Econometrica, Vol. 57, No. 3 (May 1989), 675–683

NOTES AND COMMENTS

A UNIFORM LAW OF LARGE NUMBERS FOR DEPENDENT AND HETEROGENEOUS DATA PROCESSES

BY BENEDIKT M. PÖTSCHER AND INGMAR R. PRUCHA[1]

1. INTRODUCTION

UNIFORM LAWS OF LARGE NUMBERS (ULLNs) constitute important tools in proving consistency of estimators in nonlinear econometric models. ULLNs consider sums of the form: $n^{-1}\sum_{t=1}^{n}[q_t(z_t,\theta) - Eq_t(z_t,\theta)]$, where (z_t) denotes a stochastic data generating process that takes its values in a space Z, θ is an element of the parameter space Θ, and q_t: $Z \times \Theta \to \mathbb{R}$. ULLNs provide conditions under which the above sum converges to zero uniformly over the parameter space.

Hoadley (1971) introduced a ULLN that allows for non-i.i.d. data processes. This ULLN (or some modified version of it) has been used widely in the recent econometric literature, see e.g. Bates and White (1985), Domowitz and White (1982), Levine (1983), White (1980), and White and Domowitz (1984). However, Andrews (1987) and Pötscher and Prucha (1986a, b) point out that the uniform continuity assumption maintained by this ULLN is severe. It precludes the analysis of many estimators and models of interest in economics. This suggests the need for alternative ULLNs. We note that in the above papers by Bates, Domowitz, Levine, and White the proofs of theorems regarding consistency are such that Hoadley's ULLN can be replaced by some alternative ULLN. Hence the theorems can be readily rectified and/or restored to their intended generality by the use of an alternative ULLN accompanied by a corresponding change in the catalogs of assumptions. For a detailed discussion of this issue see, e.g., Pötscher and Prucha (1986b).

The purpose of the present note is to introduce an alternative generic ULLN. The ULLN is generic in the sense that it transforms pointwise laws of large numbers into corresponding uniform ones. It maintains a set of assumptions that is relatively easy to verify and allows at the same time the analysis of a wide variety of estimators and models of interest in economics. The ULLN allows for temporal dependence and heterogeneity of the data process (z_t) as well as for heterogeneity in the functions q_t.[2] In addition to giving conditions under which $n^{-1}\sum_{t=1}^{n}[q_t(z_t,\theta) - Eq_t(z_t,\theta)]$ converges to zero uniformly, we also give additional conditions which guarantee that $n^{-1}\sum_{t=1}^{n} q_t(z_t,\theta)$ converges uniformly to a finite limit. This latter ULLN is essentially a generic generalization of ULLNs given in Bierens (1981, 1984), Pötscher and Prucha (1986a), and is closely related to results in Hansen (1982).

In a recent paper, Andrews (1987) introduces an alternative generic ULLN based on the assumption that $q_t(z,\theta)$ satisfies a Lipschitz-type smoothness condition with respect to θ. In contrast, the generic ULLN presented in this paper assumes that $q_t(z,\theta)$ satisfies a continuity-type condition jointly in z and θ. Therefore, the two ULLNs complement each other. Practically, Andrews' ULLN imposes more of a smoothness condition on the parameters, while the ULLN introduced here imposes more of a smoothness condition on the data. As an illustration of the difference between the two ULLNs, we note that the

[1]We would like to thank Charles Bates, Manfred Deistler, Ian Domowitz, Ronald Gallant, Nanhua Hu, Harry Kelejian, Whitney Newey, Ching-Zong Wei, Halbert White, Ernest Zampelli, the referees and co-editor for helpful comments. We assume responsibility, however, for any errors. The research was supported by the Max Kade Foundation and by the C. V. Starr Center's Focus Program for Capital Formation, Technological Change, Financial Structure, and Tax Policy.

[2]In Pötscher and Prucha (1986b), we derive also a more general generic ULLN than that presented here, at the expense of a more complex catalog of assumptions and a more complex proof.

former ULLN does not, e.g., readily apply to all functions $q_t(z,\theta) \equiv q(z,\theta)$ where $q(z,\theta)$ is jointly continuous, whereas the latter ULLN does. We also observe that our continuity-type smoothness condition should be easy to verify.

As a technical observation we note that the proofs of most ULLNs, including those mentioned above, are based on an approximation technique dating back to Wald (1949). This technique reduces the proof essentially to a verification of a single condition. Several authors refer to this condition (formulated within contexts of varying generality) as the first moment continuity condition; see e.g. Amemiya (1985), Andrews (1987), Hansen (1982), and Newey (1987). Therefore all of the above mentioned ULLNs differ in essence only in the way this first moment continuity condition is verified from basic assumptions.

