The Annals of Statistics
1981, Vol. 9, No. 3, 501–513

ASYMPTOTIC THEORY OF NONLINEAR LEAST SQUARES ESTIMATION[1]

BY CHIEN-FU WU

University of Wisconsin, Madison

For a linear regression model, the necessary and sufficient condition for the asymptotic consistency of the least squares estimator is known. An analogous condition for the nonlinear model is considered in this paper. The condition is proved to be necessary for the existence of any weakly consistent estimator, including the least squares estimator. It is also sufficient for the strong consistency of the nonlinear least squares estimator if the parameter space is finite. For an arbitrary compact parameter space, its sufficiency for strong consistency is proved under additional conditiöns in a sense weaker than previously assumed. The proof involves a novel use of the strong law of large numbers in $C(S)$. Asymptotic normality is also established.

1. Introduction. The method of least squares plays a central role in the inference of parameters in nonlinear regression models. Due to nonlinearity, the resulting nonlinear least squares estimators defined in (1.2) do not enjoy any tractable finite sample optimality property (e.g., the minimum variance linear unbiased estimator, etc.) A general approach to the theoretical study of nonlinear least squares is thus asymptotic. However in the past more attention was paid to the numerical methods for calculating the estimator ([6] and its references). Much of the work was done by first assuming the consistency of nonlinear least squares estimator and then proving the asymptotic normality, constructing confidence regions, etc. ([2–4], [7], [10–12], [14].) The relatively harder question of consistency was first rigorously proved by Jennrich (1969) and Malinvaud (1970). Jennrich (1969) considered the following model:

$$y_i = f(x_i, \theta_0) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

when x_i is the ith "fixed" input vector which gives rise to observation y_i, θ_0 is the unknown $p \times 1$ vector parameter from a compact parameter space $\Theta \subset R^p$, $f_i(\theta) = f(x_i, \theta)$ are continuous functions in $\theta \in \Theta$ and ϵ_i are independent identically distributed errors with mean zero and unknown variance $\sigma^2 > 0$. Any vector $\hat{\theta}_n$ in Θ which minimizes the residual sum of squares

$$S_n(\theta) = \textstyle\sum_{i=1}^n (y_i - f(x_i, \theta))^2 \tag{1.2}$$

will be called a least squares estimate of θ_0 based on $\{y_i\}_1^n$. Its existence and measurability were proved in [19]. An estimator δ_n of θ_0 is said to be strongly (weakly) consistent if $\delta_n \to \theta_0$ a.s. (in prob.) as $n \to \infty$. The strong consistency of $\hat{\theta}_n$ was proved in [19] under the following assumption:

(1.3) $n^{-1}D_n(\theta, \theta')$ converges uniformly to a continuous function $D(\theta, \theta')$ and $D(\theta, \theta_0) = 0$ if and only if $\theta = \theta_0$,

where

Received November 1979; revised June 1980.

[1] This research was supported by the National Science Foundation Grant No. MCS-79-01846.

AMS 1980 *subject classifications.* Primary 62J02, 62F12.

Key words and phrases. Nonlinear least squares estimator, nonlinear model, weak and strong consistency, asymptotic normality, strong law of large numbers in $C(S)$.

$$D_n(\theta, \theta') = \sum_{i=1}^{n} (f(x_i, \theta) - f(x_i, \theta'))^2. \tag{1.4}$$

Under slightly stronger assumptions, the asymptotic normality was proved in the same paper. Jennrich's method of proof was later extended to more complicated models in [13], [24], [25]. Essentially they all assumed (1.3). The question of hypothesis testing was considered by Gallant in a series of papers [10–12], again, by assuming the consistency of $\hat{\theta}_n$.

To assess the generality of the assumption (1.3), take the more familiar linear regression model $f(x_i, \theta) = x_i^T\theta$, where x_i is a $p \times 1$ vector. Then (1.3) is equivalent to

$$\frac{1}{n} X'_n X_n \to \text{some positive definite matrix,}$$

where $X'_n = [x_1, \ldots, x_n]$. From the recent work on the strong consistency of linear least squares estimator [9], [20], it is known that

$$(X'_n X_n)^{-1} \to 0 \tag{1.5}$$

is equivalent to the strong and weak consistency of the linear LSE under assumption (1.1) on the ϵ_i. For the nonlinear model (1.1), the analogue of (1.5) is

$$D_n(\theta, \theta') \to \infty \quad \text{as} \quad n \to \infty \quad \text{for all} \quad \theta \neq \theta'. \tag{1.6}$$

It was the attempt to relax condition (1.3) that originally motivated the author to study the problem.

In Section 2 we prove that (1.6) is necessary for the existence of *any* weakly consistent estimator of $\theta \in \Theta$ under conditions much weaker than (1.1). Since no particular form of the estimator (e.g., the least squares estimator) is assumed, this is rather a necessary condition for the "asymptotic identifiability" (or distinguishability) of the statistical problem under study. The expeditious proof of Theorem 1 is made possible by two previous results due to Hoeffding and Wolfowitz (1958) and Shepp (1965).

In Theorem 2 of Section 3 we prove that (1.6) is necessary and sufficient for the strong consistency of the least squares estimator $\hat{\theta}_n$ when Θ is a finite set. This result can also be interpreted in the context of hypothesis testing as in [16]. For general compact Θ, sufficient conditions for strong consistency are given in Theorem 3. These conditions, when specialized to the linear regression model, require that

$$\limsup_{n\to\infty} \frac{(\text{max. eigenvalue of } X'_n X_n)^{(1+c)/2}}{\text{min. eigenvalue of } X'_n X_n} < \infty \qquad \text{for some} \quad c > 0,$$

which is much weaker than (1.3). Unlike (1.3), only some weak growth rate condition $D_n(\theta, \theta_0)$ is imposed in Theorem 3. When the errors are normally distributed, the least squares estimator and the maximum likelihood estimator (MLE) are equivalent. The existing results on the asymptotic consistency of the maximum likelihood estimator for the independent not identically distributed (i.n.i.d.) case do not seem to be applicable here. For example, condition C4 (or C′4) of Hoadley (1971), when specialized to model (1.1) with normal errors, implies that $D_n(\theta, \theta_0)$ diverges to infinity at rate n for all $\theta \neq \theta_0$, which is closely related to (1.3). The crucial step in the proof of Theorem 3 involves the almost sure convergence in supremum norm of a sequence of random functions. This technique may be useful for extending the existing results on the MLE for i.n.i.d. observations. A result due to Jim Kuelbs on the strong law of large numbers in $C(S)$ (space of continuous functions on S with supremum norm) is given in the Appendix. A general result about the strong consistency of variance estimation is given in Theorem 4.

