Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 126 (2004) 423–440

journal of statistical planning and inference

www.elsevier.com/locate/jspi

# Symmetric regression quantile and its application to robust estimation for the nonlinear regression model

Lin-An Chen[a], Lanh Tat Tran[b,*], Li-Ching Lin[a]

[a] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan
[b] Department of Mathematics, Indiana University, Rawles Hall, Bloomington, IN 47405-5701, USA

Received 2 July 2001; accepted 4 September 2003

**Abstract**

Populational conditional quantiles in terms of percentage $\alpha$ are useful as indices for identifying outliers. We propose a class of symmetric quantiles for estimating unknown nonlinear regression conditional quantiles. In large samples, symmetric quantiles are more efficient than regression quantiles considered by Koenker and Bassett (Econometrica 46 (1978) 33) for small or large values of $\alpha$, when the underlying distribution is symmetric, in the sense that they have smaller asymptotic variances. Symmetric quantiles play a useful role in identifying outliers. In estimating nonlinear regression parameters by symmetric trimmed means constructed by symmetric quantiles, we show that their asymptotic variances can be very close to (or can even attain) the Cramer–Rao lower bound under symmetric heavy-tailed error distributions, whereas the usual robust and nonrobust estimators cannot.


*Keywords:* Nonlinear regression; Regression quantile; Trimmed mean

## 1. Introduction

Consider the nonlinear regression model with observations

$$y_i = g(x_i, \beta) + \varepsilon_i, \quad 1 \leqslant i \leqslant n,$$

where $g(x, b)$ is a given function defined on an Euclidean space subset $\chi \times B$, and where $\beta$ is in the interior of $B$ and $\varepsilon_i$ are independent realizations of a random variable $\varepsilon$ with a distribution function $F_\varepsilon(y)$, $y \in \mathbb{R}$. In the nonlinear regression model with

* Corresponding author. Tel.: +1-812-855-7489; fax: +1-812-855-0046.
*E-mail address:* tran@indiana.edu (L.T. Tran).

0378-3758/$ - see front matter 
doi:10.1016/j.jspi.2003.09.014

Table 1
Asymptotic variances of estimates

| $\sigma$ | LSE | $\ell_1$ | TLSE | C–R |
|---|---|---|---|---|
| 3 | 1.80 | 1.803 | 1.295 | 1.256 |
| 10 | 10.9 | 1.896 | 1.431 | 1.229 |
| 25 | 63.4 | 1.922 | 1.466 | 1.171 |
| $\infty$ | $\infty$ | 1.938 | 1.489 | 1.114 |

normal errors, among all asymptotically normally distributed estimation sequences, the LSE is known to have the best asymptotic covariance matrix (see Bunke and Bunke, 1989). However, the LSE is highly sensitive to small departures from normality and to the presence of outliers. The development of robust alternatives for analyzing the nonlinear regression model has been investigated in several papers. Oberhofer (1982), Richardson and Bhattacharyya (1987) and Wang (1995) studied the $\ell_1$-norm estimators. The trimmed least squares estimator (TLSE) based on regression quantile was proposed by Koenker and Bassett (1978). Many aspects of the TLSE have been explored by Procházka (1988), Koenker and Park (1992), and Jurečková and Procházka (1994). The consistency of M-estimator for nonlinear regression model has been studied by Liese and Vajda (1994).

Are the available nonparametric estimators really efficient when the error variable has heavy-tailed distributions? The LSE, the $\ell_1$-norm estimator and TLSE all have asymptotic normal distributions with asymptotic covariance matrices equal to

$$\tau^2 Q^{-1} \tag{1.1}$$

for some function $\tau^2$, where

$$Q = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \frac{\partial g(x_i,\beta)}{\partial \beta} \frac{\partial g(x_i,\beta)'}{\partial \beta}.$$

Under some regularity conditions imposed on the regression function $g$ and the p.d.f. $f$, the Cramer–Rao (C–R) lower bound for unbiased estimators of $\beta$ also has the form (1.1) with $\tau^{-2} = E[(f'(\varepsilon)/f(\varepsilon))^2]$ (see Bunke and Bunke, 1989; Cramer, 1989). Consider the case where $\varepsilon$ has the mixed normal distribution $0.9\mathrm{N}(0,1) + 0.1\mathrm{N}(0,\sigma^2)$ and compare the asymptotic variances of the estimators mentioned above with the C–R bound. Table 1 provides the values $\tau^2$ of the estimators. The TLSE has the optimal trimming in the sense that it has the smallest asymptotic covariance matrix.

It is obvious that none of these usual robust and nonrobust estimators have asymptotic variances close to the C–R lower bound. The TLSE under optimal trimming has asymptotic variances relatively closer to the C–R lower bound than the other two estimators. However, the discrepancies are still significant when the contaminated variance $\sigma^2$ is large.

Basically, the efficiency of an estimator depends on the ability of the estimator to deal with good observations and bad observations (outliers). It is well known that the $\ell_1$-norm does not utilize the good observations sufficiently which results in a decrease in its efficiency as is shown in the table. On the other hand, although the TLSE, under optimal trimming, improves in efficiency, the discrepancy between its asymptotic variance and the C–R lower bound still shows inadequacy in utilizing observations (see Table 2 in Section 3). This implies that regression quantiles used in detecting outliers cannot precisely classify the observations into groups of good observations and outliers. Thus, in the nonlinear regression problem it is still possible to improve the estimator's efficiency by choosing an adequate construction process for data classification.

The questions of interest are: (1) Is there another method to estimate the population nonlinear regression quantile that can detect outliers more efficiently? (2) Can we construct nonparametric weighted means based on regression quantiles to estimate regression parameters so that their asymptotic variances are close enough to the C–R lower bound when the error variable has a heavy-tailed distribution? The purpose of this paper is to address these questions. To do this, we first introduce the nonlinear symmetric quantile by extending the idea of Kim (1992) and Chen and Chiang (1996). In large sample studies, the representation of the nonlinear symmetric quantile shows that it is consistent as an estimator of the population nonlinear symmetric quantile which is the population regression quantile of Koenker and Bassett (1978) whenever the underlying distribution is symmetric. Under a heavy-tailed distribution, the asymptotic variances of the symmetric regression quantiles of small and large percentage $\alpha$ are smaller than those of the corresponding regression quantiles of Koenker and Bassett (see Jurečková and Procházka (1994) for the nonlinear regression case). This is useful in identifying outliers since they always fall below the small or above the large $\alpha$th nonlinear conditional quantiles. We demonstrate the efficiency of the symmetric quantile by considering two symmetric trimmed means. The asymptotic representation shows that when the underlying distribution is asymmetric the symmetric trimmed mean has an asymptotic bias with a form analogous to that of the trimmed mean in the linear regression model (see (5.2) of Ruppert and Carroll, 1980). However, the asymptotic bias disappears when the distribution is symmetric. The asymptotic variances of the symmetric trimmed means are analyzed using heavy-tailed distributions. We demonstrate that the asymptotic variances can be significantly closer to the C–R lower bounds in comparison with those of robust and nonrobust estimators. The trimmed mean based on symmetric regression quantiles is shown to attain the C–R lower bound when the random errors have a contaminated normal distribution.

