TAIWANESE JOURNAL OF MATHEMATICS
Vol. 6, No. 3, pp. 355-367, September 2002
This paper is available online at http://www.math.nthu.edu.tw/tjm/

ROBUST TYPE GAUSS-MARKOV THEOREM AND RAO′S FIRST ORDER EFFICIENCY FOR THE SYMMETRIC TRIMMED MEAN

Peter Thompson, Eugene K. Yang and Lin-An Chen

Abstract. In Chen and Chiang [2] and Chen, Thompson and Hung [3], the symmetric trimmed mean has been shown, for various linear models, to have the efficiency of having asymptotic covariance matrices close to the Crámer-Rao lower bounds for some heavy tail error distributions. In this paper, we investigate some further theoretical results for this symmetric trimmed mean for the linear regression model. From the nonparametric point of view, we develop a robust version of the Gauss-Markov theorem for the problem of estimating regression parameter vector β and parametric vector function $C\beta$ where the best estimators are this trimmed mean and C multiplied by it, respectively. In addition, we show that these best estimators are the best Mallows-type bounded influence linear symmetric trimmed means. Finally, from the parametric aspect, we show that the symmetric trimmed mean is Rao′s first order efficient for a heavy tail error distribution.

1. Introduction

Consider the linear regression model

$$y = X\beta + \epsilon \tag{1.1}$$

where y is a vector of observations for the dependent variable, X is an $n \times p$ matrix of observations of independent variables with $1'$s in the first column, and ϵ is a vector of independent and identically distributed error variables. We consider the problem of estimating the parameter vector β and the parametric vector function $C\beta$ of β, where C is a $q \times p$ constant matrix of rank q.

Received August 1, 2000.
Communicated by Y.-J. Lee.
2000 *Mathematics Subject Classification*: 62G35, 62J05.
Key words and phrases: Gauss-Markov theorem; Linear regression; Linear symmetric trimmed mean; Mallows-type bounded influence; Rao′s first order efficiency.

The most popular technique in estimating β and $C\beta$ is the least squares estimation. Its popularity mainly reflects its advantages in the theoretical property of using this technique. First, from the parametric point of view, besides the fact that it is a uniformly minimum variance unbiased estimator, it is Rao′s first order efficient ([10], p348) when the error variable follows a normal distribution. Secondly, from the nonparametric point of view, Gauss-Markov theorem states that it has the smallest covariance matrix in the class of unbiased linear estimators. However, the least squares estimators of β and $C\beta$ are sensitive to departures from normality and to the presence of outliers. Hence, we need to consider robust estimators.

Among the hundreds or more robust estimators for location and regression parameters investigated in the last three decades, the L-estimators, defined in terms of ordinary quantile or the regression quantile of Koenker and Bassett [7], have been an important estimator class (See Ruppert and Carroll [11], Koenker and Portnoy [8], and De Jong et. al. [4]). In terms of efficiency, L-estimators can be asymptotically efficient when the p.d.f. of the error distribution is exponentially decreasing on two tails (see Jurecková and Sen [6]). However, when the error variable follows a heavy tail distribution, the efficiency of these L-estimators no longer exist. Chen and Chiang [2] investigated the symmetric trimmed mean, constructed by a so-called symmetric regression quantile, and show that it has the efficiency of asymptotic covariance matrices close to the Cramer-Rao lower bounds for some heavy tail distributions. Results similar to this have also been obtained in Chen [1]. With this efficiency for heavy tail error distribution, a question raised is whether there are other optimal properties from either the nonparametric or parametric aspect for the symmetric trimmed mean. The aim of this paper it to answer this question.

From the nonparametric point of view, we will introduce classes of linear symmetric trimmed means for the estimation of β and $C\beta$. We then show that the estimators based on the symmetric trimmed mean are the best in terms of asymptotic covariance matrices. From this, we establish a robust version of the Gauss-Markov theorem. For describing these estimator classes, we will introduce a set of Mallows-type bounded influence symmetric trimmed means and show that this set forms a subclass of the linear symmetric trimmed means. On the other hand, from the parametric point of view, we will show that the symmetric trimmed mean satisfies the Rao′s first order efficiency when the error variable follows an extreme contaminated normal distribution.