Finally, in Section 4, we derive the asymptotic normality of $\hat{\theta}_n - \theta_0$ under much weaker growth rate conditions than those assumed in [19]. Several examples are given in various sections to illustrate the theorems.

2. A necessary condition for the existence of a weakly consistent estimator. In Section 1 we have remarked that $(X_n'X_n)^{-1} \to 0$ is both necessary and sufficient for the strong and weak consistency of the least squares estimator in the linear model case. The analogue of this condition for the nonlinear model (1.1) is $D_n(\theta, \theta') \to \infty$ as $n \to \infty$ for all $\theta \neq \theta'$. In this section we will prove the necessity of this condition for the existence of a weakly consistent estimator of θ for model (1.1). The assumptions in (1.1) on the compactness of Θ, the continuity of $f_i(\theta)$ and the moment conditions on ϵ_i are not needed for the validity of the following Theorem 1.

THEOREM 1. *Let $y_i = f_i(\theta) + \epsilon_i$, where $\theta \in \Theta$ and the parameter space is a subset of R^p, $f_i(\theta)$ are functions defined on Θ, ϵ_i are i.i.d. with the common distribution G which has a positive* (a.e.) *and absolutely continuous density g with finite Fisher information,* i.e., $\int_{-\infty}^{\infty} (g')^2/g < \infty$. *If there exists an estimator $\bar{\theta}_n$ $(y_1, y_2, \ldots, y_n)$ such that*

$$\bar{\theta}_n(y_1, \ldots, y_n) \to_{\mathscr{L}_\theta} \theta \quad \textit{in probability for all} \quad \theta \in \Theta, \tag{2.1}$$

then

$$D_n(\theta, \theta') = \textstyle\sum_{i=1}^n (f_i(\theta) - f_i(\theta'))^2 \to \infty \quad \textit{as} \quad n \to \infty \tag{2.2}$$

for all $\theta \neq \theta'$ in Θ.

PROOF. Denote $\mathbf{Y}_n = (y_1, \ldots, y_n)$ and $P_{\theta,n}$ = the probability measure of $\mathbf{Y}_n$ under θ. From a well-known result due to Hoeffding and Wolfowitz (1958) (for details, see Theorem 2.1.1 of [1]), a necessary condition for (2.1) is

$$\lim_{n\to\infty} d_n(\theta, \theta') = \lim_{n\to\infty} \int_{R^n} \left| \frac{dP_{\theta,n}}{d\mu_n}(\mathbf{Y}_n) - \frac{dP_{\theta',n}}{d\mu_n}(\mathbf{Y}_n) \right| d\mu_n(\mathbf{Y}_n) = 2, \tag{2.3}$$

for any two distinct points θ and θ' in Θ, where μ_n is any σ-finite measure on R^n such that $P_{\theta,n}$ and $P_{\theta',n}$ are absolutely continuous with respect to μ_n. Denote P_θ = the probability measure of the infinite sample $(y_i)_{i=1}^\infty$ under θ. Then (2.3) amounts to saying that P_θ and $P_{\theta'}$ are mutually singular (or disjoint) for any $\theta \neq \theta'$ in Θ. The sequence of random variables $(y_i)_{i=1}^\infty$ under θ is a translate of the sequence of random variables $(y_i)_{i=1}^\infty$ under θ'. Under the assumptions on ϵ_i, if the two probability measures of $(y_i)_{i=1}^\infty$ under θ and θ' are mutually singular, then the sum of the squares of the translates $(f_i(\theta) - f_i(\theta'))_{i=1}^\infty$ is infinite (see Theorem 1(ii) of Shepp, 1965), which is condition (2.2). □

Therefore, if there exist $\theta \neq \theta'$ in Θ such that $\lim_{n\to\infty} D_n(\theta, \theta') < \infty$, then it is impossible to find a weakly consistent estimator of $\theta \in \Theta$, no matter how this estimator is obtained. In this situation, it is the "incompetency" of the statistical problem rather than the "incompetency" of the estimation method that should bear the blame. When an estimator is found to be inconsistent, one should first check whether the associated statistical problem is *capable* of admitting a consistent estimator. If yes, find a new estimator. Otherwise, a reformulation of the problem or a new design of experiment may be needed.

EXAMPLE 1. Malinvaud (1970) considered the first order decay model $y_t = e^{-\alpha t} + \epsilon_t$, $t = 1, 2, \ldots, \alpha \in (0, 2\pi)$, and demonstrated the inconsistency of the nonlinear least squares estimator $\hat{\alpha}$. This can be easily explained by Theorem 1 since $\sum_{t=1}^\infty (e^{-\alpha t} - e^{-\beta t})^2 < \infty$ for any $\alpha \neq \beta \in (0, 2\pi)$. For linear regression model, the "directions" of inconsistency were defined and characterized in [29].

EXAMPLE 2. Another example of inconsistency is the one-compartment open model which has been used extensively in the study of pharmacokinetics [26] and chemical reaction [5],

$$y_t = f_t(\theta) + \epsilon_t = \theta_1 \frac{\theta_2}{\theta_2 - \theta_3} (e^{-\theta_3 t} - e^{-\theta_2 t}) + \epsilon_t, \qquad t = 1, 2, \ldots,$$

where $\theta = (\theta_1, \theta_2, \theta_3)$, $\theta_1, \theta_2, \theta_3 > 0$. Due to the exponential convergence to zero as $t \to \infty$, one can easily verify that $\sum_{t=1}^{\infty} (f_t(\theta) - f_t(\phi))^2 < \infty$ for all $\theta \neq \phi$ with $\theta_2, \theta_3, \phi_2, \phi_3 > 0$.

3. Sufficient conditions for strong consistency. In this section we give sufficient conditions for the strong consistency of $\hat{\theta}_n$. As in the previous sections, let $f_i(\theta) = f(x_i, \theta)$ and θ_0 be the unknown true parameter. We begin with a general lemma that provides a criterion for consistency. Denote any least squares estimator defined in (1.2) by $\hat{\theta}_n$.

