

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY**

**Box 2125, Yale Station
New Haven, Connecticut 06520**

COWLES FOUNDATION DISCUSSION PAPER NO. 790

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

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

Donald W. K. Andrews¹

October 1985

Revised: April 1986

1. INTRODUCTION

A basic tool of modern econometrics is a uniform law of large numbers (LLN). It is a primary ingredient used in proving consistency and asymptotic normality of parametric and nonparametric estimators in nonlinear econometric models. Thus, in a well-known review article, Burguete, Gallant, and Souza [8, p. 162] introduce a uniform LLN with the statement: "The following theorem is the result upon which the asymptotic theory of nonlinear econometrics rests." So pervasive is the use of uniform LLNs, that numerous authors appeal to an unspecified generic uniform LLN (e.g., see [1, 2, 11, 17, 22, 29, 32]). Others appeal to some specific result (e.g., see [4, 5, 7, 8, 10, 12, 16, 33, 34, 35]). 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--either a Lipschitz condition or a derivative condition. The latter is particularly easy to verify, and is implied by common assumptions used to prove asymptotic normality of estimators. Thus, the additional condition is not particularly restrictive. In contrast to other uniform LLNs that appear in the literature, the one given here allows the full range of heterogeneity of summands (i.e., non-identical distributions), and temporal dependence, that is available with pointwise LLNs.

To fulfill the role of being a generic uniform LLN, we require a candidate uniform LLN to satisfy several conditions. First, it must be

sufficiently general to cover most estimation procedures considered in the nonlinear econometrics literature. Second, its conditions must be easily verifiable. Third, it must be flexible enough to accommodate underlying random variables (rv's) that exhibit the entire spectrum of heterogeneity and temporal dependence that permit pointwise LLNs to hold. That is, a generic uniform LLN should not be tied to particular assumptions concerning heterogeneity or temporal dependence. This allows it to be applied in as wide a variety of stochastic environments as possible.

Given these conditions, the advantage of having available a generic uniform LLN needs little explanation. It has use in a wide variety of static and dynamic nonlinear econometric modelling contexts.

None of the uniform LLNs in the literature satisfy all of the criteria specified above for a generic uniform LLN. Whence the results presented in Sections 2 and 3 below. Before considering these new uniform LLNs, however, we discuss various uniform LLNs that are available in the literature, and how their properties relate to the criteria listed above.

The most general uniform LLNs (in some respects, at least) are Banach space LLNs (e.g., see [9, 18, 20, 23]) and empirical process uniform LLNs (e.g., see [30]). These results are quite abstract, and hence, often impose conditions that are difficult to verify, at least in comparison with the special purpose uniform LLNs that are available. They generally are context specific, and impose particular assumptions regarding heterogeneity and temporal dependence. In addition, available results are not as general with respect to conditions such as temporal dependence, as are pointwise LLNs for rv's. For example, Vapnik-Cervonenkis-type uniform LLNs for empirical processes are much more highly developed for independent situations than for dependent situations.

A number of special purpose uniform LLNs are available in the econometrics and statistical literature for use with nonlinear models (e.g., see Jennrich [21], Malinvaud [25], Gallant [14], Bierens [6, 7], Amemiya [3, Theorem 4.2.1] and Ranga Rao [31]). These uniform LLNs have conditions that are relatively easy to verify, and are applicable to most parametric estimators considered in the econometrics literature (at least for certain stochastic environments). For our purposes, however, they are restrictive in several respects.

First, they apply only to functions of independent identically distributed (iid) or stationary rv's (see Jennrich [21], Amemiya [3, Theorem 4.2.1], and Ranga Rao [31]), to variables that behave just like stationary rv's (see Burguete, Gallant, and Souza [8] and Gallant [14]), or to rv's whose average marginal distributions converge weakly to some fixed distribution in the limit (see Bierens [6, 7], Jennrich [21], Malinvaud [25], and Ranga Rao [31]). Such assumptions restrict the degree of heterogeneity and/or temporal dependence of the underlying rv's. Furthermore, such results do not allow the functions of the rv's to depend upon the observation subscript. This characteristic can be restrictive. For example, it precludes applications to maximum likelihood (ML) estimators in heterogeneous (i.e., non-identically distributed) situations.² In addition, these uniform LLNs are context specific. They impose particular assumptions with respect to heterogeneity and asymptotic weak dependence. The parameter spaces for these uniform LLNs are restricted to be subsets of Euclidean space, except in [31]. This precludes their application in nonparametric contexts.

Two other special purpose uniform LLNs that appear in the literature are those due to Hoadley [19, Theorem A.5] and Amemiya [3, Theorem 4.2.2]. These uniform LLNs permit independent non-identically distributed (inid)

summands. Hoadley's uniform LLN fulfills many of our criteria. Its conditions are relatively easy to verify. And, although it is stated only for iid rv's, the same argument as used in its proof can be applied to rv's that satisfy any pointwise LLN. Unfortunately, it does not fulfill our first criterion of general applicability. One of Hoadley's assumptions implies that for many (or most) applications of interest, the underlying rv's must be uniformly bounded. This requirement is quite restrictive. (See the Appendix for an explanation of the boundedness requirement, and a discussion of its restrictive nature.)

