sampling plans should clearly demonstrate the flexibility (or lack of) of the adaptive trimmed means.

The three adaptive trimmed means $T^\text{JAE}_n(\hat{F}_n)$, $T^\text{Var}_n(\hat{F}_n)$, and $T^\text{IQR}_n(\hat{F}_n)$ will be denoted JAE, BVAR, and BIQR respectively. To complete the definition of these adaptive trimmed means, one has to specify the interval $[\alpha_0, \alpha_1]$ of possible trimming proportions. We shall use three different intervals: $[0, .25]$, $[0, .50]$, and $[.10, .40]$. The nine different estimators will be denoted by JAE 0-25, JAE 0-50, \ldots, BIQR 10-40.

The estimator JAE 0-25 was investigated in the Princeton Robustness Study (Andrews et al. (1972)) which explains the inclusion of the 0-25 estimators. Since the estimators 0-50 include all trimmed means that trim an integer number of observations, from the mean to the median, it is also a natural class to include, even though they do not satisfy the conditions of section 3 ($\alpha_0 > 0$ and $\alpha_1 \to 1/2$ at a certain rate). Only the estimators 10-40 satisfy these conditions.

Finally, we also include all trimmed means with fixed trimming proportion $k/n$ for $k = 0, 1, \ldots, n-1$. They will be denoted by T0, T5, and so on, where the number indicates the percentage of trimming.

The sample sizes considered in this simulation are 10, 20, and 40 for all nine estimators and four sampling situations.

The simulations were performed on a SUN 3/50 workstation. For each distribution and sample size, 10,000 samples were generated. To increase accuracy, the score swindle of Johnstone and Velleman (1985) was used. The Gaussian and double exponential random variables were generated by the ziggurat method of Marsaglia and Tsang (1984). Uniform random variables were generated by the IMSL function GGUBFS which is a linear congruential generator.

Both BVAR and BIQR require Monte Carlo approximations for their bootstrap estimates of finite sample variance and interquartile range. This approximation is based on 100 bootstrap samples.

Consider now the question of performance. By the unbiasedness of the estimators under consideration, it seems reasonable to use the variance as the basis for comparison. Moreover, an adaptive estimator should be judged by its ability to use the observations to construct an estimator which is appropriate for the underlying distribution. A good yardstick is its efficiency defined to be the ratio of the variance of the best trimmed mean
to its variance (multiplied by 100 to express it as a percentage). Here we define the best trimmed mean to be the trimmed mean with fixed trimming proportion in \([0, .5]\) with the smallest variance for the sampling situation and sample size under study.

Table 4.1 contains the efficiencies of the 9 adaptive trimmed means and of all trimmed means with fixed trimming proportion \(k/10\) for \(k = 0, 1, \ldots, 4\). The sample size is 10. Results for samples of size 20 and 40 are similar and those tables are omitted for brevity. The interested reader will find them in Léger (1988). The variances of the best trimmed means for sample sizes 10 and 20 were obtained from chapter 10 of “Understanding Robust and Exploratory Data Analysis” (UREDA) edited by Hoaglin, Mosteller, and Tukey (1983) where they are computed exactly. The variances of trimmed means for samples of size 40 have not been computed exactly. Therefore, the best trimmed means for that sample size were chosen to be the trimmed mean with fixed trimming proportion \(k/40\) where \(k = 0, 1, \ldots, 19\), such that the Monte Carlo estimate of variance is minimized, with this estimate of variance being used in the calculation of the efficiencies.

For all sample sizes, the reference estimator for the normal distribution is the mean. In the case of the one-wild, the reference estimators are the 16%, 9%, and 5% trimmed means for samples of size 10, 20, and 40 respectively. The reference estimators for the slash are the 38%, 34%, and 32.5% trimmed means for the respective sample sizes. Finally, the 34%, 37%, and 40% trimmed means are the reference estimators for the double exponential.

