Commun. Korean Math. Soc. **20** (2005), No. 1, pp. 145–159

ROBUST TEST BASED ON NONLINEAR REGRESSION QUANTILE ESTIMATORS

Seung Hoe Choi, Kyung Joong Kim and Myung Sook Lee

Abstract. In this paper we consider the problem of testing statistical hypotheses for unknown parameters in nonlinear regression models and propose three asymptotically equivalent tests based on regression quantiles estimators, which are Wald test, Lagrange Multiplier test and Likelihood Ratio test. We also derive the asymptotic distributions of the three test statistics both under the null hypotheses and under a sequence of local alternatives and verify that the asymptotic relative efficiency of the proposed test statistics with classical test based on least squares depends on the error distributions of the regression models. We give some examples to illustrate that the test based on the regression quantiles estimators performs better than the test based on the least squares estimators of the least absolute deviation estimators when the disturbance has asymmetric and heavy-tailed distribution.

1. Introduction

Consider the following nonlinear regression model

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

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 $\theta_o = (\theta_1, \cdots, \theta_p)$ belong to a compact parameter space $\Theta \in \mathbb{R}^p$ and the response function $f(x, \theta)$ is a continuous on $\mathbb{R}^q \times \mathbb{R}^p$. We assume throughout that the disturbance $\{\epsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with common distribution function (df) $G(x)$ and a finite variance. Suppose that $G(x)$ possesses a continuous probability density function (pdf) $g(x)$ and the response function could be written in the form $f(x, \theta) = \theta_1 + \tilde{f}(x, (\theta_2, \ldots, \theta_p))$.

Received May 27, 2004.
2000 Mathematics Subject Classification: 62J02, 62G05.
Key words and phrases: nonlinear regression quantiles estimators, Wald test, Lagrange multiplier test, likelihood ratio test.

A statistical problem in regression model is to make inference about true parameter θ_o in some optimal way on the basis of observations $\{(x_t, y_t) : t = 1, \cdots, n\}$. The Least Square (LS) estimators plays an important role in estimation and testing hypotheses about unknown regression parameter θ. Asymptotic results for nonlinear LS estimators are given by many authors, see Seber (1989). However, in spite of the theoretical and practical merits, owing to the sensitivity of the LS estimators to modest amounts of outlier contamination the classical test statistic based on LS estimators does lose power in many non-Gaussian, especially heavy-tailed and asymmetric error distribution or contaminated distribution. In this case it is need to develop alternative test statistic based on robust estimators. Thus, during the past few years there has been an increased interest in robust estimation procedures applied to the regression model, see Huber (1981). Koenker and Basset (1978) introduced the $\beta-$Regression Quantiles (RQ) estimators, denoted by $\hat{\theta}_n(\theta)$, are defined as the value of θ minimizing the following function

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

where the "check function"

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

$G(\eta_\beta) = \beta$ and $0 < \beta < 1$. Moreover, the Least Absolute Deviation (LAD) estimators are easily seen to be a special case of the β-regression quantile when $\beta = \frac{1}{2}$. Wang (1995) and Jurečkovά and Procházka (1994) have investigated asymtotic behavior of RQ estimators in nonlinear regression models. Especially, Koenker and Basett (1982) have established the test of hypothesis based on RQ estimators when the response function is linear.

We focus in this paper on the problem of testing statistical hypotheses for unknown parameter in models (1.1). For these, in section 2 we construct a class of test statistics based on RQ estimators, which are Wald test, Lagrange Multiplier test and Likelihood Ratio test and investigate the asymptotic behavior of the statistics both under the null hypotheses and under a sequence of local alternatives. In next section, we examine the Asymptotic Relative Efficiency (ARE) of the proposed test statistics with classical test based on LS estimators for various error distribution and some examples are given to illustrate the application of the main results.

2. Test statistics based on RQ estimators:

In this section we propose three test statistics based on RQ estimators defined by (1.2) and verify that under appropriate conditions the test statistic has chi-square distribution at limit.

Let (R^m, Ξ, P_x) denotes a probability space and u be a function from R^p to R^q such that $\nabla u(\theta) = [\frac{\partial}{\partial \theta_i} u(\theta)]_{(q \times p)}$ has full rank q and $\frac{\partial}{\partial \theta_i} u(\theta)$ is continuous on parameter space Θ.

Our interest in this paper is to test a nonlinear hypothesis $u(\theta_o) = 0$. Also, in order to investigate the asymptotic power of the test statistic we consider a sequence of local alternative tending to null hypothesis at a certain rate, denoted by

$$H_n : u(\theta_o) = \frac{\gamma}{\sqrt{n}} \quad \text{where} \quad \gamma \in R^q.$$

Throughout this paper we make the following assumption on the model (1.1).

Assumption A

A_1 : For each t the response function $f_t(\theta)$ is continuously differentiable up to order 2 on Θ and x_t is bounded in probability, i.e, for every $\eta > 0$, there exists M_η such that $P\{|x_t| > M_\eta\} < \eta$ where $f_t(\theta) = f(x_t, \theta)$.

A_2 : The pdf $g(x)$ is strictly positive at $\eta_\beta = G(\beta)$.

