Journal of the Korean Statistical Society (2001), **30**: *4, pp 551–561*

Nonlinear Regression Quantile Estimators

Seung Hoe Choi[1], Hae kyung Kim[2] and Kyung Ok Park[3]

ABSTRACT

This paper deals with the asymptotic properties for statistical inferences of the parameters in nonlinear regression models. As an optimal criterion for robust estimators of the regression parameters, the regression quantile method is proposed. This paper defines the regression quantile estimators in the nonlinear models and provides simple and practical sufficient conditions for the asymptotic normality of the proposed estimators when the parameter space is compact. The efficiency of the proposed estimator is especially well compared with least squares estimator, least absolute deviation estimator under asymmetric error distribution.

Keywords: Nonlinear Regression Quantiles Estimators, Consistency, Normality, Efficiency.

1. Introduction

The general nonlinear regression model is

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

where $x_t \in \mathcal{X}$ is a vector of explanatory variable of dimension q, the true parameter $\theta_o \in R^p$ is an unknown regression parameter to be estimated, and $f : R^{p+q} \to R^1$ is a continuous function of x and θ. We will assume throughout that ε_t are independent and identically distributed (i.i.d.) random variables with finite variance.

The method of Least Square (LS), developed by Jennrich (1969) and Wu (1981) is one of the most commonly used methods for estimating the regression coefficients in model (1.1). But the extreme sensitivity of the LS estimation to modest amounts of outlier contamination makes it a very poor estimator in many

[1]Department of General Studies, Hankuk Aviation University, Koyang 411, Korea
[2]Department of Mathematics, Yonsei University, Seoul 120-749, Korea
[3]Senior Engineer, Samsung Electronic Co, Ltd, Suwon, Kyunggido, 442-742, Korea

non-Gaussian, especially long-tailed, situations. To overcome the lack of the LS estimation, there has been an increased interest in robust estimation procedures applied to the regression model.

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) and Wang(1995) gave sufficient conditions for the weak consistency and asymptotic normality of the LAD estimators in nonlinear regression models. 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 LS estimators is the same as the relative efficiency of the sample median to the sample mean. The condition $G(0) = \frac{1}{2}$ is one of sufficient conditions which ensure optimal properties for the LAD estimators. But it is not easy to select response function $f(x, \theta)$ in order to equal the ratio of the number of $\{t : y_t > f(x_t, \theta)\}$ to the number of$\{t : y_t < f(x_t, \theta)\}$. Hence 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(1978).

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 LAD estimators is easily seen to be a special case; $S_n(\theta; \frac{1}{2}) = \frac{1}{2}Q_n(\theta)$ as defined above. In some recent papers, analysis of linear models using quantiles estimation has been published by many authors : Basset and Koenker(1982, 1986), Powell(1986) and Portnoy(1991). Basset and

Koenker(1986) established the strong consistency of regression quantiles statistics in linear models with i.i.d. errors. Powell(1986) investigated asymptotic properties of proposed estimators in censored regression model. Portnoy(1991) discussed asymptotic behavior of regression quantiles under more general heteroscedasticity and dependence assumptions in linear models.

The main purpose of this paper is to provide sufficient conditions for the asymptotic properties of the regression quantile estimators in the nonlinear regression model (1.1). For this, we establish the strong consistency of nonlinear regression quantile estimators $\hat{\theta}_n$ under the some mild conditions in section 2. In section 3, we prove the asymptotic normality of using smooth function approximate to the absolute value function. Finally, we propose confidence region based on the estimators and discuss desirable asymptotic properties including the asymptotic relative efficiency of the test procedure in section 4.

2. Strong Consistency

Let $(\mathcal{X}, \Omega, P)$ be a probability space on R^q and H denote the distribution function of input vector x_t. Let $\nabla f_t(\theta) = [\frac{\partial}{\partial \theta_i} f(x_t, \theta)]_{(p\times 1)}$. We make the following assumptions in order to guarantee the existence of a sequence of strongly consistent estimators of $\hat{\theta}_n(\beta)$ for a particular value of β of the true parameter.

Assumption A

A_1 : The parameter space $\Theta(\beta)$ is a compact subspace of R^p.

A_2 : $\nabla f_t(\theta)$ are continuous on $\mathcal{X} \times \Theta(\beta)$.

A_3 : $P\{x \in \Omega : f(x, \theta_o) \neq f(x, \theta)\} > 0$ for each $\theta \neq \theta_o$.

