

THE SYMMETRIC TYPE TWO-STAGE TRIMMED LEAST SQUARES ESTIMATOR FOR THE SIMULTANEOUS EQUATIONS MODEL

Lin-An Chen, Peter Thompson* and Hui-Nien Hung*

*National Chiao Tung University and * Wabash College*

Abstract: A two-stage symmetric regression quantile is considered as an alternative for estimating the population quantile for the simultaneous equations model. We introduce a two-stage symmetric trimmed least squares estimator (LSE) based on this quantile. It is shown that, under mixed multivariate normal errors, this trimmed LSE has asymptotic variance much closer to the Cramér-Rao lower bound than some usual robust estimators. It can even achieve the Cramér-Rao lower bound when the contaminated variance goes to infinity. This suggests that the symmetric-type quantile function is as efficient in other statistical applications, such as outlier detection. A Monte Carlo study under asymmetric error distribution and a real data analysis are also presented.

Key words and phrases: Regression quantile, simultaneous equations model, trimmed least squares estimator.

1. Introduction

The two-stage LSE, developed by Anderson and Rubin (1949) in the simultaneous equations model, is highly efficient under certain conditions. However, in the presence of heavy-tail errors, the two-stage LSE's efficiency can be markedly reduced. To overcome this, some robust techniques have been applied. Amemiya (1982) and Powell (1983) investigated the large sample property of a two-stage ℓ_1 -norm estimator. Krasker (1985) considered a two-stage bounded influence estimator. Chen and Portnoy (1996) applied the regression quantile technique of Koenker and Bassett (1978) to develop a two-stage trimmed LSE. However, from the comparison of the estimator's asymptotic variance with the Cramér-Rao (C-R) lower bound under mixed normal distributions (shown in Table 1 of this paper), the existing robust estimators are not really efficient under such heavy-tail errors. Is there an efficient technique for identifying outliers so that the corresponding trimmed LSE is efficient in such a way that the asymptotic variances are much closer to the C-R lower bound than the usual robust estimators?

On the basis of trimmed means for the location model (Kim (1992)) and the linear regression model (Chen and Chiang (1996)), a symmetric-type two-stage quantile for the simultaneous equations model is introduced. Then the two-stage symmetric trimmed LSE based on this symmetric quantile is introduced. It is observed that under mixed multivariate normal errors this two-stage symmetric trimmed LSE has asymptotic variance much closer to the C-R lower bound than the usual robust estimators. Moreover, when the contaminated variances tend to infinity, the asymptotic variance of this estimator achieves the C-R lower bound. In a Monte Carlo study under asymmetric distributions and an analysis of Australian wine data, the symmetric trimmed LSE produces satisfactory results. We note here that other research on the efficiency and high breakdown point estimation has also been done; see for example, He, Jureckova, Koenker and Portnoy (1990) and Coakley and Hettmansperger (1993).

2. Two-Stage Symmetric Trimmed Least Squares Estimator

Consider the simultaneous equations model

$$y = \beta_0 1_n + Y_1 \beta_1 + Z_1 \beta_2 + \lambda, \quad (2.1)$$

where $Y = (y, Y_1)$ is an $n \times p_0$ observation matrix of p_0 endogenous variables (i.e., dependent variables), Z_1 is an $n \times (p_1 - 1)$ observation matrix of $p_1 - 1$ exogenous variables (i.e., independent variables), λ is a vector of independent and identically distributed (i.i.d.) disturbance variables, and 1_n is an n -vector of 1's. We are interested in the estimation of the parameters $\beta = (\beta_1', \beta_0, \beta_2')'$.

Let Y follow a multivariate regression model $Y = Z\Pi + V$, known as the reduced form of the simultaneous equations model, where $Z = (1_n, Z_1, Z_2)$ is the set of all exogenous variables, Z_2 is an $n \times p_2$ matrix, and the rows of V are i.i.d. random vectors of variables v_1, \dots, v_{p_0} with an unknown joint distribution function. Let $\Pi = (\pi_1, \Pi_2)$ and $V = (V_1, V_2)$ be partitioned to correspond with the dimension of (y, Y_1) . Then the reduced form can be represented as

$$(y, Y_1) = Z(\pi_1, \Pi_2) + (V_1, V_2). \quad (2.2)$$

In two-stage estimation of the model parameters in (2.1), the first stage is the estimation of Π_2 in (2.2) by an initial estimator $\hat{\Pi}_2$. Define $\hat{Y}_1 = Z\hat{\Pi}_2$. From equations (2.1), (2.2) and the identification condition (see, for example, Theil (1971)) that $V_1 - V_2\beta_1 = \lambda$, we have

$$y = D_n \beta + U, \quad (2.3)$$

where $D_n = (Z\hat{\Pi}_2, 1_n, Z_1)$ and $U = V_1 - Z(\hat{\Pi}_2 - \Pi_2)\beta_1$. We also denote $\tilde{D}_n = (Z\Pi_2, 1_n, Z_1)$. The second stage of the estimation is carried out from the above induced model form.

