*The Annals of Statistics*
1987, Vol. 15, No. 1, 20–36

# THE TRIMMED MEAN IN THE LINEAR MODEL

BY A. H. WELSH

*University of Chicago*

For the general linear model with independent errors, we propose and examine the large sample properties of an estimator of the regression parameter. In the location model, the estimator has the same properties as the trimmed mean and the robustness and efficiency properties of the trimmed mean carry over to the general model. The estimator depends on a preliminary estimate of the regression parameter and the residuals based on it. The properties of the adaptive estimator with data-determined trimming proportions are also investigated.

**1. Introduction.** The trimmed mean has long been a popular estimator of location; see Tukey and McLaughlin (1963), Bickel (1965) and Huber (1972) for accounts of its history. The popularity of the trimmed mean seems attributable to both theoretical and practical considerations. From a theoretical viewpoint, the trimmed mean is efficient under a variety of circumstances (Bickel (1965), Bickel and Lehmann (1975)) and robust for smooth distributions (Stigler (1973)). The trimmed mean has a strong intuitive appeal—the mean/median tradeoff is clear—and, being an explicit estimator, is both easy to compute and understand. Finally and most importantly, the estimator (with 10% trimming) seems to work very well on real data; see Stigler (1977), Spjøtvoll and Aastreit (1980), Hill and Dixon (1982) and Rocke, Downs and Rocke (1982). The generalisation of the trimmed mean to the linear model has proven problematical. Our purpose in this paper is to generalise the trimmed mean to the linear model in a direct, computationally simple way.

For definiteness, suppose that we observe $Y_1, \dots, Y_n$, where

$$Y_j = x_j'\theta_0 + e_j, \qquad 1 \le j \le n, \tag{1}$$

with $\{x_j' = (x_{j1}, \dots, x_{jp})\}$ a sequence of known $p$ vectors ($p \ge 1$), $\theta_0 \in \mathbf{R}^p$ an unknown parameter to be estimated and $\{e_j\}$ a sequence of independent and identically distributed random variables with common distribution function $F$. The regressors may depend on $n$ but we suppress this dependence for simplicity. We take $x_{j1} = 1$, $1 \le j \le n$, and without loss of generality suppose that $F(0) = \frac{1}{2}$. For any $\theta \in \mathbf{R}^p$, the residuals from $\theta$ are

$$e_j(\theta) = Y_j - x_j'\theta = e_j - x_j'(\theta - \theta_0), \qquad 1 \le j \le n,$$

and $e_j(\theta_0) = e_j$, $1 \le j \le n$. The location problem corresponds to the case $p = 1$ and $x_j = 1$, $1 \le j \le n$. In this context, the trimmed mean may be defined in

Received June 1985; revised December 1985.

*AMS* 1980 *subject classifications*. Primary 62G05; secondary 60F05, 62J05.

*Key words and phrases*. Adaptive estimation, linear model, quantiles, robust estimation, trimmed mean.

terms of the order statistics $Y_{n1} \leq Y_{n2} \leq \cdots \leq Y_{nn}$ as

$$T_n = ([n\beta] - [n\alpha])^{-1} \sum_{j=[n\alpha]+1}^{[n\beta]} Y_{nj}, \qquad 0 < \alpha < \tfrac{1}{2} < \beta < 1,$$

where $[\cdot]$ is the greatest integer function. An alternative functional representation in terms of the empirical distribution function $F_n(s) = n^{-1}\sum_{j=1}^{n} I(Y_j \leq s)$, $s \in \mathbf{R}$, is also useful. For any distribution function $G$, put

$$T(G) = (\beta - \alpha)^{-1} \int_{\alpha}^{\beta} G^{-1}(t)\, dt, \tag{2}$$

where $G^{-1}(t) = \inf\{s\colon G(s) \geq t\}$. Then $T_n \doteq T(F_n)$. A simple approach to extending the definition of $T_n$ to the full model (1) is to first extend the definition of order statistics and quantiles to (1) and then average the central "order statistics."

Bickel (1973) and Ruppert and Carroll (1980) construct estimators based on a preliminary estimate $\theta_n$ of the regression parameter $\theta_0$ and the resulting residuals $\{e_j(\theta_n), 1 \leq j \leq n\}$. Although Bickel's estimators have good asymptotic properties, the estimators are complicated, computationally complex and not invariant to reparametrisation. These problems are due to the componentwise construction of the $p$-vector estimator which permits each component to be trimmed differently. The estimator investigated by Ruppert and Carroll ($\rho_n$ below) is essentially the least squares estimator calculated after removing the observations whose residual is less than $\xi_{n\alpha}(\theta_n)$ or greater than $\xi_{n\beta}(\theta_n)$, where $\xi_{nq}(\theta_n)$ is the $q$th quantile of $\{e_j(\theta_n), 1 \leq j \leq n\}, 0 < q < 1$. The asymptotic properties of this estimator depend on those of $\theta_n$, so that in general the estimator is neither robust nor efficient. It turns out that when $F$ is symmetric, a particular preliminary regression parameter estimate results in a robust and efficient estimator. However, to some extent, this particular choice of preliminary estimate increases the complexity of the final estimator and also decreases its intuitive appeal.

Instead of viewing the usual quantile estimation problem as a problem of ordering observations, Koenker and Bassett (1978) viewed it as an appropriate minimisation problem. In generalising this approach, Koenker and Bassett introduced the vector regression quantiles $\eta_n(q) \in \mathbf{R}^p$, $0 < q < 1$. They proposed the trimmed mean to be the least squares estimator calculated after discarding those observations whose residual from the $\alpha$th regression quantile is negative (i.e., $e_j(\eta_n(\alpha)) < 0$) or whose residual from the $\beta$th regression quantile is positive (i.e., $e_j(\eta_n(\beta)) > 0$). Ruppert and Carroll (1980) noted that the regression quantiles are $M$-estimates (see Huber (1981), page 43 for general definitions) and showed that the Koenker–Bassett estimator has the requisite asymptotic properties. Although the regression quantiles can be computed by standard linear programming techniques, the computation of the estimator is still complicated. Moreover, it is not possible to trim any arbitrary number, $r$ say, of observations; the construction of the estimator permits $r$ to take on only certain values.