Amemiya's [3, Theorem 4.2.2] uniform LLN for iid rv's is stated as an addendum to his iid uniform LLN. He notes that it is a special case of Hoadley's [19, Theorem A.5] uniform LLN, and hence, does not prove it formally, but only describes the appropriate modification of the proof of his iid result. For independent rv's, at least, Amemiya's Theorem 4.2.2 would appear to be stronger than our results below in certain important respects. The discussion in the Appendix, however, shows that his Theorem 4.2.2 actually is not a special case of Hoadley's result, due to the boundedness restriction mentioned above. Also, it does not appear that the modification that Amemiya suggests to the proof of his iid result is sufficient to establish his iid Theorem 4.2.2. (See the Appendix for further discussion.)

To conclude, none of the uniform LLNs currently available in the literature fulfill the criteria that we have specified for generic uniform LLNs. The uniform LLN presented in Section 2 below is designed to rectify this situation. In Section 3 we present a second uniform LLN. This result utilizes a weaker smoothness condition than the generic uniform LLN of Section 2, but it restricts the degree of heterogeneity of the summands

somewhat. The weaker smoothness condition allows for isolated discontinuities, such as occur in the defining equation of Manski's [26, 27] maximum score estimator for discrete choice models.

2. A GENERIC UNIFORM LAW OF LARGE NUMBERS

This section presents a uniform LLN that is designed to be flexible, easy to verify, and sufficiently general to cover most applications of interest. We take as a basic assumption the fulfillment of a pointwise LLN. Compactness of the parameter space and a single smoothness condition, either a Lipschitz condition or a derivative condition, are all that are required to transform the pointwise result into a uniform result.

One can adopt any pointwise LLN--there are quite a number for dependent rv's. No additional restrictions are imposed regarding heterogeneity or asymptotic weak dependence of the underlying rv's. Both weak and strong LLNs can be generated. The uniform LLN applies to both finite dimensional and infinite dimensional parameter spaces. The latter arise in nonparametric contexts. The proof of the result uses a well known technique, viz., direct finite approximation (e.g., see Pollard [30]), and follows the approach of Hoadley [19], up to a point.

We now introduce some notation: Let $\{W_i : i = 1, 2, \dots\}$ be a sequence of \mathcal{W} -valued rv's defined on a probability space (Ω, \mathcal{B}, P) , where \mathcal{W} is an arbitrary set. Consider functions $q_i : \mathcal{W} \times \Theta \rightarrow \mathcal{R}^1$, $i = 1, 2, \dots$, such that $q_i(W_i, \theta)$ is a rv for each $\theta \in \Theta$ (i.e., $q_i(W_i, \theta)$ is \mathcal{B} -measurable for each $\theta \in \Theta$), where Θ is a metric space with metric d . (Below, we discuss applications where the parameter θ is both finite dimensional and infinite dimensional.) Let $B(\theta, \rho)$ be

the open ball around θ of radius ρ (i.e., $B(\theta, \rho) = \{\tilde{\theta} \in \Theta : d(\tilde{\theta}, \theta) < \rho\}$).

Define

$$(1) \quad \begin{aligned} q_i^*(W_i, \theta, \rho) &= \sup\{q_i(W_i, \tilde{\theta}) : \tilde{\theta} \in B(\theta, \rho)\}, \text{ and} \\ q_{*i}(W_i, \theta, \rho) &= \inf\{q_i(W_i, \tilde{\theta}) : \tilde{\theta} \in B(\theta, \rho)\}.^3 \end{aligned}$$

We say that a sequence of rv's, such as $\{q_i^*(W_i, \theta, \rho) : i = 1, 2, \dots\}$, satisfies a strong (weak) LLN if

$$(2) \quad \begin{aligned} \frac{1}{n} \sum_{i=1}^n (q_i^*(W_i, \theta, \rho) - E q_i^*(W_i, \theta, \rho)) &\rightarrow 0 \text{ as } n \rightarrow \infty \\ \text{a.s. [P]} &\text{ (in probability under } P \text{).} \end{aligned}$$

Consider the following conditions:

- A1 $\{q_i^*(W_i, \theta, \rho)\}$ and $\{q_{*i}(W_i, \theta, \rho)\}$ satisfy a strong (weak) LLN, for all ρ sufficiently small, for all $\theta \in \Theta$ (where ρ may depend on θ).
- A2 Θ is a compact metric space.
- A3 For each $\theta \in \Theta$, there is a constant $\tau > 0$ such that $d(\tilde{\theta}, \theta) \leq \tau$ implies

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

where $B_i : \mathcal{W} \rightarrow \mathcal{R}^+$ and $h : \mathcal{R}^+ \rightarrow \mathcal{R}^+$ are non-random functions such that $B_i(W_i)$ is a rv, $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E B_i(W_i) < \infty$, $h[y] + h[0] = 0$ as $y \rightarrow 0$, and τ , B_i , and h may depend on θ .

Our generic uniform LLN is the following:

THEOREM 1: For $\{q_i(W_i, \theta)\}$ as defined above, if A1-A3 hold, then

- (a) $\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n (q_i(W_i, \theta) - Eq_i(W_i, \theta)) \right| \rightarrow 0$ as $n \rightarrow \infty$ a.s. [P] (in probability under P), and
- (b) $\frac{1}{n} \sum_{i=1}^n Eq_i(W_i, \theta)$ is continuous on Θ , uniformly over $n \geq 1$.

The proof of Theorem 1 uses

LEMMA 1: For $\{q_i(W_i, \theta)\}$ as defined above, if A0-A3 hold, then conclusions (a) and (b) of Theorem 1 hold, where A0 is given by

$$A0 \quad \text{For all } \theta \in \Theta, \quad \lim_{\rho \rightarrow 0} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^n (Eq_i^*(W_i, \theta, \rho) - Eq_i(W_i, \theta)) \right| = 0,$$

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

The proofs of Theorem 1, Lemma 1, and other results below are given in Section 4.