Let $\hat{\beta}_I$ be an initial estimator of β for model (2.3). For $0 < \gamma < 1$, we define the γ -th symmetric type error quantile as

$$\hat{a}(\gamma) = \operatorname{argmin}_{a>0} \sum_{i=1}^n (|y_i - d'_i \hat{\beta}_I| - a)(\gamma - I(|y_i - d'_i \hat{\beta}_I| < a)). \tag{2.4}$$

Let f_1 be the marginal probability density function (p.d.f) of v_1 . For convenience, denote $f_\gamma^+ = f_1(\tilde{F}_1^{-1}(\gamma)) + f_1(-\tilde{F}_1^{-1}(\gamma))$, $f_\gamma^- = f_1(\tilde{F}_1^{-1}(\gamma)) - f_1(-\tilde{F}_1^{-1}(\gamma))$, and let v_{11}, \dots, v_{1n} be the elements of V_1 . A set of conditions on $Z, \hat{\Pi}_2, \hat{\beta}_I$ and the distribution of the error term, listed in the Appendix, is assumed to be true throughout this paper. A representation of the error quantile term $\hat{a}(\gamma)$ is stated in the following theorem.

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

$$\begin{aligned} n^{1/2}(\hat{a}(\gamma) - \tilde{F}_1^{-1}(\gamma)) &= (f_\gamma^+)^{-1} \left\{ n^{-1/2} \sum_{i=1}^n [\gamma - I(|v_{1i}| < \tilde{F}_1^{-1}(\gamma))] \right. \\ &\quad \left. + f_\gamma^- \theta' n^{1/2} \left(\begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix} (\hat{\beta}_I - \beta) + (\hat{\Pi}_2 - \Pi_2) \beta_1 \right) \right\} + o_p(1), \end{aligned}$$

where $\tilde{F}_1^{-1}(\gamma)$ satisfies $P(-\tilde{F}_1^{-1}(\gamma) < v_1 < \tilde{F}_1^{-1}(\gamma)) = \gamma$, $\theta = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n z_i$, where z'_i is the i -th row of Z .

For $0 < \gamma < 1$, a two-stage symmetric trimmed LSE $\hat{\beta}_s(\gamma)$ is defined as any vector b solving the standard equations $(D'_n A D_n) b = D'_n A y$, where $A = \operatorname{diag}(a_i : i = 1, \dots, n)$ with indicator function $a_i = I(d'_i \hat{\beta}_I - \hat{a}(\gamma) < y_i < d'_i \hat{\beta}_I + \hat{a}(\gamma))$. This $\hat{\beta}_s(\gamma)$ can also be formulated as $\hat{\beta}_s(\gamma) = (D'_n A D_n)^- D'_n A y$ where B^- is a generalized inverse of matrix B .

A representation of $\hat{\beta}_s$ is given in the following theorem.

Theorem 2.2. *If $0 < \gamma < 1$, then*

$$\begin{aligned} &\gamma \Sigma n^{1/2} (\hat{\beta}_s(\gamma) - \beta) \\ &= \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' [-\gamma Q + \tilde{F}_1^{-1}(\gamma) (f_\gamma^-)^2 (f_\gamma^+)^{-1} \theta \theta' + \tilde{F}_1^{-1}(\gamma) f_\gamma^+ Q] n^{1/2} (\hat{\Pi}_2 - \Pi_2) \beta_1 \\ &\quad + [\tilde{F}_1^{-1}(\gamma) (f_\gamma^-)^2 (f_\gamma^+)^{-1} \theta \theta' \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' + \tilde{F}_1^{-1}(\gamma) f_\gamma^+ \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' Q] n^{1/2} (\hat{\beta}_I - \beta) \\ &\quad + \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' [\tilde{F}_1^{-1}(\gamma) f_\gamma^- (f_\gamma^+)^{-1} \theta n^{-1/2} \sum_{i=1}^n (\gamma - I(|v_{1i}| \leq \tilde{F}_1^{-1}(\gamma)))] \\ &\quad + n^{-1/2} \sum_{i=1}^n z_i v_{1i} I(|v_{1i}| \leq \tilde{F}_1^{-1}(\gamma)) + o_p(1), \end{aligned}$$

where $Q = \lim_{n \rightarrow \infty} n^{-1} Z'Z$ and $\Sigma = \lim_{n \rightarrow \infty} n^{-1} \tilde{D}'_n \tilde{D}_n$.