A_3: The matrix $V_n(\theta_o(\beta)) = \frac{1}{n}\sum_{t=1}^{n} \nabla^T f_t(\theta_o(\beta)) \nabla f_t(\theta_o(\beta))$ converges to a positive definite matrix $V(\theta_o(\beta))$ as $n \to \infty$ where T denotes transpose of the matrix and $\theta_o(\beta) = (\theta_1 - \eta_\beta, \theta_2, \cdots, \theta_p)$.

A_4 : The ratio of the number of elements of the set $K_n(\beta) = \{t : f_t(\theta) \neq f_t(\theta_o(\beta)), \theta \neq \theta_o(\beta)\}$ to the sample size n in model (1.1), denoted by $\frac{k_n(\beta)}{n}$, converges to $k(\beta)$ where $0 < k(\beta) \leq 1$.

The first test statistics we considered in this paper is Wald (W) test which depends solely on nonlinear RQ estimators of unrestricted model and is defining as usual by

$$W_n(\beta) = n\tau^2(\beta)\hat{u}_n^T(\beta)[\nabla \hat{u}_n^T(\beta) V_n^{-1}(\hat{\theta}_n(\beta)) \nabla \hat{u}_n(\beta)]^{-1} \hat{u}_n(\beta)$$

where $\tau(\beta) = \frac{g(\eta_\beta)}{\sqrt{\beta(1-\beta)}}$ and $\hat{u}_n(\beta)$ and $\nabla \hat{u}_n(\beta)$ denote $u(\beta)$ and $\nabla u(\beta)$ evaluated at $\hat{\theta}_n(\beta)$, respectively.

The second statistic we used is Lagrange Multiplier (LM) test (sometimes also called the score test) which is defined on the estimated regression coefficient of restricted model and is defined by a multiplication of the gradient and the Hessian matrix of the objective function $R_n(\theta : \beta)$. However, since the component of the gradientof the function $R_n(\theta : \beta)$ is not a continuous, in order to define LM-test well it is necessary to approximate to the function $\rho_\beta(x)$ by continuously differential function. For this aim, we consider the following function

$$(\rho_\beta)_n(x) = \begin{cases} \beta x & \text{if} \quad x > \alpha_n \\ \frac{x^2}{4\alpha_n} + (\beta - \frac{1}{2})x + \frac{\alpha_n}{4} & \text{if} \quad |x| \le \alpha_n \\ (\beta - 1)x & \text{if} \quad x < -\alpha_n \end{cases}$$

where $n\alpha_n$ converges to 0 as $n \to \infty$.

Let

$$\tilde{R}_n(\theta : \beta) = \frac{1}{n}\sum_{t=1}^{n}(\rho_\beta)_n(y_t - f_t(\theta) - \eta_\beta)$$

and

$$\tilde{D}_n(\theta : \beta) = \frac{1}{n}\sum_{t=1}^{n}(\psi_\beta)_n(y_t - f_t(\theta) - \eta_\beta)\nabla f_t(\theta)$$

where $(\psi_\beta)_n(x) = (\rho_\beta)'_n(x)$. Also,

$$D_n(\theta : \beta) = \frac{1}{n}\sum_{t=1}^{n}(\psi_\beta)(y_t - f_t(\theta) - \eta_\beta)\nabla f_t(\theta)$$

which is the gradient of $R_n(\theta : \beta)$ is asymptotically equivalent to $\tilde{D}_n(\theta : \beta)$ where $(\psi_\beta) = 1 - \beta, 0, \text{or } \beta$ according as $x \le 0, x = 0, \text{or } x > 0$. See, lemma 2 in Appendix.

The LM-test in this paper is defined by

$$M_n(\beta) = \frac{n}{\beta(1-\beta)}\hat{D}^n_o(\beta)^T V_n{}^{-1}(\hat{\theta}^o_n(\beta))\hat{D}^n_o(\beta).$$

where $\hat{\theta}^o_n(\beta)$ is minimizer of $R_n(\theta : \beta)$ on $\Theta_o = \{\theta : u(\theta) = 0\}$, and $\hat{D}^n_o(\beta) = D_n(\hat{\theta}^o_n(\beta) : \beta)$.

The last test statistic is the Likelihood Ratio (LR) test which is based on the difference between the object function $R_n(\theta : \beta)$ in restricted and unrestricted model. Generally, to drive limiting distribution of LR-test we use a linear Taylor expansion or the quadratic approximation for $R_n(\theta : \beta)$. But, since $\rho_\beta(x)$ is not differentiable at zero we have to consider how to a quadratic approximation to the function $R_n(\theta : \beta)$. Now, using the quadratic form for $\tilde{R}_n(\theta : \beta)$ we define the following function

$$Q_n(\theta:\beta) = R_n(\theta_0(\beta):\beta) + (\theta-\theta_0(\beta))^T D_n(\theta_0(\beta):\beta) + \frac{g(\eta_\beta)}{2}(\theta-\theta_0(\beta))^T V_n(\theta_0(\beta))(\theta-\theta_0(\beta)).$$

In this paper the LR-test is given by

$$R_n(\beta) = n\frac{g(\eta_\beta)}{\beta(1-\beta)}[\hat{R}_n(\beta) - \hat{R}_n^o(\beta)]$$

where $\hat{R}_n(\beta)$ and $\hat{R}_n^o(\beta)$ denote $R_n(\theta:\beta)$ evaluate at $\hat{\theta}_n(\beta)$ and $\hat{\theta}_n^o(\beta)$, respectively.