Assumption B

B_1 : The distribution function $G(x)$ of the errors is continuously differen -tiable with density $g(x)$ which is strictly positive at $G^{-1}(\beta) = 0$.

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). 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_o.*

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

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

The detailed proof of this theorem is given in appendix. □

3. Asymptotic Normality

In this section, we study the asymptotic behavior of nonlinear regression quantile estimators under some mild conditions. The main idea is to approximate to the function $\varphi_\beta(x)$ by a smooth function $h_n(x)$. As such function we use

$$h_n(x) = \Big[-\frac{3}{2}{\alpha_n}^2x^3+\frac{\alpha_n}{4}x^2+\beta x+\frac{1}{4\alpha_n}\Big]I_{\{|x|\le\frac{1}{\alpha_n}\}}+\beta x I_{\{x>\frac{1}{\alpha_n}\}}+(\beta-1)xI_{\{x<-\frac{1}{\alpha_n}\}},$$

where $n^2 = o(\alpha_n^3)$, $\alpha_n = o(n)$, and $n^{\frac{1+\delta}{2}} = o(\alpha_n)$ for some $\delta > 0$. The sequence $\alpha_n = \frac{1}{n^p}$ such that $\frac{2}{3} < p < 1$ satisfies above conditions. Let $S_n^*(\theta;\beta) = \frac{1}{n}\sum_{t=1}^n h_n(r_t(\theta))$ which is close to $S_n(\theta;\beta)$ defined in (1.2) for sufficiently large n and $\tilde{\theta}_n(\beta)$ denote a minimizer of $S_n^*(\theta;\beta)$. By simple calculation we obtain

$$\nabla S_n^*(\theta;\beta) = \frac{1}{n}\sum_{t=1}^n[-h_n'(r_t(\theta))\nabla f_t(\theta)],$$

and

$$\nabla^2 S_n^*(\theta;\beta) = \frac{1}{n}\sum_{t=1}^n[h_n''(r_t(\theta))\nabla f_t(\theta)\nabla^T f_t(\theta) - h_n'(r_t(\theta))\nabla^2 f_t(\theta)],$$

where $\nabla^2 f_t(\theta) = \Big[\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_t(\theta)\Big]_{i,j=1,\cdots,p}$. We will require the following additional assumption.

Assumption C

$C_1 : V_n(\theta_o) = \frac{1}{n}\sum_{t=1}^n \nabla f_t(\theta_o)\nabla^T f_t(\theta_o)$ converges to a positive definite ma -trix $V(\theta_o)$ as $n \to \infty$.

The following theorem is the main result of this section, which present asymptotic normality of proposed estimation $\hat{\theta}_n(\beta)$.

Theorem 3.1. *If Assumptions A, B and C hold for the model (1.1), then* $\sqrt{n}(\hat{\theta}_n(\beta)$ $-\theta_o)$ *converges in distribution to a p-variate normal random vector with mean zero and variance-covariance matrix* $\frac{\beta(1-\beta)}{[g(0)]^2}V^{-1}(\theta_o)$. *That is,*

$$\sqrt{n}(\hat{\theta}_n(\beta) - \theta_o) \xrightarrow{\mathcal{L}} N_p(0, \frac{\beta(1-\beta)}{[g(0)]^2}V^{-1}(\theta_o)),$$

where $g(0)$ *is the height of the density of the error* ϵ_t *at zero.*

Proof: The proof can be briefly described as follows.

(i) $\sup_{\theta \in \Theta(\beta)} n\{S_n^*(\theta;\beta) - S_n(\theta;\beta)\} = o_p(1)$,

(ii) $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = o_p(1)$,

(iii) $\sqrt{n}(\tilde{\theta}_n(\beta) - \theta_o) \xrightarrow{\mathcal{L}} N_p(0, \frac{\beta(1-\beta)}{[g(0)]^2}V^{-1}(\theta_o))$.

The detailed proof of this result, as well as the proof for Theorem 2.1 are given in appendix. □

To see that the assumptions of the previous theorem are sufficiently to cover a class of nonlinear regression functions, we consider the following nonlinear regression models with compact parameter space.

Example 1 Consider the exponential model $y_t = f(x_t, \theta_o) + \epsilon_t$, with the regression function $f(x,\theta) = \theta_1 e^{\theta_2}, x \neq 0$ where $\theta = (\theta_1, \theta_2) \in \Theta = [0,a] \times [0,b]$ $a, b < \infty$. Then $V_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \nabla f_t(\theta) \nabla^T f_t(\theta)$ converges to

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

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

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

where $(\theta_1, \theta_2) \in \Theta^0$. Suppose that ϵ_t are i.i.d. random variables with the distribution function G for which $G(0) = \beta$ and probability density function(p.d.f.) $g(x)$ is continuous. It is easy to check that Assumptions A and B are satisfied. Thus we can guarantee the strong consistency asymptotic normality of the regression quantile estimators.

