# A Comparison of Robust Estimators Based on Two Types of Trimming

SUBHRA SANKAR DHAR[1,*] and PROBAL CHAUDHURI[1,**]

[1] *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata-700108, India. E-mail:* *subhra_r@isical.ac.in; **probal@isical.ac.in

The least trimmed squares (LTS) estimator and the trimmed mean are two well known trimming based estimators. Both estimates are popular in practice, and they are implemented in standard statistical softwares. In this paper, we compare the asymptotic variances of these two estimators in a location model, when the two location estimates have the highest breakdown point (i.e., 50%). Some interesting results concerning the asymptotic variances of the two types of estimates are obtained for distributions with exponential and polynomial tails. We extend this comparison into regression problems. Some simulation studies are performed to compare the finite sample standard errors of these estimators in location and regression models. Two examples involving real data sets are presented to compare the bootstrap standard errors of the LTS estimator and the analogue of the trimmed mean in regression problems.

*Keywords:* least trimmed squares, trimmed mean, asymptotic normality, location model, regression model, median, least absolute deviations.

## 1. Introduction

Consider the location model with observations $y_i$'s satisfying

$$y_i = \theta + e_i,\ i = 1, \cdots, n,$$

where $\theta$ is the parameter to be estimated, and the $e_i$'s are i.i.d. with a common distribution function $F$. In this location model, the trimmed mean (or the $\alpha$-trimmed mean to be more precise) is given by

$$\frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} y_{(i)},$$

where $y_{(i)}$ is the $i$ th order statistic, and $0 < \alpha < 1/2$ (see e.g., Serfling (1980), Lehmann (1983)). For $\alpha = 1/2$, the $\alpha$-trimmed mean coincides with the sample median. Rousseeuw(1984)introduced the least trimmed squares (LTS) estimator in the location model based on an alternative trimming procedure, and it is defined as

$$\hat{\theta}_{LTS} = arg \min_{\theta} \sum_{i=1}^{h} (y_i - \theta)^2_{(i)}.$$

This estimator is popularly known as $\lambda$-LTS estimator, where $\lambda = h/n$.

Both of the $\lambda$-LTS and the $\alpha$-trimmed mean estimators can be extended to data $(y_1, \mathbf{x_1})$, $(y_2, \mathbf{x_2})$, $\cdots$, $(y_n, \mathbf{x_n})$ satisfying the regression model

$$y_i = h(\mathbf{x_i}, \theta) + e_i, \; i = 1, \cdots, n,$$

where $y_i$ denotes the dependent variable, $h(\mathbf{x_i}, \theta)$ is the regression function, and $\theta \in R^{p+1}$ is the parameter to be estimated. Here $\mathbf{x_i} = (x_{i1}, x_{i2}, \cdots, x_{id}) \in R^d$ represents the vector of explanatory variables, and the $e_i$'s are i.i.d. with a common distribution function $F$. Koenker and Portnoy (1987), Ruppert and Carroll (1980), Welsh (1987) and Chen et al. (2003) proposed different regression analogues of the trimmed mean, while similar extensions of the $\lambda$-LTS estimator are discussed in Rousseeuw and Leroy (1987), Cizek (2005) and Jung (2005).

In the location model, the asymptotic breakdown point of the $\alpha$-trimmed mean is known to be $\alpha$ (See e.g., Huber(1981, p. 13)). This breakdown point increases with the amount of trimming, and the highest breakdown point 50% is achieved with the maximum amount of trimming, when the trimmed mean coincides with the sample median. Similarly, the asymptotic breakdown point of the $\lambda$-LTS estimator is the trimming proportion $= (1-\lambda)$(see Rousseeuw and Leroy(1987, p. 134)). Here also the highest breakdown point 50% is achieved when $h = [n]/2 + 1$.

The LTS estimator and the trimmed mean are implemented in many statistical softwares like S-PLUS, R, Matlab etc. The computation of the $\lambda$-LTS estimator in a location model is described in Rousseeuw and Leroy (1987, p. 171). In the regression model, the $\lambda$-LTS estimator can be computed by the sub-sampling algorithm given by Rousseeuw and Leroy(1987), but for a large data set, this procedure is very time consuming. There is an iterative procedure given in Rousseeuw and Drissen(2006), which is less time consuming. The statistical package S-PLUS computes the $\lambda$-LTS estimator using the genetic algorithm (Burns (1992)) in regression problems.

In the location model, the asymptotic variance of the $\lambda$-LTS estimator has been derived by Rousseeuw and Leroy(1987, p. 180), and that for the $\alpha$-trimmed mean is given in Lehmann(1983, p. 361) and Serfling(1980, p: 236–237). In this paper, we compare these two asymptotic variances and derive conditions under which one of the two estimators is asymptotically more efficient than the other. It is also shown that this comparison can be extended into regression models. We also report results from some numerical studies that compare finite sample variances of the estimates obtained using these two trimming procedures in location as well as regression models.

