COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 940

GENERIC UNIFORM CONVERGENCE

by

Donald W. K. Andrews[1]

August, 1989

Revised: March, 1990

ABSTRACT

This paper presents several generic uniform convergence results that include generic uniform laws of large numbers. These results provide conditions under which pointwise convergence almost surely or in probability can be strengthened to uniform convergence. The results are useful for establishing asymptotic properties of estimators and test statistics.

The results given here have the following attributes, (1) they extend results of Newey [10] to cover convergence almost surely as well as convergence in probability, (2) they apply to totally bounded parameter spaces (rather than just to compact parameter spaces), (3) they introduce a set of conditions for a generic uniform law of large numbers that has the attribute of giving the weakest conditions available for iid contexts, but which apply in dependent non–identically distributed contexts as well, and (4) they incorporate and extend the main results in the literature in a parsimonious fashion.

JEL Classification Number: 211.

Keywords: Almost sure convergence, consistency, convergence in probability, generic uniform convergence, law of large numbers, stochastic equicontinuity, uniform convergence, uniform law of large numbers.

1. INTRODUCTION

This paper extends several generic uniform convergence results and uniform laws of large numbers (LLN's) given in the literature by Andrews [1], Pötscher and Prucha [12], and Newey [10]. By definition, these "generic" results provide sufficient conditions to turn pointwise convergence (or pointwise LLN's) into uniform convergence (or uniform LLN's). As is well–known, uniform convergence and uniform LLN's are used in standard proofs of consistency of extremum estimators, including parametric, semiparametric, and nonparametric estimators. In consequence, generic uniform results are very useful tools for establishing asymptotic properties of econometric estimators.

The contribution of the present paper to the literature on generic uniform results is as follows: First, the paper follows the approach of Newey [10] and uses a stochastic equicontinuity (SE) condition. The SE condition of Newey [10] is modified, however, in a way that simplifies the condition itself and simplifies the proofs of the results. The modification also allows one to derive almost sure (a.s.) results in addition to the "in probability" results given by Newey [10]. As in Newey [10], the SE condition is shown to be a necessary condition for the generic uniform results. When replaced by more primitive conditions, the SE condition yields simpler conditions than those obtained by following the first moment continuity approach adopted in Andrews [1] and Pötscher and Prucha [12]. In some cases, the resultant conditions are also easier to verify.

Second, the results given here apply when the parameter space is non–compact. The compactness condition is replaced by a condition of total boundedness. (Recall that a subset of a metric space is totally bounded if it can be covered by a finite number of ϵ–balls $\forall \epsilon > 0$.) A totally bounded and complete parameter space is compact, so total boundedness is weaker than compactness. For Euclidean parameter spaces, total boundedness is equivalent to boundedness and compact sets are bounded and closed. In this case, the results of the paper remove the restriction that the parameter space is closed. See below for several reasons why the relaxation of this restriction is useful.[2,3]

Third, the two main sets of primitive conditions for generic uniform LLN's given by Andrews [1] and Pötscher and Prucha [12] have the drawback that they do not contain the best results available for the case when the random variables (rv's) are independent and identically distributed (iid) (e.g., see the uniform LLN of Jennrich [8, Thm. 2]). Here, we present a third set of primitive conditions for a uniform LLN (see Assumption TSE–1 of Section 3) that applies to dependent non–identically distributed (dnid) rv's. These conditions have the attribute that they obtain (and generalize) the best results available in the case where the rv's are iid. They also contain a generic analogue of Bierens' [6, Lemma 3] uniform LLN, as well as a new generic uniform LLN based on an absolute continuity condition (see Assumption TSE–1A below). We point out, however, that this third set of primitive conditions does not dominate the conditions given in Andrews [1] or Pötscher and Prucha [12]. Each of the three sets of primitive conditions involves a different tradeoff between conditions. Depending upon the context, any one of the three may be the most suitable.

Fourth, the results given here are simple and compact. They incorporate and extend the main results in the literature (or variants thereof) in a parsimonious fashion.