Let’s look at table 4.1 beginning with Jaeckel’s estimators JAE. First, note that JAE 0-25 behaves very differently from the other two. It does pretty well, especially compared with the other 0-25 estimators. It does not do as well as the other 0-25 estimators at the Gaussian and one-wild situations, but does better at the slash and double exponential. This illustrates the fact that the estimators JAE overemphasize the large trimming proportions. Whereas this hurts for the normal and the one-wild where the best trimming proportion is small, it does not hurt enough to cancel the benefits obtained at the slash and the double exponential because the trimming proportion is at most 25%. Recall that trimming proportions in the 30%-40% range minimizes the finite sample variance of the \(\alpha\)-trimmed means for the slash and the double exponential distributions. While such an emphasis on larger trimming proportion is beneficial for JAE 0-25, it is disastrous for JAE 0-50, and JAE 10-40. Note, for instance, that the efficiency at the Gaussian for a sample of size 10 is 74.9 and 74.3 for JAE 0-50 and JAE 10-40, respectively. This should be contrasted with a range of efficiencies of 72.2 to 100.0 for \(\alpha \in [0, .5]\) and a range of 72.2 to 94.9 for \(\alpha \in [.1, .4]\). Qualitatively, the results are the same for samples of size 20 and 40. A careful look at the (simulated) distribution of \(S^{JAE}(\alpha, \hat{F}_n)\) shows that it can
only be trusted for $\alpha \in [0, .25]$. For larger trimming proportions, the estimates are biased downwards. Remember from theorem 2.1 that the rate at which the largest trimming proportion may converge to 1/2 is slower for Jaeckel's estimator than for the other two bootstrap adaptive trimmed means. That asymptotic result may in part explain the relatively poor performance of JAE 0-50 and JAE 10-40. Because of their poor performance at the normal distribution, JAE 0-50 and JAE 10-40 will be ignored from the rest of the discussion.

Figure 4.1 contains plots of the efficiency of the remaining seven adaptive estimators against sample size for the four sampling situations. So for this plot, we used the results for the three different sample sizes. Notice that full lines are used for the 0-25 estimators, whereas short broken lines and long broken lines are used for the 0-50 and 10-40 estimators, respectively.

The first thing to note is that the 0-25 estimators are usually the best for the normal and one-wild while they are usually the worst for the other two distributions. Obviously, the adaptive trimmed means perform better when the set of trimming proportions considered contains the optimal one and is as small as possible. A disturbing fact is the apparent loss of efficiency as the sample size increases (except for the slash). However, this must be contrasted with the fact that the efficiency of the worst trimmed mean also decreases with the sample size. For instance, the median is the worst trimmed mean for the normal distribution and it has an efficiency of 72.3, 68.0, and 65.9 for 10, 20, and 40 observations, respectively. The same phenomenon applies with even bigger differences to the one-wild and double exponential whereas the mean has an efficiency of 0 for the slash for all sample sizes.

The plot clearly shows the difficulty of adaptation at the slash distribution with a small sample size. Since the first and last two order statistics have an infinite variance, the mean and a trimmed mean that trims only one observation on each side both have an infinite variance. So for samples of size 10, the efficiency of T0 and T10 is 0. Despite 10,000 simulated samples and a swindle, the estimated efficiency of T10 is 18.7! Since the estimators 0-25 are selecting among three trimmed means, two of which have an infinite variance that must be estimated with only 10 observations, it is not surprising to find out that they are not very efficient. But the performance is rapidly improving as the sample size increases.

Figure 4.1 shows that the bootstrap adaptive trimmed means can indeed adapt themselves rather well, even in samples as small as 10. It also shows that the adaptive trimmed means that minimizes bootstrap estimates of characteristics of the finite sample distri-
bution outperform those minimizing characteristics of the asymptotic distribution. By computing polyefficiencies as in Tukey (1979), i.e., the minimum efficiency over a set of sampling situations, we find out that the "best" bootstrap adaptive trimmed means are better than the "best" trimmed means with fixed trimming proportion even for samples of size 10. For details, see Léger (1988). Hence, even for small sample sizes, adaptation pays.

Remark 4.1: Two other classes of estimators were also included in the simulations. The first one used jackknife estimates of variance to select the trimming proportion. Their efficiencies were very close to that of the Jaeckel estimators. The reason is simple. Efron (1979, 1982) has shown that the infinitesimal jackknife estimate of variance of a statistic is a (bootstrap) estimate of the variance of the first order approximation of that statistic, and hence an estimate of the asymptotic variance of the statistic. Since the ordinary jackknife estimate is an approximation of the infinitesimal jackknife estimate, then the ordinary jackknife estimate of the variance of the $\alpha$-trimmed mean is an approximation to Jaeckel's estimate of the asymptotic variance. The second class of estimators consisted of adaptive trimmed means whose trimming proportion is chosen though cross-validation. Cross-validation is often used to select among competing models or estimators. In this case, it was not too successful. The cross-validation estimate of variance works as follows. Leaving each data point in turn, it predicts that data point by the trimmed mean of the remaining observations and takes the sum of squares of the difference between the observation and its predicted value. The problem is that if the distribution doesn't have a variance, then the cross-validation estimate of variance of the $\alpha$-trimmed mean will have an infinite expectation even though the $\alpha$-trimmed mean has a finite variance! Therefore this adaptive trimmed mean is highly non-robust. Moreover, in a theoretical study, Pruitt (1988) showed that cross-validation doesn't work in this setting.

