Journal of the Korean Statistical Society (2003), **32**: 1, pp 47–64

NONLINEAR ASYMMETRIC LEAST SQUARES ESTIMATORS†

SEUNG HOE CHOI[1], HAE KYUNG KIM[2] AND YOUNG LEE[2]

ABSTRACT

In this paper, we consider the asymptotic properties of asymmetric least squares estimators for nonlinear regression models. This paper provides sufficient conditions for strong consistency and asymptotic normality of the proposed estimators and derives asymptotic relative efficiency of the proposed estimators to the regression quantile estimators. We give some examples and results of a Monte Carlo simulation to compare the asymmetric least squares estimators with the regression quantile estimators.

AMS 2000 subject classifications. Primary 62F12; Secondary 62G05.
Keywords. Asymmetric least squares estimators, consistency, normality, efficiency.

1. INTRODUCTION

As for linear models, least squares (LS) estimators occur as the most important estimation method for nonlinear regression models. However, in spite of many favorable properties, a certain criticism on procedures based on least squares method has been pointed to robustness over outliers or slight departure from normality assumptions on errors. An alternative to LS estimators is often considered due to the fact that the distributions of errors are either grossly skewed or contaminated in many applied problems.

Alternative methods based on suitable notions of sample median or quantiles have been proposed by various authors: Oberhofer (1982), Koenker and Basset (1978), Basset and Koenker (1982, 1986), Jurečková and Procházka (1994), Buchinsky (1998), and Choi *et al.* (2001) are among those. Oberhofer (1982) proposed least absolute deviation (LAD) estimators based on sample median for

Received August 2002; accepted January 2003.
†This work was supported by Yonsei University Research Fund, 2000-1-0071.
[1]Department of General Studies, Hankuk Aviation University, Koyang 412-791, Korea
[2]Department of Mathematics, Yonsei University, Seoul 120-749, Korea

nonlinear regression models and gave sufficient conditions for consistency of the estimators. Basset and Koenker (1982, 1986), and Koenker and Basset (1978) proposed regression quantile (RQ) estimators which provide a natural generalization of the notion of sample quantiles to linear regression models and discussed asymptotic behavior of the estimators in linear models. Jurečková and Procházka (1994) studied asymptotic properties of RQ estimators and trimmed LS estimators in nonlinear regression models with intercept terms. Buchinsky (1998) provided a survey of recent theoretical development of RQ estimators. Choi *et al.* (2001) investigated asymptotic behavior of RQ estimators under the conditions different from those suggested by Jurečková and Procházka (1994).

In this paper we focus on an alternative estimation procedure that extends the concept of quantiles to more general nonlinear regression models with either skewed or contaminated errors. Consider the following nonlinear regression model

$$y_t = f(x_t, \theta_0) + \epsilon_t, \quad t = 1, \ldots, n \tag{1.1}$$

where y_t is the t^{th} observable response variable, $x_t \in \Gamma$ is a $(1 \times q)$ vector of input variable, the true parameter $\theta_0 = (\theta_1, \ldots, \theta_p)$ belongs to a parameter space $\Theta \in R^p$ and the response function $f(x, \theta)$ is continuous on $R^q \times R^p$. We assume throughout that the disturbance $\{\epsilon_t\}$ are independent and identically distributed (*iid*) random variables with a probability density function (*pdf*) $g(x)$ and a finite variance.

The β-regression quantile, denoted by $\tilde{\theta}_n(\beta)$, is defined as the value of θ minimizing the following function

$$R_n(\theta; \beta) = \frac{1}{n} \sum_{t=1}^{n} \varphi_\beta(y_t - f(x_t, \theta)) \tag{1.2}$$

where the "check function"

$$\varphi_\beta(x) = |\beta - 1(x < 0)|x = \begin{cases} \beta x, & \text{if } x \geq 0, \\ (1-\beta)x, & \text{if } x < 0, \end{cases}$$

and $0 < \beta < 1$. The LAD estimators are easily seen to be a special case of the β-regression quantile when $\beta = 1/2$.

Although the RQ estimators based on a general idea of quantiles are more efficient than LS estimators in the case of heavy-tailed or skewed error distribution, RQ estimation may be poor in the case where the distribution of the disturbance is a contaminated normal or is close to normal. For example, assume that the cumulative distribution function (*cdf*) $G(x)$ of the disturbance ϵ_t is the contaminated form

$$G(x) = \lambda\Phi(x) + (1-\lambda)\Phi(x-1)$$

where $\Phi((x-\mu)/\sigma)$ denotes the *cdf* of the normal distribution with mean μ and variance σ^2 and $\lambda \in [0,1]$. Then, the efficiency of the LS estimators to the LAD estimators, which equals the ratio of the variance and the value of the density at zero, is $1.4567/1.5567 < 1$ in the case of $\sigma = 2$ and $\lambda = 0.99$. So, the LS estimators are more efficient than the LAD estimators. Therefore, in this case it is needed to develop alternative estimators of the RQ estimators, which are based on the concept of gravity center for errors.

Newey and Powell (1987) replace the "check function" of the RQ estimators with the following loss function

$$\rho_\tau(x) = |\tau - 1(x<0)|x^2 = \begin{cases} \tau x^2, & \text{if } x \ge 0, \\ (1-\tau)x^2, & \text{if } x < 0, \end{cases}$$

where $0 < \tau < 1$. Any value of θ which minimizes the objective function

$$S_n(\theta;\tau) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau(y_t - f(x_t,\theta)) \tag{1.3}$$

is called asymmetric least squares (ALS) estimator of the true parameter θ_0 based on (y_t, x_t) and is denoted by $\hat{\theta}_n(\tau)$.

