Cowles Foundation Paper 693

*Econometrica*, Vol. 55, No 6 (November, 1987), 1465-1471

# CONSISTENCY IN NONLINEAR ECONOMETRIC MODELS: A GENERIC UNIFORM LAW OF LARGE NUMBERS

BY DONALD W. K. ANDREWS[1]

## 1. INTRODUCTION AND CONCLUSION

A UNIFORM LAW of large numbers (LLN) is a primary ingredient used in proving consistency of estimators in nonlinear econometric models. Thus, in a well-known review article, Burguete, Gallant, and Souza (1982, p. 162) introduce a uniform LLN with the statement: "The following theorem is the result upon which the asymptotic theory of nonlinear econometrics rests." The purpose of this paper is to provide a *generic* uniform LLN that is sufficiently general to incorporate most applications of uniform LLNs in the nonlinear econometrics literature.

In summary, the paper presents a result that can be used to turn state of the art pointwise LLNs into uniform LLNs over compact sets with the addition of a single smoothness condition—a Lipschitz, derivative, or continuity condition. In contrast to other uniform LLNs that appear in the literature (e.g., see Jennrich (1969), Malinvaud (1970), Gallant (1977), Bierens (1984), and Amemiya (1985)), the uniform LLN given here allows the full range of temporal dependence and heterogeneity of summands (i.e., nonidentical distributions) that is available with pointwise LLNs.

The paper also points out that a frequently used uniform LLN, due to Hoadley (1971, Theorem A.5), only applies to bounded random variables. This is very restrictive. For example, it prevents the use of Hoadley's uniform LLN in standard proofs of consistency of least squares in regression models with normally distributed errors.

A recent paper by Pötscher and Prucha (1986) follows the approach introduced here and states a *generic* uniform LLN. Their results complement those given here, since neither their assumptions nor those given here are weaker.

## 2 A GENERIC UNIFORM LAW OF LARGE NUMBERS

First, we introduce some notation: Let $\{W_i: i=1,2,\dots\}$ be a sequence of $\boldsymbol{W}$-valued random variables (rv's) defined on a probability space $(\Omega, \mathscr{B}, P)$, where $\boldsymbol{W}$ is an arbitrary set with $\sigma$-algebra $\boldsymbol{W}$. The summands for the uniform LLN are given by $\{q_i(W_i,\theta): i=1,2,\dots\}$, where $q_i$ is a function from $\boldsymbol{W}\times\Theta$ to $R^1$ for some metric space $\Theta$. Let $B(\theta,\rho)$ be the open ball around $\theta$ of radius $\rho$ (i.e., $B(\theta,\rho)=\{\tilde\theta\in\Theta: d(\tilde\theta,\theta)<\rho\}$). Define

$$
\begin{aligned}
&q_i^*(W_i,\theta,\rho)=\sup\{q_i(W_i,\tilde\theta): \tilde\theta\in B(\theta,\rho)\} \quad \text{and}\\
&q_{*i}(W_i,\theta,\rho)=\inf\{q_i(W_i,\tilde\theta). \ \tilde\theta\in B(\theta,\rho)\}.
\end{aligned} \tag{1}
$$

We say that a sequence of rv's distributed under $P$, say $\{Z_i\}$, satisfies a pointwise strong (weak) LLN if $(1/n)\sum_{i=1}^n(Z_i-EZ_i)\to 0$ as $n\to\infty$ a.s. $[P]$ (in probability under $P$).

Our preliminary generic uniform LLN is based on the following assumptions:

ASSUMPTION A1: *$\Theta$ is a compact metric space.*

ASSUMPTION A2: (a) *$q_i(W_i,\theta)$, $q_i^*(W_i,\theta,\rho)$, and $q_{*i}(W_i,\theta,\rho)$ are rv's, $\forall\theta\in\Theta$, $\forall i$, $\forall\rho$ sufficiently small (where $\rho$ may depend on $\theta$).*

[1] I would like to thank Benedict M Potscher as well as Charles Bates, Whitney K. Newey, Joon Park, Peter C. B. Phillips, David Pollard, James H Stock, Andrew Weiss, Halbert White, two referees, and the co-editor, Gary Chamberlain, for helpful comments The author gratefully acknowledges support from an Alfred P Sloan Foundation Research Fellowship and National Science Foundation Grant No SES-8419789

(b) $\{q_i^*(W_i, \theta, \rho)\}$ *and* $\{q_{*i}(W_i, \theta, \rho)\}$ *satisfy pointwise strong (weak) LLNs,* $\forall \theta \in \Theta$, $\forall \rho$ *sufficiently small (where* $\rho$ *may depend on* $\theta$).