4. Asymptotic Relative Efficiency

In this section, we will consider asymptotic confidence region for the parameter θ_o in the model (1.1), and test procedure for the hypothesis about θ_o based on the asymptotic normality of the regression quantile estimators. The Asymptotic Relative Efficiency(ARE) of two estimators having asymptotic normality, comparing the volumes of the corresponding confidence ellipsoids will be considered.

The asymptotic normality of $\sqrt{n}(\hat{\theta}_n(\beta) - \theta_o)$, derived in theorem 3.1, implies that $R_n(\hat{\theta}_n)$ has asymptotically a central chi-square distribution with p degrees of freedom, where

$$R_n(\hat{\theta}_n) = n(\hat{\theta}_n - \theta_o)^T (g(0))^2 (\beta - \beta^2)^{-1} V_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_o),$$

and $V_n(\hat{\theta}_n)$ is the $p \times p$ matrix with (i,j)-th element

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial f_t(\hat{\theta}_n)}{\partial \theta_i} \frac{\partial f_t(\hat{\theta}_n)}{\partial \theta_j}.$$

Define $C_{1-\alpha}(\hat{\theta}_n)$ as the set of θ such that

$$(\hat{\theta}_n - \theta)^T (g(0))^2 (\beta - \beta^2)^{-1} V_n(\hat{\theta}_n)(\hat{\theta}_n - \theta) \leq \frac{1}{n}\chi_p^2(\alpha),$$

where $\chi_p^2(\alpha)$ is $(1-\alpha)$th quantile of the chi-square distribution.

On the other hand, it is well known that under certain conditions, the sequence of the least squares(LS) estimators $\check{\theta}_n$ has asymptotically a normal distribution in the sense that

$$\sqrt{n}(\check{\theta}_n - \theta_o) \xrightarrow{\mathcal{L}} N_p(0, \sigma^2 V^{-1}(\theta_o)),$$

where σ^2 is the common variance of errors in the model (1.1). See Wu(1981). Thus, a $100(1-\alpha)$ percent confidence region based on the LS estimators, denoted by $C_{1-\alpha}(\check{\theta}_n)$, is the set of θ such that

$$(\check{\theta}_n - \theta)^T \sigma^{-2} V_n(\check{\theta}_n)(\check{\theta}_n - \theta) \leq \frac{1}{n}\chi_p^2(\alpha).$$

Hence the corresponding confidence ellipsoids $C_{1-\alpha}(\hat{\theta}_n)$ and $C_{1-\alpha}(\check{\theta}_n)$ have asymptotic confidence coefficient $1-\alpha$. To evaluate the ARE of two proposed estimators, we consider the ratio of the volumes of the corresponding confidence ellipsoids. See Serfling (1985) for further details. We reach the next conclusion.

Theorem 4.1. *Under the same conditions of the Theorem 3.1, the ARE of* $\hat{\theta}_n$ *with respect to* $\check{\theta}_n$ *is* $\frac{[g(0)]^2\sigma^2}{\beta(1-\beta)}$.

This result implies that the regression quantile estimators has the strictly smaller asymptotic confidence region than in LS estimators. In particular, under asymmetric error distribution, quantiles than median may reveal the information about unknown parameter θ_o in the model (1.1). Thus, the above result implies that the regression quantile estimators is relatively more efficient than LS estimators in the nonlinear regression model.

5. Appendix

Proof of theorem 2.1

First, we prove that

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

where $o_p(1)$ denotes convergence in probability. For this, 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

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

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 in Serfling(1985), we get

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

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

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

(A.1) follows from Assumption A.