Note that the LS estimators is obviously an important special case of the ASL estimators because the loss function $\rho_\tau(x)$ rotates the square function $x^2/2$ by some angle ϕ in the clockwise direction. Newey and Powell (1987) explained the advantages and disadvantages of the ALS estimators relative to the RQ estimators and investigated asymptotic behavior of the ALS estimators in linear models with *iid* random errors.

The main purpose of this paper is to provide sufficient conditions for the asymptotic properties of the ALS estimators. The outline is as follow. In Section 2, we provide the sufficient conditions for strong consistency and asymptotic normality of ALS estimation in nonlinear models. Next, we derive confidence regions based on the ALS estimators and asymptotic relative efficiency (ARE) of the ALS estimators with the RQ estimators, in Section 3. Finally, we consider some examples and provide the results of the simulation study comparing the ALS with the RQ estimators, in Section 4.

2. ASYMPTOTIC PROPERTIES OF ALS ESTIMATION

In this section we present the sufficient conditions for consistency and normality of the nonlinear ALS estimators. Let $G(x)$ denote the distribution function of the disturbances and $q_n(\theta)$ indicate the number of elements of the set $\{t : f(x_t,\theta) \neq f(x_t,\theta_0)\}$ for each $\theta \in \Theta$ with $\theta \neq \theta_0$. To simplify the notation, we denote

$$f_t(\theta) = f(x_t,\theta), \quad \nabla f_t(\theta) = \left[\frac{\partial}{\partial\theta_i} f_t(\theta)\right]_{(p\times 1)}, \quad \nabla^2 f_t(\theta) = \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} f_t(\theta)\right]_{(p\times p)}.$$

Throughout this paper, we want to make the following assumptions in the nonlinear regression model (1.1):

ASSUMPTION A.

(A1) The parameter space Θ is a compact subspace of R^q and Γ is a bounded subset of R^q.

(A2) For all t, the partial derivatives $\nabla f_t(\theta)$ and $\nabla^2 f_t(\theta)$ exist, and $\nabla f_t(\theta)$ are continuous on $\Gamma \times \Theta$.

(A3) The ratio of the number of elements of the set $\{t : f(x_t,\theta) \neq f(x_t,\theta_0)\}$ to of the sample size in the model (1.1), denoted by $q_n(\theta)/n$, converges to $q, 0 < q \leq 1$.

ASSUMPTION B. The probability density function $g(x)$ is continuous and strictly positive at zero.

The following lemma deals with uniform convergence of the modified objective function

$$Q_n(\theta;\tau) = S_n(\theta;\tau) - S_n(\theta_0;\tau). \tag{2.1}$$

Since $S_n(\theta_0;\tau)$ is independent of θ, the ALS estimator is equivalent to the value minimizing (2.1).

LEMMA 2.1. *Suppose that Assumptions (A1) and (A2) are satisfied for the model* (1.1). *Then we have*

$$Q_n(\theta;\tau) - E\{Q_n(\theta;\tau)\} = o_p(1)$$

where $o_p(1)$ denotes convergence in probability.

PROOF. The proof of this lemma is given in Appendix. □

When the distribution of ϵ_t is not symmetric about zero, then the variable y_t in the model (1.1) is asymmetrically distributed about $f_t(\theta_0)$. So, to revise this circumstance we consider the new disturbance $\bar{\epsilon}_t$ which has the *pdf* given by

$$h(x) = \begin{cases} (1-\tau)a(\tau)g(x), & \text{if } x < 0, \\ \tau a(\tau)g(x), & \text{if } x \geq 0, \end{cases}$$

where $a(\tau) = 1/\{G(0) + \tau(1 - 2G(0))\}$. Specially, let

$$\tau = \frac{\int_{-\infty}^{0} xg(x)dx}{\int_{-\infty}^{0} xg(x)dx - \int_{0}^{\infty} xg(x)dx} = \frac{\int_{-\infty}^{0} |x|g(x)dx}{E|\epsilon|}. \tag{2.2}$$

Then, by simple calculations we know that the expected value of $\bar{\epsilon}_t$ is zero. This means that τ in (2.2) transforms the random variable with non-zero mean into a random variable with zero mean. Moreover, if error terms have a *pdf* which is symmetric about zero, then we get $\tau = 1/2$ and $a(\tau) = 2$. That is, the ALS estimators $\hat{\theta}_n(\tau)$ coincide with the LS estimators in this case.

For the main result of this section we require the following condition:

ASSUMPTION C. $V_n(\theta_0) = n^{-1}\sum_{t=1}^{n} \nabla f_t(\theta_0)\nabla^T f_t(\theta_0)$ converges to a positive definite matrix $V(\theta_0)$ as $n \to \infty$, where T denotes transpose of matrix.

In the following theorem we present sufficient conditions for strong consistency of the ALS estimators.