ASSUMPTION A3: For all $\theta \in \Theta$,

$$\limsup_{\rho \to 0} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^{n} (Eq_i^*(W_i, \theta, \rho) - Eq_i(W_i, \theta)) \right| = 0$$

*and likewise with* $q_i^*(W_i, \theta, \rho)$ *replaced by* $q_{*i}(W_i, \theta, \rho)$.

THEOREM: *If Assumptions A1–A3 hold, then* (a) $(1/n)\sum_{i=1}^{n} Eq_i(W_i, \theta)$ *is continuous on* $\Theta$, *uniformly over* $n \geq 1$, *and*

(b) $$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} (q_i(W_i, \theta) - Eq_i(W_i, \theta)) \right| \to 0$$

*as* $n \to \infty$ *a.s.* $[P]$ *(in probability under* $P$)

See Section 3 for proofs of the results given in this Section.

COMMENTS: 1. The conclusion of part (a) of the Theorem is important for common proofs of consistency (e.g., see Bates and White (1985), Domowitz and White (1982), and White and Domowitz (1984)), because a fixed limit of $(1/n)\sum_{i=1}^{n} Eq_i(W_i, \theta)$ need not exist.

2. Assumption A3 can be verified directly or it can be established by verifying any of the stronger Assumptions A4, A5, or A6 given below.

3. Assumption A3 is equivalent to: For all $\theta \in \Theta$,

$$\limsup_{\rho \to 0} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^{n} E \sup_{\tilde{\theta} \in B(\theta, \rho)} |q_i(W_i, \tilde{\theta}) - q_i(W_i, \theta)| = 0.$$

When $\{W_i\}$ are identically distributed and $q_i(\cdot, \cdot)$ does not depend on $i$, this condition reduces to that of first moment continuity of $q(\cdot, \cdot)$, as used by Hansen (1982) and others.

4. The Theorem allows $\Theta$ to be an *infinite dimensional* metric space, as occurs in the application of uniform LLNs to nonparametric procedures, e.g., see Elbadawi, Gallant, and Souza (1983) and Epstein and Yatchew (1985).

5. The Theorem also holds with the normalization factor $1/n$ replaced by $1/b_n$, for any sequence of positive real numbers $\{b_n\}$ such that $b_n \to \infty$ as $n \to \infty$, provided the normalization factor $1/n$ in Assumptions A2(b) and A3 is replaced by $1/b_n$.

6. The Theorem also holds with the functions $q_i(\cdot, \cdot)$ and the rv's $\{W_i\}$ doubly indexed by $i$ and $n$, provided Assumptions A2 and A3 hold likewise. A uniform LLN with double indexing is useful in establishing Pitman drift results. Note that doubly indexed pointwise LLNs, needed for Assumption A2, most commonly are weak LLNs, not strong LLNs.

7. The fact that the rv's $\{q_i^*(W_i, \theta, \rho)\}$ and $\{q_{*i}(W_i, \theta, \rho)\}$ of Assumption A2 are suprema and infima of $\{q_i(W_i, \theta)\}$ over neighborhoods of $\theta$ introduces no particular difficulties in applying standard pointwise LLNs to verify these assumptions. One can apply pointwise LLNs directly to $\{q_i^*(W_i, \theta, \rho)\}$ and $\{q_{*i}(W_i, \theta, \rho)\}$. Or, if one desires a more concise statement of assumptions, it suffices to place sufficient moment and weak dependence conditions on the dominating rv's $\{q_i(W_i)\} \equiv \{\sup_{\theta \in \Theta} |q_i(W_i, \theta)|\}$ to ensure that $\{q_i(W_i)\}$ satisfy an LLN. Such conditions on $\{q_i(W_i)\}$ commonly are used in the literature (e.g., see Bates and White (1985, Assumptions B.3(i) and B.4(i)), Bierens (1984, Lemma 3), Burguete, Gallant, and Souza (1982, Theorem 1), and Domowitz and White (1982, Theorem 2.5)).