If the distribution function F_1 is symmetric, then $f_\gamma^- = 0$, and $\hat{\beta}_s(1 - 2\alpha), 0 < \alpha < 0.5$, has the following representation.

$$\begin{aligned}
 n^{1/2}(\hat{\beta}_s(1 - 2\alpha) - \beta) &= ((1 - 2\alpha)\Sigma)^{-1} \left\{ \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' \right. \\
 &\quad + 2F_1^{-1}(1 - \alpha)f_1(F_1^{-1}(\alpha))Q \Big] n^{1/2}(\hat{\Pi}_2 - \Pi_2)\beta_1 \\
 &\quad + 2F_1^{-1}(1 - \alpha)f_1(F_1^{-1}(\alpha))\Sigma n^{1/2}(\hat{\beta}_I - \beta) \\
 &\quad \left. + n^{-1/2} \sum_{i=1}^n \tilde{d}_i v_{1i} I(|v_{1i}| \leq F_1^{-1}(1 - \alpha)) \right\} + o_p(1). \tag{2.5}
 \end{aligned}$$

Is $\hat{\beta}_s(1 - 2\alpha)$ efficient enough for some choices of the first stage estimator $\hat{\Pi}_2$ and initial estimator $\hat{\beta}_I$? To answer this question, consider the decomposition $\Pi_2 = (\pi_2, \pi_3, \dots, \pi_{p_0})$ and, for a consistent use of symmetric type estimation, let $\hat{\pi}_j, j = 2, \dots, p_0$, be the symmetric trimmed mean of π_j that uses the ℓ_1 -norm estimator as the initial estimator. The representation of $\hat{\pi}_j$ (see Chen and Chiang (1996)) is

$$n^{1/2}(\hat{\pi}_j - \pi_j) = Q^{-1} n^{-1/2} \sum_{i=1}^n z_i \psi(v_{ji}) + o_p(1), \tag{2.6}$$

with $\psi(v_{ji}) = (1 - 2\alpha)^{-1}(f_j(F_j^{-1}(\alpha))F_j^{-1}(1 - \alpha)f_j^{-1}(0)\text{sgn}(v_{ji}) + v_{ji}I(|v_{ji}| \leq F_j^{-1}(1 - \alpha)))$, where f_j and F_j represent the p.d.f. and distribution function, respectively, of v_j . From (2.6), we have the following further representation of $\hat{\beta}_s(1 - 2\alpha)$.

$$\begin{aligned}
 &n^{1/2}(\hat{\beta}_s(1 - 2\alpha) - \beta) \\
 &= ((1 - 2\alpha)\Sigma)^{-1} \left\{ n^{-1/2} \sum_{i=1}^n \tilde{d}_i (1, -\beta'_1)(v_{1i}I(|v_{1i}| \leq F_1^{-1}(1 - \alpha)), \right. \\
 &\quad \left. \dots, v_{p_0i}I(|v_{p_0i}| \leq F_{p_0}^{-1}(1 - \alpha)) \right\}' \\
 &\quad + 2F_1^{-1}(1 - \alpha)f_1(F_1^{-1}(1 - \alpha)) \left[\begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' \right. \\
 &\quad \left. + \Sigma n^{1/2}(\hat{\beta}_I - \beta) \right] - \sum_{j=2}^{p_0} f_j(F_j^{-1}(\alpha))F_j^{-1}(1 - \alpha)f_j^{-1}(0)n^{-1/2} \sum_{i=1}^n \tilde{d}_i \text{sgn}(v_{ji}) \Big\} + o_p(1).
 \end{aligned}$$

The following theorem states an interesting property of this symmetric trimmed LSE.

Theorem 2.3. *Assume that the error vector (v_1, \dots, v_{p_0}) has a mixed p_0 -variate normal distribution*

$$(1 - \delta)N_{p_0}(0, \Gamma_1) + \delta N_{p_0}(0, s\Gamma_2), \tag{2.7}$$

where $0 < \delta < 1$, $s > 0$, and Γ_1 and Γ_2 are positive definite matrices. It is also assumed that $\hat{\Pi}_2$ is the symmetric trimmed mean with ℓ_1 -norm estimator and $\hat{\beta}_I$ has a representation with a bounded influence function. Then, as $s \rightarrow \infty$, the asymptotic covariance matrix of $\hat{\beta}_s(1 - \delta)$ achieves the C-R lower bound

$$n^{-1}(1 - \delta)^{-1}(1, -\beta'_1)\Gamma_1(1, \beta'_1)'\Sigma^{-1}. \tag{2.8}$$

This is a property of efficiency that most robust estimators do not share. For further comparison of a symmetric trimmed LSE with some robust estimators for the simultaneous equations model, we let $\hat{\beta}_I$ be the two-stage ℓ_1 -norm estimator with symmetric trimmed mean as the first-stage estimator of Π_2 . Then, from Chen and Portnoy (1996, Theorem 3.1), we have the following result.