2. A UNIFORM LAW OF LARGE NUMBERS

Let $(\Omega, \mathscr{U}, P)$ be a probability space and (Θ, ρ) a (nonempty) metric space with metric ρ. (We note that none of the subsequent results and conditions depend on the metric structure of (Θ, ρ); they only depend on the topology induced on Θ by the metric ρ. The choice of a fixed metric is only made for convenience.) Let $(z_t\colon\ t \in \mathbb{N})$ be a stochastic process defined on $(\Omega, \mathscr{U}, P)$ with values in a (nonempty) measurable space $(Z, \mathscr{B})$. For ease of exposition we first take Z to be a closed subset of $\mathbb{R}^s$ and $\mathscr{B}$ to be the corresponding Borel σ-field; more general choices for Z will be discussed later. Our generic ULLN is based on the following assumptions:

ASSUMPTION 1: *Θ is compact.*

ASSUMPTION 2 (Smoothness Condition): *$q_t(z,\theta) = \sum_{k=1}^K r_{kt}(z) s_{kt}(z,\theta)$ where the r_{kt} are $\mathscr{B}$-measurable real functions for all $t \in \mathbb{N}$ and $1 \leqslant k \leqslant K$, and the family $\{s_{kt}(z,\theta)\colon t \in \mathbb{N}\}$ of real functions is equicontinuous on $Z \times \Theta$ for all $1 \leqslant k \leqslant K$, i.e., $\sup_{t\in\mathbb{N}} |s_{kt}(z,\theta) - s_{kt}(z^0,\theta^0)| \to 0$ for $(z,\theta) \to (z^0,\theta^0)$ for all $(z^0,\theta^0) \in Z \times \Theta$ and each $1 \leqslant k \leqslant K$.*

Equicontinuity on $Z \times \Theta$ is, of course, defined w.r.t. the product topology on $Z \times \Theta$, where Z carries the standard Euclidean topology. Observe that Assumption 2 allows for discontinuities in the data z, given those discontinuities can be "separated" from the parameters θ. (As a consequence, this assumption covers, e.g., the case of the Tobit log-likelihood function where the r_{kt} represent indicator functions.) Assumption 2 is weaker than the assumption that the family $\{q_t(z,\theta)\colon t \in \mathbb{N}\}$ is equicontinuous. The latter assumption is contained as a special case and corresponds to, say, $K = 1$ and $r_{1t} \equiv 1$. However, as Newey (1987) points out, equicontinuity can often be obtained through a suitable redefinition of the data.[3]

We introduce the following notation: $D_t(z) = \sup\{|q_t(z,\theta)|\colon \theta \in \Theta\}$ and $q_t^*(z,\theta,\tau) = \sup\{q_t(z,\theta')\colon\ \rho(\theta,\theta') < \tau\}$, $q_{t*}(z,\theta,\tau) = \inf\{q_t(z,\theta')\colon\ \rho(\theta,\theta') < \tau\}$ for $\tau > 0$. Under Assumptions 1 and 2 the functions $D_t(z)$, $q_t^*(z,\theta,\tau)$ and $q_{t*}(z,\theta,\tau)$ are real valued and $\mathscr{B}$-measurable. (Note that Θ is separable and $q_t(z,\theta)$ is continuous in θ.)

ASSUMPTION 3 (Dominance Condition): (i) $\sup_n n^{-1}\sum_{t=1}^n E[D_t(z_t)^{1+\delta}] < \infty$ *for some* $\delta > 0$. (ii) $\sup_n n^{-1}\sum_{t=1}^n E|r_{kt}(z_t)| < \infty$ *for* $1 \leqslant k \leqslant K$.

[3] E.g., a set of functions q_t on $Z \times \Theta$, $Z \subseteq \mathbb{R}^s$, satisfying Assumption 2 can be written as an equicontinuous set of functions in new variables if the s_{kt} satisfy an additional condition: Define $\bar{z}_t = (w_{1t}, \ldots, w_{Kt}, z_t)$ with $w_{kt} = r_{kt}(z_t)$; then $\bar{q}_t(\bar{z}_t,\theta) = q_t(z_t,\theta)$ with $\bar{q}_t(\bar{z},\theta) = \sum_{k=1}^K w_k s_{kt}(z,\theta)$ and $\bar{z} = (w_1, \ldots, w_K, z)$. Then $\bar{q}_t(\bar{z},\theta)$ is equicontinuous if the s_{kt} are equicontinuous and $\sup_t |s_{kt}(z,\theta)| < \infty$.