The nonlinear symmetric quantile is introduced in Section 2 and its large sample properties are investigated in Section 3. Examples of weighted mean constructed by nonlinear symmetric quantile are studied in Section 4. The proofs of the theorems are presented in Appendix. Many terms in the paper depend on the sample size $n$. However, we have suppressed this index $n$ in their notations for simplicity.

## 2. Symmetric type quantile

Recall that the nonlinear regression model for the observation $(y,x)$ is $y=g(x,\beta)+\varepsilon$. For $0<\alpha<1$, the $\alpha$th conditional regression quantile of $y$ given $x$ is

$$g(x,\beta)+F^{-1}(\alpha), \tag{2.1}$$

where $F^{-1}(\alpha)$ is the ordinary quantile function of $F$. If the regression function $g$ has a constant additive term, that is $g(x,\beta)=\beta_0+g_0(x,\beta_1)$ for some constant $\beta_0$, the vector $\beta(\alpha)=\begin{pmatrix}\beta_0(\alpha)\\ \beta_1\end{pmatrix}$ with $\beta_0(\alpha)=\beta_0+F^{-1}(\alpha)$ is called the population regression quantile. In this case, the vector $\beta(\alpha)$ has been studied by Jurečková and Procházka (1994) using the technique of Koenker and Bassett (1978). When the regression function $g$ does not have a constant additive term, the population regression quantile is $\beta(\alpha)=\begin{pmatrix}F^{-1}(\alpha)\\ \beta\end{pmatrix}$. In this case, the vector $\beta(\alpha)$ has been studied by Chen (1988) also using the technique of Koenker and Bassett (1978). The estimator of $\beta(\alpha)$ by the technique of Koenker and Bassett (1978) is called the regression quantile.

For $0<\lambda<1$ and $a>0$, define $\tilde{F}(a)=P(|\varepsilon|\leqslant a)$ where $\varepsilon$ is the error variable. Define the $\lambda$th symmetric quantile of $F$ as $\tilde{F}^{-1}(\lambda)=\inf\{a:\tilde{F}(a)\geqslant\lambda\}$, and the $\lambda$th nonlinear symmetric conditional quantile as

$$\{g(x,\beta)-\tilde{F}^{-1}(\lambda),g(x,\beta)+\tilde{F}^{-1}(\lambda)\}. \tag{2.2}$$

If $F$ is a continuous function, the nonlinear symmetric conditional quantile is easily seen to satisfy

$$P(g(x,\beta)-\tilde{F}^{-1}(\lambda)\leqslant y\leqslant g(x,\beta)+\tilde{F}^{-1}(\lambda))=\lambda.$$

Furthermore, if $F$ is continuous and symmetric at 0, then, for $0<\alpha<0.5$,

$$\tilde{F}^{-1}(1-2\alpha)=F^{-1}(1-\alpha).$$

In this case, the $\alpha$th and $(1-\alpha)$th nonlinear conditional regression quantiles in (2.1) and the $(1-2\alpha)$th nonlinear symmetric conditional regression quantile in (2.2) all coincide. The following theorem follows from Chen and Chiang (1996).

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

$$\tilde{F}^{-1}(\lambda)=\arg\min_{a>0} E_F(|y-g(x,\beta)|-a)(\lambda-I\{|y-g(x,\beta)|\leqslant a|\}). \tag{2.3}$$

Let $\hat{\beta}_I$ be an initial estimator of $\beta$. Following (2.3), consider the estimator of $\tilde{F}^{-1}(\lambda)$ defined by

$$\hat{a}(\lambda)=\arg\min_{a>0}\sum_{i=1}^{n}(|y_i-g(x_i,\hat{\beta}_I)|-a)(\lambda-I\{|y_i-g(x_i,\hat{\beta}_I)|\leqslant a\}). \tag{2.4}$$

The symmetric population quantile is

$$\begin{pmatrix}\beta_0+\tilde{F}^{-1}(\lambda)\\ \beta_1\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}\tilde{F}^{-1}(\lambda)\\ \beta\end{pmatrix}$$

depending on whether the model has a constant additive term or not. It is estimated by the symmetric regression quantile which, respectively, equals

$$\hat{\beta}_I + \begin{pmatrix} \hat{a}(\lambda) \\ 0_{p-1} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \hat{a}(\lambda) \\ \hat{\beta}_I \end{pmatrix}$$

in the first and second case.

## 3. Large sample properties of the symmetric quantile

The asymptotic distribution of the estimator of the population regression quantile depends on both $\hat{\beta}_I$ and $\hat{a}(\lambda)$. Without loss of generality, we will consider the nonlinear regression model with an additive constant term. The assumptions on error variable, design vectors $x_i$, and the nonlinear function $g$ are presented in the Appendix. They are assumed to hold in the rest of this paper. Define $\tilde{d}_i = [\partial g(x_i, \beta)]/\partial\beta$. The asymptotic distribution of $\hat{a}(\lambda)$ will be investigated with $\hat{\beta}_I$ as an initial estimator.

Define

$$q_0(\lambda) = n^{-1/2} \sum_{i=1}^{n} [\lambda - I\{|\varepsilon_i| \leqslant \tilde{F}^{-1}(\lambda)\}] - (f(\tilde{F}^{-1}(\lambda)) - f(-\tilde{F}^{-1}(\lambda)))\theta' n^{1/2}(\hat{\beta}_I - \beta),$$

where $\theta = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \tilde{d}_i$.

**Theorem 3.1.** (a) *If* $0 < \lambda < 1$, *then*

$$n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) = (f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)))^{-1} q_0(\lambda) + o_p(1).$$

(b) *Suppose that* $0 < \alpha < 0.5$ *and* $F$ *is a symmetric distribution*, *then*

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

The theorem implies the consistency of $\hat{a}(\lambda)$ for $\tilde{F}^{-1}(\lambda)$ which indicates that the symmetric regression quantile $\hat{\beta}_I + \begin{pmatrix} \hat{a}(\lambda) \\ 0_{p-1} \end{pmatrix}$ is consistent for the population symmetric regression quantile $\beta + \begin{pmatrix} \tilde{F}^{-1}(\lambda) \\ 0_{p-1} \end{pmatrix}$.

When should symmetric regression quantiles be employed in statistical inference in terms of their efficiencies? We will attempt to answer this by studying (a) the symmetric trimmed means (in Section 4) based on symmetric regression quantiles and by (b) comparing the asymptotic variances of symmetric type quantiles and regression quantiles.