Remark 4.2: The bootstrap estimators BVAR and BIQR were based on 100 bootstrap replications. Preliminary simulations were done using 1,000 bootstrap samples leading, most of the time, to slightly better results. But the gains/losses were usually of the order of 1%. Such gains do not seem to be worth a ten-fold increase in computation time. Furthermore, it might even be possible to reduce the number of bootstrap samples even further. Recall that only the ordering of the variance (or interquartile range) estimates matter, not their precise values. Therefore, bootstrap sample sizes of 25 might be sufficient, although no simulations results are available to support this claim.

Remark 4.3: Note that BIQR selects the trimming proportion that minimizes the
length of a 50% bootstrap confidence interval. The theory would also hold for any other confidence level. By looking at the (simulated) distribution of the trimmed means under the various sampling plans, one finds out that the upper and lower quartiles do not seem to vary much as a function of the trimming proportion, not as much as the upper and lower deciles, for instance. Therefore, minimizing the length of a central 80% confidence interval might lead to better small sample results than minimizing the interquartile range. But one must keep in mind that there is more variability in estimating the 10th and 90th percentiles than the upper and lower quartiles, so that the (possibly) higher variability of the length of an 80% bootstrap confidence interval might actually lead to a worse choice of the trimming proportion, especially when those estimates come from only 100 bootstrap samples.

Let us now study the small sample performance of confidence intervals for the location parameter based on adaptive trimmed means.

Four different confidence intervals were considered. A standard (asymptotic) confidence interval based on JAE 0-25, a bootstrap interval based on JAE 0-25, a bootstrap interval based on BVAR 0-25, and a bootstrap-t interval also based on BVAR 0-25. The last two intervals involve a double bootstrap as one level of bootstrapping is required to select the adaptive trimming proportion. The first and second intervals are formulas (3.3) and (3.5) with all superscripts Var replaced by the superscript Jac, respectively. The third interval is (3.5), while the last one is (3.7) with the appropriate modifications.

Because of the tremendous amount of computations involved, only 500 samples of size 10 were simulated at each of three different distributions. They are the standard normal, the slash, and the double exponential. The estimates BVAR 0-25 were based on 100 bootstrap replications. The bootstrap confidence intervals were computed from 1,000 bootstrap samples. Hence, for a single sample of 10 observations, the bootstrap confidence interval based on JAE 0-25 requires 1,000 bootstrap samples of size 10, whereas the last two confidence intervals require 100,000 bootstrap samples. This must be contrasted with the standard interval based on JAE 0-25 which does not require any resampling.

For each of the simulated samples, upper 5% and 95% one-sided confidence intervals and the resulting 90% two-sided confidence interval were computed. In order to compare the four different methods, the estimated coverage probability of the three intervals were computed along with the mean and the standard deviation of the length of the 90% two-sided confidence intervals. Of course, whenever two different methods lead to intervals with very different coverage probabilities, comparing the length of the confidence intervals is not very useful, but in the other case, the length can be a useful measure.
The results of the simulations are found in table 4.2. Let's start with the case of an underlying Gaussian distribution. The coverage probabilities for the standard intervals based on JAE 0-25 are way off the mark. The actual distribution of the root $n^{1/2}[T_n(\hat{F}_n) - \theta]$ is much wider than its asymptotic approximation. Therefore, there is more than twice as much probability in each tail than claimed, leading to a coverage probability for the 90% two-sided confidence interval of 77.2%. Note that the standard deviation for these estimates is at most 2.2%. Such a result is highly undesirable, but not very surprising. As was noted in section 3, this interval fails to incorporate the adaptive nature of the estimator.

The results for the bootstrap intervals based either on JAE 0-25 or on BVAR 0-25 are better, although not great with coverage probabilities for the two-sided 90% interval of 82.8% and 83.0%. This is still well under the target value. On the other hand, the results for the bootstrap-t intervals based on BVAR 0-25 are outstanding.

The results for the slash distribution are slightly different. The standard interval are essentially as bad as they were in the case of the normal distribution, whereas the bootstrap intervals of JAE 0-25 and BVAR 0-25 are much better. They are within one standard deviation of their target for the one-sided intervals and slightly over one standard deviation for the two-sided interval. On the other hand, the bootstrap-t doesn't do as well as it did for the Gaussian. While it has reasonable coverage probabilities, it seems to be outperformed by the other two bootstrap intervals.