If the r_{kt} represent, e.g., indicator functions, Assumption 3(ii) is trivially satisfied. More primitive conditions implying Assumption 3 are

$$\sup_n n^{-1} \sum_{t=1}^{n} E|r_{kt}(z_t)|^{p(1+\delta)} < \infty$$

and

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left(\sup_{\theta \in \Theta} |s_{kt}(z_t, \theta)|^{q(1+\delta)} \right) < \infty$$

for some $\delta > 0$ and some $p > 1$, $p^{-1} + q^{-1} = 1$, and $1 \leqslant k \leqslant K$. (This follows via Hölder's inequality.)

ASSUMPTION 4 (Pointwise Laws of Large Numbers): *For all $\theta \in \Theta$ there exists a sequence of positive numbers $\tau_i = \tau_i(\theta)$, $\tau_i \to 0$, such that for each τ_i the random variables $q_t^*(z_t, \theta, \tau_i)$ and $q_{t*}(z_t, \theta, \tau_i)$ satisfy a strong law of large numbers, i.e., as $n \to \infty$*

$$n^{-1} \sum_{t=1}^{n} [q_t^*(z_t, \theta, \tau_i) - Eq_t^*(z_t, \theta, \tau_i)] \to 0, \quad P\text{-}a.s.,$$

$$n^{-1} \sum_{t=1}^{n} [q_{t*}(z_t, \theta, \tau_i) - Eq_{t*}(z_t, \theta, \tau_i)] \to 0, \quad P\text{-}a.s.$$

Since $|q_t^*(z_t, \theta, \tau_i)| \leqslant D_t(z_t)$ and $|q_{t*}(z_t, \theta, \tau_i)| \leqslant D_t(z_t)$, Assumption 4 is clearly implied by the following stronger condition: Each sequence of random variables $(f_t(z_t))$, f_t $\mathscr{B}$-measurable, with $|f_t(z_t)| \leqslant D_t(z_t)$ satisfies a strong law of large numbers. We note that even this stronger condition is satisfied for large classes of processes (z_t), e.g. for large classes of α- and ϕ-mixing processes. (For a discussion and definition of α- and ϕ-mixing processes, see e.g. Domowitz and White (1982); for laws of large numbers for such processes, see McLeish (1975).) We furthermore need the following assumption:

ASSUMPTION 5: *The family $\{H^n: n \in \mathbb{N}\}$ is tight where $H^n = n^{-1}\sum_{t=1}^{n} H_t$ and H_t denotes the distribution of z_t on Z; i.e., $\lim_{m \to \infty} \sup_n n^{-1}\sum_{t=1}^{n} P(z_t \notin K_m) = 0$ for some sequence of compact sets $K_m \subseteq Z$.*

Either one of the following two assumptions is sufficient for Assumption 5 to hold. (This is demonstrated in the proof of Theorem 1 given below.)

ASSUMPTION 5A: *There exists a monotone function h: $[0, \infty) \to [0, \infty)$ with $h(x) \to \infty$ as $x \to \infty$ such that $\sup_n n^{-1}\sum_{t=1}^{n} Eh(\|z_t\|) < \infty$.*

ASSUMPTION 5B: *The process (z_t) is asymptotically stationary in the sense that $H^n = n^{-1}\sum_{t=1}^{n} H_t$ converges weakly to some probability measure H on Z.*

Clearly Assumption 5 (and Assumption 5A) are automatically satisfied if Z is compact. Assumption 5A is a rather mild moment condition. A typical choice for h is $h(x) = x^p$ where $p > 0$ can be arbitrarily small. Another example for h is $h(x) = \ln(1 + x)$. Clearly Assumption 5B is satisfied for any identically distributed process. Assumption 5B also holds for Markov processes with stationary transition probabilities if (i) Döblin's condition holds (see Doob (1953, Ch. 5.5) or Futia (1982, Th. 3.2, 3.4, 4.9)) or (ii) Feller's condition holds, a unique invariant distribution exists and Z is compact (see Breiman (1960)).

The common feature of Assumptions 5, 5A, and 5B is that they exclude situations where some mass of the average probability distributions H^n escapes a sequence of compact sets K_m in Z. This is achieved by Assumption 5A by placing bounds on moments of certain functions of z_t, and by Assumption 5B by requiring that the H^n converge to H.

