Ann. Inst. Statist. Math.
Vol. 46, No. 4, 737–755 (1994)

# ADAPTIVE CHOICE OF TRIMMING PROPORTIONS

JANA JUREČKOVÁ[1], ROGER KOENKER[2] AND A. H. WELSH[3]

[1] *Department of Probability and Statistics, Charles University, Sokolovska 83, 18600 Prague, Czech Republic*
[2] *Department of Economics, University of Illinois, Champaign, IL 61820, U.S.A.*
[3] *Department of Statistics, The Australian National University, GPO Box 4, Canberra, ACT 2601, Australia*

(Received December 20, 1993; revised May 30, 1994)

**Abstract.** We consider Jaeckel's (1971, *Ann. Math. Statist.*, **42**, 1540–1552) proposal for choosing the trimming proportion of the trimmed mean in the more general context of choosing a trimming proportion for a trimmed $L$-estimator of location. We obtain higher order expansions which enable us to evaluate the effect of the estimated trimming proportion on the adaptive estimator. We find that $L$-estimators with smooth weight functions are to be preferred to those with discontinuous weight functions (such as the trimmed mean) because the effect of the estimated trimming proportion on the estimator is of order $n^{-1}$ rather than $n^{-3/4}$. In particular, we find that valid inferences can be based on a particular "smooth" trimmed mean with its asymptotic standard error and the Student $t$ distribution with degrees of freedom given by the Tukey and McLaughlin (1963, *Sankhyā Ser. A*, **25**, 331–352) proposal.

*Key words and phrases*: Trimmed mean, adaptive estimation, $L$-statistics.

## 1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a sample of independent and identically distributed random variables with common distribution function $F$ and let $X_{n1} \le X_{n2} \le \cdots \le X_{nn}$ denote the order statistics of this sample. The $\alpha$-trimmed mean is the mean of the observations which remain after excluding $[n\alpha]$ observations in each tail and is given by

$$M_n(\alpha) = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{ni}, \qquad 0 \le \alpha < 1/2.$$

For accounts of the history and properties of this popular estimator of location, see Tukey and McLaughlin (1963), Bickel (1965), Huber (1972), Stigler (1973) and Bickel and Lehmann (1975). Provided $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$ are unique,

$$n^{1/2}\{M_n(\alpha) - M(\alpha)\} \xrightarrow{D} N(0, V(\alpha)), \tag{1.1}$$

where

$$M(\alpha) = \frac{1}{1-2\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} y dF(y) \quad \text{and}$$

$$V(\alpha) = \frac{1}{(1-2\alpha)^2}\left\{\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} [y-M(\alpha)]^2 dF(y) + \alpha[F^{-1}(\alpha)-M(\alpha)]^2 + \alpha[F^{-1}(1-\alpha)-M(\alpha)]^2\right\}.$$

The asymptotic variance $V(\alpha)$ is readily estimated by the Winsorised variance

$$V_n(\alpha) = \frac{1}{(1-2\alpha)^2}\left\{\frac{1}{n}\sum_{i=[n\alpha]+1}^{n-[n\alpha]} [X_{ni} - M_n(\alpha)]^2 + \alpha[X_{n,[n\alpha]+1} - M_n(\alpha)]^2 + \alpha[X_{n,n-[n\alpha]} - M_n(\alpha)]^2\right\}.$$

Tukey and McLaughlin (1963) suggested that inference for the trimmed mean be based on treating $n^{1/2}\{M_n(\alpha) - M(\alpha)\}/V_n(\alpha)^{1/2}$ as having Student's $t$ distribution with $n - 2[n\alpha] - 1$ degrees of freedom. In practice, in order to use the $\alpha$-trimmed mean, we need to specify the proportion $\alpha$ of extreme observations to be trimmed in each tail. A simple approach is to use a deterministic rule such as $\alpha = 0.1$. While this rule seems to work well on real data (see Stigler (1977), Spjotvoll and Aastreit (1980), Hill and Dixon (1982) and Rocke *et al.* (1982)), it is clearly rather arbitrary and can be improved on in particular problems. Jaeckel (1971) proposed that we use the value $\hat{\alpha}$ which minimises the estimated asymptotic variance $V_n(\alpha)$ of the trimmed mean over some fixed interval $[\alpha_1, 1-\alpha_2]$ where $0 < \alpha_1 < \alpha_2 < 1/2$. He showed that this adaptive procedure is asymptotically as good as using the value $\alpha_0$ of $\alpha$ which minimises the asymptotic variance $V(\alpha)$ on $[\alpha_1, \alpha_2]$ in the sense that $M_n(\hat{\alpha}) - M_n(\alpha_0) = o_p(n^{-1/2})$. Under further conditions on $F$, Hall (1981) showed that

$$|\hat{\alpha} - \alpha_0| = O_p(n^{-1/4}) \quad \text{and} \quad |M_n(\hat{\alpha}) - M_n(\alpha_0)| = O_p(n^{-3/4}).$$

Inference for adaptive procedures, including the adaptive trimmed mean, tends to be problematic in small and moderate samples. One reason for this is that in using $n^{-1/2}V_n(\hat{\alpha})^{1/2}$ as the standard error of $M_n(\hat{\alpha})$ (as is suggested by (1.1)), we ignore the variability in $\hat{\alpha}$. Since there seems to be no reasonable basis for trying to condition on $\hat{\alpha}$, it seems desirable to investigate two other possibilities: either explicitly including the variability in $\hat{\alpha}$ in the standard error of $M_n(\hat{\alpha})$ or attempting to decrease the effect of the variability in $\hat{\alpha}$ in the standard error of $M_n(\hat{\alpha})$ sufficiently so that the variability in $\hat{\alpha}$ could be ignored in standard error calculations.

We investigate the effect of the variability in $\hat{\alpha}$ on $M_n(\hat{\alpha})$ by obtaining an expansion of the form

$$M_n(\hat{\alpha}) - M(\alpha_0) = M_n(\alpha_0) - M(\alpha_0) + (\hat{\alpha} - \alpha_0)a(F_n - F, F, \alpha_0) + \text{smaller order terms}, \tag{1.2}$$

for a specified function $a(F_n - F, F, \alpha_0)$. The second term is typically of order $n^{-1/2}O_p(|\hat{\alpha} - \alpha_0|)$ so the rate of convergence of $\hat{\alpha}$ to $\alpha_0$ determines the impact of the variability in $\hat{\alpha}$ on $M_n(\hat{\alpha})$. This observation and Hall's (1981) results raise a number of questions. Why is $\hat{\alpha}$ such a poor estimator of $\alpha_0$? Are there alternative data-driven ways of choosing an $\alpha$ which result in better rates of convergence than Jaeckel's proposal? What is the loss of asymptotic efficiency of such methods? How should the standard error be calculated? Finally, are there other trimmed location estimators for which Jaeckel's prescription works better? The purpose of this paper is to answer these basic questions.

The rate of convergence of a random variable $\hat{\alpha}_S$, which minimises a random function $S_n(\alpha)$, to $\alpha_S$, which minimises $S(\alpha)$, depends on the smoothness of $S_n(\alpha)$. Hall's result is a direct consequence of the fact that neither $V_n(\alpha)$ nor its influence function are differentiable. See Hampel *et al.* (1986) for a detailed treatment of influence functions. This suggests that rather than minimising $V_n(\alpha)$ to estimate $\alpha$, we should minimise $S_n(\alpha)$ where $S_n(\alpha)$ is a smooth function. The results of Section 3 show that

$$\begin{aligned} M_n(\hat{\alpha}_S) - M_n(\alpha_S) = M(\hat{\alpha}_S) - M(\alpha_S) &+ O_p(n^{-1/2})|\hat{\alpha}_S - \alpha_S| \\ &+ o_p(n^{-1/2}(\hat{\alpha}_S - \alpha_S)) + O_p(n^{-1}). \end{aligned}$$

