Econometrica, Vol. 59, No. 4 (July, 1991), 1161–1167

UNIFORM CONVERGENCE IN PROBABILITY AND STOCHASTIC EQUICONTINUITY[1]

BY WHITNEY K. NEWEY

1. INTRODUCTION

CONDITIONS FOR UNIFORM CONVERGENCE in probability are useful in econometrics, for showing consistency and asymptotic normality of nonlinear estimators and consistency of standard error estimates. The purpose of this paper is to provide conditions that meet two related requirements: (i) objects other than sample averages are allowed; (ii) only pointwise convergence is assumed. These requirements are motivated by nonparametric and semiparametric estimation problems, which often depend on objects that are more complicated than averages. In addition, for sample averages (i.e. uniform laws of large numbers), requirement (ii) may be useful when the data satisfies complicated dependence restrictions, such as near epoch dependence as in Gallant and White (1988) or geometric ergodicity as in Duffie and Singleton (1989). In these environments it can be easier to check pointwise convergence in probability than to check convergence of various supremums and infimums, as is required by the conditions of Andrews (1987) or Potscher and Prucha (1989). The focus here on convergence in probability, rather than almost sure convergence, is also in keeping with these requirements. For complicated objects it can be more difficult to show almost sure convergence. In any case, convergence in probability is sufficient for showing validity of asymptotic inference procedures (e.g. confidence intervals).

The paper first presents a *stochastic equicontinuity* condition that together with pointwise convergence characterizes uniform convergence in probability to equicontinuous functions on a compact set. Motivation for this condition is given by discussing its relationship to well known results on weak convergence of stochastic processes, e.g. Billingsley (1968) or Pollard (1989).

The remainder of the paper focuses on simple sufficient conditions for stochastic equicontinuity. A global Lipschitz condition is given that, together with pointwise convergence, is sufficient for uniform convergence. When specialized to sample averages, these conditions yield a uniform weak law of large numbers that is complementary to Andrews (1989, Corollary 2). It requires only pointwise convergence of sample averages, rather than convergence of certain supremums and infimums, at the expense of imposing a global, rather than a local, Lipschitz condition. A uniform weak law analogous to that of Potscher and Prucha (1989, Theorem 1) is also given.

The Lipschitz condition is also shown to be useful in two nonparametric examples. It is used to formulate simple sufficient conditions for uniform convergence in probability for U-statistics and for the nonparametric two-stage least squares criterion of Newey and Powell (1989).

2. GENERIC UNIFORM CONVERGENCE IN PROBABILITY

In order to discuss uniform convergence in probability it is necessary to introduce some notation. Let θ be a parameter vector, which can be either finite or infinite dimensional. Let $\hat{Q}_n(\theta)$ be a random function of θ and the sample size n, where explicit dependence on the data will be suppressed for notational convenience. Let $\overline{Q}_n(\theta)$ be a nonrandom function of θ and n, which might be thought of as the object which $\hat{Q}_n(\theta)$ is estimating. For example, $\overline{Q}_n(\theta)$ may be the expectation of $\hat{Q}_n(\theta)$, or the expectation of an

[1] Financial support was provided by the NSF and the Sloan Foundation. Helpful comments were provided by two referees, L. Hansen, P. Perron, J. Powell, and especially D. W. K. Andrews.

analog of $\hat{Q}_n(\theta)$ with preliminary nonparametric estimates replaced by true values. In keeping with the recent econometric literature, e.g. White (1980), $\bar{Q}_n(\theta)$ can depend on sample size to allow for moment drift.

Uniform convergence in probability over a set Θ of parameter values is

$$\sup_{\theta \in \Theta}\left|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)\right| = o_p(1). \tag{2.1}$$

To avoid measurability complications it will be assumed that probability statements such as this are for outer probability.[2]

The following conditions are important for the results to follow:

Assumption 1 (Compactness): *Θ is compact.*

Assumption 2 (Pointwise Convergence): *For each $\theta \in \Theta$, $\hat{Q}_n(\theta) - \bar{Q}_n(\theta) = o_p(1)$.*

It is difficult to do without the compactness assumption, and the pointwise convergence assumption is an obvious necessary condition for uniform convergence. The sense in which the results of this section are generic is that Assumption 2 is taken as a primitive condition. In particular cases Assumption 2 has to be verified, using some law of large numbers or other result appropriate to the form of $\hat{Q}_n(\theta)$.