In Section 2, we introduce a class of linear symmetric trimmed means. In Section 3, we derive their large sample properties and show that the symmetric trimmed mean is the best among the estimators in this trimmed mean class. In Section 4, we introduce a class of Mallows-type linear symmetric trimmed means and show that the symmetric trimmed mean is also the best among them. The Rao′s first order efficiency for the symmetric trimmed mean is shown in Section 5.

Finally, the proofs of the theorems are given in Section 6.

2. Linear Symmetric Trimmed Means

For $0 < \lambda < 1$, we call parameter vectors $\beta^-(\lambda)$ and $\beta^+(\lambda)$ the population symmetric quantiles if, for observation model $y_0 = x_0'\beta + \epsilon_0$, they satisfy

$$x_0'\beta^-(\lambda) = x_0'\beta - \tilde{F}^{-1}(\lambda) \text{ and } x_0'\beta^+(\lambda) = x_0'\beta + \tilde{F}^{-1}(\lambda)$$

where $\tilde{F}^{-1}(\lambda) = \inf\{a : P(|\epsilon_0| \quad a) \geq \lambda\}$. If F is a continuous distribution function, then $\tilde{F}^{-1}(\lambda)$ is the value satisfying $P(-\tilde{F}^{-1}(\lambda) \quad \epsilon_0 \quad \tilde{F}^{-1}(\lambda)) = \lambda$ which further implies that $P(x_0'\beta^-(\lambda) \quad y_0 \quad x_0'\beta^+(\lambda)) = \lambda$. Moreover, if we further assume that F is symmetric at 0, then $\tilde{F}^{-1}(1-2\alpha) = F^{-1}(1-\alpha)$ for $0 < \alpha < 0.5$ where F is the population quantile in the usual sense of F. This links the population symmetric quantile with the population regression quantile $\beta(\lambda)$ subjected to $x_0'\beta(\lambda) = x_0'\beta + F^{-1}(\lambda)$. Koenker and Bassett [7] proposed the so-called regression quantile to estimate the population regression quantile. Chen and Chiang [2] considered an initial-estimator-based method to estimate the population symmetric quantile. Let $\hat{\beta}_0$ be an initial estimator of β. Then $\tilde{F}^{-1}(\lambda)$ is estimated by

$$\hat{a}(\lambda) = \text{argmin}_{a>0} \sum_{i=1}^{n} (|y_i - x_i'\hat{\beta}_0| - a)(\lambda - I(|y_i - x_i'\hat{\beta}_0| \quad a))$$

and the symmetric quantile is defined as $\hat{\beta}^-(\lambda) = \hat{\beta}_0 - \begin{pmatrix} \hat{a}(\lambda) \\ 0_{p-1} \end{pmatrix}$ and $\hat{\beta}^+(\lambda) = \hat{\beta}_0 + \begin{pmatrix} \hat{a}(\lambda) \\ 0_{p-1} \end{pmatrix}$, where y_i is the i-th element of y and x_i' is the i-th row of X for $i = 1, ..., n$. The following theorem, developed by Chen and Chiang [2], provides a representation of $\hat{a}(\lambda)$.

Theorem 2.1. *Let F have the probability density function f. If f is symmetric at 0, then*

$$n^{1/2}(\hat{a}(1-2\alpha) - F^{-1}(1-\alpha)) = (2f(F^{-1}(1-\alpha)))^{-1} n^{-1/2} \sum_{i=1}^{n} [1 - 2\alpha - I(|\epsilon_i| \quad F^{-1}(1-\alpha))] + o_p(1).$$

With symmetric quantiles, $\hat{\beta}^-(\lambda)$ and $\hat{\beta}^+(\lambda)$, we then define the trimming matrix $A = \{a_{ij}, i, j = 1, ..., n$ and $a_{ij} = I(i = j$ and $x_i'\hat{\beta}^-(\lambda) \quad y_i \quad x_i'\hat{\beta}^+(\lambda))\}$. After outliers are trimmed by the symmetric quantiles, we have the following submodel

$$Ay = AX\beta + A\epsilon. \tag{2.1}$$

Since A is random, the error vector $A\epsilon$ is now not a set of independent variables.