## 2. Main Results

**Theorem 1:** *Let $y_1$, $y_2$, $\cdots$, $y_n$ be $n$ i.i.d. observations with the common distribution function $F(y - \theta)$ having the density $f(y - \theta)$. We assume $f$ to be a symmetric around zero and continuous density function that is positive on an open interval containing the interval $[F^{-1}(1/4), F^{-1}(3/4)]$. Suppose that the asymptotic variance of the $\lambda$-LTS* *estimator($\lambda \le 1/2$) and the $\alpha$-trimmed mean ($\alpha \le 1/2$) based on these $n$ observations are denoted by $\sigma_1^2(\lambda)/n$ and $\sigma_2^2(\alpha)/n$ respectively. Then, for the two estimates, each with the highest breakdown point 50% obtained using two different types of trimming, we have $\sigma_1^2(1/2) \ge \sigma_2^2(1/2)$ if and only if*

$$2f^2(0) \int_0^{F^{-1}(3/4)} t^2 f(t)dt \ge \left[1/4 - F^{-1}(3/4) f\{F^{-1}(3/4)\}\right]^2 . \tag{1}$$

Note that in the case of strict inequality in (1), by a continuity argument, there exists $\alpha_0 \in [0, 1/2)$ such that the $\sigma_1^2(\lambda) > \sigma_2^2(\alpha)$ whenever $(1-\lambda) = \alpha > \alpha_0$. An analogous result holds in the case of strict reverse inequality in (1).

Let us now consider two examples such that inequality (1) holds in one example, and it fails to hold in the other. In both the examples, the density $f$ is unimodal, continuous and symmetric around zero.

**Example 1:** Consider the standard Cauchy density function

$$f(y) = \frac{1}{\pi(1+y^2)}, -\infty \le y \le \infty.$$

Here, the (3/4)th quantile is 1, and

$$\int_0^{F^{-1}(3/4)} t^2 f(t)dt = 1/\pi - 1/4.$$

So,

$$\begin{aligned} 2f^2(0) \int_0^{F^{-1}(3/4)} t^2 f(t)dt &= 2/\pi^2(1/\pi - 1/4) \\ &> \{(\pi-2)/4\pi\}^2 \\ &= \left[1/4 - F^{-1}(3/4) f\{F^{-1}(3/4)\}\right]^2 . \end{aligned}$$

This implies that inequality (1) holds, and the median (i.e., the 1/2-trimmed mean) is asymptotically more efficient than the 1/2-LTS estimator for this standard Cauchy density function.

**Example 2:** Consider next the density function

$$\begin{aligned} f(y) &= \tfrac{\epsilon - c}{a}|y| + c, && |y| \le a \\ &= \epsilon, && a \le |y| \le b \\ &= 0, && \text{otherwise.} \end{aligned}$$

We assume that the (3/4)th quantile is $a$. Then, from the definition of the (3/4)th quantile, we can write

$$(1/2)a(c-\epsilon)+a\epsilon=1/4 \Leftrightarrow a(c+\epsilon)=1/2. \quad (2)$$

As the (3/4)th quantile is $a$, and $f$ is a symmetric density function around zero, we have

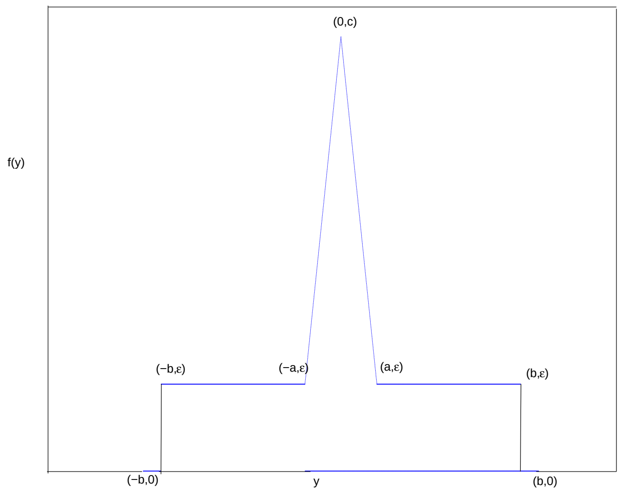
$$(b-a)\epsilon=1/4. \quad (3)$$

The strict reverse inequality of (1) holds if and only if

$$[1/4-a\epsilon]^2 > 2c^2\left[ca^3/12+\epsilon a^3/4\right]. \quad (4)$$

Many choices of $a$, $b$, $c$ and $\epsilon$ will satisfy (2), (3) and (4). For example, one can choose $\epsilon=10^{-5}$, $a=1/2(1+10^{-5})$, $b=10^5/4+a$ and $c=1$. It can be easily shown that these choices for $a$, $b$, $c$ and $\epsilon$ are satisfying (2) and (3). Further, for those choices of $a$, $b$, $c$ and $\epsilon$, $[1/4-a\epsilon]^2=(99999/400004)^2>1/17$, and $2c^2[ca^3/12+\epsilon a^3/4]=1/16(1+10^{-5})^3[1/3+10^{-5}]<1/40$ so that (4) is satisfied. Consequently, for such choices of $a$, $b$, $c$ and $\epsilon$, median is asymptotically less efficient than 1/2-LTS estimator.

**Figure 1:** Graph of $f(y)$ considered in Example 2




The following theorem gives a sufficient condition for inequality (1) to hold.

**Theorem 2:** *If $f$ is a unimodal, continuous and positive density function, which is symmetric around zero, the inequality (1) holds if* $F^{-1}(3/4)f\{F^{-1}(3/4)\}\geq 1/6$.

For the standard Cauchy density function, $F^{-1}(3/4)=1$ and $f\{F^{-1}(3/4)\}=1/2\pi$, i.e., $F^{-1}(3/4)f\{F^{-1}(3/4)\}=1/2\pi<1/6$. This implies that the above sufficient condition is not satisfied. But Example 1 already shows that the the median is asymptotically more efficient than the 1/2-LTS estimator for the standard Cauchy density function. This demonstrates that the condition in Theorem 2 is only sufficient but not necessary for inequality (1) (or equivalently the inequality $\sigma_1^2(1/2)\geq\sigma_2^2(1/2)$) to hold.