Another important property is equicontinuity of $\{\bar{Q}_n(\theta)\}_{n=1}^{\infty}$ on Θ. In what follows equicontinuity of $\{\bar{Q}_n(\theta)\}$ may be a hypothesis or conclusion, depending on the specificity of the result.

The following condition generalizes equicontinuity to random functions:

Assumption 3 (Stochastic Equicontinuity): *For every ε, $\eta > 0$ there exists random $\Delta_n(\varepsilon, \eta)$ and constant $n_0(\varepsilon, \eta)$ such that for $n \geqslant n_0(\varepsilon, \eta)$, $Prob(|\Delta_n(\varepsilon, \eta)| > \varepsilon) < \eta$ and for each θ there is an open set $\mathcal{N}(\theta, \varepsilon, \eta)$ containing θ with*

$$\sup_{\tilde{\theta} \in \mathcal{N}(\theta, \varepsilon, \eta)}\left|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)\right| \leqslant \Delta_n(\varepsilon, \eta), \qquad n \geqslant n_0(\varepsilon, \eta). \tag{2.2}$$

It is well known that pointwise convergence and equicontinuity characterize uniform convergence to a continuous function on a compact set; e.g. see Rudin (1976, Exercise 7.16) for sufficiency. The following theorem is a stochastic generalization.

Theorem 2.1: *Suppose Assumption 1 holds and $\{\bar{Q}_n(\theta)\}$ is equicontinuous. Then $\sup_{\theta \in \Theta}|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$ if and only if Assumptions 2 and 3 hold.*

Proofs are given in the Appendix.

Theorem 2.1 is related to well known results on weak convergence (i.e. convergence in distribution) of the stochastic process $\hat{Q}_n(\theta)$, with index θ. For example, suppose $\hat{Q}_n(\theta)$ is continuous, $\bar{Q}_n(\theta)$ does not depend on n, Θ is a compact metric space, and $\hat{Q}_n(\theta)$ is measurable for each θ. Let C be the metric space of continuous functions on Θ, with the uniform metric $d(\tilde{q}, q) = \sup_{\theta \in \Theta}|\tilde{q}(\theta) - q(\theta)|$.[3] In this context, uniform convergence in probability of $\hat{Q}_n(\theta)$ to $\bar{Q}(\theta)$ is convergence in probability as random elements of C. Furthermore, since $\bar{Q}(\theta)$ is nonrandom, this convergence is the same as weak convergence.

[2] The outer probability of an arbitrary set A is defined as $\inf\{E[b]$: b is measurable and $1(A) \leqslant b\}$.

[3] It follows from Θ compact and metrizable that C is separable (see Kelley (1955, p. 245)). For separable C, $\hat{Q}_n(\theta)$ is a random variable in C by $\hat{Q}_n(\bar{\theta})$ measurable at each $\bar{\theta} \in \Theta$ (see Billingsley (1968, p. 57)).

Well known necessary and sufficient conditions for weak convergence in C are convergence of the finite-dimensional distributions and tightness of the sequence of probability measures on C corresponding to $\hat{Q}_n(\theta)$ (e.g. see Billingsley (1968, p. 35)). In the context of convergence in probability on C Assumption 2 is equivalent to convergence of the finite sample distributions. Furthermore, Assumption 3 is closely related to tightness. To be precise, a straightforward generalization of Theorem 8.2 of Billingsley gives the following.

TIGHTNESS CHARACTERIZATION: *$\{\hat{Q}_n(\theta)\}$ is tight in C if and only if $\hat{Q}_n(\theta) = O_p(1)$ for all $\theta \in \Theta$ and Assumption 3 holds.*

Given this characterization, in the context of this example (which includes the important special case where $\hat{Q}_n(\theta)$ is continuous and $\Theta \subset \mathbb{R}^q$), Theorem 2.1 becomes a special case of the tightness characterization of weak convergence.