Next, we discuss three reasons why it is useful to replace the compactness assumption on the parameter space with the total boundedness assumption. First, in some cases the natural parameter space has one or more parameters restricted by a strict inequality. For example, a variance parameter may be restricted to be positive. In this case, the parameter space has to be artificially restricted if it is to be compact but need not be if it is to be totally bounded.

Second, to obtain asymptotic normality of parametric estimators, the true parameter must be an interior point of the parameter space. Hence, to obtain asymptotic normality of an estimator when the true parameter can be any point in the parameter space, the parameter space must be open (as is often assumed in asymptotic normality results in the literature). This assumption is incompatible with the standard compactness assumption used for uniform LLN's and consistency results. With a totally bounded parameter space, however, there is no incompatibility.

Third, the econometric literature on uniform LLN's and consistency in nonlinear models tends to indicate that compactness is a crucial feature of the methods used. The results given here show that it is not a crucial feature, but rather, just a convenient property.

The remainder of this paper is organized as follows: Section 2 provides the generic uniform convergence results for convergence in probability and convergence almost surely. Section 3 presents the generic uniform LLN's mentioned above. The final section, Section 4, gives two consistency results for extremum estimators that use uniform convergence properties and uniform LLN's as assumptions. These results are not new, but are included to illustrate the use of the results in Sections 2 and 3 in proving the consistency of extremum estimators when the parameter space is non–compact.

2. GENERIC UNIFORM CONVERGENCE

We start by giving sufficient conditions for a sequence $\{G_n(\theta) : n \geq 1\}$ to converge to zero in probability and almost surely uniformly over an index set Θ as $n \to \infty$. For the special case of a uniform LLN, $G_n(\theta)$ is of the form $\frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta))$, where Σ_1^n denotes $\sum_{t=1}^{n}$. Let $(\Omega,\mathcal{F},P)$ be a probability space and let Θ be a metric space with metric d. For each $\theta \in \Theta$ and $n \geq 1$, let $G_n(\theta)$ $(= G_n(\theta)_\omega)$ be a real (not necessarily measurable) function on Ω. Let $B(\theta,\delta)$ denote a closed ball in Θ of radius $\delta \geq 0$ centered at θ.

For simplicity, measurability assumptions are used in this paper only when needed.[4] All statements and assumptions concerning probabilities (or expectations) of sets (or functions) that are not measurable are understood to hold for both inner and outer probability (or inner and outer expectation). All limits are taken as $n \to \infty$ unless stated otherwise.

First, we consider "in probability" results:

DEFINITION: $\{G_n(\theta) : n \geq 1\}$ is *stochastically equicontinuous on* Θ if: $\forall \epsilon > 0\ \exists \delta > 0$ such that

$$\overline{\lim_{n\to\infty}} P\left[\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |G_n(\theta') - G_n(\theta)| > \epsilon\right] < \epsilon .$$

This definition is the standard one in the literature. For example, it is the definition used in Pollard [11] and Andrews [3,4,5].

The assumptions below use the following abbreviations: BD for totally bounded, P–WCON for pointwise weak (i.e., "in probability") convergence to zero, SE for stochastic equicontinuity, and U–WCON for uniform weak convergence to zero.

ASSUMPTION BD: Θ *is a totally bounded metric space.*

ASSUMPTION P–WCON: $G_n(\theta) \xrightarrow{p} 0\ \ \forall\theta \in \Theta$.

ASSUMPTION SE: $\{G_n(\theta) : n \geq 1]$ *is stochastically equicontinuous on* Θ .

PROPERTY U–WCON: $\sup_{\theta\in\Theta} |G_n(\theta)| \xrightarrow{p} 0$.

THEOREM 1: (a) BD, P–WCON, & SE $\Rightarrow$ U–WCON. (b) U–WCON $\Rightarrow$ P–WCON & SE.

COMMENTS: 1. Theorem 1 is quite similar to Theorem 1 of Newey [10]. It does not require compactness of Θ , however, and it uses a somewhat different SE condition than that used by Newey [10]. See Section 4 of Newey [10] for two examples of the use of Theorem 1 in non–LLN contexts.