The results about the length of the two-sided intervals are much harder to interpret. The standard interval is on average the smallest, followed by the bootstrap-t which is more than twice as large and has a much larger standard deviation. Then the ordinary bootstrap intervals follow with a mean 4 to 6 times larger than the bootstrap-t and with standard deviations much larger still. If these numbers are a good reflection of the reality (remember that the slash distribution does not have any moment and that only 500 samples of size 10 were simulated) then it certainly is tempting to use the bootstrap-t interval with a slightly lower coverage probability but a much narrower interval, on average, than the other two bootstrap intervals.

Finally, let's look at the double exponential distribution. Once again, the standard intervals are overly optimistic. The bootstrap intervals based on JAE 0-25 are better than those based on BVAR 0-25. It is not clear whether the difference is real, but if it is, it might be explained by the much higher efficiency of JAE 0-25 over BVAR 0-25 at the double exponential distribution for samples of size 10. Also, the bootstrap-t ended up between the other two bootstrap intervals. As for the lengths, even though the three bootstrap
intervals have pretty much the same coverage probability for the double-sided interval, the bootstrap-t is significantly larger and more variable than the other two. That the bootstrap-t seems to have larger intervals has been reported in other simulation studies, e.g. Loh (1987).

What should be concluded from these simulations? First, the standard intervals cannot be relied upon. They are much too optimistic. Second, the bootstrap confidence intervals have reasonable coverage probabilities, especially in view of the fact that they were based on samples of size 10. Third, the bootstrap-t intervals sometimes do much better than the ordinary bootstrap intervals, while they do worse for some other distributions. In view of the fact that the bootstrap interval based on JAE 0-25 only involves 1,000 bootstrap samples instead of the 100,000 of the other two, it seems clear that it should be the preferred choice.

This completes our study of the small sample behavior of the bootstrap adaptive trimmed means and their related confidence intervals.

5. Conclusion

In this paper we have shown that one can successfully use the bootstrap to select the trimming proportion of an adaptive trimmed mean. In short, an adaptive trimmed mean with a trimming proportion that minimizes bootstrap estimates of the finite sample variance or interquartile range of the distribution of the trimmed mean has an asymptotic variance which is the smallest among trimmed means. By using bootstrap estimates of characteristics of the finite sample distribution rather than characteristics of the asymptotic distribution, the bootstrap adaptive trimmed means introduced in this paper have smaller small sample variances.

The importance of the results in this paper is not as much the introduction of a better adaptive trimmed mean or a better location estimator for that matter, but more the illustration of the use of the bootstrap in selecting an adaptive procedure. Numerous estimators in different problems, ranging from the estimation of a finite dimensional location parameter to the infinite dimensional estimation of a density, are only specified up to a constant — the smoothing parameter. In most cases, the optimal choice of that parameter is a function of the underlying distribution of the observations and of the sample size which calls for a data-selected value of the smoothing parameter.

One way to do this is by choosing the value of the smoothing parameter such that the estimator optimizes an estimate of a certain criterion. Up until recently, estimates of characteristics of the finite sample distribution of most estimators were hard to get because
that distribution is unknown. On the other hand, in some cases, closed form estimates of characteristics of the asymptotic distribution are known, such as the asymptotic variance of the trimmed means. Thus, adaptive methods based on optimizing characteristics of the asymptotic distribution of the estimators were introduced. But the bootstrap can often be used to estimate characteristics of the finite sample distribution. By using them, it is possible to optimize estimates of a finite sample criterion, often a more meaningful criterion. Not only is it a viable alternative to using an asymptotic criterion, but sometimes it is the only alternative as no closed form estimator of characteristics of the asymptotic distribution exists.

This paper illustrates this approach by looking at trimmed means in the location problem. Trimmed means have also been defined in the regression problem, e.g., Bickel (1973), Koenker and Bassett (1978), and Ruppert and Carroll (1980), and Welsh (1987). It seems that the theory developed for bootstrap adaptive trimmed means in the location problem should easily generalize to Welsh’s definition of trimmed mean regression. The theory (which will be outlined in the appendix) shows that if the estimators vary smoothly in the smoothing parameter and the distribution of the observations, then it is possible to use the bootstrap to adaptively select the smoothing parameter. For a general theory and several other examples, see Léger and Romano (1989).