Tightness plays no direct role in Theorem 2.1, which is stated in terms of stochastic equicontinuity. In this respect Theorem 2.1 is like recent results on weak convergence such as Pollard (1989, Theorem 10.2), which have stochastic equicontinuity conditions rather than tightness as hypotheses. Indeed, when Θ is a metric space with metric $d(\cdot,\cdot)$ (e.g. for $\Theta \subset \mathbb{R}^q$), Theorem 2.1 becomes a special case of Pollard's result if $\Delta_n(\varepsilon, \eta)$ in Assumption 3 is specified as $\Delta_n(\varepsilon, \eta) = \sup_{\tilde{\theta}, \theta \in \Theta;\, d(\tilde{\theta}, \theta) < \delta} |\hat{Q}_n(\tilde{\theta}) - \hat{Q}(\theta)|$ for small enough $\delta > 0$. Also, the resulting form of Assumption 3 is a more familiar definition of stochastic equicontinuity.[4]

The remainder of the paper will focus on sufficient conditions for stochastic equicontinuity, in order to provide easily verifiable conditions for uniform convergence. One useful sufficient condition is the following Lipschitz condition. Henceforth, let h denote a function h: $[0,\infty) \to [0,\infty)$, with $h(0) = 0$ and h continuous at zero (e.g. $h(d) = d^\alpha$, $\alpha > 0$).

ASSUMPTION 3A: *Θ is a metric space and there is B_n and h such that $B_n = O_p(1)$ and for all $\tilde{\theta}$, $\theta \in \Theta$, $|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)| \leqslant B_n h(d(\tilde{\theta}, \theta))$.*

COROLLARY 2.2: *If Assumptions 1, 2, and 3A are satisfied, and $\{\bar{Q}_n(\theta)\}$ is equicontinuous, then $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$.*

Assumption 3A is similar to Andrews (1987, Assumption A4), although it is global rather than local and applies to the function $\hat{Q}_n(\theta)$ rather than individual terms of a sample average. If $\bar{Q}_n(\theta) = E[\hat{Q}_n(\theta)]$ and $E[B_n]$ is bounded, then equicontinuity of $\bar{Q}_n(\theta)$ can be dropped as a hypothesis and included as a conclusion. Also, when θ is a vector of real numbers, Assumption 3A can be replaced by the conditions that Θ is convex, $\hat{Q}_n(\theta)$ continuously differentiable, and the derivative is dominated by $B_n = O_p(1)$.

3. GENERIC WEAK UNIFORM LAWS OF LARGE NUMBERS

Specializing previous results to the case where $\hat{Q}_n(\theta)$ is a sample average gives uniform weak laws. Let the data realization be $(z_1, z_2, \ldots)$, and consider a sequence of functions $q_t(z_t, \theta)$ of a data observation z_t and a parameter vector θ. Define

$$\hat{Q}_n(\theta) = \sum_{t=1}^{n} q_t(z_t, \theta)/n, \qquad \bar{Q}_n(\theta) = \sum_{t=1}^{n} E[q_t(z_t, \theta)]/n.$$

Uniform weak laws concern conditions for equation (2.1) for these objects.

[4] This metric space modification of Theorem 2.1 was pointed out by a referee and D. W. K. Andrews.

One such result is the following corollary.

COROLLARY 3.1: *Suppose that Assumptions* 1 *and* 2 *are satisfied and* Θ *is a metric space. Also suppose there exists* $b_t(z_t)$ *and* h *such that* $\sum_{t=1}^{n} E[b_t(z_t)]/n = O(1)$, *and* $|q_t(z_t, \tilde{\theta}) - q_t(z_t, \theta)| \leqslant b_t(z_t)h(d(\tilde{\theta}, \theta))$ *for* $\tilde{\theta}, \theta \in \Theta$. *Then* $\sup_{\theta \in \Theta}|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$ *and* $\{\bar{Q}_n(\theta)\}$ *is equicontinuous.*

This result is complementary to the convergence in probability version of Andrews (1987, Corollary 2), imposing only pointwise convergence but requiring a global Lipschitz condition. Another uniform weak law is the following corollary.

