The Korean Communications
in Statistics Vol. 6, No. 2, 1999
pp. 345-355

# Asymptotic Properties of a Robust Estimator for Regression Models with Random Regressor[1]

**Sook Hee Chang[2], Hae Kyung Kim[3] and So Young Sohn[4]**

## Abstract

This paper deals with the problem of estimating regression coefficients in nonlinear regression model having random regressor. The sufficient conditions for consistency of the $L_1$-estimator with random regressor are given and discussed in this paper. An example is given to illustrate the application of the main results.

## 1. Introduction

We consider, the following regression model, defined by

$$y_i = f(x_i, \theta_o) + \varepsilon_i \qquad i = 1, 2, \cdots, n. \tag{1.1}$$

where $\theta_o$ is unknown parameter in $\mathbb{R}^p$, $\varepsilon_i$ is random error and $x_i$ is random regressor.

Let $(\Omega, A, P)$ denote the underlying probability space and assume that $y_i$ is a random variable having range $\mathbb{Y}$ and $x_i$ is a random vector having range $\mathbb{X}$ contained in $\mathbb{R}^m$, where both $\mathbb{Y}$ and $\mathbb{X}$ are borel subspaces of their respective Euclidean space. The parameter space $\Theta$ is assumed to be a compact subspace of $\mathbb{R}^p$ and the true parameter $\theta_o$ is contained in $\Theta$.

For each fixed $n$, a $L_1$-estimator, $\widehat{\theta_n}$, of $\theta_o$ is defined as the value of $\theta(\omega)$ in $\Theta$ such that

$$D_n(\theta)(\omega) = \sum_{i=1}^{n} | y_i(\omega) - f(x_i(\omega), \theta) | \tag{1.2}$$

is a minimum, for each fixed $\omega \in \Omega$.

It is well known that $L_1$-estimator is more robust than $L_2$-estimator, particulary in the presence of vertical outliers in the $y$-direction.

Various authors have provided conditions which ensure the existence, consistency of the $L_1$-estimator in the recent two decades. Most of studies are related to the properties of $L_1$-estima

1) The authors wish to acknowledge the financial support of Korea Research Foundation made in the program year of 1997
2) Lecturer, Department of Mathematics, Yonsei University, Seoul, 120-749 Korea
3) Professor, Department of Mathematics, Yonsei University, Seoul, 120-749 Korea
4) Associate Professor, Department of Industrial Systems Engineering, Yonsei University, Seoul, 120-749 Korea

-tor for linear regression models. In the literature there are only few papers dealing with the $L_1$ -estimator problem for nonlinear regression models. Obehofer(1982) studied weak consistency result of nonlinear $L_1$-estimator and Wang(1995) and Kim(1995) proved the consistency and normality of nonlinear $L_1$-estimator. These papers used fixed input variables for showing the asymptotic properties. In many statistical problems, the input variable should be considered as a random variable. Bhattacharyya(1992) showed only strong consistency property of nonlinear $L_1$ -estimator with multiplicative error term and random regressor.

We are concerned about the property of normality of nonlinear $L_1$-estimator with random regressor and additive error term and random regressor. Thus the main purpose of this paper is to show the $\sqrt{n}$-consistency and normality of $L_1$-estimator with random input variable, under simpler and more practical conditions than other papers. We know that if $\widehat{\theta_n}$ is $\sqrt{n}$-consistency, or equivalently, $\sqrt{n}(\widehat{\theta_n}-\theta_o)$is bounded in probability, $\widehat{\theta_n}$ converges to $\theta_o$ in probability at the rate of $1/\sqrt{n}$.

To show the properties of $\sqrt{n}(\widehat{\theta_n}-\theta_o)$, we will first modify the object function $D_n$ as follows:

$$G_n(\phi)=\sum_{i=1}^{n} \mid \varepsilon_i - f_i(\phi) \mid - \mid \varepsilon_i \mid \tag{1.3}$$

where $f_i(\phi)=f(x_i, \theta_o+\frac{1}{\sqrt{n}}\phi)-f(x_i,\theta_o)$ under $\phi\in\sqrt{n}(\Theta-\theta_o)$.