This representation clarifies several issues. Firstly, if $F$ is symmetric $M(\hat{\alpha}_S) - M(\alpha_S) = 0$. Otherwise $M(\hat{\alpha}_S) - M(\alpha_S) = O_p(|\hat{\alpha} - \alpha_0|)$ which is either of order $n^{-1/4}$ or of order $n^{-1/2}$. In the former case the distribution of $M_n(\hat{\alpha}_S) - M(\alpha_S)$ is determined by that of $(\hat{\alpha}_S - \alpha_S)$ while in the latter case it is determined by that of $M_n(\hat{\alpha}_S) - M_n(\alpha_S)$ and $\hat{\alpha}_S - \alpha_S$. Adaption as it is usually understood is not possible in either case unless we make the somewhat unsatisfactory claim that we are estimating the random variable $M(\hat{\alpha}_S)$ rather than the functional $M(\alpha_S)$. (Incidentally, while in general we can define the trimmed mean so that we trim $[n\alpha]$ observations in the lower tail and $n - [n\beta]$ observations in the upper tail, we have $M(\hat{\alpha}_S) - M(\alpha_S) = 0$ if and only if $F$ is symmetric and $\beta = 1 - \alpha$ *so we consider symmetric trimming only*.) In the symmetric case, we have

$$M_n(\hat{\alpha}_S) - M_n(\alpha_S) = O_p(n^{-1/2})|\hat{\alpha}_S - \alpha_S| + o_p(n^{-1/2}(\hat{\alpha}_S - \alpha_S)) + O_p(n^{-1})$$

which contains Hall's result and confirms that the rate of convergence depends on that of $\hat{\alpha}_S - \alpha_S$ and may be improved by choosing $\hat{\alpha}_S$ differently.

The loss of efficiency due to minimising some $S_n(\alpha)$ rather than $V_n(\alpha)$ can be evaluated by comparing $V(\alpha_0)$ to $V(\alpha_S)$; $V(\alpha)$ is typically nearly constant over most of its domain (see, for example, Fig. 1) so, if $S_n(\alpha)$ is close to $V_n(\alpha)$, the loss of asymptotic efficiency is generally small.

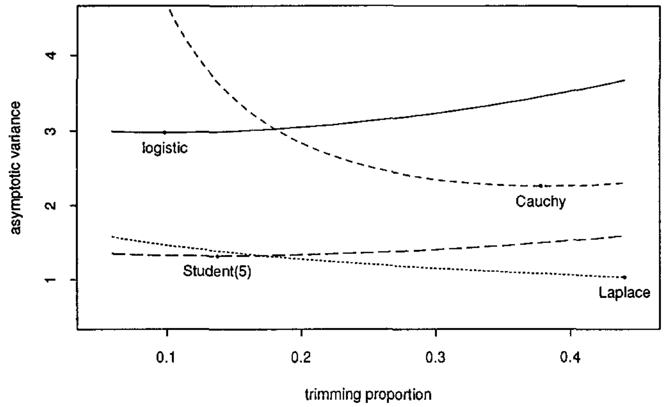


Fig. 1. Asymptotic variance of STM from several distributions.

There are two difficulties in trying to use (1.2) to obtain an expansion for the asymptotic variance of the adaptive trimmed mean $M_n(\hat{\alpha})$. Firstly, the second term in the variance expansion requires (at least) the next term in (1.2) but this seems difficult to obtain without smoothness conditions which the trimmed mean fails to satisfy. Secondly, the non-differentiability of $V_n(\alpha)$ makes it difficult to obtain a suitable expansion for $\hat{\alpha} - \alpha_0$. Moreover, even if we could obtain the requisite terms, we find that the function $a(F_n - F, F, \alpha_0)$ depends on the sparsity function $1/f(F^{-1}(q))$, so calculating the standard error would involve estimating the sparsity function. While a number of estimators of this quantity are available (see e.g., Siddiqui (1960) and Welsh (1988)), it is an estimation problem that we prefer to avoid whenever possible since convergence is inherently slower than $n^{-1/2}$. We concentrate on trying to find a simple $n^{-1/2}$ consistent method of estimating the trimming proportion which might justify neglecting the variability in $\hat{\alpha}$ in making inferences with respect to the adaptive trimmed mean.

The results of Sections 3 and 4 together show that if $T_n(\alpha)$ is a trimmed $L$-estimator with a smooth weight function such that the empirical estimator of its asymptotic variance is smooth, then Jaeckel's proposal will lead to an asymptotically efficient estimator with a faster rate of convergence. Of course, the asymptotic efficiency is over a slightly different family of estimators, but if $T_n(\alpha)$ is suitably close to $M_n(\alpha)$, the asymptotic efficiencies of the estimators may be similar; see Table 1. The contribution of the variability in $\hat{\alpha}_S$ to the asymptotic variance of $T_n(\alpha)$ is of order $n^{-2}$ and can be ignored. These are arguments in addition to those given by Stigler (1973) for using $L$-estimators with smooth weight functions in general, and smoothly trimmed means in particular. This paper can be viewed as considering two approaches to improving the rate of convergence of estimators with estimated trimming proportions.

1) Use the trimmed mean but choose $\hat{\alpha}$ in some other (asymptotically less efficient) way so that we achieve the $n^{-1}$ rate, or

2) Use an alternative estimator (such as a smoothly trimmed mean) for which (asymptotically efficient) adaption at the $n^{-1}$ rate is possible.

In either case, we believe that it is unnecessary to incorporate the variability in the trimming proportion into the standard error of the adaptive estimator. The simulation study in Section 5 shows that in terms of inference the second approach is to be preferred because the estimated asymptotic variance of the smoothly trimmed mean better reflects the variability of the adaptive estimator.

## 2. Notation and conditions

Let

$$H_\alpha(u) = \int_0^u h_\alpha(t)dt, \qquad 0 < u < 1, \quad 0 \leq \alpha < 1/2,$$

be a fixed, bounded, signed measure on $(0, 1)$ with a smooth weight function $h_\alpha$ such that $H_\alpha(1) = 1$. Then for any distribution function $G$ define an $L$-functional $T(\alpha) = \int_{-\infty}^{\infty} y dH_\alpha(G(y))$, where $G^{-1}(t) = \inf\{s : G(s) \geq t\}$. Let

$$F_n(y) = n^{-1} \sum_{j=1}^{n} I(X_j \leq y), \qquad y \in \boldsymbol{R},$$

denote the empirical distribution function of $X_1, \ldots, X_n$, and then let

$$T_n(\alpha) = \int_{-\infty}^{\infty} y dH_\alpha(F_n(y)) \approx n^{-1} \sum_{i=1}^{n} h_\alpha(i/n) X_{ni}.$$

This class of $L$-estimators includes the trimmed mean $M_n(\alpha)$ for which $h_\alpha(u) = I(\alpha \leq u \leq 1 - \alpha)/(1 - 2\alpha)$ but also allows us to consider estimators with smooth weight functions such as Stigler's (1973) smoothly trimmed mean for which

$$\begin{aligned} h_\alpha(u) = &\frac{4(u - \alpha/2)}{\alpha(2 - 3\alpha)} I(\alpha/2 \leq u \leq \alpha) + \frac{2}{\alpha(2 - 3\alpha)} I(\alpha \leq u \leq 1 - \alpha) \\ &+ \frac{4(1 - \alpha/2 - u)}{\alpha(2 - 3\alpha)} I(1 - \alpha \leq u \leq 1 - \alpha/2). \end{aligned}$$

In fact, Stigler's piecewise linear weight function is not smooth enough to achieve the best possible rate for Jaeckel's proposal. Thus we also propose to consider the very smooth

$$\begin{aligned} (2.1)\quad h_\alpha(u) = \frac{1}{(1 - 2\alpha)} [&K((u - \alpha)/\sigma) I(\alpha - \sigma \leq u \leq \alpha + \sigma) \\ &+ I(\alpha + \sigma \leq u \leq 1 - \alpha - \sigma) \\ &+ K((1 - \alpha - u)/\sigma) I(1 - \alpha - \sigma \leq u \leq 1 - \alpha + \sigma)], \end{aligned}$$

where $K(z) = \frac{15}{16}(\frac{1}{5}z^5 - \frac{2}{3}z^3 + z + \frac{8}{15}) I(-1 \leq z \leq 1) + I(z > 1)$, which enjoys the property of Tukey's biweight of having second order contact at the terminals. We will refer to the $L$-estimator with weight function defined by (2.1) as the smoothly trimmed mean.