Our approach is to examine the structure of the estimator $T_n = T(F_n)$ defined in (2) from the von Mises functional viewpoint. Under appropriate regularity conditions (de Wet and Venter (1974)), the distribution of $n^{1/2}(T_n - T(F))$ is asymptotically equivalent to that of $n^{-1/2}\sum_{j=1}^n \psi(e_j)$, where $\psi(\cdot)$ is the influence curve of the trimmed mean. Put $\xi_q = F^{-1}(q)$, $0 < q < 1$. Then

$$\psi(x) = \phi(x) - E\phi(e_1), \tag{3}$$

where

$$\phi(x) = \{\xi_\alpha I(x < \xi_\alpha) + xI(\xi_\alpha \le x \le \xi_\beta) + \xi_\beta I(x > \xi_\beta)\}/(\beta - \alpha).$$

Notice that we can write

$$\psi(x) = \big[\xi_\alpha\{I(x < \xi_\alpha) - \alpha\} + xI(\xi_\alpha \le x \le \xi_\beta) + \xi_\beta\{I(x > \xi_\beta) - (1 - \beta)\}\big]/(\beta - \alpha) - T(F).$$

The influence curve term is a sum of independent random variables and as such is easier to generalise than $T_n$. The exact representation for the trimmed mean given by Stigler (1973) indicates the error inherent in using the asymptotic approximation. We are led to the following estimator. Let $\theta_n$ be a preliminary regression parameter estimator such as, but not necessarily, the least squares estimator. Let $e_{n1}(\theta_n) \le e_{n2}(\theta_n) \le \cdots \le e_{nn}(\theta_n)$ denote the ordered residuals from $\theta_n$ and for $0 < q < 1$ put

$$\xi_{nq}(\theta_n) = \begin{cases} e_{n,nq}(\theta_n), & \text{if } nq \text{ is an integer,} \\ e_{n,[nq]+1}(\theta_n), & \text{otherwise.} \end{cases}$$

Then put

$$J_j = I\{e_j(\theta_n) \le \xi_{n\alpha}(\theta_n)\},$$
$$K_j = I\{\xi_{n\alpha}(\theta_n) < e_j(\theta_n) \le \xi_{n\beta}(\theta_n)\},$$
$$L_j = I\{e_j(\theta_n) > \xi_{n\beta}(\theta_n)\}$$

and define

$$\tau_n = A_n^- \sum_{j=1}^n x_j\big[\xi_{n\alpha}(\theta_n)\{J_j - \alpha\} + Y_jK_j + \xi_{n\beta}(\theta_n)\{L_j - (1 - \beta)\}\big],$$

where $A_n^-$ is any generalised inverse of $A_n = \sum_{j=1}^n x_jx_j'K_j$. Under mild conditions, $A_n$ is asymptotically nonsingular (see Lemma A.4 of Ruppert and Carroll (1980)). If $\theta_n$ is regression and scale equivariant and invariant to reparameterisation then so is $\tau_n$. Notice that apart from premultiplication by the random matrix $A_n^-\sum_{j=1}^n x_jx_j'$, $\tau_n$ is the least squares estimator calculated after replacing each observed $Y_j$ by

$$\xi_{n\alpha}(\theta_n)\{J_j - \alpha\} + Y_jK_j + \xi_{n\beta}(\theta_n)\{L_j - (1 - \beta)\},$$

which resembles a Winsorising observation. However, $\tau_n$ is *not* a generalisation of the Winsorised mean. It is convenient to denote the middle term of $\tau_n$ by

$$\rho_n = A_n^- \sum_{j=1}^{n} x_j Y_j K_j.$$

Then $\rho_n$ is essentially the estimator investigated by Ruppert and Carroll (1980), who showed that $\rho_n$ is not a generalisation of the trimmed mean.

Let

$$i(q) = \begin{cases} nq, & \text{if } nq \text{ is an integer,} \\ [nq] + 1, & \text{otherwise,} \end{cases}$$

and define $(D(1), \ldots, D(n))$ by

$$e_{n,j}(\theta_n) = e_{D(j)}(\theta_n), \qquad 1 \le j \le n.$$

Then we can write

$$\rho_n = A_n^- \sum_{j=i(\alpha)+1}^{i(\beta)} x_{D(j)} Y_{D(j)}$$

and

$$\tau_n = \xi_{n\alpha}(\theta_n) A_n^- \left( \sum_{j=1}^{i(\alpha)} x_{D(j)} - \alpha \sum_{j=1}^{n} x_j \right) + \rho_n$$
$$+ \xi_{n\beta}(\theta_n) A_n^- \left( \sum_{j=i(\beta)+1}^{n} x_{D(j)} - (1 - \beta) \sum_{j=1}^{n} x_j \right).$$

The above expression resembles a vector version of the "interesting variation" proposed by Bickel ((1973), page 601). However, the observations to be trimmed are selected differently here and the estimator is not constructed componentwise so that $\tau_n$ is not one of the estimators considered by Bickel. Notice that $r$ observations will be trimmed in each tail if and only if $(r-1)/n < \alpha \le r/n$ and $(n-r-1)/n < \beta \le (n-r)/n$, so that if we require $\beta = 1 - \alpha$, $r$ observations will be trimmed in each tail if and only if $\alpha = r/n$. Moreover, in the location problem

$$\tau_n = \{i(\beta) - i(\alpha)\}^{-1} \Bigg[ \xi_{n\alpha}(\theta_n)\{i(\alpha) - n\alpha\}$$
$$+ \sum_{j=i(\alpha)+1}^{i(\beta)} Y_{nj} + \xi_{n\beta}(\theta_n)\{n\beta - i(\beta)\} \Bigg],$$

where $Y_{n1} \le Y_{n2} \le \cdots \le Y_{nn}$. Thus, in the location problem $\tau_n = T_n$ if and only if $n\alpha$ and $n\beta$ are integers. The above properties are useful when $\alpha$ and $\beta$ are chosen after looking at the sample because then it is natural for $n\alpha$ and $n\beta$ to be chosen to be integers. However, if $\alpha$ and $\beta$ are specified without any knowledge of the data and $\beta = 1 - \alpha$, then it is unlikely that $n\alpha$ will be an integer and hence that the desired symmetric trimming will occur. In this case it is better to define a new estimator $\tau_n^*$, which is of the same form as $\tau_n$ but with $i(\alpha)$ replaced by $i(\alpha) - 1$ and $\rho_n$ correspondingly modified. For $\tau_n^*$, exactly $r$ observations will be trimmed in each tail if and only if $r/n < \alpha \le (r+1)/n$ and $(n-r-1)/n < \beta \le (n-r)/n$, so that with $\beta = 1 - \alpha$, $r$ observations will be trimmed in each tail if and only if $n\alpha$ is not an integer. Note that like Bickel's "interesting variation," $\tau_n^*$ does not reduce to $T_n$ for finite samples in the location problem. Since the difference between $\tau_n$ and $\tau_n^*$ is really only a difference between inequalities and strict inequalities, $\tau_n$ and $\tau_n^*$ have the same asymptotic properties under the smoothness conditions we impose on the underlying distribution. However, in finite samples, the estimator should be selected when the decision is made on how to select $\alpha$ and $\beta$.

