Bull. Korean Math. Soc. **34** (1997), No. 4, pp. 573–581

THE STRONG CONSISTENCY OF THE L_1-NORM ESTIMATORS IN CENSORED NONLINEAR REGRESSION MODELS

SEUNG HOE CHOI AND HAE KYUNG KIM

ABSTRACT. This paper is concerned with the strong consistency of the L_1-norm estimators for the nonlinear regression models when dependent variables are subject to censoring, and provides the sufficient conditions which ensure the strong consistency of L_1-norm estimators of the censored regression models.

1. Introduction

We consider in this paper the following censored nonlinear regression models

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

where $x_t \in \Omega \subset R^m$ denotes the t-th fixed input value, the true parameter vector θ_o is an interior point of parameter space $\Theta \subset R^p, f$ is a real valued function on $R^m \times \Theta, \epsilon_t$ are independent unobservable errors with finite second moment and y_t is the t-th dependent value which are censored from left at fixed censoring time c_t. Unlike general regression model, in censored regression we observe (y_t^*, x_t, δ_t), where

$$y_t^* = \begin{cases} f(x_t, \theta_o) + \epsilon_t, & \text{if } y_t < c_t \text{ (uncensored)} \\ c_t, & \text{if } y_t \geq c_t \text{ (censored)}, \end{cases}$$

Received October 15, 1996.

1991 Mathematics Subject Classification: 62J02.

Key words and phrases: Censored regression models, Strong consistency, L_1-norm estimators.

This paper was supported in part by Hankuk Aviation University Research Funds in 1997.

for some censoring c_t and the indicator variable δ_t,

$$\delta_t = \begin{cases} 0, & \text{if } y_t < c_t \\ 1, & \text{if } y_t \geq c_t \end{cases}.$$

Censored regression model, which contain only partial information about the random variable of interest, occurs frequently in medicine, biology, engineering, etc. In experiment to investigate resistance of rat for a tubercle bacillus, survival time of rats infected with a tubercle bacillus may be longer than the observation time of investigator (censoring time). In this case, the resistance of rats for a tubercle bacillus can be considered as a missing value in standard regression model. Unlike this, in censored regression analysis the researcher observation time deal with the effect of rats for a tubercle bacillus in place of the survival time of rats.

The L_1-norm estimator of the true parameter θ_o based on (y_t^*, x_t, δ_t), denoted by $\hat{\theta}_n$, is a parameter which minimizes the objective function

$$(1.2) \qquad D_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} |y_t - \min\{c_t, f(x_t, \theta)\}|.$$

Modifying (1.2), we get another objective function of the L_1-norm estimator

$$(1.3) \quad Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\{|y_t - \min\{c_t, f_t(\theta)\}| - |y_t - \min\{c_t, f_t(\theta_o)\}|\},$$

where $f_t(\theta) = f(x_t, \theta)$.

Asymptotic results for censored linear regression models, response function $f(x, \theta)$ is linear about the parameter θ, are given various authors; Amemiya(1973), Powell(1984, 1986) and Chen and Wu(1994).

The consistency and asymptotic normality of maximun likelihood estimation (MLE) were proved by Amemiya(1973). Powell(1984, 1986) introduced the censored least absolute deviation (LAD) and censored regression quantile estimators, and discussed asymptotic properties of the proposed censored estimators under some regularity conditions. In a recent paper, Chen and Wu(1994) given somewhat different sufficient conditions which employed in powell(1984) for the strong consistency of the L_1-norm estimator $\hat{\theta}_n$.

The main purpose of this paper is to provide simple and practical sufficient conditions for the strong consistency of the L_1-norm estimators in the censored nonlinear regression model (1.1). For this, we give preliminary lemmas needed in the proof of the main theorem in section 2. We present the main result which have given sufficient conditions for the strong consistency of the L_1-norm estimators in section 3.

2. Preliminaries

Let $(\Omega, \mathcal{A}, P)$ be probability space on R^m, and G_t and H_t denote the distribution function of error term ϵ_t and input variable X_t, respectively. Suppose that X_t is independent random variable and H_t is not degenerate. To simply the notations, we denote

$$\nabla f(\theta) = \left(\frac{\partial}{\partial\theta_1} f(x,\theta), \ldots, \frac{\partial}{\partial\theta_p} f(x,\theta),\right)_{(p\times 1)},$$
$$\nabla^2 f(\theta) = \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x,\theta)\right]_{(p\times p)},$$
$$V_t(\epsilon_t, \theta, \theta_o) = |y_t - \min\{c_t, f_t(\theta)\}| - |y_t - \min\{c_t, f_t(\theta_o)\}|.$$

Throuhgout this paper, we will use the following assumptions;

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

ASSUMPTION B.
B_1 : $f(x_t, \theta)$ and $f(\theta)$ are continuous on $\Omega \times \Theta$ for each t.
B_2 : x_t and ϵ_t are independent and ϵ_t has a unique median at zero.
B_3 : x_t is bounded in probability, i.e.,for every $\eta > 0$ there exists M_η such that $P\{|x_t| > M_\epsilon\} < \epsilon$.