LEMMA 1. *Suppose, for any* $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_0|\geq\delta} (S_n(\theta) - S_n(\theta_0)) > 0 \quad \text{a.s. (or in prob.)}. \tag{3.1}$$

Then, $\hat{\theta}_n \to \theta_0$ a.s. (*or in prob.*) *as* $n \to \infty$.

PROOF. Only the proof of strong consistency is given. If $\hat{\theta}_n \to \theta_0$ a.s. is not true, then there exists a $\delta > 0$ such that $P(w\colon \limsup_{n\to\infty} |\hat{\theta}_n(w) - \theta_0| \geq \delta) > 0$. From the definition of $\hat{\theta}_n$, this implies

$$P(\liminf_{n\to\infty} \inf_{|\theta-\theta_0|\geq\delta} (S_n(\theta) - S_n(\theta_0)) \leq 0) > 0,$$

contradicting (3.1). □

The same lemma is applicable to any estimation procedure which is based on the minimization or maximization of a certain function. The idea originated with Wald's proof of the strong consistency of maximum likelihood estimators.

The following Lemma 2 will be used later on. Since the condition on the denominator A_n is related to the σ_i^2 in (3.2) and is not restrictive at all, the lemma can be useful in its own right.

LEMMA 2. *Let* $\{X_i\}$ *be a sequence of independent random variables with* $EX_i = 0$ *and* $\operatorname{Var}(X_i) = \sigma_i^2$ *and*

$$A_n \to \infty, \limsup_{n\to\infty} \frac{(\sum_{1=1}^{n} \sigma_i^2)^{1/2+\delta}}{A_n} < \infty \qquad \text{for some} \quad \delta > 0. \tag{3.2}$$

Then,

$$\frac{1}{A_n} \sum_{i=1}^{n} X_i \to 0 \quad \text{a.s.}$$

PROOF. (i) Assume $\lim_{n\to\infty} \sum_1^n \sigma_i^2 = c < \infty$; from the completeness of the L^2 space and Theorem 5.3.4 [8], there exists a random variable X with $EX = 0$, $\operatorname{Var} X = c$ such that $\sum_{i=1}^{n} X_i \to X$ a.s., which implies (3.3) via (3.2).

(ii) Assume $\lim_{n\to\infty} \sum_{i=1}^{n} \sigma_i^2 = \infty$; from [8], page 126,

$$\frac{\sum_{i=1}^{n} X_i}{(\sum_{i=1}^{n} \sigma_i^2)^{1/2+\delta}} \to 0 \quad \text{a.s.},$$

which implies (3.3) via (3.2). □

When Θ is a finite set, the necessary condition for weak consistency proved in Section 2 turns out to be sufficient for strong consistency. Under slightly weaker assumption on ϵ_i

than those assumed in Theorem 1, it is also necessary for the strong consistency of the least squares estimator. This is formally stated as Theorem 2. To compare the two theorems, note that for finite Θ strong and weak consistency of $\hat{\theta}_n$ are equivalent and that the conclusion of Theorem 1 unlike that of Theorem 2 holds for any estimator.

THEOREM 2. *Suppose the parameter space* Θ *is finite and model* (1.1) *is true. Then,* $D_n(\theta, \theta_0) = \sum_{i=1}^n (f_i(\theta) - f_i(\theta_0))^2 \to \infty$ *for all* $\theta \neq \theta_0$ *implies* $\hat{\theta}_n \to \theta_0$ a.s. *Conversely, if the support of* ϵ_i *in* (1.1) *is neither bounded above nor bounded below,* $\hat{\theta}_n \to \theta_0$ a.s. *implies* $D_n(\theta, \theta_0) \to \infty$ *for all* $\theta \neq \theta_0$.

PROOF. *Sufficiency.* Since Θ is a finite set, we prove consistency by verifying (3.1) for each $\theta \neq \theta_0$. Write

$$S_n(\theta) - S_n(\theta_0) = -2 \sum_{i=1}^n d_i(\theta)\epsilon_i + \sum_{i=1}^n d_i^2(\theta), \tag{3.4}$$

where $d_i(\theta) = f(x_i, \theta) - f(x_i, \theta_0)$. From $D_n(\theta, \theta_0) = \sum_{i=1}^n d_i^2(\theta) \to \infty$ and Lemma 2, $S_n(\theta) - S_n(\theta_0) \to \infty$ a.s. for all $\theta \neq \theta_0$, thus establishing the consistency of $\hat{\theta}_n$.

Necessity. If there exists a θ_1 with $D_n(\theta_1, \theta_0) = \sum_{i=1}^n d_i^2(\theta_1) \to c < \infty$, from the completeness of the L^2-space and Theorem 5.3.4 [8], there exists a random variable X with $EX = 0$, $\operatorname{Var} X = c$ such that $\sum_{i=1}^n d_i(\theta_1)\epsilon_i \to X$ a.s. Therefore, $S_n(\theta_1) - S_n(\theta_0) \to -2X + c$ a.s. To finish the proof, it remains to show $P(-2X + c < 0) > 0$. Since c is not specified, it suffices to prove that the support of X is not bounded above. Since the convolution of one distribution with unbounded support and any other distribution has unbounded support, $\sum_{i=1}^n d_i(\theta)\epsilon_i$ has unbounded support for any n and thus establishes the result by letting $n \to \infty$. □

REMARK 1. Although the finiteness assumption on Θ is a mathematical defect, it is not quite a restriction from the practical viewpoint. In actual computation we can only search the minimum over a finite set, say, to the eighth decimal place. Theorem 2 suggests that $\{x_i\}_{i=1}^n$ should be chosen such that $D_n(\theta, \phi)$ is as large as possible for any $\theta \neq \phi$. Although the choice may not be optimal, it does guarantee that enough "information" is gathered to allow for the consistent estimation of the unknown parameter θ_0 as $n \to \infty$.

REMARK 2. If the support of ϵ_i is bounded, then $D_n(\theta, \theta_0) \to \infty$ as $n \to \infty$ may not be necessary for the strong consistency of $\hat{\theta}_n$ as the following example shows. Let $\Theta = \{\theta_0, \theta_1\}$, $f_i(\theta_0) = a_i$, $f_i(\theta_1) = a_i + \delta_i$, a_i, δ_i are known constants and ϵ_i are uniformly distributed over $[-1, 1]$. If there exists an i with $|\delta_i| > 2$, then θ_0 can be identified correctly from observing y_i. This example also applies to Theorem 1.