To illustrate this method, consider the following assumptions:

ASSUMPTION B1: $\{W_i\}$ *is a sequence of strong mixing rv's with mixing numbers* $\alpha(s)$, $s = 1, 2, \ldots$, *that satisfy* $\alpha(s) = o(s^{-\alpha/(\alpha-1)})$ *as* $s \to \infty$ *for some* $\alpha \geq 1$.[2]

[2] For example, see Domowitz and White (1982) for a definition of strong mixing—a condition of asymptotic weak dependence. Note that $\alpha = 1$ in B1 requires $\alpha(s) = 0$ for all $s$ large, i.e., "$m$-dependence" of $\{W_i\}$.

ASSUMPTION B2: (a) *$q_i^*(W_i, \theta, \rho)$, $q_{*i}(W_i, \theta, \rho)$, and $q_i(W_i)$ are r.v.'s and $q_i^*(\cdot, \theta, \rho)$ and $q_{*i}(\cdot, \theta, \rho)$ are measurable functions from $(\mathbf{W}, \mathscr{W})$ to $(R^1, \mathscr{B}_1)$, $\forall i$, $\forall \theta \in \Theta$, $\forall \rho$ sufficiently small, where $\mathscr{B}_1$ is the Borel $\sigma$-algebra on $R^1$.*

(b) $$\sup_{i \geq 1} Eq_i(W_i)^{\xi} < \infty \qquad \textit{for some } \xi > \alpha.$$

Given B1 and B2, McLeish's (1975) Theorem 2.10 implies that $\{q_i(W_i, \theta)\}$, $\{q_{*i}(W_i, \theta, \rho)\}$, and $\{q_i^*(W_i, \theta, \rho)\}$ satisfy pointwise strong LLNs for all $\theta \in \Theta$ and $\rho > 0$. Thus, we have the following corollary:

COROLLARY 1: *Assumptions B1 and B2 imply A2. Hence, under assumptions B1, B2, A1, and A3, the conclusions of the Theorem hold with the convergence of part* (b) *holding a.s.* $[P]$.

Next we give two smoothness conditions on $q_i(W_i, \theta)$, as a function of $\theta$, that imply A3. These conditions are more primitive than Assumption A3 and are easier to verify in some contexts.

ASSUMPTION A4. *For each $\theta \in \Theta$, there is a constant $\tau > 0$ such that $d(\tilde{\theta}, \theta) \leq \tau$ implies*

$$|q_i(W_i, \tilde{\theta}) - q_i(W_i, \theta)| \leq B_i(W_i) h[d(\tilde{\theta}, \theta)], \quad \forall i, \textit{ a.s. } [P],$$

*where $B_i : \mathbf{W} \to R^+$ and $h : R^+ \to R^+$ are nonrandom functions such that $B_i(W_i)$ is a rv, $\overline{\lim}_{n\to\infty} (1/n) \sum_{i=1}^n EB_i(W_i) < \infty$, $h[y] \downarrow h[0] = 0$ as $y \downarrow 0$, and $\tau$, $B_i$, $h$, and the null set may depend on $\theta$.*

ASSUMPTION A5: *$\Theta \subset R^p$; $q_i(W_i, \theta)$ is defined and differentiable in $\theta$ in a neighborhood of $\theta$ a.s. $[P]$, $\forall i$, $\forall \theta \in \Theta^*$ where $\Theta^*$ is some convex or open set that contains $\Theta$ (and where the null set may depend on $i$ and $\theta$); $\partial q_i(W_i, \theta)/\partial\theta$ and $\sup_{\tilde{\theta} \in \Theta^*} \|\partial q_i(W_i, \tilde{\theta})/\partial\theta\|$ are rv's, $\forall \theta \in \Theta$, $\forall i$; and*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^n E \sup_{\theta \in \Theta^*} \left\| \frac{\partial}{\partial \theta} q_i(W_i, \theta) \right\| < \infty.$$