The following theorem is the main result of this paper explains that asymptotically all three test statistics which are W-test, LM-test and LR-test have the same distribution and the same power. Although the three test are asymptotic equivalent, we can choose test statistic based on convenience. That is, the Wald test can be calculated using the unrestricted model and the Lagrange Multiplier test uses only the restricted model. On the other hand, the Likelihood Ratio test depends on the restricted and unrestricted estimators.

For the power of proposed tests we need the following additional assumption.

Assumption B

B_1 : There exists a fixed number $\gamma_n(\beta) \in R^p$ such that $u(\theta_o(\beta)) = \frac{\gamma_n(\beta)}{\sqrt{n}}$ and $\gamma_n(\beta)$ converges to $\gamma(\beta)$ for any sample of size n.

THEOREM 2.1 *Suppose that Assumptions A and B are satisfied on model* (1.1). *Then the three test statistics, which are W-test, LM-test and LR-test, have asymptotically a chi-square distribution with q degree of freedom and noncentrality parameter λ denoted by $\chi^2(q,\lambda)$ where*

$$\lambda = \frac{\tau^2(\beta)}{2}\gamma^T(\beta)[\nabla u(\beta)V^{-1}(\theta_o(\beta))\nabla u^T(\beta)]\gamma(\beta).$$

Under the null hypothesis we have $\lambda = 0$ where $\nabla u(\beta) = \lim_{n\to\infty}\nabla\hat{u}_n(\beta)$ and $\nabla\hat{u}_n(\beta) = \nabla u(\hat{\theta}_n(\beta))$.

The test statistics mentioned above except for LM-test involve a nuisance parameter $g(\eta_\beta)$ which is the ordinate of the pdf $g(x)$ at η_β. It is necessary to estimate the parameter $g(\eta_\beta)$ to test a null hypothesis $H_0 : u(\theta_0(\beta)) = 0$ using the test statistics based on nonlinear RQ estimators. Liu (1992) has introduced and studied the consistent estimators for $g(\eta_\beta)$. If we have a consistent estimators $\hat{g}(\eta_\beta)$ of $g(\eta_\beta)$, we reject or accept

$$H_0 : u(\theta_0(\beta)) = 0 \text{according as} LR_n(\beta) \geq \text{or} \leq \chi^2_{1-\alpha}(q, \alpha)$$

with α level of significance.

3. Asymptotic relative efficiency

In this section we consider the ARE of the proposed tests on RQ estimators with respect to a classical test statistics derived from the asymptotic normality of nonlinear LS estimators.

For this purpose, we note [Gallant (1987)] that under some conditions

$$\sqrt{n}(\tilde{\theta}_n - \theta_0(\beta)) \xrightarrow{L} N(0, \sigma^2 V^{-1}(\theta_0(\beta)))$$

and

$$\tilde{\sigma}_n^2 = \sum_{t=1}^{n} (y_t - f_t(\tilde{\theta}_n))^2 \quad \text{converges to} \quad \sigma^2$$

where $\tilde{\theta}_n$ is nonlinear LS estimators and σ^2 is the variance of the error. See. Gallant (1987). Meanwhile, let

$$S_n(\tilde{\theta}_n) = \min_{\Theta} \sum_{t=1}^{n} (y_t - f_t(\tilde{\theta}_n))^2 \text{ and } S_n(\tilde{\theta}_n^0) = \min_{\Theta_0} \sum_{t=1}^{n} (y_t - f_t(\tilde{\theta}_n))^2.$$

Then, we know that

$$LS_n = n\sigma^{-2}[S_n(\tilde{\theta}_n) - S_n(\tilde{\theta}_n^0)]$$

converges to $\chi^2(q, \lambda)$ where

$$\lambda = \frac{\sigma^2}{2} \gamma^T(\beta)[\nabla u(\beta) V^{-1}(\theta_0(\beta)) \nabla u^T(\beta)]\gamma(\beta).$$

Usually, the ARE of the given test statistic is defined [Lehmann (1986)] by the ratio of the number of observation required by the test statistics to obtain the same power against the same alternative. According to this definition we have the following result.

THEOREM 3.1 *Under the same conditions of Theorem 2.1, the asymptotic relative efficiency of the tests based on RQ estimators with respect to the test based on least squares estimators in nonlinear regression model is the ratio of $\tau(\beta)$ and σ.*

The following examples illustrate the ARE of the test statistic based on RQ estimators in nonlinear regression model when the pdf has well known function. Suppose that the disturbance $\{\epsilon_t\}$ are i.i.d. with common underlying contaminated function of $h_1(x : \mu, \sigma)$ and $h_2(x : \alpha, \beta)$ where

$$h_1(x : \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

$$h_2(x : \alpha, \beta) = \frac{1}{2\beta} \exp\left[-\frac{|x-\alpha|}{\beta}\right]$$

and $-\infty < x < \infty$.

EXAMPLE. Consider the following probability density functions:
Case 1 $g_1(x) = 0.9h_1(x : 0, 1) + 0.1h_1(x : 1, 1)$.
Case 2 $g_2(x) = 0.2h_1(x : 0, 1) + 0.8h_1(x : 1, 2)$.
Case 3 $g_3(x) = 0.9h_2(x : 0, 1) + 0.1h_2(x : 1, 1)$.

