

UNIFORM CONVERGENCE IN PROBABILITY
AND STOCHASTIC EQUICONTINUITY

Whitney K. Newey
Princeton University
Bell Communications Research

Econometric Research Program
Research Memorandum No. 342

March 1989

Financial support was provided by the NSF and the Sloan Foundation.
J. L. Powell provided helpful comments.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544-1017, USA

1. Introduction

Uniform convergence in probability is an important concept in econometric theory. It is useful for showing consistency and asymptotic normality of estimators. It is also important for large sample inference, being useful for showing that asymptotic standard error estimates are consistent.

The purpose of this paper is to provide uniform convergence results that meet two, related requirements. The first is that the result apply to objects other than sample averages. The second is that minimal pointwise convergence in probability conditions are imposed. Both of these requirements are motivated by the need for results that apply to certain nonparametric and semiparametric models. Estimators for such models often involve objects that are much more complicated than sample averages, such as preliminary nonparametric regression estimators. To show consistency of such estimators it is useful to have uniform convergence results that apply to these objects. Furthermore, for complicated objects it is helpful to keep the convergence in probability requirement to the minimum of pointwise convergence. In addition, even for sample averages, when the data satisfies complicated dependence restrictions it may be easier to check pointwise convergence in probability, rather than the convergence of various supremums and infimums as in Andrews (1989) and Potscher and Prucha (1989). Examples will be discussed below.

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 asymptotic normality arguments and the construction of asymptotic confidence intervals, so that if only asymptotic inference is a concern, then convergence in probability is all that is needed.

The paper presents a condition, referred to as uniform stochastic equicontinuity, that together with pointwise convergence characterizes uniform convergence to equicontinuous functions on a compact set. Also, an easily interpretable global Lipschitz condition is shown to be sufficient for uniform stochastic equicontinuity.

For the special case of a sample average it is possible to formulate uniform weak laws of large numbers that are complementary to those of Andrews (1989) and Potscher and Prucha (1989). In comparison with Andrews (1989), the result here require only pointwise convergence of the sample averages, rather than convergence of certain supremums and infimums, at the expense of imposing global, rather than local, Lipschitz conditions. In comparison with Potscher and Prucha (1989), only pointwise convergence in probability is required, under no additional conditions other than those given by Potscher and Prucha (1989).

Section 2 gives the general uniform convergence results. Section 3 presents generic weak uniform laws of large numbers complementary to those of Andrews (1989) and Potscher and Prucha (1989). Section 4 discusses two nonparametric examples. A uniform law of large numbers for U-statistics is given, and application of the results to the nonparametric instrumental variables environment of Newey and Powell (1989) is discussed.

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 $\bar{Q}_n(\theta)$ be a function of θ and n , which should be thought of as the limit of $\hat{Q}_n(\theta)$. For example, in some environments $\bar{Q}_n(\theta)$ can be taken to be the expectation of $\hat{Q}_n(\theta)$. In other situations $\bar{Q}_n(\theta)$ will be the expectation of an analog of $\hat{Q}_n(\theta)$ with preliminary nonparametric estimates replaced by their true values. $\bar{Q}_n(\theta)$ is allowed to depend on the sample size to allow for nonstationarities, in keeping with the recent econometric literature, e.g. White (1980).

Uniform convergence in probability over a set Θ of parameter values occurs when

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

To avoid measurability issues, it will be assumed that $\hat{Q}_n(\theta)$ is continuous. A related property that is also useful is

$$(2.2) \quad \{\bar{Q}_n(\theta)\}_{n=1}^{\infty} \text{ is equicontinuous on } \Theta.$$

In what follows equicontinuity of $\bar{Q}_n(\theta)$ may be a hypothesis or conclusion, depending on the specificity of the result.

The results make use of the following assumptions:

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 would have to be verified, using some law of large numbers or other result appropriate to the form of $\hat{Q}_n(\theta)$.