Let $S(\alpha)$ be any criterion function which we minimise over some set to choose an $\hat{\alpha}$. It is convenient and simple to choose $[\alpha_1, \alpha_2]$ for $0 < \alpha_1 < \alpha_2 < 1/2$. There is some interest in allowing $\alpha_1 = 0$ but this can only be done at the expense of moment conditions (which are undesirable from a robustness viewpoint) and a considerable increase in complexity so we will not pursue this in the present paper. If the minimum is not unique let $\alpha_S$ denote the smallest minimising value of $\alpha$ so that

$$\alpha_S = \inf\left\{u \in [\alpha_1, \alpha_2] : S(u) = \inf_{\alpha_1 \le \alpha \le \alpha_2} S(\alpha)\right\}.$$

Let $S_n(\alpha)$ be an estimator of $S(\alpha)$ and let $\hat{\alpha}_S$ denote the smallest minimising value of $S_n(\alpha)$ so that

$$\hat{\alpha}_S = \inf\left\{u \in [\alpha_1, \alpha_2] : S_n(u) = \inf_{\alpha_1 \le \alpha \le \alpha_2} S_n(\alpha)\right\}.$$

We will require that $\alpha_1 < \alpha_S < \alpha_2$ and that $S(\alpha)$ has two derivatives in a neighbourhood of $\alpha_S$ (which we denote $\boldsymbol{N}(\alpha_S)$) with $S''$ continuous and non-zero at $\alpha_S$.

Our results are based on an asymptotic representation for $T_n(\alpha)$ which holds uniformly in $\alpha$ over $[\alpha_1, \alpha_2]$. For the trimmed mean and estimators such as Stigler's smoothly trimmed mean, we will impose the following conditions on $F$ and $h_\alpha$:

M1) $h_\alpha(u) = 0$ for $u < \alpha$ or $u > 1-\alpha$, $0 < \alpha < 1/2$, $h_\alpha(u)$ is bounded for $0 \le u \le 1$ and $\alpha_1 \le \alpha \le \alpha_2$, has at most a finite number of jump discontinuities at $\alpha = s_0 < s_1(\alpha) < \cdots < s_c(\alpha) < 1 - \alpha = s_{c+1}$, where $s_j(\alpha)$ has a bounded derivative for $\alpha_1 \le \alpha \le \alpha_2$, $j = 1, 2, \ldots, c$. Moreover, $h_\alpha(u)$ satisfies the Lipschitz condition

$$\sup_{\alpha_1 \le \alpha \le \alpha_2} |h_\alpha(t+u) - h_\alpha(t)| \le K|u| \text{ for all } t+u \text{ and } t \text{ in } (s_j(\alpha), s_{j+1}(\alpha)),$$

and $\frac{\partial h_\alpha(u)}{\partial \alpha}$ exists and is continuous for $s_j(\alpha) < u < s_{j+1}(\alpha)$, $\alpha \in \boldsymbol{N}(\alpha_S)$ and $j = 0, 1, \ldots, c$ and

M2) $F$ is continuous and has a positive derivative $f$ in a neighbourhood of each jump point of $h_\alpha(u)$ such that $f$ is continuous at each jump point of $h_\alpha(u)$. Note that if $h_\alpha$ has no jump points, then a central limit theorem holds for arbitrary $F$. Indeed, this was the basis for Stigler's (1973) proposal. If $h_\alpha$ is smooth enough, we can obtain more information about the higher order terms in the expansion. We will do this for $h_\alpha$ satisfying the following conditions:

T1) $h_\alpha(u) = 0$ for $u < \alpha$ or $u > 1-\alpha$, $0 < \alpha < 1/2$, $h_\alpha(u)$ is bounded for $0 \le u \le 1$ and $\alpha_1 \le \alpha \le \alpha_2$, $h'_{\alpha_S}(u)$ and $\frac{\partial h'_\alpha(u)}{\partial \alpha}|_{\alpha=\alpha_S}$ are bounded for $0 \le u \le 1$ and continuous a.e. with respect to Lebesgue measure and $F^{-1}$.

The condition T1 does not hold for either the trimmed mean $M_n(\alpha)$ or for Stigler's (1973) smoothly trimmed mean but it does hold for (2.1) because

$$\begin{aligned} h'_\alpha(u) = \frac{1}{1-2\alpha}[&\sigma^{-1}K'((u-\alpha)/\sigma)I(\alpha-\sigma \le u \le \alpha+\sigma) \\ &- \sigma^{-1}K'((1-\alpha-u)/\sigma)I(1-\alpha-\sigma \le u \le 1-\alpha+\sigma)], \end{aligned}$$

where $K'(z) = \frac{15}{16}(z^4 - 2z^2 + 1)I(-1 \le z \le 1)$, and

$$\begin{aligned}\frac{\partial h'_\alpha(u)}{\partial \alpha} = \frac{1}{1-2\alpha}[&\sigma^{-2}K''((1-\alpha-u)/\sigma)I(1-\alpha-\sigma \le u \le 1-\alpha+\sigma) \\ &- \sigma^{-2}K''((u-\alpha)/\sigma)I(\alpha-\sigma \le u \le \alpha+\sigma)] \\ + \frac{2}{(1-2\alpha)^2}[&\sigma^{-1}K'((u-\alpha)/\sigma)I(\alpha-\sigma \le u \le \alpha+\sigma) \\ &- \sigma^{-1}K'((1-\alpha-u)/\sigma)I(1-\alpha-\sigma \le u \le 1-\alpha+\sigma)],\end{aligned}$$

where $K''(z) = \frac{15}{4}(z^3 - z)I(-1 \le z \le 1)$.

It follows from Theorem 3.2 that provided $F$ is symmetric, for any $\hat{\alpha}_S - \alpha_S = o_p(1)$ such that $\alpha_1 < \alpha_S < \alpha_2$,

$$n^{1/2}(T_n(\hat{\alpha}_S) - T(\alpha_S)) \xrightarrow{D} N(0, \sigma^2(\alpha_S, F)),$$

where $\sigma^2(\alpha, F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F(y \wedge z) - F(y)F(z)\}h_\alpha(F(y))h_\alpha(F(z))dydz$. This suggests that whether the family of estimators of interest is the trimmed mean $M_n(\alpha)$ or a more general $L$-estimator $T_n(\alpha)$, we can choose a trimming proportion $\alpha$ by minimising the empirical estimator

$$\sigma^2(\alpha, F_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F_n(y \wedge z) - F_n(y)F_n(z)\}h_\alpha(F_n(y))h_\alpha(F_n(z))dydz$$

of $\sigma^2(\alpha, F)$. That is, we can take $S(\alpha) = \sigma^2(\alpha, F)$. To guarantee that the minimum of $\sigma_n^2(\alpha)$ is a $n^{1/2}$-consistent estimator of the minimum of $\sigma^2(\alpha, F)$, we require smoothness conditions on the weight function $h_\alpha$. In particular, we require T1 given above and

T2) $h_\alpha$ has two derivatives with respect to $\alpha$ in a neighbourhood of $\alpha_S \in (\alpha_1, \alpha_2)$, $\frac{\partial^2 h_\alpha(u)}{\partial \alpha^2}$ is continuous at $\alpha_S$ for each $0 \le u \le 1$, and for $K < \infty$,

$$\sup_{0\le u\le 1}\sup_{\alpha\in N(\alpha_S)}\left|\frac{\partial h_\alpha(u)}{\partial \alpha}\right| \le K \quad \text{and} \quad \sup_{0\le u\le 1}\sup_{\alpha\in N(\alpha_S)}\left|\frac{\partial^2 h_\alpha(u)}{\partial \alpha^2}\right| \le K$$

and $\frac{\partial^2 \sigma^2(\alpha, F)}{\partial \alpha^2}$ is non-zero at $\alpha_S$.