COMMENTS: 1. Theorem 1 is proved by showing that A3 implies A0. Thus, Lemma 1 is a stronger result than Theorem 1, and hence, may be of independent interest. For example, by using a condition that differs from A3 but implies A0, a second uniform LLN is given in Section 3 below. On the other hand, assumption A3 is more primitive than A0, and hence, more easily verified, since it places assumptions directly on the functions $\{q_i(\cdot, \cdot)\}$.

2. The conclusion of part (b) of the Theorem is important for common proofs of consistency (e.g., see [5, 10, 16, 35]), because a fixed limit of $\frac{1}{n} \sum_{i=1}^n Eq_i(W_i, \theta)$ need not exist.

3. The Lipschitz condition A3 is implied by the following derivative condition, which is simpler, and which may be easier to verify in some cases:

A4 Θ is a convex subset of \mathbb{R}^p ; $q_i(W_i, \theta)$ is totally differentiable in θ in a neighborhood of θ (or, a fortiori, has continuous partial derivatives in θ in a neighborhood of θ) a.s. [P], $\forall \theta \in \Theta$, $\forall i$; and

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

Replacing A3 by A4 in Theorem 1 gives the following:

COROLLARY 1: For $\{q_i(W_i, \theta)\}$ as defined above, if A1, A2, and A4 hold, then conclusions (a) and (b) of Theorem 1 hold.

4. The smoothness conditions A3 and A4 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 $\frac{1}{n} \sum_{i=1}^n q_i(W_i, \theta)$. For example, in the nonlinear regression context, assumption 7 of Domowitz and White [10] and assumptions 2 and 6a of White and Domowitz [35] are more than enough to imply our derivative condition A4 for the least squares function $q_i(W_i, \theta) = (Y_i - g_i(X_i, \theta))^2$, where $W_i = (Y_i, X_i)$ and the model is $Y_i = g_i(X_i, \theta_0) + U_i$. In more general contexts, Burguete, Gallant, and Souza's [8] assumption 6 for their class of least mean distance estimators of nonlinear models, and assumption 8 for their class of method of moments estimators of nonlinear models, and Gallant and White's [16] assumption 8 for estimators in dynamic nonlinear models, imply our derivative condition

A4 for the relevant summands $\{q_i(W_i, \theta)\}$.

5. The assumption A2 of compactness of Θ , and the assumption of convexity of Θ in A4, need not hold for the parameter space of interest, call it $\tilde{\Theta}$. In many cases, the functions $q_i(w, \theta)$ can be extended naturally from $\tilde{\Theta}$ to sets containing $\tilde{\Theta}$. Thus, it suffices to find a compact set Θ that contains $\tilde{\Theta}$, and for which Theorem 1 applies. (If A4 is to be adopted, then the chosen Θ also must be convex.)

6. In some cases, one needs a uniform LLN for averages of the form $\frac{1}{n} \sum_{i=1}^n q_i(W_i, \hat{\gamma}, \theta)$, where $\hat{\gamma} \rightarrow \gamma_0$ as $n \rightarrow \infty$, a.s. or in probability. This often arises in problems where preliminary estimators are employed (e.g., see [8]). Although it may be obvious, we point out that the application of the above Theorem with parameter (γ, θ) and parameter space $\Gamma \times \Theta$, where Γ is some compact neighborhood of γ_0 and Θ is compact, gives the desired result.

7. Assumptions A3 and A4 avoid the boundedness implication of Hoadley's [19, Theorem A.5] assumption (b), by avoiding conditions that have to hold uniformly over $i \geq 1$ a.s. [P]. (See the Appendix for a discussion of Hoadley's assumption (b).)

8. The use of uniform LLNs for infinite dimensional, non-Euclidean parameter spaces is growing in the econometrics literature as more non-parametric estimation and testing techniques are introduced (e.g., see Elbadawi, Gallant, and Souza [11], Epstein and Yatchew [12], and Gallant [15]). Theorem 1 above can be applied in such cases. As compared to some other uniform LLNs that apply to infinite-dimensional parameters, e.g., see Ranga Rao [31], Theorem 1 has the advantage that it imposes no restrictions on heterogeneity or temporal dependence, other than those needed for pointwise LLNs.

9. The fact that the rv's $\{q_i^*(W_i, \theta, \rho)\}$ and $\{q_{*i}(W_i, \theta, \rho)\}$ of assumption A1 are suprema and infima of $\{q_i(W_i, \theta)\}$ over neighborhoods of θ , introduces no particular difficulties in applying standard pointwise LLNs. 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 a LLN.⁴ Such conditions on $\{q_i(W_i)\}$ are commonly used in the literature (e.g., see Bates and White [5, Assumption a.5i], Bierens [7, Lemma 3], Burguete, Gallant, and Souza [8, Theorem 1], Domowitz and White [10, Theorem 2.5], Gallant and White [16, Theorem 6, Assumption b]).

To illustrate this method, suppose

B1 $\{W_i\}$ are strong mixing with mixing numbers $\alpha(\ell)$, $\ell = 1, 2, \dots$, that satisfy $\alpha(\ell) = o(\ell^{-\alpha/(\alpha-1)})$ as $\ell \rightarrow \infty$, for some $\alpha \geq 1$.

B2 $\sup_{i > 1} E q_i(W_i)^\xi < \infty$, for some $\xi > \alpha$.