First we consider the limit of the objective function $Q_n(\theta)$.

LEMMA 2.1. *Suppose that Assumption A and B satisfied on the model* (1.1), *for any* θ *we have*

$$Q_n(\theta) - E[Q_n(\theta)] = o_p(1),$$

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

Proof. Enough to show that $V_t(\epsilon, \theta, \theta_o)$ is bounded by virtue of Chebyshev's inequality, which gives

$$P\{|Q_n(\theta) - E[Q_n(\theta)]| > \epsilon\} \leq \frac{\max\limits_{1\leq t\leq n} Var[V_t(\epsilon_t, \theta, \theta_o)]}{n\epsilon^2}.$$

Meanwhile, we can derive $V_t(\epsilon, \theta, \theta_o)$ into four part for each t.

(2.1)

$$V_t(\epsilon, \theta, \theta_o) \leq \begin{cases} 0, & \text{on } \Omega_1 = \{x \in \Omega : c_t < f_t(\theta), c_t < f_t(\theta_o)\ \} \\ |f_t(\theta) - f_t(\theta_o)|, & \text{on } \Omega_2 = \{x \in \Omega : c_t > f_t(\theta), c_t > f_t(\theta_o)\} \\ f_t(\theta_o) - c_t, & \text{on } \Omega_3 = \{x \in \Omega : f_t(\theta_o) < c_t < f_t(\theta)\} \\ c_t - f_t(\theta_o), & \text{on } \Omega_4 = \{x \in \Omega : f_t(\theta) < c_t < f_t(\theta_o)\}. \end{cases}$$

According to Hölder's inequality and (2.1), we get

$$V_t(\epsilon, \theta, \theta_o) \leq |f_t(\theta) - f_t(\theta_o)| \leq \|\nabla f(\bar{\theta})\| \|\theta - \theta_o\|,$$

where$\|.\|$ denote Euclidian norm and $\bar{\theta} = \lambda\theta_o + (1-\lambda)\theta, 0 \leq \lambda \leq 1$. Hence, the boundness of $V_t(\epsilon, \theta, \theta_o)$ follows from Assumption A and B_1. □

To discuss uniform convergence in probability of $Q_n(\theta)$ over parameter space Θ it is necessary to define stochastic equicontinuity given in Newey(1991). As mensioned in Newey, the stochastic equicontinuity is generalization of the ordinary equicontinuity to random variable.

The sequence $\{Q_n(\theta)\}$ is stochastic equicontinuity means for any $\epsilon, \eta > 0$ there exists random $\Lambda(\epsilon, \eta)$ and constant $\eta_o(\epsilon, \eta)$ such that for $n \geq \eta_o(\epsilon, \eta), P(|\Lambda(\epsilon, \eta)| \geq \epsilon) < \eta$ and for each θ there is an open set $N(\theta, \epsilon, \eta)$ containing θ with

$$\sup_{\tilde{\theta}\in N(\theta,\epsilon,\eta)} |Q_n(\tilde{\theta}) - Q_n(\theta)| \leq \Lambda(\epsilon, \eta), \quad n \geq \eta_o(\epsilon, \eta).$$

We also need the following lemma which gives the sufficient and necessary condition for uniform convergence. For proof, see Newey(1991).

LEMMA 2.2. *Suppose Assumption A holds. Then* $\sup_{\theta\in\Theta}|Q_n(\theta) - E[Q_n(\theta)]| = o_p(1)$ *if and only if* $\{Q_n(\theta) : \theta \in \Theta\}$ *is stochastic equicontinuity and* $Q_n(\theta) - E[Q_n(\theta)] = o_p(1)$ *for any* $\theta \in \Theta$.

The following lemma states uniform convergence of $Q_n(\theta)$.