The other important point is rather old and well-known, but can never be emphasized enough, especially in view of the availability of methods to deal with the problem. Do not make inference statements based on an adaptive procedure by assuming that the adaptively chosen parameter was a prior fixed to the adaptive choice. In the example that was investigated, this translates into not using a confidence interval that does not use the adaptive nature of the estimator in the determination of the bounds of the intervals. The confidence interval based on the asymptotic normality of the adaptive trimmed mean does not take into account its adaptivity and leads to coverage probabilities consistently at least 12% lower than the claimed 90% for the three sampling situations studied. On the other hand, bootstrap confidence intervals that share the same asymptotic properties consistently did better in small samples. Such intervals are based on the finite sample distribution of the same adaptive statistic, but for data whose distribution is different, namely, $\hat{F}_n$ instead of $F$. Hence, the adaptivity of the estimators is clearly reflected.

6. Appendix

In this section, we will outline the proofs of the results of the previous sections. The details can be found in Léger (1988).
As will soon be clear, the most important result to prove is theorem 3.2. This result is an immediate consequence of theorem 2.1 and of the analogous result where \( J_n(x, F) \) is replaced by \( J_n(x, \hat{F}_n) \) with the extra qualifier that the convergence is now in probability. The randomness of \( \hat{F}_n \) makes the latter much harder to prove. Fortunately, Beran (1984) introduced a method of proof that greatly simplifies things. The idea consists of defining a class \( C_F \) of sequences of fixed distribution functions \( \{F_n\} \) such that if \( \{F_n\} \in C_F \) then theorem 2.1 is valid with \( J_n(x, F_n) \) replacing \( J_n(x, F) \), i.e.,

\[
\sup_x |J_n(x, F_n) - J(x, F)| \to 0, \tag{A.1}
\]

where \( J(x, F) = \Phi(x/\sigma(A_0, F)) \) is the asymptotic distribution function of the adaptive trimmed mean as given in theorem 2.1. In that case, if \( \{\hat{F}_n\} \in C_F \) with probability 1, then

\[
\sup_x |J_n(x, \hat{F}_n) - J(x, F)| \to 0, \quad \text{with probability 1.}
\]

But even if \( \{\hat{F}_n\} \notin C_F \), then using Skorohod's representation theorem there often exists a random sequence of distribution functions \( \{\hat{F}^*_n\} \) such that \( \{\hat{F}^*_n\} \in C_F \) with probability 1, and \( \hat{F}^*_n \) and \( \hat{F}_n \) have the same distribution when regarded as random variables on some appropriate space (say \( D(-\infty, \infty) \)). Therefore,

\[
\sup_x |J_n(x, \hat{F}^*_n) - J(x, F)| \to 0, \quad \text{with probability 1.}
\]

But since \( \hat{F}^*_n \) and \( \hat{F}_n \) have the same distribution, this implies that

\[
\sup_x |J_n(x, \hat{F}_n) - J(x, F)| \to 0, \quad \text{in probability,}
\]
as desired. Thus the goal consists of defining a class \( C_F \) such that (A.1) holds and such that Skorohod's representation can be used. Note that by letting \( F_n = F \), theorem 2.1 will then automatically be shown.

So we are lead to study the asymptotic behavior of \( J_n(x, F_n) \) which is the distribution function of an adaptive trimmed mean. The key will be the stochastic equicontinuity of the trimmed mean process and the convergence in probability of \( \alpha_n(\hat{G}_n) \) to \( A_0(F) \) where \( \hat{G}_n \) is the empirical distribution function of a sample of size \( n \) from \( F_n \).

Consider a stochastic process \( R_n(\alpha, \hat{G}_n, F_n) \) indexed by the parameter \( \alpha \) with \( \hat{G}_n \) as defined above. We say that this stochastic process is \( F_n \)-stochastically equicontinuous at \( \alpha^* \) (Pollard (1984)) if for each \( \eta > 0 \) and \( \varepsilon > 0 \) there exists a neighborhood \( U \) of \( \alpha^* \) for which

\[
\limsup_{n \to \infty} \text{Prob}_{F_n} \left\{ \sup_{\alpha \in U} |R_n(\alpha, \hat{G}_n, F_n) - R_n(\alpha^*, \hat{G}_n, F_n)| > \eta \right\} < \varepsilon.
\]
An immediate consequence of that definition is the following lemma:

**Lemma A.1:** Suppose that

1. $\alpha_n(\hat{G}_n) \to \alpha^*$ in $F_n$-probability,
2. $R_n(\alpha, \hat{G}_n, F_n)$ is $F_n$-stochastically equicontinuous at $\alpha^*$, and,
3. The laws of $R_n(\alpha^*, \hat{G}_n, F_n)$ and $R_n(\alpha^*, \hat{F}_n, F)$ both converge weakly to the law of $R(\alpha^*, F)$.

