Bull. Korean Math. Soc. **33** (1996), No. 3, pp. 377–383

THE CONSISTENCY ESTIMATION IN NONLINEAR REGRESSION MODELS WITH NONCOMPACT PARAMETER SPACE

SEUNG HOE CHOI, HAE KYUNG KIM AND SOOK HEE JANG

1. Introduction

We consider in this paper the following nonlinear regression model

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

where y_t is the tth response, x_t is m-vector input variable, θ_o is a p-vector of unknown parameter belong to a parameter space Θ, $f: R^m \times \Theta \to R^1$ is a nonlinear known function, and ϵ_t are independent unobservable random errors with finite second moment.

The L_1-norm estimator of θ_o based on (y_t, x_t), denoted by $\hat{\theta}_n$, is a vector which minimizes the mean absolute deviation

$$D_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} |r_t(\theta)|, \tag{1.2}$$

where $r_t(\theta) = y_t - f(x_t, \theta)$. The L_1-norm estimator is a paticular case with $\rho(x) = |x|$ of a general class of robust methods based on minimizing

$$S_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \rho(r_t(\theta)), \tag{1.3}$$

Received September 21, 1995. Revised May 3, 1996.

1991 AMS Subject Classification: 62J02.

Key words and phrases: Strong consistency, L_1-norm estimator, Nonlinear regression model.

This work was supported by the Basic Science Research Institute Program Ministry of Education, 1995.

where ρ is a convex function on R.

The asymptotic properties of the nonlinear least square estimator are investigated by Jennrich(1969) and Wu (1981) when the parameter space is a compact subset of R^p. In a recent paper, Shao (1993) proved the strong consistency of nonlinear least square estimator under more general conditions. For the L_1-norm estimator, Oberhober (1982) showed weak consistency of $\hat{\theta}_n$, and Kim and Choi (1995) gave sufficient conditions for strong consistency and asymptotic normality of $\hat{\theta}_n$ when the parameter space is a compact set. Richardson and Bhattacharyya (1987) proposed sufficient condition for strong consistency when the parameter space is a noncompact set. In addition, they assumed that the regression function $f(x, \theta)$ is bounded for compactification. However, the regression function $f(x, \theta)$, in many situations, is unbounded when the parameter space is noncompact subset of R^p. For this, we now discuss an example given by many authors.

EXAMPLE 1. Consider the exponential model

$$y_t = \theta_1 e^{-\theta_2 x_t} + \epsilon_t, \ (\theta_1, \theta_2) \in \Theta = \{\theta : \theta_1 \neq 0, 0 < \theta_2 < d\},$$

where d is a fixed positive real number. Since Θ is noncompact subset of R^2 and $f(x, \theta)$ is unbounded, Oberhober's condition and Richardson's condition do not hold in this example.

The main purpose of this paper is to provide simple sufficient conditions for the strong consistency of the L_1-norm estimator and ρ-estimator, denoted by $\tilde{\theta}_n$, which is minimizing (1.3) when the parameter space is a noncompact subset of R^p and the regression function $f(x, \theta)$ is unbounded.

2. Strong Consistency

In this section we will present sufficient conditions for strong consistency of the L_1-norm estimator in model (1.1). In fixed-regression approach, there are two types of input vectors x_t's;

Deterministic regressor: The vector x are nonrandom and $\|x\| \leq b_o$, where b_o is a positive constant.

Conditional regressor: The vector x are independently and identically distributed with distribution function F and x is bounded in probability. i.e, for any $\epsilon > 0$, there exists $a > 0$ such that $P\{\|x\| > a\} < \epsilon$.

Let θ^* be any fixed parameter in Θ. Let Λ be a ray from true parameter θ_o. i.e., $\Lambda = \{\mu \in \Theta : \theta = \eta(\lambda\theta_o + (1-\lambda)\theta^*), \theta \leq \lambda \leq 1, \eta \in R^+\}$. We will use the following conditions throughout paper.