Any linear unbiased estimator has the form My with M being a $p \times n$ non-stochastic matrix satisfying $MX = I_p$. Since M is a full-rank matrix, there exist matrices H and H_0 such that $M = HH_0'$. Thus, an estimator is a linear unbiased estimator if there exists a $p \times p$ nonsingular matrix H and an $n \times p$ full-rank matrix H_0 such that the estimator can be written as

$$HH_0'y. \tag{2.2}$$

We generalize linear unbiased estimators defined on the observation vector y to estimators defined on Ay by requiring them to be of the form MAy with $M = HH_0'$.

Definition 2.1. A statistic $\hat{\beta}_{lm}$ is called a linear symmetric trimmed mean if

$$\hat{\beta}_{lm} = MAy, \tag{2.3}$$

and there are H, a $p \times p$ stochastic or nonstochastic matrix, H_0, an $n \times p$ nonstochastic matrix with decomposition $M = HH_0'$; and an initial estimator $\hat{\beta}_0$ such that the following two conditions are satisfied:

(a1) $MAX = I_p$ where I_p is the $p \times p$ identity matrix.

(a2) $n^{1/2}(\hat{\beta}_0 - \beta) = Q_{hx}^{-1} n^{-1/2} \sum_{i=1}^{n} h_i \psi(\epsilon_i) + o_p(1)$ where h_i' represents the i-th row of matrix H_0, and $\psi(\epsilon)$ has zero mean and finite varinace.

This is similar to the usual requirements for unbiased estimation except that we have introduced a trimmed observation vector to allow for robustness and considered asymptotic property instead of unbiasedness.

For estimating the parametric vector function $C\beta$, we define a class of estimators analogously.

Definition 2.2. A linear function NAy is called a linear symmetric trimmed mean for a vector parametric function $C\beta$ if (1) the matrix N can be decomposed as $N = GH_0'$ with a $q \times p$-matrix G being stochastic or nonstochastic and H_0, an $n \times p$ nonstochastic matrix, and (2) conditions (a2) and the following one are satisfied:

(a1*) $NAX = C$.

Suppose that MAy is a linear symmetric trimmed mean for the parameter vector β. Then clearly NAy with $N = CM$ is a linear symmetric trimmed mean for the vector parametric function $C\beta$. This means that the results on the optimal estimation of $C\beta$ can be derived from the estimation of β.

Two questions arise for the class of linear symmetric means. First, does this class of means contain estimators that have already appeared in the literature? The answer is affirmative because the class of linear symmetric means defined in this paper contains the symmetric trimmed mean of Chen and Chiang [2] ($H = (X'AX)^{-1}$ and $H_0 = X$), and the set of Mallows-type bounded influence symmetric trimmed means ($H = (X'WAX)^{-1}$ and $H_0' = X'W$ with W, a diagonal matrix of weights; see Section 4). Second, is there a best estimator in this class of means and can we find it if it exists? This question will be answered in the next section.

3. Large Sample Properties of the Linear Symmetric Trimmed Mean

Denote by h_i' the i-th row of H_0. Let z_i represent either the vector x_i or h_i and z_{ij} be its j-th element. The following conditions are similar to the standard ones for linear regression models given in Chen and Chiang [2] and Koenker and Portnoy [8]:

(a3) $n^{-1}\sum_{i=1}^{n} z_{ij}^4 = O(1)$ for $z = x$ or h and all j,

(a4) $n^{-1}X'X = Q_x + o(1)$, $n^{-1}H_0'X = Q_{hx} + o(1)$ and $n^{-1}H_0'H_0 = Q_h + o(1)$ where Q_x and Q_h are positive definite matrices and Q_{hx} is a full rank matrix.