COROLLARY 3.2: *Suppose that Assumptions* 1 *and* 2 *are satisfied and* Θ *is a metric space. Also, suppose* $z_t \in Z$, Z *a closed subset of* $\mathbb{R}^s$, $\{q_t(z, \theta)\}$ *is equicontinuous on* $Z \times \Theta$, *there are* $b_t(z_t)$ *and* $\gamma > 1$ *such that* $\sup_{\theta \in \Theta}|q_t(z_t, \theta)| \leqslant b_t(z_t)$ *and* $\sum_{t=1}^{n} E[b_t(z_t)^{\gamma}]/n = O(1)$, *and* $\{\sum_{t=1}^{n} H_t/n\}$ *is a tight family for* Z, *where* H_t *is the marginal distribution of* z_t. *Then* $\sup_{\theta \in \Theta}|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$ *and* $\{\bar{Q}_n(\theta)\}$ *is equicontinuous.*

This result is like the convergence in probability version of Potscher and Prucha (1989, Theorem 1), except that only Assumption 2 is required rather than the convergence of various supremums and infimums in Potscher and Prucha (1989, Assumption 4). The tightness condition in Corollary 3.2 is implied by more primitive conditions, such as boundedness of $\sum_{t=1}^{n} E[\ln(1 + \|z_t\|)]/n$; see Potscher and Prucha.

In general, stochastic equicontinuity is implied by the conditions of existing uniform laws of large numbers, since it is necessary for uniform convergence. In some cases it may be implied by a subset of the conditions, which may thus be thought of as stochastic equicontinuity assumptions. For example, if $q_t(z, \theta)$ does not depend on t and $\{z_t\}$ is stationary and ergodic, Assumption 3 is implied by the first moment continuity condition of Hansen's (1982) uniform law.

4. NONPARAMETRIC EXAMPLES

A uniform convergence result for U-statistics is useful for showing consistency of the residual-based m-estimators for nonlinear simultaneous equations models discussed in Newey (1989). Let $m(z, \tilde{z}, \theta)$ be a function of a pair of data arguments that is symmetric in the data arguments, i.e. $m(z, \tilde{z}, \theta) = m(\tilde{z}, z, \theta)$. Consider a U-statistic, depending on θ, and population analog

$$\hat{Q}_n(\theta) \equiv 2 \sum_{t=1}^{n} \sum_{s>t} m(z_t, z_s, \theta)/n(n-1), \qquad \bar{Q}(\theta) \equiv E[m(z_1, z_2, \theta)], \tag{4.1}$$

where z_t is assumed i.i.d. Results on convergence of $\hat{Q}_n(\theta)$ to $\bar{Q}(\theta)$ for fixed θ are well known, e.g. Serfling (1980). Such results can easily be turned into uniform convergence results via Corollary 2.3. An example is as follows:

COROLLARY 4.1: *Suppose that Assumption* 1 *is satisfied,* Θ *is a metric space, and* z_t, $(t = 1, 2, \ldots)$ *are i.i.d.. Also suppose* $E[|m(z_1, z_2, \theta_0)|] < \infty$ *for some* $\theta_0 \in \Theta$, *and there are* $b(z, \tilde{z})$ *and* h *such that* $E[b(z_1, z_2)] < \infty$ *and for* $\tilde{\theta}, \theta \in \Theta$, $|m(z, \tilde{z}, \tilde{\theta}) - m(z, \tilde{z}, \theta)| \leqslant b(z, \tilde{z})h(d(\tilde{\theta}, \theta))$. *Then* $\sup_{\theta \in \Theta}|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$ *and* $\bar{Q}(\theta)$ *is continuous.*

A second example is the nonparametric two-stage least squares criterion of Newey and Powell (1989). Let $\rho(z, \theta)$ be a function of a data observation and parameters, satisfying the conditional moment restriction $E[\rho(z_t, \theta_0)|x_t] = 0$ at the true parameter values θ_0, where the instruments x_t are a subvector of z_t. Also, let $\hat{E}[\rho(z,\theta)|x_t]$ be a nonparametric regression estimator of $E[\rho(z,\theta)|x]$, evaluated at x_t. Consider

$$(4.2)\qquad \hat{Q}_n(\theta) \equiv \sum_{t=1}^{n} \{\hat{E}[\rho(z,\theta)|x_t]\}^2/n, \qquad \overline{Q}(\theta) = E\left[\{E[\rho(z_t,\theta)|x_t]\}^2\right].$$