LEMMA 2.3. *Under the same condition of Lemma 2.1, we get*

$$\sup_{\theta\in\Theta}|Q_n(\theta) - E[Q_n(\theta)]| = o_p(1).$$

Proof. It suffices to show that $\{Q_n(\theta) : \theta \in \Theta\}$ is stochastic equicontinuity in accordance to the Lemma 2.2. Let $N_\delta(\theta^*) = \{\theta : |\theta - \theta^*| < \delta\}$. By means of Heine-Borel theorem, for each θ we can choose some open ball $N_\delta(\theta_k)$ containing θ. Moreover, $|Q_n(\theta_1) - Q_n(\theta_2)|$ is less than or equal to

(2.2)

$$\frac{1}{n}\sum_{t=1}^{n}\Big\{I_{\{x:x\in\Omega_2\}}|f_t(\theta_1) - f_t(\theta_2)| + I_{\{x:x\in\Omega_3\}}|c_t - f_t(\theta_1)| + I_{\{x:x\in\Omega_4\}}|c_t - f_t(\theta_2)|\Big\},$$

where I denotes indicator function. Let $\Delta_n(\tilde{\theta}, \theta) = \max_{1\le t\le n}\|\nabla f_t(\theta^*)\|, \theta^* = \lambda\tilde{\theta} + (1-\lambda)\theta, 0 \le \lambda \le 1$. By virtue to (2.2), we obtain

$$\sup_{\tilde{\theta}\in N_\delta(\theta_k)}|Q_n(\tilde{\theta}) - Q_n(\theta)| \le \Delta_n(\tilde{\theta}, \theta)\|\tilde{\theta} - \theta\|.$$

Therefore, $\Delta_n(\tilde{\theta}, \theta)$ is bounded in probability from Assumptions B. The proof is completed. □

3. Strong consistency

For strong consistency of the L_1-norm estimators, we need the additional assumption

ASSUMPTION C.

$C_1 : P\{x \in \Omega : f(x, \theta_1) \ne f(x, \theta_2)\} > 0$ for each $\theta_1, \theta_2 \in \Theta, \theta_1 \ne \theta_2$.

$C_2 : \frac{1}{n}\sum_{t=1}^{n} I_{\{c_t > f_t(\theta_o)\}}\nabla^T f_t(\theta_o)\nabla f_t(\theta_o)$ converges to a positive definite matrix as $n \to \infty$.

REMARK. As mentioned in Chen and Wu(1994), for the strong consistency of any estimators of true parameter θ_o, the number of $\{t : c_t > f(x_t, \theta_o\}$ must be large sufficiently. So the factor $I_{\{c_t > f_t(\theta_o)\}}$ appears in Assumption C_2. If the investigtor prolong the observation time infinitely, the number of $\{t : c_t > f(x_t, \theta_o\}$ is same with the number of the observation. Assumtion C_2 thus convert that $\frac{1}{n}\sum_{t=1}^{n} \nabla^T f_t(\theta_o) \nabla f_t(\theta_o)$ converges to a positive definite matrix as $n \to \infty$, which is regular condition that is often refered in nonlinear regression model. See, Kim(1995).

The main result of this paper is the following theorem which provides sufficient conditions for the strong consistency of L_1-norm estimators in the model (1.1).

THEOREM 3.1. *Suppose that Assumption A, B, and C are satisfied on the model* (1.1). *Then the L_1-norm estimators $\hat{\theta}_n$ defined in* (1.2) *is strongly consistent for θ_o.*

Proof. First we show that $Q(\theta) = \lim_{n\to\infty} E[Q_n(\theta)]$ has a unique minimizer θ_o on Θ. For this, let $a_t = c_t - f_t(\theta_o)$ and $d_t(\theta) = f_t(\theta) - f_t(\theta_o)$. By simple calculation, we obtain

$$\begin{aligned} E_{\epsilon_t}[V_t(\epsilon_t, \theta, \theta_o) I_{\{x:x\in\Omega_2\}}] &= E_{\epsilon_t}\{d_t(\theta) I_{\{x:x\in\Omega_2\}} + (|\epsilon_t - d_t(\theta)| - |\epsilon_t|) I_{\{x:x\in\Omega_2^c\}}\} \\ &= 2\int_0^{d_t(\theta)} (d_t(\theta) - \lambda) g_t(\lambda) d\lambda. \end{aligned}$$

By similar method, $E_\epsilon[Q_n(\theta)]$ becomes to

(3.1)
$$\begin{aligned} \frac{2}{n}\sum_{t=1}^{n}\Big\{ & I_{\{x:x\in\Omega_2\}} \int_0^{d_t(\theta)} (d_t(\theta) - \lambda) g_t(\lambda) d\lambda + I_{\{x:x\in\Omega_3\}} \int_0^{a_t} (a_t - \lambda) g_t(\lambda) d\lambda \\ & + I_{\{x:x\in\Omega_4\}} \Big[\int_{d_t(\theta)}^0 (\lambda - d_t(\theta)) g_t(\lambda) d\lambda + \int_{a_t}^0 (a_t - \lambda) g_t(\lambda) d\lambda \Big] \Big\}. \end{aligned}$$