It is easy to see that the vector $\sqrt{n}(\widehat{\theta_n}-\theta_o)$minimizes $G_n(\phi)$ when $\widehat{\theta_n}$ minimizes $D_n$ in (1.2).

Let $G$ denote the distribution of function $\varepsilon_i$ and $F$ the distribution function of $x_i$ having range $\mathbb{X}$. Throughout this paper, we use the following notations;

$$\nabla f(x_i,\theta)=\left[\frac{\partial}{\partial\theta_j}f(x_i,\theta)\right]_{(p\times 1)}$$

$$\frac{1}{n}\sum_{i=1}^{n}\nabla f(x_i,\theta)\nabla^T f(x_i,\theta)=V_n(\theta)$$

$$\int \nabla f(x,\theta)\nabla^T f(x,\theta)dF(x)=V(\theta).$$

## 2. $\sqrt{n}$-consistency of $L_1$-estimator

We will now formally state the regularity conditions needed to find the $\sqrt{n}$-consistency of the $L_1$-estimator.

**Assumption A**

A1. $\{(x_i, \varepsilon_i)\}$ are independent and identically distributed.

A2. For each $i$, $x_i$ and $\varepsilon_i$ are independent.

A3. $\{\varepsilon_i\}$ have a unique median at 0 with finite variance and probability density function $g(t)$ is continuous with $g(0)>0$.

A4. The function $f(x,\theta)$ is continuous on $\mathbb{X}\times\Theta$ and $P\{x\in\mathbb{X}: f(x,\theta)\neq f(x,\theta_o)\}>0$ for each fixed $\theta_o\neq\theta$.

**Theorem 2.1.** Suppose that Assumption $A$ holds for model (1.1). Then $\widehat{\theta_n}$ is $\sqrt{n}$-consistent.

**Proof of theorem 2.1.** According to the theory of probability, we know that if $\widehat{\phi_n}$ converges to 0 almost surely (a.s.), then $\widehat{\phi_n}$ is bounded in probability. Hence it suffices to show that for any $M>0$,

$$\liminf_{n\to\infty} \inf_{\|\phi-0\|>M} [G_n(\phi)-G_n(0)] > 0$$

a.s., due to $G_n(\widehat{\phi_n})-G_n(\phi)\leq 0$ for $\phi\in\sqrt{n}(\Theta-\theta_o)$. Since $G_n(0)$, it is sufficient to show

$$\liminf_{n\to\infty} \inf_{\|\phi\|>M} \frac{1}{n} G_n(\phi) > 0.$$

If we show that $E[\,|\varepsilon_i - f_i(\phi)| - |\varepsilon_i|\,]>0$ on $\|\phi\|>M$, by the strong law of large numbers (SLLN), our proof is finished.

It follows from Assumption A2 that the preceding requirement is equivalent to showing that

$$\int_{R\times X}[\,|t-f_i(\phi)| - |t|\,]dG(t)dF(x)>0.$$

Note that, since the median of $\varepsilon_i$ is zero,

$$\int_R |t|\,dG(t) = -\int_{-\infty}^0 tdG(t) + \int_0^\infty tdG(t).$$

Next, consider the integral $\int_R |t-f_i(\phi)|\,dG(t)$. Observe that, if $f_i(\phi)$, then

$$\int_R |t-f_i(\phi)|\,dG(t) = \int_{-\infty}^{f_i(\phi)}(f_i(\phi)-t)dG(t) + \int_{f_i(\phi)}^\infty (t-f_i(\phi))dG(t)$$

$$= -\int_{-\infty}^0 tdG(t) + \int_0^\infty tdG(t) + 2\int_0^{f_i(\phi)}(f_i(\phi)-t)dG(t).$$

From previous result and A3, we have the fact that

$$\int_R [\,|\,t-f_i(\phi)\,|\,-\,|\,t\,|\,]dG(t)=2\int_0^{f_i(\phi)}(f_i(\phi)-t)dG(t)>0.$$