Given the above catalog of assumptions, we have the following ULLN:

THEOREM 1: *If Assumptions 1–4 hold, and Assumption 5 or 5A or 5B holds, then $Eq_t(z_t,\theta)$ exists, is finite and is continuous on Θ, and*

(a) $$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|n^{-1}\sum_{t=1}^{n}\left[q_t(z_t,\theta)-Eq_t(z_t,\theta)\right]\right|=0 \quad P\text{-}a.s.,$$

(b) $$\left\{n^{-1}\sum_{t=1}^{n}Eq_t(z_t,\theta):\ n\in\mathbb{N}\right\} \text{ is equicontinuous on } \Theta.$$

All proofs are given in the Appendix. Theorem 1 may be viewed as a generic ULLN in the sense that it postulates in Assumption 4 only the existence of pointwise laws of large numbers for q_t^* and q_{t*} rather than to postulate sufficient conditions that imply those pointwise laws of large numbers. Specific ULLNs can now be obtained simply by verifying Assumption 4 from more primitive conditions on the process (z_t) as, e.g., α-mixing or ϕ-mixing conditions. Those mixing conditions have been frequently used in the recent econometric literature. The following corollary corresponds to these mixing conditions.

COROLLARY 1: *Given Assumptions 1, 2, 3(ii) hold, and Assumption 5 (or 5A or 5B) holds, and if, furthermore, the process (z_t) is ϕ-mixing [α-mixing] with mixing coefficients of size $-r/(2r-1)$ with $r\geq 1$ [of size $-r/(r-1)$ with $r>1$], and $\sup_t E[D_t(z_t)^{r+\delta}]<\infty$ for some $\delta>0$, then the conclusions of Theorem 1 hold.*

Theorem 1 only ensures that $n^{-1}\Sigma_{t=1}^{n}[q_t(z_t,\theta)-Eq_t(z_t,\theta)]$ converges to zero uniformly on Θ. In certain cases a stronger result, namely that also $n^{-1}\Sigma_{t=1}^{n}q_t(z_t,\theta)$ converges to some finite limit uniformly on Θ, is useful. Clearly, in order to obtain this stronger result we need some kind of asymptotic stationarity of the process. The following generic ULLN generalizes results in Bierens (1981, 1984), Pötscher and Prucha (1986a), and is closely related to results in Hansen (1982).

THEOREM 2: *Given Assumptions 1, 3(i), 4, 5B hold and $q_t(z,\theta)\equiv q(z,\theta)$ is continuous on $Z\times\Theta$, then the conclusions of Theorem 1 hold. Furthermore $\int q(z,\theta)\,dH(z)$ exists and is finite, is continuous on Θ, and*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|n^{-1}\sum_{t=1}^{n}q(z_t,\theta)-\int q(z,\theta)\,dH(z)\right|=0, \quad P\text{-}a.s.$$

Again, specific ULLNs can be obtained from Theorem 2 if Assumption 4 is replaced by, e.g., mixing conditions as in Corollary 1. Assumption 4 in Theorem 2 (as well as in Theorem 1 if $q_t\equiv q$) could also be replaced by strict stationarity and ergodicity. Of course, in this case it is possible to derive a ULLN under weaker conditions than those maintained here; see Hansen (1982).

The following corollary applies Theorems 1 and 2 to Markov processes. For a discussion of Markov processes, see e.g. Doob (1953, Ch. 5) and Futia (1982).

COROLLARY 2: *Let Assumption* 1 *hold, let* $q_t(z,\theta)\equiv q(z,\theta)$, *and let* (z_t) *be a Markov process with stationary transition probabilities. If* (*a*) *Assumptions* 2 *and* 3 *hold and the transition probabilities satisfy Döblin's condition and possess only one ergodic class, or if* (*b*) Z *is compact,* $q(z,\theta)$ *is continuous on* $Z\times\Theta$, *the transition probabilities satisfy Feller's condition and have a unique invariant distribution, then the conclusions of Theorem* 2 *hold.*

We note that given Döblin's and Feller's conditions are satisfied by the transition probabilities, then the assumption that only one ergodic class exists is equivalent to the usual uniqueness condition for the invariant distribution as, e.g., given in Futia (1982, p. 385). This follows immediately from Theorem 5.7 in Doob (1953) and Theorems 2.12, 3.3, and 4.9 in Futia (1982).

REMARK 1:[4] (a) So far we have assumed that the functions $q_t(z,\theta)$ are defined on $Z\times\Theta$ with Z a closed subset of $\mathbb{R}^s$. Now, more generally, let Z be a metrizable space with corresponding Borel σ-field. Then Theorem 1 and Corollary 1 still hold under Assumption 5 or if Assumption 5A [5B] is replaced by Assumption 5A′ [5B′] presented below, and Theorem 2 remains valid if Assumption 5B is replaced by Assumption 5B′.