Note that conditions T1 and T2 imply that $h_\alpha(\alpha) = h_\alpha(1-\alpha) = h'_\alpha(\alpha) = h'_\alpha(1-\alpha) = 0$ so from Liebnitz' formula we have that

$$\begin{aligned}\frac{\partial \sigma^2(\alpha, F)}{\partial \alpha} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}&\{F(y \wedge z) - F(y)F(z)\} \\ &\cdot\left\{\frac{\partial h_\alpha(F(y))}{\partial \alpha}h_\alpha(F(z)) + h_\alpha(F(y))\frac{\partial h_\alpha(F(z))}{\partial \alpha}\right\}dydz\end{aligned}$$

and

$$\frac{\partial^2\sigma^2(\alpha,F)}{\partial\alpha^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F(y\wedge z)-F(y)F(z)\} \cdot\left\{\frac{\partial^2 h_\alpha(F(y))}{\partial\alpha^2}h_\alpha(F(z)) + 2\frac{\partial h_\alpha(F(y))}{\partial\alpha}\frac{\partial h_\alpha(F(z))}{\partial\alpha} + h_\alpha(F(y))\frac{\partial^2 h_\alpha(F(z))}{\partial\alpha^2}\right\}dydz.$$

Hence T2 ensures that $\frac{\partial^2\sigma^2(\alpha,F)}{\partial\alpha^2}$ is continuous at $\alpha_S$.

## 3. Representation for $L$-estimators

The following theorem provides a representation for $L$-estimators (including the trimmed mean) with estimated trimming proportions. The proof is based on an argument used in Jurečková (1986).

THEOREM 3.1. *Suppose that conditions* M1–M2 *hold*, $\alpha_1 < \alpha_S < \alpha_2$ *and* $\hat{\alpha}_S - \alpha_S = o_p(1)$. *Then*

$$T_n(\hat{\alpha}_S) - T_n(\alpha_S) = T(\hat{\alpha}_S) - T(\alpha_S) + (\hat{\alpha}_S-\alpha_S)a(F_n - F, f, \alpha_S) + o_p(n^{-1/2}(\hat{\alpha}_S - \alpha_S)) + O_p(n^{-1}),$$

*where*

$$a(F_n - F, f, \alpha) = \sum_{j=0}^{c}\left\{\int_{F^{-1}(s_j(\alpha))}^{F^{-1}(s_{j+1}(\alpha))}\{F_n(y)-F(y)\}\frac{\partial h_\alpha(F(y))}{\partial\alpha}dy + \frac{h_\alpha(s_{j+1}(\alpha))}{f(F^{-1}(s_{j+1}(\alpha)))}\frac{\partial s_{j+1}(\alpha)}{\partial\alpha} \cdot\{F_n(F^{-1}(s_{j+1}(\alpha))) - s_{j+1}(\alpha)\} - \frac{h_\alpha(s_j(\alpha))}{f(F^{-1}(s_j(\alpha)))}\frac{\partial s_j(\alpha)}{\partial\alpha}\{F_n(F^{-1}(s_j(\alpha))) - s_j(\alpha)\}\right\}.$$

PROOF. We will first prove that

$$\sup_{\alpha_1\le\alpha\le\alpha_2}\left|T_n(\alpha) - T(\alpha) + \int_{-\infty}^{\infty}\{F_n(y)-F(y)\}h_\alpha(F(y))dy\right| = O_p(n^{-1}). \tag{3.1}$$

Write

$$T_n(\alpha) - T(\alpha) + \int_{-\infty}^{\infty}\{F_n(y)-F(y)\}h_\alpha(F(y))dy = -\int_{-\infty}^{\infty} W_{F_n,F}(y)\{F_n(y)-F(y)\}dy,$$

with

$$W_{G,F}(y) = \begin{cases}\dfrac{\{H_\alpha(G(y)) - H_\alpha(F(y))\}}{\{G(y)-F(y)\}} - h_\alpha(F(y)) & \text{if } G(y)\ne F(y)\\ 0 & \text{otherwise.}\end{cases}$$

There exists $-\infty < a < b < \infty$ such that for $\sup_{y\in R}|G(y)-F(y)| < \alpha_1$, we have $W_{G,F}(y) = h_\alpha(G(y)) = h_\alpha(F(y)) = 0$ for $y < a$ or $y > b$ and any $\alpha_1 \le \alpha \le \alpha_2$. That is, for $n$ large enough, the range of integration can be restricted to $[a,b]$. Without loss of generality, suppose that $h_\alpha$ has a single jump discontinuity at $\alpha$. Since $a < F^{-1}(\alpha) < b$, for $n$ sufficiently large,

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} W_{F_n,F}(y)\{F_n(y)-F(y)\}dy\right| \\
&\quad\le \int_a^b |W_{F_n,F}(y)\{F_n(y)-F(y)\}|dy \\
&\quad= \left\{\int_a^{F^{-1}(\alpha)-n^{-1/2}} + \int_{F^{-1}(\alpha)-n^{-1/2}}^{F^{-1}(\alpha)+n^{-1/2}} + \int_{F^{-1}(\alpha)+n^{-1/2}}^{b}\right\} \\
&\qquad\cdot |W_{F_n,F}(y)\{F_n(y)-F(y)\}|dy.
\end{aligned}$$

Now with $K$ a generic positive constant,

$$\begin{aligned}
&\sup_{\alpha_1\le\alpha\le\alpha_2}\int_a^{F^{-1}(\alpha)-n^{-1/2}} |W_{F_n,F}(y)\{F_n(y)-F(y)\}|dy \\
&\le \sup_{\alpha_1\le\alpha\le\alpha_2}\int_a^{F^{-1}(\alpha)-n^{-1/2}} \Bigg[I(F_n(y)>F(y)) \\
&\quad\cdot\int_0^{F_n(y)-F(y)} |h_\alpha(F(y)+u)-h_\alpha(F(y))|du \\
&\quad+ I(F_n(y)<F(y))\int_{F_n(y)-F(y)}^{0} |h_\alpha(F(y)+u)-h_\alpha(F(y))|du\Bigg]dy \\
&\le \sup_{\alpha_1\le\alpha\le\alpha_2} K\int_a^{F^{-1}(\alpha)-n^{-1/2}} \Bigg[I(F_n(y)>F(y))\int_0^{F_n(y)-F(y)} udu \\
&\quad+ I(F_n(y)<F(y))\int_{F_n(y)-F(y)}^{0}(-u)du\Bigg]dy \\
&\le K\sup_y|F_n(y)-F(y)|^2 \sup_{\alpha_1\le\alpha\le\alpha_2}|F^{-1}(\alpha)-n^{-1/2}-\alpha| \\
&= O_p(n^{-1}).
\end{aligned}$$

Similarly,

$$\sup_{\alpha_1\le\alpha\le\alpha_2}\int_{F^{-1}(\alpha)+n^{-1/2}}^{b} |W_{F_n,F}(y)\{F_n(y)-F(y)\}|dy = O_p(n^{-1}).$$

Also, $W_{G,F(\cdot)}$ is bounded so

$$\begin{aligned}
&\sup_{\alpha_1\le\alpha\le\alpha_2}\int_{F^{-1}(\alpha)-n^{-1/2}}^{F^{-1}(\alpha)+n^{-1/2}} |W_{F_n,F}(y)\{F_n(y)-F(y)\}|dy \\
&\qquad\le Kn^{-1/2}\sup_y|F_n(y)-F(y)| = O_p(n^{-1}),
\end{aligned}$$

and hence (3.1) holds.

Now write

$$\begin{aligned}
&T_n(\hat\alpha_S) - T_n(\alpha_S) - T(\hat\alpha_S) + T(\alpha_S)\\
&\quad= -\int_{-\infty}^{\infty} \{F_n(y) - F(y)\}\{h_{\hat\alpha_S}(F(y)) - h_{\alpha_S}(F(y))\}dy + O_p(n^{-1})\\
&\quad= \sum_{j=0}^{c}\left\{\int_{F^{-1}(s_j(\hat\alpha_S))}^{F^{-1}(s_{j+1}(\hat\alpha_S))} \{F_n(y) - F(y)\}h_{\hat\alpha_S}(F(y))dy\right.\\
&\qquad\left. - \int_{F^{-1}(s_j(\alpha_S))}^{F^{-1}(s_{j+1}(\alpha_S))} \{F_n(y) - F(y)\}h_{\alpha_S}(F(y))dy\right\} + O_p(n^{-1})\\
&\quad= (\hat\alpha_S - \alpha_S)a(F_n - F, f, \tilde\alpha) + O_p(n^{-1})\\
&\quad= (\hat\alpha_S - \alpha_S)a(F_n - F, f, \alpha_S)\\
&\qquad+ (\hat\alpha_S - \alpha_S)\{a(F_n - F, f, \tilde\alpha) - a(F_n - F, f, \alpha_S)\} + O_p(n^{-1}),
\end{aligned}$$

where $|\alpha_S - \tilde\alpha| \le |\alpha_S - \hat\alpha_S|$, and the result obtains. □

Under additional smoothness conditions on $h_\alpha$, we can obtain more precise information about the higher order terms in the expansion.