Moreover, from (3.1) we have

$$\nabla S_n(\theta) = \frac{2}{n}\sum_{t=1}^{n} \{I_{\{x:x\in\Omega_2\}} + I_{\{x:x\in\Omega_4\}}\} \int_0^{d_t(\theta)} g_t(\lambda) d\lambda \nabla f_t(\theta)$$

and

$$\nabla^2 S_n(\theta) = \frac{2}{n}\sum_{t=1}^{n}\{I_{\{x:x\in\Omega_2\}} + I_{\{x:x\in\Omega_4\}}\}\Big\{(g_t(d_t(\theta))\nabla^T f_t(\theta)\nabla f_t(\theta) + \int_0^{d_t(\theta)} g_t(\lambda)d\lambda\nabla^2 f_t(\theta)\Big\},$$

where $S_n(\theta) = E_\epsilon[Q_n(\theta)]$. Therefore, we obtain

$$\nabla S_n(\theta_o) = 0$$

and

$$\nabla^2 S_n(\theta_o) = \frac{2}{n}\sum_{t=1}^{n} g_t(0) I_{\{c_t > f_t(\theta_o)\}}\nabla^T f_t(\theta_o)\nabla f_t(\theta_o) \geq \frac{2k}{n}\sum_{t=1}^{n} I_{\{c_t > f_t(\theta_o)\}}\nabla^T f_t(\theta_o)\nabla f_t(\theta_o), \tag{3.2}$$

where $k = \min_{1\le t\le n} g_t(0)$. Let $R_n(\theta) = E_{x\times\epsilon}Q_n(\theta)$. Then, by means of the Assumption C_2 and (3.2) we have that $\nabla R_n(\theta_o) = 0$ and the Hessan matrix $\nabla^2 R_n(\theta)$ is positive definite for sufficiently large n. Hence, $Q_n(\theta)$ attains a local minimum at θ_o.

Next we show that this local minimum is indeed the global minimum. Let $U(\theta) = N_\delta^c(\theta_o) \cap \Theta$. By Heine-Borel theorem and compactness of $U(\theta)$, there is a finite covering of $U(\theta)$ by δ-ball. For each $\theta \in U(\theta)$ we can choose $N_\delta(\theta_i)$ such that $\theta \in N_\delta(\theta_i) \subset U(\theta)$. Since $|R_n(\theta) - R_n(\theta_i)|$ is less than or equal to

$$|R_n(\theta) - Q_n(\theta)| + |Q_n(\theta) - Q_n(\theta_i)| + |Q_n(\theta_i) - R_n(\theta_i)|,$$

from Lemmas in section 2, we have with probability greater than $1-\eta$,

$$R_n(\theta_i) < \inf_{\theta\in N_\delta(\theta_i)} R_n(\theta) + \epsilon.$$

The fact $R_n(\theta_i) > 0$ follows Theorem 2.1 in Choi(1996).

Finally we show that the L_1-norm estimators $\hat{\theta}_n$ belong to $N_\delta(\theta)$. According to above result we get

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

Because

$$|R_n(\hat{\theta}_n) - R_n(\theta_o)| \leq |R_n(\hat{\theta}_n) - Q_n(\hat{\theta}_n)| + |Q_n(\hat{\theta}_n) - Q_n(\theta_o)| + |Q_n(\theta_o) - R_n(\theta_o)|,$$

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

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

The proof is completed. □

References

[1] Amemiya, T., *Regression Analysis when the Dependent Variable is Truncated Normal*, Econometrica **41** (1973), 997-1016.

[2] Chen, X. R. and Wu, Y., *Consistency of L_1 Estimates in Censored Linear Regression Models*, Commmnication in Statistics **23** (1994), 1849-1858.

[3] Choi, S. H., Kim, H. K., and Jang, S. H., *The Consistency Estimation in Nonlinear Regression Models with Noncompact Parameter Space.*, Bulletin of the Korean Mathematical Society **33** (1996), 111-117.

[4] Kim, H. K. and Choi, S. H., *Asymtotic Properties of Nonlinear Least Absolute Deviation Estimators*, Journal of Korean Statististics Society **24** (1995), 127-139.

[5] Newey, W. K., *Uniform Convergence in Probability and Stochastic Equicontinuity*, Econometrica **5** (1991), 1161-1167.

[6] Powell, J. L., *Least Absolute Deviation Estimation for the Censored Regression Model*, Journal of Econometrics **25** (1984), 303-325.

[7] ———, *Censored Regression Quantiles*, Journal of Econometrics **27** (1986), 143-155.

Seung Hoe Choi
Department of Gemeral studies, Hankuk Ayiation University, Koyang
411-791, Korea
E-mail: shchoi@ nic.hangkong.ac.kr

Hae Kyung Kim
Department of Mathematics, Yonsei University, Seoul 120-749, Korea
E-mail: abc1217@bubble.yonsei.ac.kr