Likewise, if $f_i(\phi)<0$,

$$\int_R [\,|\,t-f_i(\phi)\,|\,-\,|\,t\,|\,]dG(t)=2\int_{f_i(\phi)}^{0}(t-f_i(\phi))dG(t)>0.$$

Due to A4, we have the fact that

$$\int_{R\times X} [\,|\,t-f_i(\phi)\,|\,-\,|\,t\,|\,]dG(t)dF(x)>0. \quad \square$$

## 3. Normality

In this section, we will study the asymptotic behavior of $\widehat{\phi}_n=\sqrt{n}(\widehat{\theta}_n-\theta_o)$. We will state the condition which is needed to show the property of normality.

### 3.1. Assumptions and main results

**Assumption B**

B1. The function $\nabla f(x,\theta)$ is continuous on $X\times\Theta$ and $\|\nabla f(x,\theta)\|\le h(x)$ for all $x$ and $\theta$, where $h$ is square integrable w.r.t. $F(x)$ on $X$.

B2. $V(\theta_o)$ is positive definite.

**Remark.** According to Jennrich (1969),

(i) $V_n(\theta)$ converges uniformly to $V(\theta)$, due to B1.

(ii) $V_n(\theta)$ converges to $V(\theta_o)$ almost surely as $\|\theta-\theta_o\|\to 0$ and $n\to\infty$.

**Theorem 3.1.** Under Assumption A and B, $\widehat{\phi}_n=\sqrt{n}(\widehat{\theta}_n-\theta_o)$ converges to $Z$ in distribution which is $p$-variate normal vector having distribution $N(0,\frac{1}{(2g(0))^2}V(\theta_o)^{-1})$.

The proof of Theorem 3.1 will be given in Section 3.2 and it is based on the following lemmas. The main idea in proof of Theorem 3.1 can be described as follows: First, since we showed that $\widehat{\phi}_n$ is bounded in probability in Section 2, without loss of generality we find the differentiable function $Q_n(\phi)$ which is the approximating function of $G_n(\phi)$ on $S$, where $S=\{\phi: \|\phi\|\le M \text{ for some } M\}$. Secondly, we try to find that the minimizer $\widetilde{\phi}_n$ of $Q_n(\phi)$

converges to $Z$ in distribution. Finally, we show that $\widehat{\phi}_n$ lies close enough to $\widetilde{\phi}_n$ to share its asymptotic behavior so that we can obtain the desired result.

### 3.2. Proof of Theorem 3.1

Let $D(\varepsilon_i)=I\{\varepsilon_i < 0\}-I\{\varepsilon_i \geq 0\}$, where $I(\cdot)$ denotes the indicator function. Define $R_i(\phi)=|\varepsilon_i-f_i(\phi)|-|\varepsilon_i|-D(\varepsilon_i)f_i(\phi)$ and $W_n(\phi)=\sum_{i=1}^{n}D(\varepsilon_i)f_i(\phi)$. Then we can write

$$G_n(\phi)=EG_n(\phi)+W_n(\phi)+\sum_{i=1}^{n}(R_i(\phi)-ER_i(\phi))$$

**Lemma 3.2.** Assume that the model (1.1) satisfies Assumptions A1-A4 and B1. Then $\sum_{i=1}^{n}[R_i(\phi)-ER_i(\phi)]\rightarrow^{p} 0$ on $S$.

