The Annals of Statistics
1986, Vol. 14, No. 4, 1591–1596

CONSISTENT ESTIMATORS IN NONLINEAR REGRESSION FOR A NONCOMPACT PARAMETER SPACE

BY G. D. RICHARDSON[1] AND B. B. BHATTACHARYYA

University of Central Florida and North Carolina State University

Sufficient conditions are given in order to ensure the existence of a sequence of strongly consistent estimators of unknown parameters in a nonlinear regression model. The primary difference between this and earlier work is in the generality of the parameter space. Indeed, the parameter space is assumed to be any separable, completely regular topological space; in particular, this includes all separable metric spaces.

1. Introduction. Consider the nonlinear regression model of the form $y_i = f_i(\theta_0) + \varepsilon_i$, where each f_i is a known *bounded*, continuous, real-valued function defined on the parameter space. The errors are assumed to be independent and identically distributed random variables each having mean zero and finite variance σ^2. The main results concerning this model are given in Section 3. Section 4 is devoted to the study of a special case.

Jennrich [3] and Malinvaud [4] were the first to give consistency proofs for estimating θ_0 in a nonlinear regression model when the parameter space is assumed to be a compact subspace of R^p. Compactness of the parameter space is needed to ensure the existence of least squares estimators. Our purpose is to modify the least squares procedure in order to establish the existence of a tractable sequence of strongly consistent estimators of θ_0 when the parameter space is not necessarily compact.

The parameter space S is assumed to be any separable, completely regular topological space. It is assumed that all topological spaces are Hausdorff. The method used in developing a strong consistency result is to embed S as a dense subspace of a compact topological space T such that each f_i has a continuous extension to T. Since a continuous real-valued function defined on a compact topological space is bounded, then necessarily each f_i must be bounded on S. However, the sequence $\{f_i\}$ is not required to be uniformly bounded on S, contrary to the case when the model is assumed to be of the form $y_i = f(x_i, \theta_0) + \varepsilon_i$, where $f\colon X \times S \to R$ is continuous and X and S are compact spaces.

The growth rate of $\sum_1^n (f_i(\alpha) - f_i(\beta))^2$ to $+\infty$ is assumed to be order n. This assumption is also made by Jennrich [3] and Malinvaud [4]. Wu [6] gives a strong consistency result when this is not necessarily the case. He replaces the assumption on the order of the growth rate by the requirement that the sequence $\{f_i\}$ satisfy a certain type of Lipschitz condition ([6], page 506).

Received November 1983; revised March 1986.

[1] Research supported by Division of Sponsored Research of University of Central Florida.

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

Key words and phrases. Nonlinear least squares estimator, noncompact space, strong consistency, Stone–Cech compactification.

However, Wu's assumptions are not comparable to those made here when the growth rate is of order n.

2. Preliminaries. It is assumed throughout this section that the model and assumptions made in the introduction are satisfied. Since the parameter space S is separable, there exists an increasing sequence $\{S_n\}$ of compact subsets of S such that $\cup_{n=1}^{\infty} S_n$ is dense in S. Indeed, each S_n could be chosen to be a finite subset of S. It is assumed throughout this work that $\{S_n\}$ is any conveniently chosen increasing sequence of compact subsets of S such that $\cup_{n=1}^{\infty} S_n$ is dense in S.

Given an underlying probability space (Ω, A, P), let $\hat{\theta}_n$: $\Omega \to S$ with $\hat{\theta}_n(\omega) = \theta_n \in S_n$ be such that $\Sigma_1^n(y_i(\omega) - f_i(\theta))^2$ assumes a minimum at $\theta = \hat{\theta}_n(\omega)$ as θ varies over S_n, $\omega \in \Omega$. This is meaningful since each S_n is compact and each f_i is a continuous function. Note that this reduces to the least squares procedure when S is compact and each S_n is chosen to be S.

Since S is a completely regular topological space, it has a Stone–Cech compactification βS. That is, S is a dense subspace of a compact space βS having the property that each bounded continuous function f: $S \to R$ has a continuous extension g: $\beta S \to R$. A proof of these results can be found in Wilansky ([5], page 147).

Following Jennrich [3], the tail cross product of $\{f_i\}$ with itself is defined when the sequence $\{1/n\Sigma_1^n f_i(\alpha) f_i(\beta)\}$ converges uniformly in $(\alpha, \beta) \in S \times S$. Note that if the above holds, then the sequence $\{1/n\Sigma_1^n(f_i(\alpha) - f_i(\beta))^2\}$ also converges uniformly on $S \times S$. Let us denote the limit function of the latter sequence by $\phi(\alpha, \beta)$. The function ϕ: $S \times S \to R$ is continuous.