Then the law of $R_n(\alpha, \hat{G}_n, \hat{G}_n, F_n)$ converges weakly to the law of $R(\alpha^*, F)$.

Note that proving that a non-adaptive statistic can be bootstrapped only requires condition 3. The other two conditions guarantee that the process is smooth enough and that the adaptive choice of the smoothing parameter converges in probability to a constant.

The program now consists of showing that conditions 1, 2 and 3 are satisfied for the trimmed mean process $M_n(\alpha, \hat{G}_n, F_n) = n^{1/2}(T(\alpha, \hat{G}_n) - T(\alpha, F_n))$ and the different adaptive trimming proportions $\alpha_n(\hat{G}_n)$. First, let's define $C_F$. We say that $\{F_n\} \in C_F$ provided that there is a constant $D$ such that for all large $n$,

$$n^{1/2} \sup_x |F_n(x) - F(x)| \leq D. \quad (A.2)$$

It can easily be shown that this condition implies that for all large $n$,

$$n^{1/2} \sup_{\alpha \in [\alpha_0, 1-\alpha_0]} |F_n^{-1}(\alpha) - F^{-1}(\alpha)| \leq \frac{D + 1}{f_0}, \quad (A.3)$$

where $\alpha_0$ and $f_0$ are defined in assumption C. Whenever we will refer to assumption C, we mean assumption C from section 2.

The next result justifies the use of sequences of fixed distribution functions $\{G_n\}$ and $\{F_n\}$ satisfying (A.2).

**Lemma A.2:** Let $\{F_n\}$ satisfy condition (C) for some distribution function $F$. Let $\hat{G}_n$ be the empirical distribution function of a sample of size $n$ from $F_n$. Then there exists $\{\hat{G}_n\}$ with the same distribution as $\{\hat{G}_n\}$ such that it satisfies (A.2) almost surely.

The proof uses results of Shorack and Wellner (1986) on empirical processes. The next theorem verifies condition 2.

**Theorem A.1:** Let $F$ satisfy assumption C and suppose that $\{F_n\} \in C_F$. Then $M_n(\alpha, \hat{G}_n, F_n)$ is $F_n$-stochastically equicontinuous for any $\alpha \in [\alpha_0, 1/2)$. 

22
Proof: Let \( \alpha_1 \in [\alpha_0, 1/2) \) be given. Consider any \( \alpha_1 < \alpha_2 < 1/2 \). After some algebra one gets

\[
|M_n(\alpha_2, \hat{G}_n, F_n) - M_n(\alpha_1, \hat{G}_n, F_n)| \\
\leq n^{1/2} \int_{\alpha_1}^{1-\alpha_1} \left| \hat{G}_n^{-1}(t) - F_n^{-1}(t) \right| \left| \frac{I(\alpha_2 \leq t \leq 1 - \alpha_2)}{1 - 2\alpha_2} - \frac{I(\alpha_1 \leq t \leq 1 - 2\alpha_1)}{1 - 2\alpha_1} \right| dt \\
\leq n^{1/2} \sup_{t \in [\alpha_1, 1-\alpha_1]} \left| \hat{G}_n^{-1}(t) - F_n^{-1}(t) \right| \\
\times \int_{\alpha_1}^{1-\alpha_1} \left| \frac{I(\alpha_2 \leq t \leq 1 - \alpha_2)}{1 - 2\alpha_2} - \frac{I(\alpha_1 \leq t \leq 1 - 2\alpha_1)}{1 - 2\alpha_1} \right| dt \\
\frac{4(\alpha_2 - \alpha_1)}{1 - 2\alpha_1} n^{1/2} \sup_{t \in [\alpha_1, 1-\alpha_1]} \left| \hat{G}_n^{-1}(t) - F_n^{-1}(t) \right|.
\]

Using Dvoretzky-Kiefer-Wolfowitz’s inequality (Dvoretzky, Kiefer, and Wolfowitz (1956)) and an argument like the one that goes from (A.2) to (A.3), we can prove that there exists a fixed constant \( E \) such that for all large \( n \),

\[
Prob_{F_n} \left\{ n^{1/2} \sup_{t \in [\alpha_0, 1-\alpha_0]} \left| \hat{G}_n^{-1}(t) - F_n^{-1}(t) \right| \leq E \right\} > 1 - \epsilon.
\]

By letting \( \alpha_2 \) be sufficiently close to \( \alpha_1 \), this result implies that the trimmed mean process is \( F_n \)-stochastically equicontinuous at \( \alpha_1 \). \( \blacksquare \)

In order to prove condition 1, the following lemma provides a sufficient condition.