For simplicity, we study the following linear regression model:

$$y_i = \beta_0 + x_i'\beta_1 + \varepsilon_i,$$

where $F$ is symmetric and $\sum_{i=1}^n x_i = 0$. Under this design, both quantiles are used to estimate the population quantile $\binom{\beta_0(\alpha)}{\beta_1}$. Recall that $\beta_0(\alpha)=\beta_0+F^{-1}(\alpha)$. The regression quantile, denoted here by $\binom{\hat\beta_0(\alpha)}{\hat\beta_1(\alpha)}$ has the following representation:

$$n^{1/2}\left(\begin{pmatrix}\hat\beta_0(\alpha)\\ \hat\beta_1(\alpha)\end{pmatrix} - \begin{pmatrix}\beta_0(\alpha)\\ \beta_1\end{pmatrix}\right)$$

$$= f^{-1}(F^{-1}(\alpha))\begin{pmatrix}1 & 0\\ 0 & Q_{11}^{-1}\end{pmatrix} n^{-1/2}\sum_{i=1}^n \begin{pmatrix}1\\ x_i\end{pmatrix}(\alpha - I\{\varepsilon_i < F^{-1}(\alpha)\}) + o_p(1),$$

where $Q_{11} = \lim_{n\to\infty}\sum_{i=1}^n x_i x_i'$. The symmetric regression quantile is $\binom{\hat\beta_{s0}(\alpha)}{\hat\beta_{s1}(\alpha)}$ with $\hat\beta_{s0}(\alpha) = \hat\beta_0 + \hat a(\alpha)$ and $\hat\beta_{s1}(\alpha) = \hat\beta_1$. Let the initial estimator $\hat\beta_I = \binom{\hat\beta_0}{\hat\beta_1}$ be the $\ell_1$-norm estimator. Using (b) of Theorem 3.1 and the representation of the $\ell_1$-norm estimator (see Koenker and Bassett, 1978), we have, for $0.5 < \alpha < 1$,

$$n^{1/2}(\hat\beta_{s0}(\alpha) - \beta_0(\alpha)) = f^{-1}(0)n^{-1/2}\sum_{i=1}^n (0.5 - I\{\varepsilon_i < 0\}) + 0.5 f^{-1}(F^{-1}(\alpha))n^{-1/2}$$

$$\times \sum_{i=1}^n (2\alpha - 1 - I\{-F^{-1}(\alpha) \leqslant \varepsilon_i \leqslant F^{-1}(\alpha)\}) + o_p(1),$$

and

$$n^{1/2}(\hat\beta_{s1}(\alpha) - \beta_1) = f^{-1}(0)n^{-1/2}Q_{11}^{-1}\sum_{i=1}^n x_i(0.5 - I\{\varepsilon_i < 0\}) + o_p(1).$$

Symmetric quantiles and regression quantiles employed to estimate $\beta_0(\alpha)$ and $\beta_1$ all have normal asymptotic distributions. Those used to estimate $\beta_1$ have asymptotic covariance matrices being $Q_{11}^{-1}$ multiplied by different constants. The efficiencies of the estimators can be compared by their constants. If the nonlinear regression model has a general form, the asymptotic variance and covariance matrices of the symmetric quantile and regression quantile are quite complicated and a direct comparison of their asymptotic variances is difficult. However, this difficulty does not occur for trimming estimators as shown in the next section. Consider the case where the error variable has the contaminated normal distribution

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

The efficiency of the symmetric quantile is defined as

$$\frac{\text{Asymptotic variance of regression quantile}}{\text{Asymptotic variance of symmetric regression quantile}}.$$

Table 2
Efficiencies of symmetric quantiles for estimating the quantile parameter $\beta_0(\alpha)$

| $\alpha$ | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | 0.98 |
|---|---|---|---|---|---|---|---|---|---|
| $\delta = 0.1$ | | | | | | | | | |
| $\sigma = 1$ | 0.87 | 0.84 | 0.84 | 0.84 | 0.87 | 0.92 | 1.01 | 1.21 | 1.47 |
| 3 | 0.91 | 0.89 | 0.90 | 0.92 | 0.98 | 1.09 | 1.30 | 1.84 | 1.90 |
| 5 | 0.91 | 0.90 | 0.91 | 0.94 | 1.01 | 1.15 | 1.44 | 2.39 | 2.02 |
| 10 | 0.90 | 0.89 | 0.90 | 0.93 | 1.01 | 1.16 | 1.49 | 2.70 | 2.03 |
| $\delta = 0.2$ | | | | | | | | | |
| $\sigma = 3$ | 0.88 | 0.87 | 0.88 | 0.92 | 1.00 | 1.14 | 1.40 | 1.78 | 1.98 |
| 5 | 0.89 | 0.88 | 0.91 | 0.97 | 1.08 | 1.28 | 1.68 | 2.04 | 2.00 |
| 10 | 0.89 | 0.89 | 0.93 | 1.01 | 1.16 | 1.45 | 1.98 | 2.10 | 2.03 |

Table 3
Efficiencies of symmetric quantiles for estimating the quantile parameter $\beta_1$

| $\alpha$ | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | 0.98 |
|---|---|---|---|---|---|---|---|---|---|
| $\delta = 0.1$ | | | | | | | | | |
| $\sigma = 1$ | 1.02 | 1.05 | 1.10 | 1.18 | 1.30 | 1.50 | 1.86 | 2.87 | 5.38 |
| 3 | 1.07 | 1.12 | 1.19 | 1.31 | 1.51 | 1.88 | 2.78 | 7.14 | 43.0 |
| 5 | 1.07 | 1.13 | 1.21 | 1.35 | 1.59 | 2.07 | 3.40 | 15.5 | 331 |
| 10 | 1.06 | 1.11 | 1.20 | 1.35 | 1.62 | 2.17 | 3.92 | 42.9 | 1318 |
| $\delta = 0.2$ | | | | | | | | | |
| $\sigma = 3$ | 1.04 | 1.09 | 1.19 | 1.34 | 1.61 | 2.16 | 3.73 | 15.2 | 66.6 |
| 5 | 1.04 | 1.11 | 1.23 | 1.43 | 1.82 | 2.75 | 6.79 | 114 | 178 |
| 10 | 1.05 | 1.13 | 1.27 | 1.53 | 2.06 | 3.60 | 17.4 | 502 | 681 |

The parameters to estimate are $\beta_0(\alpha)$ and $\beta_1$, respectively. Tables 2 and 3 list the efficiencies for the cases $\delta$=0.1 and 0.2, where $\sigma$=1, 3, 5, 10 and $\alpha$=0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95 and 0.98. Note that the results for $\alpha$ and $1 - \alpha$ are identical.

Based on Tables 2 and 3, regarding the estimation of the population regression quantile $\binom{\beta_0(\alpha)}{\beta_1}$, we notice the following:

(a) In estimating $\beta_0(\alpha)$, symmetric quantiles are more efficient than regression quantiles when $\alpha$ is small or large. Regression quantiles are more efficient than symmetric quantiles when $\alpha$ is close to 0.5 on either side.

(b) In estimating the slope parameters $\beta_1$, symmetric quantiles are more efficient than regression quantiles uniformly in $\alpha$.

(c) In estimating the population quantile vector $\binom{\beta_0(\alpha)}{\beta_1}$, the symmetric quantile is more efficient than the regression quantile when $\alpha$ is either small or large. It seems that symmetric quantiles are more suitable in classifying the data set into groups of good data and outliers. With suitable choice of the trimming percentage, high efficiency of estimation is attainable by proper weighting of those observations lying outside the estimated symmetric conditional quantile.

Similar to the ordinary quantile function or the regression quantile, the symmetric quantile has many applications in the study of influence functions. In the next section, we consider its application in parameter estimation for the case of two weighted means.

## 4. Weighted means based on symmetric quantile

The weighted means are defined based on a linearized model. This linear approximation is a single step Gauss–Newton method (see Kennedy and Gentle, 1980) based on a root-$n$ consistent estimator. Estimation based on linear approximation can be found in Fox et al. (1980) and Cook and Weisberg (1982). The one-step Huber's M-estimator by Bickel (1975) is an example of this technique for the linear regression model.

