〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 18, No. 2, 1989

The Strong Consistency of Nonlinear Least Squares Estimators[+]

Hae Kyung Kim*

ABSTRACT

This paper is concerned with the strong consistency of the least squares estimators for the nonlinear regression models. A simple and practical sufficient condition for the strong consistency of the least squares estimators is given. It is also discussed that the extension of the strong consistency to a wide class of regression functions can be established by imposing some condition on the input values. Some examples are given to illustrate the application of main result

1. Introduction

Consider the following nonlinear regression model for a univariate response Y,

$$Y_t = f(x_t, \theta) + \varepsilon_t, \quad t = 1, \cdots, n \tag{1.1}$$

where $x_t \in \Xi \subset R^1$ denotes the t-th fixed input value, θ is the parameter vector from a parameter space $\Theta \in R^p$, f is a continuous function $f: R^{m+1} \to R^1$ and ε_t are independent random variables such that $E\,\varepsilon_t = 0$ and $E\varepsilon_t^2 = \sigma_t^2$ with $0 < \sigma < \infty$. Suppose Ξ is a compact subset of R^1 and Θ is a compact convex subset of R^p.

The parameter θ is unknown and the regression problem is to make inference about θ in some optimal way, on the basis of observations on Y_t and x_t, t= $1, \cdots, n$. Considerable attention has been devoted in literature to the least squares estimation under the assum-

* Department of Mathematics, Yonsei University, Seoul 120-746, Korea
+ This research was supported by the Korean Ministry of Education (1989).

ption that the errors ε_t are independent and identically distributed (i.i.d). For a nonlinear model of the form (1.1), the least squares estimators (LSE's) $\hat{\theta}_n$ are the values of the parameter θ which minimize the sum of squares function

$$S_n(\theta)=\sum_{t=1}^{n}(Y_t - f(x_t,\theta))^2. \tag{1.2}$$

An estimator $\tilde{\theta}_n$ of θ_o, the true value of θ, is said to be strongly consistent if $\tilde{\theta}_n$ converges to θ_o almost surely (a.s.) as n$\rightarrow\infty$. Asymptotic results for nonlinear least squares estimation are given by various authors: Jennrich (1969), Malinvaud(1970) and Wu (1981) among them.

The existence and the measurability of LSE were proved by Jennrich (1969). In the same paper the strong consistency of LSE was proved under the assumption of i.i.d. errors ε_t and suitable conditions of $f(x,\theta)$. This results was later extended to the case where ε_t is generated by a stationary time series in Hannan (1971).

In Wu (1981), he has also established the strong consistency of LSE $\hat{\theta}_n$ under some-what different conditions from those employed in Jennrich (1969). As mentioned in Wu (1981), for the strong consistency of $\hat{\theta}_n$, Wu's assumptions impose a much weaker condition than Jennrich's on the growth rate of $\sum_{t=1}^{n}[f(x_t,\theta)-f(x_t,\theta_o)]^2$. In this point of view Wu's result may be applicable to a wider class of regression functions than Jennrich's. However, from a practical point of view, there are still two main difficulties: First, in many situations, these ideal conditions are violated, and the valued results are not applicable as will be shown in Section 4. Second, in applying the result the verification of the conditions may cause some difficulty. In practice, these conditions are quite tedious to verify in applications and few would bother to do so.

The main purpose of this paper is to provide simple sufficient conditions for the strong consistency of the LSE in the nonlinear regression model (1.1), and to extend this result to a large class of regression functions by imposing some condition on the inputs $\{x_t\}$. For these, in Section 2 we provide some assumptions and preliminary lemmas needed in the proof of the main theorem. In section 3 we give the main result which gives sufficient conditions for the strong consistency of the LSE. Some examples of the application of the main result are contained in Section 4.

2. Assumptions and Preliminaries

We start this section by introducting some conditions on regression function and inputs, which ensure the strong consistency of LSE.

