East Asian Math. J 14 (1998), No 2, pp 421–427

THE CONSISTENCY OF NONLINEAR REGRESSION MINIMIZING L_p-NORM

SEUNG HOE CHOI AND KYUNG OK PARK

ABSTRACT. In this paper we provide sufficient conditions which ensure the strong consistency of L_p-norm estimation in nonlinear regression model when the probability distribution of the errors term is symmetric about zero. The least absolute deviation and least square estimation are discussed as special cases of the proposed estimation.

1. Introduction

We consider in this paper the following nonlinear regression model

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

where $x_t \in \Gamma$ is an independent and identically distributed(i.i.d.) random variable, $f : R^m \times R^q \to R$ is a continuous function, the parameter vector θ_o which is interior point of the parameter space $\Theta \subseteq R^q$ is unknown and to be estimated, and the unobservable errors ϵ_t are i.i.d.random variables with continuous probability density function $g(x)$ and $E|\epsilon|^p < \infty$.

The L_p-norm estimation of the true parameter θ_o based on (y_t, x_t), denoted by $\hat{\theta}_{pn}$, is a vector which minimizes the objective function

$$S_{pn}(\theta) = \frac{1}{n}\sum_{t=1}^{n} |y_t - f(x_t, \theta)|^p, \tag{1.2}$$

for $1 \le p < \infty$. Modifying (1.2), we get another objective function of the L_p-norm estimation

$$Q_{pn}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\{|y_t - f(x_t, \theta)|^p - |y_t - f(x_t, \theta_o)|^p\}.$$

Received November 10, 1998.

The asymptotic results of nonlinear L_1-norm and L_2-norm estimation are given by various authors. The asymptotic behavior of Least Absolute Deviation(LAD) estimation ($p = 1$) in the nonlinear regression models was proved by Oberhor(1982). The existence, strong consistency and asymptotic normality of the nonlinear Least Square(LS) estimation $p = 2$ were given by various authors: Jennrich(1969) and Wu(1981) among them. Optimal properties and asymptotic properties of the L_p-norm estimation in linear regression models was proposed by Nyquist(1983).

The main purpose of this paper is to provide simple sufficient conditions for strong consistency of nonlinear L_p-norm estimation by imposing some conditions on input vectors x_t and error terms ϵ_t when p is an element of the natural number.

2. Strong consistency

Let $(\Gamma, \mathcal{A}, P)$ be probability space on R^m and H be a common distribution function of input vector x_t. To simplify the notation, we denote

$$\nabla f_t(\theta) = \left[\frac{\partial}{\partial\theta_1} f(x_t, \theta), \cdots, \frac{\partial}{\partial\theta_q} f(x_t, \theta)\right]_{(q\times 1)},$$

and

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

Throughout this paper, we will use the following assumptions;

Assumption A

A_1 . The parameter space Θ is a compact subspace of R^q.

A_2 : $\nabla f_t(\theta)$ is continuous on $\Gamma \times \Theta$ for each t.

A_3 : $g(x)$ is symmetric and strictly positive at zero.

A_4 : x_t is bounded in probability, i.e.,for every $\eta > 0$ there exists M_η such that $P\{|x_t| > M_\eta\} < \eta$.

First we consider the uniform convergence of the another objective function $Q_{pn}(\theta)$.

LEMMA 2 1. *Suppose that $A_1 - A_3$ in Assumption A are satisfied on the model (1.1). Then we have*

$$|Q_{pn}(\theta) - E\{Q_{pn}(\theta)\}| = o_p(1),$$

where $o_p(1)$ stands for convergence in probability.

Proof. From the fact $|u|^p - |v|^p \le 2^p|u-v|^p + (2^p-1)|v|^p$, we obtain

$$|Q_{pn}(\theta)| \le \frac{1}{n}\sum_{t=1}^{n} 2^p|d_t(\theta)|^p + (2^p-1)|\epsilon_t|^p,$$

where $d_t(\theta) = f(x_t, \theta_o) - f(x_t, \theta)$. Let $X_t(\theta) = 2^p|d_t(\theta)|^p + (2^p-1)|\epsilon_t|^p$. Using the Hölder's inequality, we get

$$|X_t(\theta)| \le 2^p\|\nabla f_t(\theta^*)\|^p\|\theta_o - \theta\|^p + |\epsilon_t|^p,$$

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

$$P\{|Q_{pn}(\theta) - E\{S_{pn}(\theta)\}| > \epsilon\} \le \frac{\max\limits_{1\le t\le n} VarX_t(\theta)}{n\epsilon^2}.$$

Moreover, the boundedness of $VarX_t(\theta)$ follows from Assumption A. Hence the proof is completed.