REMARK 3. For the sufficiency part of Theorem 2, we need only assume that ϵ_i are independent with $E\epsilon_i = 0$ and $\sup_{i\geq 1} E\epsilon_i^2 < \infty$.

For finite Θ, one can verify condition (3.1) for each $\theta \neq \theta_0$. This is why the proof of strong consistency in Theorem 4 is not difficult. When Θ is not discrete, it becomes necessary, in verifying (3.1), to compare the random function $S_n(\theta_0)$ with infinitely many other random functions $S_n(\phi)$ in the neighborhood of θ distinct from θ_0. It is thus important to establish results on the uniform convergence (with probability 1) of a sequence of random functions. Under assumption (1.3), Jennrich (1969) obtained one such result in his theorem 4. But when $D_n(\theta, \theta_0)$ diverges to infinity at a rate slower than n, his method of proof fails. This is why we have to resort to the probability theory of Banach space valued random variables. Two such results are given in the Appendix.

For compact Θ, the strong consistency of $\hat{\theta}_n$ will be established under the following assumptions.

ASSUMPTION A. (i) For any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{\{\sum_{i=1}^n \sup_{|\theta-\theta_0|\ge\delta} (f_i(\theta) - f_i(\theta_0))^2\}^{(1+c)/2}}{\inf_{|\theta-\theta_0|\ge\delta} \sum_{i=1}^n (f_i(\theta) - f_i(\theta_0))^2} < \infty \tag{3.5}$$

for some $c > 0$,

(ii) $f_i(\theta)$ are Lipschitz functions on Θ and

$$\Lambda(f_i) = \sup_{\theta_1\ne\theta_2} \frac{|f_i(\theta_1) - f_i(\theta_2)|}{|\theta_1 - \theta_2|} \le M \sup_{|\theta-\theta_0|\ge\bar\delta} |f_i(\theta) - f_i(\theta_0)| \tag{3.6}$$

for some $\bar\delta > 0$ and for all i, where M is independent of i and $|\theta_1 - \theta_2|$ is the Euclidean distance between θ_1 and θ_2.

ASSUMPTION A′. For any $\theta \ne \theta_0$, there exists an $r_\theta > 0$ such that

$$\text{(i)} \quad \limsup_{n\to\infty} \frac{\{\sum_{i=1}^n \sup_{|\phi-\theta|\le r_\theta} (f_i(\phi) - f_i(\theta_0))^2\}^{(1+c)/2}}{\inf_{|\phi-\theta|\le r_\theta} \sum_{i=1}^n (f_i(\phi) - f_i(\theta_0))^2} < \infty \tag{3.5$'$}$$

for some $c > 0$,

(ii) $f_i(\theta)$ are Lipschitz functions on $B(\theta, r_\theta) = \{\phi \in \Theta, |\phi - \theta| \le r_\theta\}$ and

$$\sup_{\phi\ne\phi';\phi,\phi'\in B(\theta,r_\theta)} \frac{|f_i(\phi) - f_i(\phi')|}{|\phi - \phi'|} \le M' \sup_{\phi\in B(\theta,r_\theta)} |f_i(\phi) - f_i(\theta_0)| \tag{3.6$'$}$$

for all i, where M' is independent of i but may depend on (θ, r_θ).

Assumption A′ is a local version of Assumption A and may be cumbersome to verify. But (3.5)′ is weaker than (3.5) (see Example 3). To compare our assumptions and Jennrich's assumption (1.3), consider the linear model $f_i(\theta) = x_i^T\theta$ as in Section 1. Assumptions A or A′ are reduced to

$$\limsup_{n\to\infty} \frac{(\text{max eigenvalue of } X_n'X_n)^{(1+c)/2}}{\text{min eigenvalue of } X_n'X_n} < \infty \qquad \text{for some} \quad c > 0,$$

while (1.3) implies the more stringent condition that

$$\lim_{n\to\infty} \left(\frac{1}{n} \text{ max and min eigenvalues of } X_n'X_n\right) \text{ exist and both are positive and finite.}$$

In general, Assumptions A or A′ do not impose any condition on the growth rate of $D_n(\theta, \theta_0)$ as $n \to \infty$. However, when $D_n(\theta, \theta_0)$ grows to infinity at rate n as was assumed in Jennrich (1969), our assumptions are not comparable to his. Denote any least squares estimator defined in (1.2) by $\hat\theta_n$.

THEOREM 3. *Under Assumptions* A *or* A′, $D_n(\theta, \theta_0) = \sum_{i=1}^n (f_i(\theta) - f_i(\theta_0))^2 \to \infty$ *for all* $\theta \ne \theta_0$ *implies* $\hat\theta_n \to \theta_0$ a.s.

PROOF. We first give a proof under Assumption A. Write

$$S_n(\theta) - S_n(\theta_0) = D_n(\theta, \theta_0)(1 - 2\textstyle\sum_{i=1}^n d_i(\theta)\epsilon_i/D_n(\theta, \theta_0)),$$

where $d_i(\theta) = f_i(\theta) - f_i(\theta_0)$. From $D_n(\theta, \theta_0) \to \infty$ and (3.5), $\inf_{|\theta-\theta_0|\ge\delta} D_n(\theta, \theta_0) \to \infty$ as $n \to \infty$ for any $\delta > 0$. Therefore from Lemma 1, it suffices for the strong consistency of $\hat\theta_n$ to prove

$$\sup_{|\theta-\theta_0|\ge\delta} |\textstyle\sum_{i=1}^n d_i(\theta)\epsilon_i| / \inf_{|\theta-\theta_0|\ge\delta} D_n(\theta, \theta_0) \to 0 \text{ a.s.} \tag{3.7}$$

for any small $\delta > 0$. In view of (3.5), (3.7) is implied by

(3.8) $$\sup_{|\theta-\theta_0|\geq\delta} |\sum_{i=1}^n d_i(\theta)\epsilon_i| / \{\sum_{i=1}^n \sup_{|\theta-\theta_0|\geq\delta} d_i^2(\theta)\}^{(1+c)/2} \to 0 \text{ a.s.}$$

for some $c > 0$. To apply Corollary A in the appendix, take the compact set $S = \{\theta \in \Theta, |\theta - \theta_0| \geq \delta\}$ and the Lipschitz function g_j to be d_j. By choosing δ smaller than the $\bar{\delta}$ in (3.6), the d_j functions satisfy condition (8) in the Appendix. Therefore Corollary A implies (3.8).