Bickel (1975) defined a one-step $M$-estimator (type 1) as

$$\mu_n = \theta_n + \left\{ \sum_{j=1}^{n} x_j x_j' \psi'\big(e_j(\theta_n)/s_n(\theta_n)\big) \right\}^{-} \sum_{j=1}^{n} x_j s_n(\theta_n) \psi\big(e_j(\theta_n)/s_n(\theta_n)\big),$$

where $s_n(\theta_n)$ is a robust estimator of the scale of the residuals $e_j(\theta_n)$, $1 \le j \le n$, such as the median absolute deviation from the median and $\psi$ is some real function. The Huber $M$-estimator uses

$$\psi(x) = -MI(x < -M) + xI(-M \le x \le M) + MI(x > M),$$

for some $M < \infty$. With

$$J_j^* = I\{e_j(\theta_n) < -Ms_n(\theta_n)\},$$
$$K_j^* = I\{-Ms_n(\theta_n) \le e_j(\theta_n) \le Ms_n(\theta_n)\}$$

and

$$L_j^* = I\{e_j(\theta_n) > Ms_n(\theta_n)\},$$

the one-step (type 1) Huber $M$-estimator can be written as

$$\mu_n = \left\{ \sum_{j=1}^{n} x_j x_j' K_j^* \right\}^{-} \sum_{j=1}^{n} x_j \{-Ms_n(\theta_n) J_j^* + Y_j K_j^* + Ms_n(\theta_n) L_j^*\}.$$

While $\mu_n$ is of roughly the same form as $\tau_n^*$, there is an important fundamental difference between them; whether or not an observation is "trimmed" in $\mu_n$ depends strictly on the relative magnitude of its residual from $\theta_n$, while whether or not an observation is trimmed in $\tau_n$ or $\tau_n^*$ depends solely on the position of its residual in the sample of ordered residuals from $\theta_n$. If we alter the definition of $\mu_n$ by replacing $-Ms_n(\theta_n)$ by $\xi_{n\alpha}(\theta_n)$ and $Ms_n(\theta_n)$ by $\xi_{n\beta}(\theta_n)$, the resulting estimator is not strictly an $M$-estimator. Moreover, this estimator weights the trimmed observations differently from $\tau_n^*$ and does not have the same asymptotic distribution as $\tau_n^*$.

In the location problem, Jaeckel (1971) and Shorack (1974) have investigated the properties of the trimmed mean with data-determined trimming proportions. Asymptotically, these estimators have better efficiency properties over a class of distributions than the trimmed mean with fixed trimming proportions. We will develop analogous results for the estimator $\tau_n$ and consider a simple method of determining the trimming proportions. To prevent the proliferation of subscripts, we adopt the convention that a caret (ˆ) means that a quantity is calculated at the data-determined trimming proportions ($\hat{\alpha}$, $\hat{\beta}$).

The results of this paper are presented and discussed in Section 2 and proved in Section 4. We apply $\tau_n$ to two sets of data in Section 3. All probability statements are made at the true parameter value $\theta_0$ and all limits, unless otherwise state, are taken as $n \to \infty$.

**2. Results.** To derive asymptotic results for $\tau_n$, we impose the following conditions, which we denote C:

(C1) $n^{1/2}(\theta_n - \theta_0)$ is bounded in probability;

(C2) $x_{j1} = 1$ for all $j$, $\sum_{j=1}^{n} x_{jk} = 0$, $k = 2, \ldots, p$, for each $n$, and there exists a positive definite matrix $\Gamma$ such that

$$\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} x_j x_j' = \Gamma;$$

(C3) $F$ has a continuous density $f$ that is positive on the support of $F$.

We denote by C′ the same set of conditions with (C3) strengthened to

(C3′) $F$ has a uniformly continuous, positive and bounded density.

Note that the usual least squares estimator satisfies (C1) if (C2) holds and if $\mathrm{Var}(e_1) < \infty$. The first two conditions in (C2) simplify the calculation of the asymptotic bias when $F$ is asymmetric. Otherwise, the second condition of (C2) is not needed. Finally, the last part of (C2) ensures that $\lim_{n \to \infty} \max_{1 \le j \le n} n^{-1} x_j' x_j \to 0$. The conditions C are the same as those imposed by Ruppert and Carroll (1980) and the conditions C′ are essentially the same as those imposed by Bickel (1973).

The first theorem establishes that $\tau_n$ is indeed a generalisation of the trimmed mean and, moreover, that $\tau_n$ is asymptotically equivalent to the estimator proposed by Koenker and Bassett (1978). The proof is similar to that of Theorem 1 of Ruppert and Carroll (1980) and is outlined in Section 4.

THEOREM 1. *Suppose that the conditions* C *hold. Then with*

$$\tau_0 = \theta_0 + (T(F), 0, \ldots, 0)' \in \mathbf{R}^p,$$

$$n^{1/2}(\tau_n - \tau_0) - n^{1/2}\Gamma^{-1} \sum_{j=1}^{n} x_j \psi(e_j) \to_P 0.$$

*It follows that*

$$n^{1/2}(\tau_n - \tau_0) \to_D N(0, \sigma^2(\alpha, \beta)\Gamma^{-1}),$$

*where*

$$\sigma^2(\alpha, \beta) = (\beta - \alpha)^{-2}\left[\int_{\xi_\alpha}^{\xi_\beta}(t - T(F))^2\, dF(t) + \alpha\kappa_\alpha^2 + (1-\beta)\kappa_\beta^2 - \left\{(1-\beta)\kappa_\beta + \alpha\kappa_\alpha\right\}^2\right],$$

*with* $\kappa_q = \xi_q - T(F)$, $q = \alpha, \beta$.

To carry out statistical inference, we need to be able to estimate the asymptotic variance $\sigma^2(\alpha, \beta)\Gamma^{-1}$. The second theorem, which may be proved by a similar argument to that used to prove Theorem 5 of Ruppert and Carroll (1980), shows how to construct a consistent estimator.

