Bull. Korean Math. Soc. **36** (1999), No. 3, pp. 451–457

THE STRONG CONSISTENCY OF NONLINEAR REGRESSION QUANTILES ESTIMATORS

SEUNG HOE CHOI AND HAE KYUNG KIM

ABSTRACT. This paper provides sufficient conditions which ensure the strong consistency of regression quantiles estimators of nonlinear regression models. The main result is supported by the application of an asymptotic property of the least absolute deviation estimators as a special case of the proposed estimators. Some example is given to illustrate the application of the main result.

1. Introduction

In this paper we consider the following nonlinear regression model

$$y_t = f(x_t, \theta_o) + \epsilon_t, \quad t = 1, 2, \cdots, n \tag{1.1}$$

where f is a known response function, x_t which belong to bounded subspace Ω of R^q is a observed input vector, the error term ϵ_t are independent and identically distributed (i.i.d.) random variables with finite variance. The parameter vector θ_o which is interior point in Θ is unknown and to be estimated.

The most common technique used to estimate the true parameters in model (1.1) is the method of least squares developed by Jennrich (1969) and Wu (1981). However, on the occasions of the error terms contain some outliers or depart from normal distribution the least squares method is poor estimators due to the extreme sensitivity of the least squares estimators to some outlier. To overcome this defect, the search for the robust procedures alternative the least squares method has generated considerable interest in statistical inference.

Received June 5, 1998.
1991 Mathematics Subject Classification: 62J02.
Key words and phrases: nonlinear regression model, strong consistency, regression quantiles estimator.

The Least Absolute Deviation (LAD) estimators based on sample median is defined by any vector minimizing the sum of absolute deviations

$$D_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} |y_t - f_t(\theta)|,$$

where $f_t(\theta) = f(x_t, \theta)$. Oberhofer (1982) gave sufficient conditions for the weak consistency of the LAD estimators in nonlinear regression models. Wang (1995) derived the asymptotic normality of the nonlinear LAD estimators and Kim and Choi (1995) investigated the asymptotic properties of the nonlinear LAD estimators and explained that the relative efficiency of the LAD estimators to the least square estimators is the same as the relative efficiency of the sample median to the sample mean.

Meanwhile, in case of a distribution function of errors is positively skewed (or negatively skewed) other quantiles than median (50th quantile) may reveal the information about the unknown parameter θ_o in model (1.1). Regression quantiles which provide a natural generalization of the notion of sample quantile to the general regression model were proposed by Koenker and Basset (1982). Quantile-based estimators have long been known in the statistical literature as 'L-estimators' for their relative efficiency for heavy-tailed error distributions.

The β-th regression quantiles estimators $(0 < \beta < 1)$ of the true parameter θ_o based on (y_t, x_t), denoted by $\hat{\theta}_n(\beta)$, is a parameter which minimizes the objective function

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

where the "check" function

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

Since the check function $\varphi_\beta(x)$ rotates the absolute function $\frac{|x|}{2}$ by some angle ϕ in the clockwise direction $(\beta < \frac{1}{2})$, the least absolute deviation estimators is an obviously important special case of the regression quantiles estimators. In some recent papers, analysis of linear models using quantiles estimation has been published by many authors : Basset and Koenker (1982, 1986) and Portnoy (1991). Basset and Koenker (1986) established the strong consistency of regression quantiles statistics in linear models with i.i.d. errors. Portnoy (1991) discussed asymptotic behavior of regression quantiles under more general heteroscedasticity and dependence assumptions in linear models.

The main object of this paper is to provide simple and sufficient conditions for the strong consistency of the regression quantiles estimators $\hat{\theta}_n(\beta)$ in nonlinear regression model (1.1).

2. Strong Consistency

We start this section by introducing some conditions which ensure the strong consistency of the regression quantile estimator in the nonlinear regression model (1.1). Let P be a probability measure on R^q and H denote the distribution function of input vector x_t. Let $\nabla f_t(\theta) = [\frac{\partial}{\partial \theta_i} f(x_t, \theta)]_{(p\times 1)}$.