When $G(\beta) = 0$, simple calculation shows that

Table 3.1. ARE of Proposed Test

Distribution	$g_1(x)$	$g_2(x)$	$g_3(x)$
LS(σ)	0.91	1.612	1.2806
LAD($\tau(\frac{1}{2})$)	1.3045	2.25	1.0671
RQ($\tau(\beta)$)	1.3077	1.09	0.82

From Theorem 3.1 and example above we conclude that the test based on RQ estimators performs better than the test based on LS estimators or LAD estimators when the error has asymmetric and heavy-tailed distribution.

4. Conclusion

In this paper we considered the test statistics based on nonlinear RQ estimators which are Wald test, Lagrange Multiplier test and Likelihood Ratio test and proved that the proposed test statistics are asymptotically equivalent. We also verified that the asymptotic relative efficiency of the proposed test statistics with classical test based on least squares depends on the error distribution of the regression model and that in the case of asymmetric and heavy-tailed distribution the test based on RQ estimation is superior to the test based on LS estimators or LAD estimators.

5. Appendix

To obtain the asymptotic distribution of W-test we need the asymptotic properties of RQ estimators $\hat{\theta}_n(\beta)$. The proof of the following Lemma is given by Jurečková and Procházka (1994).

LEMMA 1. *Suppose that Assumption A hold the model (1.1). Then the RQ estimators* $\hat{\theta}_n(\beta)$ *is strongly consistent for* $\theta_0(\beta)$ *and* $\sqrt{n}(\hat{\theta}_n(\beta) - \theta_0(\beta))$ *converges in distribution to a* $p-$*variate normal random vector with a mean of zero and covariance matrix* $\tau^{-2}(\beta)V^{-1}(\theta_0(\beta))$.

In next lemma we explain the consistency of the restricted RQ estimators $\hat{\theta}_n^0(\beta)$.

LEMMA 2. *Under the conditions of Lemma 1, we have that* $\hat{\theta}_n^0(\beta)$ *converges almost surely to* $\theta_0(\beta)$.

PROOF. First, we consider an another objective function

$$A_n(\theta : \beta) = R_n(\theta : \beta) - R_n(\theta_0 : \beta).$$

Then, Chebyshev's inequality and Assumption A imply that $A_n(\theta : \beta)$ converges to $A(\theta : \beta)$ which has unique minimizer at $\theta_0(\beta)$ and

$$A(\theta : \beta) = \lim_{n\to\infty} E[A_n(\theta : \beta)].$$

On the other hand, the compactness of Θ_0 follows that the sequence of the estimators $\{\hat{\theta}_n^0(\beta)\}$ has at least one limit point $\theta^*(\beta)$. Let $\{\hat{\theta}_{n_k}^0(\beta)\}$ be a subsequence of $\{\hat{\theta}_n^0(\beta)\}$ which converges to $\theta^*(\beta)$. Then, the continuity of the function $\rho_\beta(x)$ implies that

$$A_n(\theta^*(\beta) : \beta) = \lim_{k\to\infty} A_{n_k}(\hat{\theta}_{n_k}^0(\beta) : \beta) \leq A_{n_k}(\theta_0(\beta) : \beta) = A(\theta_0(\beta) : \beta).$$

Hence, the proof is completed. □

In the next lemma we consider approximately linearity of $D_n(\theta\beta)$ and verify that $D_n(\theta : \beta)$ and $\tilde{D}_n(\theta : \beta)$ are asymptotically equivalent.

LEMMA 3. *Under the conditions of Lemma 1, we have*

(i) $\sup_{\theta\in\Theta}\{\hat{D}_n(\theta : \beta) - D_n(\theta : \beta)\} = o_p(n^{-1})$

(ii) $$\sup_{\theta\in M_n(\theta_0(\beta))} \sqrt{n}\{\tilde{D}_n(\theta : \beta) - D_n(\theta : \beta) - g(\eta_\beta)V_n(\theta_0(\beta))(\theta - \theta_0(\beta))\} = o_p(1)$$

where $M_n(\theta_0(\beta)) = \{\theta : \sqrt{n}||\theta - \theta_0(\beta)|| \leq M, M \in R^+\}$ *and* $o_p(1)$ *denotes convergence in probability.*

PROOF. For (i), note that

$$[\psi_\beta - (\psi_\beta)_{\alpha_n}](x) = \begin{cases} \frac{\alpha_n - x}{2\alpha_n} & \text{if} \quad 0 < x \leq \alpha_n \\ \frac{-\alpha_n - x}{2\alpha_n} & \text{if} \quad -\alpha_n < x \leq 0 \\ 0 & \text{if} \quad |x| \geq 0 \end{cases}$$

and

$$[\psi_\beta - (\psi_\beta)_{\alpha_n}](x) \leq \frac{1}{2} I_{[|x| \leq \alpha_n]}.$$