THEOREM 2. *Suppose that the conditions* C *hold. Let*

$$\bar{e}_K = \{n(\beta - \alpha)\}^{-1} \sum_{j=1}^{n} e_j(\theta_n) K_j,$$

$$\kappa_{nq}(\theta_n) = \xi_{nq}(\theta_n) - \bar{e}_K, \qquad 0 < q < 1,$$

*and define*

$$S_n^2(\alpha, \beta) = (\beta - \alpha)^{-2}\left[(n-p)^{-1} \sum_{j=1}^{n} \left\{e_j(\theta_n) - \bar{e}_K\right\}^2 K_j + \alpha\kappa_{n\alpha}^2(\theta_n) + (1-\beta)\kappa_{n\beta}^2(\theta_n) - \left\{\alpha\kappa_{n\alpha}(\theta_n) + (1-\beta)\kappa_{n\beta}(\theta_n)\right\}^2\right].$$

*Then*

$$S_n^2(\alpha, \beta)\left(n^{-1} \sum_{j=1}^{n} x_j x_j'\right)^{-1} \to^P \sigma^2(\alpha, \beta)\Gamma^{-1}.$$

Of course, for $\tau_n^*$, the quantities $\bar{e}_K$ and $S_n^2$ should be appropriately modified. In the location problem, these results reduce to the analogous results for the trimmed mean. Indeed, Theorem 1 implies that asymptotically $\tau_n$ has the same robustness and efficiency properties in the linear model context as the trimmed mean has in the location problem. This is also true of the estimators of Bickel (1973) and Koenker and Bassett (1978). The estimator constructed in Theorem 2 depends only on the preliminary estimator $\theta_n$ and not on $\tau_n$. This simplifies computation, but of course the efficacy of the estimator depends on $\theta_n$. It is easy to construct estimators depending on $\tau_n$ too. The quantity $T(F)$ represents the asymptotic bias of the estimator. As in the case of $M$-estimators (Carroll (1979)), the bias involves the intercept but not the slopes. If $F$ is symmetric then $T(F)$ is the center of symmetry, which, without loss of generality we have taken to be zero. In this case it is usual to take $\beta = 1 - \alpha$, $0 < \alpha < \frac{1}{2}$, in the definition of $\tau_n$.

We now consider the properties of $\tau_n$ with $(\alpha, \beta)$ replaced by the data-determined quantity $(\hat{\alpha}, \hat{\beta})$. We show (Theorem 3) that under conditions C′, if $(\hat{\alpha}, \hat{\beta}) \to_P (\alpha_0, \beta_0)$, then $\tau_n$ calculated with $(\hat{\alpha}, \hat{\beta})$ (which we denote $\hat{\tau}_n$) has the same asymptotic distribution as $\tau_n$ calculated with $(\alpha_0, \beta_0)$.

The proof of Theorem 3 depends on uniform convergence arguments. In particular, we require the following result, which can be extrated from the proof of Theorem A.4 of Koul (1969). Bickel (1973) noted that Koul's conditions can be weakened slightly; we adopt the conditions imposed by Bickel (1973).

LEMMA 1 (Koul (1969); Bickel (1973)). *Let* $\{c_{jn}\}$ *be any sequence of constants such that* $\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n} c_{jn}^2 < \infty$ *and* $\lim_{n\to\infty} \max_{1\le j\le n} n^{-1/2}|c_{jn}| = 0$ *hold. Then if* C′ *holds,*

$$\sup_{-\infty<s<\infty}\ \sup_{|t|\le n^{-1/2}M}\left| n^{-1/2}\sum_{j=1}^{n} c_{jn}\{I(e_j \le s + t'x_j) - F(s + t'x_j) - I(e_j \le s) + F(s)\}\right| \to_P 0$$

*for any* $M < \infty$.

We also require a slight modification of this result. The proof of the following lemma, which extends Theorems A.3 and A.4 of Koul (1969), is given in Section 4.

LEMMA 2. *Let g be a real, continuous, nonnegative function defined on* $\mathbf{R}^+$. *Suppose that the conditions* C′ *hold. Then*

$$\sup_{s_1\le s\le s_2}\ \sup_{|t|\le n^{-1/2}M}\left| n^{-1/2}\sum_{j=1}^{n} x_j\{g(e_j)I(0 \le e_j \le s + t'x_j) - Eg(e_1)I(0 \le e_1 \le s + t'x_j) - g(e_j)I(0 \le e_j \le s) + Eg(e_1)I(0 \le e_1 \le s)\}\right| \to_P 0,$$

*for any* $M < \infty$, $0 < s_1 < s_2 < \infty$.

The above results lead to the following theorem for the adaptive estimator.

THEOREM 3. *Suppose that* $(\hat{\alpha}, \hat{\beta}) \to_P (\alpha_0, \beta_0)$, $0 < \alpha_0 < \frac{1}{2} < \beta_0 < 1$, *and that the conditions* C′ *hold. Then with* $\hat{\tau}_0 = \theta_0 + (\hat{T}(F), 0, \ldots, 0)' \in \mathbf{R}^p$, *where* $\hat{T}(F) = (\hat{\beta} - \hat{\alpha})^{-1}\int_{\xi_{\hat{\alpha}}}^{\xi_{\hat{\beta}}} t\, dF(t)$, *and* $\psi$ *evaluated at* $(\alpha_0, \beta_0)$,

$$n^{1/2}(\hat{\tau}_n - \hat{\tau}_0) - n^{-1/2}\Gamma^{-1/2}\sum_{j=1}^{n} x_j\psi(e_j) \to_P 0.$$

It follows that

$$n^{1/2}(\hat{\tau}_n - \hat{\tau}_0) \to_D N(0, \sigma^2(\alpha_0, \beta_0)\Gamma^{-1}).$$

Of course, we can also extend Theorem 2 to show that we can consistently estimate $\sigma^2(\alpha_0, \beta_0)$. In particular, we have the following theorem.

THEOREM 4. *Suppose that the conditions* C′ *hold. Then with* $S_n^2(\alpha, \beta)$ *defined in Theorem* 2,

$$\sup_{\substack{\alpha_1 \le \alpha \le \alpha_2 \\ \beta_1 \le \beta \le \beta_2}} \left| S_n^2(\alpha, \beta) - \sigma^2(\alpha, \beta) \right| \to_P 0,$$

*for any* $0 < \alpha_1 < \alpha_2 < \frac{1}{2}$, $\frac{1}{2} < \beta_1 < \beta_2 < 1$. *If in addition* $(\hat{\alpha}, \hat{\beta}) \to_P (\alpha_0, \beta_0)$, $0 < \alpha_0 < \frac{1}{2} < \beta_0 < 1$, *then* $S_n^2(\hat{\alpha}, \hat{\beta}) \to_P \sigma^2(\alpha_0, \beta_0)$.

The above theorems are analogous to those obtained for the location problem by Jaeckel (1971) and Shorack (1974). If $F$ is symmetric, the bias term is zero if $\hat{\beta} = 1 - \hat{\alpha}$ for each $n$ and $\beta_0 = 1 - \alpha_0$.

Theorems 3 and 4 hold very generally in that the only requirement of the random sequences $\{\hat{\alpha}\}$ and $\{\hat{\beta}\}$ is that $\hat{\alpha} \to_P \alpha_0$ and $\hat{\beta} \to_P \beta_0$, $0 < \alpha_0 < \frac{1}{2} < \beta_0 < 1$. In the special case that $F$ is symmetric, we can show that, under mild regularity conditions, the sequence of minima of $S_n^2(\alpha, 1 - \alpha)$ satisfies the requirement that $\hat{\alpha} \to_P \alpha_0$. The proof depends on the uniform convergence result of Theorem 4 and, being standard, is omitted.