(a5) The probability density function f is symmetric at zero. f and its derivative are both bounded and bounded away from 0 in a neighborhood of $F^{-1}(\alpha)$ for $\alpha \in (0, 1)$.

Throughout this paper we let $\lambda = 1-2\alpha, 0 < \alpha < 0.5$, for the convenient setting of the quantile function of F. The following theorem gives a Bahadur representation of the linear symmetric trimmed mean.

Theorem 3.1. *Under conditions* (a1)-(a5), *we have*

$$n^{1/2}(\hat{\beta}_{lm} - \beta) = (1-2\alpha)^{-1} Q_{hx}^{-1} n^{-1/2} \sum_{i=1}^{n} h_i \phi_\psi(\epsilon_i) + o_p(1)$$

where $\phi_\psi(\epsilon) = 2F^{-1}(1-\alpha) f(F^{-1}(1-\alpha))\psi(\epsilon) + \epsilon I(|\epsilon| \quad F^{-1}(1-\alpha))$.

The limiting distribution of the linear symmetric trimmed mean follows from the central limit theorem (see, e.g. Serfling [12], p. 30).

Corollary 3.2. *Under the conditions of Theorem* 3.1, *the normalized linear symmetric trimmed mean* $n^{1/2}(\hat{\beta}_{lm} - \beta)$ *has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix* :

$$\sigma^2(\alpha) Q_{hx}^{-1} Q_h Q_{hx}^{-1\prime}$$

where

$$\sigma^2(\alpha) = (1-2\alpha)^{-2} \mathrm{E}(\phi_\psi^2(\epsilon)).$$

Next we consider the question of optimal linear symmetric trimmed estimation. For any two positive definite $p \times p$ matrices Q_1 and Q_2, we say that Q_1 is smaller than or equal to Q_2 if $Q_2 - Q_1$ is positive semidefinite. An estimator is said to be the best in an estimator-class if it is in this class and its asymptotic covariance matrix is smaller than or equal to that of any estimator in this class. The following lemma implies that any linear symmetric trimmed mean with asymptotic covariance matrix

$$\sigma^2(\alpha)Q_x^{-1} \tag{3.1}$$

is the best estimator in this class.

Lemma 3.3. *For any matrices Q_{hx} and Q_h induced from conditions* (a1) *and* (a4), *the difference*

$$Q_{hx}^{-1}Q_hQ_{hx}^{-1\prime} - Q_x^{-1},$$

is positive semidefinite.

The symmetric trimmed mean proposed by Chen and Chiang [2] is defined by

$$\hat{\beta}_s = (X'AX)^{-1}X'Ay. \tag{3.2}$$

Put $H = (X'AX)^{-1}$, and $H_0 = X$ which implies that (a1) holds. If we further let $\hat{\beta}_0$ satisfy condition (a2), then the symmetric trimmed mean is a linear symmetric trimmed mean. Moreover, Chen and Chiang [2] proved that $n^{1/2}(\hat{\beta}_s - \beta)$ has an asymptotic normal distribution with zero mean and covariance matrix of the form (3.1). Then Lemma 3.3 implies the optimal property of $\hat{\beta}_s$.

Theorem 3.4. *Under conditions* (a1)-(a5), *the symmetric trimmed mean $\hat{\beta}_s$ defined in* (3.2) *is a best linear symmetric trimmed mean.*

For estimating the parametric vector function $C\beta$, we have the following theorem.

Theorem 3.5. *Under conditions* (a1*) *and* (a2)-(a5), *we have*

(a) $n^{1/2}(NAy - C\beta)) = n^{-1/2}(1-2\alpha)^{-1}CQ_{hx}^{-1}\sum_{i=1}^n h_i\phi_\psi(\epsilon_i) + o_p(1)$, *and*

(b) *the normalized linear symmetric trimmed mean $n^{1/2}(NAy - C\beta)$ has an asymptotic normal distribution with zero mean and asymptotic covariance matrix*

$$\sigma^2(\alpha)CQ_{hx}^{-1}Q_hQ_{hx}^{-1\prime}C'.$$

Lemma 3.3. also implies that the product of C and the symmetric trimmed mean is asymptotically best in a class of linear functions of the trimmed observation vector Ay.