It is assumed throughout this work that g_i: $\beta S \to R$ denotes the continuous extension of f_i to βS. Suppose that the tail cross product of $\{f_i\}$ with itself exists. Since $\{g_i\}$ is a continuous extension of $\{f_i\}$ and the topology of uniform convergence on $\beta S \times \beta S$ is complete, it is easy to verify that the tail cross product of $\{g_i\}$ with itself also exists. Hence, the sequence $\{1/n\Sigma_1^n(g_i(\alpha) - g_i(\beta))^2\}$ also converges uniformly to $\psi(\alpha, \beta)$, $\alpha, \beta \in \beta S$ with ψ the continuous extension of ϕ to $\beta S \times \beta S$.

The proof given by Jennrich for Theorem 4 [3] may now be duplicated to verify the next result.

LEMMA 2.1 [3]. *Assume that the tail cross product of* $\{f_i\}$ *with itself exists. Then* $1/n\Sigma_1^n g_i(\alpha)\varepsilon_i \to 0$ *uniformly in* $\alpha \in \beta S$, *almost surely.*

3. Strong consistency. Let us refer to the nonlinear regression model $y_i = f_i(\theta_0) + \varepsilon_i$ as *model 1*. This section is devoted to extending the following result due to Jennrich [3] to the noncompact parameter space case.

THEOREM 3.1 (Jennrich [3]). *Suppose that model 1 and the assumptions made in the introduction are satisfied. Moreover, assume* (1) *the tail cross product of* $\{f_i\}$ *with itself exists*, (2) $\phi(\theta_0, \theta) = 0$ *iff* $\theta = \theta_0$, *and* (3) S *is a compact subspace of* R^p. *Then* $\hat{\theta}_n \to \theta_0$ *almost surely.*

The sequence $\{\hat{\theta}_n\}$ in Theorem 3.1 denotes the least squares estimators of θ_0. The next theorem is an extension of Theorem 3.1.

THEOREM 3.2. *Suppose that model 1 and the assumptions made in the introduction are satisfied. Moreover, assume* (1) *the tail cross product of* $\{f_i\}$ *with itself exists and* (2) *if* $\alpha \in \beta S$, *then* $\psi(\theta_0, \alpha) = 0$ *iff* $\alpha = \theta_0$. *Then* $\hat{\theta}_n \to \theta_0$ *almost surely.*

PROOF. Let $Q_n(\theta) = 1/n\sum_1^n(y_i - f_i(\theta))^2$, $\theta \in S$, and let $G_n(\alpha) = 1/n\sum_1^n(y_i - g_i(\alpha))^2$, $\alpha \in \beta S$. Then

$$\begin{aligned} G_n(\alpha) &= \frac{1}{n}\sum_1^n\left(g_i(\theta_0) + \varepsilon_i - g_i(\alpha)\right)^2 \\ &= \frac{1}{n}\sum_1^n\left(g_i(\theta_0) - g_i(\alpha)\right)^2 + \frac{2}{n}\sum_1^n\left(g_i(\theta_0) - g_i(\alpha)\right)\varepsilon_i + \frac{1}{n}\sum_1^n\varepsilon_i^2. \end{aligned}$$

Lemma 2.1 implies that $2/n\sum_1^n(g_i(\theta_0) - g_i(\alpha))\varepsilon_i \to 0$ uniformly in $\alpha \in \beta S$, almost surely. Hence it follows that $G_n(\alpha) \to \psi(\theta_0,\alpha) + \sigma^2$ uniformly in $\alpha \in \beta S$, almost surely.

Let V be any open neighborhood of θ_0 in βS. Assumption (2) implies that $\inf\{\psi(\theta_0, \alpha)|\alpha \in V^c\} = \delta > 0$. However, since $\psi(\theta_0, \alpha)$ is continuous at θ_0, there exists a neighborhood W of θ_0 in βS such that $\sup\{\psi(\theta_0, \alpha)|\alpha \in W\} \leq \delta/2$. Recall that $\{S_n\}$ is an increasing sequence of subsets of S such that $\bigcup_1^\infty S_n$ is dense in S; hence $W \cap S_n$ is nonempty for all n sufficiently large. It follows that, eventually $\hat{\theta}_n$ belongs to $V \cap S$, almost surely. Thus $\hat{\theta}_n \to \theta_0$ almost surely. □

REMARK. A sufficient condition for $\hat{\theta}_n$ to converge to θ_0 almost surely is the existence of a $\delta > 0$ and a compact subset K of S such that $\phi(\theta_0, \theta) = 0$ iff $\theta = \theta_0$ and $\phi(\theta_0, \theta) \geq \delta$ when $\theta \in K^c$. Now if we suppose that S is a locally compact, but not a compact topological space, then one-point compactification of S exists and let b denote this additional point. If $\phi(\theta_0, \theta)$ has a continuous extension $\phi_1(\theta_0, \alpha)$ to this compactified space, then $K = \{\theta \in S|\phi(\theta_0, \theta) \leq \delta\}$ is a compact subset of S if $\phi_1(\theta_0, b) > \delta > 0$. Clearly $\phi(\theta_0, \theta) \geq \delta$ when $\theta \in K^c$. Hence, in this case $\hat{\theta}_n \to \theta_0$ almost surely.