2. Theorem 1 can be obtained as a special case of a recent result of Pollard [11, Thm. 10.2] regarding the weak convergence of a sequence of stochastic processes to a limit process that is not necessarily degenerate. The proof of Theorem 1 given here, however, is very much simpler than that required for Pollard's result.

PROOF OF THEOREM 1: For part (a), suppose BD, P–WCON and SE hold. Given $\epsilon > 0$, take δ as in the definition of stochastic equicontinuity. Using BD, let $\{B(\theta_j, \delta) : j = 1, \ldots, J\}$ be a finite cover of Θ . Using P–WCON and SE,

$$\overline{\lim_{n\to\infty}} P\left[\sup_{\theta\in\Theta}|G_n(\theta)| > 2\epsilon\right] \le \overline{\lim_{n\to\infty}} P\left[\max_{j\le J}\sup_{\theta'\in B(\theta_j,\delta)}(|G_n(\theta')-G_n(\theta_j)|+|G_n(\theta_j)|) > 2\epsilon\right]$$

$$\le \overline{\lim_{n\to\infty}} P\left[\sup_{\theta\in\Theta}\sup_{\theta'\in B(\theta,\delta)}|G_n(\theta')-G_n(\theta)| > \epsilon\right] + \overline{\lim_{n\to\infty}} P\left[\max_{j\le J}|G_n(\theta_j)| > \epsilon\right] < \epsilon . \tag{2.1}$$

For part (b), U–WCON ⇒ P–WCON is immediate and U–WCON ⇒ SE follows from

$$P\left[\sup_{\theta\in\Theta}\sup_{\theta'\in B(\theta,\delta)}|G_n(\theta')-G_n(\theta)| > \epsilon\right] \le P\left[2\sup_{\theta\in\Theta}|G_n(\theta)| > \epsilon\right] \longrightarrow 0 . \ \square \tag{2.2}$$

Next, we consider "almost sure" results:

DEFINITION: $\{G_n(\theta) : n \ge 1\}$ is *strongly stochastically equicontinuous on* Θ if $\{\sup_{m\ge n}|G_m(\theta)| : n \ge 1\}$ is stochastically equicontinuous on Θ.

The assumptions below use the abbreviations: P–SCON for pointwise strong (i.e., almost sure) convergence, SSE for strong stochastic equicontinuity, and U–SCON for uniform strong convergence.

ASSUMPTION P–SCON: $G_n(\theta) \to 0$ a.s. $\forall\theta\in\Theta$.

ASSUMPTION SSE: $\{G_n(\theta) : n \ge 1\}$ *is strongly stochastically equicontinuous on* Θ.

PROPERTY U–SCON: $\sup_{\theta\in\Theta}|G_n(\theta)| \to 0$ a.s.

THEOREM 2: (a) BD, P–SCON, & SSE ⇒ U–SCON. (b) U–SCON ⇒ P–SCON & SSE.

PROOF OF THEOREM 2: Theorem 2 follows from Theorem 1 and the well–known result

$$X_n \to 0 \text{ a.s.} \Leftrightarrow \sup_{m\ge n}|X_m| \xrightarrow{p} 0 . \ \square \tag{2.3}$$

For completeness, we provide a simple, well–known proof of (2.3): We have $\{|X_n| > \epsilon \text{ i.o.}\} \uparrow \{X_n \not\to 0\}$ as $\epsilon \downarrow 0$, where i.o. abbreviates "infinitely often." Thus, by the monotone convergence theorem, $X_n \to 0$ a.s. $\Leftrightarrow P(|X_n| > \epsilon \text{ i.o.}) = 0\ \forall\epsilon > 0$.

Next, $\{\sup_{m\geq n}|X_m| > \epsilon\} = \bigcup_{m\geq n}\{|X_m| > \epsilon\} \downarrow \{|X_m| > \epsilon \text{ i.o.}\}$ as $n \to \infty$. Hence, $\lim_{n\to\infty} P(\sup_{m\geq n}|X_m| > \epsilon) = P(|X_n| > \epsilon \text{ i.o.})\ \forall\epsilon > 0$ and (2.3) follows.