ASSUMPTION 5A′: *There exists a sequence of compact sets* $K_m\subseteq Z$, *a* $\mathcal{B}$*-measurable function* $g\colon Z\to[0,\infty]$, *and real numbers* $r_m\to\infty$ *such that*: (i) $\{z\in Z\colon g(z)<r_m\}\subseteq K_m$ *and* (ii) $\sup_n n^{-1}\sum_{t=1}^n Eg(z_t)<\infty$.

ASSUMPTION 5B′: *The process* (z_t) *is asymptotically stationary in the sense that the probability measures* $H^n=n^{-1}\sum_{t=1}^n H_t$ *converge weakly to some probability measure* H *on* Z. *Furthermore,* H *and each of the* H^n *are tight* (*i.e.,* H *and each of the* H^n *have a* σ*-compact support*).

If, moreover, Z is a Borel subset of a separable and completely metrizable space, then Assumption 5B′ reduces to Assumption 5B in view of Theorem 3.2 in Parthasarathy (1967, p. 29). Clearly, if Z is a closed subset of $\mathbb{R}^s$, then Assumption 5A implies Assumption 5A′ upon choosing $K_m=Z\cap\{z\in\mathbb{R}^s\colon \|z\|\leqslant m\}$, $g(z)=h(\|z\|)$ and $r_m=h(m)$.

(b) Corollary 2 remains valid for any metrizable space Z that is a Borel subset of a separable, completely metrizable space. (Of course, for part (b) Z has also to be compact.)

(c) If the functions $q_t(z,\theta)$ can be extended to $W\times\Theta$, $Z\subseteq W$, such that Assumption 2 holds on W, we can substitute W for Z and try to perform the analysis on W rather than Z. This may be helpful if the results presented in this paper apply to W but not to Z.

REMARK 2: (a) Theorem 1 remains valid if the random variables z_t and the functions q_t are allowed to depend on t and n, given the following modifications are made: In Assumption 2 the functions $q_t(z,\theta)$ are replaced by $q_t^n(z,\theta)=\sum_{k=1}^K r_{kt}^n(z)s_{kt}^n(z,\theta)$ and the families $\{s_{kt}^n\colon t\leqslant n,\ t\in\mathbb{N},\ n\in\mathbb{N}\}$ are assumed to be equicontinuous; furthermore Assumptions 3, 4, 5, 5A, and 5B (or more generally 5A′, 5B′) are to be applied to the double indexed random variables and functions accordingly.

(b) "Weak" ULLNs corresponding to the "strong" ULLNs derived in this paper can be obtained by replacing almost sure convergence in Assumption 4 by convergence in probability.

Department of Econometrics, University of Technology Vienna, 1040, Vienna, Austria
and
Department of Economics, University of Maryland, College Park, MD 20742, U.S.A.

Manuscript received April, 1987; final revision received May, 1988.

[4]For a proof of the subsequent results see the Appendix.

APPENDIX

LEMMA A1: *Let $s_t(z,\theta)$ be a sequence of equicontinuous, real functions on $Z \times \Theta$ where Z is a metrizable space and (Θ, ρ) is a metric space; let $\varnothing \neq K \subseteq Z$ be compact. Then for each $\theta \in \Theta$ and each sequence $\tau_i \to 0$, $\tau_i > 0$, we have:*

$$\sup_{t \in \mathbb{N}} \sup_{z \in K} \sup_{\theta' \in B(\theta,\tau_i)} |s_t(z,\theta') - s_t(z,\theta)| \to 0 \quad \textit{as} \quad i \to \infty,$$

where $B(\theta,\tau_i) = \{\theta' \in \Theta: \rho(\theta,\theta') < \tau_i\}$.

PROOF: First note that the sup operators can be interchanged. Suppose the lemma is not true; then for some $\varepsilon > 0$ and some $\theta^0 \in \Theta$ we can find a sequence $\theta^i \to \theta^0$ for which $\sup_t \sup_{z \in K} |s_t(z,\theta^i) - s_t(z,\theta^0)| > \varepsilon$. But then there exists for each $i \in \mathbb{N}$ a point $z^i \in K$ such that $\sup_t |s_t(z^i,\theta^i) - s_t(z^i,\theta^0)| > \varepsilon$. Since K is compact we can find a subsequence $z^{i(k)}$ converging to some $z^0 \in K$. The last inequality then implies

$$\varepsilon < \sup_t |s_t(z^{i(k)},\theta^{i(k)}) - s_t(z^{i(k)},\theta^0)| \leqslant \sup_t |s_t(z^{i(k)},\theta^{i(k)}) - s_t(z^0,\theta^0)| + \sup_t |s_t(z^0,\theta^0) - s_t(z^{i(k)},\theta^0)|.$$