Theorem 3.6. *Under the conditions of Theorem* 3.5, *a best linear symmetric trimmed mean for estimating* $C\beta$ *is*

$$C\hat{\beta}_s,$$

where $\hat{\beta}_s$ *is a symmetric trimmed mean.*

In the class of linear estimators based on the trimmed observation vector Ay, we have shown that for estimating the parameter vector β and the vector parametric function $C\beta$, the symmetric trimmed mean and its product with C are both best linear symmetric trimmed means. This establishes a robust version of the Gauss-Markov theorem.

4. Mallows-Type Bounded Influence Symmetric Trimmed Means

Like most robust estimators, the symmetric trimmed mean has a Bahadur representation with influence function of the form

$$Q_x^{-1} x_i \gamma(\epsilon_i)$$

with function γ bounded in the domain of error variable ϵ if the initial estimator $\hat{\beta}_0$ also has influence function bounded in error variable. Then the influence function of the symmetric trimmed mean is bounded in the error space but not bounded in the space of explanatory variables. In statistical literature, the Mallows-type bounded influence technique has been applied by De Jongh et. al. [4] on the trimmed least squares estimator of Koenker and Bassett [7] and a trimmed mean proposed by Welsh [13]. Here we use the Mallows-type bounded influence technique to construct some bounded influence linear symmetric trimmed means.

Let $w_i, i = 1, ..., n$, be real numbers and denote the $n \times n$ matrix W by the diagonal matrix of the set $\{w_i, i = 1, ..., n\}$. The Mallows-type bounded influence symmetric trimmed mean is defined as

$$\hat{\beta}_w = (X'WAX)^{-1} X'WAy.$$

Here trimming matrix A is for bounding the influence of error variable and the weighted matrix W is for bounding the influence of explanatory variables (For construction of weights, see De Jongh et. al. [4]). In addition, the following assumption is valid.

(a6) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n w_i x_i x_i' = Q_w$, $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n w_i^2 x_i x_i' = Q_{ww}$, where Q_w and Q_{ww} are $p \times p$ positive definite matrices.

By letting $H = (X'WAX)^{-1}$ and $H_0' = X'W$, we see from the following lemma that the Mallows-type bounded influence symmetric trimmed mean is a linear symmetric trimmed mean.

Lemma 4.1. *Under conditions* (a1)-(a6),

$$n^{-1}X'WAX = (1-2\alpha)Q_w + o_p(1).$$

From condition (a2), $\hat{\beta}_0$ is required to have the following representation

$$n^{1/2}(\hat{\beta}_0 - \beta) = Q_w^{-1} n^{-1/2} \sum_{i=1}^{n} w_i x_i \psi(\epsilon_i) + o_p(1).$$

Theorem 4.2. *The Mallows-type bounded influence symmetric trimmed means form a subclass of linear symmetric trimmed means.*

The following theorem states a representation for this bounded influence symmetric trimmed mean.

Theorem 4.3.

$$n^{1/2}(\hat{\beta}_w - \beta) = (1-2\alpha)^{-1} Q_w^{-1} n^{-1/2} \sum_{i=1}^{n} w_i x_i \phi_\psi(\epsilon_i) + o_p(1),$$

and

$$n^{1/2}(\hat{\beta}_w - \beta) \to N(0, \sigma^2(\alpha) Q_w^{-1} Q_{ww} Q_w^{-1}).$$

As the fact that the symmetric trimmed mean is a Mallows-type bounded influence trimmed mean ($W = I_n$), we then can state the following theorem.