Corollary 2.4. *If the distribution functions $F_j, j = 1, \dots, p_0$ are symmetric, $\hat{\Pi}_2$ is the symmetric trimmed mean for the reduced form (2.2) with ℓ_1 -norm initial estimator, and $\hat{\beta}_I$ is the two-stage ℓ_1 -norm estimator, then*

$$n^{1/2}(\hat{\beta}_s - \beta) = \Sigma^{-1}n^{-1/2} \sum_{i=1}^n \tilde{d}_i(1, -\beta'_1)(\psi(v_{1i}), \dots, \psi(v_{p_0i}))' + o_p(1).$$

It is observed that the two-stage estimators, least squares, ℓ_1 -norm, and trimmed LSE based on the regression quantiles, and the symmetric trimmed LSE, all have multivariate normal distributions with zero mean and covariance matrices of the form $\sigma^2(\alpha)\Sigma^{-1}$ where $\sigma^2(\alpha) = (1, -\beta'_1)\text{Cov}(\psi(v_1), \dots, \psi(v_{p_0}))'(1, -\beta'_1)'$ for the two-stage symmetric trimmed LSE. Hence to compare the efficiencies of these estimators one only needs to compare the values of $\sigma^2(\alpha)$.

For a simple comparison, let us consider a simultaneous equations model with $p_0 = 2$ and $p_1 = 2$. The random error is assumed to have a contaminated bivariate normal distribution $(1 - \delta)N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) + \delta N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s^2 & 0 \\ 0 & s^2 \end{pmatrix}\right)$.

The parameter values are set for $\rho = .2$, $\beta_1 = 1$ and $s = 3, 5, 10, 25$ and ∞ . Table 1 provides the efficiencies of the two-stage LSE, two-stage ℓ_1 norm estimator, two-stage trimmed LSE and two-stage symmetric trimmed LSE, where an estimator's efficiency is defined as (Asymptotic variance of estimator) / (Cramer-Rao lower bound).

Conclusions from the above table of asymptotic variances are the following.

- (a) The two-stage LSE is not as efficient as the other three estimators, and performs poorly as the contaminated variance becomes high. The inefficiency of the two-stage LSE is due to the influence function being unbounded in the space of error variables.

- (b) Except for the two-stage LSE, the other three estimators have influence functions bounded in the space of the error variables (see Chen and Portnoy (1996) and (2.5)), so their asymptotic variances are all bounded by a constant. However, it is seen that the two-stage ℓ_1 -norm estimator is less efficient than the two-stage trimmed LSE, which is in turn less efficient than the two-stage symmetric trimmed LSE.

Table 1. Asymptotic variances and efficiencies of the two-stage estimators.

(δ, s)	2 stage LSE	2 stage ℓ_1	2 stage TLSE	2 stage STLSE	C-R bound
(.1, 3)	3.24 (.599)	3.190 (.608)	2.689 (.721)	2.185 (.888)	1.941
(.1, 5)	6.44 (.293)	3.283 (.574)	2.805 (.672)	2.137 (.883)	1.887
(.1, 10)	21.44 (.085)	3.356 (.543)	2.893 (.630)	2.015 (.904)	1.823
(.1, 25)	126.4 (.014)	3.400 (.526)	2.947 (.607)	1.900 (.941)	1.789
(.1, ∞)	∞ (.000)	3.423 (.519)	2.965 (.600)	1.777 (1)	1.777
(.2, 3)	4.88 (.474)	3.753 (.616)	3.270 (.707)	2.808 (.824)	2.314
(.2, 5)	11.28 (.196)	3.995 (.554)	3.545 (.625)	2.724 (.813)	2.216
(.2, 10)	41.28 (.050)	4.193 (.498)	3.765 (.555)	2.489 (.840)	2.092
(.2, 25)	251.2 (.008)	4.318 (.468)	3.902 (.518)	2.257 (.896)	2.024
(.2, ∞)	∞ (.000)	4.383 (.456)	3.950 (.500)	2.000 (1)	2.000

- (c) Except for the two-stage symmetric trimmed LSE, the asymptotic variances of the other three estimators are not close to the C-R lower bound. As stated in Theorem 2.3, the asymptotic variance of the two-stage symmetric trimmed LSE achieves the C-R lower bound when the contaminated variance of the mixed normal distribution reaches infinity.
- (d) When s goes up, the asymptotic variance of the two-stage symmetric trimmed LSE goes down like the C-R lower bound. This exceptional property does not appear in the usual estimators, whether they are robust or not. It also shows that the symmetric quantile $\hat{a}(\lambda)$ is a powerful outlier detector and will serve as a good tool for other statistical applications.