Since under equicontinuity both expressions on the r.h.s. converge to zero as $k \to \infty$, this yields a contradiction. *Q.E.D.*

PROOF OF THEOREM 1: Clearly, $Eq_t^*(z_t,\theta,\tau_i)$, $Eq_{t*}(z_t,\theta,\tau_i)$, and $Eq_t(z_t,\theta)$ exist by Assumption 3; the latter expectation is also continuous by Assumptions 2 and 3 and dominated convergence. Let K_m be any sequence of compact sets in Z and $B_i(\theta) = \{\theta' \in \Theta: \rho(\theta,\theta') < \tau_i\}$ with $\tau_i = \tau_i(\theta)$ as in Assumption 4. We first verify the first moment continuity condition, i.e., we show that for each $\theta \in \Theta$,

$$\text{(A.1)} \qquad \lim_{i \to \infty} \sup_n n^{-1} \sum_{t=1}^n E \sup\{|q_t(z_t,\theta') - q_t(z_t,\theta)|: \theta' \in B_i(\theta)\} = 0.$$

Clearly, the l.h.s. of (A.1) is (for $m \in \mathbb{N}$) not greater than $A_m + B_m$ where

$$A_m = \overline{\lim_{i \to \infty}} \sup_n n^{-1} \sum_{t=1}^n E\left[1_{K_m}(z_t) \sup\{|q_t(z_t,\theta') - q_t(z_t,\theta)|: \theta' \in B_i(\theta)\}\right] \quad \text{and}$$

$$B_m = \overline{\lim_{i \to \infty}} \sup_n n^{-1} \sum_{t=1}^n E\left[1_{Z-K_m}(z_t) \sup\{|q_t(z_t,\theta') - q_t(z_t,\theta)|: \theta' \in B_i(\theta)\}\right].$$

If $K_m = \varnothing$, then $A_m = 0$. Otherwise, using the definition of q_t it is readily seen that

$$A_m \leqslant \sum_{k=1}^K \overline{\lim_{i \to \infty}} \sup_n n^{-1} \sum_{t=1}^n E\left[1_{K_m}(z_t) |r_{kt}(z_t)| \times \sup\{|s_{kt}(z_t,\theta') - s_{kt}(z_t,\theta)|: \theta' \in B_i(\theta)\}\right]$$

$$\leqslant \sum_{k=1}^K \left[\overline{\lim_{i \to \infty}} \sup_t \sup_{z \in K_m} \sup\{|s_{kt}(z,\theta') - s_{kt}(z,\theta)|: \theta' \in B_i(\theta)\}\right] \times \left[\sup_n n^{-1} \sum_{t=1}^n E\left(1_{K_m}(z_t) |r_{kt}(z_t)|\right)\right].$$

The first expression in brackets on the r.h.s. of the last inequality is zero by Lemma A1; the second expression is bounded by Assumption 3(ii); hence $A_m = 0$. By Assumption 3(i) and the triangle inequality we see that $B_m \leqslant 2\sup_n n^{-1}\Sigma_{t=1}^n E[1_{Z-K_m}(z_t)D_t(z_t)]$. Applying twice Hölder's inequality with $p = 1+\delta$ and $q = 1+\delta^{-1}$ we get further:

$$B_m \leqslant 2\left[\sup_n n^{-1}\sum_{t=1}^n E\left[D_t(z_t)^p\right]\right]^{1/p}\left[\sup_n n^{-1}\sum_{t=1}^n P(z_t \notin K_m)\right]^{1/q}.$$

The first term in brackets is finite by Assumption 3(i). Under Assumption 5, choosing the sets K_m as in this assumption, we have $B_m \to 0$ as $m \to \infty$. To see that Assumption 5A implies Assumption 5, put $K_m = \{z \in Z\colon \|z\| \leqslant m\}$, which is compact since Z is closed; by Markov's inequality eventually

$$\sup_n n^{-1}\sum_{t=1}^n P(z_t \notin K_m) \leqslant \sup_n n^{-1}\sum_{t=1}^n P(\|z_t\| \geqslant m) \leqslant \sup_n n^{-1}\sum_{t=1}^n Eh(\|z_t\|)/h(m) \to 0.$$