CONDITION A.

Condition A_1. $\lim_{\|\theta\| \to \infty} |f(x,\theta)| = \infty$.

Condition A_2. For each $\theta \in \Lambda$, there exists a constant γ_Λ such that

$$\lim_{\|\mu\| \to \infty} f(x,\mu) = \gamma_\Lambda.$$

Many regression models which are occurred in statistical problems satisfy the condition A_1 or A_2.

Note that Θ is a compact set, we can construct a function h on R^p such that $h|_\Theta = f$ and $\lim_{\|\theta\| \to \infty} g(x,\theta) = f(x,a)$, where $a \in B(\Theta) \cap \Lambda$ and $B(\Theta)$ is boundary set of parameter space. Hence, the function $f(x,\theta)$ has a continuous extention which satisfies the condition A_2.

Let G_t denote the distribution function of ϵ_t and P_X the probability measure on R^m. For the strong consistency, we will assume the following

ASSUMPTION B.

B_1. For each t, the function $f_t(\theta) = f(x_t, \theta)$ is continuous .

B_2. ϵ_t and X are independent and ϵ_t has a unique median at zero.

B_3. $P_X\{x \in R^m | f(x,\theta) \neq f(x,\theta_o)\} > 0$ for each $\theta \neq \theta_o$.

B_4. There exists a $b(x)$ such that $|f(x,\theta)f(x,\theta')| \leq b(x)$ and $E(b^2(x)) < \infty$, for finite θ and θ'.

REMARK. The assumption B_4 always holds in deterministic regressor because of the assumption B_1. For finite $\theta \neq \theta_o$, the assumption B_4 implies that $\frac{1}{n}\sum_{t=1}^n \{d_t(\theta)\}^2$ converges to a positive continuous function uniformly due to the assumption B_3 and that $\lim_{n\to\infty} \inf_\theta \frac{1}{n}\sum_{t=1}^n \{d_t(\theta)\}^2 > 0$ given in Shao (1993), where $d_t(\theta) = f(x_t,\theta) - f(x_t,\theta_o)$.

In following theorem we provide the sufficient conditions for the strong consistency of the L_1-norm estimator in model (1.1).

THEOREM 2.1. *Under the condition A, assume that the model* (1.1) *satisfies the assumption B. Then the L_1-norm estimator $\hat{\theta}_n$ defined on* (1.2) *is strong consistent for θ_o.*

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

$$\lim_{n\to\infty} \inf_{\|\theta-\theta_o\|>\delta} \{D_n(\theta) - D_n(\theta_o)\} > 0 \text{ a.e.} \tag{2.1}$$

First, assume the condition A_1. Let $N = \{t : |\epsilon_t| > a\}$, where $a > 1$. Then

$$\begin{aligned} D_n(\theta_o) &= \frac{1}{n}\{\sum_{t\in N^c} |\epsilon_t| + \sum_{t\in N} |\epsilon_t|\} \\ &\le a + \frac{1}{n}\sum_{t\in N} \epsilon_t^2. \end{aligned}$$

Since ϵ_t has finite variance, there exists a positive constant M such that $D_n(\theta_o) \le M$ for sufficiently large n. Let $B_m = \{\theta : \|\theta - \theta_o\| \le \eta_m\}$ where η_m is a strictly increasing sequence. Then B_m is increasing sequence and $\lim_{m\to\infty} B_m = R^p$. Due to the condition A_1, with probability greater than $1-\epsilon$, we get

$$|d_t(\theta)| \to \infty \quad \text{as} \quad \|\theta\| \to \infty$$

for any ϵ and all t. Thus, we can choose m' such that $D_n(\theta_n) = \frac{1}{n}\sum_{t=1}^n |\epsilon_t + d_t(\theta_n)| > M$, for all $\theta \in B_{m'}^c$. Therefore the L_1-norm estimator $\hat{\theta}_n$ belong to $B_{m'}$. (2.1) follows boundness of $\{\hat{\theta}_n\}$ and the result of theorem 2.1 in Kim and Choi (1995).