## 2.1. Some results for error distributions with exponential and polynomial tails

Since the asymptotic variance of an estimate of a location parameter greatly depends on the behavior of the tail of the error distribution, we consider some examples of error distributions with exponential and polynomial tails. Three very well known standard distributions, namely, double exponential, normal and Cauchy distributions are covered by these two families of distributions. Using Theorem 2 above, we now derive some results related to inequality (1) for different density functions with exponential and polynomial tails.

**Result 3:** *Consider the exponential power family consisting of densities of the form*

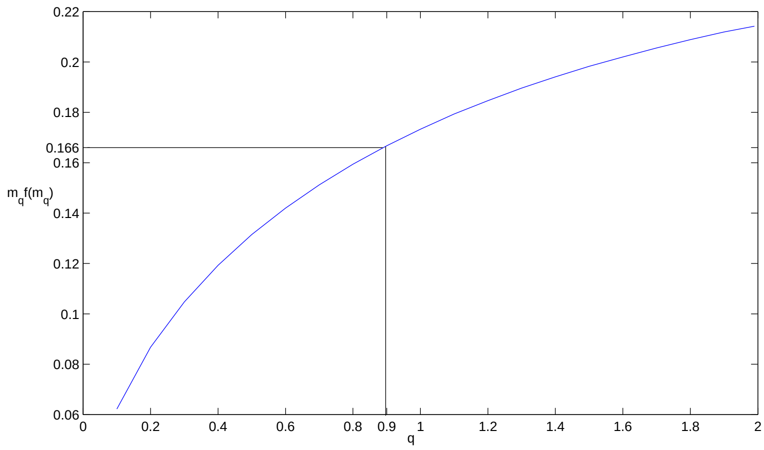
$$f(y) = \frac{1}{2\Gamma(1+q^{-1})} e^{-|y|^q}, -\infty \leq y \leq \infty, q \geq 0.$$

*For such densities, it can be proved that the sufficient condition of Theorem 2 is satisfied for $q \geq 2$ (see the proof in the Appendix).*

Note that $q = 1$ corresponds to the standard double exponential density. The (3/4)th quantile of the standard double exponential density is $\ln 2$. So, $f\{F^{-1}(3/4)\} = 1/4$, and $F^{-1}(3/4)f\{F^{-1}(3/4)\} = 1/4 \times \ln 2 > 1/6$. In other words, the sufficient condition in Theorem 2 is satisfied for the standard double exponential density function.

We have studied the case $q < 2$ numerically, and the graph plotting $m_q f(m_q)$ against different values of $q < 2$, where $m_q$ is the (3/4)-th quantile of the above exponential power density, is presented in Figure 2. This graph indicates that when $q$ lies between .9(approximately) and 2, the sufficient condition in Theorem 2 holds. In other words, for those values of $q$, $\sigma_2^2(1/2) < \sigma_1^2(1/2)$.

**Figure 2:** Graph of $m_q f(m_q)$



**Result 4:** *Consider the polynomial tail family consisting of densities of the form*

$$f(y) = \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+y^2)^k}, -\infty \leq y \leq \infty, k \geq 0.$$

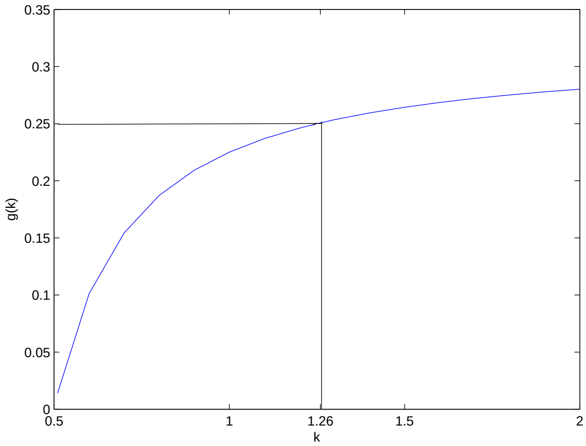
*For any density in this family, the sufficient condition of Theorem 2 is satisfied whenever*

$$g(k) := \int_0^{\tan^{-1}(t_k)} \frac{\Gamma(k)}{2^{k-1}\Gamma(1/2)\Gamma(k-1/2)} (1+\cos 2\theta)^{k-1} d\theta \geq 1/4,$$

*where* $t_k = \sqrt{(3/2)^{1/k} - 1}$ *(see the proof in the Appendix).*

Now, we numerically study $g(k)$ and try to determine those values of $k$ for which $g(k) \geq 1/4$. Figure 3 presents the graph of $g(k)$ plotted against different values of $k > 0.5$. It is indicated by the graph that $g(k) \geq 1/4$ when $k > 1.26$ (approximately). Consequently, for those values of $k$, $\sigma_2^2(1/2) < \sigma_1^2(1/2)$.

**Figure 3:** Graph of $g(k)$




## 3. Extension into Regression Problems

Let us now recall the regression model introduced earlier:

$$y_i = h(\mathbf{x_i}, \theta) + e_i, \; i = 1, \cdots, n.$$

The vector $\theta = (\beta_0, \beta_1, \ldots, \beta_p)$ of unknown parameters is assumed to belong to a compact parameter space $\Theta \subseteq R^{p+1}$. Further assumptions on the regression function and $F$ will be stated and discussed later.