COROLLARY 2. *The derivative condition A5 implies the Lipschitz condition A4, which in turn, implies Assumption A3. Thus, under Assumptions A1, A2, and A4 or A5, the conclusions of the Theorem hold.*

COMMENTS: 1. The smoothness conditions A4 and A5 are quite convenient, because they usually are implied by assumptions invoked to prove asymptotic normality of estimators that are based on optimization procedures (whose optimands are functions of the average $(1/n)\sum_{i=1}^n q_i(W_i, \theta)$). For example, Assumption 7 of Domowitz and White (1982), Assumptions 2 and 6(a) of White and Domowitz (1984), and Assumptions 6 and 8 of Burguete, Gallant, and Souza (1982) imply the derivative condition A5.

2. The Corollary also holds if the normalization factor $1/n$ is replaced by $1/b_n$ in Assumptions A2–A5 for $b_n$ as defined above or if the functions $q_i(\cdot, \cdot)$ and the rv's $\{W_i\}$ are doubly indexed by $i$ and $n$ in Assumptions A2–A5.

3. An alternative uniform LLN to those given above is that due to Hoadley (1971, Theorem A.5). As we now show, however, one of Hoadley's assumptions requires that the summands be uniformly bounded, in most cases of interest. In contrast, Assumptions A3, A4, and A5 above avoid this very restrictive boundedness requirement.

Suppose $\{W_i\}$ are i.i.d. and the functions $\{q_i\}$ do not depend on $i$. Let $\Theta \subset R^p$ be compact. Hoadley assumes (among other things) that $\{q(W_i, \theta)\}$ are continuous in $\theta$ uniformly in $i$, a.s. $[P]$. This assumption means that there is a set $C \in \mathscr{B}$ with $P(C) = 1$

such that for all $\omega \in C$, given any $\varepsilon > 0$ and $\theta \in \Theta$ there exists a $\delta > 0$ (which may depend on $\omega$, $\varepsilon$, and $\theta$) such that $\|\tilde{\theta} - \theta\| < \delta$ implies

$$(2) \qquad D_\omega \equiv \sup_{i \geq 1} |q(W_{i\omega}, \tilde{\theta}) - q(W_{i\omega}, \theta)| < \varepsilon,$$

where $W_{i\omega}$ denotes the value of $W_i$ when the sample path $\omega \in \Omega$ is realized.

The restrictiveness of this assumption arises because continuity of $q(W_i, \theta)$ uniformly over $i$ a.s. $[P]$ requires continuity of $q(w, \theta)$ uniformly over different values of $w$, since $W_{i\omega}$ changes with $i$. The typical sequence $\{W_{i\omega}: i = 1, 2, \ldots\}$, for $\omega \in C$, forms a dense subset of the support of $W_i$, and so, if $q(w, \theta)$ is continuous in $w$ for each $\theta$, the above assumption is equivalent to assuming continuity of $q(w, \theta)$ in $\theta$ uniformly over all $w$ in the support of $W_i$. For the functions $q(w, \theta)$ and the rv's $W_i$ that are considered in the nonlinear econometrics literature, this usually requires $q(w, \theta)$ to be uniformly bounded. If $q(w, \theta)$ is unbounded, as occurs with the least squares criterion function and with the criterion functions that correspond to ML estimators for models with normal errors, then Hoadley's assumption usually fails to hold.[3]

For example, consider the least squares estimator in the location model: $Y_i = \theta_0 + U_i$, $i = 1, \ldots, n$. For this model $W_i = (Y_i, 1)$ and $q(w, \theta) = (y - \theta)^2$. For simplicity, consider the case where $\theta = \theta_0$. We have

$$(3) \qquad D_\omega = \sup_{i \geq 1} |2(\theta_0 - \tilde{\theta}) U_{i\omega} + (\theta_0 - \tilde{\theta})^2| \geq \sup_{i \geq 1} \left| 2|\theta_0 - \tilde{\theta}| \cdot |U_{i\omega}| - (\theta_0 - \tilde{\theta})^2 \right|.$$