Assumption A

The parameter space Θ is compact subspace of R^p.

Assumption B

B_1 : The response function $f(x_t, \theta)$ and the partial derivatives $\nabla f_t(\theta)$ are continuous on $\Theta \times \Omega$.

B_2 : The distribution function $G(x)$ of the errors is continuously differentiable with density $g(x)$ which is strictly positive at $G^{-1}(\beta) = 0$.

B_3 : $V_n(\theta_o) = \frac{1}{n}\sum_{t=1}^{n} \nabla f_t(\theta_o) \nabla^T f_t(\theta_o)$ converges to a positive definite matrix $V(\theta_o)$ as $n \to \infty$.

B_4 : $P\{x \in \Omega : f(x, \theta_o) \neq f(x, \theta)\} > 0$ for each $\theta \neq \theta_o$.

Modifying (1.2), we have another objective function of the nonlinear regression quantiles estimators

$$Q_n(\theta; \beta) = S_n(\theta; \beta) - S_n(\theta_o; \beta). \tag{2.1}$$

Since $S_n(\theta_o; \beta)$ is independent of θ, the regression quantiles estimators $\hat{\theta}_n(\beta)$ defined in (1.2) is equivalent to the minimizer of (2.1). Before we proceed to consider the main result, we present the following lemma needed in the proof of the main theorem.

LEMMA 2.1. *Suppose that model* (1.1) *satisfies Assumptions A and B. Then for any* θ,

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

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

Proof. Define the random variable $Z_t(\theta)$ as following

$$Z_t(\theta) = \begin{cases} 1 & \text{if } y_t \leq f_t(\theta), \\ 0 & \text{otherwise.} \end{cases}$$

Then we can rewrite

$$\begin{aligned} Q_n(\theta;\beta) &= \frac{1}{n}\sum_{t=1}^{n}[\varphi_\beta(r_t(\theta)) - \varphi_\beta(\varepsilon_t)] \\ &= \frac{1}{n}\sum_{t=1}^{n}[(\beta-1)(r_t(\theta)I_{\{r_t(\theta)\leq 0\}} - \varepsilon_t I_{\{\varepsilon_t\leq 0\}}) \\ &\qquad +\beta(r_t(\theta)I_{\{r_t(\theta)>0\}} - \varepsilon_t I_{\{\varepsilon_t>0\}})] \\ &= \frac{1}{n}\sum_{t=1}^{n}[(\beta - Z_t(\theta))r_t(\theta) + (Z_t(\theta_0)-\beta)\varepsilon_t], \end{aligned}$$

where $r_t(\theta) = y_t - f_t(\theta)$. Let $X_t = (\beta - Z_t(\theta))r_t(\theta) + (Z_t(\theta_0) - \beta)\varepsilon_t$. According to Hölder's inequality, we get

$$|X_t| \leq (\beta+2)|\epsilon_t| + |\beta+1| \|\nabla f(\bar{\theta})\| \|\theta - \theta_o\|,$$

where$\|.\|$ denote Euclidean norm and $\bar{\theta} = \lambda\theta_o + (1-\lambda)\theta, 0 \leq \lambda \leq 1$. On the other hand, Chebyshev's inequality gives

$$P\{|Q_n(\theta,\beta) - E\{Q_n(\theta,\beta)\}| > \epsilon\} \leq \frac{\max\limits_{1\leq t\leq n} Var X_t}{n\epsilon^2}.$$

The proof follows from Assumption A and B_1. □

The following theorem is the main result of this section, which provides sufficient conditions for the strong consistency of regression quantiles estimators.