Ruppert and Carroll (1980) considered two methods for defining a regression analogue of the $\alpha$-trimmed mean (see also Koenker and Portnoy (1987)). The first one was originally suggested by Koenker and Bassett(1978) and uses their concept of regression quantiles, and its computation is non iterative. However, the computation of the second one is iterative, which is based on using the residuals from a preliminary estimator as in Bickel(1973). Welsh (1987) proposed another analogue of the $\alpha$-trimmed mean using the von Mises functional approach, and its computation is also iterative. When the trimming proportion $\alpha$ equals 1/2, all of the above three regression analogues of the $\alpha$-trimmed mean reduce to the median regression or the least absolute deviations (LAD) estimator $\hat{\theta}_{LAD}$ (say), which minimizes the sum of the absolute deviations. These three regression analogues of the $\alpha$-trimmed mean have the same asymptotic dispersion under appropriate conditions. Chen et al. (2003) proposed a nonlinear regression version of the $\alpha$-trimmed mean, and they used Welsh's approach. They derived the large sample properties of their proposed estimates under some appropriate assumptions considering a fixed design model. One can derive the same asymptotic normality results in a random design model for this estimate when the required conditions hold *almost surely* in the $\mathbf{x_i}$'s.

On the other hand, Rousseeuw and Leroy (1987, p. 132) defined the $\lambda$-LTS estimator in the regression model. Also, Jung (2005) extended the $\lambda$-LTS estimator into the multivariate regression model. Cizek (2005) extended the $\lambda$-LTS estimator into the nonlinear regression model, which is defined by

$$\hat{\theta}_{LTS} = arg \min_{\theta \in R^p} \sum_{i=1}^{h} \{r^2_{(i)}(\theta)\},$$

where the residual $r_i(\theta)$ is defined by

$$r_i(\theta) = y_i - h(\mathbf{x_i}, \theta).$$

We now state some conditions for deriving the asymptotic efficiency of the regression analogue of the $\lambda$-LTS estimator considered by Cizek (2005)with respect to that of the $\alpha$-trimmed mean considered by Chen et al. (2003) in the special case $\lambda = \alpha = 1/2$. We begin with the distributional conditions on the $\mathbf{x_i}$ and the $e_i$. We assume that the $\mathbf{x_i}$'s are i.i.d. having finite second moment, mutually independent with the $e_i$'s. We assume the conditions on the error distribution function $F$ proposed by Cizek(2005, See, p. 3971) and the finiteness of the fourth moment of $F$.

Next, we state the conditions on the regression function. We suppose that the regression function $h(\mathbf{x_i}, \theta)$ has finite second moment, and it is thrice continuously differentiable with respect to $\theta$. The function and its derivatives are assumed to be uniformly bounded in $\mathbf{x_i}$. Further, the first order and the second order derivatives of $h(\mathbf{x_i}, \theta)$ are assumed to have finite second and fourth moments respectively. Also, we suppose that $\sup_{\theta \in \Theta} |h(\mathbf{x_i}, \theta)|$ and $\sup_{\phi : |\phi - \theta| < \delta} |\nabla_\phi h(\mathbf{x_i}, \phi)|$ have finite $r_\beta$th moment ($r_\beta \geq 2$) for some $\delta > 0$. $W_h = E[\{\nabla_\theta h(\mathbf{x_i}, \theta)\}\{\nabla_\theta h(\mathbf{x_i}, \theta)\}^T]$ is assumed to be a nonsingular positive definite matrix.

We assume that $\Theta$ is a compact space. For computation of iterative estimates of the regression analogue of the trimmed mean considered by Chen et al. (2003), we assume the existence of an initial estimator $\theta_n$ such that $n^{1/2}(\theta_n - \theta)$ is bounded in probability.

We assume an identifiability condition used by Cizek(2005), which is stated as follows. For any $\epsilon > 0$ and $\{\phi : \phi \in \Theta, |\phi - \theta| \geq \epsilon\}$ compact, there exists $\alpha(\epsilon) > 0$ such that

$$\min_{\phi:|\phi-\theta|\geq\epsilon} E_\theta[r_i^2(\phi)I\{r_i^2(\phi) \leq G_\phi^{-1}(\lambda)\}] - E_\theta[r_i^2(\theta)I\{r_i^2(\theta) \leq G_\theta^{-1}(\lambda)\}] \geq \alpha(\epsilon).$$

We now state a corollary that follows from Theorem 1.

**Corollary 5:** *Let the error distribution function $F$ with the density function $f$, the explanatory variable $\mathbf{x_i}$ and the regression function $h(\mathbf{x_i}, \theta)$ satisfy the preceding conditions. Suppose that the asymptotic dispersions of the $\lambda$-LTS estimator considered by Cizek (2005) and the $\alpha$-trimmed mean estimator (given by Chen et al. (2003)) for the vector $\theta$ based on $n$ observations are denoted by $n^{-1}\Sigma_1(\lambda)$ and $n^{-1}\Sigma_2(\alpha)$ respectively. Then, for $\lambda = \alpha = 1/2$, we have*

$$\Sigma_2(1/2) \leq_L \Sigma_1(1/2) \tag{5}$$

*(where $\leq_L$ denotes the Lowner ordering i.e., $\Sigma_1(1/2) - \Sigma_2(1/2)$ is non negative definite) if and only if inequality (1) in Section 2 holds.*

It is straight-forward to verify that the sufficient condition in Theorem 2 namely, $F^{-1}(3/4)f\{F^{-1}(3/4)\} \geq 1/6$ implies that inequality (5) holds. Note also that inequality (5) holds when the error density function has exponential tail as discussed in Result 3. This result is true for the same values of $q$ as in Result 3 and the discussion following that result. Further, inequality (5) holds when the error density function has polynomial tail as considered in Result 4. However, in that case, only for $k > 5/2$, the fourth moment of the density function with polynomial tail is finite, which is a condition required by Chen et al. (2003).