The condition that, together with pointwise convergence, characterizes uniform convergence to equicontinuous functions on a compact set is:

Assumption 3: (Uniform Stochastic Equicontinuity) For every $\varepsilon > 0$ there is $\Delta_n(\varepsilon)$ such that $\limsup_{n \rightarrow \infty} \text{Prob}(|\Delta_n(\varepsilon)| > \varepsilon) < \varepsilon$, and for each θ there is an open set $\mathcal{N}(\theta, \varepsilon)$ containing θ such that

$$(2.3) \quad \lim_{n \rightarrow \infty} \text{Prob}(\sup_{\theta' \in \mathcal{N}(\theta, \varepsilon)} |\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| \leq \Delta_n(\varepsilon)) = 1.$$

The word uniform refers to the fact that $\Delta_n(\varepsilon)$ does not depend on θ , and the word equicontinuity to the fact that the neighborhood $\mathcal{N}(\theta, \varepsilon)$ does not depend on sample size. If $\Delta_n(\varepsilon)$ were allowed to depend on θ , then this condition would be equivalent to stochastic equicontinuity at each θ , in the sense of Pollard (1985). In the special case where $\hat{Q}_n(\theta)$ is nonstochastic and $\Delta_n(\varepsilon) = \varepsilon$, this assumption is equicontinuity of the sequence $\{\hat{Q}_n(\theta)\}$. It is well known that pointwise convergence and equicontinuity characterizes convergence to a continuous function on a compact set; e.g. see Rudin (1976, Exercise 7.16) for sufficiency. Generalizing to the stochastic case gives

Theorem 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.

The proofs are given in the Appendix.

Although uniform stochastic equicontinuity may appear to be a complicated condition, it is easy to see that it is implied by a global, stochastic Lipschitz condition.

Corollary 2.1: Suppose Assumptions 1 and 2 are satisfied and $\bar{Q}_n(\theta)$ is equicontinuous. Also suppose that Θ is a metric space with metric $d(\theta, \theta')$ and there exists B_n such that for all $\theta, \theta' \in \Theta$, $|\hat{Q}_n(\theta) - \hat{Q}_n(\theta')| \leq B_n d(\theta, \theta')$ and $B_n = o_p(1)$. Then $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$.

The Lipschitz condition of Theorem 2 is similar to that of Andrews (1989), although here it is global rather than local and applied to the whole function $\hat{Q}_n(\theta)$ rather than one component of a sample average. It is easy to check that 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, it is easy to see that when θ is a vector of real numbers, this Lipschitz condition is implied by Θ convex, $\hat{Q}_n(\theta)$ continuously differentiable, and norm of the derivative dominated by B_n .

3. Generic Weak Uniform Laws of Large Numbers

In this section the theorems of Section 1 are applied to give uniform laws of large numbers. Suppose that the data is a realization of some stochastic process z_1, z_2, \dots , and consider a sequence of functions $q_t(z_t, \theta)$ of a data observation z_t and a parameter vector θ . Define

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

Uniform weak laws of large numbers concern conditions under which $\hat{Q}_n(\theta) - \bar{Q}_n(\theta)$ converges uniformly in probability to zero.

The first result is like Andrews (1989) generic uniform law.

Corollary 3.1: Suppose that Assumptions 1 and 2 are satisfied and Θ is a metric space with metric $d(\theta, \theta')$. Also suppose there exists $d_t(z_t)$ such that for all $\theta, \theta' \in \Theta$, $|q_t(z_t, \theta) - q_t(z_t, \theta')| \leq d_t(z_t)d(\theta, \theta')$ and $\sum_{t=1}^n E[d_t(z_t)]/n$ is bounded. Then $\bar{Q}_n(\theta)$ is equicontinuous and $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$.

This result differs from the convergence in probability version of Andrews (1989) in that only pointwise convergence of the sample average is assumed, rather than convergence of the average of supremum and infimum of $q_t(z_t, \theta)$ over small sets to the population average. This change may make the conditions somewhat easier to check, particularly in situations where weak dependence conditions are imposed.

The second result is like the generic law of Potscher and Prucha (1989).