By the Taylor expansion theorem, there exists a function $\theta_i : \mathbb{R}^p \to \mathbb{R}$, $0 < \theta_i < 1$, such that

$$\beta_0 + g(x_i, \beta) = \hat{\beta}_0 + g(x_i, \hat{\beta}_1) + (\beta - \hat{\beta}_I)' d_i + 0.5(\beta_1 - \hat{\beta}_1)' \frac{\partial^2 g(x_i, b)}{\partial b \partial b'}|_{b=\beta_i} (\beta_1 - \hat{\beta}_1),$$

where

$$\hat{\beta}_I' = (\hat{\beta}_0, \hat{\beta}_1'), \quad \beta_i = \beta_1 + \theta_i(\hat{\beta}_1)\hat{\beta}_1, \quad \text{and} \quad d_i = \begin{pmatrix} 1 \\ \dfrac{\partial g(x_i, \hat{\beta}_1)}{\partial \hat{\beta}_1} \end{pmatrix}.$$

The approximate linearized regression model is

$$y_i = \hat{\beta}_0 + g(x_i, \hat{\beta}_1) + d_i' \beta^* + \varepsilon_i^*,$$

where $\beta^*$ represents the term

$$\beta - \hat{\beta}_I \quad \text{and} \quad \varepsilon_i^* = \varepsilon_i + 0.5(\beta_1 - \hat{\beta}_1)' \left. \frac{\partial^2 g(x_i, b)}{\partial b \partial b'} \right|_{b=\beta_i} (\beta_1 - \hat{\beta}_1).$$

The trimmed mean of $\beta$ is defined based on this linearized regression model. Let $0 < \lambda < 1$ and $y_i^* = y_i - (\hat{\beta}_0 + g(x_i, \hat{\beta}_1))$. Define the trimming matrix

$$A = \mathrm{diag}(a_i : a_i = I\{-\hat{a}_\lambda < y_i^* < \hat{a}_\lambda\}, i = 1, \ldots, n).$$

In addition, denote

$$D_n = \begin{bmatrix} d_1' \\ \vdots \\ d_n' \end{bmatrix}.$$

The symmetric trimmed mean for estimating $\beta$ is defined as

$$\hat{\beta}_{\mathrm{t}}(\lambda) = \hat{\beta}_I + L_n(\lambda) \quad \text{with } L_n(\lambda) = (D_n' A D_n)^{-1} D_n' A y^*,$$

where the vector $y^* = (y_1^*, y_2^*, \ldots, y_n^*)'$. The symmetric trimmed mean, defined through a linearization of the nonlinear regression function and the residuals based on an initial estimator has some advantages:

(a) This estimator is defined explicitly while most robust and nonrobust estimators are defined implicitly.

(b) This estimator appears to be simpler to compute in comparison with the trimmed means of Jurečková and Procházka (1994) for nonlinear regression and Koenker and Bassett (1978) for multiple linear regression.

Denote

$$g(\lambda) = \tilde{F}^{-1}(\lambda)(f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)))^{-1}(f(\tilde{F}^{-1}(\lambda)) - f(-\tilde{F}^{-1}(\lambda))),$$

$$q_1(\lambda) = Q(f(\tilde{F}^{-1}(\lambda)) + f(-\tilde{F}^{-1}(\lambda)))\tilde{F}^{-1}(\lambda) - (f(\tilde{F}^{-1}(\lambda)) - f(-\tilde{F}^{-1}(\lambda)))g(\lambda)\theta\theta',$$

$$q_2(\lambda) = g(\lambda)\theta n^{-1/2} \sum_{i=1}^{n} (\lambda - I\{|\varepsilon_i| \leqslant \tilde{F}^{-1}(\lambda)\}),$$

$$q_3(\lambda) = n^{-1/2} \sum_{i=1}^{n} \tilde{d}_i(\varepsilon_i I\{|\varepsilon_i| \leqslant \tilde{F}^{-1}(\lambda)\} - \tau).$$

**Theorem 4.1.** *With $q_1(\lambda)$, $q_2(\lambda)$ and $q_3(\lambda)$ as denoted above, we have*

$$n^{1/2}(\hat{\beta}_{\mathrm{t}}(\lambda) - (\beta + \gamma)) = \lambda^{-1} Q^{-1}\{q_1(\lambda) n^{1/2}(\hat{\beta}_I - \beta) + q_2(\lambda) + q_3(\lambda)\} + o_{\mathrm{p}}(1)$$

*where $\gamma = \lambda^{-1} Q^{-1}\theta\tau$ with*

$$\tau = \int_{-\tilde{F}^{-1}(\lambda)}^{\tilde{F}^{-1}(\lambda)} \varepsilon \, \mathrm{d}F(\varepsilon).$$

This result reveals that the symmetric trimmed mean $\hat{\beta}_{\mathrm{t}}(\lambda)$ is not generally consistent for regression parameter vector $\beta$ where consistency holds only if, under our assumptions, the term $\tau$ disappears. The following corollary displays the desired results.

**Corollary 4.2.** *Suppose that $F$ is symmetric and $\lambda = 1 - 2\alpha$ with $0 < \alpha < 0.5$. Then*

$$n^{1/2}(\hat{\beta}_{\mathrm{t}}(1-2\alpha) - \beta) = (1-2\alpha)^{-1}\left[2f(F^{-1}(1-\alpha))F^{-1}(1-\alpha)n^{1/2}(\hat{\beta}_I - \beta) + Q^{-1}n^{-1/2}\sum_{i=1}^{n}\tilde{d}_i\varepsilon_i I\{-F^{-1}(1-\alpha) \leqslant \varepsilon_i \leqslant F^{-1}(1-\alpha)\}\right] + o_{\mathrm{p}}(1). \tag{4.1}$$

Table 4
Asymptotic variances of estimates under $G = \mathrm{N}(0, \sigma^2)$

| $\delta$ | $\sigma$ | LS | $\ell_1$ | TLSE | $\hat{\beta}_t$ | C–R |
|---|---|---|---|---|---|---|
| 0.1 | 3 | 1.8 | 1.803 | 1.295(0.10) | 1.305(0.02) | 1.256 |
| | 5 | 3.4 | 1.855 | 1.373(0.13) | 1.287(0.03) | 1.253 |
| | 10 | 10.9 | 1.896 | 1.431(0.15) | 1.229(0.04) | 1.209 |
| | 25 | 63.4 | 1.922 | 1.466(0.16) | 1.171(0.05) | 1.161 |
| | $\infty$ | $\infty$ | 1.938 | 1.489(0.16) | 1.114(0.05) | 1.113 |
| 0.2 | 3 | 2.60 | 2.091 | 1.600(0.16) | 1.632(0.04) | 1.532 |
| | 5 | 5.8 | 2.226 | 1.770(0.20) | 1.605(0.06) | 1.540 |
| | 10 | 20.8 | 2.336 | 1.905(0.23) | 1.492(0.08) | 1.455 |
| | 25 | 125 | 2.406 | 1.988(0.25) | 1.377(0.09) | 1.356 |
| | $\infty$ | $\infty$ | 2.453 | 2.044(0.24) | 1.256(0.10) | 1.255 |