# 4. Some Numerical Results

In earlier sections, we compared the asymptotic variances of the least trimmed squares estimator and the trimmed mean. In order to compare the finite sample variances of these two estimators, we have done some Monte Carlo simulations for the location and the regression models. In all simulation studies, we have used sample size $n = 50$, and 1000 Monte Carlo samples of that size were generated in each case.

## 4.1. A simulation study for the location model

Here, we compute the finite sample variances of the 1/2-LTS and the 1/2-trimmed mean (i.e., median) estimators of the location parameter. We consider double exponential (DE(0, 1)), normal (N(0, 1)) and Cauchy (C(0, 1)) densities as the error density functions, where the location parameter and the scale parameter are zero and one respectively. The standard deviations of 1000 Monte Carlo replications of the 1/2-LTS estimator and the median are computed. These standard deviations can be taken as estimates of finite sample standard errors (S.E.) of the two estimators. We have carried out this and subsequent simulation studies by using the statistical package S-PLUS. The results are presented in the table below.

**Table 1:** Standard errors of the 1/2-LTS estimator and the median (i.e., 1/2-trimmed mean ) for location parameter.

| Distribution | S.E. of median | S.E. of 1/2-LTS estimator |
|---|---|---|
| $N(0,1)$ | 0.174993 | 0.358970 |
| $DE(0,1)$ | 0.150333 | 0.219434 |
| $C(0,1)$ | 0.223357 | 0.291299 |

## 4.2. A simulation study for linear regression model

We next consider some linear regression models with covariates having independent uniform (0, 1) distributions. Estimates of the regression parameters and the intercept in the linear model and their standard errors are computed. The error distributions considered are again double exponential (DE(0, 1)), normal (N(0, 1)) and Cauchy (C(0, 1)) distributions. The results are summerized in the table below.

**Table 2:** Standard errors of the intercept and the regression parameter estimates constructed by two different trimming procedures.

| Error Density | Number of Covariates | Parameters | S.E. of the median regression estimator | S.E. of the 1/2-LTS estimator | Eigen values of Dispersion($\hat{\theta}_{LTS}$) - Dispersion($\hat{\theta}_{LAD}$) |
|---|---|---|---|---|---|
| $N(0,1)$ | 1 | $\beta_0$ | 0.127326 | 0.452115 | 0.115306 |
| | | $\beta_1$ | 0.358591 | 1.037669 | and 0.319836 |
| $N(0,1)$ | 2 | $\beta_0$ | 0.248571 | 0.825463 | 0.118407, |
| | | $\beta_1$ | 0.396139 | 1.353462 | 0.429544 |
| | | $\beta_2$ | 0.446782 | 1.341363 | and 0.622083 |
| $DE(0,1)$ | 1 | $\beta_0$ | 0.152057 | 0.255491 | 0.178943 |
| | | $\beta_1$ | 0.353128 | 0.852322 | and 0.366729 |
| $DE(0,1)$ | 2 | $\beta_0$ | 0.232524 | 0.555277 | 0.169523, |
| | | $\beta_1$ | 0.354152 | 0.958907 | 0.447322 |
| | | $\beta_2$ | 0.351264 | 0.954411 | and 0.683286 |
| $C(0,1)$ | 1 | $\beta_0$ | 0.294205 | 0.353859 | 0.083214 |
| | | $\beta_1$ | 0.501512 | 1.105278 | and 0.279521 |
| $C(0,1)$ | 2 | $\beta_0$ | 0.497496 | 0.712445 | 0.091943, |
| | | $\beta_1$ | 0.748346 | 1.128293 | 0.356873 |
| | | $\beta_2$ | 0.829979 | 1.139245 | and 0.537984 |

## 4.3. Analysis of some real data

Here, we look at two examples involving real data to compare the 1/2-LTS and the median regression estimates. Monte Carlo replication is not possible in this case, and we have used bootstrap resampling procedure for comparing the standard errors of the two types of estimators. We used 1000 bootstrap samples in each of the following two examples.

### *4.3.1. Example 1: Pilot-Plant Data*

This data set is from Daniel and Wood (1971). For this data, Rousseeuw and Leroy (1987) considered a linear model

$$y = \beta_0 + \beta_1 x + e,$$

where $y$ is the acid content determined by titration, and $x$ is the organic acid content determined by extraction and weighing. In the table below, we give the bootstrap standard errors (S.E.) of the 1/2-LTS and the median regression estimators of $\beta_0$ and $\beta_1$.