See [4, 5, 10, or 35] for the definition of strong mixing, which is a condition of asymptotic weak dependence. Note that $\alpha = 1$ requires $\alpha(\ell) = 0$ for all ℓ large, i.e., "m-dependence" of $\{W_i\}$.

Given B1 and B2, McLeish's [28] Theorem 2.10 implies that $\{q_i(W_i, \theta)\}$, $\{q_{*i}(W_i, \theta, \rho)\}$, and $\{q_i^*(W_i, \theta, \rho)\}$ satisfy strong LLNs for all $\theta \in \Theta$ and $\rho > 0$. Thus, we have:

COROLLARY 2: For $\{q_i(W_i, \theta)\}$ as defined above, if B1, B2, A2, and either A3 or A4 hold, then the conclusion of Theorem 1 holds, and the convergence in part (a) of the conclusion holds a.s. [P].

An attractive feature of Theorem 1 above is that one can generate analogous corollaries with the strong mixing assumption B1 replaced by an assumption of ergodicity, or stochastic stability (see Bierens [7]), or any of a number of other mixing conditions.

3. A SECOND UNIFORM LAW OF LARGE NUMBERS

In Section 2 we stated a generic uniform LLN that holds for any sequence of random functions $\{q_i(W_i, \theta)\}$ that satisfies pointwise LLNs and a smoothness condition. We argued that this smoothness condition is quite convenient because it is implied by common assumptions used in the literature to establish asymptotic normality. This argument notwithstanding, there may be cases where one does not wish to place as strong a smoothness condition on the random functions as A3, but one is willing to place more restrictions on the heterogeneity of the marginal distributions of $\{W_i\}$ than are needed for pointwise LLNs.

In this section, we present a result that accommodates this tradeoff of assumptions. A noteworthy feature of our result is that for the special case of stationary ergodic sequences of random functions $\{q(W_i, \theta)\}$ (where $q(\cdot, \cdot)$ does not depend on i), it employs weaker conditions than those special purpose uniform LLNs that appear in the econometrics literature. Its conditions allow $q_i(W_i, \theta)$ to have some discontinuities in θ . This permits applications to such estimators as Manski's [26, 27] maximum score estimator (in both stationary and non-stationary contexts).

Let $a.e.[\mu]$ denote "almost everywhere with respect to the measure μ ." That is, a condition holds $a.e.[\mu]$ if it holds except on a set with μ -measure zero. We introduce the following assumption:

- A5 (a) The marginal distributions of W_1, W_2, \dots under P are each dominated by a measure μ .
- (b) $\frac{1}{n} \sum_{i=1}^n q_i(w, \theta) p_i(w)$ is continuous in θ at $\theta = \theta^*$ uniformly in n a.e. $[\mu]$, for each $\theta^* \in \Theta$, where $p_i(w)$ is the density of W_i with respect to μ .
- (c) $\int \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} |q_i(w, \theta)| p_i(w) d\mu(w) < \infty$.

Discussion of this assumption follows the next result:

THEOREM 2: For $\{q_i(W_i, \theta)\}$ as defined in Section 2, if A1, A2, and A5 hold, then conclusions (a) and (b) of Theorem 1 hold.

COMMENTS: 1. Assumption A5 imposes a weaker smoothness condition than A3, since it requires continuity, rather a Lipschitz condition, to hold. On the other hand, each part (a), (b), and (c) of A5 restricts the degree of heterogeneity of $\{W_i\}$. For parts (a) and (b) this is not difficult to see. For part (c), notice that if $\sup_{w \in \Omega} \sup_{\theta \in \Theta} |q_i(w, \theta)| < \infty$, then (c) requires $\int \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n p_i(w) d\mu(w) < \infty$. Without the "sup" this always holds, but with the "sup" it need not hold. For example, if W_i puts probability mass one on the point $1/i$, then

$$\int \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n p_i(w) d\mu(w) = \sum_{t=1}^{\infty} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n p_i(1/t) = \sum_{t=1}^{\infty} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n 1(i=t) = \sum_{t=1}^{\infty} 1/t = \infty.$$

Thus, part (c) also restricts the heterogeneity of $\{W_i\}$.

2. If $\{W_i\}$ is stationary ergodic and $q_i(\cdot, \cdot)$ does not depend on i , then A5 simplifies to yield conditions that are weaker than those of the following uniform LLNs: Jennrich [21] Theorem 2, Malinvaud [25, pp. 967-968], Gallant [14] Theorem 1, Ranga Rao [31] Theorem 6.2, Bierens [7] Lemma 2, and Amemiya [3] Theorem 4.2.1.

The simplification of A5 that occurs is the following: A5(a) is automatically satisfied by taking μ equal to the probability measure P_{W_1} of W_1 , A5(b) is equivalent to $q(w, \theta)$ being continuous in θ at $\theta = \theta^*$ a.s. $[P_{W_1}]$, $\forall \theta^* \in \Theta$, and A5(c) becomes $E \sup_{\theta \in \Theta} |q(W_1, \theta)| < \infty$. These conditions are weaker than those of the results listed above, because the continuity of $q(w, \theta)$ need not hold for all w . In consequence, $q(w, \theta)$ can have isolated discontinuities, as occurs, for example, when $q(w, \theta)$ corresponds to the defining equation of Manski's [26, 27] maximum score (MS) estimator. None of the uniform LLNs listed above allow such discontinuities.⁵