Assume that the condition A_2. Let Γ be the set of all ray from θ_o, denoted by Λ, and $\Lambda_m = \{\mu \in \Lambda : \delta < \|\mu - \theta_o\| \le \eta_m\}$. It is enough to show that

$$\inf_{\Gamma} \lim_{m\to\infty} \inf_{\mu\in\Lambda_m} \{D_n(\mu) - D_n(\theta_o)\} > 0 \quad \text{a.e.}$$

for sufficiently large n. For this, let $H_m^\Lambda = \Lambda_m \cap \Lambda_{m-1}^c (m \ge 2)$. In virtue to the condition A_2, there exists m_Λ such that $d_t(\mu) \cong f_t(\theta_o) - \gamma_\Lambda$ for $\mu \in H_m^\Lambda$ and $m \ge m_\Lambda$. ($\cong$ denotes asymptotically equivalent). Due to the assumption B_4, with probability greater than $1-\epsilon$ we can choose b(x) such that $d_t^2(\mu) \leq b_t(x)$ and $Eb_t^2(x) < \infty$ for each $\mu \in H_m^\Lambda$ and $m \geq m_\Lambda$. From the strong law of large number (SLLN) for non - i.i.d case,

$$D_n(\mu) - D_n(\theta_o) = \frac{1}{n}\sum_{t=1}^{n} E\{|\epsilon_t + f_t(\theta_o) - \gamma_\Lambda| - |\epsilon_t|\} + o(1).$$

Note that

$$\begin{aligned} &E\{|\epsilon_t + f_t(\theta_o) - \gamma_\Lambda| - |\epsilon_t|\} \\ &\qquad = \int_I \{|\epsilon - c_1(\mu)| - |\epsilon|\} dG_t(\epsilon) dF(x_1) \\ &\qquad = \int_I \{s(r_1(\mu))\epsilon - s(\epsilon)\epsilon - s(r_1(\mu))c_1(\mu)\} dG_t(\epsilon) dF(x_1), \end{aligned}$$

where $I = R^m \times R, s(x) = sign(x)$, and $c_t(\mu) = \gamma_\Lambda - f_t(\theta_o)$. Next, by a simple calculation, we obtain

$$E\{|\epsilon_t - c_t(\mu)| - |\epsilon_t|\} = \begin{cases} 2\int_{R^m}\int_0^{c_1(\mu)} (c_1(\mu) - \epsilon) dG_t(\epsilon) dF(x_1), & \text{if } c_1(\mu) > 0 \\ 2\int_{R^m}\int_{c_1(\mu)}^0 (\epsilon - c_1(\mu)) dG_t(\epsilon) dF(x_1), & \text{if } c_1(\mu) < 0 \end{cases}$$

On the other hand, for $m \geq m_\Lambda$ we have

$$\inf_{H_m^\Lambda} E\{|\epsilon_t - c_1(\mu)| - |\epsilon_t|\} \geq \inf_{H_m^\Lambda} \int_w \int_R \{|\epsilon - c_1(\mu)| - |\epsilon|\} dG_t(\epsilon) dF(x_1),$$

where $w = \{x \in R^m | f(x,\mu) \neq f(x,\theta_o)\}$. Moreover, there exists a positive number s less than $c_1(\mu)$ such that $\int_0^s dG_t(\epsilon) > 0$ because of the assumption B_2. Hence,

$$\inf_{H_m^\Lambda} \frac{1}{n}\sum_{t=1}^{n} E\{|\epsilon_t - c_t(\mu)| - |\epsilon_t|\} \geq \tau_\Lambda,$$

where τ_Λ is a positive number. Let $m^* = \sup\{m_\Lambda : \Lambda \in \Gamma\}$. Then, for $m > m^*$

$$\inf_{\|\theta-\theta_o\| \geq \eta_{m^*}} \{D_n(\theta) - D_n(\theta_o)\} = \inf_{\Gamma} \inf_{\mu \in \Lambda_{m^*}} \{D_n(\mu) - D_n(\theta_o)\} > \tau_1 \quad \text{a.e.},$$

where $\tau_1 = \inf\{\tau_\Lambda : \Lambda \in \Gamma\}$.