The following notation is used : $D_u f(x,\theta)= \frac{\partial}{\partial\theta_u} f(x,\theta)$

$D_{uv}f(x,\theta)=D_v[D_u f(x,\theta)]$, $f'(x,\theta)= \frac{\partial}{\partial x} f(x,\theta)$ and $B_\delta= \{\theta\in\Theta: |\theta-\theta_0|\geq\delta\}$. Furthermore, throughout the paper we make the following assumptions: for the model (1.1),

Assumption 1. f' and the all firt order partial derivatives $D_u f$ and $D_u f'$ are continuous on $\Xi \times \Theta$.

Assumption 2. $f(x,\theta)$ is such that for any $\delta>0$, either

(a) $\sup_{\theta\in B_\delta} | f(x,\theta)-f(x,\theta_0)| >0$ on Ξ, or

(b) $\sup_{\theta\in B_\delta} | f'(x,\theta)-f'(x,\theta_0)|>0$ on τ

whenever $\tau= \{x \in \Xi: \sup_{\theta\in B_\delta} | f(x,\theta)-f(x,\theta_0)| = 0\}$ is a nonempty closed set.

Assumption 3. The sequence of inputs $\{x_t\}$ generates Cesàro summable sequences with respect to a probability measure μ defined on the Borel subsets of Ξ, i.e., for every real valued function g with $\int_\Xi | g(x)| d\mu(x)<\infty$,

$$\lim_{n=\infty} \frac{1}{n} \sum_{t=1}^{n} g(x_t)= \int_\Xi g(x)\, d\mu(x).$$

Remark 2.1. Assumption 3 is the regularity condition on the limiting behavior of inputs for the weak convergence of measures. One simple way of generating such sequence is to choose inputs as a random sample from some distribution function $H(x)$ defined on Ξ. In this case every realization generate Cesàro summable with respect to $H(x)$ by the Strong Law of Large Numbers.

Before we proceed to consider the main result, we present the following several preliminary lemmas, as they will be required subsequently.

Lemma 2.1. Let $g(x,\theta)$ be a real valued continuous function on X $\times$ Y of two Euclidean spaces. If B is a bounded subset of Y and the function h is defined by $h(x)= \sup_{\theta\in B} g(x,\theta)$, then $h(x)$ is continuous on X.

Proof. Follows easily from uniform continuity of g on B. ■

Lemma 2.2. Let the model (1.1) and Assumption 1 hold. If the condition (a) of Assumption 2 holds, then there exists a constant M independent of x such that

$$\sup_{\theta_1 \neq \theta_2 \in B_\delta} \frac{|f(x,\theta_1)-f(x,\theta_2)|}{|\theta_1-\theta_2|} \leq M \sup_{\theta \in B_\delta} |f(x,\theta)-f(x,\theta_o)| \tag{2.1}$$

for some $\delta>0$ and for all x.

Proof. Note that for a fixed $\delta>0$, $\sup_{\theta \in B_\delta} | f(x,\theta)-f(x,\theta_o) |$ is, by Lemma 2.1, continuous function on Ξ. Then, it follows from the condition (a) and the compactness of Ξ that there is an $\varepsilon = \varepsilon(\delta)>0$ such that

$$\sup_{\theta \in B_\delta} | f(x,\theta)-f(x,\theta_o) | \geq \varepsilon$$

uniformaly in x. Moreover, by virtue of the Mean Value Theorem on R^p, for any $\theta_1 \neq \theta_2$ in B_δ,

$$f(x,\theta_1)-f(x,\theta_2) = \sum_{u=1}^{P} D_u f(x,\xi)\ (\theta_1^u-\theta_2^u)$$

so that

$$\sup_{\theta_1 \neq \theta_2 \in B_\delta} \frac{|f(x,\theta_1)-f(x,\theta_2)|}{|\theta_1-\theta_2|} = \sup_{\theta_1 \neq \theta_2 \in B_\delta} \left| \sum_{u=1}^{P} D_u f(x,\xi) \frac{\theta_1^u-\theta_2^u}{|\theta_1-\theta_2|} \right|$$