Now, we define

$$X_{ti}(\theta : \beta) = \eta[\psi_\beta - (\psi_\beta)_{\alpha_n}](r_t(\theta))(f_t)_i(\theta)$$

where $r_t(\theta) = y_t - f_t(\theta) - \eta_{beta}$. Then, we have

$$E|X_{ti}(\theta : \beta)| \leq \frac{n\alpha_n}{2} g_t(\alpha_n^*)(f_t)_i(\theta)$$

and

$$\frac{Var X_{ti}(\theta : \beta)}{n} \leq \frac{n\alpha_n}{4} g_t(\alpha_n^{**})(f_t)_i^2(\theta),$$

where α_n^* and α_n^{**} belong to $[-\alpha_n, \alpha_n]$. Therefore, from Kolmogolov's SLLN and Assumption A we obtain that

$$\sum_{t=1}^{n} [\psi_\beta - (\psi_\beta)_{\alpha_n}](r_t(\theta))(f_t)_i(\theta) \overset{a.s}{\to} 0.$$

This result follows the proof of (i).

For (ii), since $M_n(\theta_0)$ is closed subset of parameter space and $(\psi_\beta)_{\alpha_n}$ $(r_t(\theta))$ is continuous on $M_n(\theta_0)$, due to the first result it is sufficient to show that

$$\sqrt{n}\{\tilde{D}_n(\theta : \beta) - \tilde{D}_n(\theta_0 : \beta) - g(\eta_\beta)V_n(\theta_0)(\theta - \theta_0)\} = o_p(1)$$

for some $\theta \in M_n(\theta_0)$. Let $V_n^{(j)}(\theta)$ be the jth-row of $V_n(\theta)$. By a somewhat tedious calculation we show that

$$E\sqrt{n}\{\tilde{D}_{nj}(\theta : \beta) - \tilde{D}_{nj}(\theta_0 : \beta) - g(\eta_\beta)V_n^{(j)}(\theta_0)(\theta - \theta_0)\}$$

is less than

$$(A.1) \qquad \frac{M}{n} \sum_{t=1}^{n} \sum_{k=1}^{p} \Big| \frac{G_t(\eta_\beta) - G_t(\eta_\beta + d_t(\theta) - \alpha_n)}{d_t(\theta) - \alpha_n} (f_t)_k(\theta)_\lambda (f_t)_j(\theta)$$
$$+ g_t(\eta_\beta)(f_t)_j(\theta_0)(f_t)_k(\theta_0)\Big|,$$

where $d_t(\theta) = f_t(\theta) - f_t(\theta_0)$, $\theta_\lambda = \lambda\theta + (1-\lambda)\theta_0$, and $0 < \lambda < 1$. Furthermore, since

$$\frac{G_t(\eta_\beta + d_t(\theta) - \alpha_n) - G_t(\eta_\beta)}{d_t(\theta) - \alpha_n}$$

converges to $g(\eta_\beta)$ as $n \to \infty$, the Toeplitz lemma implies that the (A.1) converges in probability to zero. Suppose that $(f_t)_j(\theta) \neq 0$ where $(f_t)_j(\theta) = \frac{\partial}{\partial\theta_j} f_t(\theta)$.

From a simple calculation we get

$$Var\{\sqrt{n}\tilde{D}_{nj}(\theta:\beta) - \tilde{D}_{nj}(\theta_0:\beta)\}$$

is less than

$$\max_{1\le t\le n} E[(T_t)_j^2(\theta:\beta)]\frac{1}{n}\sum_{t=1}^{n}[(f_t)_j(\theta)]^2,$$

where $(T_t)_j(\theta:\beta) = \{\psi_\beta(r_t(\theta)) - \psi_\beta(r_t(\theta_0))f_t(\theta_0))f_t(\theta_0)(f_t)_j^{-1}(\theta)\}$ and $\tilde{D}_{nj}(\theta:\beta) = \frac{\partial}{\partial\theta_j}\tilde{D}_n(\theta:\beta)$. Hence, the variance of $\sqrt{n}\{D_{nj}(\theta:\beta) - \tilde{D}_{nj}(\theta:\beta_0)\}$ converges to zero due to $T_t)_j(\theta:\beta) = o_p(1)$ and Assumption A. The proof is thus shown. □

The next lemma shows that $Q_n(\theta:\beta)$ provides a useful approximation to $R_n(\theta:\beta)$.

LEMMA 4. *Under the same conditions of Lemma 1, we get*

(*i*) $\tilde{\theta}_n(\beta)$ *and* $\tilde{\theta}_n^0(\beta)$ *converges almost surely to* $\theta_0(\beta)$
(*ii*) $\sup_{\theta\in M_n(\theta_0(\beta))}|Q_n(\theta:\beta) - R_n(\theta:\beta)| = o_p(n^{-1})$

where $\tilde{\theta}_n(\beta)$ *and* $\tilde{\theta}_n^0(\beta)$ *be minimizer of* $Q_n(\theta:\beta)$ *on* Θ *and on* Θ_0, *respectively.*

