Statistica Sinica **7**(1997), 669-686

AN EFFICIENT CLASS OF WEIGHTED TRIMMED MEANS FOR LINEAR REGRESSION MODELS

Lin-An Chen

National Chiao Tung University

Abstract: We propose and study a class of weighted trimmed means based on the symmetric quantile functions for the location and linear regression models. A robustness comparison with the underlying distribution of a symmetric-type heavy tail is given. The weighted trimmed mean in optimal trimming under symmetric distributions is shown to have an asymptotic variance very close to the Cramér-Rao lower bound. For fixed weight setting, the weighted trimmed mean is still relatively more efficient in terms of asymptotic variance than the trimmed mean based on regression quantiles. From the parametric point of view, the computationally easy weighted trimmed mean is shown to be an efficient alternative to maximum likelihood estimation which is usually computationally difficult for most underlying distributions except the ideal case of normal ones. From the nonparametric point of view, this weighted trimmed mean is shown to be an efficient alternative robust estimator. A methodology for confidence ellipsoids and hypothesis testing based on the weighted trimmed mean is also introduced.

Key words and phrases: Initial estimator, symmetric quantile, weighted trimmed mean.

1. Introduction

Many nonadaptive robust estimators have been proposed for the estimation of location and linear regression parameters under the assumption of heavy tail error distribution. Some studies dealing with this topic include Ruppert and Carroll (1980), Koenker and Bassett (1978), Welsh (1987a, b), Frees (1991) and Koul and Mukherjee (1994). In terms of their asymptotic variances, Koenker (1982) and Ruppert and Carroll (1980) have shown that the usual robust estimators are very competitive. However, the list of Cramér-Rao (CR) lower bounds in Chen and Chiang (1996) showed that under a heavy tail distribution such as the contaminated normal distribution none of these estimators is really efficient when the contaminated variance is large. The problem of constructing asymptotically efficient estimators for location and regression parameters has also attracted considerable attention. Adaptive location R-, M- and L-estimators were treated by Beran (1974), Stone (1975) and Sacks (1975). Adaptive procedures extended to the linear model have been proposed by Bickel (1982), Manski (1984), Manski and

Hsieh (1987), Portnoy and Koenker (1989) and Welsh (1991). However, adaptive estimators that deal with the estimation of regression parameters must estimate a score function, which includes the derivative of the logarithm of an unknown density function, which make them computationally complicated. Moreover, unlike most nonadaptive estimators, the adaptive estimators cannot naturally be generalized to other statistical problems, especially when the Fisher information is not known. Chen and Chiang (1996), as a nonadaptive estimator, proposed a class of symmetric trimmed means constructed by a symmetric quantile and showed that in optimal trimming this estimator is more efficient than robust estimators such as ℓ_1-norm, Huber's M-estimator, and the trimmed mean based on regression quantiles in almost all cases of contaminated variances. Moreover, when the contaminated variance is large enough, the asymptotic variance of the symmetric trimmed mean may even be close to the CR lower bounds. However, the impressive property of the asymptotis variance of the symmetric trimmed mean being close to the CR lower bound does not occur when the contaminated variance is not large.

In light of the fact that the symmetric trimmed mean is inefficient when the contaminated variance is not large, a situation that occurs frequently in practice, our purpose in this paper is to obtain a class of estimators that reduces this inefficiency. Toward this end we propose a class of weighted trimmed means and show that it, indeed, reduces the asymptotic variance for all cases of contaminated variance. This class of weighted trimmed means is worth while applying to parametric and nonparametric estimation. With respect to parametric estimation, at optimal weight settings for observations of the dependent variable, the asymptotic variances of the weighted trimmed means are very close to the CR lower bounds. So this also provides a computationally easy alternative to maximum likelihood estimation whenever the latter as is often the case, is too complicated computationally. From the nonparametric point of view, a comparison of the weighted trimmed mean and the trimmed mean based on regression quantiles for a fixed setting of the weights shows that the weighted trimmed mean is relatively more efficient for contaminated normal distributions. An analysis of real data which obviously follows a model with asymmetric errors is also given for comparison of these two estimators. For statistical inference, we also sketch a large sample methodology for confidence ellipsoids and hypothesis testing based on the weighted trimmed mean.

In Section 2 we state several assumptions needed to develop a large sample representation of the weighted trimmed mean and introduce the weighted trimmed mean itself. In Section 3 we develop a large sample representation of the weighted trimmed mean based on the Bahadur representation. In Section 4 we compare the asymptotic variances of the weighted trimmed mean, the symmetric trimmed mean, and the trimmed mean based on regression quantiles for the case of optimal weight settings and also the case of fixed weights, associated with the CR lower bounds. In Section 5 we present a large sample methodology for confidence ellipsoids and hypothesis testing. Section 6 contains a real data analysis and Section 7 is an appendix containing the proofs of the theorems used in the paper.

2. Assumptions and the Construction of the Weighted Trimmed Mean

Consider the linear regression model

$$y_i = x_i'\beta + \epsilon_i, i = 1, \ldots, n, \tag{2.1}$$

where the y_i's are observations of the dependent random variable, the x_i's are design p-vectors, and the ϵ_i's are random errors that are independent and identically distributed (i.i.d.) with a distribution function F of zero mean and constant variance.

If x_i is a scalar of one then (2.1) turns out to be a location model; so the properties that we will develop in this paper for the weighted trimmed mean also hold for the location model. Our purpose is to estimate the parameter β. Like the symmetric mean, the weighted trimmed mean is constructed based on a symmetric quantile that depends on an initial estimator $\hat{\beta}_0$. We now list a set of assumptions about the design vectors, the distribution function, and the initial estimator that are needed for our results:

(A.1) $\max_{i,j} n^{-1/4}|x_{ij}| = O(1)$.

(A.2) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n x_i x_i' = Q$ and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n x_i = \theta$, where Q is a $p \times p$ positive definite matrix and θ is a finite p-vector.

(A.3) $n^{-1}\max_j \sum_{i=1}^n x_{ij}^4 = O(1)$.

(A.4) The distribution function F has a continuous density function f which is positive on the support of the random error variable.

(A.5) The derivative f' is bounded in a neighborhood of $b\tilde{F}^{-1}(\lambda)$ and $-b\tilde{F}^{-1}(\lambda)$ for $0 < \lambda < 1$ and $1 \le b < \infty$, where $\tilde{F}^{-1}(\lambda)$ satisfies $\lambda = P(-\tilde{F}^{-1}(\lambda) < \epsilon < \tilde{F}^{-1}(\lambda))$.

(A.6) $n^{1/2}(\hat{\beta}_0 - \beta) = O_p(1)$.

In the remainder of this paper we assume that conditions (A.1)-(A.6) are all satisfied.