**Proof.** It suffices to show that $E\left[\sum_{i=1}^{n}[R_i(\phi)-ER_i(\phi)]\right]^2$ converges to 0 in order to prove Lemma 3.2. Using the following inequality

$$|R_i(\phi)| \leq 2|f_i(\phi)| I\{|\varepsilon_i| \leq |f_i(\phi)|\}$$

and the independence of $R_i(\phi)$ and $R_j(\phi)$ where $i$ and $j$ are distinct, we have

$$\begin{aligned} E\Big[\sum_{i=1}^{n}[R_i(\phi)-ER_i(\phi)]\Big]^2 &\leq \sum_{i=1}^{n}E[2|f_i(\phi)| I\{|\varepsilon_i| \leq |f_i(\phi)|\}]^2 \\ &\leq 4\sum_{i=1}^{n}E|f_i(\phi)|^2 EI\{|\varepsilon_i| \leq |f_i(\phi)|\}. \end{aligned}$$

Note that

$$\max_{1\leq i\leq n} f_i^2(\phi)=\max_{1\leq i\leq n}\phi^T\Big[\frac{1}{n}\nabla f(x_i,\overline{\theta_o})\nabla^T f(x_i,\overline{\theta_o})\Big]\phi,$$

where $\overline{\theta_o}$ lies in the interior of the line segment joining $\theta_o$ and $\theta_o+\phi/\sqrt{n}$. Since $(1/n)\sum_{i=1}^{n}\nabla f(x_i,\overline{\theta_o})\nabla^T f(x_i,\overline{\theta_o})$ converges to $V(\theta_o)$ as $n\rightarrow\infty$, by Wu(1981), $\max_{1\leq i\leq n}(1/n)\nabla f(x_i,\overline{\theta_o})\nabla^T f(x_i,\overline{\theta_o})$ converges to 0 as $n\rightarrow\infty$. Therefore, we know that

$$\max_{1\leq i\leq n} f_i^2(\phi)\rightarrow 0$$

as $n\rightarrow\infty$, and hence

$$\max_{1\leq i\leq n}|f_i(\phi)|\rightarrow 0$$

as $n \to \infty$. Define $U(t) = EI\{ \mid \varepsilon_i \mid \le t \}$. We know that $U(t) \to 0$ as $t \to 0$. Finally, we know that

$$\begin{aligned} E\Big[ \sum_{i=1}^{n} [R_i(\phi) - ER_i(\phi)] \Big]^2 &\le 4 \sum_{i=1}^{n} E \mid f_i(\phi) \mid^2 EI\{ \mid \varepsilon_i \mid \le \mid f_i(\phi) \mid \} \\ &\le 4U( \max_{1 \le i \le n} \mid f_i(\phi) \mid ) \phi^T V(\theta_o) \phi \\ &\to 0 \end{aligned}$$

as $n \to \infty$, which completes the proof. □

**Lemma 3.3.** Under Assumptions A1-A4 and B1, we obtain

$$\lim_{n \to \infty} E\Big[ \sum_{i=1}^{n} [ \mid \varepsilon_i - f_i(\phi) \mid - \mid \varepsilon_i \mid ] \Big] = g(0) \phi^T V(\theta_o) \phi \text{on} \ S.$$

**Proof.** Assumptions A1 and A2 imply that

$$E[ \mid \varepsilon_i - f_i(\phi) \mid - \mid \varepsilon_i \mid ] = \int_{R \times X} ( \mid t - f_i(\phi) \mid - \mid t \mid ) dG(t) dF(x).$$

We know that

$$\int_R [ \mid t - f_i(\phi) \mid - \mid t \mid ] dG(t) = 2 \int_0^{f_i(\phi)} (f_i(\phi) - t) g(t) dt .$$

Because $g(t)$ is continuous at $t = 0$ and $(f_i(\phi) - t) g(t)$ is non-negative in $t \in (0, f_i(\phi))$, when $\mid f_i(\phi) \mid$ is smaller enough, or equivalently, $f_i(\phi)$ is contained in a neighborhood of 0, we know that

$$\int_R [ \mid t - f_i(\phi) \mid - \mid t \mid ] dG(t) = g(0) f_i^2(\phi) + o( f_i^2(\phi)).$$