The preceding remark is also related to the following result of Malinvaud [4]. The estimators $\hat{\theta}_n$ in Theorem 3.3 are those obtained by using the least squares procedure and are assumed to exist.

THEOREM 3.3 (Malinvaud [4]). *Suppose that model 1 and assumptions made in the introduction are satisfied. Assume* (1) *S is any subspace of R^p and suppose that the least squares estimators exist,* (2) *there exist a* $\delta > 0$, *an* n_0, *and a compact subset K of S such that* $1/n\sum_1^n(f_i(\theta) - f_i(\theta_0))^2 \geq 4\sigma^2 + \delta$ *for each* $\theta \in K^c$, $n \geq n_0$, (3) $\{1/n\sum_1^n(f_i(\alpha) - f_i(\beta))^2\}$ *converges uniformly in* $(\alpha, \beta) \in K \times K$ *to* $\phi(\alpha, \beta)$, *and* (4) $\phi(\theta_0, \theta) = 0$ *for* $\theta \in K$ *only when* $\theta = \theta_0$. *Then* $\hat{\theta}_n \to \theta_0$ *in probability.*

It should be mentioned that the Stone–Cech compactification βS of S was used in the development of these results because (1) βS always exists and (2) each bounded continuous f_i: $S \to R$ has a continuous extension to βS. In general, other compactifications of S do not always possess these two properties. For example, if S is a totally bounded metric space and each f_i: $S \to R$ is uniformly continuous, then the completion of S is in fact a compactification and may be used in place of βS since each such f_i has a continuous extension to the completion space. Similarly, if S is a separable, locally compact topological space such that each f_i: $S \to R$ has a continuous extension to the one-point compactification space, then this compactification may be used in place of βS.

The following is an example when the one-point compactification of S may not be used.

EXAMPLE 3.1. Let $S = [0,1) \cup (1,2]$ be a subspace of R and define

$$f_i(\theta) = \begin{cases} \theta, & 0 \le \theta < 1 \\ \theta + \dfrac{1}{i}, & 1 < \theta \le 2 \end{cases} \qquad i \ge 1.$$

The model is $y_i = f_i(\theta_0) + \varepsilon_i$, where $\{\varepsilon_i\}$ are independent and identically distributed random variables with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma^2 < +\infty$. It is straightforward to show that $1/n\sum_1^n f_i(\alpha)f_i(\beta) \to \alpha\beta$ and $1/n\sum_1^n(f_i(\alpha)f_i(\beta))^2 \to \phi(\alpha, \beta) = (\alpha - \beta)^2$, each uniformly in $(\alpha, \beta) \in S \times S$. Since $\phi(\theta_0, \theta) = (\theta_0 - \theta)^2 \ge \delta$ when θ is sufficiently close to 1, then it follows that $\psi(\theta_0, \alpha) > 0$ for each $\alpha \in \beta S - S$. Hence, by Theorem 3.2, $\hat{\theta}_n \to \theta_0$ almost surely. Note that the one-point compactification cannot be used since each f_i fails to have a continuous extension to it.

4. Related model. Quite often in applications, each regression function $f_i(\theta)$ in model 1 can be written as $f(x_i, \theta)$, $i \ge 1$. This leads to the study of a closely related model. Let $\mathscr{X}$ be any subspace of R^m and $\mathscr{B}_m$ be the set of Borel subsets of $\mathscr{X}$. *Model 2* is of the form $y_i = f(x_i, \theta_0) + \varepsilon_i$ with the following assumptions: (1) $\{x_i\}$ is any sequence selected from $\mathscr{X}$ such that the corresponding sequence $\{\mu_n\}$ of empirical probability measures converges weakly to some probability measure μ on $(\mathscr{X}, \mathscr{B}_m)$; (2) $\{\varepsilon_i\}$ are independent and identically distributed random variables each having mean zero and finite variance σ^2; (3) S is a separable, completely regular topological space; (4) f: $\mathscr{X} \times S \to R$ is a bounded, continuous function; and (5) a compactification T of S exists such that f has a continuous extension g: $\mathscr{X} \times T \to R$.

Malinvaud [4] and Gallant [2] have given consistency results for these models when $(\mathscr{X}, \mathscr{B}_m)$ and S are compact subspaces of finite-dimensional Euclidean spaces. Theorem 3.2 is used to extend these results to the noncompact case.