To see that the defining equation of the MS estimator satisfies A5 in the general non-iid setting, consider the MS estimator for the binary choice model. In this case, the function $q_i(W_i, \theta)$ is $Y_i 1(X_i' \theta < 0) + (1 - Y_i) 1(X_i' \theta \geq 0)$, where $W_i = (Y_i, X_i)$, Y_i is a zero-one response variable, and X_i is a vector of exogenous variables. Suppose (i) $\{W_i\}$ satisfies A5(a), (ii) $\int \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n p_i(w) d\mu(w) < \infty$, and (iii) X_i has absolutely continuous distribution for all i . Since $q(\cdot, \cdot)$ is bounded, condition (ii) implies A5(c). Since $q(\cdot, \cdot)$ does not depend on i , and (ii) implies that $\sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n p_i(w) < \infty$ a.e. $[\mu]$, A5(b) holds if $q(w, \theta)$ is continuous in θ at $\theta = \theta^*$ a.e. $[\mu]$, for each $\theta^* \in \Theta$. For given θ^* , $q(w, \theta)$ is continuous at θ^* for all $w = (y, x)$ for

which $x'\theta^* \neq 0$. By (iii), the dominating measure μ can be taken such that the μ -measure of the set $\{w : x'\theta^* = 0\}$ is zero, and so, A5(b) holds. Thus, under assumptions (i)-(iii), assumption A5 holds. By Theorem 2, then, a uniform LLN can be obtained for the defining equation of the MS estimator under any (additional) conditions that are sufficient for pointwise LLNs to hold.

This result extends straightforwardly to the MS estimator for multinomial choice models (see Manski [26] and Amemiya [3, pp. 339-348]). Also, under similar conditions assumption A5 holds for the defining equations of the least absolute deviations estimator of various nonlinear models, when it is defined as the solution to a system of equations.⁶ In this case, its defining equations involve the sign function, and hence, exhibit discontinuities.

3. Assumption A5(b) 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 θ at $\theta = \theta^*$ uniformly in i a.e. $[\mu]$, $\forall \theta^* \in \Theta$. The latter assumption (ii) is an analogue of Hoadley's [19, Theorem A.5] assumption (b). Assumption (ii) circumvents the restriction to bounded $q(w, \theta)$ functions, however, that arises with Hoadley's assumption (b) (see the Appendix).

4. Assumption A5(c) can be replaced by a traditional domination condition such as: $p_i(w) \leq p(w)$ and $|q_i(w, \theta)| \leq q(w, \theta)$, $\forall i$, a.e. $[\mu]$, for some functions $p(w)$ and $q(w, \theta)$ that satisfy

$$\int \sup_{\theta \in \Theta} q(w, \theta) p(w) d\mu(w) < \infty.$$

4. PROOFS

PROOF OF THEOREM 1: Given Lemma 1, it suffices to show that A3 implies

A0. Using assumption A3, we have

$$\begin{aligned}
 K &\equiv \overline{\lim}_{\rho \rightarrow 0} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n |(\text{Eq}_i^*(W_i, \theta, \rho) - \text{Eq}_i(W_i, \theta))| \leq \overline{\lim}_{\rho \rightarrow 0} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n E|q_i^*(W_i, \theta, \rho) - q_i(W_i, \theta)| \\
 (3) \quad &\leq \overline{\lim}_{\rho \rightarrow 0} \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n E B_i(W_i) \cdot h[\rho] = 0.
 \end{aligned}$$

The same argument holds with $q_i^*(W_i, \theta, \rho)$ replaced by $q_{*i}(W_i, \theta, \rho)$.

Thus, A0 holds.

Q.E.D.

PROOF OF LEMMA 1: Part (b) follows straightforwardly from A0. To show part (a),

let $Q_i(\theta) = \text{Eq}_i(W_i, \theta)$. By A0, given $\epsilon > 0$ and $\theta \in \Theta$, we can choose $\rho(\theta)$ so small that for all $n \geq 1$,

$$(4) \quad \frac{1}{n} \sum_{i=1}^n Q_i(\theta) - \epsilon \leq \frac{1}{n} \sum_{i=1}^n \text{Eq}_{*i}(W_i, \theta, \rho(\theta)) \leq \frac{1}{n} \sum_{i=1}^n \text{Eq}_i^*(W_i, \theta, \rho(\theta)) \leq \frac{1}{n} \sum_{i=1}^n Q_i(\theta) + \epsilon.$$

The collection of balls $\{B(\theta, \rho(\theta)) : \theta \in \Theta\}$ is an open cover of the compact set Θ , and hence, by the Heine-Borel Theorem, has a finite subcover $\{B(\theta_\ell, \rho(\theta_\ell)) : \ell = 1, \dots, L\}$. We now have: for all $\theta \in \Theta$,

$$\begin{aligned}
 (5) \quad \min_{\ell < L} \frac{1}{n} \sum_{i=1}^n [q_{*i}(W_i, \theta_\ell, \rho(\theta_\ell)) - \text{Eq}_{*i}(W_i, \theta_\ell, \rho(\theta_\ell))] - \epsilon &\leq \frac{1}{n} \sum_{i=1}^n (q_i(W_i, \theta) - Q_i(\theta)) \\
 &\leq \max_{\ell < L} \frac{1}{n} \sum_{i=1}^n [q_i^*(W_i, \theta_\ell, \rho(\theta_\ell)) - \text{Eq}_i^*(W_i, \theta_\ell, \rho(\theta_\ell))] + \epsilon.
 \end{aligned}$$

The upper and lower bounds above are maxima and minima over finite numbers of rv's, and hence, converge to ϵ and $-\epsilon$, respectively, by A1. Since $\epsilon > 0$ is arbitrary, the proof of part (a) is complete. *Q.E.D.*