$$\leq \sum_{u=1}^{P} | D_u f(x,\xi) | \tag{2.2}$$

for some ξ, which may depend on x, on the line segment joining θ_1 and θ_2, where θ_1^u is the uth component of θ_1, etc., Furthermore, due to the continuity of $D_u f$ on $\Xi \times \Theta$, there is a constant K, which is independent of u, such that $|D_u f(x,\theta)| \leq K$ uniformly in (x,θ), so that

$$\sum_{u=1}^{P} | D_u f(x,\xi) | \leq M \sup_{\theta \in B_\delta} | f(x,\theta)-f(x,\theta_o) | \tag{2.3}$$

with $M=\frac{pK}{\varepsilon}$ which is independent of x. Combining (2.2) and (2.3) we arrive at (2.1) and the proof is completed. ■

Lemma 2.3. Under the same assumptions of Lemma 2.2, if the condition (b), instead of (a) Assumption 2 holds, then the following two results hold:

(i) τ is a finite set.

(ii) there exists a constant M independent of x, satifying (2.1).

Proof. (i) Let x_o be a point of τ. It suffices to prove the result (i) for the interior point x_o of τ, the proof for boundary point is quite analogous. Since the function $|f'(x,\theta)-f'(x,\theta_o)|$ is continuous on $\Xi\times\Theta$, it follows from the condition (b) of Assumption 2, that there are an $\eta>0$ and a neighborhood $N_{(o)}\times B_{(o)}$ where $x_o\in N_{(o)}\subset\Xi$ and $B_{(o)}\subset B_\delta$, such that

$$|f'(x,\theta)-f'(x,\theta_o)|\geq\eta$$

for all $(x,\theta)\in N_{(o)}\times B_{(o)}$. Without loss of generality, we may suppose that $\{N_{(t)}\mid x_t\in\tau\}$ are disjoint. Let x be an arbitrary point of $N_{(o)}-\{x_o\}$ such that $(x,\theta)\in N_{(o)}\times B_{(o)}$, then we obtain easily a ξ_θ which may depend on θ and lies between x and x_o, such that

$$\begin{aligned} f(x,\theta)-f(x,\theta_o)&=f(x_o,\theta)-f(x_o,\theta_o)+(x-x_o)[f'(\xi_\theta,\theta)-f'(\xi_\theta,\theta_o)]\\ &=(x-x_o)[f'(\xi_\theta,\theta)-f'(\xi_\theta,\theta_o)]\end{aligned}$$

The last equality follows from the condition (b). Consequently, we have

$$\sup_{\theta\in B_{(o)}}|f(x,\theta)-f(x,\theta_o)|=|x-x_o|\sup_{\theta\in B_{(o)}}|f'(\xi_\theta,\theta)-f'(\xi_\theta,\theta_o)| \tag{2.5}$$

for all $x\in N_{(o)}-\{x_o\}$, so that from (2.4) and (2.5)

$$\sup_{\theta\in B_\delta}|f(x,\theta)-f(x,\theta_o)|\geq\sup_{\theta\in B_{(o)}}|f(x,\theta)-f(x,\theta_o)|>0$$

for all $x\in N_{(o)}-\{x_o\}$. This implies that x_o is an isolated point. Since x_o is arbitrary, τ consists only of the isolated points.

On the other hand, since τ is compact it follows from the Heine-Borel Theorem that the open covering $\{N_{(t)}\mid x_t\in\tau\}$ of τ has a finite subcovering. That is, there exist a finite number of neighborhoods $N_{(t_j)}$ $j=1,\cdots,m$, say, such that $\tau\subset\bigcup_{j=1}^{m}N_{(t_j)}$. Since $\tau\cap N_{(t_j)}=\{x_{(t_j)}\}$, $j=1,\cdots,m,\tau$ consists of only finite number of points, and part (i) is proved.