Corollary 2.3 can be used to establish uniform convergence in probability of $\hat{Q}_n(\theta)$ to $\overline{Q}(\theta)$, and hence to show consistency of an estimator of θ_0 obtained by minimizing $\hat{Q}_n(\theta)$ over Θ. To be concrete, consider a series estimator, $\hat{E}[\rho(z,\theta)|x_t] = P'_{tK}(P'P)^- P'\rho(\theta)$, where $P'_{tK} = (p_{1K}(x_t),\ldots,p_{KK}(x_t))$, for approximating functions $\{p_{kK}(x)\}$ $(K=1,2,\ldots)$ (e.g. power or Fourier series, or B-splines), $P=[P_{1K},\ldots,P_{nK}]'$, $(\cdot)^-$ denotes a generalized inverse, and $\rho(\theta) = (\rho(z_1,\theta),\ldots,\rho(z_n,\theta))'$.

Corollary 4.2: *Suppose that Assumption* 1 *is satisfied*, Θ *is a metric space, and* z_t $(t=1,2,\ldots)$ *are i.i.d. Also suppose that* (*i*) $Var[\rho(z,\theta)|x]$ *is bounded for each* $\theta \in \Theta$, *for any* $g(x)$ *with* $E[g(x_t)^2] < \infty$ *there exists* $\{\pi_K\}$ *such that* $\lim_{K\to\infty} E[\{g(x_t) - P'_{tK}\pi_K\}^2] = 0$, *and* $K = K(n)$ *such that* $K(n) \to \infty$ *and* $K(n)/n \to 0$; (*ii*) *there is* $b(z)$ *and continuous* h *with* $|\rho(z,\tilde{\theta}) - \rho(z,\theta)| \leqslant b(z)h(d(\tilde{\theta},\theta))$ *and* $E[b(z_1)^2] < \infty$. *Then* $\sup_{\theta\in\Theta}|\hat{Q}_n(\theta) - \overline{Q}_n(\theta)| = o_p(1)$ *and* $\overline{Q}(\theta)$ *is continuous.*

Condition (i) and (ii) are used to establish Assumptions 2 and 3A, respectively. A more general result, that allows for other types of nonparametric regression estimators (e.g. nearest neighbor), is given in Newey and Powell (1989).

Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Manuscript received April, 1989; final revision received August, 1990.

APPENDIX

Lemma A.1: *Define* $\hat{R}_n(\theta) \equiv \hat{Q}_n(\theta) - \overline{Q}_n(\theta)$. *If* $\{\overline{Q}_n(\theta)\}$ *is equicontinuous and* Θ *compact, then Assumption* 3 *is satisfied if and only if it is satisfied with* $\hat{Q}_n(\theta)$ *replaced by* $\hat{R}_n(\theta)$.

Proof: The proof is a consequence of the simple triangle inequalities

$$(A.1)\qquad |(a-b)-(c-d)| \leqslant |a-c| + |b-d|, \qquad |a-c| \leqslant |(a-b)-(c-d)| + |b-d|,$$

for $a = \hat{Q}_n(\tilde{\theta})$, $b = \overline{Q}_n(\tilde{\theta})$, $c = \hat{Q}_n(\theta)$, and $d = \overline{Q}_n(\theta)$: For the if part for Assumption 3, consider ε, $\eta > 0$. Let $\Delta_n(\varepsilon/2,\eta)$ and $\mathscr{N}(\theta,\varepsilon/2,\eta)$ be as in Assumption 3, and let $\mathscr{N}'(\theta,\varepsilon/2,\eta) \subseteq \mathscr{N}(\theta,\varepsilon/2,\eta)$ be an open set such that $\sup_{n,\tilde{\theta}\in\mathscr{N}'}|\overline{Q}_n(\tilde{\theta}) - \overline{Q}_n(\theta)| < \varepsilon/2$. Then for $\Delta'_n(\varepsilon,\eta) = \Delta_n(\varepsilon/2,\eta) + \varepsilon/2$, $\Pr(\Delta'_n > \varepsilon) = \Pr(\Delta_n > \varepsilon/2) < \eta$, while by the first half of equation (A.1), for $n \geqslant n_0(\varepsilon/2,\eta)$, $\sup_{\tilde{\theta}\in\mathscr{N}'}|\hat{R}_n(\tilde{\theta}) - \hat{R}_n(\theta)| \leqslant \Delta_n(\varepsilon/2,\eta) + \varepsilon/2 = \Delta'_n(\varepsilon,\eta)$. The "only if" part for Assumption 3 follows by the second half of equation (A.1) by the same argument, interchanging $\hat{R}_n$ and $\hat{Q}_n$. *Q.E.D.*