THEOREM 2.1. *For the model* (1.1), *suppose that Assumptions A and B are fulfilled. Then the regression quantiles estimators $\hat{\theta}_n(\beta)$ defined in* (1.2) *is strongly consistent for θ_0.*

Proof. For any $\delta > 0$, it is sufficient to show that

$$\lim_{n\to\infty} \inf_{\|\theta-\theta_o\|>\delta} \{Q_n(\theta;\beta)\} > 0 \ a.e. \tag{2.2}$$

From the lemma 2.1 we have

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

where E_ϵ denotes the expected value of the error term ϵ_t. Note that

$$\begin{aligned} E_\epsilon X_t &= \int_R (\beta - I_{\{\lambda\le d_t(\theta)\}})(\lambda - d_t(\theta)) + (I_{\{\lambda\le d_t(\theta_o)\}} - \beta)\lambda dG(\lambda) \\ &= \int_{-\infty}^{0} \lambda\, dG(\lambda) - \int_{-\infty}^{d_t(\theta)} \lambda\, dG(\lambda) + d_t(\theta)G(d_t(\theta)) - \beta d_t(\theta), \end{aligned}$$

where $d_t(\theta) = f_t(\theta) - f_t(\theta_o)$.

First we prove that θ_o is a local minimizer of $Q(\theta;\beta) = \lim\limits_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_\epsilon X_t$. By simple calculation, we get

$$Q(\theta;\beta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\Big[\int_{d_t(\theta)}^{0} \lambda\, dG(\lambda) + d_t(\theta)G(d_t(\theta)) - \beta d_t(\theta)\Big].$$

$$\nabla Q(\theta;\beta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}[\nabla f_t(\theta)(G(d_t(\theta)) - \beta)].$$

$$\nabla^2 Q(\theta;\beta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}[\nabla f_t(\theta)(G(d_t(\theta)) - \beta) + g(d_t(\theta))\nabla f_t(\theta)\nabla^T f_t(\theta)].$$

Furthermore, $\nabla Q(\theta_o;\beta) = 0$ and $\nabla^2 Q(\theta_o;\beta)$ is positive definite matrix. Hence $Q(\theta;\beta)$ attains a local minimum at θ_o.

Next we show that this local minimizer θ_o is indeed the global minimizer. Let $N_\tau(\theta_o) = \{\theta : \|\theta - \theta_o\| < \tau\}$. Since $R(\theta) = N_\tau^c(\theta_o) \cap \Theta$ is compact, there exists θ^* such that

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

We consider

$$\begin{aligned} Q(\theta;\beta) &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\Big[\int_{d_t(\theta)}^{0} \lambda\, dG(\lambda) + d_t(\theta)G(d_t(\theta)) - \beta d_t(\theta)\Big] \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\int_{d_t(\theta)}^{0} (\lambda - d_t(\theta))\, dG(\lambda). \end{aligned}$$

If $d_t(\theta) < 0$, then $\lambda - d_t(\theta)$ is positive in $(d_t(\theta), 0)$. Thus there exist ξ_1 and ξ_2 such that $d_t(\theta) < \xi_1 < \xi_2 < 0$. From Assumption B_2, since $g(\lambda)$ is strictly positive on $[\xi_1, \xi_2]$, there exists a $\eta_1 > 0$ such that $g(\lambda) > \eta_1$ on $[\xi_1, \xi_2]$. Thus we obtain

$$(2.3)\quad \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\int_{d_t(\theta)}^{0}(\lambda - d_t(\theta))g(\lambda)\,d\lambda \;>\; \int_{\xi_1}^{\xi_2}(\lambda - d_t(\theta))g(\lambda)\,d\lambda$$

$$> \eta_1\int_{\xi_1}^{\xi_2}(\lambda - d_t(\theta))\,d\lambda.$$

Likewise if $d_t(\theta) > 0$, we have similar result. Thus, we have $Q(\theta;\beta) > Q(\theta_o;\beta)$ on $R(\theta)$ because the last term is positive.