Next, in virtue of (A.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. Let $Q(\theta;\beta) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_\epsilon X_t$ and $N_\tau(\theta_o) = \{\theta : \|\theta-\theta_o\| \le \tau\}$. Since $R(\theta) = N_\tau^c(\theta_o)\cap\Theta$ is compact, there exists θ^* such that

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

Note that

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

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_1, 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

$$\begin{aligned} \lim_{n\to\infty}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}\int_{d_t(\theta)}^{0}(\lambda - d_t(\theta))g(\lambda)\, d\lambda &> \int_{\xi_1}^{\xi_2}(\lambda - d_t(\theta))g(\lambda)\, d\lambda \\ &> \eta_1\int_{\xi_1}^{\xi_2}(\lambda - d_t(\theta))\, d\lambda. \end{aligned} \qquad (A.2)$$

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 A_3 and above fact we obtain

$$\inf_{\|\theta-\theta_o\|>\delta} E_{\epsilon\times x}Q_n(\theta;\beta) \ge \inf_{\|\theta-\theta_o\|>\delta}\int_\omega\int_R X_1 dG(\lambda)dH(x), \qquad (A.3)$$

where $\omega = \{x\in\Omega | f(x,\theta) \ne f(x,\theta_o)\}$. In virtue of (A.2) and (A.3), we get for sufficiently large n

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

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

Proof of theorem 3.1

For the first aim, note that $n[S_n^*(\theta;\beta) - S_n(\theta;\beta)] \le \sum_{t=1}^{n}\frac{1}{4\alpha_n}I_{\{|r_t(\theta)|\le\frac{1}{\alpha_n}\}}$. Thus Chebyshev's inequality gives

$$P[|n\{S_n^*(\theta;\beta) - S_n(\theta;\beta)\}| > \epsilon] \le \frac{n\max_{1\le t\le n} Var I_{\{|r_t(\theta)|\le\frac{1}{\alpha_n}\}}}{16\alpha_n^2\epsilon^2}.$$

In addition, since $E[I_{\{|r_t(\theta)|\le \frac{1}{\alpha_n}\}}] = \frac{2}{\alpha_n}(g(\eta_n))$, where $\eta_n = o(1)$ and $n^2 = o(\alpha_n^3)$, we have

$$n\{S_n^*(\theta;\beta) - S_n(\theta;\beta)\} = o_p(1).$$

On the other hand, because $\{S_n^*(\theta;\beta) - S_n(\theta;\beta)\}$ is continuous and $\Theta(\beta)$ is compact, there is a θ^* on $\Theta(\beta)$ such that

$$\sup_{\theta\in\Theta(\beta)} \{S_n^*(\theta;\beta) - S_n(\theta;\beta)\} = \{S_n^*(\theta^*;\beta) - S_n(\theta^*;\beta)\}.$$

Hence we obtain

$$\sup_{\theta\in\Theta(\beta)} n\{S_n^*(\theta;\beta) - S_n(\theta;\beta)\} = o_p(1). \tag{A.4}$$

Thus the proof of (i) is completed.

For the second purpose, note that

$$S_n^*(\hat{\theta}_n;\beta) - S_n^*(\tilde{\theta}_n;\beta) \le [S_n^*(\hat{\theta}_n;\beta) - S_n(\hat{\theta}_n;\beta)] + [S_n(\tilde{\theta}_n;\beta) - S_n^*(\tilde{\theta}_n;\beta)].$$

From the fact (A.4), it follows that

$$n\{S_n^*(\hat{\theta}_n;\beta) - S_n^*(\tilde{\theta}_n;\beta)\} = o_p(1).$$

Also by a Taylor expansion, we have $S_n^*(\hat{\theta}_n;\beta) - S_n^*(\tilde{\theta}_n;\beta)$ is equal to

$$\nabla S_n^*(\tilde{\theta}_n;\beta)(\hat{\theta}_n - \tilde{\theta}_n) + \frac{1}{2}(\hat{\theta}_n - \tilde{\theta}_n)^T \nabla^2 S_n^*(\bar{\theta}_n;\beta)(\hat{\theta}_n - \tilde{\theta}_n),$$

where $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and $\tilde{\theta}_n$. Since $\nabla S_n^*(\tilde{\theta}_n;\beta)$ is equal to zero and $\nabla^2 S_n^*(\bar{\theta}_n;\beta)$ is symmetric matrix, Courant-Fisher minimax characterization and Assumption C_1 yield

$$n(\hat{\theta}_n - \tilde{\theta}_n)^T(\hat{\theta}_n - \tilde{\theta}_n) \le \frac{2n}{\lambda_n}\{S_n^*(\hat{\theta}_n;\beta) - S_n^*(\tilde{\theta}_n;\beta)\},$$

where λ_n is the smallest eigenvalue of $\nabla^2 S_n^*(\bar{\theta}_n;\beta)$. Thus, it suffices to prove that $\nabla^2 S_n^*(\bar{\theta}_n;\beta)$ converges to a positive definite. Since