In order to study two-stage estimators for the simultaneous equations model with asymmetric error distributions, we perform a Monte Carlo simulation for the simple simultaneous equations model $y = \beta_0 + \beta_1 y_1 + \beta_2 x_1 + \beta_3 x_2 + \lambda$ with reduced form $(y, y_1) = (1, x_1, x_2)(\Pi_1, \Pi_2) + (v_1, v_2)$. We denote by $(exp_1(1), exp_2(1))$ a vector of two independent exponential random variables with mean 1. We assume that the error vector in the reduced form follows the mixed model

$$(v_1, v_2) = (1 - \delta) \begin{pmatrix} \sqrt{1 - \rho^2} \rho \\ 0 & 1 \end{pmatrix} \begin{pmatrix} exp_1(1) - 1 \\ exp_2(1) - 1 \end{pmatrix} + \delta s \begin{pmatrix} exp_1(1) - 1 \\ exp_2(1) - 1 \end{pmatrix}.$$

This design ensures that (v_1, v_2) has an asymmetric distribution with mean 0, drawn with probability $(1 - \delta)$ from a distribution with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and with probability δ from a distribution with covariance matrix $s^2 I_2$, where large values of s may produce outliers.

We also design (Π_1, Π_2) such that $\beta_j = 0.5, j = 0, 1, 2, 3$ and $\rho = 0.3$. Moreover, we take sample size $n = 40$ and samples (x_1, x_2) randomly generated from a bivariate normal distribution. With 1000 replications, we generate the observations (y, y_1, x_1, x_2) , obeying the assumptions above, and compute the two-stage ℓ_1 norm estimates, the trimmed LSE and the symmetric trimmed LSE. Tables 2, 3 and 4 display the results in terms of average mean squares errors (MSE).

Table 2. MSE for two-stage ℓ_1 -norm estimator.

$\delta = 0$	$\delta = 0.1$ $s = 3$	$s = 5$	$s = 10$	$s = 20$
2.627	20.22	10.55	15.62	77.69

Table 3. MSE for two-stage trimmed LSE.

trim percent	$\delta = 0$	$\delta = 0.1$ $s = 3$	$s = 5$	$s = 10$	$s = 20$
5%	0.383	0.343	1.166	0.528	0.593
10%	0.463	0.922	6.167	0.502	5.548
15%	1.115	0.417	1.318	0.701	2.894
20%	0.441	0.596	1.178	0.571	14.24
25%	11.49	1.018	0.395	0.492	1.337
30%	7.083	1.295	4.952	0.488	1.226

We have the following brief conclusions for the results in Tables 2, 3 and 4.

- (a) The two-stage symmetric trimmed LSE is not only relatively more efficient than the other two robust estimators, it is also more stable with respect to

the number of observations being trimmed.

- (b) The disadvantages of the other two estimators are that (1) the two-stage ℓ_1 -norm estimator estimates the population median, which is not identical to the mean that we aim to estimate, and (2) the outliers produced by asymmetric distributions are, in general, unbalanced with respect to the mean. This makes the trimmed LSE with equal trimming percentage remove both bad and good observations in the trimming process.

Table 4. MSE for two-stage symmetric trimmed LSE.

trim #	$\delta = 0$	$\delta = 0.1$ $s = 3$	$s = 5$	$s = 10$	$s = 20$
4	0.081	0.121	0.133	0.508	0.991
5	0.083	0.129	0.147	0.517	1.530
6	0.081	0.113	0.141	0.509	1.214
7	0.078	0.124	0.148	0.504	1.384
8	0.072	0.136	0.147	0.438	1.205
9	0.074	0.123	0.140	0.554	1.392
10	0.077	0.125	0.143	0.439	1.059

3. An Example

In this section, the results obtained are contrasted with various estimates on the wine industry data (see Maddala (1988)) of 20 observations in Australia from the year 1955-1956 to 1974-1975. A reasonable demand model is

$$y_1 = \beta_0 + \beta_{pw}y_2 + \beta_{pb}z_1 + \beta_d z_2 + \beta_e z_3 + \lambda$$

where all the variables are in logs. In this model, y_1 and y_2 are endogenous variables representing consumption and the price of wine, respectively, whereas, $z_i, i = 1, 2$ and 3 , are exogenous variables representing the price of beer, disposable income, and advertising expenditure. According to the theory of economics, it is anticipated that the true regression equation or the underlying main trend that represents their relationship should have a negative sign for the parameter β_{pw} and a positive sign for β_{pb}, β_d and β_e . There is an available instrumental variable z_4 , which is the index for storage costs. Detailed specifications for this model are given in Maddala (1988).