(ii) As we have shown in part (i), there exist a finite number of neighborhoods N_{t_j}, $j=1,\cdots,m$, such that $\tau\subset\bigcup_{j=1}^{m}N_{t_j}$. Let x be a point of Ξ. First we suppose $x\in\Xi-\bigcup_{j=1}^{m}N_{t_j}$. Then, clearly there exists, in view of Lemma 2.2, a constant M^*, independent of x, satisfying (2.1) for all x on $\Xi-\bigcup_{j=1}^{m}N_{t_j}$. Next, if $x\in\bigcup_{j=1}^{m}N_{t_j}$, then there is an element of $\{N_{t_j}\ j=1,\cdots,m\}$, say N_{t_k} which is a neighborhbood of $x_k\in\tau$, such that $x\in N_{t_k}$. Noting that for $\delta>0$, $\sup_{\theta\in B_\delta}|f(x_k,\theta)-f(x_k,\theta_o)|=0$ implies $f(x_k,\theta)-f(x_k,\theta_o)=0$ for any $\theta\neq\theta_o$ in B_δ

so that by virtue of the Triangle Inequality $|f(x_k,\theta_1)-f(x_k,\theta_2)| = 0$ for any $\theta_1, \theta_2 \in B_\delta$, $\theta_1 \neq \theta_2$, we obtain

$$|f(x,\theta_1)-f(x,\theta_2)| \leq |x-x_k| |f'(\xi,\theta_1)-f'(\xi,\theta_2)| \tag{2.6}$$

for all $x \in N_{x_k}$ and for any $\theta_1, \theta_2 \in B_\delta$, $\theta_1 \neq \theta_2$, where ξ is a point which depends on θ_1 and θ_2, on the line segment with end points x and x_k. Furthermore, the right-hand side of (2.6) is less than or equal to

$$|x-x_k| \, |\theta_1-\theta_2| \sum_{u=1}^{p} |D_u f'(\xi,\theta^*)|$$

for some θ^* on the line segment with end points θ_1 and θ_2. Owing to the boundedness of $D_u f'$, the last expression can be also made less than $|x-x_k| \, |\theta_1-\theta_2| \, pK$ with some constant K which does not depend on θ and x. Thus,

$$\sup_{\theta_1=\theta_2 \in B_\delta} \frac{|f(x,\theta_1)-f(x,\theta_2)|}{\theta_1-\theta_2} \leq |x-x_k| \, \mathrm{pK} \tag{2.7}$$

for all $x \in N_{x_k}$. On the other hand, by an argument similar to the one used in the proof of Lemma 2.2, we also obtain that for any $x \in N_{x_k}$

$$\begin{aligned} \sup_{\theta \in \hat{B}_\delta} |f(x,\theta)-f(x,\theta_o)| &= |x-x_k| \sup_{\theta \in B_\delta} |f'(x_\theta,\theta)-f'(x_\theta,\theta_o)| \\ &\geq |x-x_k| \, \varepsilon_\delta \end{aligned} \tag{2.8}$$

for some $\varepsilon_\delta > 0$ and some point x_θ on the line segment joining x and x_o. By combining (2.7) and (2.8), we obtain

$$\sup_{\theta \in B_\delta} \frac{|f(x,\theta_1)-f(x,\theta_2)|}{|\theta_1-\theta_2|} \leq M_k \sup_{\theta \in B_\delta} |f(x,\theta)-f(x,\theta_o)|$$

for all x on N_{x_k}, with $M_k = pK/\varepsilon_\delta$ which is independent of x. Let $M = \max\{M^*, M_1, \cdots, M_m\}$. Then the desired inequality (2.1) holds for all $x \in \Xi$ with M. ■

The following lemma, the proof of which is given in the appendix in Malinvaud (1970), shall be used in the proof of Theorem 3.1.