We now introduce Lipschitz conditions that are sufficient for SE and SSE. For the special case of uniform LLN's, alternative sufficient conditions for SE and SSE are given in Section 3. The latter include both Lipschitz and non–Lipschitz conditions.

ASSUMPTION SE–1: (a) $G_n(\theta) = \hat{Q}_n(\theta) - Q_n(\theta)$, *where* $Q_n(\theta)$ *is a nonrandom function that is continuous in* θ *uniformly over* $\theta \in \Theta$ *and* $n \geq 1$.

(b) $|\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| \leq B_n h(d(\theta', \theta))\ \ \forall\theta', \theta \in \Theta$ a.s. *for some nonrandom function* h *such that* $h(y) \downarrow 0$ *as* $y \downarrow 0$.

(c) *Either* $\sup_{n\geq 1} EB_n < \infty$ *or* $B_n = O_p(1)$.

ASSUMPTION SSE–1: (a) SE–1(a) *and* (b) *hold.*

(b) *Either* $\sup_{n\geq 1} EB_n < \infty$ *and* $B_n - EB_n \to 0$ a.s. *or* $\sup_{m\geq n} B_m = O_p(1)$.

LEMMA 1: (a) SE–1 ⇒ SE. (b) SSE–1 ⇒ SSE.

COMMENT: If $G_n(\theta)$ satisfies a Lipschitz condition, one can take $Q_n(\theta) = 0$ in SE–1 and SSE–1.

PROOF OF LEMMA 1: The second condition of SSE–1(b) is implied by the first. Hence, SSE–1 ⇒ SSE follows from

$$
\begin{aligned}
&\overline{\lim_{n\to\infty}}\, P\left[\sup_{\theta\in\Theta}\ \sup_{\theta'\in B(\theta,\delta)}\left|\sup_{m\geq n}|G_m(\theta')| - \sup_{m\geq n}|G_m(\theta)|\right| > \epsilon\right] \\
&\leq \overline{\lim_{n\to\infty}}\, P\left[\sup_{m\geq n} B_m h(\delta) > \epsilon/2\right] + 1\left[\sup_{n\geq 1}\ \sup_{\theta\in\Theta}\ \sup_{\theta'\in B(\theta,\delta)}|Q_n(\theta') - Q_n(\theta)| > \epsilon/2\right] < \epsilon
\end{aligned}
\tag{2.4}
$$

for δ sufficiently small. The proof of SE–1 ⇒ SE is analogous. □

3. GENERIC UNIFORM LAWS OF LARGE NUMBERS

3.1. Preliminaries

Throughout this section, we take $G_n(\theta) = \frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta))$, where Z_t is assumed to be a $\mathcal{Z}$–valued random variable (rv) for some measurable space $(\mathcal{Z}, \mathcal{B})$ and $q_t(z,\theta)$ is a measurable function from $\mathcal{Z}$ to R for each $\theta \in \Theta$ and each $t \geq 1$. For $G_n(\theta)$ as above, we give primitive conditions on $\{Z_t : t \geq 1\}$ and $\{q_t(z,\theta) : t \geq 1\}$ such that SE and SSE hold. Then, Theorems 1 and 2 yield generic weak and strong LLN's.

For $G_n(\theta)$ as above, P–WCON, P–SCON, U–WCON, and U–SCON correspond to a pointwise weak LLN (P–WLLN), a pointwise strong LLN (P–SLLN), a uniform weak LLN (U–WLLN), and a uniform strong LLN (U–SLLN). For reasons of convention, we adopt the latter terminology in this section and define:

ASSUMPTION P–WLLN: $\frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta)) \xrightarrow{p} 0 \ \forall\theta \in \Theta$.

ASSUMPTION P–SLLN: $\frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta)) \rightarrow 0$ a.s. $\forall\theta \in \Theta$.