The weighted trimmed mean will be constructed by means of the symmetric quantile, which is distinct from the regression quantile (see Koenker and Bassett (1978)) generalized from the ordinary quantile function. Let $0 < \lambda < 1$ and let $\hat{\beta}_0$ be an initial estimator of β. The symmetric quantile at percentage λ introduced by Chen and Chiang (1996) is defined as a pair $(x'\hat{\beta}_0 - \hat{a}(\lambda), x'\hat{\beta}_0 + \hat{a}(\lambda))$ for which $\hat{a}(\lambda)$ satisfies

$$\hat{a}(\lambda) = argmin_{a>0} \sum_{i=1}^n (|y_i - x_i'\hat{\beta}_0| - a)(\lambda - I(|y_i - x_i'\hat{\beta}_0| \le a)), \tag{2.2}$$

where $I(\cdot)$ represents the indicator function with values zero and one. Here $x'\hat{\beta}_0 - \hat{a}(\lambda)$ and $x'\hat{\beta}_0 + \hat{a}(\lambda)$ represent the lower and upper quantile bounds. The following lemma, which provides a representation for the term $\hat{a}(\lambda)$, was proved by Chen and Chiang (1996).

Lemma 2.1. *If* $0 < \lambda < 1$, *then*

$$n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) = (f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)))^{-1}\Big[n^{-1/2}\sum_{i=1}^{n}(\lambda - I(|\epsilon_i| \le \tilde{F}^{-1}(\lambda))) + (f(\tilde{F}^{-1}(\lambda)) - f(-\tilde{F}^{-1}(\lambda)))\theta' n^{1/2}(\hat{\beta}_0 - \beta)\Big] + o_p(1).$$

Based on the symmetric quantile function, we define the weighted trimmed mean as follows.

Definition 2.2. If $1 \le b < \infty,\ 1 < c < \infty$, and $\hat{a}(\lambda)$ is the solution of (2.2), the weighted trimmed mean is defined as

$$\hat{\beta}(\lambda, b, c) = \Big(\sum_{i=1}^{n} x_i x_i' J_i\Big)^{-1} \sum_{i=1}^{n} x_i[y_i J_i + \hat{a}(\lambda)\mathrm{sgn}(y_i - x_i'\hat{\beta}_0)L_i(\lambda) + (c-b)^{-1}(c\hat{a}(\lambda) - |y_i - x_i'\hat{\beta}_0|)\mathrm{sgn}(y_i - x_i'\hat{\beta}_0)K_i(\lambda)], \qquad (2.3)$$

where $J_i(\lambda) = I(|y_i - x_i'\hat{\beta}_0| \le \hat{a}(\lambda))$, $L_i(\lambda) = I(\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le b\hat{a}(\lambda))$, $K_i(\lambda) = I(b\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le c\hat{a}(\lambda))$, and $\mathrm{sgn}(\cdot)$ is the sign function with values -1 and $+1$.

Apart from pre-multiplication by the random matrix $(\sum_{i=1}^{n} x_i x_i' J_i(\lambda))^{-1} \sum_{i=1}^{n} x_i x_i'$, $\hat{\beta}(\lambda, b, c)$ is the least squares estimator calculated after replacing y_i by $y_i J_i(\lambda) + \hat{a}(\lambda)\,\mathrm{sgn}(y_i - x_i'\hat{\beta}_0)L_i(\lambda) + (c-b)^{-1}(c\hat{a}(\lambda) - |y_i - x_i'\hat{\beta}_0|)\mathrm{sgn}(y_i - x_i'\hat{\beta}_0)K_i(\lambda)$, which resembles a reweighted observation.

3. Bahadur Representation and Limiting Distribution

In this section we state the Bahadur representation of the weighted trimmed mean in terms of the initial estimator and give an exact form of the representation with the ℓ_1-norm as initial estimator. We use the ℓ_1-norm because the differences in the performance of the symmetric trimmed mean with various robust initial estimators is small enough to be negligible (see Chen and Chiang (1996)). So, for practical purposes, we can use any convenient robust initial estimator. We now state the main theorem for representation of $\hat{\beta}(\lambda, b, c)$.

Theorem 3.1. *The weighted trimmed mean has the following representation*

$$n^{1/2}(\hat{\beta}(\lambda, b, c) - \beta) = (\lambda Q)^{-1}\Big\{n^{-1/2}\sum_{i=1}^{n} x_i[\epsilon_i J_i^0(\lambda) + \tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i)L_i^0(\lambda)$$

$$+(c-b)^{-1}(c\tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i)-\epsilon_i)K_i^0(\lambda)]$$
$$+(c-b)^{-1}P(b\tilde{F}^{-1}(\lambda)<|\epsilon_i|<c\tilde{F}^{-1}(\lambda))n^{1/2}(\hat{\beta}_0-\beta)\Big\}$$
$$+o_p(1),$$

where $J_i^0(\lambda)=I(|\epsilon_i|\le\tilde{F}^{-1}(\lambda))$, $L_i^0(\lambda)=I(\tilde{F}^{-1}(\lambda)<|\epsilon_i|\le c\tilde{F}^{-1}(\lambda))$ *and* $K_i^0(\lambda) = I(b\tilde{F}^{-1}(\lambda)<|\epsilon_i|\le c\tilde{F}^{-1}(\lambda))$.

The second term indicates the contribution of the initial estimator to the weighted trimmed mean. Suppose that the initial estimator has the representation

$$n^{1/2}(\hat{\beta}_0-\beta)=Qn^{-1/2}\sum_{i=1}^{n}x_ih(\epsilon_i)+o_p(1) \tag{3.1}$$

for some function h. For example, $h(\epsilon)$ is ϵ if least squares provides the initial estimator, and it is $0.5f^{-1}(F^{-1}(0.5))\mathrm{sgn}(\epsilon)$ if the ℓ_1th norm is the initial estimator (see Ruppert and Carroll (1980)). Here we select the ℓ_1th norm as our initial estimator. We then have the following representation of $\hat{\beta}(\lambda,b,c)$.

Theorem 3.2. *Let $\hat{\beta}_0$ be the ℓ_1th norm estimator. We then have*

(a) $n^{1/2}(\hat{\beta}(\lambda,b,c)-(\beta+(\lambda Q)^{-1}\theta\delta))=(\lambda Q)^{-1}n^{-1/2}\sum_{i=1}^{n}x_i(\psi(\epsilon_i)-\delta)+o_p(1)$, *where*

$$\psi(\epsilon_i)=(\epsilon_i+h_{bc}\mathrm{sgn}(\epsilon_i>F^{-1}(0.5)))J_i^0(\lambda)+(\tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i)$$
$$+h_{bc}\mathrm{sgn}(\epsilon_i>F^{-1}(0.5)))L_i^0(\lambda)+((c-b)^{-1}(c\tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i)-\epsilon_i)$$
$$+h_{bc}\mathrm{sgn}(\epsilon_i>F^{-1}(0.5)))K_i^0+h_{bc}\mathrm{sgn}(\epsilon_i>F^{-1}(0.5))I(|\epsilon_i|>c\tilde{F}^{-1}(\lambda)),$$

and where $h_{bc}=0.5(c-b)^{-1}f^{-1}(F^{-1}(0.5))P(b\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda))$ *and*

$$\delta=\int_{-\tilde{F}^{-1}(\lambda)}^{\tilde{F}^{-1}(\lambda)}\epsilon dF+\tilde{F}^{-1}(\lambda)(P(\tilde{F}^{-1}(\lambda)<\epsilon<b\tilde{F}^{-1}(\lambda))$$
$$-P(-b\tilde{F}^{-1}(\lambda)<\epsilon<-\tilde{F}^{-1}(\lambda)))$$
$$+(c-b)^{-1}\Big[c\tilde{F}^{-1}(\lambda)(P(b\tilde{F}^{-1}(\lambda)<\epsilon<c\tilde{F}^{-1}(\lambda))$$
$$-P(-c\tilde{F}^{-1}(\lambda)<\epsilon<-b\tilde{F}^{-1}(\lambda)))-\int_{\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda)}\epsilon dF\Big].$$