Maddala analyzed the data using the LSE and two-stage LSE (2SLS). Besides these two nonrobust estimators, we carry out a comparison with several robust estimators, all of whose first-stage estimators are based on the ℓ_1 -norm. These include the two-stage ℓ_1 -norm estimator ($2\ell_1$) and Huber's M-estimator (M) that solves the problem $\sum_{i=1}^n \psi((y_{1i} - d_i' b)/k) d_i = 0$, where $\psi(a) =$

$\max(-1.25, \min(a, 1.25))$, $k = k_0/0.6745$ and k_0 is the median of the absolute values of the residuals based on $2\ell_1$. The estimates are listed in Table 5.

Two conclusions can be drawn from Table 5.

- (a) It is quite obvious that the LSE, two-stage LSE and Huber's M-estimate do not catch the right trend since the signs of their estimates are partly wrong.
- (b) The two stage ℓ_1 -norm estimate finds the right trend in the data since the signs of the estimates are as expected.

Table 5. Some estimates in the Australian data.

Estimates	β_0	β_{pw}	β_{pb}	β_d	β_e
LSE	-23.65	1.158	-0.275	3.212	-0.603
2SLS	-26.19	0.643	-0.139	4.082	-0.985
M	-27.84	0.511	0.088	4.333	-1.063
$2\ell_1$	-25.87	-0.619	1.868	2.642	0.830

Tables 6 and 7 give the estimates, with various trimming proportions, of the two-stage trimmed LSE based on regression quantiles and the two-stage symmetric trimmed LSE.

Table 6. Two-stage trimmed LSE for the Australian data.

trim percent	β_0	β_{pw}	β_{pb}	β_d	β_e
0.05	-29.50	-4.088	-3.533	11.14	-3.776
0.10	-29.85	-3.496	-2.619	9.625	-2.666
0.15	-20.59	1.176	-0.306	2.382	0.008
0.20	-29.45	0.984	0.880	3.511	-0.534
0.25	-31.56	0.346	1.206	4.085	-0.573
0.30	-26.02	5.485	5.672	-4.699	3.162

Explanation and conclusion drawn from Tables 6 and 7.

- (a) The two-stage trimmed LSE is computed by sequentially increasing the trimming proportion α . Unfortunately, none of them give the signs suggested by theory.
- (b) The two-stage symmetric trimmed LSE is computed by removing observations with larger absolute residuals, with the number increasing sequentially. After 6, 7, 9 or more observations are removed, the estimates match the signs correctly. The symmetric type trimmed LSE is much better than the trimmed LSE based on regression quantiles for this data set.
- (c) The two-stage symmetric trimmed LSE with fifteen observations trimmed is identical to $2\ell_1$. This illustrates the theorem (Theorem. 3.1) of Koenker and

Bassett (1978) that a quantile can be formulated as the LSE in an observation subset of size equal to the number of parameters, five in this example.

Table 7. Two-stage symmetric trimmed LSE for the Australian data.

trim #	β_0	β_{pw}	β_{pb}	β_d	β_e
4	-24.43	0.348	0.529	2.918	0.181
5	-23.77	0.086	0.552	2.919	0.277
6	-23.59	-0.297	0.504	3.264	0.123
7	-22.61	-0.569	0.766	2.959	0.441
8	-17.61	0.095	0.189	1.877	0.817
9	-18.37	-0.046	0.392	1.954	0.847
10	-21.17	-0.122	0.788	2.293	0.700
11	-16.60	-0.355	0.514	1.577	1.247
12	-21.17	-0.906	1.566	1.936	1.406
13	-23.30	-0.914	1.828	2.197	1.272
14	-25.18	-0.723	1.884	2.531	0.941
15	-25.87	-0.619	1.868	2.642	0.830

Acknowledgements

The authors thank the Associate Editor and two referees whose valuable comments led to a substantial improvement of the exposition.

Appendix

The following assumptions are similar to the standard ones for linear regression models as given in Koenker and Portnoy (1987) and Welsh (1987).

- (a) Conditions concerning the matrix Z , estimators $\hat{\Pi}_2$ and $\hat{\beta}_I$ are these.
- (a.1) $n^{-1} \sum_{i=1}^n z_{ij}^A = O(1)$ for all j , and $n^{-1/4} \max_{ij} |z_{ij}| = O(1)$;
- (a.2) $n^{-1} \sum_{i=1}^n z_i = \theta + o(1)$ and $n^{-1} Z'Z = Q + o(1)$, where Q is positive definite;
- (a.3) $\Sigma = \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' Q \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}$ is a positive definite matrix;
- (a.4) $n^{1/2}(\hat{\Pi}_2 - \Pi_2) = O_p(1)$ and $n^{1/2}(\hat{\beta}_I - \beta) = O_p(1)$.
- (b) Both f_1 and f_1' are bounded away from 0 in a neighborhood of $\tilde{F}_1^{-1}(\gamma)$ for $\gamma \in (0, 1)$. Moreover, f_j , and f_j' , $j = 2, \dots, p_0$, are bounded away from 0 in a neighborhood of $F_j^{-1}(0)$.