Lemma A.3: Let \( I \) and \( I_n \) for \( n = 1, 2, \ldots \) be subsets of \( \mathbb{R} \). Suppose that \( I_n \) is asymptotically dense for \( I \), i.e., given \( \epsilon > 0 \), for each \( x \in I \) there exist \( N = N(x) \) and \( x_n \in I_n \) such that for all \( n > N \), \( |x - x_n| < \epsilon \). Let \( S(\alpha) \) be a fixed continuous function of \( \alpha \) for all \( \alpha \in I \), and suppose that

\[
\sup_{\alpha \in I_n} |\hat{S}(\alpha) - S(\alpha)| \to 0, \quad \text{in probability}, \quad (A.4)
\]

where \( \hat{S}(\cdot) \) is a random estimate of \( S(\cdot) \). Then

\[
\alpha_n \to A_0 \quad \text{in probability},
\]

where \( \alpha_n = \arg \min_{\alpha \in I_n} \hat{S}(\alpha) \), \( A_0 = \arg \min_{\alpha \in I} S(\alpha) \), and \( S(\alpha) \) is assumed to have a unique global minimum over the closure of \( I \).

This lemma is a slight generalization of lemma 3 in Jaeckel (1971). Here \( \hat{S}(\alpha) = S_n(\alpha, \hat{G}_n) \). To show that (A.4) holds, we first consider fixed \( \{G_n\} \in C_F \) and then use lemma A.2. The next result and its proof are the key to showing (A.4) for the three
adaptive choices of the trimming proportion. Moreover, as an immediate consequence, condition 3 follows.

**Lemma A.4:** Let \( F \) satisfy assumptions (C.1) and (C.2), and suppose that \( \{F_n\} \) satisfies (A.2). If \( b_n \to 0 \), also assume (C.4). Then if \( n^{1/2}b_n \to \infty \) as \( n \to \infty \),

\[
\sup_{x} \sup_{\alpha \in I_{n}} |L_{n}(x; \alpha, F_{n}) - L(x; \alpha, F)| \to 0, \quad \text{as} \quad n \to \infty,
\]

where \( I_{n} = [\alpha_{0}, 1/2 - b_{n}] \), \( L_{n}(x; \alpha, F_{n}) \) is the distribution function of \( M_{n}(\alpha, \hat{G}_{n}, F_{n}) \) treating \( \alpha \) as fixed, and \( L(x; \alpha, F) = \Phi(x/\sigma(\alpha, F)) \) is its asymptotic distribution function.

To prove this result, we differentiate the process \( M_{n} \) for each value of \( \alpha \) as is done in Boos (1979) and Serfling (1980). The linear approximation is shown to converge weakly uniformly in \( \alpha \). This involves showing that the asymptotic variance of the \( \alpha \)-trimmed mean based on \( F_{n} \) converges to \( \sigma^{2}(\alpha, F) \) uniformly in \( \alpha \). Then the remainder is shown to converge to 0 in \( F_{n} \)-probability uniformly in \( \alpha \). The inequality of Dvoretzky, Kiefer, and Wolfowitz (1956) is used to give an upper bound to the rate at which the remainder converges to 0.

From lemma A.4, and the continuity of \( L(x; \alpha, F) \), we obtain the corresponding uniform convergence of the quantile functions, from which we conclude that \( S_{n}^{\text{lar}} \) satisfies (A.4).

To show that \( S_{n}^{\text{Var}} \) satisfies (A.4), write it as the sum of the variance of the linear approximation, the variance of the remainder, and the covariance between the two. We then show that the variance of the remainder converges to 0 (and hence the covariance term) using the rate at which the remainder vanishes in the proof of lemma A.4. Finally in that proof, we also show that the variance of the linear approximation (based on samples from \( F_{n} \)) converges uniformly (in \( \alpha \)) to the asymptotic variance (based on samples from \( F \)).