PROOF OF COROLLARY 1: We just need to show that A4 implies A3. Since Θ is convex and $q_i(W_i, \theta)$ is totally differentiable in θ in a neighborhood of θ a.s.[P], the mean value theorem applies for each $\theta \in \Theta$ and we have: For all $\theta \in \Theta$ and $\tilde{\theta} \in \Theta$,

$$(6) \quad |q_i(W_i, \tilde{\theta}) - q_i(W_i, \theta)| \leq \sup_{\theta^* \in \Theta} \left\| \frac{\partial}{\partial \theta} q_i(W_i, \theta^*) \right\| \cdot \|\tilde{\theta} - \theta\| \quad \text{a.s. [P]},$$

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

$$B_i(W_i) = \sup_{\theta^* \in \Theta} \left\| \frac{\partial}{\partial \theta} q_i(W_i, \theta^*) \right\|, \quad \text{and we are done.} \quad \text{Q.E.D.}$$

PROOF OF THEOREM 2: Given Lemma 1, it suffices to show that A5 implies A0. Under A5, we have

$$\begin{aligned} K &= \overline{\lim}_{\rho \rightarrow 0} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^n \int (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) du(w) \right| \\ (7) \quad &\leq \overline{\lim}_{\rho \rightarrow 0} \int \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^n (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) \right| du(w) \\ &= \int \left[\overline{\lim}_{\rho \rightarrow 0} \sup_{n \geq 1} \left| \frac{1}{n} \sum_{i=1}^n (q_i^*(w, \theta, \rho) - q_i(w, \theta)) p_i(w) \right| \right] du(w) = 0, \end{aligned}$$

where K is as in the proof of Theorem 1, the second equality holds by the dominated convergence theorem using A5(c), and the third equality holds by A5(b). The same argument holds with $q_i^*(q, \theta, \rho)$ replaced by $q_{*i}(w, \theta, \rho)$. Thus, A0 holds. Q.E.D.

APPENDIX

We now show how the uniform boundedness restriction in Hoadley's [19, Theorem A.5] uniform LLN arises from his assumptions (in cases of particular interest). We adopt the notation of Section 2. For ease of exposition, we suppose $\{W_i\}$ are iid, and the functions $\{q_i\}$ do not depend on i . Let $\Theta \subset \mathbb{R}^p$ be compact. Hoadley's uniform LLN applies to the rv's $\{q(W_i, \theta)\}$ and establishes the result of uniform convergence given in part (a) of Theorem 1.

To reach this conclusion, Hoadley assumes (among other things) that $\{q(W_i, \theta)\}$ are continuous in θ uniformly in i , a.s. [P]. This assumption means that there is a set $B \in \mathcal{B}$ with $P(B) = 1$ such that for all $\omega \in B$, given any $\epsilon > 0$ and $\theta_1 \in \Theta$ there exists a $\delta > 0$, which may depend on ω , ϵ , and θ_1 , such that $\|\theta - \theta_1\| < \delta$ implies

$$(8) \quad D_\omega = \sup_{i \geq 1} |q(W_{i\omega}, \theta) - q(W_{i\omega}, \theta_1)| < \delta,$$

where $W_{i\omega}$ denotes the value of W_i when the sample path $\omega \in \Omega$ is realized. (Inspection of Hoadley's proof shows that the assumption actually is used as stated above.)

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 $\{W_{i\omega} : i = 1, 2, \dots\}$, for $\omega \in B$, forms a dense subset of the support of W_i , and so, if $q(w, \theta)$ is continuous in w for each θ , the above assumption is equivalent to assuming continuity of $q(w, \theta)$ in θ 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 non-

linear econometrics literature, this is a very restrictive assumption.

For example, for LS estimators of nonlinear regression models, the function q commonly is taken to be $q(W_i, \theta) = (Y_i - g(X_i, \theta))^2$, where $W_i = (Y_i, X_i)$ and the model is $Y_i = g(X_i, \theta_0) + U_i$ (e.g., see [10, 35]). In ML estimation, q often is taken to be the logarithm of the likelihood function of an observation (e.g., see [5, 24]).⁷ In the regression case, q is unbounded, and hence, Hoadley's uniform LLN cannot be applied, if the errors have unbounded support (e.g., if the errors are normal) or if the regression function is unbounded. In the ML case, q is unbounded if the score function is unbounded (as occurs in models with normal errors). The functions q corresponding to numerous other estimators violate the boundedness restriction implied by Hoadley's assumption.

To illustrate the boundedness restriction in a particular example, consider the least squares estimator in the simplest of all regression models, viz., a regression model with only a constant term:

$$Y_i = \theta_0 + U_i, \quad i = 1, \dots, n.$$

For this model $W_i = (Y_i, 1)$ and the criterion function $q(w, \theta)$ is taken to be $(y - \theta)^2$. Again for simplicity, consider equation (8) for the case $\theta_1 = \theta_0$. We have

$$(9) \quad D_\omega = \sup_{i > 1} \left| 2(\theta_0 - \theta)U_{i\omega} + (\theta_0 - \theta)^2 \right| \geq \sup_{i > 1} \left| 2|\theta_0 - \theta| \cdot |U_{i\omega}| - (\theta_0 - \theta)^2 \right|.$$

If the errors have unbounded support (as in the case of normality), then the right-hand-side equals infinity for all ω in a set with probability one (for $\theta \neq \theta_0$). This follows, because

$\lim_{n \rightarrow \infty} \max_{i \leq n} |U_i| = \infty$ a.s. [P], by results for extreme order statistics (e.g., see Galambos [13, 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., [10, 35]) have been designed carefully so that their results hold with any uniform strong LLN. In consequence, the uniform LLN presented in Section 2 or Section 3 above can be used to replace that of Hoadley in these papers, and the results of the papers hold in the full generality that is intended. For such papers, then, the restrictiveness of Hoadley's uniform LLN is a nuisance, rather than a serious problem.