Proof of Theorem 2.1: By Lemma A.1 it suffices to show the result with $\hat{R}_n(\theta)$ replacing $\hat{Q}_n(\theta)$ in Assumption 3. For the if part, consider ε, $\eta > 0$, $\Delta_n(\varepsilon/2,\eta)$, and $\mathscr{N}(\theta,\varepsilon/2,\eta)$. By Θ compact there is an open subcovering $\{\mathscr{N}(\theta_j)\}_{j=1}^{J}$ of such $\mathscr{N}(\theta,\varepsilon/2,\eta)$. Then by the triangle inequality,

$$\sup_{\theta\in\Theta}\left|\hat{R}_n(\theta)\right| \leqslant \max_j\left|\hat{R}_n(\theta_j)\right| + \sup_{j,\theta\in\mathscr{N}(\theta_j)}\left|\hat{R}_n(\theta) - \hat{R}_n(\theta_j)\right| \leqslant o_p(1) + \Delta_n(\varepsilon/2,\eta),$$

where the second inequality follows by Assumptions 2 and 3. Thus,

$$\Pr\left(\sup_{\theta\in\Theta}\left|\hat{R}_n(\theta)\right| > \varepsilon\right) \leqslant \Pr\left(\Delta_n(\varepsilon/2,\eta) > \varepsilon/2\right) + \Pr\left(o_p(1) > \varepsilon/2\right) < \eta$$

for large enough n. For the only if part, note that Assumption 2 is a trivial consequence of uniform convergence. Also, $\Delta_n(\varepsilon,\eta) \equiv 2\sup_{\theta\in\Theta}|\hat{R}_n(\theta)| = o_p(1)$ and for any θ and open neighborhood $\mathscr{N}$, $\sup_{\theta\in\mathscr{N}}|\hat{R}_n(\tilde{\theta}) - \hat{R}_n(\theta)| \leqslant \Delta_n(\varepsilon,\eta)$, giving Assumption 3. *Q.E.D.*

PROOF OF TIGHTNESS CHARACTERIZATION: Omitted for brevity, but available from author upon request.

PROOF OF COROLLARY 2.2: By Theorem 2.1 it suffices to show Assumption 3. Consider ε, $\eta > 0$. By $B_n = O_p(1)$, there is M such that for all n, $\Pr(B_n > \varepsilon M) < \eta$, so the first condition of Assumption 3 is satisfied for $\Delta_n(\varepsilon,\eta) = B_n/M$. Choose δ small enough that $h(d) < 1/M$ for all $0 \leqslant d < \delta$ and let $\mathscr{N}(\theta,\varepsilon,\eta) = \{\tilde{\theta}\in\Theta: d(\tilde{\theta},\theta) < \delta\}$. Then $\sup_{\mathscr{N}}|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)| \leqslant B_n \sup_{0\leqslant d<\delta} h(d) \leqslant \Delta_n(\varepsilon,\eta)$, giving the second condition of Assumption 3. *Q.E.D.*

PROOF OF COROLLARY 3.1: Let $B_n = \Sigma_{t=1}^n b_t(z_t)/n$. By hypothesis, $E[B_n] = O(1)$ and $|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)| \leqslant \Sigma_{t=1}^n |q_t(z_t,\tilde{\theta}) - q_t(z_t,\theta)|/n \leqslant B_n h(d(\tilde{\theta},\theta))$, so that the conclusion follows by Corollary 2.2 and the remarks that follow it. *Q.E.D.*

PROOF OF COROLLARY 3.2: Let $D_n(K) \equiv \Sigma_{t=1}^n 1(z_t \notin K) b_t(z_t)/n$ for a compact set $K \subseteq Z$. By the Holder inequality,

$$E[D_n(K)] \leqslant \sum_{t=1}^{n} \{\Pr(z_t \notin K)\}^{(\gamma-1)/\gamma}\{E[b_t(z_t)^\gamma]\}^{1/\gamma}/n \tag{A.2}$$
$$\leqslant \left\{\sum_{t=1}^{n} \Pr(z_t\notin K)/n\right\}^{(\gamma-1)/\gamma}\left\{\sum_{t=1}^{n} E[b_t(z_t)^\gamma]/n\right\}^{1/\gamma}$$