If the errors have unbounded support (as in the case of normality), then the right-hand side equals infinity for all $\omega$ in a set with probability one (for $\tilde{\theta} \neq \theta_0$), because $\lim_{n\to\infty} \max_{i \leq n} |U_i| = \infty$ a.s. $[P]$ (e.g., see Galambos (1978, Corollary 4.3.1)). Thus, Hoadley's continuity assumption fails.

Those papers in the nonlinear econometrics literature that make use of Hoadley's uniform LLN (e.g., Bates and White (1985), Domowitz and White (1982), Levine (1983), and White and Domowitz (1984))[4] have been designed carefully so that their results hold with any uniform strong LLN. In consequence, any of the generic uniform LLNs presented here can be used to replace that of Hoadley in these papers and with the requisite alterations of assumptions the results of the papers hold in the generality that is intended. For such papers, then, the restrictiveness of Hoadley's uniform LLN is not a serious problem.

We now consider cases where one does not wish to place as strong a smoothness condition on the random functions as Assumptions A4 or A5, but one is willing to place more restrictions on the heterogeneity of the marginal distributions of $\{W_i\}$ than are needed for pointwise LLNs.

Let $\mu$ be a $\sigma$-finite measure that dominates each of the marginal distributions of $W_i$, $i = 1, 2, \ldots$. (Such a measure exists, e.g., $\mu = \sum_{i=1}^{\infty} c_i P_i$, where the constants $c_i > 0$ satisfy $\sum_{i=1}^{\infty} c_i = 1$ and $P_i$ is the probability measure of $W_i$ under $P$.) Let $p_i(w)$ denote the density

[3] Hoadley (1971) does not apply his uniform LLN to the logarithm of the likelihood function in his proof of consistency of ML estimators. Instead, he applies it to bounded rv's. Thus, his ML consistency results are not affected by the boundedness implication of the assumptions of his uniform LLN.

His asymptotic normality results, however, also use his uniform LLN. For these results, he assumes that the matrix of second partial derivatives of the score function are continuous in $\theta$, uniformly over $i = 1, 2, \ldots$, a.s. $[P]$ (see his Condition N4). This can be quite restrictive. For example, in a linear regression model with i.i.d. normal errors and i.i.d. random regressors, it requires the support of the regressors to be bounded.

[4] Amemiya (1985, Theorem 4.2.2, p. 117) states a uniform LLN for i.n.i.d. rv's and references Hoadley's (1971) uniform LLN for a proof. The above discussion illustrates that his Theorem 4.2.2 is not a special case of Hoadley's result. In addition, the modification that Amemiya suggests to the proof of his i.i.d. result is not sufficient to establish his i.n.i.d. result. Counter-examples to his i.n.i.d. result can be constructed.

of $W_i$ with respect to $\mu$ and let a.e. $[\mu]$ abbreviate "almost everywhere with respect to the measure $\mu$." We introduce the following condition:

ASSUMPTION A6: (a) *$q_i(w, \theta)p_i(w)$ is continuous in $\theta$ at $\theta = \theta^*$ uniformly in $i$ a.e. $[\mu]$, for each $\theta^* \in \Theta$, and $q_i(w, \theta)$ is a $\mathscr{W}$/Borel-measurable function for each $i$ and each $\theta \in \Theta$.*

(b) $$\int \sup_{i \geq 1, \theta \in \Theta} |q_i(w, \theta)| p_i(w)\, d\mu(w) < \infty.$$

COROLLARY 3: *Assumption A6 implies Assumption A3. Hence, under Assumptions A1, A2, and A6, the conclusions of the Theorem hold.*

COMMENTS: 1. For independent nonidentically distributed (i.n.i.d.) contexts, this result is a close analogue of Hoadley's (1971) uniform LLN, but it replaces Hoadley's unnatural and restrictive continuity assumption by an assumption that does not require uniform boundedness of the summands.