THEOREM 3.2. *Suppose that condition* T1 *holds*, $\alpha_1 < \alpha_S < \alpha_2$ *and* $\hat\alpha_S - \alpha_S = o_p(1)$. *Then*

$$\begin{aligned}
T_n(\hat\alpha_S) - T_n(\alpha_S) = T(\hat\alpha_S) - T(\alpha_S)&\\
- (\hat\alpha_S - \alpha_S)\int_{-\infty}^{\infty} \{F_n(y) - F(y)\}\frac{\partial h_{\alpha_S}(F(y))}{\partial \alpha_S}dy&\\
+ O_p\{n^{-1/2}(\hat\alpha_S - \alpha_S)^2\} + o_p(n^{-1}).&
\end{aligned}$$

PROOF. As in the proof of Theorem 3.1, write

$$\begin{aligned}
T_n(\alpha) - T(\alpha) + \int_{-\infty}^{\infty} \{F_n(y) - F(y)\}h_\alpha(F(y))dy&\\
+ \frac{1}{2}\int_{-\infty}^{\infty} \{F_n(y) - F(y)\}^2 h'_\alpha(F(y))dy&\\
= -\int_{-\infty}^{\infty} W_{F_n,F}(y)\{F_n(y) - F(y)\}^2 dy,&
\end{aligned}$$

where now

$$W_{G,F}(y) = \begin{cases} \dfrac{\{H_\alpha(F(y)) - H_\alpha(G(y))\}}{\{G(y) - F(y)\}^2} & \\ \qquad -\dfrac{h_\alpha(F(y))}{\{G(y) - F(y)\}} - \dfrac{1}{2}h'_\alpha(F(y)) & \text{if } G(y) \ne F(y)\\ 0 & \text{otherwise.}\end{cases}$$

Then, by the dominated convergence theorem,

$$\sup_{\alpha_1 \le \alpha \le \alpha_2} \left| T_n(\alpha) - T(\alpha) + \int_{-\infty}^{\infty} \{F_n(y) - F(y)\} h_\alpha(F(y)) dy \right.$$
$$\left. + \frac{1}{2} \int_{-\infty}^{\infty} \{F_n(y) - F(y)\}^2 h'_\alpha(F(y)) dy \right| = o_p(n^{-1}).$$

Now arguing as in the proof of Theorem 3.1,

$$\begin{aligned} & T_n(\hat{\alpha}_S) - T_n(\alpha_S) - T(\hat{\alpha}_S) - T(\alpha_S) \\ &= -\int_{-\infty}^{\infty} \{F_n(y) - F(y)\}\{h_{\hat{\alpha}_S}(F(y)) - h_{\alpha_S}(F(y))\} dy \\ &\quad - \frac{1}{2} \int_{-\infty}^{\infty} \{F_n(y) - F(y)\}^2 \{h'_{\hat{\alpha}_S}(F(y)) - h'_{\alpha_S}(F(y))\} dy + o_p(n^{-1}) \\ &= -(\hat{\alpha}_S - \alpha_S) \int_{-\infty}^{\infty} \{F_n(y) - F(y)\} \frac{\partial h_{\alpha_S}(F(y))}{\partial \alpha_S} dy \\ &\quad + O_p\{n^{-1/2}(\hat{\alpha}_S - \alpha_S)^2\} + o_p(n^{-1}). \end{aligned}$$

□

Table 1. Asymptotic efficiency of the smooth trimmed mean.

Standard Normal

| | $\alpha = 0.05$ | $\alpha = 0.1$ | $\alpha = 0.2$ | $\alpha = 0.4$ |
|---|---|---|---|---|
| $\sigma = 0$ | 0.973 | 0.949 | 0.880 | 0.727 |
| $\sigma = 0.01$ | 0.974 | 0.943 | 0.874 | 0.722 |
| $\sigma = 0.05$ | 0.976 | 0.945 | 0.876 | 0.726 |

Student on 12 degrees of freedom

| | $\alpha = 0.05$ | $\alpha = 0.1$ | $\alpha = 0.2$ | $\alpha = 0.4$ |
|---|---|---|---|---|
| $\sigma = 0$ | 0.995 | 0.995 | 0.949 | 0.803 |
| $\sigma = 0.01$ | 0.995 | 0.988 | 0.942 | 0.797 |
| $\sigma = 0.05$ | 0.996 | 0.989 | 0.944 | 0.801 |

Student on 3 degrees of freedom

| | $\alpha = 0.05$ | $\alpha = 0.1$ | $\alpha = 0.2$ | $\alpha = 0.4$ |
|---|---|---|---|---|
| $\sigma = 0$ | 0.848 | 0.941 | 0.996 | 0.915 |
| $\sigma = 0.01$ | 0.849 | 0.932 | 0.987 | 0.909 |
| $\sigma = 0.05$ | 0.842 | 0.930 | 0.988 | 0.912 |

Student on 1 degree of freedom

| | $\alpha = 0.05$ | $\alpha = 0.1$ | $\alpha = 0.2$ | $\alpha = 0.4$ |
|---|---|---|---|---|
| $\sigma = 0$ | 0.226 | 0.428 | 0.704 | 0.883 |
| $\sigma = 0.01$ | 0.226 | 0.418 | 0.696 | 0.876 |
| $\sigma = 0.05$ | 0.202 | 0.408 | 0.691 | 0.877 |

As noted in Section 2, it follows from Theorems 3.1 and 3.2 that provided $F$ is symmetric, for any $\hat{\alpha}_S - \alpha_S = o_p(1)$ such that $\alpha_1 < \alpha_S < \alpha_2$, $n^{1/2}(T_n(\hat{\alpha}_S) - T(\alpha_S)) \xrightarrow{D} N(0, \sigma^2(\alpha_S))$, where $\sigma^2(\alpha) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F(y \wedge z) - F(y)F(z)\}h_\alpha(F(y)) \cdot h_\alpha(F(z))dydz$. For $h_\alpha$ given by (2.1), we can use this result to calculate the asymptotic efficiency of the smooth trimmed mean for different values of $(\alpha, \sigma)$. The results of calculations at selected underlying distributions (corresponding to those used in the simulation study in Section 5) are given in Table 1. The efficiency of the smooth trimmed mean (with small $\sigma$) is similar to that of the trimmed mean.

## 4. Choosing the trimming proportions

First consider the general structure described in Section 2 in which $S(\alpha)$ is any criterion function which we can minimise to choose an $\alpha = \alpha_S$. Let $S_n(\alpha)$ be an estimator of $S(\alpha)$ and let $\hat{\alpha}_S$ denote the smallest minimising value of $S_n(\alpha)$. A standard argument (see Lemma 3 of Jaeckel (1971) or Theorem 1 of Hall (1981)) can be used to show that $\hat{\alpha}_S - \alpha_S = o_p(1)$. If $\alpha_1 < \alpha_S < \alpha_2$, $S_n(\alpha)$ has two derivatives in a neighbourhood of $\alpha_S$, $\sup_{\alpha \in N(\alpha_S)} |S_n''(\alpha) - S''(\alpha)| = o_p(1)$ and $S''$ is continuous and non-zero at $\alpha_S$, then for any $\hat{\alpha}_S$ which satisfies $S_n'(\hat{\alpha}_S) = o_p(n^{-1/2})$,

$$\hat{\alpha}_S - \alpha_S = -\frac{S_n'(\alpha_S)}{S''(\alpha_S)} + o_p(\hat{\alpha}_S - \alpha_S). \tag{4.1}$$

When $S$ is smooth, as the next result establishes, $\hat{\alpha}_S - \alpha_S = o_p(n^{-1/2})$. Further, if the functional $S(\alpha)$ is the asymptotic variance of a smooth $L$-estimator (e.g. of the estimator defined in (2.1)), we may still approximate the right-hand side of (4.1) by a functional of an empirical process and thus eventually obtain the (non-normal) limiting distribution of $\hat{\alpha}_S - \alpha_S$. More precisely, in such a case we have