THEOREM 5. *Suppose that $F$ is symmetric and* $\sigma^2(\alpha, 1 - \alpha)$ *has a unique minimum at* $\alpha_0$ *in the interval* $[\alpha_1, \alpha_2]$, *where* $0 < \alpha_1 < \alpha_2 < \frac{1}{2}$. *Then if* $\hat{\alpha} = \{q\colon S_n^2(q, 1 - q) = \inf_{\alpha_1 \le \alpha \le \alpha_2} S_n^2(\alpha, 1 - \alpha)\}$,

$$\hat{\alpha} \to_P \alpha_0.$$

The result remains true if we replace $S_n^2(\alpha, 1 - \alpha)$ by any estimator of $\sigma^2(\alpha, 1 - \alpha)$ that satisfies Theorem 4. In particular, an analogue of the variance estimator proposed by Jaeckel (1971) is given by

$$R_n^2(\alpha) = (1 - 2\alpha)^{-2}\Bigg[(n - p)^{-1} \sum_{j=1}^{n} \{e_j(\theta_n) - \bar{e}_k\}^2 K_j + \alpha\kappa_{n\alpha}(\theta_n) + \alpha\kappa_{n,1-\alpha}(\theta_n)\Bigg]$$

and the value of $\alpha$ that minimises $R_n^2(\alpha)$ converges in probability to $\alpha_0$. As with Jaeckel's estimator in the location problem, calculating the minimising $\alpha$ involves evaluating the variance estimator at each $\alpha$ satisfying $\alpha_1 \le \alpha \le \alpha_2$ and $n\alpha$ is an integer. (With $R_n^2$ appropriately modified for $\tau_n^*$, the minimum on $\alpha_1 \le \alpha \le \alpha_2$ occurs at $n\alpha +$ , where $n\alpha$ is an integer, which is much less convenient.)

Together with Theorems 3 and 4, Theorem 5 yields an analogue of Theorem 1 of Jaeckel (1971). If the underlying distribution is normal, $\alpha_0 = 0$, which is precluded in our results. (This is also true in the location problem.) If $\sigma^2(\alpha, 1-\alpha)$ has several isolated minima, then we can define $\alpha_0$ to be the smallest minimum of $\sigma^2(\alpha, 1-\alpha)$ to obtain a slight variation of Theorem 5. Finally, the advantage of having the variance estimator defined only in terms of the initial estimator and not in terms of $\hat{\tau}_n$ becomes clear when $\hat{\alpha}$ is chosen to minimise the variance estimator; the calculation of $\hat{\alpha}$ involves calculating the variance estimator on a finite grid of points, but we do *not* have to calculate $\hat{\tau}_n$ at each point on the grid. Thus $\hat{\alpha}$ is obtained before we calculate $\hat{\tau}_n$ rather than simultaneously and this greatly reduces the calculations. An alternative is to use $\tau_n$ at a fixed $\alpha$ ($\beta = 1-\alpha$) instead of $\theta_n$ to determine $\hat{\alpha}$. This approach is useful if the preliminary estimate $\theta_n$ is poor.

**3. Examples.** In this section, we consider the application of the trimmed mean proposed in this paper to two sets of data, namely, the stackloss data given by Brownlee (1965) and the water salinity data given by Ruppert and Carroll (1980). Both sets of data are analysed in Ruppert and Carroll (1980) and their results provide a basis for comparison.

For simplicity, we will adopt the least squares estimate as the initial estimate and we will restrict attention to symmetric trimming. We will consider three estimates; $\tau_n^*$ with $\alpha = 0.1$, $\hat{\tau}_n$ with $\hat{\alpha}$ estimated by minimising $R_n(\alpha)$ on $[0.05, 0.35]$ and $\hat{\tau}_n$ with $\hat{\alpha}$ determined by examining a residual plot of the residuals from the least squares fit and making a subjective decision on the number of observations to trim. Notice that if we decide to trim $r$ observations in each tail, we take $\alpha = r/n$ (if we want to trim $r$ in the lower tail and $m$ in the upper tail we would take $\alpha = r/n$ and $\beta = (n-m)/n$). Following Ruppert and Carroll (1980), we quote the interquartile range (IQR) of the residuals as a crude means of assessing fit. The calculations were carried out using the MINITAB package in conjunction with a separate FORTRAN program to evaluate $R_n^2(\alpha)$ on the Decsystem-20 computer of the University of Chicago. The results are tabulated in Table 1.

The stackloss data involve the regression of stackloss on air flow, temperature and acid. The 10% trimmed mean trims observations 21 and 9 in the lower tail and observations 4 and 3 from the upper tail. The function $R_n^2(\alpha)$ is minimised at $\alpha = 2/21$ for which $R_n^2(2/21) = 8.643$, and the same four observations as for $\tau_n^*$ are trimmed. The residual plot of the least squares residuals suggests that observations 4 and 21 and possibly also observations 3 and 9 should be trimmed (if we were not restricting attention to symmetric trimming, we might trim observations 3, 4 and 21). The result of trimming four observations is of course the same as $\tau_n$ above. It is interesting to note that using a different initial estimator Ruppert and Carroll (1980) trimmed observations 1, 3, 9 and 21 and achieved a slightly better fit. The 15% Koenker and Bassett (1978) trimmed mean trims observations 4, 9 and 21.

For the water salinity data, Ruppert and Carroll (1980) regressed salinity on salinity lagged by two weeks, river discharge and a linear time trend. The 10%

TABLE 1
*Results for the stackloss and salinity data.*

| Estimate | Intercept | Air flow | Temperature | Acid | Asympt. variance | IQR |
|---|---|---|---|---|---|---|
| | | | Stackloss Data | | | |
| $\theta_n$ | −39.92 | 0.716 | 1.295 | −0.152 | 10.519 | 3.124 |
| $\tau_n^*(0.1)$ | −40.90 | 0.852 | 0.865 | −0.128 | 8.869 | 2.856 |
| $\hat{\tau}_n(2/21)$ | −40.79 | 0.851 | 0.869 | −0.129 | 8.643 | 2.875 |

| Estimate | Intercept | Lagged salinity | Time trend | Flow | Asympt. variance | IQR |
|---|---|---|---|---|---|---|
| | | | Salinity Data | | | |
| $\theta_n$ | 9.590 | 0.777 | −0.026 | −0.295 | 1.770 | 1.377 |
| $\tau_n^*(0.1)$ | 12.353 | 0.765 | −0.088 | −0.401 | 1.852 | 1.172 |
| $\hat{\tau}_n(3/28)$ | 13.738 | 0.749 | −0.095 | −0.452 | 1.367 | 1.013 |
| $\hat{\tau}_n(2/28)$ | 12.424 | 0.751 | −0.047 | −0.402 | 1.788 | 1.115 |