Proof of Theorem 2.1. Let $H(t) = n^{-1/2} \sum_{i=1}^n [\gamma - I(|v_{1i} - n^{-1/2} z_i' t_n| < n^{-1/2} t_0 + \tilde{F}_1^{-1}(\gamma))]$, where $t = (t_0, t_n)'$. From Chen and Chiang (1996), we have,

for $k > 0$,

$$H(T) - H(0) - n^{-1} \sum_{i=1}^n (f_1(\tilde{F}_1^{-1}(\gamma))(1, z'_i) + f_1(-\tilde{F}_1^{-1}(\gamma))(1, -z'_i))T = o_p(1) \quad (\text{A.1})$$

for any random vector T with $T = O_p(1)$. Following Ruppert and Carroll (1980), we have

$$n^{-1/2} \sum_{i=1}^n [\gamma - I(|y_i - d'_i \hat{\beta}_I| < \hat{a}(\gamma))] = o_p(1). \quad (\text{A.2})$$

By rearrangement, the following equations hold:

$$y - D_n \hat{\beta}_I = -n^{-1/2} Z_n T_n + V_1 \quad (\text{A.3})$$

with

$$T_n = n^{1/2} \left(\begin{bmatrix} \hat{\Pi}_2 & I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix} (\hat{\beta}_I - \beta) + (\hat{\Pi}_2 - \Pi_2) \beta_1 \right). \quad (\text{A.4})$$

We use the method of Jureckova (1977, proof of Lemma 5.2) and (A.1), to show that for $\delta > 0$ there exist positive values η , k and N_0 such that

$$P(\inf_{|t_0| \geq k} n^{-1/2} \left| \sum_{i=1}^n [\gamma - I(|v_{1i} - n^{-1/2} z'_i T_n| < n^{-1/2} t_0 + \tilde{F}_1^{-1}(\gamma))] \right| < \eta) \leq \delta, \quad (\text{A.5})$$

where T_n is any sequence of random vectors with $T_n = O_p(1)$. Combining equations (A.2) and (A.5), we have

$$n^{1/2}(\hat{a}(\gamma) - \tilde{F}_1^{-1}(\gamma)) = O_p(1). \quad (\text{A.6})$$

The theorem follows by combining equations (A.4), (A.6) and (A.1).

Proof of Theorem 2.2. From equation (2.3), $\hat{\beta}_s$ can be formulated as

$$n^{-1}(D'_n A D_n)(\hat{\beta}_s - \beta) = -n^{-1} D'_n A Z (\hat{\Pi}_2 - \Pi_2) \beta_1 + n^{-1} D'_n A V_1.$$

From Jureckova (1984) we have

$$n^{-1}(D'_n A D_n) = \gamma \Sigma + o_p(1) \text{ and } n^{-1} D'_n A Z = \gamma \begin{bmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' Q + o_p(1). \quad (\text{A.7})$$

The rest is to find a representation for the term $n^{-1/2} Z' A V_1$ from $n^{-1} D'_n A V_1 = n^{-1} \begin{bmatrix} \hat{\Pi}_2 & I_{p_1} \\ 0_{p_2 \times p_1} \end{bmatrix}' Z' A V_1$. From (A.3) and (A.4) we can see that

$$\begin{aligned} n^{-1/2} Z' A V_1 &= n^{-1/2} \sum_{i=1}^n z_i v_{1i} \{ [I(v_{1i} < n^{-1/2}(1, z'_i)(T_0, T'_n)' + \tilde{F}_1^{-1}(\gamma)) \\ &\quad - I(v_{1i} < \tilde{F}_1^{-1}(\gamma))] - [I(v_{1i} < n^{-1/2}(-1, z'_i)(T_0, T'_n)' - \tilde{F}_1^{-1}(\gamma)) \\ &\quad - I(v_{1i} < -\tilde{F}_1^{-1}(\gamma))] + n^{-1/2} \sum_{i=1}^n z_i v_{1i} I(|v_{1i}| < \tilde{F}_1^{-1}(\gamma)) \}, \end{aligned}$$

where T_n is defined as in (A.4) and $T_0 = n^{1/2}(\hat{a}(\gamma) - \tilde{F}_1^{-1}(\gamma))$.