THEOREM 2.2. *Suppose that Assumptions A, B and C hold for the model (1.1). Then the ALS estimators $\hat{\theta}_n(\tau)$ converges almost surely to the true parameter θ_0.*

PROOF. For any $\delta > 0$, it is enough to show that

$$\lim_{n\to\infty} \inf_{\|\theta-\theta_0\|>\delta} \{Q_n(\theta;\tau)\} > 0 \quad a.e.$$

A detailed proof of this is given in Appendix. □

The following theorem provides asymptotic normality of the ALS estimators. The normality of the proposed estimators can be obtained by an expansion of $\nabla S_n(\theta;\tau)$ or a linear expansion of $f_t(\theta)$. In this paper we used the former expansion which is cited often in nonlinear models, see Seber and Wild (1989).

THEOREM 2.3. *Under the same conditions as in Theorem 2.2,* $\sqrt{n}(\hat{\theta}_n(\tau) - \theta_0)$ *converges in distribution to a p-variate normal vector with mean zero and variance-covariance matrix* $\omega(\tau)V^{-1}(\theta_0)$. *That is,*

$$\sqrt{n}(\hat{\theta}_n(\tau) - \theta_0) \xrightarrow{\mathcal{L}} N\big(0, \omega(\tau)V^{-1}(\theta_0)\big),$$

where

$$\omega(\tau) = \frac{(1-2\tau)m(\tau) + \tau^2(\sigma^2 + \mu^2)}{\big\{G(0) + \tau(1 - 2G(0))\big\}^2}$$

and $m(\tau) = \int_{-\infty}^{0} x^2 dG(x)$.

PROOF. From a first order Taylor series expansion we get

$$\nabla S_n(\hat{\theta}_n(\tau); \tau) - \nabla S_n(\theta_0; \tau) = \nabla^2 S_n(\theta_n^*(\tau); \tau)(\hat{\theta}_n(\tau) - \theta_0)$$

where $\|\theta_n^*(\tau) - \theta_0\| \le \|\hat{\theta}_n(\tau) - \theta_0\|$. To derive the normality of the ALS estimator we have to show that

$$\sqrt{n}\nabla S_n(\theta_0; \tau) \xrightarrow{\mathcal{L}} N\big(0, 4\omega_1 V(\theta_0)\big),$$

$$\nabla^2 S_n(\theta_0; \tau) - 2\omega_2 V(\theta_0) = o_p(1),$$

where $\omega_1 = (1-2\tau)m(\tau) + \tau^2(\sigma^2 + \mu^2)$ and $\omega_2 = G(0) + \tau(1 - 2G(0))$. The rest of the proof of this theorem is given in Appendix. □

Note that if ϵ_t has the normal distribution with mean zero and variance σ^2, then we obtain

$$\tau = 1/2 \text{ and } \omega = \sigma^2.$$

Hence, for the LS estimator $\hat{\theta}_n(1/2)$, Theorem 2.3 implies that

$$\sqrt{n}(\hat{\theta}_n(\tfrac{1}{2}) - \theta_0) \xrightarrow{\mathcal{L}} N\big(0, \sigma^2 V^{-1}(\theta_0)\big),$$

which is the same result as in Jennrich (1969) and Wu (1981).

REMARK. To distinguish between $\hat{\theta}_n(\tau_1)$ and $\hat{\theta}_n(\tau_2)$ for $\tau_1 \ne \tau_2$ we let the response function have an intercept term. That is,

$$f(x, \theta) = \theta_1 + \tilde{f}(x, (\theta_2, \ldots, \theta_p)). \tag{2.3}$$

For this, let the *pdf* of the new disturbance ϵ_t^* be defined by

$$h^*(x) = \begin{cases} (1-\tau^*)a(\tau^*)g(x), & \text{if } x < \mu(\tau^*), \\ \tau^* a(\tau^*)g(x), & \text{if } x \ge \mu(\tau^*), \end{cases}$$

where

$$\tau^* = \frac{\int_{-\infty}^{\mu(\tau^*)}(x-\mu(\tau^*))g(x)dx}{\int_{-\infty}^{\mu(\tau^*)}(x-\mu(\tau^*))g(x)dx - \int_{\mu(\tau^*)}^{\infty}(x-\mu(\tau^*))g(x)dx}$$
$$= \frac{\int_{-\infty}^{\mu(\tau^*)}|x-\mu(\tau^*)|g(x)dx}{E|\epsilon-\mu(\tau^*)|},$$

and $a(\tau^*) = 1/\left[G(\mu(\tau^*)) + \tau^*\{1-2G(\mu(\tau^*))\}\right]$. Then, an easy calculation shows that the disturbance ϵ_t^* has non-zero mean $\mu(\tau^*)$. Newey and Powell (1987) referred to $\mu(\tau^*)$ as the $(\tau^*)^{th}$ expectile and suggested the existence and the properties of $\mu(\tau^*)$.

Now, we obtain the interesting properties of the ALS estimators by converting $S_n(\theta;\tau)$ in (1.3), $Q_n(\theta;\tau)$ in (2.1) and Assumption B into $S_n(\theta;\tau^*), Q_n(\theta;\tau^*)$ and Assumption B*, respectively. We define

$$S_n(\theta;\tau^*) = \frac{1}{n}\sum_{t=1}^{n}\rho_\tau(y_t - f(x_t,\theta) - \mu(\tau^*)),$$
$$Q_n(\theta;\tau^*) = S_n(\theta;\tau^*) - S_n(\theta_0;\tau^*).$$

The following assumption takes the place of Assumption B in nonlinear models with intercept terms.

ASSUMPTION B*. The probability density function $g(x)$ is continuous and strictly positive at $\mu(\tau^*)$.