Corollary 3.2: Suppose that Assumptions 1 and 2 are satisfied and $\Theta \subset \mathbb{R}^k$. Also suppose $q_t(z, \theta)$ is equicontinuous on $Z \times \Theta$, there is $\gamma > 1$ and $d_t(z_t)$ such that $\sup_{\theta \in \Theta} |q_t(z_t, \theta)| \leq d_t(z_t)$ and $\sum_{t=1}^n E[d_t(z_t)^\gamma]/n = O(1)$, and there exists an increasing sequence of compact sets K_j such that $\cup_{j=1}^\infty K_j = \mathbb{R}^S$ and $\lim_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{t=1}^n \text{Prob}(z_t \notin K_j)/n = 0$. Then $\bar{Q}_n(\theta)$ is equicontinuous and $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_p(1)$.

See Potscher and Prucha (1989) for more primitive conditions that imply the condition involving the compact sets K_j . This result differs from the convergence in probability version of their result in that only pointwise convergence of the sample average is assumed. This modification does not require strengthening the other conditions. However, it is important to be reminded that convergence in probability is under consideration here, and that both Potscher and Prucha (1989) and Andrews (1989) also consider

almost sure convergence.

Since uniform stochastic equicontinuity is implied by uniform convergence and equicontinuity, Assumption 3 is implied by the conditions of Andrews (1989), as well as the usual uniform law of large numbers for stationary, ergodic environments. A direct proof that Assumption 3, with $\Delta_n(\varepsilon) = \varepsilon$, is implied by these other conditions is also straightforward.

4. Examples

The first example is a uniform convergence in probability result for U-statistics. Such a result is useful in showing consistency of the residual-based method of moments estimator for nonlinear simultaneous equations models developed in Newey (1988). For notational simplicity we restrict attention to U-statistics of order 2. Let $m(z, \tilde{z}, \theta)$ be a function of a pair of data arguments that is symmetric in the data arguments, i.e. $m(z, \tilde{z}, \theta) = m(\tilde{z}, z, \theta)$. A U-statistic, depending on the parameters θ , takes the form

$$(4.1) \quad \hat{Q}_n(\theta) = 2 \sum_{t=1}^n \sum_{s>t} m(z_t, z_s, \theta) / n(n-1)$$

Results on the pointwise convergence of $2 \sum_{t=1}^n \sum_{s>t} m(z_t, z_s, \theta) / n(n-1)$ to $\bar{Q}(\theta) = E[m(z_t, z_s, \theta)]$ are well known, e.g. Serfling (1980). Such results can easily be turned into uniform convergence results via Corollary 2.1. For example:

Corollary 4.1: Suppose that Assumption 1 is satisfied and that z_t , $(t=1,2,\dots)$ are i.i.d.. Also suppose that $E[|m(z_1, z_2, \theta)|] < \infty$ for all θ and there is a function $d(z, \tilde{z})$ such that for $\theta, \theta' \in \Theta$, $|m(z, \tilde{z}, \theta) - m(z, \tilde{z}, \theta')| \leq d(z, \tilde{z}) \|\theta - \theta'\|$, and $E[d(z_1, z_2)] < \infty$. Then $E[m(z_1, z_2, \theta)]$ is continuous and

$$\sup_{\theta \in \Theta} |2 \sum_{t=1}^n \sum_{s>t} m(z_s, z_t, \theta) / n(n-1) - E[m(z_1, z_2, \theta)]| = o_p(1).$$

A second, more specific example is the nonparametric two-stage least squares estimation problem of Newey and Powell (1989). In this example θ is interpreted as function, appearing in the equation

$$y_t = \theta_0(x_t) + \varepsilon_t, \quad E[\varepsilon_t | z_t] = 0.$$

The variables y_t and x_t are the endogenous variables, and z_t is an instrumental variable. The nonparametric two-stage least squares estimate $\hat{\theta}$ of θ_0 is obtained by minimizing $\hat{Q}_n(\theta)$, where

$$\hat{Q}_n(\theta) = \sum_{t=1}^n (y_t - \hat{E}[\theta | z_t])^2 / n,$$

and $\hat{E}[\theta | z]$ is an estimate of the conditional expectation of θ given z .