PROOF. First, from integration for $\psi_\beta(x)$ we get

$$E_\epsilon D_{nj}(\theta_0(\beta):\beta) = \frac{1}{n}\sum_{t=1}^{n}\{\beta - G_t(\eta_\beta)\}(f_t)_j(\theta_0)(\beta)$$

and

$$VarD_{nj}(\theta_0:\beta) \le \frac{\beta(1-\beta)}{n}.$$

Thus, Assumption A implies that $D_{nj}(\theta_0 : \beta)$ converges to zero. Therefore, from Lemma 1 we obtain

$$\{Q_n(\theta : \beta) - Q_n(\theta_0(\beta) : \beta)\} = \frac{g(\eta_\beta)}{2}(\theta - \theta_0(\beta))^T V_n(\theta_0(\beta))(\theta - \theta_0(\beta)) + o_p(1).$$

Hence, under Assumption A we have

$$\{Q_n(\theta : \beta) - Q_n(\theta_0(\beta) : \beta)\} > 0$$

for a sufficiently large n. This result implies that the estimators $\tilde{\theta}_n^0(\beta)$ converges to$\theta_0(\beta)$. the proof for the consistency of $\tilde{\theta}_n(\beta)$ is omitted here.

Next, using the quadratic form of $R_n(\theta : \beta)$ we have

$$\nabla Q_n(\theta : \beta) = D_n(\theta_0(\beta) : \beta) + g(\eta_\beta) V_n(\theta_0(\beta))(\theta - \theta_0(\beta)).$$

On the other hand, all points on the line segment joining θ_0 and θ belong to $M_n(\theta_0)$ are expressed by means of $\theta_\lambda = \lambda\theta + (1 - \lambda)\theta_0$. In virtue of the Chain Rule and Hölder inequality we get

$$\begin{aligned} n|Q_n(\theta : \beta) - R_n(\theta : \beta)| &\leq n \textstyle\sum_{j=1}^p |(\theta_j - \theta_{oj})\{Q_{nj}(\theta_\lambda : \beta) - D_{nj}(\theta_\lambda : \beta)\} \\ &\leq \sqrt{n}||\theta - \theta_0|| \, ||\sqrt{n}\{\nabla Q_n(\theta : \beta) - D_n(\theta : \beta)||, \end{aligned}$$

where $Q_{nj}(\theta : \beta) = [\frac{\partial}{\partial \theta_j} Q_n(\theta : \beta)]$. Since the last expression converges to zero because of Lemma 3, the lemma holds.

The following lemma states that the estimators $\hat{\theta}_n(\beta)$ and $\tilde{\theta}_n(\beta)$ are asymptotic equivalent with rate $\sqrt{n}$.

LEMMA 5 *Under the same conditions of Lemma 1, we get*

(i) $\sqrt{n}(\hat{\theta}_n(\beta) - \tilde{\theta}_n(\beta)) = o_p(1)$
(ii) $\tilde{Q}_n(\beta) - \hat{Q}_n(\beta) = o_p(n^{-1})$ *and* $\tilde{Q}_n^0(\beta) - \hat{Q}_n^0(\beta) = o_p(n^{-1})$.

PROOF. First, since the estimators $\tilde{\theta}_n(\beta)$ converge almost surely to θ_0 in Lemma 4 we have with probability greater than $1 - \delta$

$$||\tilde{\theta}_n(\beta) - \theta_0(\beta)|| \leq \frac{\gamma}{\sqrt{n}},$$

for sufficiently large n and $\gamma > 0$. Furthermore, for any $\epsilon^* > 0$ Lemma 4 yields

$$R_n(\tilde{\theta}_n(\beta) : \beta) < R_n(\theta : \beta) + \epsilon^*$$

on the set $\Gamma_n(\tilde{\theta}_n(\beta)) = \{\theta \in \Theta : ||\tilde{\theta}_n(\beta) - \theta_0(\beta)|| \leq \frac{\gamma}{\sqrt{n}}\}$. That is, $R_n(\theta : \beta)$ has unique minimizer in $\Gamma_n(\tilde{\theta}_n(\beta))$. Therefore, it is sufficient to show that $\{R_n(\theta(\beta) : \beta) - R_n(\theta_0(\beta) : \beta)$ is strictly positive for any $\theta(\beta)$ in $\Gamma_n^C(\tilde{\theta}_n(\beta))$, which is the complement set of $\Gamma_n(\tilde{\theta}_n(\beta))$.

Suppose that there exists $\theta_n(\beta) \in \Gamma_n^C(\tilde{\theta}_n(\beta))$, such that

$$(A.2) \qquad R_n(\theta_n(\beta) : \beta) \leq R_n(\theta_0(\beta : \beta).$$

Let

$$X_t(\beta, \theta_n, \theta_0) = \varphi_\beta(r_t(\theta_n(\beta))) - \varphi_\beta(r_t(\theta_0)).$$

Since $\varphi_\beta(x)$ is convex, $X_t(\beta, \theta_n(\beta), \theta_0)$ is less than $\varphi_\beta(f_t(\theta_0) - f_t(\theta_n(\beta)))$. Thus, according to Hölder's inequality we get

$$\begin{aligned} X_t(\beta, \theta_n, \theta_0) &\leq \varphi_\beta(d_t(\theta_n(\beta))) \\ &\leq \tau_\beta ||\nabla f_t(\theta_n^*(\beta))|| \, ||\theta_n(\beta) - \theta_0(\beta)||, \end{aligned}$$

where $\tau_\beta = max\{\beta, 1-\beta\}$ and $\theta_n^*(\beta) = \lambda\theta_n(\beta) + (1-\lambda)\theta_0$. By virtue of Chebyshev's inequality we obtain

$$R_n(\theta_n(\beta) : \beta) - R_n(\theta_0(\beta) : \beta) = \frac{1}{n}\sum_{t=1}^{n} EX_t(\beta, \theta_n, \theta_0) + o_p(1).$$