For the strong consistency of the nonlinear L_p-norm estimation, we need the additional assumption

Assumption B

$B_1 : P\{x \in \Gamma . f(x, \theta_1) \ne f(x, \theta_2)\} > 0$ for each $\theta_1, \theta_2 \in \Theta, \theta_1 \ne \theta_2$.

$B_2 \cdot V_n(\theta_o) = \frac{1}{n}\sum_{t=1}^{n} \nabla f_t(\theta)\nabla^T f_t(\theta)$ converges to a positive definite matrix

$V(\theta_o)$ as $n \to \infty$.

The main result of this paper is the following theorem which provides sufficient conditions for the strong consistency of L_p-norm estimation in the nonlinear model (1.1).

THEOREM 2.1. *Suppose that Assumptions A and B are satisfied on the model (1.1). Then the nonlinear L_p-norm estimation $\hat{\theta}_{pn}$ defined in (1.2) is strongly consistent for θ_o.*

proof. Kim and Choi(1995) and Wu(1981) proved for the case LAD($p = 1$) and LS($p = 2$) estimation respectively. Here we only consider the case $1 < p < \infty, p \neq 2$. It suffices to show that

(i) $Q_p(\theta) = \lim_{n\to\infty} E[Q_{pn}(\theta)]$ has a unique minimizer θ_o on Θ,

(ii) The local minimizer θ_o is indeed the global minimizer,

(iii) The L_p-norm estimation $\hat{\theta}_{pn}$ belongs to $N_\delta(\theta) = \{\theta : \|\theta - \theta_o\| < \delta\}$.

For the first aim, let C and D denote the set of an even number and an odd number respectively. If $d_t(\theta) > 0$, note that

$$E_\epsilon[|\lambda + d_t(\theta)|^p - |\lambda|^p] = \begin{cases} \sum_{s=1}^{p} \binom{p}{s} \int_R \lambda^{p-s} d_t^s(\theta) g(\lambda) d\lambda, & \text{if} \quad p \in C, \\ \sum_{s=1}^{p} \binom{p}{s} \{2\int_{d_t(\theta)}^{\infty} \lambda^{p-s} d_t^s(\theta) - \int_R \lambda^{p-s} d_t^s(\theta) \\ \quad + \int_{d_t(\theta)}^{0} 2\lambda^{p-s} d_t^s(\theta)\} g(\lambda) d\lambda, & \text{if} \quad p \in D. \end{cases} \tag{2.1}$$

Let $U_p(\theta) = \lim_{n\to\infty} E_\epsilon Q_{pn}(\theta)$ In the case of $p \in C$, by simple calculation we have

$$\nabla U_p(\theta_o) = \lim_{n\to\infty} \frac{-p}{n} \sum_{t=1}^{n} \int_R \lambda^{p-1} g(\lambda) d\lambda \nabla f_t(\theta_o),$$

and

$$\nabla^2 U_p(\theta_o) = \lim_{n\to\infty} \Big[\frac{-p}{n} \sum_{t=1}^{n} \int_R \lambda^{p-1} g(\lambda) d\lambda \nabla^2 f_t(\theta_o) + \frac{2p(p-1)}{n} \sum_{t=1}^{n} \int_R \lambda^{p-2} g(\lambda) d\lambda \nabla f_t(\theta_o) \nabla^T f_t(\theta_o) \Big].$$

Using a similar method, in the case of $p \in D$ we get

$$\nabla U_p(\theta_o) = \lim_{n\to\infty} \frac{p}{n} \sum_{t=1}^{n} \left[\int_{-\infty}^{0} - \int_{0}^{\infty} \right] \lambda^{p-1} g(\lambda) d\lambda \nabla f_t(\theta_o),$$

and

$$\nabla^2 U_p(\theta_o) = \lim_{n\to\infty} \left[\frac{p}{n} \sum_{t=1}^{n} \left(-2 \int_{0}^{\infty} + \int_{R} \right) \lambda^{p-1} g(\lambda) d\lambda \nabla^2 f_t(\theta_o) \right.$$
$$\left. + \frac{p(p-1)}{n} \sum_{t=1}^{n} \left(\int_{0}^{\infty} - \int_{-\infty}^{0} \right) \lambda^{p-2} g(\lambda) d\lambda \nabla f_t(\theta_o) \nabla^T f_t(\theta_o) \right]$$

Then, by means of the above formula and the assumptions A_3 and B_2 we have that $\nabla U_p(\theta_o) = 0$ and the Hessan matrix $\nabla^2 U_p(\theta_o)$ is positive definite. Hence, $U_P(\theta)$ attains a local minimizer at θ_o.