To show consistency of $\hat{\theta}$ it is quite useful to show uniform convergence of $\hat{Q}_n(\theta)$ to $\bar{Q}(\theta) = E[(y_t - E[\theta | z_t])^2]$. It is difficult to see how one would apply results from previous papers, but it is straightforward to apply Corollary 2.1. For pointwise convergence, existence of second moments of y_t and $E[\theta | z_t]$, a law of large numbers, and $\sum_{t=1}^n (\hat{E}[\theta | z_t] - E[\theta | z_t])^2 / n = o_p(1)$ for each θ suffices. Also for $d(\theta, \theta') = \sup_x |\theta(x) - \theta'(x)|$, it is easy to check that the Lipschitz condition is satisfied if $\sum_{t=1}^n (\hat{E}[\theta | z_t])^2 / n \leq \sup_x |\theta(x)|^2$; see Newey and Powell (1989) for details.

Appendix

Proof of Theorem 1: First sufficiency of Assumptions 2 and 3 will be shown. Pick $\varepsilon > 0$. By equicontinuity of $\bar{Q}_n(\theta)$, for every θ there exists a neighborhood $\mathcal{N}(\theta, \varepsilon)$ of θ , contained in the neighborhood of Assumption 3, such that for all n , $\sup_{\theta' \in \mathcal{N}(\theta, \varepsilon)} |\bar{Q}_n(\theta') - \bar{Q}_n(\theta)| \leq \varepsilon$. Then by the triangle in equality and Assumptions 2 and 3,

$$(A.1) \quad \begin{aligned} \sup_{\theta' \in \mathcal{N}(\theta, \varepsilon)} |\hat{Q}_n(\theta') - \bar{Q}_n(\theta')| &\leq \sup_{\theta' \in \mathcal{N}(\theta, \varepsilon)} |\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| \\ &+ |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| + \sup_{\theta' \in \mathcal{N}(\theta, \varepsilon)} |\bar{Q}_n(\theta) - \bar{Q}_n(\theta')| \leq \varepsilon + \Delta_n(\varepsilon) + o_p(1). \end{aligned}$$

Since $\cup_{\theta \in \Theta} \mathcal{N}(\theta, \varepsilon)$ is an open covering of Θ and Θ is compact, there is a finite subcovering $\cup_{j=1}^J \mathcal{N}(\theta_j, \varepsilon)$. Thus, by eq. (A.1),

$$(A.2) \quad \begin{aligned} \sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| &= \max_j \sup_{\theta \in \mathcal{N}(\theta_j, \varepsilon)} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| \\ &\leq \max_j \{\varepsilon + |\Delta_n(\varepsilon)| + o_p(1)_j\} \leq \varepsilon + |\Delta_n(\varepsilon)| + o_p(1). \end{aligned}$$

Consider $\delta > 0$ and choose $\varepsilon \leq \delta/3$. Note that $\text{Prob}(|\Delta_n(\varepsilon)| > \delta/3) \leq \text{Prob}(|\Delta_n(\varepsilon)| > \varepsilon)$. Then by eq. (A.2),

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \text{Prob}(\sup_{\theta \in \Theta} |Q_n(\theta) - \bar{Q}_n(\theta)| > \delta) \\ &\leq \text{Prob}(\varepsilon > \delta/3) + \text{Prob}(|\Delta_n(\varepsilon)| > \delta/3) + \text{Prob}(|o_p(1)| > \delta/3) \\ &\leq 0 + \text{Prob}(|\Delta_n(\varepsilon)| > \varepsilon) + o(1) \leq \varepsilon + o(1) < \delta + o(1). \end{aligned}$$