Theorem 4.4. *The symmetric trimmed mean is the best Mallows-type bounded influence symmetric trimmed mean.*

The Mallows-type bounded influence symmetric trimmed mean for vector parametric function $C\beta$ is defined as

$$C\hat{\beta}_w,$$

indexed in weight matrix W. Let $G = C(X'WAX)^{-1}$. It is seen that

$$G \to (1-2\alpha)^{-1} C Q_w^{-1} \text{ in probability.}$$

Moreover, by letting $H_0 = WX$,

$$GH_0AX = C.$$

Condition (a1$*$) holds and then we have the following theorem.

Theorem 4.5. *The Mallows-type bounded influence symmetric trimmed means $C\hat{\beta}_w$ are also linear symmetric trimmed means.*

Their large sample properties are easily obtained from Theorem 4.3.

Theorem 4.6.

(a) *A representation of the Mallows-type bounded influence symmetric trimmed means* $C\hat{\beta}_w$ *is*

$$n^{1/2}(C\hat{\beta}_w - C\beta) = (1-2\alpha)^{-1}CQ_w^{-1}n^{-1/2}\sum_{i=1}^{n} w_i x_i \phi_\psi(\epsilon_i) + o_p(1).$$

(b) *The normalized Mallows-type bounded influence symmetric trimmed mean* $n^{1/2}(C\hat{\beta}_w - C\beta)$ *has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix* :

$$\sigma^2(\alpha)CQ_w^{-1}Q_{ww}Q_w^{-1}C'.$$

The symmetric trimmed mean $C\hat{\beta}_s$ for estimating vector parametric function $C\beta$ is also a Mallows-type bounded influence symmetric trimmed mean ($W = I_n$).

Theorem 4.7. *The symmetric trimmed mean* $C\hat{\beta}_s$ *is the best Mallows-type bounded influence symmetric trimmed mean.*

These results are based solely on considerations of the asymptotic variances and ignore the fact that symmetric trimmed mean does not have bounded influence in the space of explanatory variables. It confirms that bounded influence is achieved at the cost of efficiency.

5. Rao′s First Order Efficiency for the Symmetric Trimmed Mean

Let z be an observation vector with joint probability density function f_z having unknown parameter vector θ. Rao ([10], p348) defined an estimator $\hat{\theta}$ with first order efficiency for estimating parameter vector θ if there is a constant matrix B such that

$$n^{1/2}||(\hat{\theta} - \theta) - B\frac{\partial \ln f_z(z,\theta)}{\partial \theta}|| = o_p(1),$$

where $||\ ||$ is the Euclidean norm in R^n.

Now, let y follow regression model (1.1) and redenote the symmetric trimmed mean of (3.2) by $\hat{\beta}_s(1-2\alpha)$ where $1-2\alpha$ is the percentage of observations retained for estimator computing. Suppose that error variable ϵ has a contaminated normal distribution as

$$(1-\delta)N(0,\sigma^2) + \delta N(0,\gamma\sigma^2), \tag{5.1}$$

where $0 < \delta < 1$, $\gamma > 0$. The contaminated normal distribution of (5.1) satisfies $\epsilon f(\epsilon) \to 0$ as $\epsilon \to \infty$ and as $\gamma \to \infty$, $F_\epsilon^{-1}(1-\delta/2) \to \infty$. Then as $\gamma \to \infty$, we have

$$\frac{\partial \ln f(y,\theta)}{\partial \theta} = -\sigma^{-2}\sum_{i=1}^{n} x_i \epsilon_i^*, \tag{5.2}$$

asymptotically, where ϵ_i^* are i.i.d. with distribution $N(0, \sigma^2)$. Only an estimator which has a representation proportional to (5.2) is Rao′s first order efficient. We now state that the symmetric trimmed mean is first order efficient in this extreme heavy tail error distribution.