Consider $\theta \in N_\delta^C(\theta_o) = \{\theta : |\|\theta - \theta_o\| \geq \delta\}$ for the second purpose. Due to $d_t(\theta) > 0$ and Assumptions A_3 there exist positive constant η, ξ_1 and ξ_2 such that

(i) $0 < \eta < d_t^s(\theta), 1 < s \leq p$,

(ii) $g(x)$ is positive on $[\xi_1, \xi_2]$.

Thus, we have

$$\sum_{s=1}^{p} \binom{p}{s} \int_R \lambda^{p-s} d_t^s(\theta) g(\lambda) d\lambda > \eta \sum_{s=1}^{p} \binom{p}{s} \int_R \lambda^{p-s} g(\lambda) d\lambda$$
$$> 2\eta \sum_{p-s\in C} \binom{p}{s} \int_{\xi_1}^{\xi_2} \lambda^{p-s} g(\lambda) d\lambda > 0,$$

where $p \in C$. On the other hand, in the case of $p \in D$ we know that $E_\epsilon[|\lambda + d_t(\theta)|^p - |\lambda|^p]$ equal to

$$2\left[\sum_{p-s\in C} \binom{p}{s} \int_{d_t(\theta)}^{0} + \sum_{p-s\in D} \binom{p}{s} \int_{d_t(\theta)}^{\infty} \right] \lambda^{p-s} d_t^s(\theta) dG(\lambda)$$
$$+ 2 \int_{d_t(\theta)}^{0} \lambda^p g(\lambda) d\lambda$$

Since $U(\theta,\lambda) = d_t^p(\theta) + \lambda^p \geq 0$ on $[-d_t(\theta), d_t(\theta)]$, we get

$$\begin{aligned} E_\epsilon[|\lambda + d_t(\theta)|^p - |\lambda|^p] >\ & 2\int_{-d_t(\theta)}^{0} (d_t^p(\theta) + \lambda^p) g(\lambda) d\lambda \\ >\ & 2\int_{\xi_1}^{\xi_2} (d_t^p(\theta) + \lambda^p) g(\lambda) > 0, \end{aligned}$$

where $0 < \xi_1 < \xi_2 < d_t(\theta)$. Likewise if $d_t(\theta) > 0$, we have the similar result. Therefore,

$$U_p(\theta) \geq \min_{1\leq t\leq n} \sum_{s=1}^{p} \binom{p}{s} \int_R \lambda^{p-s} d_t^s(\theta) g(\lambda) d\lambda > 0.$$

Now, we have to consider the last aim. According to above result we get

$$U_p(\theta_o) < \inf_{\theta\in N_\delta^c\cap\Theta} U_p(\theta).$$

Because

$$\begin{aligned} & |U_p(\hat\theta_{pn}) - U_p(\theta_o)| \\ \leq\ & |U_p(\hat\theta_{pn}) - Q_{pn}(\hat\theta_{pn})| + |Q_{pn}(\hat\theta_{pn}) - Q_{pn}(\theta_o)| + |Q_{pn}(\theta_o) - U_p(\theta_o)|, \end{aligned}$$

by means of Lemma 2 1, we obtain with probability greater than $1-\eta$

$$U_p(\hat\theta_{pn}) < U_p(\theta_o) + \epsilon < \inf_{\theta\in N_\delta^c\cap\Theta} U_p(\theta) + \epsilon.$$

The proof follows the following fact

$$Q_p(\theta) = \int_{R^m}\int_R \{|\lambda + d(\theta)|^p - |\lambda|^p\} dG(\lambda) dH(x).$$

REMARK 2.1. If p is the element of $R\backslash N$ we require the supplementary assumption:

the density function $g(x)$ is a monotone decreasing on R^+.

References

[1] R. I. Jennrich, *Asymptotic properties of nonlinear least squares estimations*, The Annal of Mathematical Statistics **40** (1969), 633–643
[2] H K Kim and S. H. Choi, *Asymptotic Properties of Nonlinear Least Absolute Deviation estimation*, Journal of Korean Statististics Society **24** (1995), 127–139
[3] E. Malinvaud, *The consistency of nonlinear regression*, Annals of Mathematical Statistics **41** (1970), 956–969
[4] H Nyquist, *The optimal L_p-norm estimation in linear regression models*, Communication of Statistics-Theory and Method **12** (1983), 2511–2524
[5] W. Oberhofer, *The consistency of nonlinear regression minimizing the L_1-norm*, The Annals of Statistics **10** (1982), 316–319
[6] C F Wu, *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

Kyung Ok Park
Department of Mathematics
Yonsei University
Seoul120-749, Korea