Furthermore, Assumption 5B implies Assumption 5 as is seen from Theorem 6.2 in Billingsley (1968) observing that Z, as a closed subset of $\mathbb{R}^s$, is separable and complete. Hence we have shown (A.1) under each set of assumptions, which in turn implies part (b) of Theorem 1. Given (A.1) the proof of part (a) is now standard: Observe that (A.1) implies for $\varepsilon > 0$ and all $\theta \in \Theta$ the existence of an index $i(\theta, \varepsilon)$ such that for $i = i(\theta, \varepsilon)$:

$$\text{(A.2a)} \qquad \sup_n \left| n^{-1}\sum_{t=1}^n E\left(q_t^*(z_t,\theta,\tau_i) - q_t(z_t,\theta)\right)\right| < \varepsilon,$$

$$\text{(A.2b)} \qquad \sup_n \left| n^{-1}\sum_{t=1}^n E\left(q_{t*}(z_t,\theta,\tau_i) - q_t(z_t,\theta)\right)\right| < \varepsilon.$$

The family $\{B_{i(\theta,\varepsilon)}(\theta)\colon \theta \in \Theta\}$ is an open cover of Θ; hence by Assumption 1 there exist finitely many $\theta(\nu)$, $1 \leqslant \nu \leqslant L$, such that the corresponding balls $B(\nu) = B_{i(\theta(\nu),\varepsilon)}(\theta(\nu))$ cover Θ. It then follows using (A.2) that for all $\theta \in \Theta$:

$$\begin{aligned} -2\varepsilon + \min_\nu n^{-1}\sum_{t=1}^n &\left[q_{t*}(z_t,\theta(\nu),\tau(\nu)) - Eq_{t*}(z_t,\theta(\nu),\tau(\nu))\right] \\ &\leqslant n^{-1}\sum_{t=1}^n \left[q_t(z_t,\theta) - Eq_t(z_t,\theta)\right] \\ &\leqslant \max_\nu n^{-1}\sum_{t=1}^n \left[q_t^*(z_t,\theta(\nu),\tau(\nu)) - Eq_t^*(z_t,\theta(\nu),\tau(\nu))\right] + 2\varepsilon \end{aligned}$$

where $\tau(\nu) = \tau_{i(\theta(\nu),\varepsilon)}(\theta(\nu))$. Part (a) of the theorem now follows from Assumption 4 since ε is arbitrary. *Q.E.D.*

Proof of Corollary 1: Assumption 3(i) is directly implied by the assumptions of the corollary. Assumption 4 is readily seen to follow from Theorem 2.10 of McLeish (1975): Observe that q_t^* and q_{t*} are $\mathscr{B}$-measurable; hence $q_t^*(z_t,\theta,\tau_i) - Eq_t^*(z_t,\theta,\tau_i)$ and $q_{t*}(z_t,\theta,\tau_i) - Eq_{t*}(z_t,\theta,\tau_i)$ are α-mixing (ϕ-mixing) processes with mixing coefficients no greater than those of (z_t). Note that McLeish's definition of mixing coefficients is slightly weaker than the usual definition employed here. *Q.E.D.*

Proof of Theorem 2: Clearly, the third set of assumptions of Theorem 1 is satisfied (taking $K = 1$ and $r_{1t} \equiv 1$). It hence suffices to prove that $\int q(z,\theta)\,dH$ exists, is finite, and that it is the uniform limit of $n^{-1}\Sigma_{t=1}^n Eq(z_t,\theta) = \int q(z,\theta)\,dH^n$. Assumption 3 implies that $\sup_n \int |q(z,\theta)|^{1+\delta}\,dH^n < \infty$ for all $\theta \in \Theta$. Since $H^n \to H$ weakly, it follows analogously as in Theorem 5.4 of Billingsley (1968) that $\int q(z,\theta)\,dH$ exists, is finite, and that $\int q(z,\theta)\,dH^n \to \int q(z,\theta)\,dH$ for each $\theta \in \Theta$; part (b) of Theorem 1 implies that the family $\{\int q(z,\theta)\,dH^n\colon n \in \mathbb{N}\}$ is equicontinuous on Θ. The uniform convergence now follows from the theorem of Ascoli-Arzela. *Q.E.D.*