As for Assumption A′, the set $\{\theta \in \Theta, |\theta - \theta_0| \geq \delta\}$ is compact for any $\delta > 0$. From Lemma 1 and the finite covering property of compact set, it suffices to prove that, for any $\theta \neq \theta_0$, there exists an $r_\theta > 0$ such that

(3.9) $$\lim_{n\to\infty} \inf_{|\phi-\theta|\leq r_\theta} (S_n(\phi) - S_n(\theta_0)) = \infty \text{ a.s.}$$

By using the same argument outlined above, one can easily show that (3.9) holds under Assumption A′. This completes the proof. □

Remark 3 after Theorem 2 is also applicable to Theorem 3.

EXAMPLE 3. Consider the power curve model $y_t = (t + \theta)^d + \epsilon_t$ $t = 1, 2, \ldots$, where d is a known constant and the parameter space is a compact subset of R^1. Then $D_n(\theta, \theta_0) = \sum_{t=1}^n ((t + \theta)^d - (t + \theta_0)^d)^2 \to \infty$ as $n \to \infty$ for all $\theta \neq \theta_0$ iff $d \geq 1/2$. Therefore for $d < 1/2$ no weakly consistent estimator exists. But for $d \geq 1/2$, both Assumptions A and A′ are satisfied. Therefore the least squares estimator $\hat{\theta}_n$ is strongly consistent. Note that, except for the linear case $d = 1$, Jennrich's condition (1.3) is not applicable here.

EXAMPLE 4. Consider $y_t = \theta_1 t^{-\theta_2} + \epsilon_t$, $t = 1, 2, \ldots$, with parameters $\theta = (\theta_1, \theta_2)$ and the true parameter $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)})$ lies in the interior of the parameter space $\Theta = [0, a] \times [0, b]$, $a, b < \infty$. It can be verified that $D_n(\theta, \theta')$ is of the same order as $n^{-2\min(\theta_2,\theta_2')+1}$ when $\min(\theta_2, \theta_2') \neq 1/2$ and as $\log n$ when $\min(\theta_2, \theta_2') = 1/2$, where $\theta = (\theta_1, \theta_2)$, $\theta' = (\theta_1', \theta_2')$ and $\theta \neq \theta'$. Therefore, no weakly consistent estimator exists when $\theta_2^{(0)} > 1/2$. For $\theta_2^{(0)} < 1/2$, it is not hard to verify that Assumption A′ is satisfied but *not* Assumption A. Therefore, $\hat{\theta}_n$ is strongly consistent. The really interesting (or disappointing) case is $\theta_2^{(0)} = 1/2$, for which condition (3.5)′ is not satisfied. By choosing $\theta = (\theta_1, 1/2)$ with $\theta_1 \neq \theta_1^{(0)}$, the denominator of (3.5)′ is of the order $\log n$ while the numerator is of the order $n^{r\theta(1+c)}$. Theorem 3 does not guarantee the strong consistency of $\hat{\theta}_n$. But if $\theta_1^{(0)}$ is a known constant, then Assumption A′ is still satisfied and the strong consistency of $\hat{\theta}_n$ follows. Again, condition (1.3) is not satisfied except for $\min(\theta_2, \theta_2') = 0$.

The next theorem provides a very general condition under which $n^{-1} S_n(\tilde{\theta}_n)$ converges to σ^2 a.s. for any strongly consistent estimator $\tilde{\theta}_n$ of θ_0. In the linear case, the estimator of σ^2 is strongly consistent without any condition on the consistency of $\tilde{\theta}_n$ or on $\{f_i\}$. See Schmidt (1976).

THEOREM 4. *Suppose the sequence of functions* $\{f_i(\theta)\}_{i=1}^\infty$ *are equicontinuous in* θ. *Then, for any estimator* $\tilde{\theta}_n \to \theta_0$ a.s., $n^{-1} S_n(\tilde{\theta}_n) \to \sigma^2$ a.s.

A proof can be easily obtained by writing $n^{-1}(S_n(\tilde{\theta}_n) - S_n(\theta_0))$ as in (3.4) and is thus omitted.

4. Asymptotic normality. The asymptotic normality of $\hat{\theta}_n$ was proved by Jennrich (1969, page 639) under assumptions on the first and second order derivatives of $f_i(\theta)$ which are similar to condition (1.3). Similar results were also obtained by Malinvaud (1970b) for more complicated models. Under much weaker growth rate conditions than Assumption (c) of [19], we derive the asymptotic normality result in this section. The basic idea is again to use Corollary A. For each $\theta \in \Theta$, let

$$f_i'(\theta) = \left(\frac{\partial}{\partial\theta_j} f_i(\theta)\right)_j, \qquad f_i''(\theta) = \left[\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right]_{j,k}, \qquad j, k = 1, \ldots, p.$$

ASSUMPTION B.

(i) $f_i'(\theta)$ and $f_i''(\theta)$ exist for all θ near θ_0; the true parameter θ_0 is in the interior of Θ and there exist $\tau_n \uparrow \infty$ such that $1/\tau_n \sum_{i=1}^n f_i'(\theta_0) f_i'(\theta_0)^T$ converges to a positive definite matrix $\sum$ as $n \to \infty$;

(ii) $\max_{1\le i\le n} f_i'(\theta_0)^T (\sum_{i=1}^n f_i'(\theta_0) f_i'(\theta_0)^T)^{-1} f_i'(\theta_0) \to 0 \quad$ as $\quad n \to \infty$;

(iii) $\sum_{i=1}^n f_i'(\theta_1) f_i'(\theta_1)^T (\sum_{i=1}^n f_i'(\theta_0) f_i'(\theta_0)^T)^{-1}$ converges to the identity matrix uniformly as $n \to \infty$ and $|\theta_1 - \theta_0| \to 0$;

(iv) there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{\tau_n} \sum_{i=1}^n \sup_{|\theta-\theta_0|\le\delta} \left(\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right)^2 < \infty \tag{4.1}$$

for all j, k;

(v) if, for a pair (j, k),

$$\sum_{i=1}^\infty \sup_{|\theta-\theta_0|\ge\delta} \left(\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right)^2 = \infty,$$

then there exists an M independent of i such that

$$\sup_{s\ne t;\, s,t\in S} \left| \frac{\partial^2 f_i(s)}{\partial s_j\partial s_k} - \frac{\partial^2 f_i(t)}{\partial t_j\partial t_k} \right| \Big/ |s - t| \le M \sup_{\theta\in S} \left| \frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k} \right| \tag{4.2}$$

for all i, where $S = \{\theta \in \Theta, |\theta - \theta_0| \le \delta\}$ and δ is the same as in (4.1).