Lemma 2.4. Let $B \subset R^p$ be compact and let $g_t(\theta)$, $t=1,\cdots,n$, be Lipschitz functions on B satisfying

$$\sup_{\theta_1 \neq \theta_2 \in B} \frac{|g_t(\theta_1)-g_t(\theta_2)|}{|\theta_1-\theta_2|} \leq M \sup_{\theta \in B} |g_t(\theta)| < \infty \tag{2.9}$$

for some constant M independent of t. If $\sum_{t=1}^{n} \sup_{\theta \in B} |g_t(\theta)^2| \to \infty$ as $n \to \infty$, then for the independent random variables ε_t with zero mean and $\sup_t \operatorname{Var} \varepsilon_t < \infty$,

$$\lim_{n \to \infty} \frac{\sup_{\theta \in B} |\sum_{t=1}^{n} g_t(\theta)\ \varepsilon_t|}{\left[\sum_{t=1}^{n} \sup_{\theta \in B} |g_t(\theta)|\right]^{\frac{1+c}{2}}} = 0 \quad a.s.,$$

as $n \to \infty$, where c is any positive real.

3. Strong consistency

The main result of this paper is the following theorem, which provides sufficient conditions for the strong consistency of LSE. The proof proceed using the criterion for consistency utilized by Wu(1981).

Theorem 3.1. For the model (1.1), suppose that Assumptions 1, 2 and 3 are fulfilled. Then the LSE $\hat{\theta}_n$ defined in (1.2) is strongly consistent for θ_o.

Proof. To establish the strong consistency of $\hat{\theta}_n$, it suffices [Lemma 1 in Wu (1981)] to show that for any $\delta>0$

$$\liminf_{n \to \infty} \inf_{\theta \in B_\delta} [S_n(\theta)-S_n(\theta_o)] > 0 \ a.s.. \tag{3.1}$$

Note that $\inf_{\theta \in B_\delta} [S_n(\theta)-S_n(\theta_o)]$ is greater than or equal to

$$\inf_{\theta \in B_\delta} \sum_{t=1}^{n}[f(x_t,\theta)-f(x_t,\theta_o)]^2 \left\{ 1-2 \frac{\sup_{\theta \in B_\delta} \sum_{t=1}^{n}[f(x_t,\theta)-f(x_t,\theta_o)]\ \varepsilon_t}{\inf_{\theta \in B_\delta} \sum_{t=1}^{n}[f(x_t,\theta)-f(x_t,\theta_o)]^2} \right\}.$$

To prove (3.1) it is enough to show that for any $\delta>0$

$$\text{(a)} \quad \inf_{\theta \in B_\delta} \sum_{t=1}^{n'} [f(x_t,\theta)-f(x_t,\theta_o)]^2 \to \infty$$

and

$$\text{(b)} \quad \frac{\sup \sum_{t=1}^{n} [f(x_t,\theta)-f(x_t,\theta_o)]\, \varepsilon_t}{\inf \sum_{t=1}^{n} [f(x_t,\theta)-f(x_t,\theta_o)]^2} \to 0 \ ,$$

as $n \to \infty$.

But(a) follows from the fact [Theorem 1 in Gallant (1977)] that under Assumption 3

$$\frac{1}{n} \sum_{t=1}^{n} [f(x_t,\theta)-f(x_t,\theta_o)]^2 \to \int_{\Xi} [f(x,\theta)-f(x,\theta_o)]^2 d\mu(x) \qquad (3.2)$$

uniformly for all θ in Θ. To prove (b) it suffices to show that for any $\delta > 0$

$$\text{(b1)} \quad \frac{\sup_{\theta \in B_\delta} \sum_{t=1}^{n} [f(x_t,\theta)-f(x_t, \theta_o)]\, \varepsilon_t}{\sum_{t=1}^{n} \sup_{\theta \in B_\delta} [f(x_t,\theta)-f(x_t,\theta_o)]^2} \to 0 \ a.s.,$$

and

$$\text{(b2)} \quad \frac{\sum_{t=1}^{n} \sup_{\theta \in B_\delta} [f(x_t,\theta)-f(x_t,\theta_o)]^2}{\inf_{\theta \in B_\delta} \sum_{t=1}^{n} [f(x_t,\theta)-f(x_t,\theta_o)]^2} \quad \text{coverges } a.s.,$$

But (b1) follows from Lemma 2.4 when applied with $g_t(\theta) = f(x_t,\theta) - f(x_t,\theta_o)$ and $B = B_\delta$.