The following theorem provides sufficient conditions for the asymptotic properties of the ALS estimators in the model (2.3) and the proof of the following theorem is an easy modification of the proofs of Theorem 2.2 and Theorem 2.3. So, we omitted the proof here.

THEOREM 2.4. *Suppose that Assumptions A, B* and C hold for the model* (2.3). *Then we have*

$$\hat\theta_n(\tau^*) \xrightarrow{a.s.} (\theta_1 + \mu(\tau^*), \theta_2, \ldots, \theta_p),$$
$$\sqrt{n}(\hat\theta_n(\tau^*) - (\theta_1 + \mu(\tau^*), \theta_2, \ldots, \theta_p)) \xrightarrow{\mathcal{L}} N\big(0, \omega(\tau^*)V^{-1}(\theta_0)\big),$$

where

$$\omega(\tau^*) = \frac{(1-2\tau^*)m(\tau^*) + (\tau^*)^2\{\sigma^2 + (\mu-\mu(\tau^*))^2\}}{\{G(\mu(\tau^*)) + \tau^*(1-2G(\mu(\tau^*)))\}^2}$$

and $m(\tau^*) = \int_{-\infty}^{0}(x-\mu(\tau^*))^2 dG(x)$.

3. ASYMPTOTIC RELATIVE EFFICIENCY

In this section we consider asymptotic relative efficiency of the ALS estimators to the RQ estimators, based on the volumes of the corresponding confidence regions for a specified value of the limiting confidence coefficient. First, we investigate asymptotic normality of the RQ estimators to induce ARE of the ALS estimators to the RQ estimators. For this aim, we add the following assumption.

ASSUMPTION D. The probability density function $g(x)$ is continuously differentiable and strictly positive at $\beta = G(0)$.

The asymptotic normality of the RQ estimators is given in the following result which is conformable with Theorem 2.3 in Jurečková and Procházka (1994) in the case of nonlinear regression models without intercept terms. Choi *et al.* (2001) proposed sufficient conditions for the asymptotic properties of the RQ estimators.

THEOREM 3.1. *For the model (1.1), suppose that Assumptions A, C and D are fulfilled. Then the RQ estimators $\tilde{\theta}_n(\beta)$ has asymptotically normal distribution with mean zero and variance-covariance matrix $\eta V^{-1}(\theta_0)$, i.e.,*

$$\sqrt{n}(\hat{\theta}_n(\beta) - \theta_0) \xrightarrow{\mathcal{L}} N\big(0, \eta V^{-1}(\theta_0)\big), \qquad (3.1)$$

where $\eta = \{\beta(1-\beta)\}/(g(0))^2$.

PROOF. For a detailed proof, see Choi *et al.* (2001). □

As mentioned in Serfling (1980) and Sen and Singer (1993), asymptotic relative efficiency of two estimators is defined by the ratio of volumes of the corresponding confidence ellipsoids for a specified value of the limiting confidence coefficient. To compare the ALS with the RQ estimators, let

$$U_n(\tau) = n\omega^{-1}(\hat{\theta}_n(\tau) - \theta_0)^T V_n(\hat{\theta}_n(\tau))(\hat{\theta}_n(\tau) - \theta_0)$$

and

$$U_n(\beta) = n\eta^{-1}(\tilde{\theta}_n(\beta) - \theta_0)^T V_n(\tilde{\theta}_n(\beta))(\tilde{\theta}_n(\beta) - \theta_0).$$

On the other hand, the formula (3.1) and Theorem 2.3 imply that $U_n(\tau)$ and $U_n(\beta)$ have asymptotically a chi-square distribution with p degrees of freedom, denoted by χ_p^2. Let c_α be defined by $P(\chi_p^2 > c_\alpha) = \alpha$. Then the confidence regions

$$E_n(\tau) = \{\theta : U_n(\tau) < c_\alpha\}$$

and

$$E_n(\beta) = \{\theta : U_n(\beta) < c_\alpha\}$$

have asymptotic confidence coefficient $1 - \alpha$ and have volumes

$$\frac{\pi^{p/2}(c_\alpha/n)^{p/2}\omega^{1/2}}{\Gamma(p/2+1)|V_n(\hat{\theta}_n(\tau))|^{1/2}}$$

and

$$\frac{\pi^{p/2}(c_\alpha/n)^{p/2}\eta^{1/2}}{\Gamma(p/2+1)|V_n(\tilde{\theta}_n(\beta))|^{1/2}}.$$

Hence, from these results we have the next consequence.

THEOREM 3.2. *Under the same conditions as in Theorem 2.3 and Theorem 3.1, the asymptotic relative efficiency of the ALS estimators with respect to the RQ estimators is the ratio of* $\omega(\tau)$ *and* $\eta(\beta)$*, that is,*

$$\lim_{n\to\infty} e(\hat{\theta}_n(\tau)|\tilde{\theta}_n(\beta)) = \frac{\omega(\tau)}{\eta(\beta)}$$

where $\omega(\tau)$ *and* $\eta(\beta)$ *are given in Theorem 2.3 and Theorem 3.1, respectively.*

PROOF. The theorem immediately follows from Theorem 2.3 and Theorem 3.1. □

Now, we explain the ARE of the ALS estimators to the RQ estimators when the random errors have the following distributions.