For τ_n satisfying $\lim_{n\to\infty} \tau_{n-1}/\tau_n = 1$, from Lemma 3 at the end of the section, the above assumption (ii) is implied by (i). Assumptions (iv) and (v) are needed for the application of Corollary A.

THEOREM 5. *Let $\hat\theta_n$ be a strongly consistent least squares estimator of θ_o under model* (1.1). *Under Assumption* B

$$\sqrt{\tau_n}(\hat\theta_n - \theta_0) \to_{\mathscr{L}_{\theta_0}} N(0, \sigma^2 \textstyle\sum^{-1}). \tag{4.3}$$

PROOF. Since $\hat\theta_n \to \theta_0$ a.s., with almost every w in the sample space $\hat\theta_n(w)$ eventually takes its values in a convex compact neighborhood of θ_0 which is interior to Θ. It is thus legitimate to expand $S_n'(\theta)$ in the neighborhood of $\hat\theta_n$. Note that the first two derivatives of $S_n(\theta)$ are

$$S_n'(\theta) = \sum_{i=1}^n (f_i(\theta) - f_i(\theta_0) - \epsilon_i) f_i'(\theta), \tag{4.4}$$

and

$$S_n''(\theta) = \sum_{i=1}^n f_i'(\theta) f_i'(\theta)^T - \sum_{i=1}^n f_i''(\theta)\epsilon_i + \sum_{i=1}^n (f_i(\theta) - f_i(\theta_0)) f_i''(\theta). \tag{4.5}$$

From the mean-value theorem, there exists a $\lambda_n \in [0, 1]$ such that

$$\begin{aligned} S_n'(\theta_0) &= -\sum_{i=1}^n f_i'(\theta_0)\epsilon_i \\ &= S_n'(\hat\theta_n) + S_n''(\theta_n^*)(\theta_0 - \hat\theta_n), \end{aligned} \tag{4.6}$$

where $\theta_n^* = (1 - \lambda_n)\hat{\theta}_n + \lambda_n\theta_0$ is measurable from Lemma 3 of [19]. Since $\hat{\theta}_n$ is in the interior of Θ eventually, $S_n'(\hat{\theta}_n) = 0$. Now, (4.6) can be rewritten as

$$\sum_{i=1}^n f_i'(\theta_0)\epsilon_i = Z_n\left(\sum_{i=1}^n f_i'(\theta_0)f_i'(\theta_0)^T\right)(\hat{\theta}_n - \theta_0),$$

where

$$\begin{aligned} Z_n = {} & \sum_{i=1}^n f_i'(\theta_n^*)f_i'(\theta_n^*)^T\left(\sum_{i=1}^n f_i'(\theta_0)f_i'(\theta_0)^T\right)^{-1} \\ & - \sum_{i=1}^n f_i''(\theta_n^*)\epsilon_i\left(\sum_{i=1}^n f_i'(\theta_0)f_i'(\theta_0)^T\right)^{-1} \\ & + \sum_{i=1}^n (f_i(\theta_n^*) - f_i(\theta_0))f_i''(\theta_n^*)\left(\sum_{i=1}^n f_i'(\theta_0)f_i'(\theta_0)^T\right)^{-1}. \end{aligned}$$

We want to show $Z_n \to I$ a.s. The first term of Z_n converges to I a.s. because of $\theta_n^* \to \theta_0$ a.s. and condition (iii) of Assumption B. The third term converges to zero a.s. because of $\theta_n^* \to \theta_0$ a.s., conditions (i), (iii), (iv) and Cauchy-Schwarz inequality. To prove that the second term of Z_n converges to 0 a.s., from condition (i) of Assumption B and $\theta_n^* \to \theta_0$ a.s., it suffices to prove that, with probability 1,

$$\frac{1}{\tau_n}\sum_{i=1}^n f_i''(\theta)\,\epsilon_i \to 0 \qquad \text{as} \qquad n \to \infty \tag{4.7}$$

uniformly on $S = \{\theta \in \Theta, |\theta - \theta_0| \le \delta\}$ defined in (4.2). The (j, k) entry of the random matrix in (4.7) is

$$\frac{1}{\tau_n}\sup_{\theta\in S}\left|\sum_{i=1}^n \frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\epsilon_i\right| \to 0 \text{ a.s.} \qquad \text{as} \quad n \to \infty. \tag{4.8}$$

(i) $$\sum_{i=1}^\infty \sup_{\theta\in S}\left|\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right|^2 < \infty:$$

from a theorem of Itô and Nisio (1968), there exists a $C(S)$ valued random variable η such that

$$\sup_{\theta\in S}\left|\sum_{i=1}^n \frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\epsilon_i - \eta(\theta)\right| \to 0 \text{ a.s.} \qquad \text{as} \quad n \to \infty.$$

Since $\tau_n \to \infty$, this implies (4.8).

(ii) $$\sum_{i=1}^\infty \sup_{\theta\in S}\left|\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right|^2 = \infty:$$

from condition (iv) of Assumption B, (4.8) is implied by the following

$$\sup_{\theta\in S}\left|\sum_{i=1}^n \frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\epsilon_i\right| \Big/ \sum_{i=1}^n \sup_{\theta\in S}\left(\frac{\partial^2 f_i(\theta)}{\partial\theta_j\partial\theta_k}\right)^2 \to 0 \text{ a.s.} \qquad \text{as} \quad n \to \infty \tag{4.9}$$

which is guaranteed by Corollary A because of condition (v) of Assumption B.

To finish the proof, it remains to show that

$$\left(\sum_{i=1}^n f_i'(\theta_0)f_i'(\theta_0)^T\right)^{-1/2}\left(\sum_{i=1}^n f_i'(\theta_0)\epsilon_i\right) \to_{\mathcal{L}_{\theta_0}} N(0, \sigma^2 I),$$

which follows from Proposition 2.2 of Huber (1973) under condition (ii) of Assumption B. □

The strongly consistent estimator of Example 3 is also asymptotically normal since Assumption B is satisfied. However, condition (i) of Assumption B is not satisfied by Example 4. This again demonstrates the difficulty of the asymptotic theory when $D_n(\theta, \theta_0)$ goes to infinity at a rate different from n.