(b) $n^{1/2}(\hat{\beta}(\lambda,b,c)-(\beta+(\lambda Q)^{-1}\theta\delta)) \;\rightarrow\; N(0,\ \sigma^2(\lambda)Q)$, *where* $\sigma^2(\lambda)=\lambda^{-2}(E((\psi(\epsilon))^2)-\delta^2)$ *and where*

$$E((\psi(\epsilon))^2)=h_{bc}^2+\int_{-\tilde{F}^{-1}(\lambda)}^{\tilde{F}^{-1}(\lambda)}\epsilon^2dF+(\tilde{F}^{-1}(\lambda))^2P(\tilde{F}^{-1}(\lambda)<|\epsilon|<b\tilde{F}^{-1}(\lambda))$$
$$+(c-b)^{-2}(c^2(\tilde{F}^{-1}(\lambda))^2P(b\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda))$$

$$+\int_{b\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda)}\epsilon^2 dF-2c\tilde{F}^{-1}(\lambda)\int_{b\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda)}|\epsilon|dF)$$
$$+2h_{bc}[\int_{-\tilde{F}^{-1}(\lambda)}^{\tilde{F}^{-1}(\lambda)}\epsilon\mathrm{sgn}(\epsilon>F^{-1}(0.5))dF$$
$$+\tilde{F}^{-1}(\lambda)\int_{\tilde{F}^{-1}(\lambda)<|\epsilon|<b\tilde{F}^{-1}(\lambda)}\mathrm{sgn}(\epsilon)\mathrm{sgn}(\epsilon>F^{-1}(0.5))dF$$
$$+(c-b)^{-1}\int_{b\tilde{F}^{-1}(\lambda)<|\epsilon|<c\tilde{F}^{-1}(\lambda)}(c\tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon)-\epsilon)\mathrm{sgn}(\epsilon>F^{-1}(0.5))].$$

Since $\psi(\epsilon_i)$ are i.i.d. with mean δ, the weighted trimmed mean has an asymptotic normal distribution with asymptotic bias $(\lambda Q)^{-1}\theta\delta$. Consider the special design that Ruppert and Carroll (1980) and Welsh (1987a) considered:

$$\sum_{i=1}^{n}x_{ij}=0, j=2,\ldots,p, \text{ and } x_{i1}=1,\ i=1,\ldots,n. \tag{3.2}$$

This yields $\theta=(1,0,\ldots,0)'$ and $Q=[Q_{ij}]_{i,j=1,2}$ which is partitioned such that $Q_{11}=1$, Q_{12} and Q_{21} are row and column $(p-1)$-vectors of zeros, and Q_{22} is a $(p-1)\times(p-1)$ positive definite matrix. Then the asymptotic bias of $\hat{\beta}(\lambda,b,c)$ is $\lambda^{-1}\delta(1,0,\ldots,0)'$. The bias of $\hat{\beta}(\lambda,b,c)$ for β involves only the intercept and not the slopes. Consider the general design in which F is symmetric at zero and assume that the ℓ_1-norm is the initial estimator. Then $\delta=0$ and the weighted trimmed mean is asymptotically unbiased.

Corollary 3.3. *Let $\hat{\beta}_0$ be the ℓ_1-norm estimator and $\lambda=1-2\alpha$, $0<\alpha<0.5$. We also assume that F is symmetric at zero. Then $n^{1/2}(\hat{\beta}(1-2\alpha,b,c)-\beta)\rightarrow N(0,\ \sigma^2(\alpha)Q^{-1})$ in distribution as $n\rightarrow\infty$, where*

$$\begin{aligned}\sigma^2(\alpha)=(1-2\alpha)^{-2}\Big\{&\int\epsilon^2(J_0(\alpha)+(c-b)^{-2}K_0(\alpha))dF\\&+(c-b)^{-1}\int|\epsilon|[\alpha_{bc}f^{-1}(0)J_0(\alpha)-2(c-b)^{-1}(cF^{-1}(1-\alpha)\\&+0.5\alpha_{bc}f^{-1}(0))K_0(\alpha)]dF+(1-2\alpha)(0.5(c-b)^{-1}\alpha_{bc}f^{-1}(0))^2\\&+(F^{-1}(1-\alpha)+0.5(c-b)^{-1}\alpha_{bc}f^{-1}(0))^2\alpha_b+(c-b)^{-2}[cF^{-1}(1-\alpha)\\&+0.5\alpha_{bc}f^{-1}(0)]^2\alpha_{bc}+(c-b)^{-2}(0.5f^{-1}(0)\alpha_{bc})^2\alpha_c\Big\};\end{aligned}$$

and where

$$\begin{aligned}J_0(\alpha)&=I(|\epsilon|\le F^{-1}(1-\alpha)),\ L_0(\alpha)=I(F^{-1}(1-\alpha)<|\epsilon|\le bF^{-1}(1-\alpha)),\\K_0(\alpha)&=I(bF^{-1}(1-\alpha)<|\epsilon|\le cF^{-1}(1-\alpha)),\\\alpha_b&=P(F^{-1}(1-\alpha)<|\epsilon|\le bF^{-1}(1-\alpha)),\\\alpha_{bc}&=P(bF^{-1}(1-\alpha)<|\epsilon|\le cF^{-1}(1-\alpha)),\ \text{and}\ \alpha_c=P(|\epsilon|>cF^{-1}(1-\alpha)).\end{aligned}$$

The efficiency of the class of weighted trimmed means is not clear at this point. We will display a table of the asymptotic variance of some robust estimators including $\hat{\beta}(\lambda, b, c)$ for comparison of their efficiencies.

4. The Efficiency of the Weighted Trimmed Mean

To verify the efficiency of the weighted trimmed mean, we list here a table of asymptotic variances of the weighted trimmed mean and the symmetric trimmed mean, both of which use the ℓ_1-norm as the initial estimator, and the trimmed mean based on regression quantiles, all in their optimal percentages, under the following contaminated normal distribution:

$$(1-\delta)N(0,\ 1)+\delta N(0,\ \sigma^2).$$

For the usual robust estimators, we select only the trimmed mean for comparison because with respect to optimal trimming, the trimmed mean is quite efficient.