THEOREM 4.1. *Suppose that* T1–T2 *hold. Then with* $S(\alpha) = \sigma^2(\alpha, F)$ *and* $S_n(\alpha) = \sigma^2(\alpha, F_n)$

$$\hat{\alpha}_S - \alpha_S = -\frac{\eta_n(\alpha_S)}{S''(\alpha_S)} + o_p(n^{-1/2}),$$

*where*

$$\begin{aligned}\eta_n(\alpha) = \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} &\{\{F_n(y \wedge z) - F(y \wedge z)\}J_\alpha(y,z)\\ &+ \{F_n(y) - F(y)\}L_\alpha(y,z)\\ &+ \{F_n(z) - F(z)\}L_\alpha(z,y)\}dydz,\end{aligned}$$

*with*

$$J_\alpha(y,z) = \frac{\partial h_\alpha(F(y))}{\partial\alpha}h_\alpha(F(z)) + h_\alpha(F(y))\frac{\partial h_\alpha(F(z))}{\partial\alpha}$$

*and*

$$\begin{aligned}L_\alpha(y,z) = \{&F(y\wedge z) - F(y)F(z)\}\\ &\cdot\left\{\frac{\partial h'_\alpha(F(y))}{\partial\alpha}h_\alpha(F(z)) + h'_\alpha(F(y))\frac{\partial h_\alpha(F(z))}{\partial\alpha}\right\}\\ &- F(z)\left\{\frac{\partial h_\alpha(F(y))}{\partial\alpha}h_\alpha(F(z)) + h_\alpha(F(y))\frac{\partial h_\alpha(F(z))}{\partial\alpha}\right\}.\end{aligned}$$

*Clearly,* $\hat\alpha_S - \alpha_S = O_p(n^{-1/2})$.

PROOF. That $\sup_{\alpha\in N(\alpha_S)}|S''_n(\alpha) - S''(\alpha)| = o_p(1)$ under T1–T2 follows from the dominated convergence theorem. Write $S'_n(\alpha_S) - S'(\alpha_S) - \eta_n(\alpha_S) = \sum_{k=1}^4 Q_k(\alpha_S)$ where

$$\begin{aligned}Q_1(\alpha) = \iint &\{F_n(y\wedge z) - F_n(y)F_n(z)\}\{U_\alpha(F_n(y),F_n(z)) - U_\alpha(F(y),F(z))\\ &\cdot\{F_n(y) - F(y)\}U_\alpha^{(1)}(F(y),F(z))\\ &- \{F_n(z) - F(z)\}U_\alpha^{(2)}(F(y),F(z))\}dydz,\\ Q_2(\alpha) = \iint &\{F_n(y\wedge z) - F_n(y)F_n(z) - F(y\wedge z) + F(y)F(z)\}\\ &\cdot\{F_n(y) - F(y)\}U_\alpha^{(1)}(F(y),F(z))dydz,\\ Q_3(\alpha) = \iint &\{F_n(y\wedge z) - F_n(y)F_n(z) - F(y\wedge z) + F(y)F(z)\}\\ &\cdot\{F_n(z) - F(z)\}U_\alpha^{(2)}(F(y),F(z))dydz\end{aligned}$$

and

$$Q_4(\alpha) = -\iint\{F_n(y) - F(y)\}\{F_n(z) - F(z)\}U_\alpha(F(y),F(z))dydz,$$

where $U_\alpha(u,v) = \frac{\partial h_\alpha(u)}{\partial\alpha}h_\alpha(v) + h_\alpha(u)\frac{\partial h_\alpha(v)}{\partial\alpha}$, $U_\alpha^{(1)}(u,v)$ and $U_\alpha^{(2)}(u,v)$ denote the derivatives of $U_\alpha(u,v)$ with respect to $u$ and $v$ respectively and all the integrals are taken over the range $(F^{-1}(\alpha), F^{-1}(1-\alpha))$. As in the proof of Theorem 3.1, for $n$ large enough, the range of integration can be restricted to $[a,b]$, where $-\infty < a < b < \infty$.

Now write

$$\begin{aligned}U_\alpha(G(y),G(z)) &- U_\alpha(F(y),F(z)) - \{G(y) - F(y)\}U_\alpha^{(1)}(F(y),F(z))\\ &- \{G(z) - F(z)\}U_\alpha^{(2)}(F(y),F(z))\\ &= V_{G,F}(y)h_\alpha(F(z)) + \frac{\partial h_\alpha(F(y))}{\partial\alpha}W_{G,F}(z)\\ &\quad + \{G(y) - F(y)\}\frac{\partial h'_\alpha(F(y))}{\partial\alpha}\{h_\alpha(G(z)) - h_\alpha(F(z))\}\\ &= V_{G,F}(z)h_\alpha(F(y)) + \frac{\partial h_\alpha(F(z))}{\partial\alpha}W_{G,F}(y)\\ &\quad + \{G(z) - F(z)\}\frac{\partial h'_\alpha(F(z))}{\partial\alpha}\{h_\alpha(G(y)) - h_\alpha(F(y))\}\end{aligned}$$

with

$$W_{G,F}(y) = \begin{cases} \{h_\alpha(G(y)) - h_\alpha(F(y))\}/\{G(y) - F(y)\} & \\ \quad -h'_\alpha(F(y)) & \text{if } G(y) \neq F(y) \\ 0 & \text{otherwise,} \end{cases}$$

and

$$V_{G,F}(y) = \begin{cases} \left\{ \dfrac{\partial h_\alpha(G(y))}{\partial \alpha} - \dfrac{\partial h_\alpha(F(y))}{\partial \alpha} \right\} / \{G(y) - F(y)\} & \\ \quad -\dfrac{\partial h'_\alpha(F(y))}{\partial \alpha} & \text{if } G(y) \neq F(y) \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} |Q_1(\alpha)| \leq O_p(n^{-1/2}) \int_a^b \int_a^b \Bigg\{ & |V_{F_n,F}(y) h_\alpha(F(z))| + \left| \frac{\partial h_\alpha(F(y))}{\partial \alpha} W_{F_n,F}(z) \right| \\ & + \left| \frac{\partial h'_\alpha(F(y))}{\partial \alpha} \{h_\alpha(F_n(z)) - h_\alpha(F(z))\} \right| \\ & + |V_{F_n,F}(z) h_\alpha(F(y))| \left| \frac{\partial h_\alpha(F(z))}{\partial \alpha} W_{F_n,F}(y) \right| \\ & + \left| \frac{\partial h'_\alpha(F(z))}{\partial \alpha} \{h_\alpha(F_n(y)) - h_\alpha(F(y))\} \right| \Bigg\} dydz \\ = o_p(n^{-1/2}), \qquad & \end{aligned}$$

by the dominated convergence theorem. Similarly $\sup_{\alpha_1 \leq \alpha \leq \alpha_2} |Q_k| = o_p(n^{-1/2})$, $k = 2, 3$. The fourth term is of order $n^{-1}$ in probability and the result obtains. □

There exists a Brownian bridge $B_n$ dependent on $X_1, \ldots, X_n$ such that

$$\begin{aligned} \eta_n(\alpha) = 2n^{-1/2} & \int_{F^{-1}(\alpha)}^{1-F^{-1}(\alpha)} \int_{F^{-1}(\alpha)}^{1-F^{-1}(\alpha)} I[y < z] \\ & \cdot \{B_n(F(y))(J_\alpha(y,z) + L_\alpha(y,z)) + B_n(F(z)) L_\alpha(z,y)\} dydz \\ & + o_p(n^{-1/2}). \end{aligned}$$

This implies that $\hat{\alpha}_S - \alpha = O_p(n^{-1/2})$. Hence, the order of variability in $\hat{\alpha}_S$ is considerably smaller in the smoothly trimmed case and we have a sufficient reason to ignore it.

## 5. Simulation

In this section we report the results of a small simulation experiment designed to evaluate the finite sample performance of several versions of the adaptive trimmed mean. As we emphasized above, an important aspect of performance will be our ability to make credible inferences and construct valid confidence intervals without specifically including the variability in $\hat{\alpha}_S$.