EXAMPLE 3.1 (*Contaminated normal distribution*). The *cdf* of the disturbance ϵ_t is given by

$$G(x) = \lambda\Phi(x) + (1-\lambda)\Phi(x-1), \quad -\infty < x < \infty,$$

where λ is a constant in the interval $[0,1]$. Let $\lambda = 0.9$. Then, we get

$$\tau \fallingdotseq 0.3219, \ \beta = G(0) \fallingdotseq 0.4659, \ \omega \fallingdotseq 1.6902, \text{ and } \eta \fallingdotseq 1.6937.$$

Thus, we have

$$\lim_{n\to\infty} e(\hat{\theta}_n(\tau)|\tilde{\theta}_n(\beta)) = \frac{1.6902}{1.6937} < 1.$$

EXAMPLE 3.2 (*Contaminated double exponential distribution*). Let $\Xi(x:\gamma, \kappa)$ be double exponential distribution with mean γ and variance 2κ. Consider the following distribution

$$G(x) = \lambda\Xi(x:0,1) + (1-\lambda)\Xi(x:1,1), \qquad -\infty < x < \infty.$$

If $\lambda = 0.7$, simple calculations show that

$$\tau \risingdotseq 0.5611 \text{ and } \beta = G(0) \risingdotseq 0.9018.$$

Also, since

$$\omega \risingdotseq 0.8201 \text{ and } \eta \risingdotseq 0.1081$$

we obtain

$$\lim_{n\to\infty} e(\hat{\theta}_n(\tau)|\tilde{\theta}_n(\beta)) = \frac{0.8201}{0.1081} > 1.$$

EXAMPLE 3.3 (*Mixed distribution of normal and double exponential*). Consider the following distribution

$$G(x) = \lambda\Phi(x) + (1-\lambda)\Xi(x:1,1), \qquad -\infty < x < \infty.$$

If $\lambda = 0.9$, by a similar method we get

$$\tau \risingdotseq 0.1727 \text{ and } \beta = G(0) \risingdotseq 0.4684.$$

From the result

$$\omega \risingdotseq 0.0789 \text{ and } \eta \risingdotseq 1.02,$$

we have

$$\lim_{n\to\infty} e(\hat{\theta}_n(\tau)|\tilde{\theta}_n(\beta)) \risingdotseq \frac{0.0789}{1.02} < 1.$$

Theorem 3.2 and the above examples imply that the ALS estimators is more efficient than the RQ estimators when the distribution of errors is a contaminated normal or is close to normal.

4. Monte Carlo Simulation

In this section, to compare the four estimators (LS, LAD, RQ, ALS) we perform a Monte Carlo simulation. We consider in this experiment the following model

$$y_t = \theta_1 e^{-\theta_2 x_t} + \epsilon_t, \quad t = 1, \dots, 15. \tag{4.1}$$

The disturbance ϵ_t are generated as random variates from one of the following distribution:

1. the standard normal distribution, denoted by $N(0,1)$;
2. the double exponential distribution with mean zero and variance 18, denoted by $DE(0,3)$;
3. the normal distribution with mean one and variance one, denoted by $N(1,1)$;
4. the double exponential distribution with mean one and variance 18, denoted by $DE(1,3)$.

In all cases the parameter θ_1 is set to equal $1,000$, the other parameter θ_2 equals to 1.5, and x_t equals to t. The following table reports the results for the four distributions when $1,000$ simulation runs were executed. In each cell, the first figure gives the mean of the estimates and the second figure presents the average of the mean squared errors.

TABLE 4.1 *Estimate and mean square error*

Distribution	*LSE*	*LAD*	*ALS*	*RQ*
N(0,1)	**1.500006**	1.500055	1.50511	1.502926
	(1.573831e-05)	(1.879849e-05)	(5.4532598e-05)	(4.532599e-05)
DE(0;3)	1.499257	**1.499345**	1.503353	1.501119
	(0.000322)	(0.000391)	(0.000346)	(0.00041)
N(1,1)	1.494144	1.495652	**1.497482**	1.497086
	(4.983533e-05)	(3.764651e-05)	(2.747208e-05)	(3.396625e-05)
DE(1;3)	1.493909	1.495516	1.495293	**1.495987**
	(6.595429e-05)	(5.443909e-05)	(5.21137e-05)	(5.25472e-05)

In Table 4.1, with respect to the average of the estimates we see that the ALS estimator performs better than the RQ estimator when the error has an asymmetric normal distribution. On the other hand, in the case of an asymmetric and heavy-tailed distribution the RQ estimator is superior to the ALS estimator.

From Theorem 3.2, Table 4.1 and examples in Section 3, we obtain the following table which identifies the estimator with the best performance for each error distribution.

TABLE 4.2 *Best estimator*

Best estimator	*error distribution*
LSE	symmetric normal distribution
ALS	asymmetric normal distribution
LAD	symmetric and heavy-tailed distribution
RQ	asymmetric and heavy-tailed distribution

5. CONCLUSIONS

In this paper we proposed the ALS estimators for the regression parameters in nonlinear regression models with either grossly skewed or contaminated error distribution. The proposed estimators have been proved to be strongly consistent and asymptotically normal under some mild conditions. From the asymptotic efficiency based confidence regions and a large simulation study we concluded that the ALS estimators perform better than the RQ estimators when the distribution of errors is a contaminated normal or is close to normal. In the case of an asymmetric and heavy-tailed distribution, the RQ estimation is superior to the ALS.