Table 1. Comparison of asymptotic variances

2α	σ	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_{wst}$	CR	2α	σ	$\hat{\beta}_t$	$\hat{\beta}_{st}$	$\hat{\beta}_{wst}$	CR
0.1	1	1.000	1.000	1.000	1.000	0.2	15	1.9512	1.4345	1.4045	1.4030
	3	1.2959	1.3053	1.2660	1.2562		25	1.9886	1.3775	1.3565	1.3557
	5	1.3735	1.2879	1.2579	1.2528		50	2.0169	1.3255	1.3130	1.3127
	10	1.4318	1.2295	1.2107	1.2085		∞	2.0438	1.2562	1.2557	1.2557
	15	1.4511	1.2001	1.1854	1.1847	0.3	3	1.9642	2.0360	1.9182	1.8742
	25	1.4666	1.1719	1.1619	1.1616		5	2.2752	2.0134	1.9342	1.9104
	50	1.4782	1.1465	1.1409	1.1407		10	2.5349	1.8349	1.7774	1.7717
	∞	1.4892	1.1139	1.1137	1.1137		15	2.6272	1.7435	1.6934	1.6904
0.2	3	1.6001	1.6325	1.5574	1.5326		25	2.7032	1.6479	1.6117	1.6101
	5	1.7707	1.6050	1.5506	1.5381		50	2.7617	1.5595	1.5376	1.5369
	10	1.9051	1.4928	1.4544	1.4513		∞	2.8179	1.4397	1.4387	1.4387

Designation: $\hat{\beta}_t$=Trimmed Mean, $\hat{\beta}_{st}$=Symmetric trimmed mean and $\hat{\beta}_{wst}$=Weighted symmetric trimmed mean, also mention CR.

Here we list some conclusions that may be drawn from Table 1:

(a) The class of weighted trimmed means is more efficient than either the class of trimmed means or that of symmetric trimmed means. Moreover, the class of weighted trimmed means has asymptotic variances equal to the CR lower bounds when the contaminated variance is large enough, thus attaining the efficiencies of maximum likelihood estimation. This efficiency for the weighted trimmed mean is not surprising since it has more tuning constants to set and these are here chosen optimaly.

(b) Compared with the symmetric trimmed mean, the weighted trimmed mean produces a significant improvement in the asymptotic variances, especially in cases of small contaminated variances.
(c) As is the case with the symmetric trimmed mean, the asymptotic variance of the weighted trimmed mean decreases as the contaminated variances increases. This exceptional property does not hold for the usual class of estimators in their optimal settings, robust or nonrobust.

We continue here considering the contaminated normal distribution. However, the parameters of the estimators under consideration are fixed as constants. The trimmed means based on regression quantiles are set to have trimming percentage $2\alpha = 0.1,\ 0.2$ and 0.3 and the weighted trimmed means are set to have trimming percentage $2\alpha = 0.1$ and 0.15, and with constant $(b,\ c) = (1.2,\ 1.7)$. The following table gives the asymptotic variances of these two estimators corresponding to the above trimming percentages and weight settings.

Table 2. Comparison of asymptotic variances

δ	σ	$\hat{\beta}_t(0.1)$	$\hat{\beta}_t(0.2)$	$\hat{\beta}_t(0.3)$	$\hat{\beta}_{swt}(v_1)$	$\hat{\beta}_{swt}(v_2)$
0.1	3	1.29	1.34	1.44	1.14	1.24
	5	1.37	1.40	1.49	1.28	1.35
	10	1.45	1.44	1.53	1.26	1.31
0.2	3	1.62	1.61	1.70	1.63	1.58
	5	1.92	1.77	1.83	2.08	1.60
	10	2.33	1.91	1.94	4.60	1.77

Designation:(1) () in $\hat{\beta}_t(\ )$ is the trimming percentage and (, ,) in $\hat{\beta}_{swt}(\ ,\ ,\ )$ is the fixed vector of $2\alpha,\ b,\ c$. (2) $v_1 = (.1, 1.2, 1.7)$ and $v_2 = (.15, 1.2, 1.7)$

These results show that the weighted trimmed mean with parameter setting $(2\alpha,\ b,\ c) = (.15,\ 1.2,\ 1.7)$ is quite efficient for the contaminated normal distributions.

The performance of the symmetric weighted trimmed means shown in the tables above reveals its importance in application with respect to both parametric and nonparametric estimation. Because computation of maximum likelihood estimation for most underlying distributions, symmetric or asymmetric, except for the normal one, is complicated, this easy computation implies that the symmetric weighted trimmed mean is an efficient alternative for parametric estimation. For nonparametric estimation, the weighted trimmed mean also performs relatively more efficiently than the trimmed mean based on regression quantiles. An example analyzing data with outliers and obviously with asymmetric error distribution will be given in Section 6.

5. Large Sample Inference

Here we sketch a large sample methodology for confidence ellipsoids and hypothesis testing based on the weighted trimmed mean for the case of a symmetric distribution. To do this, we first need to estimate the asymptotic covariance matrix of $\hat{\beta}(1-2\alpha, b, c)$. For simplicity here we let $e_i = y_i - x_i'\hat{\beta}_0$ and $\hat{\alpha}_{st} = n^{-1}\sum_{i=1}^{n} I(s\hat{a}(1-2\alpha) < e_i < t\hat{a}(1-2\alpha))$. By assumption (A.2), we estimate Q by $n^{-1}\sum_{i=1}^{n} x_i x_i'$. Furthermore, let

$$\begin{aligned}\hat{\sigma}^2(\alpha, f^{-1}(0)) = (1-2\alpha)^{-2}\Big\{ & n^{-1}\sum_{i=1}^{n} e_i^2 (J_i(1-2\alpha) + (c-b)^{-2}K_i(1-2\alpha)) \\ & + (c-b)^{-1}n^{-1}\sum_{i=1}^{n}|e_i|[\hat{\alpha}_{bc}f^{-1}(0)J_i(1-2\alpha) \\ & + \hat{a}(1-2\alpha)\hat{\alpha}_{bc}f^{-1}(0)L_i(1-2\alpha) - 2(c-b)^{-1}(c\hat{a}(1-2\alpha) \\ & + 0.5\hat{\alpha}_{bc}f^{-1}(0))K_i(1-2\alpha)] \\ & + (0.5(c-b)^{-1}\hat{\alpha}_{bc}f^{-1}(0))^2 + (\hat{a}(1-2\alpha))^2\hat{\alpha}_{1b} \\ & + (c-b)^{-2}[(c\hat{a}(1-2\alpha))^2 + c\hat{a}(1-2\alpha)\hat{\alpha}_{bc}f^{-1}(0)]\hat{\alpha}_{bc}\Big\},\end{aligned}$$

where J_i, L_i, and K_i are defined in Definition 2.2.