trimmed mean trims observations 17 and 15 from the lower tail and observations 16 and 9 from the upper tail. In this example, $R_n^2(\alpha)$ is minimized at $\alpha = 3/28$ for which $R_n^2(3/28) = 1.367$. In addition to the observations trimmed by $\tau_n^*$, observations 11 and 13 are trimmed. Finally, the residual plot of the least squares residuals shows that such outliers as may be present are not too extreme but the conservative statistician would probably trim two observations in each tail as was done by $\tau_n^*$. In this case with $\alpha = 2/28$, $R_n^2(2/28) = 1.788$. Ruppert and Carroll (1980) trimmed observations 1, 11, 13, 15, 16 and 17 with their estimator and observations 1, 13, 15 and 17 with the 15% Koenker and Bassett (1978) trimmed mean. The fit for $\hat{\tau}_n$ with $\alpha = 3/28$ is slightly better than that achieved by either of the above estimators.

Of course, it is difficult to assess the performance of the estimators based on their application to two real data sets. Nonetheless, the performance of $\tau_n^*$ and $\hat{\tau}_n$ is certainly comparable to that of the estimators studied by Ruppert and Carroll (1980). Since we are particularly interested in protecting against deviations from the normal assumptions, trimming based on the least squares residuals has a strong intuitive appeal, and makes the calculation of $\tau_n^*$ and $\hat{\tau}_n$ extremely simple. However, in small samples one-step estimators are sensitive to the initial estimator and (since in general the least squares estimate is not a good preliminary estimate), at the expense of computational simplicity, a more robust preliminary estimate, such as for example the least absolute deviations estimate, should be used. An alternative is to iterate on the trimmed mean, but this seems less attractive since the effect of a bad initial estimate may persist. The estimator $\tau_n$ is not robust against outliers in the design space. The adaptive estimator $\hat{\tau}_n$ permits a flexible approach to the analysis of data that is useful in practice.

**4. Proofs.** For any $r \times c$ matrix $D$, let $|D|$ denote the Euclidean norm of $D$, i.e.,

$$|D| = \text{trace}(D'D).$$

To simplify notation, let

$$T_{nq} = \theta_n - \theta_0 + \left(\xi_{nq}(\theta_n) - \xi_q, 0, \dots, 0\right)', \qquad 0 < q < 1,$$

and define

$$\begin{aligned} J_j(T_{n\alpha}) &= I\{e_j(\theta_n) \le \xi_{n\alpha}(\theta_n)\} = I\{e_j \le \xi_\alpha + x_j'T_{n\alpha}\}, \\ K_j(T_{n\alpha}, T_{n\beta}) &= I\{\xi_{n\alpha}(\theta_n) < e_j(\theta_n) \le \xi_{n\beta}(\theta_n)\} \\ &= I\{\xi_\alpha + x_j'T_{n\alpha} < e_j \le \xi_\beta + x_j'T_{n\beta}\} \end{aligned}$$

and

$$L_j(T_{n\beta}) = I\{e_j(\theta_n) > \xi_{n\beta}(\theta_n)\} = I\{e_j > \xi_\beta + x_j'T_{n\beta}\}, \qquad 1 \le j \le n.$$

PROOF OF THEOREM 1. We have

$$\begin{aligned} &\left| n^{-1/2}A_n(\tau_n - \theta_0) - (\beta - \alpha)n^{-1/2} \sum_{j=1}^n x_j\{\phi(e_j) - E\phi(e_j) + T(F)\} \right| \\ &\le \left| n^{-1/2} \sum_{j=1}^n x_j\left[\xi_{n\alpha}(\theta_n)\{J_j(T_{n\alpha}) - \alpha\} - \xi_\alpha\{J_j(0) - \alpha\}\right] \right. \\ &\qquad\qquad \left. - n^{1/2}\Gamma T_{n\alpha}\xi_\alpha f(\xi_\alpha) \right| \\ &\quad + \left| n^{-1/2} \sum_{j=1}^n x_j e_j\{K_j(T_{n\alpha}, T_{n\beta}) - K_j(0,0)\} \right. \\ &\qquad\qquad \left. + n^{1/2}\Gamma\{T_{n\alpha}\xi_\alpha f(\xi_\alpha) - T_{n\beta}\xi_\beta f(\xi_\beta)\} \right| \\ &\quad + \left| n^{-1/2} \sum_{j=1}^n x_j\left[\xi_{n\beta}(\theta_n)\{L_j(T_{n\beta}) - (1-\beta)\} - \xi_\beta\{L_j(0) - (1-\beta)\}\right] \right. \\ &\qquad\qquad \left. + n^{1/2}\Gamma T_{n\beta}\xi_\beta f(\xi_\beta) \right|, \end{aligned} \tag{4.1}$$

and the result follows since each term on the right-hand side of (4.1) converges in probability to zero by arguments similar to those given in the proof of Theorem 1 of Ruppert and Carroll (1980). □

Let $\delta_{jn}(t) = n^{-1/2}x_j't$, $1 \le j \le n$, $t \in \mathbf{R}^p$, and then define

$$\begin{aligned} V_n(t,s) &= n^{-1/2} \sum_{j=1}^n x_j g(e_j) I\{0 \le e_j \le s + \delta_{jn}(t)\}, \\ W_n(t,s) &= V_n(t,s) - EV_n(t,s) \end{aligned}$$

and

$$H(s) = Eg(e_1)I\{0 \le e_1 \le s\} = \int_0^s g(t)f(t)\,dt.$$

Notice that $EV_n(t, s) = n^{-1/2}\sum_{j=1}^n x_j H(s + \delta_{jn}(t))$ and that $h(s) = H'(s) = g(s)f(s)$, which is continuous on a set containing $[0, s_2]$.