**Table 3:** Bootstrap standard errors of the intercept and the regression parameter estimates for Pilot-Plant Data

| Parameters | S.E. of the 1/2-LTS estimator | S.E. of the median regression estimator | Eigen values of Dispersion($\hat{\theta}_{LTS}$) -Dispersion($\hat{\theta}_{LAD}$) |
|---|---|---|---|
| $\beta_0$ | 2.97381 | 0.740263 | 0.000019 |
| $\beta_1$ | 0.02066 | 0.007729 | and 1.181196 |

### *4.3.2. Example 2: Education Expenditure Data*

This data set is discussed in Chatterjee and Price (1977, p. 108). It contains education expenditure variable for fifty U.S. states. For this data, Rousseeuw and Leroy (1987) considered a linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e,$$

where $y$ is the per capita expenditure on public education in a state, $x_1$ is the number of residents per thousand residing in urban areas in 1970, $x_2$ is the per capita personal income in 1973, and $x_3$ is the number of residents per thousand under eighteen years of age in 1974. The following table contains the bootstrap standard errors of the 1/2-LTS and the median regression estimators of $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$.

**Table 4:** Bootstrap standard errors of the intercept and the regression parameter estimates for Education Expenditure Data

| Parameters | S.E. of the 1/2-LTS estimator | S.E. of the median regression estimator | Eigen values of Dispersion($\hat{\theta}_{LTS}$) -Dispersion($\hat{\theta}_{LAD}$) |
|---|---|---|---|
| $\beta_0$ | 347.30350 | 224.41719 | 0.00004, |
| $\beta_1$ | 0.14386 | 0.08587 | 0.00787, |
| $\beta_2$ | 0.03658 | 0.02171 | 0.06825 |
| $\beta_3$ | 0.92686 | 0.59062 | and 103916.2 |

# 5. Some Concluding Remarks

Sample median and the 1/2-LTS estimator have the same asymptotic breakdown points in the location model. However, in regression model, the asymptotic breakdown points of the 1/2-LTS estimator and the median regression estimator (i.e., the regression analogue of the 1/2-trimmed mean) are 50% and 0% respectively. This is due to the fact that the 1/2-LTS estimator is more robust against the outliers in the explanatory variables compared to the median regression estimator.

We have seen in Sections 2 and 3 that the median and its analogues in regression models are asymptotically more efficient compared to the corresponding 1/2-LTS estimators for several error densities. In Section 4, it is observed through numerical study that the finite sample variances of the median and the least absolute deviations estimators are also smaller than those of the 1/2-LTS estimators in a number of cases.

The computation of least absolute deviations estimator (i.e., the regression analogue of the 1/2-trimmed mean) is much simpler than the computation of the 1/2-LTS estimator in regression problems.

# 6. Appendix: Proofs

*Proof of Theorem 1:* Under the assumptions made on the distribution function $F$ and the density function $f$, it follows that

$$\sigma_1^2(\lambda) = \frac{2\int_0^{F^{-1}(1/2+\lambda/2)} t^2 f(t)dt}{[\lambda - 2F^{-1}(1/2+\lambda/2)f\{F^{-1}(1/2+\lambda/2)\}]^2}.$$

Rousseeuw and Leroy(1987, p.180) stated this result under the assumption that $f$ is a symmetric density function. They did not give the complete derivation, and referred to the proof given by Yohai and Maronna (1976). Yohai and Maronna (1976) assumed the differentiability of the density function $f$ and the finiteness of the Fisher information associated with it. However, by specializing Cizek's (2005) arguments to the location model, one can derive the above expression for the $\sigma_1^2(\lambda)$ under the assumptions stated in our theorem. Next, note that

$$\sigma_2^2(\alpha) = \frac{2[G(\alpha) + \alpha\{F^{-1}(1-\alpha)\}^2]}{(1-2\alpha)^2}$$

whenever $f$ is a symmetric and continuous density function, which is positive on an open interval containing the interval $[F^{-1}(\alpha), F^{-1}(1-\alpha)]$, and $G(\alpha) = \int_0^{F^{-1}(1-\alpha)} t^2 f(t)dt$. This expression of $\sigma_2^2(\alpha)$ is given in Lehmann(1983, p.361) and Serfling(1980, pp.236–237).

For $\alpha = 1/2$, the $\alpha$-trimmed mean is same as the sample median, and in that case, it can be shown that $\sigma_2^2(1/2) = (1/4)f^{-2}(0)$, which is the limiting value of the earlier expression of $\sigma_2^2(\alpha)$ as $\alpha \longrightarrow 1/2$.

Now,

$$\begin{aligned}
\sigma_1^2(1/2) - \sigma_2^2(1/2) &= \frac{2\int_0^{F^{-1}(3/4)} t^2 f(t)dt}{[1/2 - 2F^{-1}(3/4)f\{F^{-1}(3/4)\}]^2} - \frac{1}{4f^2(0)} \\
&= \frac{2\int_0^{F^{-1}(3/4)} t^2 f(t)dt}{4[1/4 - F^{-1}(3/4)f\{F^{-1}(3/4)\}]^2} - \frac{1}{4f^2(0)} \\
&= \frac{2f^2(0)\int_0^{F^{-1}(3/4)} t^2 f(t)dt - [1/4 - F^{-1}(3/4)f\{F^{-1}(3/4)\}]^2}{4f^2(0)[1/4 - F^{-1}(3/4)f\{F^{-1}(3/4)\}]^2}.
\end{aligned}$$

Inequality (1) implies that the numerator is nonnegative. Since the denominator is a perfect square, the theorem is proved. □

*Proof of Theorem 2:* For a unimodal, continuous and positive density function $f$, which is symmetric around zero and has the (3/4)th quantile $m = F^{-1}(3/4)$, we have the following:

$$mf(m) \leq 1/4 \tag{6}$$

and

$$mf(0) \geq 1/4. \tag{7}$$

In order to prove (6), note that the area under the graph of the function $f(t)$ from $t = 0$ to $t = m$ is 1/4 using the definition of the (3/4)th quantile. The function $f(t)$ is a decreasing function of $t$, and $mf(m)$ is the area of a rectangle contained in the region under the graph of the function $f(t)$ from $t = 0$ to $t = m$. This implies that $mf(m) \leq 1/4$.