If we further assume that $\hat{\beta}_I$ is the $\ell_1$-norm estimator of $\beta$, then the representation of $\ell_1$-norm estimator (see Ruppert and Carroll, 1980) implies that

$$n^{1/2}(\hat{\beta}_t(1-2\alpha)-\beta) = (1-2\alpha)^{-1}Q^{-1}n^{-1/2}\sum_{i=1}^{n}\tilde{d}_i[(\varepsilon_i + \mathrm{sgn}(\varepsilon_i)f^{-1}(0)$$
$$\times f(F^{-1}(1-\alpha)))I\{|\varepsilon_i| \leqslant F^{-1}(1-\alpha)\}$$
$$+ f^{-1}(0)f(F^{-1}(1-\alpha))F^{-1}(1-\alpha)$$
$$\times \mathrm{sgn}(\varepsilon_i)I\{|\varepsilon_i| > F^{-1}(1-\alpha)\}] + o_p(1),$$

which has an asymptotic normal distribution with zero means and asymptotic covariance matrix $\sigma_s^2(\alpha)Q^{-1}$, where

$$\sigma_s^2(\alpha) = (1-2\alpha)^{-2}\left[(f(F^{-1}(1-\alpha))F^{-1}(1-\alpha)f^{-1}(0))^2 + 2\int_0^{F^{-1}(1-\alpha)} \varepsilon^2 \, \mathrm{d}F \right.$$
$$\left. + 4f(F^{-1}(1-\alpha))F^{-1}(1-\alpha))f^{-1}(0)\int_0^{F^{-1}(1-\alpha)} \varepsilon \, \mathrm{d}F\right].$$

Note that the symmetric trimmed mean also has an asymptotic normal distribution with zero means and covariance matrix of form (1.1) with $\tau^2 = \sigma_s^2(\alpha)$. Thus, in comparing the efficiencies of these estimators we only need to compare the values of $\tau^2$ and the C–R lower bound. Consider the error variable with standard normal distribution contaminated by a distribution $G$ with location parameter 0 and scale parameter $\sigma^2$, that is, the error variable has the distribution

$$(1-\delta)\mathrm{N}(0,1) + \delta G(0, \sigma^2).$$

We list in Table 4 the asymptotic variances for the $\ell_1$-norm estimator, the TLSE, and the symmetric trimmed mean $\hat{\beta}_t$. The values in parentheses are the trimming proportions corresponding to the trimmed means which achieve the smallest asymptotic variances.

Table 5
Asymptotic variances of estimates for $G = \text{Cauchy}(0, \sigma^2)$

| $\delta$ | $\sigma$ | $\ell_1$ | TLSE | $\hat{\beta}_t$ | C–R |
|---|---|---|---|---|---|
| 0.1 | 3 | 1.829 | 1.340(0.11) | 1.246(0.02) | 1.216 |
| | 5 | 1.872 | 1.398(0.13) | 1.244(0.03) | 1.219 |
| | 10 | 1.905 | 1.443(0.14) | 1.207(0.04) | 1.190 |
| | 25 | 1.925 | 1.471(0.15) | 1.161(0.04) | 1.153 |
| 0.2 | 3 | 2.157 | 1.694(0.18) | 1.518(0.05) | 1.459 |
| | 5 | 2.269 | 1.826(0.21) | 1.523(0.06) | 1.472 |
| | 10 | 2.359 | 1.933(0.23) | 1.449(0.08) | 1.415 |
| | 25 | 2.415 | 2.000(0.24) | 1.356(0.09) | 1.338 |

In Table 5, we list the asymptotic variances of the estimators considered above except the LSE for $G = \text{Cauchy}(0, \sigma^2)$.

From Tables 4 and 5 we draw several conclusions:

(a) For given $\delta$, the TLSE asymptotic variances increase with the variance of the contaminated distribution, whereas the symmetric trimmed mean behaves in nearly the opposite way. This interesting property implies that the power of the symmetric quantiles to detect contaminated data gradually increases with $\sigma^2$.

(b) The symmetric trimmed mean is not only more efficient than the $\ell_1$-norm and TLSE, but also has an asymptotic variance as small as the C–R lower bound when the contaminated variance goes to infinity.

Similar to the regression quantile of Koenker and Bassett (1978), symmetric quantiles have many applications. We consider here a refined weighted mean based on symmetric quantiles. We will show that the efficiencies of symmetric trimmed means still can be improved.

**Definition 4.3.** Let $0 < \lambda < 1, 0 \leqslant b \leqslant 1$ and $\hat{a}(\lambda)$ be the solution of (2.4). The symmetric Winsorized mean indexed by $(\lambda, b)$ is defined as

$$\hat{\beta}(\lambda, b) = \hat{\beta}_I + \ell_n(\lambda, b),$$

with $\ell_n(\lambda, b) = (D_n' A D_n)^{-1}(D_n' A y^* + b\hat{a}(\lambda) D_n' A^* 1_{\text{sgn}})$ and where $1_{\text{sgn}}$ is $n$-vector of $\text{sgn}(y_i^*)$ and $A^* = I_n - A$.

The symmetric $(\lambda, b)$th Winsorized mean and the symmetric $\lambda$th trimmed mean has the following relation:

$$\hat{\beta}(\lambda, b) = \hat{\beta}_t + b\hat{a}(\lambda)(D_n' A D_n)^{-1} D_n' A^* 1_{\text{sgn}}.$$

Denote by

$$g_1^*(\alpha) = 2(1-b) f(F^{-1}(1-\alpha)) F^{-1}(1-\alpha) Q,$$

Table 6
Asymptotic variances of estimators based on symmetric quantile

| $\sigma$ | $\gamma$ | $\hat{\beta}_{\mathrm{t}}$ | $\hat{\beta}_{\mathrm{w}}$ | C–R | $\gamma$ | $\hat{\beta}_{\mathrm{t}}$ | $\hat{\beta}_{\mathrm{w}}$ | C–R |
|---|---|---|---|---|---|---|---|---|
| 3 | 0.1 | 1.305 | 1.274 | 1.256 | 0.2 | 1.632 | 1.566 | 1.532 |
| 5 | | 1.287 | 1.277 | 1.253 | | 1.605 | 1.586 | 1.540 |
| 10 | | 1.229 | 1.227 | 1.209 | | 1.492 | 1.490 | 1.455 |

and

$$g_2^*(\alpha) = n^{-1/2} \sum_{i=1}^{n} \tilde{d}_i [\varepsilon_i I\{|\varepsilon_i| \leqslant F^{-1}(1-\alpha)\} + bF^{-1}(1-\alpha)\,\mathrm{sgn}(\varepsilon_i) I\{|\varepsilon_i| > F^{-1}(1-\alpha)\}].$$

**Theorem 4.4.** *Let* $\lambda = 1 - 2\alpha$ *with* $0 < \alpha < 0.5$. *If* $F$ *is symmetric around* 0, *then*

$$n^{1/2}(\hat{\beta}(1-2\alpha, b) - \beta) = (1-2\alpha)^{-1} Q^{-1} \{g_1^*(\alpha) n^{1/2}(\hat{\beta}_I - \beta) + g_2^*(\alpha)\} + o_{\mathrm{p}}(1).$$