PROPERTY U–WLLN: $\sup_{\theta\in\Theta}\left|\frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta))\right| \xrightarrow{p} 0$.

PROPERTY U–SLLN: $\sup_{\theta\in\Theta}\left|\frac{1}{n}\Sigma_1^n(q_t(Z_t, \theta) - Eq_t(Z_t, \theta))\right| \rightarrow 0$ a.s.

For example, Assumptions P–WLLN and P–SLLN can be verified for dependent non–identically distributed rv's using results of Andrews [2] and McLeish [9].

Next, we introduce a useful continuity property of $\frac{1}{n}\Sigma_1^n Eq_t(Z_t, \theta)$ in θ. It is implied by the sufficient conditions given below for a U–WLLN or a U–SLLN.

PROPERTY CTY: $\frac{1}{n}\Sigma_1^n Eq_t(Z_t, \theta)$ is continuous in θ uniformly over $\theta \in \Theta$ and $n \geq 1$.

3.2. Generic Uniform LLN'*s Via a Lipschitz Condition*

The first set of sufficient conditions for a U–WLLN and U–SLLN utilizes a Lipschitz condition, as in Andrews [1]:

ASSUMPTION W–LIP: (a) $|q_t(Z_t, \theta') - q_t(Z_t, \theta)| \le B_t(Z_t)h(d(\theta', \theta))$ $\forall \theta', \theta \in \Theta$ a.s. *for some measurable functions* $\{B_t : t \ge 1\}$ *and some nonrandom function* h *that satisfies* $h(y) \downarrow 0$ *as* $y \downarrow 0$.

(b) $\sup_{n \ge 1} \frac{1}{n}\Sigma_1^n EB_t(Z_t) < \infty$.

ASSUMPTION S–LIP: (a) W–LIP holds.

(b) $\frac{1}{n}\Sigma_1^n(B_t(Z_t) - EB_t(Z_t)) \longrightarrow 0$ a.s.

LEMMA 2: (a) W–LIP ⇒ SE & CTY. (b) S–LIP ⇒ SSE & CTY.

PROOF OF LEMMA 2: S–LIP ⇒ W–LIP ⇒ CTY is immediate. Also, W–LIP ⇒ SE–1 and S–LIP ⇒ SSE–1 hold with $\hat{Q}_n(\theta) = \frac{1}{n}\Sigma_1^n q_t(Z_t, \theta)$, $Q_n(\theta) = \frac{1}{n}\Sigma_1^n Eq_t(Z_t, \theta)$, and $B_n = \frac{1}{n}\Sigma_1^n B_t(Z_t)$. Lemma 1 now gives the result. □

Theorems 1 and 2 and Lemma 2 give the following U–WLLN and U–SLLN results:

THEOREM 3: (a) BD, P–WLLN, & W–LIP ⇒ U–WLLN & CTY.

(b) BD, P–SLLN, & S–LIP ⇒ U–SLLN & CTY.

COMMENTS: 1. Theorem 3 is similar to Corollary 2 of Andrews [1] except that Θ is allowed to be totally bounded rather than compact and P–WLLN, P–SLLN, and S–LIP(b) arise in place of conditions of pointwise weak and strong LLN's for the supremum and infimum of $q_t(Z_t, \theta)$ over small neighborhoods of θ $\forall \theta \in \Theta$. The conditions P–WLLN etc. are neater and more elegant than the latter, but are usually not any more general. Theorem 3(a) is also very similar to Corollary 3.1 of Newey [10]. It differs only in that it allows Θ to be totally bounded, whereas Newey's U–WLLN requires Θ to be compact.

2. In spite of the description of W–LIP and S–LIP as Lipschitz conditions, these conditions are *not* smoothness conditions. The reason is that $h(\cdot)$ is arbitrary (provided $h(y) \downarrow 0$ as $y \downarrow 0$). For example, W–LIP and S–LIP are implied by continuity of $q_t(z,\theta)$ in θ uniformly over $\theta \in \Theta$, $z \in \mathcal{Z}$, and $t \ge 1$ and the latter is not a smoothness condition.[5] (A consequence of this result is that when Θ and $\mathcal{Z}$ are compact, Pötscher and

Prucha's [12] equicontinuity condition on $\{q_t(z,\theta) : t \geq 1\}$ implies W–LIP and S–LIP and, in turn, is close to being implied by the latter.)