Finally, according to Assumption B1 and the fact that $\theta_o + \frac{1}{\sqrt{n}} \phi$ converges to $\theta_o$ as $n \to \infty$ on $S$, we conclude that

$$\begin{aligned} \lim_{n \to \infty} E\Big[ \sum_{i=1}^{n} [ \mid \varepsilon_i - f_i(\phi) \mid - \mid \varepsilon_i \mid ] \Big] &= \lim_{n \to \infty} \sum_{i=1}^{n} \int_X g(0) f_i^2(\phi) dF(x) \\ &= g(0) \phi^T V(\theta_o) \phi \end{aligned}$$ □

**Lemma 3.4.** Under Assumptions A and B , $\sum_{i=1}^{n} D(\varepsilon_i) f_i(\phi)$ converges to $Z^* \phi$ in distribution where $Z^*$ is $p$-variate normal random vector with mean 0 and covariance matrix $V(\theta_o)$ on $S$.

**Proof.** If we show that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(\varepsilon_i) \nabla^T f(x_i, \theta_o)$ converges to $Z^*$ in distribution, the required result follows from the fact that

(i) $\sum_{i=1}^{n} D(\varepsilon_i) f_i(\phi) = \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} D(\varepsilon_i)\nabla^T f_i(x_i, \overline{\theta_o})\right]\phi$, where $\overline{\theta_o}$ lies in the line segment joining $\theta_o$ and $\theta_o + \frac{1}{\sqrt{n}}\phi$.

(ii) $\nabla f(x_i, \overline{\theta_o}) \rightarrow \nabla f(x_i, \theta_o)$ as $n \rightarrow \infty$, due to B1.

For the sake of convenience, let $W_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(\varepsilon_i)\nabla^T f(x_i, \theta_o)$. According to Cramer-Wold device, we will show that , for any nonzero vector $\lambda = (\lambda_1, \lambda_2, \cdot\cdot\cdot, \lambda_p)^T$, $W_n^* \lambda$ converges to $z_*$(univariate normal random variable) in distribution. Now if $W_n^* \lambda$ satisfies Lindeberg condtion, by Lindeberg-Feller Central Limit Theorem, we will have the fact that $W_n^* \lambda$ converges to $z_*$ in distribution. Finally, to show that the Lindeberg condition holds, let $a_i = \sum_{j=1}^{p} \lambda_j \frac{\partial}{\partial \theta_j} f(x_i, \theta_o)$ and $T_i = \frac{1}{\sqrt{n}} a_i D(\varepsilon_i)$. Then $W_n^* \lambda = \sum_{j=1}^{n} T_i$. Let $H_i$ be the distribution function of $T_i$. From the definition of $T_i$, we know that $T_i$ is independent and $E(T_i) = 0$ and $Var(T_i) = \frac{1}{n} Ea_i^2 < \infty$. Set $B_n^2 = Var(W_n^* \lambda)$. For any $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^{n} \int_{\{|T_i| \geq \eta B_n\}} T_i^2 \, dH_i(x)$$

$$= \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^{n} E[T_i^2 I(|T_i| > \eta B_n)]$$

$$= \lim_{n\to\infty} \frac{1}{\sum_{i=1}^{n} Ea_i^2} \sum_{i=1}^{n} E\left[a_i^2 D^2(\varepsilon_i) I\left(\left| a_i \frac{D(\varepsilon_i)}{\sqrt{n}} \right| > \eta \sqrt{\sum_{i=1}^{n} \frac{Ea_i^2}{n}}\right)\right]$$

$$= \lim_{n\to\infty} \frac{1}{\sum_{i=1}^{n} Ea_i^2} \sum_{i=1}^{n} E\left[a_i^2 D^2(\varepsilon_i) I\left(|D(\varepsilon_i)| > \eta \frac{1}{|a_i|} \sqrt{\sum_{i=1}^{n} Ea_i^2}\right)\right]$$

$$\leq \lim_{n\to\infty} \frac{1}{\sum_{i=1}^{n} Ea_i^2} \sum_{i=1}^{n} E\left[a_i^2 D^2(\varepsilon_i) I\left(|D(\varepsilon_i)| > \eta \frac{1}{\max |a_i|} \sqrt{\sum_{i=1}^{n} Ea_i^2}\right)\right]$$

$$= E\left[I\left(|D(\varepsilon_i| > \eta \frac{1}{\max |a_i|} \sqrt{\sum_{i=1}^{n} Ea_i^2}\right)\right] .$$