Theorem 5.1. $\hat{\sigma}^2(\alpha,\ f^{-1}(0)) \to \sigma^2(\alpha)$ *in probability. To obtain an estimator of* $\sigma^2(\alpha)$, *we still need to obtain an estimator of the density* $f(0)$.

Lemma 5.5. *For* $h > 0$,

$$n^{-1}\sum_{i=1}^{n} I(-2^{-1}h < e_i < 2^{-1}h) \ \to\ F(h/2) - F(-h/2) \ \textit{in probability.}$$

Since F is differentiable on its support, we have $\lim_{h\to 0} h^{-1}(F(h/2) - F(-h/2)) = f(0)$. Let h_0 be some suitable choice of bandwith h. Useful rules for determining h_0 are stated in Scott (1992). Estimates of the density function f in the regression case were also introduced by Koenker and Portnoy (1987) and Welsh (1991). A reasonable estimator of density $f(0)$ is $\hat{f}(0) = (nh_0)^{-1}n^{-1}\sum_{i=1}^{n} I(-2^{-1}h_0 < e_i < 2^{-1}h_0)$. We then have the following estimator of $\sigma^2(\alpha) : S^2(\alpha) = \hat{\sigma}^2(\alpha, \hat{f}^{-1}(0))$. For $0 < u < 1$, let $F_u(r_1, r_2)$ denote the $(1-u)$ quantile of the F distribution, with r_1 and r_2 degrees of freedom, and let $d_u(r_1, r_2) = (1-2\alpha)^{-1}S^2(\alpha)r_1F_u(r_1, r_2)$. Suppose for some integer ℓ, K is a matrix of size $\ell \times p$, and K has rank ℓ. Let m be the number of residuals e_i lying outside the interval $(-c\hat{a}(1-2\alpha),\ c\hat{a}(1-2\alpha))$. Then the region of β

$$(\hat{\beta}(1-2\alpha, b, c) - \beta)'K'\Big[K(\sum_{i=1}^{n} x_i x_i' J_i(1-2\alpha))^{-1}K'\Big]^{-1}K(\hat{\beta}(1-2\alpha, b, c) - \beta)$$
$$\geq d_u(\ell, n-m-p)$$

has a probability of approximately u. If $K = I_p$, the confidence ellipsoid

$$(\hat{\beta}(1-2\alpha,b,c)-\beta)'\Big(\sum_{i=1}^{n} x_i x_i' J_i(1-2\alpha)\Big)(\hat{\beta}(1-2\alpha,b,c)-\beta) \le d_u(\ell, n-m-p)$$

for β has an asymptotic confidence coefficient of approximately $1-u$. Moreover, if we test $H_0: \ K\beta = v$ by rejecting H_0 whenever

$$(K\hat{\beta}(1-2\alpha,b,c)-v)'\Big[K(\sum_{i=1}^{n} x_i x_i' J_i(1-2\alpha))^{-1}K'\Big]^{-1}(K\hat{\beta}(1-2\alpha,b,c)-v) \ge d_u(\ell, n-m-p)$$

it has an asymptotic size of u.

6. Example

Let us look at an example of real data with outliers and asymmetric errors to compare the symmetric weighted trimmed mean with the trimmed mean based on regression quantiles. The example we now consider is a data set of international phone calls that appeared in the Belgian Statistcal Survey and can be seen in Rousseeuw and Leroy (1987). The plot of the phone calls (in tens of millions) is indexed by "A" in Figures 1 as indicated hereunder.

It seems to show an upward trend over years. However, the tendency contains heavy contamination from year 64 to 69 (1964 - 1969). For this data set, we first study the performance of the trimmed means. The trimmed mean has parameters of trimming percentages on two sides to be determined. We first set the equally trimming percentage as $\alpha = 0.05$ (0.95), 0.15 (0.85), 0.25 (0.75), 0.35 (0.65) and 0.45 (0.55). The predicted regression functions based on trimmed means associated with these α's and numbers of observations being trimmed are plotted in Figure 1. The trimmed mean with trimming percentage $\alpha = 0.35$ (0.65) is still not satifactory in explaining the main trend of the data. The reasons for this poor performance includes: (a) There is still one (65) in the 6 extreme observations (64-69) not being trimmed. (b) Too many observations representing the data of the main trend are removed. This impairs the estimation of the regression parameters. Although the trimmed mean with $\alpha = 0.45$ (0.55) performs much better than the others, it is still impaired by removing too many good observations (21 of 24 observations are removed).

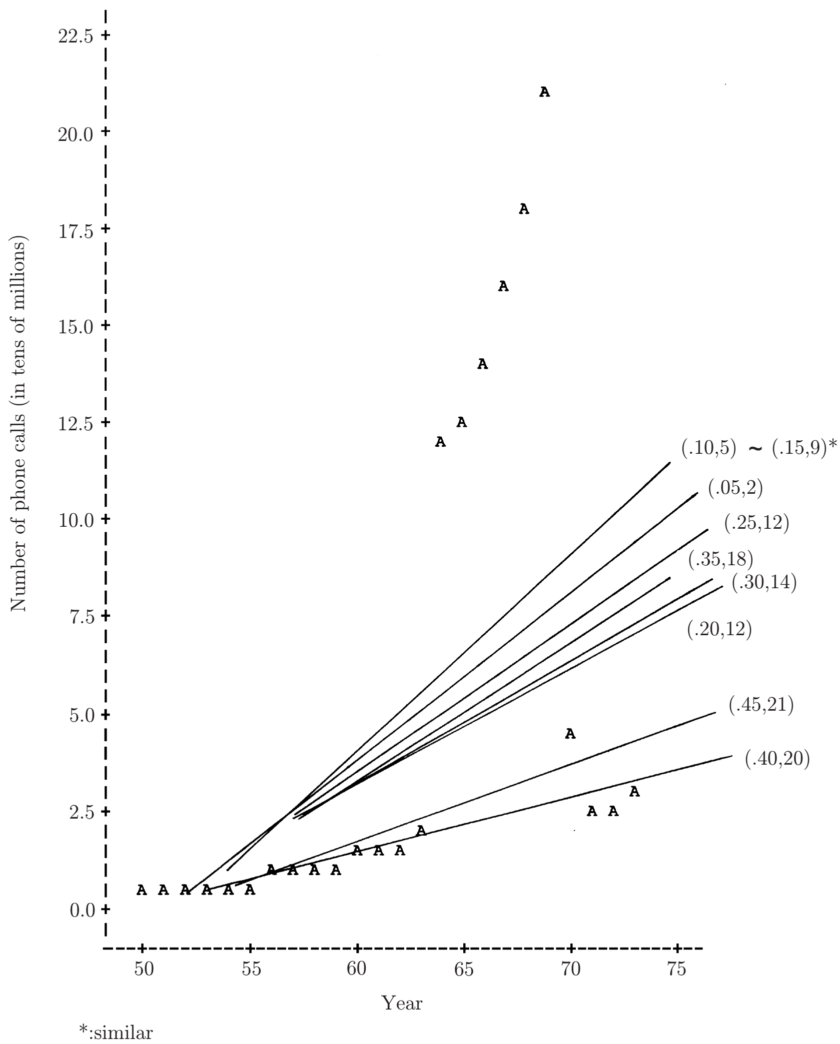

Figure 1. Sequential trimmed means based on regression quantiles

Unlike the trimmed mean that is suffering from removing good observations, the symmetric weighted trimmed mean would remove observations only if they have larger absolute residuals computed from an initial estimate. Let the ℓ_1-norm be the initial estimate. We also fix the tuning constant with $(b, c) = (1.2, 1.7)$ because it performed well in the estimation under errors of contaminated normal distributions. Suppose that the residuals from the ℓ_1-norm estimate ordered by their absolute values are $e_1, \ldots, e_n$. We then compute a sequence of symmetric weighted trimmed means by setting the sequential trimming percentage λ_i as $\lambda_i = |e_i|$. The following table gives the estimates associated with the years those corresponding observations are removed.