In this case, the Lipschitz condition for $f(x_t,\theta) - f(x_t,\theta_o)$ on B_δ is an immediate consequence of the boundedness of $D_u f$ in the following inequality:

$$|\, f(x_t,\theta) - f(x_t,\theta') \,| = |\, \sum_{u=1}^{P} D_u f(x_t,\bar{\theta})\ (\theta^u - \theta'^u) \,|$$

$$\leq \sum_{u=1}^{P} |\, D_u f(x_t,\bar{\theta}) \,|\ \ |\, \theta - \theta' \,|$$

for some $\bar{\theta}$ on the line segment from θ to θ'. The above equality follows from the Mean Value Theorem on R^P. To show (b2) we first note that $\sum_{t=1}^{n} [f(x_t,\theta) - f(x_t,\theta_o)]$ is continuous on the compact set B_δ. Thus there exists a point θ_n^* on B_δ, which depends on n, such that

$$\inf_{\theta \in B_\delta} \sum_{t=1}^{n} [f(x_t,\theta) - f(x_t,\theta_o)] = \sum_{t=1}^{n} [f(x_t,\theta_n^*) - f(x_t,\theta_o)]^2. \qquad (3.3)$$

Furthermore, owing to the boundedness the sequence $\{\theta_n^*\}$ has a convergent subsequence.

Let θ^{**} be any cluster point of $\{\theta_n^*\}$, and let $\{n_\nu\}$ be a subsequernce of $\{n\}$ such that $\{\theta_{n_\nu}^*\}$ converges to θ^{**}. Then, by the same argument as in (3.2)

$$\frac{1}{n}\sum_{t=1}^{n}[f(x_t,\theta_{n_\nu}^*)-f(x_t,\theta_o)]^2 \to \int_{\Xi}[f(x,\theta^{**})-f(x,\theta_o)]^2\,d\mu(x). \tag{3.4}$$

Now, we let θ^* be a point on B_δ such that

$$\int_{\Xi}[f(x,\theta^*)-f(x,\theta_o)]^2\,d\mu(x)= inf\int_{\Xi}[f(x,\theta^{**})-f(x,\theta_o)]^2\,d\mu(x), \tag{3.5}$$

the infimum being taken over the set of all cluster points of θ_n^*. On the other hand, from the continuity of $\sup_{\theta\in B_\delta}[f(x,\theta)-f(x,\theta_o)]^2$ on Ξ, there exists a constant L, independent of x, such that $\sup_{\theta\in B_\delta}[f(x,\theta)-f(x,\theta_o)]^2\leq$L. Thus we obtain

$$\sum_{t=1}^{n}\sup_{\theta\in B_\delta}[f(x_t,\theta)-f(x_t,\theta_o)]^2\leq nL. \tag{3.6}$$

It follows from (3.3) through (3.6) that

$$\frac{\sum_{t=1}^{n}\sup_{\theta\in B_\delta}[f(x_t,\theta)-f(x_t,\theta_o)]^2}{\inf_{\theta\in B_\delta}\sum_{t=1}^{n}[f(x_t,\theta)-f(x_t,\theta_o)]^2}\leq\frac{L}{\int_{\Xi}[f(x,\theta^*)-f(x,\theta_o)]^2d\mu(x)}<\infty,$$

and the theorem is proved. ■

Remark 3.1. In the linear model

$$Y_t= X_t'\theta+\varepsilon_t,\ t=1,\cdots,n, \tag{3.7}$$

where $X_t=(x_{t1},\cdots,x_{tp})'$ and $x_{t1}=1$, the LSE $\hat\theta_n$ of θ_o based on the design matrix X$=[X_1,\cdots,X_n]'$ and response vector Y$=(Y_t,\cdots,Y_n)'$, is given by