Since $\frac{\sqrt{\sum_{i=1}^{n} Ea_i^2}}{\max |a_i|} = \frac{\sqrt{n \cdot Var(T_i)}}{\max |a_i|}$ diverges to $\infty$,

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^{n} \int_{\{|T_i| \geq \eta B_n\}} T_i^2 \, dH_i(x) = 0 .$$

Thus Lindeberg condition holds. Moreover, we obtain

$$\begin{aligned} Ea_i^2 &= E\Big[\sum_{j=1}^{p} \lambda_j \frac{\partial}{\partial \theta_j} f(x_i, \theta_o)\Big]^2 \\ &= \lambda^T V(\theta_o)\, \lambda. \quad \square \end{aligned}$$

We finally find the approximating function, $Q_n(\phi)$ to $G_n(\phi)$ such that

$$G_n(\phi) = Q_n(\phi) + \gamma_n(\phi) \quad \text{where} \quad \gamma_n(\phi) \to^p 0 \quad \text{on} \quad S\,. \tag{3.1}$$

Then we can define $Q_n(\phi)$ as

$$Q_n(\phi) = g(0)\phi^T V(\theta_o)\phi + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(\varepsilon_i) \nabla f(x_i, \theta_o)\phi.$$

**Proof of Theorem 3.1.** Let $\widetilde{\phi_n}$ be minimizer of $Q_n(\phi)$. Note that $\widehat{\phi_n} = [\,\widehat{\phi_n} - \widetilde{\phi_n}] + \widetilde{\phi_n}$. The theorem will be proved if we show that

(i) $\widetilde{\phi_n}$ converges to $Z$ in distribution.

(ii) $\widehat{\phi_n} - \widetilde{\phi_n}$ converges to zero in probability.

To (i), we can find the minimizer of $Q_n(\phi)$, by taking the gradient and setting it equal to 0:

$$\nabla Q_n(\phi) = 2g(0)\, V(\theta_o)\phi + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(\varepsilon_i) \nabla f(x_i, \theta_o) = 0.$$

Solving above equation, we can find that the minimizer $\widetilde{\phi_n}$ of $Q_n(\phi)$ is

$$-\frac{1}{2g(0)}\, V(\theta_o)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(\varepsilon_i) \nabla f(x_i, \theta_o).$$

In Lemma 3.4., we showed that $(1/\sqrt{n}) \sum_{i=1}^{n} D(\varepsilon_i) \nabla f(x_i, \theta_o)$ converges to $Z^*$ which is $p$-variate normal random vector in distribution with mean 0 and covariance matrix $V(\theta_o)$. Hence $\widetilde{\phi_n}$ converges to $Z$ in distribution which is $p$-variate normal random vector with mean 0 and covariance matrix $\frac{1}{(2g(0))^2} V(\theta_o)^{-1}$. Now, we will show that $Q_n(\phi)$ has unique minimizer.

If $\phi_1 \neq \phi_2$, we get $(\phi_1 - \phi_2)^T [\nabla Q_n(\phi_1) - \nabla Q_n(\phi_2)] = 2g(0)(\phi_1 - \phi_2)^T V(\theta_o)(\phi_1 - \phi_2)$. Since $g(0)$ is positive and $V(\theta_0)$ is positive definite, $Q_n(\phi)$ is strictly convex . It guarantees uniqueness of $\widetilde{\phi_n}$. From uniqueness and the fact that $\widetilde{\phi_n}$ converges in distribution, it is bounded in probability. Hence, with high probability, $\widetilde{\phi_n} \in S$.