Theorem 5.1. *We assume that initial estimator $\hat{\beta}_0$ has a representation with bounded influence function. Then, as $\gamma \to \infty$, $\hat{\beta}_s(1-\delta)$ is first order efficient with matrix $B = -(1-\delta)^{-1}\sigma^2 n^{-1/2} Q^{-1}$.*

It says that the symmetric trimmed mean with percentage of observations being removed exactly equal to that of the outliers in model (5.1) is first order efficient. This is the first result which bears an estimator with Rao′s first order efficiency in a heavy tail error distribution. It also further implies that the symmetric trimmed mean is one that can completely remove all outliers and retain all good observations.

6. Concluding Remarks

In this paper we have shown that the symmetric trimmed mean has several properties. (a) It is asymptotically the best among the class of linear symmetric trimmed means. This extends the property of being a best linear unbised estimation for the least squares estimator to robust estimation. (b) It is asymptotically the best among the class of Mallows-type bounded influence symmetric trimmed means. (c) It satisfies Rao′s first order efficiency. An interesting question raised from the above results is whether there is another trimmed mean or M-estimator that satisfies all or some of these properties. Moreover, since Léger and Romano [9] claimed that the adaptive trimmed mean of Welsh [13] can asymptotically achieve optimal trimming percentage, it would also be interesting to see if this property also holds for the symmetric trimmed mean.

7. Appendix

Proof of Theorem 3.1. Inserting (2.1) in equation (2.3), we have

$$n^{1/2}(\hat{\beta}_{lw} - \beta) = n^{1/2} H H_0' A \epsilon,$$

and from condition (a4) and Lemmas 3.2 of Jureçková [5] or Chen and Chiang [2], we have

$$n^{-1}\sum_{i=1}^{n} h_i x_i' I(|y_i - x_i'\hat{\beta}_0| \quad \hat{a}(1-2\alpha)) = (1-2\alpha)Q_{hx} + o_p(1),$$

which implies that

$$nH = (1-2\alpha)^{-1} Q_{hx} + o_p(1). \tag{7.1}$$

Moreover, from Lemma 3.1 of Jurecková [5], we have the following that develop a representation of $n^{-1/2} H_0' A \epsilon$

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} h_i \epsilon_i [I(|\epsilon_i - n^{-1/2} x_i' T_n| \\ & \quad n^{-1/2} T_0 + F^{-1}(1-\alpha)) - I(|\epsilon_i| \quad F^{-1}(1-\alpha))] \\ & = 2f(F^{-1}(1-\alpha)) F^{-1}(1-\alpha) n^{-1} \sum_{i=1}^{n} h_i x_i' T_n + o_p(1), \end{aligned} \tag{7.2}$$

for any sequences T_0 and T_n satisfying $T_0 = O_p(1)$ and $T_n = O_p(1)$. Equation (7.2) can be applied to obtain the following extension:

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} h_i \epsilon_i I(|\epsilon_i - n^{-1/2} x_i' T_n| \quad n^{-1/2} T_0 + F^{-1}(1-\alpha)) \\ & = n^{-1/2} \sum_{i=1}^{n} h_i \epsilon_i [I(|\epsilon_i - n^{-1/2} x_i' T_n| \quad n^{-1/2} T_0 + F^{-1}(1-\alpha)) \\ & -I(|\epsilon_i| \quad F^{-1}(1-\alpha))] + n^{-1/2} \sum_{i=1}^{n} h_i \epsilon_i I(|\epsilon_i| \quad F^{-1}(1-\alpha)) + o_p(1). \end{aligned} \tag{7.3}$$

The theorem is followed from (7.1) and the result of (7.3) with replacing T_n by $n^{-1/2}(\hat{\beta}_0 - \beta)$ and T_0 by $n^{1/2}(\hat{a}(1-2\alpha) - F^{-1}(1-\alpha))$.