To prove (7), we need to use the fact that $mf(0)$ is the area of a rectangle that contains the region under the graph $f(t)$ from $t = 0$ to $t = m$. This implies that $mf(0) \geq 1/4$.

Now, setting $z = mf(m)$, the given condition $z \geq 1/6$ and (6) imply that

$$\begin{array}{cl} (z-1/6) & (z-3/8) \leq 0 \text{ (since } z \leq 1/4 < 3/8) \\ \Leftrightarrow & z^2 - (3/8)z - (1/6)z + 1/16 \leq 0 \\ \Leftrightarrow & z^2 - (13/24)z + 1/16 \leq 0 \\ \Leftrightarrow & (1/24)z \geq 1/16 - (1/2)z + z^2 \\ \Leftrightarrow & (2/3)(1/16)z \geq [(1/4) - z]^2 \end{array}$$

Putting back $z = mf(m)$ and using (7), we have that the condition $z = mf(m) \geq 1/6$ implies the following.

$$\begin{array}{cl} (2/3) & f^2(0)m^2\{mf(m)\} \geq [1/4 - mf(m)]^2 \\ \Leftrightarrow & 2f^2(0)(m^3/3)f(m) \geq [1/4 - mf(m)]^2 \\ \Rightarrow & 2f^2(0)\int_0^m t^2 f(t)dt \geq [1/4 - mf(m)]^2 \text{ (since } f(t) \geq f(m) \text{ for all } t \leq m). \end{array}$$

This implies that inequality (1) holds whenever $mf(m) \geq 1/6$. □

*Proof of Result 3:* If the random variable $u_q$ has the exponential power density with power $= q$, then $|u_q|^q$ will have Gamma density with shape parameter $= 1/q$. So, $|u_q|^q$ is stochastically smaller than $|u_{q'}|^{q'}$ for all $q \geq q'$ since stochastic ordering holds in the Gamma distribution with respect to the shape parameter. This implies that $m_q^q$ is a decreasing function of $q$, where $m_q$ is the (3/4)th quantile of the density having exponential tail with power $= q$.

From the definition of the (3/4)th quantile, we get,

$$\begin{array}{rl} \int_0^{m_q} \frac{1}{2\Gamma(1+q^{-1})} e^{-y^q} dy = 1/4 \;\Leftrightarrow & \int_0^{m_q} e^{-y^q} dy = \Gamma(1+q^{-1})/2 \\ \Rightarrow & m_q \geq \Gamma(1+q^{-1})/2 \end{array} \qquad (8)$$

The last implication follows from the fact that $e^{-y^q} \leq 1$ for $y \geq 0$.

Now,

$$\begin{array}{rcl} m_q\{1/2\Gamma(1+q^{-1})\}e^{-m_q^q} & \geq & \{\Gamma(1+q^{-1})/2\}\{1/2\Gamma(1+q^{-1})\}e^{-m_q^q} \text{ (using(8))} \\ & = & (1/4)e^{-m_q^q}. \end{array}$$

In order to satisfy the sufficient condition of Theorem 2, we have to show that

$$e^{-m_q^q} \geq 2/3 \Leftrightarrow m_q^q \leq \ln 3 - \ln 2.$$

Note that, for $q = 2$, $m_2$ is the (3/4)th quantile of the normal density function with mean zero and variance 1/2. We now show that $m_2 \leq 1/2$. From the definition of the

(3/4)th quantile, in order to show this, it is enough to prove that

$$(1/\sqrt{\pi})\int_0^{1/2} e^{-y^2}dy \geq 1/4 \Leftrightarrow \int_0^{1/2} e^{-y^2}dy \geq \sqrt{\pi}/4.$$

Now, using integration by parts, we have

$$\begin{aligned}\int_0^{1/2} e^{-y^2}dy &= (1/2)e^{-1/4} + 2\int_0^{1/2} y^2 e^{-y^2}dy \\ &\geq (1/2)e^{-1/4} + 2e^{-1/4}(1/3)(1/8) \\ &\geq \sqrt{\pi}/4.\end{aligned}$$

So, we have established that $m_2^2 \leq 1/4$, which is smaller than $\ln 3 - \ln 2$. We have already proved that $m_q^q$ is a decreasing function of $q$. This implies that $m_q^q \leq \ln 3 - \ln 2$ for all $q \geq 2$.

*Proof of Result 4:* The sufficient condition of Theorem 2 holds if

$$m_k \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+m_k^2)^k} \geq 1/6,$$

where $m_k$ is the (3/4)th quantile of the density having polynomial tail with power $= k$.

From the definition of the (3/4)th quantile, we can write

$$\begin{aligned}&\int_0^{m_k} \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+y^2)^k}dy = 1/4 \\ \Leftrightarrow &\int_0^{m_k} \frac{1}{(1+y^2)^k}dy = \frac{\Gamma(1/2)\Gamma(k-1/2)}{4\Gamma(k)} \\ \Rightarrow \ & m_k \geq \frac{\Gamma(1/2)\Gamma(k-1/2)}{4\Gamma(k)}. \qquad (9)\end{aligned}$$

The last implication follows from the fact that $(1+y^2)^{-k} \leq 1$.