To show (ii), we will prove that for every $\delta > 0$, $P(\| \widehat{\phi_n} - \widetilde{\phi_n} \| > \delta) \to 0$ as $n \to \infty$. To prove, first we rewrite $(1/\sqrt{n})\sum_{i=1}^{n} D(\varepsilon_i)\nabla f(x_i, \theta_o)$ in terms of the minimizer $\widetilde{\phi_n}$ of $Q_n(\phi)$, that is, $(1/\sqrt{n})\sum_{i=1}^{n} D(\varepsilon_i)\nabla f(x_i, \theta_o) = -2g(0)V(\theta_o)\widetilde{\phi_n}$.

Using the fact that $2x^T y = \| x \|^2 + \| y \|^2 - \| x - y \|^2$, we can rewrite $Q_n(\phi)$ as

$$Q_n(\phi) = g(0)(\phi - \widetilde{\phi_n})^T V(\theta_o)(\phi - \widetilde{\phi_n}) - g(0)\widetilde{\phi_n}^T V(\theta_0)\widetilde{\phi_n}.$$

Define the closed ball, $B(\widetilde{\phi_n}, \delta)$ which is centered at $\widetilde{\phi_n}$ with radius $\delta$. Now let us examine the behavior of $G_n(\phi)$ outside $B(\widetilde{\phi_n}, \delta)$. Suppose that $\phi = \widetilde{\phi_n} + \alpha\beta$ with $\alpha > \delta$ and $\beta$ is a $p\times 1$ unit vector. Define $\phi^*$ to be the boundary point of $B(\widetilde{\phi_n}, \delta)$ lying on the segment from $\widetilde{\phi_n}$ to $\phi$, that is, $\phi^* = \widetilde{\phi_n} + \delta\beta$. Put $\triangle_n = \sup_{\phi \in S} \| \gamma_n(\phi) \|$. Obviously, $\triangle_n \to^p 0$. The convexity of $Q_n(\phi)$ implies

$$\begin{aligned} \frac{\delta}{\alpha} Q_n(\phi) + (1 - \frac{\delta}{\alpha})Q_n(\widetilde{\phi_n}) &\geq Q_n(\phi^*) \\ &\geq g(0)\delta^2\beta^T V(\theta_o)\beta - g(0)\widetilde{\phi_n}^T V(\theta_o)\widetilde{\phi_n} \\ &\geq g(0)\delta^2\beta V(\theta_o)\beta + Q_n(\widetilde{\phi_n}) . \end{aligned}$$

Rewrite this as

$$\frac{\delta}{\alpha} Q_n(\phi) - \frac{\delta}{\alpha} Q_n(\widetilde{\phi_n}) \geq g(0)\delta^2\beta^T V(\theta_o)\beta$$

implying that

$$Q_n(\phi) \geq \left(\frac{\beta}{\delta}\right)[\delta^2 g(0)\beta^T V(\theta_o)\beta] + Q_n(\widetilde{\phi_n}).$$

Using formula (4) and definition of $\triangle_n$,

$$G_n(\phi) \geq \alpha\delta g(0)\beta^T V(\theta_o)\beta - 2\triangle_n + G_n(\widetilde{\phi_n}).$$

We notice that the last expression is independent of $\phi$. Therefore

$$\inf_{\| \phi - \widetilde{\phi_n} \| > \delta} G_n(\phi) \geq [\alpha\delta g(0)\beta^T V(\theta_o)\beta - 2\triangle_n] + G_n(\widetilde{\phi_n}).$$

We already knew that $\triangle_n \to^p 0$, and $V(\theta_o)$ is positive definite, and $\alpha\delta g(0)$ is positive. Hence, for sufficiently large $n$, $2\triangle_n < \alpha\delta g(0)\beta^T V(\theta_o)\beta$. Therefore,

$$\inf_{\| \phi - \widetilde{\phi_n} \| > \delta} G_n(\phi) \geq G_n(\widetilde{\phi_n}).$$

This means that the minimizer $\widehat{\phi_n}$ of $G_n(\phi)$ cannot be any $\phi$ such that $\| \phi - \widetilde{\phi_n} \| > \delta$, for sufficiently large $n$. Hence given any $\delta > 0$, with probability tending to 1,

$\| \widehat{\Phi_n} - \widetilde{\Phi_n} \| \leq \delta$, for sufficiently large $n$. This is the desired result. The proof is finished. □

For the applications of the Theorem 2.1 and Theorem 3.1, we consider nonlinear regression model with random regressor.