Proof of Lemma 3.3. Write $\text{plim}(B_n) = B$ if B_n converges to B in probability. Let

$$P = HH_0 - (X'AX)^{-1} X'.$$

Now $PAX = HH_0' AX - (X'AX)^{-1} X'AX = 0$. Hence

$$\begin{aligned} & Q_{hx}^{-1} Q_h Q_{hx}^{-1\prime} = (1-2\alpha)^{-1} \text{plim}(HH_0' A (HH_0' A)') \\ & = (1-2\alpha)^{-1} \text{plim}((PA + (X'AX)^{-1} X'A)(PA + (X'AX)^{-1} X'A)'), \\ & = (1-2\alpha)^{-1} [\text{plim}(PAP') + \text{plim}((X'AX)^{-1} X'AX (X'AX)^{-1})], \\ & = (1-2\alpha)^{-1} \text{plim}(PAP') + (1-2\alpha)^{-2} Q_x^{-1}, \\ & \geq (1-2\alpha)^{-2} Q_x^{-1}. \end{aligned}$$

Proofs of Theorem 3.5, Lemma 4.1, Theorem 4.3 and Theorem 4.6 can all be analogously derived through the line for Theorem 3.1 and are then skipped.

Proof of Theorem 5.1. A representation of the symmetric trimmed mean in Chen and Chiang [2] is

$$\begin{aligned} & n^{1/2}(\hat{\beta}_s(1-2\alpha)-\beta) \\ & = \lambda^{-1}Q_x^{-1}[2F^{-1}(1-\alpha)f(F^{-1}(1-\alpha))Q_x n^{1/2}(\hat{\beta}_0-\beta) \\ & \quad + n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i I(|\epsilon_i| \quad F^{-1}(1-\alpha))] + o_p(1). \end{aligned}$$

However, as $\gamma \to \infty$, $F^{-1}(1-\delta/2) \to \infty$ and then

$$n^{1/2}(\hat{\beta}_s(1-\delta)-\beta) = (1-\delta)^{-1}Q_x^{-1}n^{-1/2}\sum_{i=1}^{n} x_i\epsilon_i^* + o_p(1),$$

where ϵ_i^* are i.i.d. with distribution $N(0,\sigma^2)$ which proves the theorem by comparing with (5.2).

Acknowledgements

The authors thank the referee whose valuable comments led to a substantial improvement of the exposition.

References

1. L. A. Chen, An efficient class of weighted trimmed means for linear regression models. *Statistica Sinica* **7** (1997), 669-686.

2. L. A. Chen and Y. C. Chiang, Symmetric type quantile and trimmed means for location and linear regression model. *Journal of Nonparametric Statistics* **7** (1996), 171-185.

3. L. A. Chen, P. Thompson and H.-N. Hung, The symmetric type two-stage trimmed least squares estimator for the simultaneous equations model, *Statistica Sinica* **10** (2000), 1243-1255.

4. P. J. De Jongh, T. De Wet and A. H. Welsh, Mallows-Type Bounded-Influence-Regression Trimmed Means, *Journal of the American Statistical Association* **83** (1988), 805-810.

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

6. J. Jurecková and P. K. Sen, *Robust statistical procedures - asymptotics and interrelations*. Wiley, New York, 1996.

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

8. R. Koenker and S. Portnoy, L-estimation for linear model. *Journal of the American Statistical Association* **82** (1987), 851-857.

9. C. Léger and J. P. Romano, Bootstrap adaptive estimation: the trimmed-mean example. *The Canadian Journal of Statistics* **18** (1990), 297-314.

10. C. R. Rao, *Linear statistical inference and its applications*. Wiley, New York, 1973.

11. D. Ruppert, and R. J. Carroll, Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* **75** (1980), 828-838.

12. R. J. Serfling, *Approximation theorems of mathematical Statistics*. Wiley, New York 1980.

13. A. H. Welsh, The trimmed mean in the linear model. *Annals of Statistics* **15** (1987), 20-36.

Peter Thompson
Mathematics Department
Wabash College
Crawfordsville, IN 47933
U.S.A.

Eugene K. Yang
Institute of Applied Mathematics
National Tsing Hua University
Hsinchu, Taiwan

Lin-An Chen
Institute of Statistics
National Chiao Tung University
Hsinchu, Taiwan