APPENDIX

PROOF OF LEMMA 2.1. Let

$$A_t(\tau) = |\tau - 1(y_t < f_t(\theta))|(\epsilon_t - d_t(\theta))^2 - |\tau - 1(y_t < f_t(\theta_0))|\epsilon_t^2.$$

Then $A_t(\tau)$ equals to

$$\begin{cases} (1-\tau)(2\epsilon_t - d_t(\theta))(-d_t(\theta)), & y_t < f_t(\theta), \quad y_t < f_t(\theta_0), \\ \tau(2\epsilon_t - d_t(\theta))(-d_t(\theta)), & y_t > f_t(\theta), \quad y_t > f_t(\theta_0), \\ (1-\tau)d_t^2(\theta) - (\tau - 1)2\epsilon_t d_t(\theta) + (1-2\tau)\epsilon_t^2, & y_t < f_t(\theta), \quad y_t > f_t(\theta_0), \\ \tau d_t^2(\theta) - \tau 2\epsilon_t d_t(\theta) + (2\tau - 1)\epsilon_t^2, & y_t > f_t(\theta), \quad y_t < f_t(\theta_0). \end{cases}$$

Moreover,

$$|A_t(\tau)| \le (|2\epsilon_t| + |d_t(\theta)|)|d_t(\theta)|$$

or

$$|A_t(\tau)| \le |d_t(\theta)|^2 + 2|\epsilon_t| \cdot |d_t(\theta)| + |\epsilon_t|^2.$$

From the assumption of the error terms, $E(\epsilon_t^2) < \infty$, and Assumption A, we know that there exists $M > 0$ such that $|A_t(\tau)| < M$. In virtue of Chebyshev inequality we obtain

$$P\Big\{|Q_n(\theta;\tau) - EQ_n(\theta;\tau)| > \varepsilon\Big\} \le \frac{\sum_{t=1}^n E_\epsilon(A_t(\tau) - EA_t(\tau))^2}{n^2\epsilon^2} \le \frac{\max_{1\le t\le n}\mathrm{Var}A_t(\tau)}{n\epsilon^2}.$$

Hence, the proof is completed. □

PROOF OF THEOREM 2.2. From Lemma 2.1 we get

$$Q_n(\theta\,;\tau) = \frac{1}{n}\sum_{t=1}^n E_\epsilon A_t(\tau) + o_p(1).$$

First, we prove that θ_0 is a local minimizer of

$$Q(\theta;\tau) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n E_\epsilon Q_n(\theta;\tau).$$

By simple calculations we have

$$\begin{aligned}\nabla Q(\theta;\tau) &= \lim_{n\to\infty} -\frac{2}{n}\sum_{t=1}^n \int_{-\infty}^{\infty} \big\{|\tau - 1(\lambda < d_t(\theta))|(\lambda - d_t(\theta))\big\} dG_t(\lambda)\nabla f_t(\theta) \\ &= \lim_{n\to\infty} -\frac{2}{n}\sum_{t=1}^n \Bigg\{\int_{-\infty}^{d_t(\theta)} (1-\tau)(\lambda - d_t(\theta))dG_t(\lambda) \\ &\qquad + \int_{d_t(\theta)}^{\infty} \tau(\lambda - d_t(\theta))dG_t(\lambda)\Bigg\}\nabla f_t(\theta)\end{aligned}$$

and

$$\begin{aligned}\nabla^2 Q(\theta;\tau) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \Bigg[&2\Bigg\{\int_{-\infty}^{d_t(\theta)} (1-\tau)dG_t(\lambda) + \int_{d_t(\theta)}^{\infty}\tau dG_t(\lambda)\Bigg\}\nabla f_t^T(\theta)\nabla f_t(\theta) \\ &-2\Bigg\{\int_{-\infty}^{\infty} |\{\tau - 1(\lambda < d_t(\theta))\}|(\lambda - d_t(\theta))dG_t(\lambda)\Bigg\}\nabla^2 f_t(\theta)\Bigg].\end{aligned}$$

Let $\delta(\tau) = \min\{\tau, 1-\tau\}$. Then we can choose $\eta > 0$ such that $0 < \eta < \delta(\tau) < 1$. Since

$$\int_{-\infty}^{0} (1-\tau) dG_t(\lambda) + \int_{0}^{\infty} \tau dG_t(\lambda) > \eta > 0,$$

$\nabla^2 Q(\theta_0;\tau)$ is a positive definite matrix. Hence $\nabla Q(\theta_0;\tau) = 0$ implies that θ_0 is a local minimizer of $Q(\theta;\tau)$.

Second, we show that this local minimizer θ_0 is indeed the global minimizer. Let $N_\delta(\theta_0) = \{\theta : \|\theta - \theta_0\| < \delta\}$. Then, from the fact that $R^*(\theta) = N_\delta^c(\theta_0) \cap \Theta$ is compact we have θ^* such that

$$Q(\theta^*;\tau) = \inf_{\theta \in R^*(\theta)} Q(\theta;\tau).$$

If $d_t(\theta) > 0$, we obtain

$$\begin{aligned} Q(\theta;\tau) &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E_\epsilon\{|\tau - 1(\epsilon_t < d_t(\theta))|(\epsilon_t - d_t(\theta))^2 - |\tau - 1(\epsilon_t < 0)|\epsilon_t^2\} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E_\epsilon B_t(\tau) \end{aligned}$$

where

$$B_t(\tau) = \int_{-\infty}^{\infty} \{|\tau - 1(\lambda < d_t(\theta))|(\lambda - d_t(\theta))^2 - |\tau - 1(\lambda < 0)|\lambda^2\} dG_t(\lambda).$$