5.1 *Monte-Carlo swindle*

Since all of the estimators under consideration are translation equivariant and their associated variance estimators are scale equivariant we may exploit the Monte-Carlo swindle of Relles (1970) and Gross (1973) for error distributions from the normal/independent family (see Efron and Olshen (1978) for a nice characterization of this family).

In the location setting we may draw $y_i = z_i/v_i$, $i = 1, \ldots, n$, where $z_i$ are independent standard normal random variables and the $v_i$ are independent draws from another distribution. Here $v_i \sim \sqrt{\chi^2_\nu/\nu}$ so the $y_i$'s have a Student's $t$ distribution on $\nu$ degrees of freedom. Conditional on the $v_i$'s an efficient estimator of location may be constructed as $\hat{\mu} = (\sum_{i-1}^n v_i^2)^{-1} \sum_{i=1}^n v_i^2 y_i$. Then, $\hat{\mu} \sim N(0, \sigma_v^2)$ where $\sigma_v^2 = (\sum_{i-1}^n v_i^2)^{-1}$, and for any other scale equivariant estimator of location $T_n$ and associated estimator of scale $s_n$,

$$P[T_n > ks_n] = P[\hat{\mu} > ks_n - T_n + \hat{\mu}] = 1 - \Phi((ks_n - T_n - \hat{\mu})/\sigma_v)$$

and by symmetry,

$$P[T_n > ks_n] = \Phi((-ks_n + T_n - \hat{\mu})/\sigma_v).$$

Averaging these two probabilities over the number of Monte-Carlo replications, say $R$, of the experiment for several values of $k$ yields estimates $\hat{p}_R(k_i)$ for $i = 1, \ldots, K$. Finally, regressing $\text{logit}(\hat{p}_R(k_i))$ on $k$, we may interpolate to find $k^*$ such that $\hat{p}_R(k^*) = .025$. For the adaptive estimators where trimming is random it is natural to adopt the Tukey and McLaughlin's (1963) suggestion to use $t$ on $n - 2[n\hat{\alpha}_S] - 1$ critical values. Since $\hat{\alpha}_S$ is location and scale invariant, we may incorporate the $t$ critical value into the definition of $s_n$ above and estimate a multiplicative factor, say $\lambda^*$, required to inflate (or deflate) the nominal Tukey-McLaughlin critical values to achieve correct size. These adjustment factors are reported in Table 3 below. Expected confidence interval lengths (ECIL's) may also be estimated. Again following Gross (1977), $ECIL = 2k^* E(s_n/\hat{\sigma}_0)E(\hat{\sigma}_0)$ since $s_n/\hat{\sigma}_0$ and $\hat{\sigma}_0^2 = \sum v_i^2(y_i - \hat{\mu})^2$ are conditionally independent. Averaging $s_n/\hat{\sigma}_0$ over Monte-Carlo replications and noting that $\hat{\sigma}_0^2$ is conditionally $\chi^2_{n-1}$, $E\hat{\sigma}_0 = \sqrt{2/n-1}\Gamma(n/2)/\Gamma((n-1)/2)$ yields an estimate of expected confidence interval length. Again, the Tukey-McLaughlin critical values may be incorporated into $s_n$ and the same argument used to compute "swindled" ECIL's based on the estimated inflation factors $\lambda^*$. Thus we simply compute

$$ECIL = 2\lambda^* E(k(\hat{\alpha})s_n) = 2\lambda^* E(k(\hat{\alpha}_S)s_n/\hat{\sigma}_0)E(\hat{\sigma}_0),$$

where $k(\alpha)$ denotes the nominal (level .025) critical value for the $t$ on $n - 2[n\alpha] - 1$ degrees of freedom.

5.2 *Monte-Carlo design*

We consider 9 Monte-Carlo configurations: three sample sizes, $n = 25, 50, 100$, and three distribution functions, all Student's $t$ on 1, 3, and 12 degrees of freedom. The number of Monte-Carlo replications, $R$, is 1000 for each configuration.

Four estimators are considered: two fixed-$\alpha$ trimmed means with $\alpha = .1$ and $\alpha = .25$, and two adaptive trimmed means. In the tables $JATM(\alpha_2)$ will refer to a slightly modified version of the Jaeckel adaptive trimmed mean implemented in the Princeton Robustness Study (Andrews *et al.* (1972)) with $\hat{\alpha}$ chosen to minimise $V_n(\alpha)$ over $[\alpha_1, \alpha_2]$. The lower bound $\alpha_1$ is chosen to be .06 in all cases, and we consider two choices of $\alpha_2$: .25 and .44. Similarly, $SATM(\alpha_2)$ will refer to the smooth trimmed mean based on the weight function $h_\alpha(u)$ as in (2.1) with $\sigma = 0.01$ and with $\hat{\alpha}$ chosen to minimize

$$s_n^2(\alpha) = \iint (F_n(y \wedge z) - F_n(y)F_n(z))h_\alpha(F_n(y))h_\alpha(F_n(z))dF_n(y)dF_n(z)$$

over $[\alpha_1, \alpha_2]$. Some experimentation with choosing $\hat{\alpha}$ as for the $SATM$, but then using the classical trimmed mean with trimming proportion $\hat{\alpha}$, indicated that the location estimators were essentially identical. Several modifications of the Andrews *et al.* (1972) version of the Jaeckel adaptive trimmed mean were required—notably their centering by the Winsorized mean (!) was altered to accord with $V_n(\alpha)$ in (1.1). The smoothed estimates based on the weight function (2.1) were computed by explicitly computing integrals defining $T_n(\alpha)$ and $s_n^2(\alpha)$ over the piecewise constant segments of $F_n$. (Fortran source for both estimators is available on request.) Both estimators evaluate an estimate of the asymptotic variance on a grid and choose $\hat{\alpha}_S$ as the minimizer on the grid. The grid is 50 equally spaced values between $\alpha_1$ and $\alpha_2$.

The choice of the upper bound $\alpha_2$ for the $\alpha$ grid is a rather delicate issue. This is largely a consequence of the fact that in all the symmetric situations we have considered the asymptotic variance $\sigma^2(\alpha)$ is quite flat as a function of $\alpha$. This is illustrated for several familiar distributions in Fig. 1. Moreover, as $\alpha \to 1/2$ the "effective sample size" $n_\alpha = n - 2[n\alpha]$ tends to zero. This has two effects: there is a perceptible downward bias in $s_n^2$ for large values of $\alpha$ when we normalize by $n_\alpha$, and secondly, the variability of $s_n^2(\alpha)$ increases with $\alpha$. The former effect may be treated by normalizing by $n_\alpha - 1$ rather than $n_\alpha$; this treatment already appears in the code of Andrews *et al.* (1972). The latter problem is more difficult to treat, but it is clear that it contributes significantly to the tendency of both adaptive procedures to overestimate $\alpha$, and when this occurs it appears to produce an underestimation of $\sigma^2(\alpha_S)$. The latter phenomenon is reflected in rather large correction factors for the critical values when $\alpha_2 = .40$ and $n$ is moderate, say less than 100. Substantially better performance is achieved by restricting $\alpha$ to the range $[0, .25]$ at least over the range of conditions investigated in our experiment. Since even in relatively extreme situations like the Cauchy, the optimal trimming proportion is only about one-third, the restriction of $\hat{\alpha}_S$ to values below this level is less severe than might, at first, be thought.

### 5.3 *Results*

In Table 2 we report the raw mean-squared errors of the four estimators in the 9 experimental configurations. A table of "swindled mses", mean square errors of $(T_n - \hat{\mu})$ is essentially identical and therefore has been omitted. As noted above,

Table 2. Mean-squared error performance.