Finally, from Assumption B_4 and above fact we obtain

$$(2.4)\qquad \inf_{\|\theta-\theta_o\|>\delta} E_{\epsilon\times x}Q_n(\theta;\beta) \geq \inf_{\|\theta-\theta_o\|>\delta}\int_{\omega}\int_{R} X_1 dG(\lambda)dH(x),$$

where $\omega = \{x \in \Omega | f(x,\theta) \neq f(x,\theta_o)\}$. In virtue of (2.3) and (2.4), we get for sufficiently large n

$$\inf_{\|\theta-\theta_o\|>\delta} E_{\epsilon\times x}Q_n(\theta;\beta) \geq \eta_2,$$

where η_2 is a positive real number. The proof is completed. □

To see that the assumptions of the main theorem are sufficient to cover a class of nonlinear regression functions, we now consider the following example.

EXAMPLE. Let p be a differential function from R^m to R^+. Consider the model $y_t = f(x_t, \theta_o) + \epsilon_t$ where $\theta_o \in \Theta = [0, a_1] \times [0, a_2], a_1, a_2 < \infty$ and $f(x,\theta) = \theta_1 p(x)^{\theta_2}$. Many authors consider the nonlinear model with $p(x) = e^x$ and $p(x) = x$ for $m = 1$. Assume that $\{\epsilon_t\}$ are i.i.d. random variables with the continuous p.d.f $g(x)$ and distribution function $G(x)$ for which $G(0) = \beta$. Then $V_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\nabla f_t(\theta)\nabla^T f_t(\theta)$ converges to $V(\theta)$ where

$$V(\theta) = \begin{bmatrix} \int p(x)^{2\theta_2}\,dG(x) & \int \theta_1 p(x)^{2\theta_2}\ln p(x)\,dG(x) \\ \int \theta_1 p(x)^{2\theta_2}\ln p(x)\,dG(x) & \int (\theta_1 p(x)^{\theta_2}\ln p(x))^2\,dG(x) \end{bmatrix}.$$

For a non-zero vector $\alpha = (\alpha_1, \alpha_2)$

$$\alpha V(\theta)\alpha^T = \int (\alpha_1 + \theta_1 \ln p(x)\alpha_2)^2 p(x)^{2\theta_2}\, dG(x) > 0,$$

where $(\theta_1, \theta_2) \in \Theta^0$. It is not hard to verify that Assumptions A and B are satisfied. Thus we can guarantee the strong consistency of the regression quantile estimators.

ACKNOWLEDGEMENTS. We thank to the referee for suggestions and remarks in improving this paper.

References

[1] Basset, G. and Koenker, R., *An empirical quantile function for linear models with i.i.d. errors*, Journal of the American Statistical Association **77** (1982), 407-415.

[2] ———, *Strong consistency of regression quantiles and related empirical process*, Econometric Theory **2** (1986), 191-201.

[3] Jennrich, R. I., *Asymptotic properties of nonlinear least squares estimators*, The Annal of Mathematical Statistics **40** (1969), 633-643.

[4] Kim, H. K. and Choi, S. H., *Asymptotic properties of nonlinear Least Absolute Deviation estimators*, Journal of Korean Statistics Society **24** (1995), 127-139.

[5] Oberhofer, W., *The consistency of nonlinear regression minimizing the L_1-norm*, The Annals of Statistics **10** (1982), 316-319.

[6] Portnoy, S., *Asymptotic behavior of regression quantiles in non-stationary, dependent cases*, Journal of Multivariate Analysis **50** (1991), 100-113.

[7] Wang, J., *Asymptotic normality of L_1-estimators in nonlinear regression*, Journal of Multivariate Analysis **54** (1995), 227-238.

[8] Wu, C. F., *Asymptotic theory of nonlinear least squares estimation*, The Annals of Statistics **9** (1981), 501-513.

SEUNG HOE CHOI, DEPARTMENT OF GENERAL STUDIES, HANKUK AVIATION UNIVERSITY, KOYANG 411, KOREA
E-mail: shchoi@mail.hangkong.ac.kr

HAE KYUNG KIM, DEPARTMENT OF MATHEMATICS, YONSEI UNIVERSITY, SEOUL 120, KOREA
E-mail: abc1217@yonsei.ac.kr