On the other hand, the simple calculation implies that the expectation of the $X_t(\beta, \theta_n(\beta), \theta_0)$ is

$$\begin{cases} \int_{\eta_\beta}^{\eta_\beta + d_t(\theta_n(\beta))} (d_t(\theta_n(\beta)) - \lambda + \eta_\beta) dG_t(\lambda), & \text{if} \quad d_t(\theta_n(\beta)) > 0 \\ \int_{\eta_\beta + d_t(\theta_n(\beta))}^{\eta_\beta} (\lambda - d_t(\theta_n(\beta)) - \eta_\beta) dG_t(\lambda), & \text{if} \quad d_t(\theta_n(\beta)) < 0. \end{cases}$$

Thus, the expectation of the $X_t(\beta, \theta_n(\beta), \theta_0)$ is greater than

$$\begin{aligned} &\zeta_n(\beta) \equiv \min\{g_t(\eta_\beta^*)[G_t(\eta_\beta + d_t(\theta_n(\beta))) - G_t(\eta_\beta)], \\ &g_t(\eta_\beta^{**})[G_t(\eta_\beta) - G_t(\eta_\beta + d_t(\theta_n(\beta)))]\} \end{aligned}$$

where η_β^* and η_β^{**} are internal dividing points between η_β and $\eta_\beta + d_t(\theta_n(\theta))$. Therefore, Assumption A_4 implies that

$$R_n(\theta_n(\beta) : \beta) - R_n(\theta_0 : \beta) \geq \frac{k_n(\beta)}{n}\zeta_n(\beta).$$

This is a contradiction to $(A.2)$ due to Assumption A_4 and completes the proof of (i).

Next, we can rewrite that

$$\begin{aligned} &n\{\tilde{Q}_n(\beta) - \hat{Q}_n(\beta)\} \\ &= \sqrt{n}(\tilde{\theta}_n(\beta) - \hat{\theta}_n(\beta))\sqrt{n}D_n(\theta_0(\beta) : \beta) \\ &+ \sqrt{n}(\tilde{\theta}_n(\beta) - \hat{\theta}_n(\beta))g(\eta_\beta)V^{-1}(\theta_0(\beta))\{\sqrt{n}(\tilde{\theta}_n(\beta) - \theta_0(\beta)) \\ &- \sqrt{n}(\hat{\theta}_n(\beta) - \theta_0(\beta))\}. \end{aligned}$$

Moreover, since $E\psi_\beta(\epsilon_t - \eta_\beta) = 0$ and $Var\psi_\beta(\epsilon_t - \eta_\beta) = \beta(1-\beta)$ by means of Linderberg Central Limit Theorem we have

$$(A.3) \qquad D_n(\theta_0(\beta):\beta) \xrightarrow{L} N(0, \beta(1-\beta)(\theta)V(\theta_0(\beta))).$$

Hence, the asymptotic behavior of $D_n(\theta_0(\beta):\beta), \tilde{\theta}_n(\beta)$ and $\hat{\theta}_n(\beta)$ complete the proof of (ii). □

PROOF OF THEOREM 2.1 First, we consider Wald-Test. Let $\theta^*(\beta)$ be the line segment joining $\hat{\theta}_n(\beta)$ and $\theta_0(\beta)$. Then, by virtue of the first order Taylor's formula we have

$$\sqrt{n}[\hat{u}_n(\beta) - u(\theta_0(\beta))] = \sqrt{n}\nabla u_n^*(\beta)[\hat{\theta}_n(\beta) - \theta_0(\beta)]$$

where $\nabla u_n^*(\beta)$ denote $\nabla u(\beta)$ evaluated at $\theta^*(\beta)$. The asymptotic normality of nonlinear RQ estimators and Sluksky's theorem imply that

$$\sqrt{n}\hat{u}_n(\beta) \xrightarrow{L} N_q(\gamma(\beta), \tau^{-2}(\beta)\nabla u(\theta_0(\beta))V^{-1}(\theta_0(\beta))\nabla u^T(\theta_0(\beta)))$$

The proof follows consistency of $\hat{\theta}_n(\beta)$.

For the second time we investigate the limiting distribution of LM-test. Note that $M_n(\beta)$ is asymptotic equivalent to

$$[\sqrt{n}\hat{D}_n^0(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\nabla\hat{u}_n(\beta)]$$
$$[\beta(1-\beta)\nabla\hat{n}_n(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\nabla\hat{u}_n(\beta)]^{-1}$$
$$[\nabla\hat{u}_n(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\sqrt{n}\hat{D}_n^0(\beta)].$$

The mean value theorem of $\hat{D}_n^0$ and $\hat{u}_n^0(\beta)$ imply that

$$\nabla u_n^*(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\sqrt{n}\hat{D}_n^0(\beta)$$

is equal to

$$\nabla u_n^*(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\sqrt{n}\hat{D}_n^0(\beta) - g(\eta_\beta)\sqrt{n}u(\theta_0(\beta)).$$

Hence,

$$\nabla u_n^*(\beta)V_n^{-1}(\hat{\theta}_n^0(\beta))\sqrt{n}\hat{D}_n^0(\beta)$$
$$\to N_q(-g(\eta_\beta)\gamma(\beta), \beta(1-\beta)\nabla u^T(\beta)V(\theta_0(\beta))\nabla u(\beta)).$$

The asymptotic distribution of $M_n(\beta)$ follows $(A.3)$.