To conclude this section, we give

LEMMA 3. *Let* $(x_i)_{i=1}^n$ *be* n $p \times 1$ *vectors such that there exist* $\tau_n \uparrow \infty$ *and* $\lim_{n\to\infty} \tau_{n-1}/\tau_n = 1$ *with* $\tau_n^{-1} \sum_{i=1}^n x_i x_i^T$ *converging to a positive definite matrix* Σ.
Then

$$\max_{1\le i\le n} x_i^T \left(\textstyle\sum_{i=1}^n x_i x_i^T\right)^{-1} x_i \to 0 \quad \text{as} \quad n \to \infty. \tag{4.10}$$

PROOF. Since $\tau_n^{-1} \sum_{i=1}^n x_i x_i^T \to \Sigma$, (4.10) is implied by $\max_{1\le i\le n} (1/\tau_n) x_i^T \Sigma^{-1} x_i \to 0$ which is in turn implied by

$$\max_{1\le i\le n} \frac{x_i^T x_i}{\tau_n} \to 0 \quad \text{as} \quad n \to \infty. \tag{4.11}$$

From $\operatorname{tr}(\tau_n^{-1} \sum_{i=1}^n x_i x_i^T) = \tau_n^{-1} \sum_{i=1}^n x_i^T x_i \to C = \operatorname{tr} \Sigma > 0$, for any $\epsilon > 0$ there exists an N such that

$$\left| \frac{\tau_{n-1}}{\tau_n} - 1 \right| \le \epsilon, \quad \left| \frac{1}{\tau_n} \textstyle\sum_{i=1}^n x_i^T x_i - C \right| \le \epsilon \qquad \text{for all} \quad n \ge N.$$

Now from

$$\begin{aligned} \frac{1}{\tau_n} \max_{N\le i\le n} x_i^T x_i &\le \frac{1}{\tau_n} \max_{N\le i\le n} \{(C+\epsilon)\tau_i - (C-\epsilon)\tau_{i-1}\} \\ &\le \max_{N\le i\le n} \left\{ C \frac{\tau_i}{\tau_n} \left(1 - \frac{\tau_{i-1}}{\tau_i}\right) + 2\epsilon \frac{\tau_i}{\tau_n} \right\} \\ &\le (C+2)\epsilon \end{aligned}$$

(4.11) follows easily since $\tau_n \uparrow \infty$. □

For the linear model considered in Section 1, Huber (1973) has proved that the asymptotic normality of the least squares estimator is equivalent to the Lindberg type condition $\max_{1\le i\le n} x_i^T(\sum_{i=1}^n x_i x_i^T)^{-1} x_i \to 0$. Such a condition is automatically satisfied under the assumptions on $(x_i)_{i=1}^\infty$ in Lemma 3.

APPENDIX

A strong law of large numbers in $C(S)$. As shown in the main text, the most difficult step in the proof of consistency and normality involves the uniform convergence of a sequence of $C(S)$-valued random variables where $C(S)$ is the space of continuous functions on a compact metric space S with the supremum norm. Since the desired results can not be found in the literature, we list them in the following Lemma A and Corollary A. Lemma A and its proof were provided by Professor Jim Kuelbs in response to some discussions regarding the material of the paper.

Let (S, d) be a compact metric space and $C(S)$ the Banach space of real-valued continuous functions on S with the supremum norm

$$\|x\|_\infty = \sup_{s\in S} |x(s)|.$$

For a d-continuous metric ρ on S let $N(S, \rho, \epsilon)$ denote the minimal number of ρ-balls of radius less than or equal to ϵ which cover S, and set

$$H(S, \rho, \epsilon) = \log N(S, \rho, \epsilon).$$

We let

$$\mathrm{Lip}(\rho) = \{x \in C(S) \colon \Lambda(x) = \sup_{s \neq t} \frac{|x(s) - x(t)|}{\rho(s, t)} < \infty\},$$

and for $x \in \mathrm{Lip}(\rho)$ we define

$$\|x\|_\rho = \Lambda(x) + |x(a)|$$

where a is some fixed point in S.

LEMMA A. *Let (S, d) denote a compact metric space and suppose ρ is a d-continuous metric on S with*

$$\int_0^\delta H^{1/2}(S, \rho, u)\, du < \infty \qquad \text{for some} \quad \delta > 0. \tag{1}$$

Let $f\colon [0, \infty) \to [0, \infty)$ be increasing with

$$\int_c^\infty f^{-2}(u)\, du < \infty \qquad \text{for some} \quad c > 0 \tag{2}$$

and for $\{g_j \colon j \ge 1\} \subseteq C(S)$ set

$$D_n = \sum_{j=1}^n \|g_j\|_\infty^2. \tag{3}$$

Then, for $\{g_j\} \subseteq \mathrm{Lip}(\rho)$ with

$$\sum_{j=1}^\infty \frac{\Lambda^2(g_j)}{f^2(D_j)} < \infty, \tag{4}$$

and $\{\epsilon_j \colon j \ge 1\}$ independent with $E\epsilon_j = 0$ and $\sup_j E(\epsilon_j^2) < \infty$, we have

$$\lim_{n\to\infty} \Big\|\sum_{j=1}^n g_j\epsilon_j / f(D_n)\Big\|_\infty = 0 \tag{5}$$

whenever $D_n \to \infty$.

PROOF. Since (1) holds the identity mapping v from $\mathrm{Lip}(\rho)$ into $C(S)$ is a type 2 mapping (see Zinn (1977), proof of Corollary 1). Hence we have $A < \infty$ such that for all n

$$E\|X_1 + \cdots + X_n\|_\infty^2 \le A \sum_{j=1}^n E\|X_j\|_\rho^2, \tag{6}$$

whenever $X_1, \ldots, X_n$ are independent $\mathrm{Lip}(\rho)$ valued random variables with mean zero.