$$E[h_n''(r_t(\bar{\theta}_n))] = g(0) + o(1),$$

and

$$Var[h_n''(r_t(\bar{\theta}_n))] \le \frac{\alpha_n}{2} g(0) + o(1),$$

Chebyshev's inequality gives

$$\frac{1}{n}\sum_{t=1}^{n} h_n''(r_t(\bar{\theta}_n)) = g(0) + o(1)$$

because of $\alpha_n = o(n)$. Now, we consider proof of the final object. From the Mean Value Theorem, we obtain

$$\nabla S_n{}^*(\tilde{\theta_n};\beta) - \nabla S_n{}^*(\theta_o;\beta) = \nabla^2 S_n{}^*(\bar{\theta}_n;\beta)(\tilde{\theta_n} - \theta_o),$$

where $\bar{\theta}_n$ lies in the interior of the line segment joining θ_o and $\tilde{\theta_n}$. It follows that

$$\sqrt{n}(\tilde{\theta_n} - \theta_o) = -\nabla^2 S_n^*(\bar{\theta}_n;\beta)^{-1}\sqrt{n}\nabla S_n^*(\theta_o;\beta).$$

Because of $\tilde{\theta_n}$ converges to θ_o almost surely(a.s.), it suffices to prove normality of $-\sqrt{n}\nabla S_n^*(\theta_o;\beta)$. Since

$$E[\sqrt{n}(h_n'(\epsilon_t))I_{\{|\epsilon_t|\le\frac{1}{\alpha_n}\}}]^{1+\delta} \le \frac{2(n\beta^2)^{\frac{1+\delta}{2}}}{\alpha_n}(g(\xi_n)),$$

where $h_n'(x) = \Big[-\frac{3}{2}\alpha_n{}^2x^2 + \frac{\alpha_n}{2}x + \beta\Big]I_{\{|x|\le\frac{1}{\alpha_n}\}} + \beta I_{\{x>\frac{1}{\alpha_n}\}} + (\beta-1)I_{\{x<-\frac{1}{\alpha_n}\}}$, $\xi_n = o(1)$, Markov's theorem gives

$$\frac{1}{n}\sum_{t=1}^{n}[\sqrt{n}h_n'(r_t(\varepsilon_t))I_{\{|r_t(\varepsilon_t)|\le\frac{1}{\alpha_n}\}}] = o_p(1)$$

because of $n^{\frac{1+\delta}{2}} = o(\alpha_n)$. Hence we have $-\sqrt{n}\nabla S_n^*(\theta_o;\beta)$ converges to

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\beta I_{\{\varepsilon_t>\frac{1}{\alpha_n}\}} + (\beta-1)I_{\{\varepsilon_t<-\frac{1}{\alpha_n}\}}]\nabla f_t(\theta_o).$$

For any nonzero vector $\lambda = (\lambda_1,\cdots,\lambda_p)^T$, we first prove the asymptotic normality of $\sum_{t=1}^{n}\lambda^T U_t$, where

$$A_t = \beta I_{\{\varepsilon_t>\frac{1}{\alpha_n}\}} + (\beta-1)I_{\{\varepsilon_t<-\frac{1}{\alpha_n}\}}$$

and

$$U_t = \frac{1}{\sqrt{n}}\sum_{k=1}^{p}\lambda_k\frac{\partial}{\partial\theta_k}f_t(\theta_o)A_t.$$

Since $E(U_t)$ converges to zero and $Var(U_t)$ converges to $\frac{b_t^2}{n}\beta(1-\beta)$ $(<\infty)$ as $n\to\infty$, where $b_t = \sum_{k=1}^{p}\lambda_k\frac{\partial}{\partial\theta_k}f_t(\theta_o)$. By the application of Linderberg Central Limit Theorem and Cramér-Wold device, we have

$$b_t^2 = \Big[\sum_{k=1}^{p}\lambda_k\frac{\partial}{\partial\theta_k}f_t(\theta_o)\Big]^2 = \lambda^T[\nabla f_t(\theta_o)\nabla^T f_t(\theta_o)]\lambda.$$

We conclude that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t \nabla f_t(\theta_o) \xrightarrow{d} N_p(0, \beta(1-\beta)V(\theta_o)).$$

We reach the result of this theorem.

REFERENCES

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

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

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

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

Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, 46, 33-49.

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

Powell, J. L.(1986). Censored regression quantiles, *Journal of Econometrics*, 32, 143-155.

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

Serfling, R. J. (1985). *Approximation theorems of mathematical statistics, John Wiley & Sons.*

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

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