Table 4. Sequential weighted trimmed mean for international calls

Step	$\hat{\beta}_0$	$\hat{\beta}_1$	Years with obs removed
1	-27.16	.525	*None*
2	-25.53	.496	*None*
3	-22.57	.442	*None*
4	-17.86	.355	69
5	-17.38	.347	69
6	-6.348	.130	64, 65, 66, 67, 68, 69
7	-5.851	.121	64, 65, 66, 67, 68, 69
8	-4.976	.105	64, 65, 66, 67, 68, 69
9	-3.632	.081	64, 65, 66, 67, 68, 69
10	-5.545	.115	64, 65, 66, 67, 68, 69, 70, 71, 72, 73
11	-5.282	.111	64, 65, 66, 67, 68, 69, 70, 71, 72, 73
12	-4.969	.105	64, 65, 66, 67, 68, 69, 70, 71, 72, 73
13	-4.429	.095	64, 65, 66, 67, 68, 69, 70, 71, 72, 73
14	-4.484	.095	64, 65, 66, 67, 68, 69, 70, 71, 72, 73

Properties of the sequential weighted trimmed means based on this example include:

(a) The stability of the sequential weighted trimmed means strongly depends on whether extreme outliers remain in the data set for computing estimates. The sequence becomes unstable when outliers are not all removed (Step 1 to 5) and stable when outliers are all removed (Steps after 6).

(b) Removing larger outliers is possible without hurting the estimate by removing good observations. Also, when the residual quantile estimate $\hat{a}(\lambda)$ becomes large as it does for larger steps, the good observations are all retained.

For this data set that can be plotted on a plane, one might argue that a trimmed mean with asymmetric trimming by removing only the observations lying above the upper regression quantile may perform better. Indeed it turns out that the performance of the asymmetric trimmings is relatively better than the equally trimming ones. However, note that asymmetric trimming still suffers from removing some good observations. To see this, we list here the observations that lie above the upper qunatiles in the following table, where the years in $(\cdot)$ correspond to good observations.

Table 5. Observations lying above the regression quantile

α	Years with obs removed
0.95	69
0.85	(50, 51), 68, 69
0.75	(50, 51, 52), 67, 68, 69
0.65	(50, 51), 64, 66, 67, 68, 69
0.55	(50, 51, 52, 53), 64, 65, 66, 67, 68, 69, 70

We have two points to note on the asymmetric trimmed mean. First, the trimmed mean with $\alpha = 0.65$ is still not satisfactory to explain the main trend of the data because the extreme of year 65 has not been removed. It would perform better if $\alpha = 0.45$. However, it has suffered from removing good observations of years 50-53.

7. Appendix

To prove Theorem 3.1, we need several lemmas.

Lemma 7.1. *For a fixed real value k,*

$$n^{-1/2}\sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \leq k\hat{a}(\lambda)) = n^{-1/2}\sum_{i=1}^{n} x_i I(\epsilon_i \leq k\tilde{F}^{-1}(\lambda))$$
$$f(k\tilde{F}^{-1}(\lambda))[kn^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda))\theta + Qn^{1/2}(\hat{\beta}_0 - \beta)] + o_p(1).$$

Proof. From Lemma 2.1 of Jureckova (1984), we have

$$n^{-1/2}\sum_{i=1}^{n} x_i[I(\epsilon_i \leq k\tilde{F}^{-1}(\lambda) + n^{-1/2}(k, x_i')T_n) - I(\epsilon_i \leq k\tilde{F}^{-1}(\lambda))]$$
$$= f(k\tilde{F}^{-1}(\lambda))n^{-1/2}\sum_{i=1}^{n} x_i(k, x_i')T_n + o_p(1), \tag{7.1}$$

for any sequence of $(p + 1)$-dimensional random vector T_n which satisfies $T_n = O_p(1)$.

With equation (2.1), we have

$$y_i - x_i'\hat{\beta}_0 - k\hat{a}(\lambda) = \epsilon_i - [k\tilde{F}^{-1}(\lambda) + n^{-1/2}(k, x_i')n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda),\ (\hat{\beta}_0 - \beta)')']. \tag{7.2}$$

Also, Lemma 2.1 and assumption (A.6) imply that

$$n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda),\ (\hat{\beta}_0 - \beta)')' = O_p(1). \tag{7.3}$$

Replacing T_n by $n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda),\ (\hat{\beta}_0 - \beta)')'$ and substituting the relation (7.2) into (7.1) we have

$$n^{-1/2}\sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \leq k\hat{a}(\lambda)) = n^{-1/2}\sum_{i=1}^{n} x_i I(\epsilon_i \leq k\tilde{F}^{-1}(\lambda))$$
$$+f(k\tilde{F}^{-1}(\lambda))n^{-1/2}\sum_{i=1}^{n} x_i(k(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) + x_i'(\hat{\beta}_0 - \beta)) + o_p(1). \tag{7.4}$$

Lemma 7.1 follows from (7.4) and assumption (A.2).

Lemma 7.2. *For constant k,*

$$n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i I(y_i - x_i'\hat{\beta}_0 \le k\hat{a}(\lambda)) = n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i I(\epsilon_i \le k\tilde{F}^{-1}(\lambda))$$
$$+k\tilde{F}^{-1}(\lambda)f(k\tilde{F}^{-1}(\lambda))[kn^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda))\theta + Qn^{1/2}(\hat{\beta}_0 - \beta)] + o_p(1).$$

Proof. From Lemma 3.1 of Jureckova, we have

$$n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i[I(\epsilon_i \le k\tilde{F}^{-1}(\lambda) + n^{-1/2}(k, x_i')T_n) - I(\epsilon_i \le k\tilde{F}^{-1}(\lambda))]$$
$$= k\tilde{F}^{-1}(\lambda)f(k\tilde{F}^{-1}(\lambda))n^{-1/2}\sum_{i=1}^{n} x_i(k, x_i')T_n + o_p(1), \tag{7.5}$$

for any sequence of $(p+1)$-dimensional random vectors T_n which satisfies $T_n = O_p(1)$. Then the lemma follows from (7.2), (7.3), and (7.5).