Consider ε, $\eta > 0$. By hypothesis $\Sigma_{t=1}^n E[b_t(z_t)^\gamma]/n$ is bounded and there exists K such that $\Sigma_{t=1}^n \Pr(z_t \notin K)/n$ is arbitrarily small for all n, so that K can be chosen so that $E[D_n(K)] < \varepsilon\eta/4$ for all n. Also, as in Potscher and Prucha (1989), $q_t(z,\theta)$ is continuous on $K\times\Theta$ uniformly in t for each compact K, so for any θ there exists $\mathscr{N}$ such that $\sup_{t,(z,\tilde{\theta})\in K\times\mathscr{N}}|q_t(z,\tilde{\theta}) - q_t(z,\theta)| < \varepsilon/2$, implying

$$\sup_{\tilde{\theta}\in\mathscr{N}}|q_t(z,\tilde{\theta}) - q_t(z,\theta)| < \varepsilon/2 + 2\cdot 1(z\notin K)b_t(z) \qquad (t = 1,2,\dots). \tag{A.3}$$

Let $\Delta_n(\varepsilon,\eta) = \varepsilon/2 + 2D_n(K)$. By equation (A.3) and the triangle inequality, $\sup_{\tilde{\theta}\in\mathscr{N}}|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)| < \Delta_n(\varepsilon,\eta)$. Also,

$$\Pr(\Delta_n(\varepsilon,\eta) > \varepsilon) = \Pr(2D_n(K) > \varepsilon/2) \leqslant E[D_n(K)]/(\varepsilon/4) < \eta,$$

giving Assumption 3. Furthermore,

$$\sup_{\tilde{\theta}\in\mathscr{N}}|\bar{Q}_n(\tilde{\theta}) - \bar{Q}_n(\theta)| = \sup_{\tilde{\theta}\in\mathscr{N}}\left|E\left[\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)\right]\right|$$
$$\leqslant E\left[\sup_{\tilde{\theta}\in\mathscr{N}}|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)|\right] \leqslant E[\Delta_n(\varepsilon,\eta)] < \eta,$$

so $\{\bar{Q}_n(\theta)\}$ is equicontinuous. Theorem 2.1 then gives the conclusion. *Q.E.D.*

PROOF OF COROLLARY 4.1: Assumption 2 is satisfied by Theorem A of Serfling (1980). By the i.i.d. assumption, $E[\hat{Q}_n(\theta)] = \bar{Q}(\theta)$. Also, for $B_n \equiv 2\Sigma_{t=1}^n\Sigma_{s>t} b(z_s, z_t)/n(n-1)$, $E[B_n] = E[d(z_1,z_2)] < \infty$, and

$$\left|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)\right| \leqslant 2\sum_{t=1}^{n}\sum_{s>t}\left|m(z_s,z_t,\tilde{\theta}) - m(z_s,z_t,\theta)\right|/n(n-1) \leqslant B_n h(d(\tilde{\theta},\theta)),$$

so that the conclusion follows by Corollary 2.2 and its following remarks. *Q.E.D.*

PROOF OF COROLLARY 4.2: To show that Assumption 2 holds, let $y = \rho(\theta)$, $g = (E[\rho(z,\theta)|x_1],\dots,E[\rho(z,\theta)|x_n])'$, and $W = P(P'P)^- P'$. Note that

$$\left|\hat{Q}_n(\theta) - (g'g/n)\right| = \left|\|Wy\|^2 - \|g\|^2\right|/n \leqslant \left(\|Wy - g\|^2 + 2\|g\|\cdot\|Wy - g\|\right)/n,$$

and that $\|g\|^2/n = O_p(1)$ by the Markov inequality, so that to show $|\hat{Q}_n(\theta) - (g'g/n)| = o_p(1)$ it suffices to show $\|Wy - g\|^2/n = o_p(1)$. Next for $X = (x_1, \ldots, x_n)'$, by $WP = P$, W idempotent with rank (= trace) no bigger than K, and condition (i),