PROOF OF LEMMA 2. Arguing as in the proof of Theorem A.3 of Koul (1969), it follows that for each fixed $t_0 \in \mathbf{R}^P$,

$$n^{-1/2} \sum_{j=1}^{n} x_j \Big[ g(e_j) I\{0 \le e_j \le s + n^{-1/2} t_0' x_j\} - Eg(e_1) I\{0 \le e_1 \le s + n^{-1/2} t_0' x_j\} \Big] \tag{4.2}$$

converges weakly in $D^P[s_1, s_2]$ to a Gaussian process. ($D^P[s_1, s_2]$ is the space of $p$-dimensional vector functions such that each component function is an element of $D[s_1, s_2]$.) The conclusion of the lemma is that for any $M < \infty$,

$$\sup_{s_1 \le s \le s_2} \sup_{|t| \le M} |W_n(t, s) - W_n(0, s)| \to_P 0. \tag{4.3}$$

It follows that Lemma A.4 of Ruppert and Carroll (1980) that for each fixed $s \in [s_1, s_2]$, $\sup_{|t| \le M} |W_n(t, s) - W_n(0, s)| \to_P 0$. Hence, (4.3) will obtain if we can show that

$$\sup_{|r-s| \le \delta} \sup_{|t| \le M} |W_n(t, r) - W_n(t, s) - W_n(0, r) + W_n(0, s)| \to_P 0, \tag{4.4}$$

as $n \to \infty$, $\delta \downarrow 0$. It follows from (4.2) that for each fixed $t_0 \in \mathbf{R}^p$,

$$\sup_{|r-s| \le \delta} |W_n(t_0, r) - W_n(t_0, s) - W_n(0, r) + W_n(0, s)| \to_P 0, \tag{4.5}$$

as $n \to \infty$, $\delta \downarrow 0$, so it remains to show that, given $\varepsilon > 0$, we can find an $\varepsilon_1 > 0$ such that

$$\lim_{\delta \to 0} \lim_{n \to \infty} P\Big\{ \sup_{|r-s| \le \delta} \sup_{|t - t_0| \le \varepsilon_1} |W_n(t, r) - W_n(t, s) - W_n(t_0, r) + W_n(t_0, s)| \ge \varepsilon \Big\} = 0 \tag{4.6}$$

for then, given $\varepsilon > 0$, we can choose $\varepsilon_1$ such that (4.6) holds and then choose a grid of points $\{t_i\colon 1 \le i \le k\}$ such that for any $|t| \le M$, $|t - t_i| \le \varepsilon_1$ for some $1 \le i \le k$, so that

$$\begin{aligned} &P\Big\{ \sup_{|r-s| \le \delta} \sup_{|t| \le M} |W_n(t, r) - W_n(t, s) - W_n(0, r) + W_n(0, s)| \ge \varepsilon \Big\} \\ &\quad \le \sum_{i=1}^{k} P\Big\{ \sup_{|r-s| \le \delta} \sup_{|t - t_i| \le \varepsilon_1} |W_n(t, r) - W_n(t, s) - W_n(t_i, r) + W_n(t_i, s)| \ge \varepsilon/2 \Big\} \\ &\quad + \sum_{i=1}^{r} P\Big\{ \sup_{|r-s| \le \delta} |W_n(t_i, r) - W_n(t_i, s) - W_n(0, r) + W_n(0, s)| \ge \varepsilon/2 \Big\} \end{aligned}$$

converges to zero, yielding (4.4).

Now, notice that for each fixed $t_0$,

$$W_n(t, r) - W_n(t, s) - W_n(t_0, r) + W_n(t_0, s)$$
$$= \{V_n(t, r) - V_n(t, s) - V_n(t_0, r) + V_n(t_0, s)\}$$
$$-E\{V_n(t, r) - V_n(t, s) - V_n(t_0, r) + V_n(t_0, s)\}$$

and

$$\lim_{\delta \downarrow 0} \lim_{n \to \infty} \sup_{|r-s| \le \delta} \sup_{|t-t_0| \le M} |E\{V_n(t, r) - V_n(t, s) - V_n(t_0, r) + V_n(t_0, s)\}|$$
$$\le \lim_{\delta \downarrow 0} \lim_{n \to \infty} \sup_{|r-s| \le \delta} \sup_{|t-t_0| \le M} n^{-1/2} \sum_{j=1}^{n} |x_j| \big| H(r + \delta_{jn}(t)) - H(r + \delta_{jn}(t_0))$$
$$+H(s + \delta_{jn}(t)) + H(s + \delta_{jn}(t_0)) \big|$$
$$\le M \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} |x_j|^2 \lim_{\delta \downarrow 0} \sup_{|r-s| \le \delta} |h(r) - h(s)|$$
$$= 0,$$

as $h$ is uniformly continuous on $[s_1, s_2]$, so it remains to show that given $\varepsilon > 0$, we can choose $\varepsilon_1 > 0$ such that for each fixed $t_0$,

$$\lim_{\delta \downarrow 0} \lim_{n \to \infty} P\left\{ \sup_{|r-s| \le \delta} \sup_{|t-t_0| \le \varepsilon_1} |V_n(t, r) - V_n(t, s) - V_n(t_0, r) + V_n(t_0, s)| \ge \varepsilon \right\} = 0. \tag{4.7}$$

However, (4.7) may be proved by the argument of Lemma A.3 of Koul (1969). □

PROOF OF THEOREM 3. It follows from Lemma 4.6 of Bickel (1973) and Theorem A.3 of Koul (1969) that

$$|n^{-1}\hat{A}_n - (\beta_0 - \alpha_0)\Gamma| \to_P 0.$$

Now

$$\Big| n^{-1/2} \sum_{j=1}^{n} x_j \big[ \xi_{n\hat{\alpha}}(\theta_n) \{ I(e_j(\theta_n) \le \xi_{n\hat{\alpha}}(\theta_n)) - \hat{\alpha} \}$$
$$- \xi_{\hat{\alpha}} \{ I(e_j \le \xi_{\hat{\alpha}}) - \hat{\alpha} \} \big] - n^{1/2} \Gamma T_{n\hat{\alpha}} \xi_{\hat{\alpha}} f(\xi_{\alpha_0}) \Big|$$
$$\le |\xi_{n\hat{\alpha}}(\theta_n)| \Big| n^{-1/2} \sum_{j=1}^{n} x_j \big[ I\{ e_j \le \xi_{n\hat{\alpha}}(\theta_n) + x_j'(\theta_n - \theta_0) \}$$
$$- F(\xi_{n\hat{\alpha}}(\theta_n) + x_j'(\theta_n - \theta_0))$$
$$- I(e_j \le \xi_{n\hat{\alpha}}(\theta_n)) + F(\xi_{n\hat{\alpha}}(\theta_n)) \big] \Big|$$

$$
\begin{aligned}
(4.8)\qquad & +|\xi_{n\hat\alpha}(\theta_n)|\left|n^{-1/2}\sum_{j=1}^{n} x_j\{F(\xi_{\hat\alpha}+x_j'T_{n\hat\alpha})-F(\xi_{\hat\alpha})\}\right.\\
&\qquad\qquad \left. -n^{1/2}\Gamma T_{n\hat\alpha}f(\xi_{\alpha_0})\right|\\
&+|\xi_{n\hat\alpha}(\theta_n)|\left|n^{-1/2}\sum_{j=1}^{n} x_j\{I(e_j\le \xi_{n\hat\alpha}(\theta_n))-F(\xi_{n\hat\alpha}(\theta_n))\right.\\
&\qquad\qquad \left. -I(e_j\le\xi_{\hat\alpha})+\hat\alpha\}\right|\\
&+|\xi_{n\hat\alpha}(\theta_n)-\xi_{\hat\alpha}|\left|n^{-1/2}\sum_{j=1}^{n} x_j\{I(e_j\le\xi_{\hat\alpha})-\hat\alpha\}\right.\\
&\qquad\qquad \left. +n^{1/2}\Gamma T_{n\hat\alpha}f(\xi_{\alpha_0})\right|,
\end{aligned}
$$