For i.i.d. contexts where $q_i$ does not depend on $i$, A6 reduces to the usual assumptions: $q(w, \theta)$ is continuous in $\theta$ at $\theta = \theta^*$ a.e. $[\mu]$, $\forall \theta^* \in \Theta$, and $\int \sup_{\theta \in \Theta} |q(w, \theta)| p(w)\, d\mu(w) < \infty$.

2. Assumption A6(a) can be replaced by the following simpler, but stronger, conditions: (i) $\sup_{i \geq 1} p_i(w) < \infty$ *a.e.* $[\mu]$ and (ii) $q_i(w, \theta)$ is continuous in $\theta$ at $\theta = \theta^*$ uniformly in $i$ a.e. $[\mu]$, $\forall \theta^* \in \Theta$. Assumption (ii) is an analogue of Hoadley's (1971, Theorem A.5) Assumption (b). Assumption (ii) circumvents the restriction to bounded $q_i(w, \theta)$ functions, however, because it assumes continuity of $q_i(w, \theta)$ uniformly in $i$ for each fixed $w$ (independent of $i$) in a set with $\mu$-measure one, whereas Hoadley's Assumption (b) requires continuity of $q_i(W_{i\omega}, \theta)$ uniformly in $i$ for each $\omega$ in a set with $P$-probability one, where $W_{i\omega}$ depends on $i$. As seen above, it is the dependence of $W_{i\omega}$ on $i$ that causes Hoadley's Assumption (b) to be restrictive.

3. Assumption A6 allows $q_i(w, \theta)$ to have isolated discontinuities, as occurs, for example, when $q_i(w, \theta)$ corresponds to the defining equation of Manski's (1975, 1985) maximum score (MS) estimator. To see that the defining equation of the MS estimator satisfies A6 in the general non-i.i.d. setting, consider the MS estimator for the binary choice model. In this case, $q_i(W_i, \theta)$ equals $Y_i 1(X_i'\theta < 0) + (1 - Y_i)1(X_i'\theta \geq 0)$, where $W_i = (Y_i, X_i)$, $Y_i$ is a zero-one response variable, $X_i$ is a vector of exogenous variables, and $\Theta$ equals $\{\theta \in R^K : \|\theta\| = 1\}$. Suppose (1) $\int \sup_{i \geq 1} p_i(w)\, d\mu(w) < \infty$ and (2) $\mu(\{w : x'\theta = 0\}) = 0$, $\forall \theta \in \Theta$. Since $q(\cdot, \cdot)$ is bounded, condition (1) implies A6(b) Since $q(\cdot, \cdot)$ does not depend on $i$ and $\sup_{i \geq 1} p_i(w) < \infty$ a.e. $[\mu]$ by (1), A6(a) holds if $q(w, \theta)$ is continuous in $\theta$ at $\theta = \theta^*$ a.e. $[\mu]$, for each $\theta^* \in \Theta$. For any $\theta^* \neq \mathbf{0}$, $q(w, \theta)$ is continuous at $\theta^*$ for all $w = (y, x)$ for which $x'\theta^* \neq 0$. Assumption A6(a) now follows by assumption (2).

4. Assumption A6 does not depend on the choice of the dominating measure $\mu$.

5. The Corollary also holds if the functions $q_i(\cdot, \cdot)$ and the rv's $\{W_i\}$ are doubly indexed by $i$ and $n$ in Assumptions A2, A3, and A6 and if the supremum in Assumption A6(b) is taken over $n \geq 1$ as well as over $i \geq 1$.