Remembering that \( S_{n}^{\text{Jae}}(\alpha, F_{n}) \) is the asymptotic variance of a trimmed mean of a sample from \( F_{n} \), use (2.2) to show that each of the three terms converge uniformly in \( \alpha \) to obtain (A.4).
References


Table 4.1

Efficiencies of estimators based on a sample of size 10

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Normal</th>
<th>One-Wild</th>
<th>Slash</th>
<th>Double Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAE 0-25</td>
<td>92.0</td>
<td>98.6</td>
<td>64.6</td>
<td>89.5</td>
</tr>
<tr>
<td>BVAR 0-25</td>
<td>97.7</td>
<td>93.8</td>
<td>32.7</td>
<td>78.7</td>
</tr>
<tr>
<td>BIQR 0-25</td>
<td>95.5</td>
<td>96.3</td>
<td>57.1</td>
<td>83.1</td>
</tr>
<tr>
<td>JAE 0-50</td>
<td>74.9</td>
<td>85.6</td>
<td>97.1</td>
<td>95.3</td>
</tr>
<tr>
<td>BVAR 0-50</td>
<td>95.7</td>
<td>88.5</td>
<td>59.3</td>
<td>81.7</td>
</tr>
<tr>
<td>BIQR 0-50</td>
<td>87.6</td>
<td>90.8</td>
<td>85.6</td>
<td>89.7</td>
</tr>
<tr>
<td>JAE 10-40</td>
<td>74.3</td>
<td>85.5</td>
<td>97.6</td>
<td>95.9</td>
</tr>
<tr>
<td>BVAR 10-40</td>
<td>91.0</td>
<td>92.0</td>
<td>81.7</td>
<td>91.0</td>
</tr>
<tr>
<td>BIQR 10-40</td>
<td>84.0</td>
<td>92.2</td>
<td>91.8</td>
<td>94.5</td>
</tr>
<tr>
<td>T0</td>
<td>100.0</td>
<td>12.9</td>
<td>0.0</td>
<td>69.9</td>
</tr>
<tr>
<td>T10</td>
<td>94.9</td>
<td>98.8</td>
<td>18.7</td>
<td>86.4</td>
</tr>
<tr>
<td>T20</td>
<td>88.3</td>
<td>98.7</td>
<td>74.9</td>
<td>95.5</td>
</tr>
<tr>
<td>T30</td>
<td>80.8</td>
<td>92.8</td>
<td>96.1</td>
<td>99.0</td>
</tr>
<tr>
<td>T40</td>
<td>72.2</td>
<td>84.0</td>
<td>98.4</td>
<td>95.7</td>
</tr>
</tbody>
</table>
Table 4.2
Coverage and length of confidence intervals based on a sample of size 10

Normal Distribution

<table>
<thead>
<tr>
<th>Type of Interval</th>
<th>Coverage</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5% One-sided</td>
<td>95% One-sided</td>
</tr>
<tr>
<td>Stand JAE 0-25</td>
<td>11.6</td>
<td>88.8</td>
</tr>
<tr>
<td>Boot JAE 0-25</td>
<td>8.4</td>
<td>91.2</td>
</tr>
<tr>
<td>Boot BVAR 0-25</td>
<td>8.0</td>
<td>91.0</td>
</tr>
<tr>
<td>Boot-t BVAR 0-25</td>
<td>5.2</td>
<td>94.4</td>
</tr>
</tbody>
</table>

Slash Distribution

<table>
<thead>
<tr>
<th>Type of Interval</th>
<th>Coverage</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5% One-sided</td>
<td>95% One-sided</td>
</tr>
<tr>
<td>Stand JAE 0-25</td>
<td>12.6</td>
<td>90.0</td>
</tr>
<tr>
<td>Boot JAE 0-25</td>
<td>4.2</td>
<td>96.8</td>
</tr>
<tr>
<td>Boot BVAR 0-25</td>
<td>5.4</td>
<td>96.6</td>
</tr>
<tr>
<td>Boot-t BVAR 0-25</td>
<td>7.0</td>
<td>93.4</td>
</tr>
</tbody>
</table>

Double Exponential Distribution

<table>
<thead>
<tr>
<th>Type of Interval</th>
<th>Coverage</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5% One-sided</td>
<td>95% One-sided</td>
</tr>
<tr>
<td>Stand JAE 0-25</td>
<td>12.2</td>
<td>90.8</td>
</tr>
<tr>
<td>Boot JAE 0-25</td>
<td>6.6</td>
<td>93.8</td>
</tr>
<tr>
<td>Boot BVAR 0-25</td>
<td>9.2</td>
<td>92.6</td>
</tr>
<tr>
<td>Boot-t BVAR 0-25</td>
<td>8.4</td>
<td>93.6</td>
</tr>
</tbody>
</table>
Figure 4.1: Efficiencies of the best adaptive trimmed means

Normal

One-wild

Slash

Double Exponential

1: Jae 0-25, 2: Bvar 0-25, 3: Biqr 0-25, 4: Bvar 0-50
5: Biqr 0-50, 6: Bvar 10-40, 7: Biqr 10-40