By simple calculations and (2.2), $B_t(\tau)$ equals to

$$\begin{aligned} &(1-\tau)\left\{\int_0^{d_t(\theta)} (d_t(\theta) - \lambda)^2 dG_t(\lambda) + \int_{-\infty}^{0} d_t^2(\theta) dG_t(\lambda)\right\} \\ &+ \tau\left\{\int_0^{\infty} d_t^2(\theta) dG_t(\lambda) - \int_0^{d_t(\theta)} (d_t(\theta) - \lambda)^2 dG_t(\lambda)\right\}. \end{aligned}$$

Moreover, $-\lambda(\lambda - 2d_t(\theta))$ is positive on $(0, d_t(\theta))$. Hence $B_t(\tau)$ is strictly positive and

$$\inf_{\|\theta - \theta_0\| > \delta} Q_n(\theta;\tau) > 0. \tag{A.1}$$

By a similar method we have the same conclusion in the case where $d_t(\theta) < 0$. Hence, the theorem follows from Assumption (A3) and (A.1). □

PROOF OF THEOREM 2.3. First, we consider

$$\sqrt{n}\nabla S_n(\theta_0;\tau) \xrightarrow{\mathcal{L}} N\big(0, 4\omega_1 V(\theta_0)\big).$$

Let $c_t(\tau) = -2|\tau - 1(\epsilon_t < 0)|\epsilon_t$ and $d_t(\tau) = 2|\tau - 1(\epsilon_t < 0)|$. Then, from (1.3) we obtain

$$\sqrt{n}\nabla S_n(\theta_0;\tau) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} c_t(\tau)\cdot\nabla f_t(\theta_0).$$

Also, in virtue of (2.2) we have $E(c_t(\tau)) = 0$. Since the disturbance ϵ_t are independent, we obtain

$$\begin{aligned}\mathrm{Var}(c_t(\tau)) &= 4\Big\{\int_{-\infty}^{0} x^2 dG_t(x) - 2\tau\int_{-\infty}^{0} x^2 dG_t(x) + \tau^2\int_{-\infty}^{\infty} x^2 dG_t(x)\Big\}\\ &= 4\big\{(1-2\tau)m(\tau) + \tau^2(\sigma^2+\mu^2)\big\}.\end{aligned}$$

Let $D = (d_1, \ldots, d_p)^T$ be any nonzero vector. Then

$$D^T\sqrt{n}\nabla S_n(\theta_0;\tau) = \sum_{t=1}^{n}\sum_{k=1}^{p}\frac{1}{\sqrt{n}}c_t(\tau)d_k(f_t)_k(\theta_0)$$

where $(f_t)_k(\theta_0) = \partial f_t(\theta_0)/\partial\theta_k$. Let $Z_{nt} = n^{-1/2}b_t c_t(\tau)$ and $b_t = \sum_{k=1}^{p} d_k(f_t)_k(\theta_0)$. Then, $E(Z_{nt}) = 0$ and $\mathrm{Var}(Z_{nt}) = n^{-1}b_t^2\,\mathrm{Var}(c_t(\tau))$. Let $B_n^2 = \sum_{t=1}^{n}\mathrm{Var}(Z_{nt}) = n^{-1}\sum_{t=1}^{n} b_t^2\mathrm{Var}(c_t(\tau))$. For arbitrary $\varepsilon > 0$

$$\frac{1}{B_n^2}\sum_{t=1}^{n} E[Z_{nt}^2 I_{\{|Z_{nt}|>\varepsilon B_n\}}(x)]$$

is less than

$$\frac{1}{\sum_{t=1}^{n} b_t^2\mathrm{Var}(c_t(\tau))}\sum_{t=1}^{n} E\left[b_t^2 c_t^2(\tau) I_{\left\{|c_t(\tau)|>\frac{\varepsilon}{b_t^*}\sqrt{\sum_{t=1}^{n} b_t^2\mathrm{Var}(c_t(\tau))}\right\}}(x)\right],$$

where $b_t^* = \max_{1\le t\le n} b_t$. Therefore,

$$\frac{1}{B_n^2}\sum_{t=1}^{n} E[Z_{nt}^2 I_{\{|Z_{nt}|>\varepsilon B_n\}}(x)]$$

converges to zero because $(b_t^*)^{-1}\sqrt{n\sum_{t=1}^{n}\mathrm{Var}(Z_{nt})}$ diverges to ∞ as $n\to\infty$. Hence, by the Lindeberg form of central limit theorem

$$\frac{\sum_{t=1}^{n} Z_{nt}}{B_n} \xrightarrow{\mathcal{L}} N(0,1).$$

That is,

$$\frac{\sqrt{n}D^T\nabla S_n(\theta_0;\tau)}{\sum_{t=1}^n \mathrm{Var}(Z_{nt})} \xrightarrow{\mathcal{L}} N(0,1).$$

Since

$$\sum_{t=1}^n \mathrm{Var}(Z_{nt}) = \frac{1}{n}\sum_{t=1}^n D^T\nabla f_t(\theta_0)\nabla^T f_t(\theta_0)D\mathrm{Var}(c_t(\tau)),$$

we have

$$\sqrt{n}D^T\nabla S_n(\theta_0;\tau) \xrightarrow{\mathcal{L}} N\big(0, 4\omega_1 D^T V(\theta_0)D\big).$$