## 3 PROOFS

PROOF OF THEOREM: Part (a) follows straightforwardly from A3. To show part (b), let $Q_i(\theta) = Eq_i(W_i, \theta)$ By A3, given $\varepsilon > 0$ and $\theta \in \Theta$, we can choose $\rho(\theta)$ so small that for all $n \geq 1$,

$$(4) \qquad \frac{1}{n}\sum_{i=1}^{n} Q_i(\theta) - \varepsilon \leq \frac{1}{n}\sum_{i=1}^{n} Eq_{*i}(W_i, \theta, \rho(\theta)) \leq \frac{1}{n}\sum_{i=1}^{n} Eq_i^*(W_i, \theta, \rho(\theta)) \leq \frac{1}{n}\sum_{i=1}^{n} Q_i(\theta) + \varepsilon.$$

The collection of balls $\{B(\theta, \rho(\theta)): \theta \in \Theta\}$ is an open cover of the compact set $\Theta$, and hence, has a finite subcover $\{B(\theta_l, \rho(\theta_l)): l = 1, \ldots, L\}$.

For any $\theta \in B(\theta_l, \rho(\theta_l))$, we have

$$(5) \qquad \frac{1}{n}\sum_{t=1}^{n}(q_t(W_t,\theta)-Q_t(\theta)) \leq \frac{1}{n}\sum_{t=1}^{n}[q_t^*(W_t,\theta_l,\rho(\theta_l))-Eq_{*t}(W_t,\theta_l,\rho(\theta_l))]$$
$$\leq \frac{1}{n}\sum_{t=1}^{n}[q_t^*(W_t,\theta_l,\rho(\theta_l))-Eq_t^*(W_t,\theta_l,\rho(\theta_l))]+2\varepsilon$$

and

$$\frac{1}{n}\sum_{t=1}^{n}(q_t(W_t,\theta)-Q_t(\theta)) \geq \frac{1}{n}\sum_{t=1}^{n}[q_{*t}(W_t,\theta_l,\rho(\theta_l))-Eq_{*t}(W_t,\theta_l,\rho(\theta_l))]-2\varepsilon.$$

We now have: For all $\theta \in \Theta$,

$$(6) \qquad \min_{l\leq L}\frac{1}{n}\sum_{t=1}^{n}[q_{*t}(W_t,\theta_l,\rho(\theta_l))-Eq_{*t}(W_t,\theta_l,\rho(\theta_l))]-2\varepsilon$$
$$\leq \frac{1}{n}\sum_{t=1}^{n}(q_t(W_t,\theta)-Q_t(\theta))$$
$$\leq \max_{l\leq L}\frac{1}{n}\sum_{t=1}^{n}[q_t^*(W_t,\theta_l,\rho(\theta_l))-Eq_t^*(W_t,\theta_l,\rho(\theta_l))]+2\varepsilon$$

The upper and lower bounds above are maxima and minima over finite numbers of rv's, and hence, converge to $2\varepsilon$ and $-2\varepsilon$, respectively, by Assumption A2. Since $\varepsilon > 0$ is arbitrary, the proof of part (b) is complete. *Q.E.D.*

PROOF OF COROLLARY 2: Using Assumption A4, we have

$$(7) \qquad K \equiv \overline{\lim_{\rho\to 0}}\sup_{n\geq 1}\left|\frac{1}{n}\sum_{t=1}^{n}(Eq_t^*(W_t,\theta,\rho)-Eq_t(W_t,\theta))\right|$$
$$\leq \overline{\lim_{\rho\to 0}}\sup_{n\geq 1}\frac{1}{n}\sum_{t=1}^{n}E\left|q_t^*(W_t,\theta,\rho)-q_t(W_t,\theta)\right|$$
$$\leq \overline{\lim_{\rho\to 0}}\sup_{n\geq 1}\frac{1}{n}\sum_{t=1}^{n}EB_t(W_t)\cdot h[\rho]=0.$$

The same argument holds with $q_t^*(W_t, \theta, \rho)$ replaced by $q_{*t}(W_t, \theta, \rho)$. Thus, A3 holds.