$$\hat\theta_n=(X'X)^{-1}X'Y$$

provided that X′X is nonsingular. It is known [Lai, Robbins and Wei (1978)] that the sufficient condition for the strong consistency of $\hat\theta_n$ is

$$(X'X)^{-1}\to O \tag{3.8}$$

as $n \to \infty$. Earlier, Drygas (1976) established the strong consistency of $\hat{\theta}_n$ under the alternative assumptions that there exist positive constants $k_n \to \infty$ and a positive definite matrix Σ such that

$$\frac{1}{k_n}(X'X) \to \Sigma \tag{3.9}$$

as $n \to \infty$. Although (3.9) reduces to (3.8) where $p=1$, it is much stronger than (3.8) when $p>1$.

In the simple linear model ($p=2$), our result is applicable. In this case, Assumptions 1 and 2 are straightforward and the condition (3.9) is implied by Assumption 3: With $k_n = n$

$$\frac{1}{n}(X'X) = \left[\frac{1}{n}\sum_{t=1} x_{tu}\, x_{tv}\right]_{u,v} \to \Sigma^* = \left[\int_{\Xi} x_u x_v \, d\mu(x)\right]_{u,v}$$

for $u, \nu = 1, 2$, where $x_1 = 1$ and $x_2 = x$ are independent variables in the linear model. In this case, the matrix Σ^* is positive definite since for any nonzero $\lambda = (\lambda_1, \lambda_2)'$ in R^2, $\lambda'\Sigma^*\lambda = \int_{\Xi} (\lambda_1 + \lambda_2 x)^2 \, d\mu(x) > 0$ unless $\lambda_1 + \lambda_2 x = 0$ a.s..

4. Some examples

For the applications of the main result, we now consider several regression functions. Throughout the example we assume that the input values $\{x_i\}$ which give rise to observation $Y_1, \cdots, Y_n$ are chosen as a realization of a random sample on $[0,T]$, $T<\infty$.

Example 1. Consider the multiple exponential model with the regression function

$$f(x,\theta) = \sum_{t=1}^{k} \alpha_t \, e^{-\beta_t x}.$$

Evidently, the regression function satisfies the assumptions of Theorem 3.1. That is, the derivative f' and all partial derivatives $D_u f$, $D_u f'$ exist and are continuous on $[0,T] \times \Theta$. Moreover, it is easy to show that $\sup_{\theta \in B\delta} |f(x,\theta) - f(x,\theta_0)| > 0$ on $[0, T]$ unless all $\alpha_1 = 1$. Thus, we can conclude that the LSE $\hat{\theta}_n$ is strongly consistent for θ_0 under the given sampling scheme on $[0,T]$.

Example 2. Consider the first order decay model $Y_t = e^{-\beta x_t} + \varepsilon_t$ where $0 < \beta \in \Theta$, Θ is a compact set in R^1, and the ε_t are i.i.d. random variables having expected value zero and positive standard deviation. In this case, $\tau = \{0\}$ is nonempty. However, it is easy to see that $\sup_{\theta \in B_\delta} | \beta_o e^{-\beta_o x} - \beta e^{-\beta x} | > 0$ at $x = 0$. That is, the condition (a) of Assumption 2 is not satisfied but the condition (b) is. The other conditions of Theorem 3.1 are fulfilled obviously. We can guarantee, therefore, the strong consistency of the LSE β under the sampling scheme.

For the first order growth model growth model $y_t = e^{\alpha t} + \varepsilon_t$, it is also easy to show that Theorem 3.1 ensures the strong consistency of the LSE $\hat{\alpha}_n$ under the given sampling.