To prove (5) note that by Kronecker's lemma we need only prove

$$\sum_{j\ge1} \epsilon_j g_j / f(D_j) \tag{7}$$

converges with probability one in $C(S)$. Now by Itô-Nisio (1968), (7) converges with probability one since it converges in mean square, i.e., by (6)

$$\begin{aligned} E\Big\|\sum_{j=m}^n \epsilon_j g_j / f(D_j)\Big\|_\infty^2 &\le A \sum_{j=m}^n E\|\epsilon_j g_j\|_\rho^2 / f^2(D_j) \\ &\le A \sup_{j\ge1} E(\epsilon_j^2) \sum_{j=m}^n (\Lambda(g_j) + |g_j(a)|)^2 / f^2(D_j) \\ &\le 2A \sup_{j\ge1} E(\epsilon_j^2) \left[\sum_{j=m}^n \|g_j\|_\infty^2 / f^2(D_j) + \sum_{j=m}^n \frac{\Lambda^2(g_j)}{f^2(D_j)}\right] \\ &= 2A \sup_{j\ge1} E(\epsilon_j^2) \left[\sum_{j=m}^n \frac{D_j - D_{j-1}}{f^2(D_j)} + \sum_{j=m}^n \frac{\Lambda^2(g_j)}{f^2(D_j)}\right] \\ &\le 2A \sup_{j\ge1} E(\epsilon_j^2) \left[\int_{D_{m-1}}^{D_n} \frac{1}{f^2(t)}\, dt + \sum_{j=m}^n \Lambda^2(g_j)/f^2(D_j)\right] \end{aligned}$$

which converges to zero as $n, m \to \infty$ because of (2), (4). Thus the proof is complete. □

A special case of Lemma A has been repeatedly used in the main text. Let S be a compact subset of R^p and $\rho(s, t)$ be the Euclidean metric in R^p. Then (1) is satisfied. Condition (2) is satisfied for $f(s) = x^{(1+c)/2}$. Let $\{g_j\}$ be Lipschitz functions on S satisfying: there exists a constant M independent of j such that

$$\text{(8)} \qquad \sup_{s \neq t} \frac{|g_j(s) - g_j(t)|}{|s - t|} \leq M \sup_s |g_j(s)| < \infty,$$

where $|s - t|$ is the Euclidean distance between s and t. Then (4) is satisfied and the conclusion of Lemma A holds. This is stated as

COROLLARY A. *Let S be a compact subset of R^p and $g_j(x)$ Lipschitz functions satisfying* (8). *Then for the independent random variables $\{\epsilon_j\}$ with $E\,\epsilon_j = 0$ and $\sup_j E(\epsilon_j^2) < \infty$, we have, for any $c > 0$,*

$$\lim_{n \to \infty} \sup_{s \in S} \left| \sum_{j=1}^n g_j(s)\epsilon_j \right| / D_n^{(1+c)/2} = 0 \text{ a.s.}$$

where $D_n = \sum_{j=1}^n \sup_{s \in S} |g_j(s)|^2 \to \infty$ as $n \to \infty$.

Acknowledgments. The author is very grateful to Jim Kuelbs for generously providing Lemma A and its proof. He also wishes to thank Peter Bickel, Lucien LeCam, Jim Reeds and Larry Shepp for helpful discussions.

REFERENCES

[1] AKAHIRA, M. and TAKEUCHI, K. (1979). The concepts of asymptotic efficiency in statistical estimation theory. Unpublished lecture notes.

[2] BARNETT, W. A. (1976). Maximum likelihood and iterated Aitken estimation of nonlinear systems of equations. J. Amer. Statist. Assoc. **71** 354–60.

[3] BEALE, E. M. L. (1960). Confidence regions in non-linear estimation. *J. Roy. Statist. Soc. Ser. B.* **22** 41–76.

[4] BLISS, C. I. and JAMES, A. T. (1966). Fitting the rectangular hyperbola. *Biometrics* **22** 573–602.

[5] BOX, G. E. P. and LUCAS, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* **46** 77–90.

[6] CHAMBERS, J. M. (1977). *Computational Methods for Data Analysis.* Wiley, New York.

[7] CHAMBERS, J. M. and ERTEL, J. E. (1975) Distribution of nonlinear estimates. In *Proc. Social Statist. Section, Amer. Statist. Assoc.* 51–56.

[8] CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic Press, New York.

[9] DRYGAS, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **34** 119–27.

[10] GALLANT, A. R. (1975a). The power of the likelihood ratio test of location in nonlinear regression models. *J. Amer. Statist. Assoc.* **70** 198–203.

[11] GALLANT, A. R. (1975b). Testing a subset of the parameters of a nonlinear regression model. *J. Amer. Statist. Assoc.* **70** 927–32.

[12] GALLANT, A. R. (1975c). Seemingly unrelated nonlinear regressions. *J. Econometrics* **3** 35–50.

[13] HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probability* **8** 767–80.

[14] HARTLEY, H. O. and BOOKER, A. (1965). Non-linear least squares estimation. *Ann. Math. Statist.* **36** 638–50.

[15] HOADLEY, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.* **42** 1977–91.

[16] HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions. *Ann Math. Statist.* **29** 700–18.

[17] HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1,** 799–821.

[18] ITÔ, K. and NISIO, M. (1968). On the convergence of sums of independent Banach space valued random variables. *Osaka Math. J.* **5** 35–48.

[19] JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–43.

[20] LAI, T. L., ROBBINS, H. and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. U.S.A.* **75** 3034–6.

[21] MALINVAUD, E. (1970a). The consistency of nonlinear regressions. *Ann. Math. Statist.* **41** 956-69.
[22] MALINVAUD, E. (1970b). *Statistical Methods of Econometrics. Rev. ed.* North Holland, Amsterdam.
[23] NEVEU, J (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.
[24] PHILIPS, P. C. B. (1976). The iterated minimum distance estimator and the quasi-maximum likelihood estimator. *Econometrica* **44** 449-59.
[25] ROBINSON, P. M. (1972). Nonlinear regression for multiple time series. *J. Appl. Probability* **9** 758-68.
[26] RODDA, B. E., SAMPSON, C. B. and SMITH, D. W. (1975). The one compartment open model: some statistical aspects of parameter estimation. *Appl. Statist.* **24** 309-18.
[27] SCHMIDT, W. H. (1976). Strong consistency of variance estimation and asymptotic theory for tests of the linear hypotheses in multivariate linear models. *Math. Operationsforsch. Statist.* **7** 701-5.
[28] SHEPP, L. A. (1965). Distinguishing a sequence of random variables from a translate of itself. *Ann. Math. Statist.* **36** 1107-12.
[29] WU, C. F. (1980). Characterizing the consistent directions of least squares estimates. *Ann. Statist.* **8** 789-801.
[30] ZINN, J. (1977). A note on the central limit theorem in Banach spaces. *Ann. Probability* **5** 283-6.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
1210 W. DAYTON ST.
MADISON, WISCONSIN 53706