which converges in probability to zero by Lemma 1, Theorem 4.6 of Bickel (1973) and the weak convergence properties of the process

$$E_n(s)=n^{-1/2}\sum_{j=1}^{n} x_j\{I(e_j\le s)-F(s)\}$$

with the construction of Skorokhod (1956). Similarly,

$$
\begin{aligned}
&\left|n^{-1/2}\sum_{j=1}^{n} x_je_j\left[\Gamma\{\xi_{n\hat\alpha}(\theta_n)<e_j(\theta_n)\le\xi_{n\hat\beta}(\theta_n)\}\right.\right.\\
(4.9)\qquad &\qquad\qquad \left. -I\{\xi_\alpha<e_j\le\xi_{\hat\beta}\}\right]\\
&\left. +n^{1/2}\Gamma\{T_{n\hat\alpha}\xi_{\hat\alpha}f(\xi_{\alpha_0})-T_{n\hat\beta}\xi_{\hat\beta}f(\xi_{\beta_0})\}\right|\to_P 0.
\end{aligned}
$$

Let $\hat\psi$ be $\psi$ (defined in (3)) but evaluated at $(\hat\alpha,\hat\beta)$ and let $\psi_0$ be $\psi$ evaluated at $(\alpha_0,\beta_0)$. Then as in the proof of Theorem 1,

$$\left|n^{-1/2}A_n(\hat\tau_n-\theta_0)-(\hat\beta-\hat\alpha)n^{-1/2}\sum_{j=1}^{n} x_j\{\hat\psi(e_j)+\hat T(F)\}\right|\to_P 0$$

by (4.8) and (4.9). The result will follow if we can show that

$$n^{-1/2}\sum_{j=1}^{n} x_j\{(\hat\beta-\hat\alpha)\hat\psi(e_j)-(\beta_0-\alpha_0)\psi_0(e_j)\}\to_P 0.$$

Let

$$\eta_1(r)=n^{-1/2}\sum_{j=1}^{n} x_j\left[r\{I(e_j<r)-F(r)\}-(-e_j)I\{0\le -e_j\le -r\}+H(r)\right],$$

and

$$\eta_2(s) = n^{-1/2} \sum_{j=1}^{n} x_j \big[ s\{ I(e_j > s) - (1 - F(s)) \} + e_j I\{0 \le e_j \le s\} - H(s) \big],$$

$$-\infty < r_1 \le r \le r_2 < 0, 0 < s_1 \le s \le s_2 < \infty.$$

Since

$$n^{-1/2} \sum_{j=1}^{n} x_j(\beta - \alpha)\psi(e_j) = \eta_1(\xi_\alpha) + \eta_2(\xi_\beta),$$

the proof may be completed by the Skorokhod (1956) construction argument. □

PROOF OF THEOREM 4. By Lemma 4.6 of Bickel (1973),

$$\sup_{q_1 \le q \le 1 - q_1} \left| \xi_{nq}(\theta_n) - \xi_q \right| \to_P 0 \quad \text{for any } 0 < q_1 < \tfrac{1}{2}.$$

Let $A = [\alpha_1, \alpha_2]$ and $B = [\beta_1, \beta_2]$. Then to show that

$$\sup_{\alpha \in A, \beta \in B} \left| \bar{e}_K - T(F) \right| \to_P 0,$$

we write

$$\sup_{\alpha \in A, \beta \in B} \left| n^{-1} \sum_{j=1}^{n} e_j I\{ \xi_{n\alpha}(\theta_n) < e_j(\theta_n) \le \xi_{n\beta}(\theta_n) \} - T(F) \right|$$

$$\le \sup_{\alpha \in A} \left| n^{-1} \sum_{j=1}^{n} (-e_j) I\{0 \le -e_j < -\xi_{n\alpha}(\theta_n) - x_j'(\theta_n - \theta_0)\} - H(\xi_\alpha) \right|$$

$$+ \sup_{\beta \in B} \left| n^{-1} \sum_{j=1}^{n} e_j I\{0 \le e_j \le \xi_{n\beta}(\theta_n) + x_j'(\theta_n - \theta_0)\} - H(\xi_\beta) \right|$$

and apply Lemma 2. Similarly

$$\sup_{\alpha \in A, \beta \in B} \left| n^{-1} \sum_{j=1}^{n} e_j(\theta_n)^2 I\{ \xi_{n\alpha}(\theta_n) < e_j(\theta_n) \le \xi_{n\beta}(\theta_n) \} - \int_{\xi_\alpha}^{\xi_\beta} t^2 \, dF(t) \right| \to_P 0,$$

and the result obtains. □

**Acknowledgments.** I am grateful to the participants in a workshop held at the University of Chicago for helpful comments and to the referees, Associate Editor and Editor for their thorough and helpful reviews.

## REFERENCES

BICKEL, P. J. (1965). On some robust estimators of location. *Ann. Math. Statist.* **36** 847–858.

BICKEL, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for non-parametric models. II. *Ann. Statist.* **3** 1045–1069.

BROWNLEE, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*, 2nd ed. Wiley, New York.

CARROLL, R. J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *J. Amer. Statist. Assoc.* **74** 674–679.

DE WET, T. and VENTER, J. H. (1974). An asymptotic representation of trimmed means with applications. *S. African Statist. J.* **8** 127–134.

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

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

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

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

KOENKER, R. W. and BASSETT, G. W. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

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

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

SHORACK, G. R. (1974). Random means. *Ann. Statist.* **2** 661–675.

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** 261–290.

SPJØTVOLL, 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, G. (1963). Less vulnerable confidence and significance procedures for location based on a single sample. *Sankyā Ser. A* **35** 331–352.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637

## DISCUSSION

P. J. DE JONGH AND T. DE WET

*Institute for Maritime Technology*

This paper proposes a new way of defining trimmed means in the linear model, which differs from earlier proposals by Bickel (1973), Koenker and Bassett (1978) and Ruppert and Carroll (1980). We find the idea of the proposal very interesting. It has the "right" equivariance and asymptotic properties and is thus an attractive (large sample) extension of the trimmed mean in the location case. These properties also hold for the Koenker–Bassett (1978) estimator, but the Welsh estimator has the potential advantage of computational simplicity (if least squares is used as a preliminary estimator). Our remarks will concern the small sample behaviour of the proposed estimator. We wish to