Student on 1 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.4193 | 0.1388 | 0.1499 | 0.1188 | 0.1555 | 0.1190 |
| $n = 50$ | 0.1122 | 0.0549 | 0.0572 | 0.0487 | 0.0598 | 0.0486 |
| $n = 100$ | 0.0506 | 0.0272 | 0.0280 | 0.0249 | 0.0284 | 0.0247 |

Student on 3 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.0713 | 0.0645 | 0.0653 | 0.0706 | 0.0662 | 0.0713 |
| $n = 50$ | 0.0339 | 0.0310 | 0.0321 | 0.0321 | 0.0321 | 0.0327 |
| $n = 100$ | 0.0160 | 0.0154 | 0.0156 | 0.0165 | 0.0155 | 0.0168 |

Student on 12 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.0419 | 0.0454 | 0.0445 | 0.0487 | 0.0444 | 0.0503 |
| $n = 50$ | 0.0237 | 0.0249 | 0.0242 | 0.0266 | 0.0242 | 0.0272 |
| $n = 100$ | 0.0117 | 0.0130 | 0.0120 | 0.0133 | 0.0119 | 0.0136 |

two versions of both adaptive estimators are reported in the table, one with the upper bound $\alpha_2 = .25$, and the other with $\alpha_2 = .44$.

Perhaps the most striking feature of Table 2 is the excellent performance of the 25% fixed trimmed mean. Clearly 10% trimming is insufficient to cope with the (admittedly somewhat extreme) Cauchy situation, but 25% trimming sacrifices very little in the (near Gaussian) $t$ on 12 situation. The mse performance of the adaptive trimmed means is quite good. In the three Cauchy situations the SATM (.44) and JATM (.44) are slightly better than the 25% fixed trimmed mean—one might expect this since the optimal $\alpha$ for the Cauchy is about .37. In the $t$ on 3 situation 25% trimming is nearly optimal, so the excellent performance of 25% fixed trimming is not surprising. Here the SATM's sacrifice little to 25% fixed trimming and perform somewhat better than the Jaeckel estimators. In the $t$ on 12 case, 10% trimming is excellent, but again the smoothly adaptive trimming is essentially as good with similar performance from the adaptive Jaeckel's.

In Table 3 we report estimated correction factors for the Tukey-McLaughlin critical values for a two-sided 5% test based on the test statistic $T_n/s_n$. Thus, an entry of one indicates no adjustment is necessary to achieve correct nominal size. An entry less than one indicates that the Tukey-McLaughlin critical values are conservative, leading us to reject less often than indicated by the nominal size of the test. For Cauchy observations the Tukey-McLaughlin critical values for fixed trimming are quite conservative, as they are for adaptive estimators when $\alpha_2 = .25$. However when $\alpha_2$ is allowed to be as large as .44, we begin to see the tendency for the adaptive procedures to overestimate the optimal $\alpha$ and the underestimation of $\sigma^2(\alpha_S)$ for large $\alpha$ is reflected in the large estimated correction

Table 3. Tukey-McLaughlin correction factors.

Student on 1 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.8192 | 0.8930 | 0.9345 | 1.0096 | 0.9965 | 1.4194 |
| $n = 50$ | 0.8843 | 0.8668 | 0.9182 | 1.0707 | 0.9607 | 1.2550 |
| $n = 100$ | 0.9018 | 0.9522 | 0.9487 | 1.0866 | 0.9828 | 1.1827 |

Student on 3 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.8954 | 0.9722 | 1.0389 | 1.0861 | 1.0993 | 1.4408 |
| $n = 50$ | 0.9826 | 0.9427 | 1.0246 | 1.1551 | 1.0597 | 1.3274 |
| $n = 100$ | 0.9789 | 0.9831 | 1.0179 | 1.1512 | 1.0447 | 1.2582 |

Student on 12 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.9180 | 0.9838 | 1.0598 | 1.1928 | 1.1140 | 1.5558 |
| $n = 50$ | 0.9886 | 0.9589 | 1.0398 | 1.1849 | 1.0765 | 1.3600 |
| $n = 100$ | 0.9772 | 0.9868 | 1.0137 | 1.1386 | 1.0389 | 1.2326 |

Table 4. Estimated (expected, size-adjusted) confidence interval lengths.

Student on 1 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 2.0501 | 1.3486 | 1.3778 | 1.4861 | 1.3868 | 1.7366 |
| $n = 50$ | 1.1563 | 0.8633 | 0.8745 | 0.8903 | 0.8878 | 0.9387 |
| $n = 100$ | 0.7993 | 0.6060 | 0.6120 | 0.6111 | 0.6166 | 0.6221 |

Student on 3 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 1.0272 | 1.0462 | 1.0056 | 1.1228 | 1.0177 | 1.3144 |
| $n = 50$ | 0.6957 | 0.6937 | 0.6883 | 0.7419 | 0.6906 | 0.7793 |
| $n = 100$ | 0.4902 | 0.4798 | 0.4832 | 0.5131 | 0.4846 | 0.5291 |

Student on 12 degrees of freedom

| | TM(.1) | TM(.25) | SATM(.25) | SATM(.44) | JATM(.25) | JATM(.44) |
|---|---|---|---|---|---|---|
| $n = 25$ | 0.8671 | 0.9496 | 0.8858 | 1.0603 | 0.8874 | 1.2312 |
| $n = 50$ | 0.5950 | 0.6376 | 0.6081 | 0.6758 | 0.6081 | 0.7174 |
| $n = 100$ | 0.4143 | 0.4382 | 0.4209 | 0.4567 | 0.4205 | 0.4714 |

factors for the adaptive methods when $\alpha_2 = .44$. When $\hat{\alpha}_S$ is constrained to be less than 25%, the correction factors are more reasonable, enough so that were we to use the Tukey-McLaughlin critical values we would not be far off. The

SATM correction factors are consistently nearer one than those for the JATM's confirming our expectation that the smoothed estimate of the variability of the trimmed mean is more reliable than the classical Winsorized variance estimate.

In Table 4 we report expected confidence interval lengths for the experiment. Here the performance of the adaptive trimmed means is quite good for $\alpha_2 = .25$, but significantly worse for $\alpha_2 = .44$. The smoothing, it will be noted, yields somewhat more reliable estimates of variability and slightly better expected confidence interval lengths than the Jaeckel version of the adaptive trimmed mean, as suggested by the preceding theory.

## References


Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press, New Jersey.

Bickel, P. (1965). On some robust estimates of location, *Ann. Math. Statist.*, **36**, 847–858.

Bickel, P. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. II, *Ann. Statist.*, **3**, 1045–1069.

Efron, B. and Olshen, R. A. (1978). How broad is the class of normal scale mixtures?, *Ann. Statist.*, **6**, 1159–1164.

Gross, A. M. (1973). A Monte Carlo swindle for estimators of location, *Applied Statistics*, **22**, 347–353.

Gross, A. M. (1977). Confidence intervals for bisquare regression estimates, *J. Amer. Statist. Assoc.*, **72**, 341–354.

Hall, P. (1981). Large sample properties of Jaeckel's adaptive trimmed mean, *Ann. Inst. Statist. Math.*, **33**, 449–462.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based in Influence Functions*, Wiley, New York.

Hill, M. A. and Dixon, W. J. (1982). Robustness in real life: a study of clinical laboratory data, *Biometricsi*, **38**, 377–396.

Huber, P. J. (1972). Robust statistics: a review, *Ann. Math. Statist.*, **43**, 1041–1067.

Jaeckel, L. A. (1971). Some flexible estimates of location, *Ann. Math. Statist.*, **42**, 1540–1552.

Jurečková, J. (1986). Asymptotic representations of $L$-estimators and their relations to $M$-estimators, *Sequential Anal.*, **5**, 317–338.

Relles, D. A. (1970). Variance reduction techniques for Monte Carlo sampling from Student distributions, *Technometrics*, **12**, 499–515.

Rocke, D. M., Downs, G. W. and Rocke, A. J. (1982). Are robust estimators really necessary?, *Technometrics*, **24**, 95–101.

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population, *J. Res. Nat. Bur. Standards*, **64B**, 145–150.

Spjotvoll, E. and Aastreit, A. H. (1980). Comparison of robust estimators on data from field experiments, *Scand. J. Statist.*, **7**, 1–13.

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean, *Ann. Statist.*, **1**, 472–477.

Stigler, S. M. (1977). Do robust estimators work with real data?, *Ann. Statist.*, **5**, 1055–1077.

Tukey, J. W. and McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample (Trimming/Winsorisation 1), *Sankhyā Ser. A*, **25**, 331–352.

Welsh, A. H. (1988). Asymptotically efficient estimation of the sparsity function at a point, *Statist. Probab. Lett.*, **6**, 427–432.