Lemma 7.3. *For constant k,*

$$n^{-1}\sum_{i=1}^{n} x_i x_i' I(y_i - x_i'\hat{\beta}_0 \le k\hat{a}(\lambda)) = P(\epsilon \le k\tilde{F}^{-1}(\lambda))Q + o_p(1).$$

Proof. Lemma 3.2 of Jureckova implies that

$$n^{-1}\sum_{i=1}^{n} x_i x_i' I(\epsilon_i \le k\tilde{F}^{-1}(\lambda) + n^{-1/2}(k, x_i')T_n) = P(\epsilon \le k\tilde{F}^{-1}(\lambda))Q + o_p(1) \tag{7.6}$$

for any sequence T_n which satisfies $T_n = O_p(1)$. Then the lemma follows from (7.2), (7.3) and (7.6)

We now prove the main theorem.

Proof of Theorem 3.1. From (2.1) and (2.3), the weighted trimmed mean can be formulated as

$$n^{-1}\sum_{i=1}^{n} x_i x_i' I(|y_i - x_i'\hat{\beta}_0| \le \hat{a}(\lambda))(\hat{\beta}(\lambda, b, c) - \beta) = A + B + C,$$

where

$$A = \sum_{i=1}^{n} x_i\epsilon_i I(|y_i - x_i'\hat{\beta}_0| \le \hat{a}(\lambda)),$$
$$B = \hat{a}(\lambda)\sum_{i=1}^{n} x_i \operatorname{sgn}(y_i - x_i'\hat{\beta}_0) I(\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le b\hat{a}(\lambda)),$$
$$C = (c-b)^{-1}\sum_{i=1}^{n} x_i(c\hat{a}(\lambda) - |y_i - x_i'\hat{\beta}_0|)\operatorname{sgn}(y_i - x_i'\hat{\beta}_0) I(b\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le c\hat{a}(\lambda)).$$

Since A can be rewritten as

$$\sum_{i=1}^{n} x_i\epsilon_i I(y_i - x_i'\hat{\beta}_0 \le \hat{a}(\lambda)) - \sum_{i=1}^{n} x_i\epsilon_i I(y_i - x_i'\hat{\beta}_0 \le -\hat{a}(\lambda)),$$

Lemma 7.2 with $k = -1$ and $k = 1$ gives A as follows:

$$\begin{aligned} n^{-1/2}A = & \tilde{F}^{-1}(\lambda)(f(\tilde{F}^{-1}(\lambda)) - f(-\tilde{F}^{-1}(\lambda)))\theta n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) \\ & +\tilde{F}^{-1}(\lambda)(f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)))Qn^{1/2}(\hat{\beta}_0 - \beta) \\ & +n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i I(|\epsilon_i| \le \tilde{F}^{-1}(\lambda)) + o_p(1). \end{aligned}$$

Let $B_0 = \tilde{F}^{-1}(\lambda)(\hat{a}(\lambda))^{-1}B$, then B_0 can be expanded as

$$\begin{aligned} \tilde{F}^{-1}(\lambda) \Big[& \sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \le b\hat{a}(\lambda)) - \sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \le \hat{a}(\lambda)) \\ & -\sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \le -\hat{a}(\lambda)) + \sum_{i=1}^{n} x_i I(y_i - x_i'\hat{\beta}_0 \le -b\hat{a}(\lambda))\Big]. \end{aligned} \qquad (7.7)$$

Applying Lemma 7.1 to (7.7) with $k = 1, \ -1, \ b$, and $-b$, we have

$$\begin{aligned} n^{-1/2}B_0 = \tilde{F}^{-1}(\lambda)\{ & [bf(b\tilde{F}^{-1}(\lambda)) - f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)) \\ & -bf(-b\tilde{F}^{-1}(\lambda))]\theta n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) + [f(b\tilde{F}^{-1}(\lambda)) - f(\tilde{F}^{-1}(\lambda)) \\ & -f(-\tilde{F}^{-1}(\lambda)) + f(-b\tilde{F}^{-1}(\lambda))]Qn^{1/2}(\hat{\beta}_0 - \beta) \\ & +n^{-1/2}\sum_{i=1}^{n} x_i \mathrm{sgn}(\epsilon_i) I(\tilde{F}^{-1}(\lambda) < |\epsilon_i| \le b\tilde{F}^{-1}(\lambda))\} + o_p(1). \end{aligned}$$

Furthermore, Lemma 2.1 gives $n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) = O_p(1)$, We then have

$$n^{-1/2}B = n^{-1/2}B_0 + o_p(1).$$

Consider the representation of C. Decompose C as

$$C = C_1 + C_2 + C_3,$$

where

$$\begin{aligned} C_1 &= c(c-b)^{-1}\hat{a}(\lambda)\sum_{i=1}^{n} x_i I(b\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le c\hat{a}(\lambda)), \\ C_2 &= (c-b)^{-1}\sum_{i=1}^{n} x_i x_i' I(b\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le c\hat{a}(\lambda))(\hat{\beta}_0 - \beta), \\ C_3 &= -(c-b)^{-1}\sum_{i=1}^{n} x_i\epsilon_i I(b\hat{a}(\lambda) < |y_i - x_i'\hat{\beta}_0| \le c\hat{a}(\lambda)). \end{aligned}$$

Using an argument similar to the above, we have

$$\begin{aligned}n^{-1/2}C_1 = c(c-b)^{-1}\tilde{F}^{-1}(\lambda)\{[cf(c\tilde{F}^{-1}(\lambda)) - bf(b\tilde{F}^{-1}(\lambda)) + bf(-b\tilde{F}^{-1}(\lambda))\\ -cf(-c\tilde{F}^{-1}(\lambda))]\theta n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) + [f(c\tilde{F}^{-1}(\lambda)) - f(b\tilde{F}^{-1}(\lambda))\\ -f(-b\tilde{F}^{-1}(\lambda)) + f(-c\tilde{F}^{-1}(\lambda))]Qn^{1/2}(\hat{\beta}_0 - \beta)\\ +n^{-1/2}\sum_{i=1}^{n} x_i \mathrm{sgn}(\epsilon_i) I(b\tilde{F}^{-1}(\lambda) < |\epsilon_i| \le c\tilde{F}^{-1}(\lambda))\} + o_p.\end{aligned}$$

It follows on applying Lemma 7.3 with $k = b,\ -b,\ c$ and $-c$ to C_2 that

$$n^{-1/2}C_2 = (c-b)^{-1}\lambda_{bc}Qn^{1/2}(\hat{\beta}_0 - \beta) + o_p(1).$$