Sufficiency follows since δ can be arbitrarily small. Turning to necessity, note that Assumption 2 is implied by uniform convergence. Suppose that Assumption 3 is not satisfied. Then for some ε there is no Δ_n satisfying the hypotheses, so that for $\Delta_n = \varepsilon$, there exists some θ such that for all

neighborhoods \mathcal{N} of θ , $\text{Prob}(\sup_{\theta' \in \mathcal{N}} |\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| > \varepsilon)$ does not converge to zero. Choose \mathcal{N} such that $\sup_{\theta' \in \mathcal{N}} |\bar{Q}_n(\theta') - \bar{Q}_n(\theta)| < \varepsilon/3$ for all n large enough, so that for such n ,

$$\begin{aligned} & \text{Prob}(\sup_{\theta' \in \mathcal{N}} |\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| > \varepsilon) \\ & \leq \text{Prob}(\sup_{\theta' \in \mathcal{N}} |\hat{Q}_n(\theta') - \bar{Q}_n(\theta')| > \varepsilon/3) + \text{Prob}(|\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| > \varepsilon/3). \end{aligned}$$

Then since the second term following the inequality converges to zero, it follows that $\text{Prob}(\sup_{\theta' \in \mathcal{N}} |\hat{Q}_n(\theta') - \bar{Q}_n(\theta')| > \varepsilon/3)$ does not converge to zero. Thus, convergence is not uniform over \mathcal{N} , and consequently not over Θ . ■

Proof of Corollary 2.1: Consider $\varepsilon > 0$. Choose $\delta(\varepsilon)$ so that $\limsup_{n \rightarrow \infty} \text{Prob}(B_n > \varepsilon/\delta(\varepsilon)) < \varepsilon$, and take $\mathcal{N}(\theta, \varepsilon)$ to be an open ball centered at θ with radius $\delta(\varepsilon)$. Then note that Assumption 3 is satisfied for $\Delta_n(\varepsilon) = B_n \delta(\varepsilon)$, so the conclusion follows by Theorem 1. ■

Proof of Corollary 3.1: By the triangle inequality,

$$\begin{aligned} |\hat{Q}_n(\theta) - \hat{Q}_n(\theta')| & \leq \sum_{t=1}^n |q_t(z_t, \theta) - q_t(z_t, \theta')|/n \leq [\sum_{t=1}^n d_t(z_t)/n] \|\theta - \theta'\|, \\ |\bar{Q}_n(\theta) - \bar{Q}_n(\theta')| & \leq E[|\hat{Q}_n(\theta) - \hat{Q}_n(\theta')|] \leq \{\sum_{t=1}^n E[d_t(z_t)]/n\} \|\theta - \theta'\|. \end{aligned}$$

The hypotheses of Corollary 2.1 are satisfied for $B_n = \sum_{t=1}^n d_t(z_t)/n$. ■

Proof of Corollary 3.2: Consider the compact set $S_j = K_j \times \Theta$. It follows as in Potscher and Prucha (1989) that $q_t(z, \theta)$ is uniformly continuous on S_j , uniformly in t , so that for any θ and $\varepsilon > 0$ there exists a neighborhood $\mathcal{N}(\theta, \varepsilon, j)$ such that $\sup_{t, \theta' \in \mathcal{N}(\theta, \varepsilon, j), z \in K_j} |q_t(z, \theta') - q_t(z, \theta)| < \varepsilon/2$, giving

$$\begin{aligned}
(A.4) \quad \sup_{\theta' \in \mathcal{N}(\theta, \varepsilon, j)} |Q_n(\theta') - Q_n(\theta)| &\leq \sum_{t=1}^n \sup_{\theta' \in \mathcal{N}(\theta, \varepsilon, j)} |q_t(z_t, \theta') - q_t(z_t, \theta)|/n \\
&\leq \sum_{t=1}^n [\varepsilon + 2 \cdot 1(z_t \notin K_j) d_t(z_t)]/n \\
&\leq \varepsilon/2 + [\sum_{t=1}^n 1(z_t \notin K_j)/n]^{(\gamma-1)/\gamma} [\sum_{t=1}^n d_t(z_t)^\gamma/n]^{1/\gamma}.
\end{aligned}$$