From the theorem 2.1 in Kim and Choi (1995), we have

$$\inf_{\delta \leq \|\theta-\theta_o\| \leq \eta_{m^*}} \{D_n(\theta) - D_n(\theta_o)\} \geq \tau_2 \quad \text{a.e.},$$

where τ_2 is a positive number. Hence, for sufficiently large n we have

$$\inf_{\|\theta-\theta_o\| > \delta} \{D_n(\theta) - D_n(\theta_o)\} > \tau \text{ a.e.},$$

where $\tau = \min\{\tau_1, \tau_2\}$. The proof is completed.

For the applications of the theorem 2.1, we consider now the nonlinear regression model with noncompact parameter space.

EXAMPLE 2. Let s be a fixed positive real number. Consider the logistic model

$$y_t = f(x, \theta) = \frac{\theta_3}{1 + e^{-\theta_1(x-\theta_2)}} + \epsilon_t,$$

where $\theta_o \in \Theta = (0, \infty) \times (-s, s) \times (0, \infty)$ and $-\infty < x < \infty$. Assume that ϵ_t are independent random variable having median zero uniquely. We can check easily that the condition A_2, and the assumption B_1 and B_4 are satisfied. Since $f(x, \theta) = f(x, \theta')$ if and only if $\theta_1 = \theta_1', \theta_2 = \theta_2'$, and $\theta_3 = \theta_3'$, the regression function satisfy the assumption B_3. Under same conditions, we can show that the L_1 norm estimator in exponential model converges to θ_o almost surely.

For the sufficient condition of ρ-estimator, we impose upon an identifiable. The true parameter θ_o is identifiable if for each neighborhood V of θ_o, there exists n_o and $\epsilon > 0$ such that $E(S_n(\theta)) - E(S_n(\theta_o)) \geq \epsilon$ for each $\theta \in V^c, n \geq n_o$. (See [4].) The next result concerns with the strong consistency of the ρ-estimator.

THEOREM 2.2. *Assume that the model* (1.1) *satisfies the assumptions* B_1 *and* B_2, θ_o *is identifiable, and that* $E\rho^2(r_t(\theta)) < \infty$ *for finite* $\theta \in \Theta$. *Then the* ρ*-estimator* $\tilde{\theta}_n$ *which minimizing* (1.3) *converges almost surely to* θ_o.

Proof. The proof is similar to that of theorem 2.1 because of the property of convex and continuous function.

References

1. Jennrich, R. I., *Asymptotic Propeties of Nonlinear Least Squares Estimators*, Annal of Mathematical Statistics **40** (1969), 633-643
2. Kim, H. K. and Choi, S. H., *Asymtotic Properties of Nonlinear Least Absolute Deviation Estimators*, Journal of Korean Statististics Society **24** (1995), 127-139.
3. Oberhofer, W., *The consistency of nolinear regression minimizing the* L_1*-Norm*, Annal of Statistics **43** (1982), 316-319.
4. Richardson, G. D. and Bhattacharyya, B. B., *Consistent* L_1*-Estimators in Nonlinear Regression for a Noncompact Parameter Space*, Sankhyā **49 Series A** (1987), 377-387.
5. Wu, C. F., *Asymptotic Theory of Nonlinear Least Squares Estimation*, Annal of Statistics **9** (1981), 501-513.
6. Shao, J., *Consistency of Least Squares Estimator and its Jackknife Variance Estimator in Nonlinear Models*, The Canadian Journal of Statistics **20** (1992), 415-428.

DEPARTMENT OF MATHEMATICS, YONSEI UNIVERSITY, SEOUL 120-749, KOREA