Again, using Lemma 7.2 on C_3, we get

$$\begin{aligned}n^{-1/2}C_3 = -(c-b)^{-1}\{[c^2\tilde{F}^{-1}(\lambda)(f(c\tilde{F}^{-1}(\lambda)) - f(-c\tilde{F}^{-1}(\lambda))) - b^2\tilde{F}^{-1}(\lambda)(f(b\tilde{F}^{-1}(\lambda))\\ -f(-b\tilde{F}^{-1}(\lambda)))]\theta n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) + [c\tilde{F}^{-1}(\lambda)(f(c\tilde{F}^{-1}(\lambda))\\ +f(-c\tilde{F}^{-1}(\lambda))) - b\tilde{F}^{-1}(\lambda)(f(b\tilde{F}^{-1}(\lambda)) + f(-b\tilde{F}^{-1}(\lambda)))]Qn^{1/2}(\hat{\beta}_0 - \beta)\\ +n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i I(b\tilde{F}^{-1}(\lambda) < |\epsilon_i| \le c\tilde{F}^{-1}(\lambda))\} + o_p(1).\end{aligned}$$

Take the sum $n^{-1/2}(C_1 + C_2 + C_3) + n^{-1/2}A + n^{-1/2}B$ and make some simple rearrangements. We then have

$$\begin{aligned}&n^{-1/2}(A+B+C)\\ &= (c-b)^{-1}\lambda_{bc}Qn^{1/2}(\hat{\beta}_0 - \beta) + n^{-1/2}\sum_{i=1}^{n} x_i\\ &\Big[\epsilon_i I(|\epsilon_i| \le \tilde{F}^{-1}(\lambda)) + \tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i) I(\tilde{F}^{-1}(\lambda) < |\epsilon_i| \le\ b\tilde{F}^{-1}(\lambda))\\ &+(c-b)^{-1}(c\tilde{F}^{-1}(\lambda)\mathrm{sgn}(\epsilon_i) - \epsilon_i) I(b\tilde{F}^{-1}(\lambda) < |\epsilon_i| \le c\tilde{F}^{-1}(\lambda))\Big] + o_p(1). \quad (7.8)\end{aligned}$$

The theorem then follows from (7.8) and the following result from Lemma 7.3:

$$n^{-1}\sum_{i=1}^{n} x_i x' i I(|y_i - x_i'\hat{\beta}_0| \le \hat{a}(\lambda)) = \lambda Q + o_p(1).$$

Now we turn to proving Theorem 5.1. Theorem 5.1 follows from the following lemma.

Lemma 7.4. (a) *Let* $-\infty < s < t < \infty$. *Then*

$$n^{-1}\sum_{i=1}^{n} I(s\hat{a}(1-2\alpha) < y_i - x_i'\hat{\beta}_0 < t\hat{a}(1-2\alpha)) \to P(sF^{-1}(1-\alpha) < \epsilon < tF^{-1}(1-\alpha))$$

in probability.
(b) *Let* $0 \le s < t < \infty$. *Then*

$$n^{-1}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta}_0)^c I(s\hat{a}(1-2\alpha) < y_i - x_i'\hat{\beta}_0 < t\hat{a}(1-2\alpha))$$
$$\to \int \epsilon^c I(sF^{-1}(1-\alpha) < \epsilon < tF^{-1}(1-\alpha))dF$$

in probability for $c = 1$ *and* 2.

Proof. We consider only case (b) for $c = 1$. The proofs of the other cases are similar. Replacing y_i by $x_i'\beta + \epsilon_i$, the equation in (b) with $c = 1$ is

$$n^{-1}\sum_{i=1}^{n}\epsilon_i I(s\hat{a}(1-2\alpha) < y_i - x_i'\hat{\beta}_0 < t\hat{a}(1-2\alpha)) - n^{-1}\sum_{i=1}^{n}x_i I(s\hat{a}(1-2\alpha) < y_i - x_i'\hat{\beta}_0 < t\hat{a}(1-2\alpha))(\hat{\beta}_0 - \beta). \tag{7.9}$$

Lemma 7.1 implies that

$$n^{-1}\Big[\sum_{i=1}^{n}x_i I(s\hat{a}(1-2\alpha) < y_i - x_i'\hat{\beta}_0 < t\hat{a}(1-2\alpha)) - \sum_{i=1}^{n}x_i I(sF^{-1}(1-2\alpha) < \epsilon_i < tF^{-1}(1-2\alpha))\Big] = O_p(n^{-1/2}).$$

With addition of assumption (A.6), the second term of (7.9) is $O_p(n^{-1/2})$. Moreover, Lemma 7.2 implies that the first term of (7.9) is asymptotically equal to

$$n^{-1}\sum_{i=1}^{n}\epsilon_i I(sF^{-1}(1-\alpha) < \epsilon_i < tF^{-1}(1-\alpha)) = O_p(n^{-1/2}).$$

These results then imply that (b) with $c = 1$ holds.

Acknowledgements

I am grateful to the referees for their most useful comments that have greately improved the presentation and the quality of this manuscript.

References

Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2**, 63-74.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647-671.

Chen, L. A. and Chiang, Y. C. (1996). Symmetric type quantile and trimmed means for location and linear regression model. *J. Nonparametr. Statist.* **7**, 171-185.

Frees, E. W. (1991). Trimmed slope estimates for simple linear regression. *J. Statist. Plann. Inference* **27**, 203-221.

Jureckova, J. (1984). Regression quantile and trimmed least square estimator under general design. *Kybernetika* **20**, 345-357.

Jureckova, J. and Sen, P. K. (1984). On adaptive scale-equivariant M-estimators in linear model. *Statist. Decisions Supplemental Issue* **1**, 31-46.

Koenker, R. and Bassett, G. J. (1978). Regression quantile. *Econometrica* **46**, 33-50.

Koenker, R. and Portnoy, S. (1987). *L*-estimators for linear models. *J. Amer. Statist. Assoc.* **82**, 851-857.

Koenker, R. (1982). Robust method in econometrics. *Econometric Rev.* **1**, 213-255.

Koul, H. L. and Mukherjee, K. (1994). Regression quantiles and related processes under long range dependent errors. *J. Multivariate Anal.* **51**, 318-337.

Manski, C. F. (1984). Adaptive estimation of nonlinear regression models. *Econometric Rev.* **3**, 145-194.

Manski, C. F. and Hsieh, D. A. (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Ann. Statist.* **15**, 541-551.

Portnoy, S. and Koenker, R. (1989). Adaptive *L*-estimation for linear models. *Ann. Statist.* **17**, 362-381.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley.

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

Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3**, 285-298.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualization.* Wiley.

Stone, C. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3**, 267-284.

Welsh, A. H. (1987a). The trimmed mean in the linear model. *Ann. Statist.* **15**, 20-36.

Welsh, A. H. (1987b). One-step *L*-estimators for the linear model. *Ann. Statist.* **15**, 626-641.

Welsh, A. H. (1991). Asymptotically efficient adaptive *L*-estimators in linear model. *Statist. Sinica* **1**, 203-228.

Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan.

E-mail: lachen@cc.nctu.edu.tw

(Received February 1994; accepted June 1996)