Let $\hat{\beta}_I$ be the $\ell_1$-norm estimator. From Jurečková and Procházka (1994), it is seen that $n^{1/2}(\hat{\beta}(1-2\alpha, b) - \beta)$ has the normal asymptotic distribution with zero means and covariance matrix $\sigma_w^2 Q^{-1}$, where

$$\sigma_w^2 = (1-2\alpha)^{-2} \Big[2\alpha(bF^{-1}(1-\alpha) + (1-b)f^{-1}(0)F^{-1}(1-\alpha)f(F^{-1}(1-\alpha)))^2 + 2\int_0^{F^{-1}(1-\alpha)} \varepsilon^2 \,\mathrm{d}F + 4(1-b)f^{-1}(0)F^{-1}(1-\alpha) \times f(F^{-1}(1-\alpha)) \int_0^{F^{-1}(1-\alpha)} \varepsilon \,\mathrm{d}F + (1-2\alpha)(1-b)^2 (f^{-1}(0) \times F^{-1}(1-\alpha) f(F^{-1}(1-\alpha)))^2\Big].$$

We now give the asymptotic variances of two weighted means associated with the C–R lower bound in Table 6.

An inspection of the estimators' asymptotic variances reveals that the efficiencies of the symmetric trimmed means have improved. It is shown that the high efficiencies of the symmetric trimmed mean and the symmetric Winsorized mean depend only on optimal settings of the turning constants $\alpha$ and $b$. The adaptive estimator selected with the smallest bootstrap estimate of the finite sample variance can also achieve high efficiency (see Léger and Romano (1990) for information on the adaptive trimmed mean for location estimation).

The following theorem shows that the asymptotic variance of the symmetric trimmed mean can attain the C–R lower bound (as indicated in Table 4) when $\varepsilon$ has a contaminated normal distribution.

**Theorem 4.5.** *Suppose that the error variable* $\varepsilon$ *has the contaminated normal distribution*

$$(1-\delta)\mathrm{N}(0,\sigma^2)+\delta\mathrm{N}(0,\gamma\sigma^2) \tag{4.2}$$

*for* $\gamma>0$ *and some known* $\delta$, $0<\delta<1$. *In addition, assume that* $\hat{\beta}_I$ *has a bounded influence function. Then the asymptotic covariance matrix of* $\hat{\beta}_{\mathrm{t}}(1-\delta)$ *attains the C–R lower bound*

$$n^{-1}(1-\delta)^{-1}\sigma^2 Q^{-1}, \tag{4.3}$$

*as* $\gamma\to\infty$.

Theorem 4.5 has a practical meaning only in the rare cases where the level of contamination is known.

## Acknowledgements

The authors would like to express their appreciation to the Associate Editor and two anonymous referees for their valuable comments.

## Appendix

We now list the assumptions employed in the paper.

(a.1) $n^{-1}\sum_{i=1}^n \tilde{d}_i\tilde{d}_i' = Q+\mathrm{o}(1)$ and $n^{-1}\sum_{i=1}^n \tilde{d}_i=\theta+\mathrm{o}(1)$ where $Q$ is positive definite and $\theta$ is a finite vector.

(a.2) $n^{-1}\sum_{i=1}^n(\partial g(x_i,\beta)/\partial\beta_j)^4=\mathrm{O}(1)$, $n^{-1}\sum_{i=1}^n(\partial^2 g(x_i,\beta)/\partial\beta_j\beta_k)^2=\mathrm{O}(1)$.

(a.3) $n^{-1}\max_{\|\beta\|\leqslant b}\sum_{i=1}^n|\partial g(x_i,\beta)/\partial\beta_j|^2=\mathrm{O}(1)$ for some $b>0$.

$$n^{-1/4}\max_{\|\beta\|\leqslant b}|\partial g(x_i,\beta)/\partial\beta_j|=\mathrm{O}(1),$$

$$n^{-1/2}\max_{\|\beta\|\leqslant b}|\partial^2 g(x_i,\beta)/\partial\beta_j\partial\beta_k|=\mathrm{O}(1),$$

$$n^{-1/2}\max_{\|\beta\|\leqslant b}|\partial^3 g(x_i,\beta)/\partial\beta_j\partial\beta_k\partial\beta_h|=\mathrm{O}(1).$$

(a.4) The probability density function $f$ of $\varepsilon$ is bounded away from 0 in a neighborhood of $F^{-1}(\lambda)$ and $\tilde{F}^{-1}(\lambda)$, for some $0<\lambda<1$. In addition, $\varepsilon$ has a finite fourth population moment.

(a.5) $n^{1/2}(\hat{\beta}_I-\beta)=\mathrm{O}_{\mathrm{p}}(1)$.

To prove Theorem 3.1, we need several lemmas. Let, by replacing $d_i$ and $[\partial^2 g(x_i,b)]/\partial b\partial b'$ by $d_i(b)$ and $G_i(b)$, respectively,

$$h_i(c,t_1,t_3)=\begin{pmatrix} c \\ d_i(\beta+n^{-1/2}t_3)\end{pmatrix}' t_1,$$

$$g_i(t_2,t_3)=t_2' G_i(\beta+n^{-1/2}t_3)t_2,$$

and

$$S(t_1,t_2,t_3)=n^{-1/2}\sum_{i=1}^{n}(\lambda-I\{-\tilde{F}^{-1}(\lambda)+n^{-1/2}h_i(-1,t_1,t_3)$$
$$-0.5n^{-1}g_i(t_2,t_3)\leqslant\varepsilon_i\leqslant\tilde{F}^{-1}(\lambda)+n^{-1/2}h_i(1,t_1,t_3)$$
$$-0.5n^{-1}g_i(t_2,t_3)\}).$$

**Lemma A.1.** *For any* $b>0$,

$$\max_{\|t_j\|\leqslant b, j=1,2,3}\Bigg|S(t_1,t_2,t_3)-S(0,0,0)+n^{-1}\sum_{i=1}^{n}[f(\tilde{F}^{-1}(\lambda))h_i(1,t_1,t_3)$$
$$-f(-\tilde{F}^{-1}(\lambda))h_i(-1,t_1,t_3)+(f(\tilde{F}^{-1}(\lambda))$$
$$-f(-\tilde{F}^{-1}(\lambda)))0.5n^{-1/2}g_i(t_2,t_3)]\Bigg|=o_p(1).$$

**Proof.** Let $\alpha_1=P(\varepsilon<-\tilde{F}^{-1}(\lambda))$ and

$$S_1(t_1,t_2,t_3)=n^{-1/2}\sum_{i=1}^{n}(\lambda+\alpha_1-I\{\varepsilon_i\leqslant\tilde{F}^{-1}(\lambda)$$
$$+n^{-1/2}h_i(1,t_1,t_3)-0.5n^{-1}g_i(t_2,t_3)\})$$

and

$$S_2(t_1,t_2,t_3)=n^{-1/2}\sum_{i=1}^{n}(\alpha_1-I\{\varepsilon_i\leqslant-\tilde{F}^{-1}(\lambda)+n^{-1/2}h_i(-1,t_1,t_3)$$
$$-0.5n^{-1}g_i(t_2,t_3)\}).$$

So, $S(t_1,t_2,t_3)=S_1(t_1,t_2,t_3)-S_2(t_1,t_2,t_3)$.