Next, we consider

$$\begin{aligned}&\int_0^{t_k} \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+y^2)^k}dy \qquad (10) \\ = &\int_0^{\tan^{-1}(t_k)} \frac{\Gamma(k)}{2^{k-1}\Gamma(1/2)\Gamma(k-1/2)}(1+\cos 2\theta)^{k-1}d\theta = g(k).\end{aligned}$$

From the definition of the (3/4)th quantile and using our assumption that $g(k) \geq 1/4$, (10) implies that $m_k \leq t_k = \sqrt{(3/2)^{1/k} - 1}$.

Now,

$$\begin{aligned}
m_k \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+m_k^2)^k} &\geq \frac{\Gamma(1/2)\Gamma(k-1/2)}{4\Gamma(k)} \frac{\Gamma(k)}{\Gamma(1/2)\Gamma(k-1/2)} \frac{1}{(1+m_k^2)^k} \quad \text{(using (9))} \\
&= (1/4)(1+m_k^2)^{-k} \\
&\geq (1/4)(2/3) \quad (\text{since } m_k \leq \sqrt{(3/2)^{1/k}-1}) \\
&= 1/6.
\end{aligned}$$

□

*Proof of Corollary 5:* The distributional conditions on the $\mathbf{x_i}$'s in Cizek (2005) follows from the assumptions that the $\mathbf{x_i}$'s are i.i.d. having finite second moment. Also, the assumptions on the regression function $h(\mathbf{x_i}, \theta)$ given in Section 3 imply directly the assumptions $(H)$ on the regression function in Cizek (2005). Under these assumptions along with the assumptions on the $e_i$'s given in Section 3, we have that $\Sigma_1(\lambda) = \sigma_1^2(\lambda) W_h^{-1}$, where $\sigma_1^2(\lambda)$ is as in Theorem 1 (see Cizek (2005, p.3976)).

In view of the strong law of large number and the conditions on the regression function $h(\mathbf{x_i}, \theta)$ given in section 3 along with the assumption that the $\mathbf{x_i}$'s are i.i.d., it follows that the conditions (a1)-(a4) on the regression function assumed by Chen et al. (2003) hold *almost surely* in the $\mathbf{x_i}$'s. Also, the distributional conditions on the $e_i$'s along with the finiteness of the fourth moment of $F$ imply the conditions on the $e_i$'s in Chen et al. (2003). Hence, we have $\Sigma_2(\alpha) = \sigma_2^2(\alpha) W_h^{-1}$, where $\sigma_2^2(\alpha)$ is as in Theorem 1 (see Chen et al.(2003, p.565)).

The two variances $\sigma_1^2(\lambda)$ and $\sigma_2^2(\alpha)$ were introduced in section 2, and their expressions were given in the proof of Theorem 1. Inequality (1) in Theorem 1 now implies that $\Sigma_2(1/2) \leq_L \Sigma_1(1/2)$ as $W_h^{-1}$ is a positive definite matrix.

□

# References


[1] Bickel, P. J. (1973) On some analogues to linear combinations of order statistics in the linear model. *The Annals of Statistics.*, **1**, 597–616.

[2] Burns, P. J. (1992) A genetic algorithm for robust regression estimation. *StatSci Technical Note, Seattle, Washington, U.S.A.*.

[3] Chatterjee, S. and Price, B. (1977) *Regression Analysis by Example.* John Wiley & Sons, New York.

[4] Chen, L. A., Huang, J. Y., Lin, M. L. and Yang, E. K. (2003) Welsh's trimmed mean for the nonlinear regression model. *Sankhyā*, **65**, 560–576.

[5] Cizek, P. (2005) Least trimmed squares in nonlinear regression under dependence. *Journal of Statistical Planning and Inference*, **136**, 3967–3988.

[6] Daniel, C. and Wood, F. S. (1971) *Fitting Equations to Data.* John Wiley & Sons, New York.

[7] Huber, P. J. (1981) *Robust Statistics*. John Wiley & Sons, New York.
[8] Jung, K. M. (2005) Multivariate least-trimmed squares regression estimator.*Computational Statistics and Data Analysis*, **48**, 307–316.
[9] Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
[10] Koenker, R. and Portnoy, S. (1987) L-estimation for linear models. *Journal of the American Statistical Association*, **82**, 851–857.
[11] Lehmann, E. L. (1983)*Theory of Point Estimation*. John Wiley & Sons , New York.
[12] Rousseeuw, P. J.(1984) Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871–880.
[13] Rousseeuw, P. J. and Diressen, K. (2006) Computing LTS regression for large data set.*Data Mining and Knowledge Discovery*, **12**, 29–45.
[14] Rousseeuw, P. J. and Leroy, A. M. (1987) *Robust Regression and Outlier Detection*. John Wiley & Sons , New York.
[15] Ruppert, D. and Carroll, R. J.(1980) Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, **75**, 828–838.
[16] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons , New York.
[17] Welsh, A. H. (1987) The trimmed mean in the linear model. *The Annals of Statistics*, **15**, 20–36.
[18] Yohai, V and Maronna, R.(1976) Location estimators based on linear combinations of modified order statistics. *Communication in Statistics (Theory and Methods)*, **5**, 481–486.