Note that $\sum_{t=1}^n d_t(z_t)^\gamma/n = O_p(1)$ by the Markov inequality. Choose M such that $\limsup_{n \rightarrow \infty} \text{Prob}(\sum_{t=1}^n d_t(z_t)^\gamma/n > M) < 3/2$, j such that $\limsup_{n \rightarrow \infty} [\sum_{t=1}^n \text{Prob}(z_t \notin K_j)/n] \{(\varepsilon/2)^{\gamma/(\gamma-1)} / M^{1/(\gamma-1)}\}^{-1} \leq \varepsilon/3$, and let $\mathcal{N}(\theta, \varepsilon) = \mathcal{N}(\theta, \varepsilon, j)$ and $\Delta_n(\varepsilon)$ be the expression following the last inequality in eq. (A.4). Then

$$\begin{aligned}
\text{Prob}(\Delta_n(\varepsilon) > \varepsilon) &\leq \text{Prob}([\sum_{t=1}^n 1(z_t \notin K_j)/n]^{(\gamma-1)/\gamma} [\sum_{t=1}^n d_t(z_t)^\gamma/n]^{1/\gamma} > \varepsilon/2) \\
&\leq \text{Prob}(\sum_{t=1}^n 1(z_t \notin K_j)/n \geq (\varepsilon/2)^{\gamma/(\gamma-1)} / M^{1/(\gamma-1)}) \\
&\quad + \text{Prob}([\sum_{t=1}^n d_t(z_t)^\gamma/n]^{1/\gamma} \geq M^{1/\gamma}) \\
&\leq \sum_{t=1}^n E[1(z_t \notin K_j)/n] \{(\varepsilon/2)^{\gamma/(\gamma-1)} / M^{1/(\gamma-1)}\}^{-1} + \varepsilon/3 + o(1) \\
&\leq 2\varepsilon/3 + o(1),
\end{aligned}$$

giving Assumption 3. Equicontinuity of $\bar{Q}_n(\theta)$ follows by a similar argument, so that the conclusion follows from Theorem 1. \blacksquare

Proof of Corollary 4.1: Take $\hat{Q}_n(\theta)$ as in eq. (4.1) and $\bar{Q}(\theta) = E[m(z_1, z_2, \theta)]$. Assumption 2 is satisfied by Theorem A of Serfling (1980). By the i.i.d. assumption, the distribution of (z_t, z_s) , $t \neq s$, is invariant to t and s , implying $E[\hat{Q}_n(\theta)] = \bar{Q}(\theta)$. Furthermore, for $B_n = 2 \sum_{t=1}^n \sum_{s>t} d(z_s, z_t)/n(n-1)$,

$$E[B_n] = E[d(z_1, z_2)] < \infty,$$

$$|\hat{Q}_n(\theta) - \hat{Q}_n(\theta')| \leq 2 \sum_{t=1}^n \sum_{s>t} |m(z_s, z_t, \theta) - m(z_s, z_t, \theta')| / n(n-1) \leq B_n \|\theta - \theta'\|,$$

$$|\bar{Q}(\theta) - \bar{Q}(\theta')| \leq E[|\hat{Q}_n(\theta) - \hat{Q}_n(\theta')|] \leq E[B_n] \|\theta - \theta'\|.$$

It follows by the last inequality that $\bar{Q}(\theta)$ is continuous, so that by the second inequality the hypotheses of Theorem 2 are satisfied. ■

References

- Andrews, D.W.K. (1989): "Consistency in Nonlinear Econometric Models," *Econometrica*, forthcoming.
- Newey, W.K. (1988): "Efficient Estimation of Semiparametric Models Via Moment Restrictions," mimeo, Princeton University.
- Newey, W.K. and J.L. Powell (1989): "Nonparametric Instrumental Variables Estimation," mimeo, University of Wisconsin.
- Pollard, D. (1985): *Convergence of Stochastic Process*. New York: Springer-Verlag.
- Potscher, B. and I.R. Prucha (1989): "The Uniform Law of Large Numbers and Dependent and Heterogeneous Processes," *Econometrica*, forthcoming.
- Rudin, W. (1976): *Principles of Mathematical Analysis*, New York: McGraw-Hill.
- Serfling (1980): *Approximation Theorems of Mathematical Statistics*, New York: Wiley.