PROOF OF COROLLARY 2: To prove part (a) we show that Assumptions 4 and 5B hold (hence Theorem 1 applies), and that $n^{-1}\Sigma_{t=1}^{n} Eq(z_t,\theta) \to \int q(z,\theta)\,dH$ (where uniform convergence follows from the theorem of Ascoli-Arzela). Since Döblin's condition holds it follows from Futia (1982, Theorems 3.2, 3.4, 4.9) that H^n converges to some invariant distribution H in the variation norm and hence also weakly. Thus Assumption 5B is satisfied. Since only one ergodic class is assumed to exist, H is then the unique invariant distribution; cp. Doob (1953, Theorem 5.7). By Assumption 3(i) we have $\sup_n \int |q(z,\theta)|^{1+\delta}\,dH^n < \infty$. Since furthermore H^n converges to H in variation norm it follows similarly as in the proof of Theorem 2 that $\int |q(z,\theta)|\,dH < \infty$ and $n^{-1}\Sigma_{t=1}^{n} Eq(z_t,\theta) = \int q(z,\theta)\,dH^n \to \int q(z,\theta)\,dH$. An analogous result holds for q^* and q_*. It now follows from Doob (1953, Theorem 6.2) applied to the functions $q^*(z_t,\theta,\tau_i)$ and $q_*(z_t,\theta,\tau_i)$ that Assumption 4 is satisfied. This proves part (a). To prove part (b) note first that Assumption 3(i) trivially holds since $Z \times \Theta$ is compact. Now Assumption 5B follows from Proposition 1 in Breiman (1960) after integrating w.r.t. the initial distribution. Assumption 4 follows from the Theorem in Breiman (1960) (which also holds for arbitrary initial distributions) since q^* and q_* are continuous on Z and since Assumption 5B implies

$$n^{-1}\sum_{t=1}^{n} Eq^*(z_t,\theta,\tau_i) \to \int q^*(z,\theta,\tau_i)\,dH \quad \text{and}$$

$$n^{-1}\sum_{t=1}^{n} Eq_*(z_t,\theta,\tau_i) \to \int q_*(z,\theta,\tau_i)\,dH.$$

Applying Theorem 2 proves part (b). *Q.E.D.*

We note that the proofs of Theorem 1, 2, and Corollary 1 do not utilize the structure of Z beyond the fact that it is a metrizable space—except in the verification of Assumption 5 from Assumptions 5A or 5B. Therefore the following lemma proves the nontrivial portions of Remark 1(a). Remark 1(b) follows from Remark 1(a) upon observing that the results from Doob (1953) and Futia (1982) used in the proof of Corollary 2 apply to Borel subsets of separable and completely metrizable spaces and that the results in Breiman (1960) hold for compact metrizable spaces.

LEMMA A2: *Let Z be a metrizable space and $\mathscr{B}$ its Borel σ-field. Then Assumption 5A′ as well as Assumption 5B′ imply Assumption 5.*

PROOF: For the sets K_m of Assumption 5A′ we have $r_m > 0$ eventually and that

$$\sup_n n^{-1}\sum_{t=1}^{n} P(z_t \notin K_m) \leqslant \sup_n n^{-1}\sum_{t=1}^{n} P(g(z_t) \geqslant r_m)$$
$$\leqslant \sup_n n^{-1}\sum_{t=1}^{n} Eg(z_t)/r_m.$$

This proves the first claim. Under Assumption 5B′, Theorem 8 in Appendix III of Billingsley (1968) implies that $\{H^n\colon n \in \mathbb{N}\}$ is tight. This proves the second claim. *Q.E.D.*

REFERENCES

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

ANDREWS, D. W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

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

BIERENS, H. J. (1981): "Robust Methods and Asymptotic Theory in Nonlinear Econometrics," *Lecture Notes in Economics and Mathematical Systems*, Vol. 192. Berlin: Springer Verlag.

——— (1984): "Model Specification Testing of Time Series Regression," *Journal of Econometrics*, 26, 323–353.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

BREIMAN, L. (1960): "The Strong Law of Large Numbers for a Class of Markov Chains," *Annals of Mathematical Statistics*, 31, 801–803.

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

DOOB, J. L. (1953): *Stochastic Processes*. New York: Wiley.

FUTIA, C. A. (1982): "Invariant Distributions and the Limiting Behavior of Markovian Economic Models," *Econometrica*, 50, 377–408.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments 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, 1977–1991.

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

MCLEISH, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 829–839.

NEWEY, W. K. (1987): "Expository Notes on Uniform Laws of Large Numbers," Princeton University, mimeo.

PARTHASARATHY, K. R. (1967): *Probability Measures on Metric Spaces*. New York: Academic Press.

PÖTSCHER, B. M., AND I. R. PRUCHA (1986a): "A Class of Partially Adaptive One-Step *M*-Estimators for the Nonlinear Regression Model with Dependent Observations," *Journal of Econometrics*, 32, 219–251.

——— (1986b): "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results," Department of Economics, University of Maryland, Working Paper No. 86-9, June, 1986.

WALD, A. (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 20, 595–600.

WHITE, H. (1980): "Nonlinear Regression on Cross-Section Data," *Econometrica*, 48, 721–746.

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