From Lemma 3.1 of Jureckova (1984) we have

$$\begin{aligned} n^{-1/2} Z' AV_1 &= \tilde{F}_1^{-1}(\gamma)(f_1(\tilde{F}_1^{-1}(\gamma)) - f_1(-\tilde{F}_1^{-1}(\gamma)))\theta n^{1/2}(\hat{a}(\gamma) - \tilde{F}_1^{-1}(\gamma)) \\ &\quad + \tilde{F}_1^{-1}(\gamma)(f_1(\tilde{F}_1^{-1}(\gamma)) + f_1(-\tilde{F}_1^{-1}(\gamma)))Qn^{1/2}\left[\hat{\Pi}_2 \begin{matrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{matrix}\right](\hat{\beta}_I - \beta) \\ &\quad + (\hat{\Pi}_2 - \Pi_2)\beta_1] + n^{-1/2} \sum_{i=1}^n z_i v_{1i} I(|v_{1i}| < \tilde{F}_1^{-1}(\gamma)) + o_p(1). \end{aligned}$$

Replacing $n^{1/2}(\hat{a}(\gamma) - \tilde{F}_1^{-1}(\gamma))$ with the representation in Theorem 2.1, by (A.7) and condition (a.4), Theorem 2.2 is proved.

Proof of Theorem 2.3. Denote by g_{Γ_1} and g_s the p.d.f.'s of $N_{p_0}(0, \Gamma_1)$ and the mixed distribution of (2.7), respectively. The C-R bound for β is

$$(1, -\beta_1') (E_{\tilde{g}_s} \frac{\partial \ln \tilde{g}_s(v_1, \dots, v_{p_0})}{\partial (v_1, \dots, v_{p_0})'})^{-1} \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix} (\tilde{D}'_n \tilde{D}_n)^{-1}$$

which converges to the matrix in (2.8) as $s \rightarrow \infty$. On the other hand, the contaminated normal distribution at (2.7) satisfies $v_j f_j(v_j) \rightarrow 0$ as $v_j \rightarrow \infty$. Since both $\hat{\Pi}_2$ and $\hat{\beta}_I$ have bounded influence functions, the asymptotic covariance matrix of $\hat{\beta}_s(1 - \delta)$ is

$$\begin{aligned} &n^{-1}(1 - \delta)^{-2} (1, -\beta_1') \lim_{n \rightarrow \infty} E_{g_{\Gamma_1}} (v_1 I(|v_1| \leq F_1^{-1}(1 - \delta/2)), \dots, v_{p_0} I(|v_{p_0}| \leq F_{p_0}^{-1}(1 - \delta/2))) \\ &\cdot (1 - \delta/2)' (v_1 I(|v_1| \leq F_1^{-1}(1 - \delta/2)), \dots, v_{p_0} I(|v_{p_0}| \leq F_{p_0}^{-1}(1 - \delta/2))) (1, -\beta_1')' \Sigma^{-1}. \end{aligned}$$

As $s \rightarrow \infty$, $F_j^{-1}(1 - \delta/2) \rightarrow \infty$ for $j = 1, \dots, p_0$. Then the above covariance matrix reduces to (2.8). This proves the theorem.

References

- Amemyia, T. (1982). Two stage least deviations estimators. *Econometrica* **50**, 689-711.
- Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* **20**, 46-63.
- Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428-433.
- Chen, L.-A. and Portnoy, S. (1996). Regression quantiles and trimmed least squares estimators for structural equations models. *Commun. Statist. A* **25**, 1005-1032.
- Chen, L.-A. and Chiang, Y. C. (1996). Symmetric quantile and trimmed means for location and linear regression model. *J. Nonparametr. Statist.* **7**, 171-185.
- Coakley, C. W. and Hettmansperger, T. P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.* **88**, 872-880.
- He, X., Jureckova, J., Koenker, R. and Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58**, 1195-1214.

- Jureckova, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear model. *Ann. Statist.* **5**, 464-472.
- Jureckova, J. (1984). Regression quantiles and trimmed least squares estimator under general design. *Kybernetika* **20**, 345-357.
- Kim, S. J. (1992). The metrically trimmed mean as a robust estimator of location. *Ann. Statist.* **20**, 1534-1547.
- Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**, 33-50.
- Koenker, R. W. and Portnoy, S. (1987). L estimation for linear model. *J. Amer. Statist. Assoc.* **82**, 851-857.
- Krasker, W. S. (1985). Two stage bounded-influence estimators for simultaneous equations models. *J. Business Economic Statist.* **4**, 432-444.
- Maddala, G. S. (1988). *Introduction to Econometrics*. Macmillan, New York.
- Powell, J. L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica* **51**, 1569-1576.
- Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model, *J. Amer. Statist. Assoc.* **75**, 828-838.
- Theil, H. (1971). *Principles of Econometrics*. Wiley, New York.
- Welsh, A. H. (1987). The trimmed mean in the linear model. *Ann. Statist.* **15**, 20-36.

Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, hsinchu 30050, Taiwan.

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

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

E-mail: thomsop@wabash.edu

(Received April 1998; accepted December 1999)