$$\begin{aligned}
E\left[\|Wy - g\|^2\right] &= E\left[E\left[\|Wy - g\|^2|X\right]\right] = E\left[E\left[\|W(y-g) - (I-W)g\|^2|X\right]\right] \\
&= E\left[\operatorname{trace}\{E[W(y-g)(y-g)'W|X]\} + g'(I-W)g\right] \\
&= E\left[\operatorname{trace}\{W \operatorname{Var}(y|X)W\} + \{g - P\pi_K\}'(I-W)\{g - P\pi_K\}\right] \\
&\leqslant \sup_x \operatorname{var}(\rho(z,\theta)|x) E[\operatorname{trace}(W)] \\
&\quad + nE\left[\{g(x_t) - P_{tK}'\pi_K\}^2\right] = O(K) + o(n).
\end{aligned}$$

Thus, $\|Wy - g\|^2/n = o_p(1)$ follows by the Markov inequality. Then since $g'g/n = \bar{Q}(\theta) + o_p(1)$ by the weak law of large numbers, the triangle inequality gives Assumption 2. To show Assumption 4A, let $b = (b(z_1), \ldots, b(z_n))'$ and $\tilde{B}_n = (\|b\|^2 + 2\|\rho(\theta_0)\| \cdot \|b\|)/n$. Note (i) and (ii) imply $\|b\|^2/n = O_p(1)$ and $\|\rho(\theta_0)\|^2/n = O_p(1)$, so $\tilde{B}_n = O_p(1)$. Also, by W idempotent (implying $\|Wa\| \leqslant \|a\|$ for conformable a), compactness, and continuity of $h(\cdot)$ (implying $h(d(\tilde{\theta},\theta))$ is bounded on $\Theta \times \Theta$), there is a constant C such that

$$\begin{aligned}
\left|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)\right| &= \left|\left[\left\|W\rho(\tilde{\theta})\right\|^2 - \|W\rho(\theta)\|^2\right|/n\right. \\
&\leqslant \left(\left\|\rho(\tilde{\theta}) - \rho(\theta)\right\|^2 + 2\|\rho(\theta)\| \cdot \left\|\rho(\tilde{\theta}) - \rho(\theta)\right\|\right)/n \\
&\leqslant \left\{\|b\|^2 h(d(\tilde{\theta},\theta)) + 2\|b\|^2 h(d(\theta,\theta_0)) + 2\|\rho(\theta_0)\| \cdot \|b\|\right\} \\
&\quad \times h(d(\tilde{\theta},\theta))/n \\
&\leqslant B_n h(d(\tilde{\theta},\theta)),
\end{aligned}$$

with $B_n = C\tilde{B}_n$ for some constant C. Assumption 4A follows by $B_n = C \cdot O_p(1)$. Continuity of $\bar{Q}(\theta)$ follows by a similar argument, which gives

$$\begin{aligned}
\left|\bar{Q}(\tilde{\theta}) - \bar{Q}(\theta)\right| &\leqslant E\left[\left|E\left[\rho(z,\tilde{\theta})\middle|x\right]^2 - E[\rho(z,\theta)|x]^2\right|\right] \\
&\leqslant CE\left[b(z)^2 + 2|\rho(z,\theta_0)|\,|b(z)|\right] h(d(\tilde{\theta},\theta)). \qquad Q.E.D.
\end{aligned}$$

REFERENCES

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

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

DUFFIE, D., AND K. J. SINGLETON (1989): "Simulated Moments Estimation of Markov Models of Asset Prices," mimeo, Graduate School of Business, Stanford University.

GALLANT, A. R., AND H. WHITE (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil-Blackwell.

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

KELLEY, J. L. (1955): *General Topology*. New York: Springer-Verlag.

NEWEY, W. K., (1989): "Locally Efficient, Residual Based Estimation of Nonlinear Simultaneous Equations Models," mimeo, Princeton University.

NEWEY, W. K., AND J. L. POWELL (1989): "Nonparametric Instrumental Variables Estimation," mimeo, University of Wisconsin.

POLLARD, D. (1989): *Empirical Processes: Theory and Applications*. Philadelphia: CBMS/NSF Regional Conference Series.

POTSCHER, B. M., AND I. R. PRUCHA (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Processes," *Econometrica*, 57, 675–683.

RUDIN, W. (1976): *Principles of Mathematical Analysis*. New York: McGraw-Hill.

SERFLING (1980): *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.