THEOREM 4.1. *Suppose that model 2 and assumptions previously made hold. Moreover, assume that* $\mu(x \in \mathscr{X} | g(x, \theta_0) \ne g(x, \alpha)\} > 0$ *whenever* $\alpha \in T$, $\alpha \ne \theta_0$. *Then* $\hat{\theta}_n \to \theta_0$ *almost surely.*

PROOF. Since $g(x, \alpha)g(x, \beta)$ is a bounded, continuous function defined on $\mathscr{X} \times T \times T$, where T is compact, then $\{g(\cdot, \alpha)g(\cdot, \beta) | \alpha, \beta \in T\}$ is an equicontinuous family on $\mathscr{X}$. It follows, by a result of Rango Rao (e.g., see problem 2.8 of Billingsley [1]), that

$$\frac{1}{n}\sum_{1}^{n} g(x_1, \alpha)g(x_i, \beta) = \int g(x, \alpha)g(x, \beta)\, d\mu_n(x) \to \int g(x, \alpha)g(x, \beta)\, d\mu(x)$$

uniformly in $(\alpha, \beta) \in T \times T$. Hence, condition (1) of Theorem 3.2 holds. This also implies that

$$\frac{1}{n}\sum_{1}^{n} (g(x_i, \alpha) - g(x_i, \beta))^2 \to \int (g(x, \alpha) - g(x, \beta))^2\, d\mu(x) = \psi(\alpha, \beta)$$

uniformly in $(\alpha, \beta) \in T \times T$. The hypothesis implies that if $\alpha \in T$, then $\psi(\theta_0, \alpha) = 0$ iff $\alpha = \theta_0$. Hence, condition (2) of Theorem 3.2 holds and thus $\hat{\theta}_n \to \theta_0$ almost surely. □

Let us conclude with an extension to the noncompact case of the example given by Jennrich ([3], page 642).

EXAMPLE 4.1. Let b be a fixed positive real number and let the parameter space $S = \{(\theta_1, \theta_2) | 0 < \theta_1 < b, \theta_2 > 0\}$ be a subspace of R^2. Suppose that $\mathscr{X} = (0, \infty)$ and $\{x_i\}$ is a sequence selected in $\mathscr{X}$ such that $\mu_n \to \mu$ weakly on $(\mathscr{X}, \mathscr{B}_1)$, where μ is nondegenerate.

Consider the model $y_i = f(x_i, \theta_0) + \varepsilon_i$, where $\theta_0 \in S$ and $f(x, \theta) = \theta_1 e^{-\theta_2 x}$, $\theta' = (\theta_1, \theta_2) \in S$. Assume that $\{\varepsilon_i\}$ are independent and identically distributed random variables each having mean zero and finite variance. Let

$$T = \{(\alpha_1, \alpha_2) | 0 \le \alpha_1 \le b, 0 \le \alpha_2 \le +\infty\}$$

be equipped with the product topology; then T is a compactification of S. Note that $g(x, \alpha) = \alpha_1 e^{-\alpha_2 x}$ is a continuous extension of $f(x, \theta)$ to $\mathscr{X} \times T$. Hence, the conditions of model 2 are satisfied. The argument used by Jennrich ([3], page 642) shows that, since μ is not degenerate, $\mu\{x \in \mathscr{X} | g(x, \alpha) \neq g(x, \theta_0)\} > 0$ when $\alpha \in T$ and $\alpha \neq \theta_0$. It follows from Theorem 4.1 that $\hat{\theta}_n \to \theta_0$ almost surely.

REMARK. In the present paper under suitable regularity conditions, the sequence of least squares estimators over an increasing sequence of compact sets whose union is dense in the noncompact parameter space has been shown to be strongly consistent. This process ensures the existence and the measurability of the sequence of estimators. However, present study does not address the problem of selection of these compact sets. One possible way is to choose monotonically increasing sets of grid points in closed balls of increasingly large radii with a convenient center α_0 using the metric $\phi^{1/2}$ or some variable metric like ρ_n where

$$\rho_n(\alpha, \beta) = \left(\frac{1}{n}\sum_{1}^{n} (f_t(\alpha) - f_t(\beta))^2\right)^{1/2}.$$

Now ρ_n, even though variable, will become stable with increasing sample size as it converges uniformly to $\phi^{1/2}$. Of course, in any particular situation with a given value of n, we would like to choose as large a compact subset of S as possible.

Acknowledgment. The authors wish to thank the referees for numerous improvements in the original manuscript, in particular, simplification of our original proofs of Theorems 3.2 and 4.1.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
[2] GALLANT, A. R. (1975). Seemingly unrelated nonlinear regressions. *J. Econometrics* **3** 35–50.
[3] JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.
[4] MALINVAUD, E. (1970). The consistency of nonlinear regressions. *Ann. Math Statist.* **41** 956–969.
[5] WILANSKY, A. (1970). *Topology for Analysis*. Ginn, Waltham, Mass.
[6] WU, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FLORIDA 32816

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8203