Example 3. Consider the regression function as a complicated nonlinear model that occurs in the pressure data of nuclear explosion:

$$Y_t = \alpha_1 e^{-\beta_1 x_t} + \alpha_2 e^{-\beta_2 x_t} + \alpha_3 e^{-\beta_3 x_t} \cos\sqrt{ax+b} + \varepsilon_t, \quad t = 1, \cdots, n. \tag{4.1}$$

Assume that $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, a, b) \in \Theta$ and Θ is compact in $R^p (p = 9)$. Furthermore, we assume here that ε_t are i.i.d. errors with mean zero and finite variance. For the regression function $f(x, \theta)$ in the model (4.1), it is easy to check that f' and all $D_u f$, $D_u f'$ exist and are contiuous on $[0, T] \times \Theta$, where $ax+b > 0$ on $[0, T]$. Moreover $\sup_{\theta \in B_\delta} | f(x, \theta) - f(x, \theta_o) | > 0$ on $[0,T]$. Thus, under the sampling scheme, the LSE is strongly consistent by Theorem 3.1.

Remark 4.1. Although Assumption 3 is a mathematical defect, it is not quite a restriction from the practical viewpoint. Rather, the condition plays an important role in allowing the strong consistency of the least squares estimator to a wide class of regression functions, as we shall see in the following examples.

Example 4. Consider the power curve model [Wu (1981), Example 3] $Y_t = (t+\theta)^d + \varepsilon_t$, $t = 1, 2, \cdots$, where $d (\neq 0)$ is a known constant and θ ranges over a compact subset Θ in R^1. We first note that Wu demonstrated the inconsistency of the LSE $\hat{\theta}_n$ for when $d < 1/2$. However, for the regression function $f(x, \theta) = (x+\theta)^d$, the derivatives f', $D_u f$ and $D_u f'$ exist and are continuous on $[0,T] \times \Theta$ unless $x + \theta = 0$. Moreover, $\sup_{\theta \in B_\delta} | f(x, \theta) - f(x, \theta_o) | = \sup_{\theta \in B_\delta} | (x+\theta)^d - (x+\theta_o)^d | > 0$ on $[0,T]$. Therefore, if we impose the our sampling scheme on the sequence of inputs (in the sense of Assumption 3), Theorem 3.1 guarantees the strong consistency of the LSE $\hat{\theta}_n$.

Example 5. Let us consider again the model in [Wu (1981), Example 4] with the regression function $f(x,\theta) = \theta_1 x^{-\theta_2}$, $x \neq 0$ where $\theta = (\theta_1, \theta_2) \in \Theta = [0, a] \times [0, b]$ a, b $< \infty$. It is not hard to verify that Assumptions 1 and 2 are satisfied. Therefore, under the our sampling scheme on [0,T] the LSE $\hat{\theta}_n$ is strongly consistent. Now we note that Wu's result does not guarantee the strong consistency of the LSE $\hat{\theta}_n$ when $\theta_2 = 1/2$.

Remark 4.2. In each of the examples above, we need only assume that ε_t are independent with $E\ \varepsilon_t = 0$ and $\sup_{t \geq 1} E\ \varepsilon_t^2 < \infty$.

References

(1) Hannan, E.J. (1971) Nonlinear time series regression. *Journal of Applied Probability*, 8, 767-780.

(2) Drygas, H. (1976) Weak and strong consistency of the least square estimates in regression models. Z. *Wahrscheinlickeitstheorie verw. Gebiete*, 34, 119-127.

(3) Gallant, A.R. (1977) Three-stage least-squares estimation for a system of simultaneous, nonlinear, implicit equations. *Journal of Econometrics*, 5, 71-81.

(4) Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators, *Annals of Mathematical Statistics*, 40, No. 2, 633-643.

(5) Lai, T.L., Robbins, H. and Wei, C.Z. (1978) Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy Sciences, U.S.A.* 75, 3034-3036.

(6) Malinvaud, E. (1970) The consistency of nonlinear regression. *Annals Mathematical Statistics*, 41, 956-969.

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