In the last we consider the asymptotic distribution of $R_n(\beta)$. Since $\{\hat{R}_n^0(\beta) - \hat{R}_n(\beta)\}$ equal to

$$(\hat{R}_n^0(\beta) - \hat{Q}_n^0(\beta)) + (\hat{Q}_n^0(\beta) - \tilde{Q}_n^0(\beta))$$
$$+ (\tilde{Q}_n^0(\beta) - \tilde{Q}_n(\beta)) + (\tilde{Q}_n(\beta) - \hat{Q}_n(\beta))$$
$$+ (\hat{Q}_n(\beta) - \hat{R}_n(\beta)),$$

The Lemma 2, Lemma 3 and Lemma 4 cited above mean that

$$n[\hat{R}_n^0(\beta) - \hat{R}_n(\beta)] \simeq n[(\tilde{Q}_n^0(\beta) - \tilde{Q}_n(\beta)]$$

where $\simeq$ denotes asymptotic equivalent. Also, first order Taylor theorem for $\nabla Q_n(\tilde{\theta}_n^0(\beta):\beta)$ signifies that

$$(\tilde{\theta}_n^0(\beta) - \tilde{\theta}_n(\beta)) = \frac{1}{g(\eta_\beta)} V_n^{-1}(\theta_{on}^*(\beta)) \nabla Q_n(\tilde{\theta}_n^0(\beta):\beta).$$

Thus, this result and Taylor theorem imply that

$$(\tilde{Q}_n^0(\beta) - (\tilde{Q}_n(\beta) \simeq \frac{1}{g(\eta_\beta)} \nabla^T Q_n(\tilde{\theta}_n^0(\beta):\beta) V_n^{-1}(\theta_0(\beta)) \nabla^T Q_n(\tilde{\theta}_n^0(\beta):\beta)$$

So, we have

$$R_n(\beta) \simeq \sqrt{n} \nabla^T Q_n(\tilde{\theta}_n^0(\beta):\beta)[\frac{1}{\beta(1-\beta)} V_n^{-1}(\theta_0(\beta))] \sqrt{n} \nabla Q_n(\tilde{\theta}_n^0(\beta):\beta).$$

Moreover, $\nabla \hat{u}_n(\beta) V_n^{-1}(\theta_0(\beta)) \nabla Q_n(\tilde{\theta}_n^0(\beta):\beta)$ is asymptotic equivalent to

$$\nabla \hat{u}_n(\beta) V_n^{-1}(\theta_0(\beta)) D_n(\theta_0(\beta):\beta) - g(\eta_\beta) u(\theta_0(\beta)).$$

Applying $Q_n(\tilde{\theta}_n^0(\beta):\beta)$ and $\nabla \hat{u}_n(\beta)$ to Lagrange theorem, we obtain

$$\begin{aligned} &\sqrt{n} Q_n(\tilde{\theta}_n^0(\beta):\beta) \\ &\simeq \nabla \hat{u}_n^T(\beta)[\nabla \hat{u}_n(\beta) V_n^{-1}(\theta_0(\beta)) \nabla \hat{u}_n^T(\beta)]^{-1} \\ &- g(\eta_\beta) \nabla \hat{u}_n^T(\beta)[\nabla \hat{u}_n(\beta) V_n^{-1}(\theta_0(\beta)) \nabla \hat{u}_n^T(\beta)]^{-1} \\ &\sqrt{n} u(\theta_0(\beta)) \end{aligned}$$

Hence, Assumption B and (A.3) mean that the proof completed. □

References

[1] A. R. Gallant, *Nonlinear Regression Models* , John Wiley and Sons, 1987.

[2] Huber, *Robust Statistics* , John Wiley and Sons, 1981.

[3] R. Koenker and G. Bassett, *Regression Quantiles*, Econometrica **46** (1978), 33–50.

[4] ———, *Robust Tests for Heterocedasticity based on Regression Quantiles*, Econometrica **50** (1982), 43–61.

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

[6] E. L. Lehmann, *Testing Statistical Hypotheses* , John Wiley and Sons, 1986.

[7] Z. J. Liu, *Nonparametric Estimates of the Nuisance Parameter in the LAD Tests*, Communication in Statistics-A, **21** (1992), 861–881.

[8] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, John Wiley and Sons, 1989.

[9] J. Wang, *Asymptotic normality of L_1-estimators in nonlinear regression*, J. Multivariate Anal. **54** (1995), 227–238.

Seung Hoe Choi and Kyung Joong Kim
Department of General Studies
Hankuk Aviation University
Koyang 411, Korea
E-mail: shchoi@mail.hangkong.ac.kr
kj_kim@mail.hangkong.ac.kr

Myung Sook Lee
Department of Mathematics
Yonsei University
Seoul 120-749, Korea
E-mail: jhmslee@yonsei.ac.kr