Next we show that A5 implies A4. Since $\Theta^*$ is open or convex and $q_t(W_t, \theta)$ is differentiable in $\theta$ in a neighborhood of $\theta$ a.s. $[P]$, $\forall \theta \in \Theta^*$, the mean value theorem applies for each $\theta \in \Theta$ and we have: For all $\theta \in \Theta$, $\exists \tau > 0$ such that $\forall \tilde{\theta} \in B(\theta, \tau)$,

$$(8) \qquad |q_t(W_t,\tilde{\theta})-q_t(W_t,\theta)| \leq \sup_{\theta^*\in\Theta^*}\left\|\frac{\partial}{\partial\theta}q_t(W_t,\theta^*)\right\|\cdot\|\tilde{\theta}-\theta\| \quad \text{a.s. } [P],$$

using the Cauchy-Schwartz inequality. Set $h[y] = y$ and

$$B_t(W_t)=\sup_{\theta^*\in\Theta^*}\left\|\frac{\partial}{\partial\theta}q_t(W_t,\theta^*)\right\|$$

to get A4. *Q.E.D*

PROOF OF COROLLARY 3: Under Assumption A6, we have

$$K = \overline{\lim_{\rho \to 0}} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^{n} \int (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) \, d\mu(w) \right| \tag{9}$$

$$\leq \overline{\lim_{\rho \to 0}} \int \sup_{i \geq 1} \left| (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) \right| \mu(w)$$

$$= \int \left[ \overline{\lim_{\rho \to 0}} \sup_{i \geq 1} \left| (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) \right| \right] d\mu(w) = 0,$$

where $K$ is as in the proof of Corollary 2, the second equality holds by the dominated convergence theorem using A6(b), and the third equality holds by A6(a). The same argument holds with $q_i^*(w, \theta, \rho)$ replaced by $q_{*i}(w, \theta, \rho)$. Thus, A3 holds. *Q.E.D.*

*Cowles Foundation for Research in Economics, Dept. of Economics, Yale University, New Haven, CT 06520, U.S.A.*

*Manuscript received February, 1986; final revision received March, 1987*

## REFERENCES

AMEMIYA, T (1985) *Advanced Econometrics.* Cambridge, MA Harvard University Press.

BATES, C., AND H. WHITE (1985) "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory,* 1, 151-178.

BIERENS, H. J. (1984): "Model Specification Testing of Time Series Regressions," *Journal of Econometrics,* 26, 323-353

BURGUETE, J. F, A. R. GALLANT, AND G. SOUZA (1982) "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews,* 1, 151-190

DOMOWITZ, I., AND H. WHITE (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics,* 20, 35-58

ELBADAWI, I., A. R. GALLANT, AND G. SOUZA (1983). "An Elasticity Can Be Estimated Consistently Without *A Priori* Knowledge of Functional Form," *Econometrica,* 51, 1731-1751.

EPSTEIN, L., AND A. YATCHEW (1985) "Nonparametric Hypothesis Testing Procedures and Applications to Demand Analysis," *Journal of Econometrics,* 30, 149-169.

GALAMBOS, J (1978) *The Asymptotic Theory of Extreme Order Statistics* New York: Wiley.

GALLANT, A. R. (1977)· "Three-Stage Least Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," *Journal of Econometrics,* 5, 71-88.

HANSEN, L P. (1982) "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica,* 50, 1029-1054

HOADLEY, B. (1971): "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," *Annals of Mathematical Statistics,* 42, 1977-1990

JENNRICH, R I. (1969): "Asymptotic Properties of Non-Linear Least Squares Estimation," *Annals of Mathematical Statistics,* 40, 633-643.

LEVINE, D. (1983): "A Remark on Serial Correlation in Maximum Likelihood," *Journal of Econometrics,* 23, 337-342.

MALINVAUD, E. (1970). "The Consistency of Nonlinear Regressions," *Annals of Mathematical Statistics,* 41, 956-969

MANSKI, C. F (1975)· "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics,* 3, 205-228.

——— (1985). "Semiparametric Analysis of Discrete Response," *Journal of Econometrics,* 27, 313-333.

MCLEISH, D L (1975) "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability,* 3, 826-836

POTSCHER, B. M, AND I R. PRUCHA (1986) "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results," Department of Economics Working Paper No. 86-9, June, 1986, University of Maryland.

WHITE, H, AND I DOMOWITZ (1984) "Nonlinear Regression with Dependent Observations," *Econometrica,* 52, 143-161.