Thus, Cramér-Wold device implies that

$$\sqrt{n}\nabla S_n(\theta_0;\tau) \xrightarrow{\mathcal{L}} N\big(0, 4\omega_1 V(\theta_0)\big).$$

Next, to prove the second result we show that

$$\nabla^2 S_n(\theta_0;\tau) = \frac{1}{n}\sum_{t=1}^n \{d_t(\tau)\nabla f_t(\theta_0)\nabla^T f_t(\theta_0) - c_t(\tau)\nabla^2 f_t(\theta_0)\}$$

converges to $2\{(1-\tau)G(0)+\tau(1-G(0))\}V(\theta_0)$. Let

$$\begin{aligned} \mathrm{I}_1 &= \frac{1}{n}\sum_{t=1}^n d_t(\tau)\nabla f_t(\theta_0)\nabla^T f_t(\theta_0), \\ \mathrm{I}_2 &= \frac{1}{n}\sum_{t=1}^n c_t(\tau)\nabla^2 f_t(\theta_0). \end{aligned}$$

Since $c_t(\tau)$ has mean zero and a finite variance, Chebyshev inequality implies that

$$\mathrm{I}_2 = \frac{1}{n}\sum_{t=1}^n c_t(\tau)\nabla^2 f_t(\theta_0)$$

converges to 0 as $n \to \infty$. Let

$$U_t^{ij}(\theta_0) = d_t(\tau)(f_t)_i(\theta_0)(f_t)_j(\theta_0).$$

Then, for an arbitrary $\varepsilon > 0$

$$\begin{aligned} P\left[\frac{1}{n}\left|\sum_{t=1}^n \{U_t^{ij}(\theta_0) - E(U_t^{ij}(\theta_0))\}\right| > \varepsilon\right] &\le \frac{1}{n^2\varepsilon^2}\sum_{t=1}^n \mathrm{Var}(U_t^{ij}(\theta_0)) \\ &\le \frac{\max_{1\le t\le n}\mathrm{Var}(U_t^{ij}(\theta_0))}{n\varepsilon^2}. \end{aligned}$$

The fact

$$\frac{1}{n}\left|\sum_{t=1}^{n}\{U_t^{ij}(\theta_0) - E(U_t^{ij}(\theta_0))\}\right| = o_p(1)$$

follows from Chebyshev inequality. Thus,

$$\frac{1}{n}\sum_{t=1}^{n} 2|\tau - 1(\epsilon_t < 0)|\nabla f_t(\theta_0)\nabla^T f_t(\theta_0)$$

converges to $E\{2|\tau - 1(\epsilon_t < 0)|\}V(\theta_0)$ as $n \to \infty$. On the other hand, since

$$E\{2|\tau - 1(\epsilon_t < 0)|\} = 2\{(1-\tau)G(0) + \tau(1 - G(0))\},$$

we get

$$\mathrm{I}_1 - 2\{(1-\tau)G(0) + \tau(1 - G(0))\}V(\theta_0) = o_p(1).$$

By the first and second results, we have

$$\sqrt{n}(\hat{\theta}_n(\tau) - \theta_0) \xrightarrow{\mathcal{L}} N\big(0, \frac{\omega_1}{\omega_2}V^{-1}(\theta_0)\big).$$

Hence, the proof of Theorem 2.3 is completed. □

Acknowledgements

We thank to the two referees for their constructive suggestions and comments, which led to improvements of the earlier version of this paper.

References

Basset, G. and Koenker, R. (1982). "An empirical quantile function for linear models with iid errors", *Journal of the American Statistical Association*, **77**, 407–415.

Basset, G. and Koenker, R. (1986). "Strong consistency of regression quantile and related empirical process", *Econometric Theory*, **2**, 633–643.

Buchinsky, M. (1998). "Recent advances in quantiles regression model", *The Journal of Human Resources*, **27**, 88–126.

Choi, S. H., Kim, H. K. and Park, K. O. (2001). "Nonlinear regression quantiles estimation", *Journal of the Korean Statistical Society*, **30**, 551–556.

Jennrich, R. (1969). "Asymptotic properties of nonlinear least squares estimations", *The Annals of Mathematical Statistics*, **40**, 633–643.

Jurečková, J. and Procházka, B. (1994). "Regression quantiles and trimmed least squares estimators in nonlinear regression model", *Nonparametric Statistics*, **3**, 201–222.

Koenker, R. and Basset, G. (1978). "Regression Quantiles", *Econometrica*, **46**, 33–50.

Newey, W. K. and Powell, L. (1987). "Asymmetric least squares estimation and testing", *Econometrica*, **55**, 819–847.

OBERHOFER, W. (1982). "The consistency of nonlinear regression minimizing the L_1-norm", *The Annals of Statistics*, **10**, 316–319.

SEBER, G. A. F. AND WILD, C. J. (1989). *Nonlinear Regression*, John Wiley & Sons.

SEN, P. K. AND SINGER, J. M. (1993). *Large Sample Methods in Statistics*, Chapman and Hall.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.

WU, C. F. J. (1981). "Asymptotic theory of nonlinear least squares estimation", *The Annals of Statistics*, **9**, 501–513.