We now briefly discuss the inid uniform LLN of Amemiya [3, Theorem 4.2.2, p. 117], because this result appears to be stronger than our Theorems 1 and 2 in some respects. Amemiya's inid uniform LLN is given, more or less, as a corollary to his iid uniform LLN. He states that the inid result is a special case of Hoadley's uniform LLN, and in consequence, does not give a formal proof, but only mentions the modification of his iid proof that is needed. In view of the boundedness restriction that arises with Hoadley's uniform LLN, we see that Amemiya's assumptions actually are much weaker than Hoadley's, and hence, Hoadley's proof does not apply.

With regard to the modification of his iid proof, Amemiya states that the proof of the iid case can be generalized to the extent that

$$T^{-1} \sum_{t=1}^T g_t(\theta_i) \quad \text{and} \quad T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| \quad (\text{using his notation})$$

can be subjected to laws of large numbers. By following this method of proof, however, one also needs to show that

$$H \equiv \lim_{n \rightarrow \infty} \limsup_{T \rightarrow \infty} \max_{i \leq n} T^{-1} \sum_{t=1}^T E \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| = 0. \quad \text{Under the assumptions}$$

of his Theorem 4.2.2, this does not necessarily hold, as the following example illustrates.

For brevity, we adopt Amemiya's notation without comment. Let $m : \mathbb{R} \rightarrow [0,1]$ be a bounded, even, continuous function such that $m(z) = 1$ for $z = 0$, and $m(z) = 0$ for all $|z| > 1$. Let $\{y_t\}$ be a sequence of independent rv's with $P(y_t = 1/t) = P(y_t = -1/t) = 1/2$. Let $g(y, \theta) = m(\theta/y^2) \cdot \text{sgn}(y)$. Take Θ to be a compact subset of \mathbb{R} that contains 0. $\{y_t\}$, $g(\cdot, \cdot)$, and Θ satisfy the conditions of Amemiya's Theorem 4.2.2, but $H \neq 0$. To see the latter, suppose 0 is in Θ_i^n . Let $\tilde{\theta}$ equal θ_i , unless $\theta_i = 0$. If $\theta_i = 0$, take $\tilde{\theta}$ to be any non-zero element of Θ_i^n . Then, we have

$$(10) \quad \sup_{\theta \in \Theta_i^n} |g(y_t, \theta) - g(y_t, \theta_i)| \geq |g(y_t, 0) - g(y_t, \tilde{\theta})| = |1 - m(\tilde{\theta}/y_t^2) \text{sgn}(y_t)| = 1,$$

where the second equality holds for t sufficiently large. This implies that $H \neq 0$, as claimed. In addition, (10) implies that

$$\sum_{i=1}^n P(T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| > \epsilon) = 1 \text{ for } T \text{ large and } \epsilon < 1,$$

which also illustrates that Amemiya's iid method of proof does not apply.

To conclude, while Amemiya's Theorem 4.2.2 may be correct, its proof seems to be lacking as yet, and hence, it does not dominate our results.

FOOTNOTES

1. I would like to thank Charles Bates, Joon Park, Peter C. B. Phillips, David Pollard, James H. Stock, and Halbert White for helpful comments. The author gratefully acknowledges support from a Sloan Foundation Research Fellowship and National Science Foundation grant SES-8419789.
2. If the response variables and exogenous variables are assumed to be unconditionally stationary, then this restriction does not occur. The reason being, for a single observation the conditional log likelihood function of the response variables, given the current period exogenous variables and all past variables, depends on the observation subscript only through the variables.
In other heterogeneous situations, it is possible to avoid indexing the functions by the observation subscript, by adding an element to the underlying rv vector that captures the desired changes of the function over observations. To apply the LLNs referenced immediately above in the text, one then needs to assume that the average marginal distribution of the elongated rv vector converges weakly to some fixed distribution. This may be difficult to justify.
3. Note that $q_i^*(W_i, \theta, \rho)$ and $q_{*i}(W_i, \theta, \rho)$ are measurable, because we assume below that the metric space Θ is compact, and hence, is separable. This implies that $q_i^*(W_i, \theta, \rho)$ and $q_{*i}(W_i, \theta, \rho)$ equal supremum and infimum over countable sets, respectively.
4. For the same reason as in footnote 3, $q_i(W_i)$ is measurable. It also is finite a.s. [P], by compactness of Θ (assumption A2), and continuity of $q_i(W_i, \theta)$ in θ a.s. [P] (assumption A3 or A4).
5. Amemiya [3, Theorem 9.6.2] shows that his iid uniform LLN (Theorem 4.2.1) can be extended to cover the maximum score estimator. But, unlike Theorem 2 above, Amemiya's argument is specific to the case at hand, viz., the maximum score estimator, and is somewhat lengthy.
6. The least absolute deviations (LAD) estimator can be defined as the solution to a system of equations that involves the discontinuous sign function, or as the solution to a minimization problem that involves the continuous absolute value function. One reason for adopting the former definition, when proving consistency, is that it allows consistency to be established without any moment conditions on the errors in nonlinear regression and nonlinear simultaneous equations models. Using the latter definition, one needs to assume that the errors have one moment finite to apply the standard methods of proof. This argument for the use of the system of equations definition is enhanced by the fact that robustness with respect to fat-tailed error distributions is an important attribute of LAD estimators.