Let $F_0$ satisfy $\alpha_0=P(\varepsilon<F_0)$ and

$$S_a(t_1,t_2,t_3)=n^{-1/2}\sum_{i=1}^{n}(\alpha_0-I\{\varepsilon_i\leqslant F_0+n^{-1/2}h_i(c,t_1,t_3)-0.5n^{-1}g_i(t_2,t_3)\}).$$

From Jurečková (1984) and Chen (1988, pp. 72–75),

$$\max_{\|t_j\|\leqslant b, j=1,2,3}\Bigg|S_a(t_1,t_2,t_3)-S_a(0,0,0)+n^{-1}f(F_0)\sum_{i=1}^{n}[h_i(c,t_1,t_3)$$
$$-0.5n^{-1/2}g_i(t_2,t_3)]\Bigg|=o_p(1). \tag{A.1}$$

Substituting $(\alpha_0,F_0)$ by $(\lambda+\alpha_1,\tilde{F}^{-1}(\lambda))$ and then by $(\alpha_1,-\tilde{F}^{-1}(\lambda))$ in (A.1), we obtain the representations of $S_1(t_1,t_2,t_3)$ and $S_2(t_1,t_2,t_3)$, respectively. Combining these two representations, we then have the lemma. □

**Lemma A.2.** $n^{-1/2}\sum_{i=1}^{n}(\lambda - I\{|y_i^*| < \hat{a}(\lambda)\}) = o_p(1)$.

**Proof.** We follow Ruppert and Carroll (1980). Let

$$G(c) = \sum_{i=1}^{n}(|y_i^*| - (\hat{a}(\lambda) + c))(\lambda - I\{|y_i^*| \leqslant \hat{a}(\lambda)\}).$$

The right derivative of $G$ at $c$ is

$$G^+(c) = (1-\lambda)\sum_{i=1}^{n} I\{|y_i^*| \leqslant \hat{a}(\lambda) + c\} - \lambda \sum_{i=1}^{n} I\{|y_i^*| > \hat{a}(\lambda) + c\}.$$

We need to show that $n^{-1/2}G^+(0) = o_p(1)$. Clearly, $G^+(c)$ is nondecreasing. So, for small $\delta > 0$,

$$G^+(-\delta) \leqslant G^+(0) \leqslant G^+(\delta).$$

Since $G$ achieves its minimum at 0, we have

$$G^+(0) \leqslant \lim_{\delta \to 0^+}(G^+(\delta) - G^+(-\delta)) = \sum_{i=1}^{n} I\{|y_i^*| = \hat{a}(\lambda)\}. \qquad \square \tag{A.2}$$

The lemma follows since the term on the right-hand side of (A.2) is bounded.

**Lemma A.3.** *If* $0 < \lambda < 1$, *then* $n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) = O_p(1)$.

**Proof.** The following inequality holds:

$$\begin{aligned}
&P(|n^{1/2}(\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda))| \geqslant k) \\
&\quad \leqslant P\left(\min_{|t_0| \geqslant k} n^{1/2}\left|\sum_{i=1}^{n}(\lambda - I\{|\varepsilon_i - n^{-1/2}d_i(\hat{\beta}_I)'T_1| \right.\right. \\
&\quad \left.\left. \leqslant n^{-1/2}t_0 + 0.5n^{-1}\tilde{\beta}_I' G_i(\hat{\beta}_I)\tilde{\beta}_I\})\right| < \eta\right) \\
&\quad\quad + P\left(n^{-1/2}\left|\sum_{i=1}^{n}(\lambda - I\{|\varepsilon_i - n^{-1/2}d_i(\hat{\beta}_I)'T_1| \right.\right. \\
&\quad \left.\left. \leqslant (\hat{a}(\lambda) - \tilde{F}^{-1}(\lambda)) + n^{-1}\tilde{\beta}_I' G_i(\hat{\beta}_I)\tilde{\beta}_I\})\right| \geqslant \eta\right),
\end{aligned} \tag{A.3}$$

where $T_1$ is an arbitrary sequence of random vectors with $T_1 = O_p(1)$ and $\tilde{\beta}_I = n^{1/2}(\hat{\beta}_I - \beta)$. Using the method of Jurečková (1977, Lemma 5.2) and Lemma A.1, one can show that for $\delta > 0$ there exist numbers $\eta, k$ and $N_0$ such that for $n \geqslant N_0$, the first term on the right-hand side of (A.3) is less than or equal to $\delta$. The proof follows by Lemma A.1. $\square$

**Proof of Theorem 2.2.** From Lemma A.3,

$$n^{1/2}\left(\begin{pmatrix}\hat{a}(\lambda)\\ \hat{\beta}_I\end{pmatrix}-\begin{pmatrix}\tilde{F}^{-1}(\lambda)\\ \beta\end{pmatrix}\right)=O_p(1).$$

Then Lemmas A.1 and A.3 imply that

$$-n^{-1/2}\sum_{i=1}^{n}(\lambda-I\{-\tilde{F}^{-1}(\lambda)\leqslant\varepsilon_i\leqslant\tilde{F}^{-1}(\lambda)\})$$

$$=n^{-1}\sum_{i=1}^{n}\left[\left(f(\tilde{F}^{-1}(\lambda))\begin{pmatrix}1\\ d_i(\hat{\beta}_I)\end{pmatrix}-f(-\tilde{F}^{-1}(\lambda))\begin{pmatrix}-1\\ d_i(\hat{\beta}_I)\end{pmatrix}\right)'\right.$$

$$\left.\times\, n^{1/2}\begin{pmatrix}\hat{a}(\lambda)-\tilde{F}^{-1}(\lambda)\\ \hat{\beta}_I-\beta\end{pmatrix}\right]+o_p(1),$$

from which the theorem follows. □

**Proof of Theorem 4.1.** From the setting of $L_n(\lambda)$, we have

$$n^{1/2}(\hat{\beta}_t-\beta)=(n^{-1}D_n'AD_n)^{-1}n^{-1/2}D_n'A\varepsilon.$$

Also, the following representation can be found in Ruppert and Carroll (1980) or Jurečková (1984)

$$n^{-1}D_n'AD_n=\lambda Q+o_p(1). \tag{A.4}$$

It is not hard to show that $n^{-1/2}D_n'A\varepsilon=n^{-1/2}\tilde{D}_n'A\varepsilon+o_p(1)$ with

$$\tilde{D}_n=\begin{pmatrix}\tilde{d}_1'\\ \vdots\\ \tilde{d}_n'\end{pmatrix}.$$