**Example 3.1.** We consider $y_i = \dfrac{1}{\theta_1 + \theta_2 x_i} + \varepsilon_i$, where $\theta = (\theta_1, \theta_2) \in \Theta = [1, a_1] \times [1, a_2]$. $a_1, a_2 < \infty$. Assume that $\{\varepsilon_i\}$ are i.i.d. random variable with distribution function $G(x)$ for with continuous probability density function $g(x)$ such that $g(0) > 0$ and $G(0) = \dfrac{1}{2}$. And assume that input variable $\{x_i\}$ is random sample with uniform distribution $F(x)$ which has probability density function

$$k(x) = \begin{cases} 1 & \text{on } [0, 1] \\ 0 & \text{otherwise} \end{cases}.$$

Partial derivatives $\dfrac{\partial f}{\partial \theta_1} = \dfrac{1}{(\theta_1 + \theta_2 x_i)^2}$, $\dfrac{\partial f}{\partial \theta_2} = \dfrac{x_i}{(\theta_1 + \theta_2 x_i)^2}$ are continuous in $(x, \theta)$ and $\| \nabla f(x, \theta) \| \leq h(x)$ where $h(x) = \sqrt{\dfrac{2}{(1+x)^4}}$ is square integrable. Hence $V_n(\theta)$ converges to $V(\theta)$ where

$$V(\theta) = \begin{pmatrix} \int \dfrac{1}{(\theta_1 + \theta_2 x)^4} dF(x) & \int \dfrac{x}{(\theta_1 + \theta_2 x)^4} dF(x) \\ \int \dfrac{x}{(\theta_1 + \theta_2 x)^4} dF(x) & \int \dfrac{x^2}{(\theta_1 + \theta_2 x)^4} dF(x) \end{pmatrix}.$$

For a nonzero vector $c = (c_1\ c_2)$,

$$\begin{aligned} cV(\theta)c^T &= (c_1\ c_2) \begin{pmatrix} \int \dfrac{1}{(\theta_1 + \theta_2 x)^4} dF(x) & \int \dfrac{x}{(\theta_1 + \theta_2 x)^4} dF(x) \\ \int \dfrac{x}{(\theta_1 + \theta_2 x)^4} dF(x) & \int \dfrac{x^2}{(\theta_1 + \theta_2 x)^4} dF(x) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\ &= \int \left( \dfrac{c_1 + c_2 x}{(\theta_1 + \theta_2 x)^2} \right)^2 dF(x) > 0. \end{aligned}$$

Hence $V(\theta)$ is positive definite. Obviously, the regression function satisfies the assumptions of the previous theorem. Hence $L_1$-estimator $\widehat{\theta_n}$ converges to $\theta_o$ in probability at the rate of $\dfrac{1}{\sqrt{n}}$ and $\sqrt{n}(\widehat{\theta_n} - \theta_o)$ has asymptotically normal distribution. □

## References


[1] Bhattacharyya, B.B and Otsuka, Y. and Richardson, G.D. (1992), Strong consistency in nonlinear regression with multiplicative error, *Communications in Statistics-Theory Meth.* 21, 2825-2831.

[2] Jennrich, R. (1969), Asymptotic properties of nonlinear least squares estimations, *The Annals of Mathematical Statistics*. 40, 633-643.

[3] Kim, Hae kyung and Choi Seung hoe (1995), Asymptotic properties of nonlinear least absolute deviation estimators, *Journal of the Korean Statistical Society* 24. No.1, 127-139.

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

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

[6] Wang, F.T and Scott, D.W, (1994), The $L_1$-method for roust nonparametric regression, *Journal of The American Statistical Association* 89, 65-76.

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