7. Hoadley [19] 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 θ , uniformly over $i = 1, 2, \dots$ a.s. [P] (see his condition N4). This can be quite restrictive. For example, in a linear regression model with iid normal errors and iid random regressors, it requires the support of the regressors to be bounded.

REFERENCES

- [1] Amemiya, T.: "The Nonlinear Two-Stage Least-Squares Estimator," Journal of Econometrics, 2 (1974), 105-110.
- [2] Amemiya, T.: "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," Econometrica, 45 (1977), 955-968.
- [3] Amemiya, T.: Advanced Econometrics. Cambridge, MA: Harvard University Press, 1985.
- [4] Andrews, D. W. K.: "Robust and Efficient Estimation of Nonlinear Regression Models with Dependent Errors," manuscript, Cowles Foundation, Yale University (1983).
- [5] Bates, C. and H. White: "A Unified Theory of Consistent Estimation for Parametric Models," Econometric Theory, 1 (1985), 151-178.
- [6] Bierens, H. J.: "Model Specification Testing of Time Series Regressions," Research Memorandum No. 8321, Department of Economics, University of Amsterdam (1983).
- [7] Bierens, H. J.: "Model Specification Testing of Time Series Regressions," Journal of Econometrics, 26 (1984), 323-353.
- [8] Burguete, J. F., A. R. Gallant, and G. Souza: "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," Econometric Reviews, 1 (1982), 151-190.
- [9] Dehling, H. and W. Philipp: "Almost Sure Invariance Principles for Weakly Dependent Vector-Valued Random Variables," Annals of Probability, 10 (1982), 689-701.
- [10] Domowitz, I. and H. White: "Misspecified Models with Dependent Observations," Journal of Econometrics, 20 (1982), 35-58.
- [11] Elbadawi, I., A. R. Gallant, and G. Souza: "An Elasticity Can Be Estimated Consistently without a a priori Knowledge of Functional Form," Econometrica, 51 (1983), 1731-1751.
- [12] Epstein, L. and A. Yatchew: "Nonparametric Hypothesis Testing Procedures and Applications to Demand Analysis," Discussion Paper, Department of Economics, University of Toronto, 1984.
- [13] Galambos, J.: The Asymptotic Theory of Extreme Order Statistics. New York: Wiley, 1978.
- [14] Gallant, A. R.: "Three-Stage Least Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," Journal of Econometrics, 5 (1977), 71-88.
- [15] Gallant, A. R.: "Identification and Consistency in Semi-Nonparametric Regression," Invited Paper for the World Congress of the Econometric Society, Cambridge, MA, 1985.

- [16] Gallant, A. R. and H. White: "Consistency and Asymptotic Normality for Parametric Estimation with Dependent Observations," UCSD Department of Economics Discussion Paper, 1985.
- [17] Hannan, E. J.: "Non-Linear Time Series Regression," Journal of Applied Probability, 8 (1971), 767-780.
- [18] Heinkel, B.: "On the Law of Large Numbers in 2-Uniformly Smooth Banach Spaces," Annals of Probability, 12 (1984), 851-857.
- [19] Hoadley, B.: "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," Annals of Mathematical Statistics, 42 (1971), 1977-1990.
- [20] Hoffman-Jorgensen, J. and G. Pisier: "The Law of Large Numbers and the Central Limit Theorem in Banach Spaces," Annals of Probability, 4 (1976), 587-599.
- [21] Jennrich, R. I.: "Asymptotic Properties of Non-Linear Least Squares Estimation," Annals of Mathematical Statistics, 40 (1969), 633-643.
- [22] Jorgenson, D. W. and J. Laffont: "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," Annals of Economic and Social Measurement, 3 (1974), 615-640.
- [23] Kuelbs, J. and J. Zinn: "Some Stability Results for Vector Valued Random Variables," Annals of Probability, 7 (1979), 75-84.
- [24] Levine, D.: "A Remark on Serial Correlation in Maximum Likelihood," Journal of Econometrics, 23 (1983), 337-342.
- [25] Malinvaud, E.: "The Consistency of Nonlinear Regressions," Annals of Mathematical Statistics, 41 (1970), 956-969.
- [26] Manski, C. F.: "Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics, 3 (1975), 205-228.
- [27] Manski, C. F.: "Semiparametric Analysis of Discrete Response," Journal of Econometrics, 27 (1985), 313-333.
- [28] McLeish, D. L.: "A Maximal Inequality and Dependent Strong Laws," Annals of Probability, 3 (1975), 826-836.
- [29] Phillips, P. C. B.: "On the Consistency of Non-Linear FIML," Econometrica, 50 (1982), 1307-1324.
- [30] Pollard, D.: Convergence of Stochastic Processes. New York: Springer-Verlag, 1984.
- [31] Ranga Rao, R.: "Relations between Weak and Uniform Convergence of Measures with Applications," Annals of Mathematical Statistics, 33 (1962), 659-680.
- [32] Robinson, P. M.: "Non-Linear Regression for Multiple Time-Series," Journal of Applied Probability, 9 (1972), 758-768.

- [33] White, H.: "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48 (1980), 817-838.
- [34] White, H.: "Nonlinear Regression on Cross-Section Data," Econometrica, 48 (1980), 721-746.
- [35] White, H. and I. Domowitz: "Nonlinear Regression with Dependent Observations," Econometrica, 52 (1984), 143-161.