Let

$$U_j(t_1,t_2,t_3)=n^{-1/2}\sum_{i=1}^{n}\tilde{d}_{ij}\varepsilon_i I\{\varepsilon_i<a+n^{-1/2}h_i(c,t_1,t_2)-0.5n^{-1}g_i(t_2,t_3)\},$$

where $\tilde{d}_{ij}$ represents the $j$th element of $\tilde{d}_i$ and $(a,c)$ is either $(\tilde{F}^{-1}(\lambda),1)$ or $(-\tilde{F}^{-1}(\lambda),-1)$. Along the line of Chen (1988) and Jurečková (1984), we see that

$$U_j(T_1,T_2,T_3)-U_j(0,0,0)=n^{-1}af(a)\sum_{i=1}^{n}\tilde{d}_i[h_i(c,T_1,T_2)-0.5n^{-1}g_i(T_2,T_3)]$$

$$+o_p(1). \tag{A.5}$$

for any sequence $T=(T_1',T_2',T_3')'$ with $T=O_p(1)$. Using (A.5), the theorem follows by imposing $T_1=n^{1/2}\binom{\hat{a}(\lambda)-\tilde{F}^{-1}(\lambda)}{\hat{\beta}_I-\beta}$, $T_2=n^{1/2}(\hat{\beta}_1-\beta_1)$ and $T_3=n^{1/2}(\hat{\beta}_I-\beta)$ in the following representation:

$$\begin{aligned}n^{-1/2}\tilde{D}_n'A\varepsilon = n^{-1/2}\sum_{i=1}^n \tilde{d}_i\varepsilon_i[I\{\varepsilon_i<\tilde{F}^{-1}(\lambda)+n^{-1/2}h_i(1,T_1,T_2)-0.5n^{-1}g_i(T_2,T_3)\}\\ -I\{\varepsilon_i<\tilde{F}^{-1}(\lambda)\}]-n^{-1/2}\sum_{i=1}^n \tilde{d}_i\varepsilon_i[I\{\varepsilon_i<-\tilde{F}^{-1}(\lambda)\\ +n^{-1/2}h_i(-1,T_1,T_2)-0.5n^{-1}g_i(T_2,T_3)\}-I\{\varepsilon_i<-\tilde{F}^{-1}(\lambda)\}]\\ +n^{-1/2}\sum_{i=1}^n \tilde{d}_i\varepsilon_i I\{-\tilde{F}^{-1}(\lambda)<\varepsilon_i<\tilde{F}^{-1}(\lambda)\}. \qquad \square \qquad (A.6)\end{aligned}$$

**Proof of Theorem 4.4.** Clearly,

$$n^{1/2}(\hat{\beta}(1-2\alpha,b)-\beta)=(n^{-1}D_n'AD_n)^{-1}(n^{-1/2}D_n'A\varepsilon+n^{-1/2}b\hat{a}(1-2\alpha)D_n'A^*1_{\text{sgn}}).$$

By Theorem 3.1, (A.4) and (A.6), we only need to consider $n^{-1/2}\tilde{D}_n'A*1_{\text{sgn}}$ with

$$n^{-1/2}\tilde{D}_n'A*1_{\text{sgn}}=n^{-1/2}\sum_{i=1}^n \tilde{d}_i(I\{y_i^*>\hat{a}(1-2\alpha)\}-I\{y_i^*<-\hat{a}(1-2\alpha)\}.$$

By using arguments similar to those of (A.5) and (A.6), we obtain

$$\begin{aligned}n^{-1/2}\tilde{D}_n'A*1_{\text{sgn}} = -n^{-1/2}2f(F^{-1}(1-\alpha))Qn^{1/2}(\hat{\beta}_I-\beta)\\ +n^{-1/2}\sum_{i=1}^n \tilde{d}_i\,\text{sgn}(\varepsilon_i)I\{|\varepsilon_i>F^{-1}(1-\alpha)\}.\end{aligned}$$

The proof of Theorem 4.4 follows.

**Proof of Theorem 4.5.** Denote by $\tilde{g}_\gamma$ the contaminated distribution (4.2). The C–R bound for $\beta$ is

$$(1-\delta)^{-1}\left(E_{\tilde{g}_\gamma}\left(\frac{\partial \ln \tilde{g}_\gamma(\varepsilon)}{\partial\varepsilon}\right)^2\right)^{-1}Q^{-1}$$

which converges to the C–R lower bound given in (4.3) as $\gamma\to\infty$. On the other hand, the contaminated normal distribution of (4.2) satisfies $\varepsilon f(\varepsilon)\to 0$ as $\varepsilon\to\infty$. Since, $\hat{\beta}_I$ has a bounded influence function, from (4.1), the asymptotic covariance matrix of $\hat{\beta}_s(1-\delta)$ is

$$n^{-1}Q^{-1}(1-\delta)^{-2}E_{g_\sigma}\varepsilon^2 I\{|\varepsilon|\leqslant F_\varepsilon^{-1}(1-\delta/2)\},$$

where $g_\sigma$ is distribution of $N(0,\sigma^2)$. However, as $\gamma\to\infty$, $F_\varepsilon^{-1}(1-\delta/2)\to\infty$. Then the above variance is also the quantity of (4.3). This completes the proof of the theorem.

## References


Bickel, P.J., 1975. One-step Huber estimates in the linear model. J. Amer. Statist. Assoc. 70, 428–433.

Bunke, H., Bunke, O., 1989. Nonlinear Regression, Functional Relations and Robust Methods. Wiley, New York.

Chen, L.A., 1988. Regression quantiles and trimmed least squares estimators for structural equations and nonlinear regression model. Ph.D. Thesis, University of Illinois.

Chen, L.A., Chiang, Y.C., 1996. Symmetric type quantile and trimmed means for location and linear regression model. J. Nonparametric Statist. 7, 171–185.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman & Hall, New York.

Cramer, J.S., 1989. Econometric Applications of Maximum Likelihood Methods. Cambridge University Press, Cambridge.

Fox, T., Hinkley, D., Larntz, K., 1980. Jackknifing in nonlinear regression. Technometrics 22, 29–33.

Jurečková, J., 1977. Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann. Statist. 5, 464–472.

Jurečková, J., 1984. Regression quantile and trimmed least square estimator under general design. Kybernetika 20, 345–357.

Jurečková, J., Procházka, B., 1994. Regression quantiles and trimmed least squares estimator in nonlinear regression model. J. Nonparametric Statist. 3, 201–222.

Kennedy, W., Gentle, J., 1980. Statistical Computing. Dekker, New York.

Kim, S.J., 1992. The metrically trimmed mean as a robust estimator of location. Ann. Statist. 20, 1534–1547.

Koenker, R., Bassett, G.J., 1978. Regression quantile. Econometrica 46, 33–50.

Koenker, R., Park, B.J., 1992. An interior point algorithm for nonlinear quantile regression. Faculty Working Paper 92-0127. College of Commerce and Business Administration, University of Illinois, Urbana-Champaign.

Léger, C., Romano, J.P., 1990. Bootstrap adaptive estimation: The trimmed-mean example. Canad. J. Statist. 18, 297–314.

Liese, F., Vajda, I., 1994. Consistency of M-estimates in general regression models. J. Multivariate Anal. 50, 93–114.

Oberhofer, W., 1982. The consistency of nonlinear regression minimizing the $\ell_1$-norm. Ann. Statist. 10, 316–319.

Procházka, B., 1988. Regression quantiles and trimmed least squares estimator in the nonlinear regression model. Comput. Statist. Data Anal. 6, 358–391.

Richardson, G.D., Bhattacharyya, B.B., 1987. Consistent $\ell_1$-estimators in nonlinear regression for a noncompact parameter space. Shankhya A 49, 377–387.

Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. J. Amer. Statist. Assoc. 75, 828–838.

Wang, J., 1995. Asymptotic normality of $\ell_1$-estimators in nonlinear regression. J. Multivariate Anal. 54, 227–238.