3. In many cases, uniform LLN's are used in proofs of consistency and asymptotic normality of extremum estimators. Standard proofs of the asymptotic normality of such estimators rely on the differentiability of the functions $\{q_t(z,\theta) : t \geq 1\}$ in θ. An attribute of the conditions W–LIP and S–LIP are that they are implied by such assumptions. In contrast, some other sets of sufficient conditions for uniform LLN's, including those of Pötscher and Prucha [12] and Assumption TSE–2A below, impose continuity assumptions on $q_t(z,\theta)$ in (z,θ) that are not implied by the standard differentiability in θ assumptions used to obtain asymptotic normality.

On the other hand, if one is interested only in consistency and not in asymptotic normality, then it is useful to have at one's disposal alternative generic uniform LLN's. A variety of such uniform LLN's are described below and in the references given below.

3.3. Generic Uniform WLLN*'s Via Termwise Stochastic Equicontinuity*

Here we introduce a condition called *termwise stochastic equicontinuity* (TSE) which, together with a domination (DM) condition, is sufficient for SE. The TSE condition is not primitive, since it involves an interaction between the functions $\{q_t(z,\theta) : t \geq 1\}$ and the marginal distributions of $\{Z_t : t \geq 1\}$. It is easy to obtain primitive sufficient conditions that imply TSE, however, and a variety is given below.

ASSUMPTION TSE: $\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} \frac{1}{n}\Sigma_1^n P\left[\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |q_t(Z_t, \theta') - q_t(Z_t, \theta)| > \epsilon\right] = 0$ $\forall \epsilon > 0$.

ASSUMPTION DM: $\lim_{M\to\infty} \overline{\lim_{n\to\infty}} \frac{1}{n}\Sigma_1^n E d_t(Z_t) 1(d_t(Z_t) > M) = 0$ *for some nonrandom functions* $\{d_t : t \geq 1\}$ *that satisfy* $d_t(z) \geq \sup_{\theta\in\Theta} |q_t(z,\theta)|$ $\forall z \in \mathcal{Z}$, $\forall t \geq 1$.

Assumption DM is implied by $\overline{\lim_{n\to\infty}} \frac{1}{n}\Sigma_1^n E d_t^{\gamma}(Z_t) < \infty$ for some $\gamma > 1$. Also, for identically distributed $\{Z_t : t \geq 1\}$, if $d_t(z)$ can be taken to be independent of t (as occurs, for example, if $q_t(z,\theta)$ does not depend on t), then DM is implied by $Ed(Z_t) < \infty$.

LEMMA 3: DM & TSE $\Rightarrow$ SE & CTY.

PROOF OF LEMMA 3: Let $Y_{t\delta} = \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |q_t(Z_t, \theta') - q_t(Z_t, \theta)|$ and $d_t = d_t(Z_t)$. Given $\epsilon > 0$, take $M < \infty$ and $\delta > 0$ such that $\overline{\lim}_{n\to\infty} \frac{2}{n}\Sigma_1^n E d_t 1(d_t > M/2) < \epsilon^2/6$ and $\overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n P(Y_{t\delta} > \epsilon^2/6) < \epsilon^2/(6M)$. Then, we have

$$\begin{aligned}
\overline{\lim}_{n\to\infty} P\Big[\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |G_n(\theta') - G_n(\theta)| > \epsilon\Big] &\leq \overline{\lim}_{n\to\infty} P\Big[\tfrac{1}{n}\Sigma_1^n (Y_{t\delta} + EY_{t\delta}) > \epsilon\Big] \\
&\leq \frac{2}{\epsilon}\, \overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n EY_{t\delta} \\
&= \frac{2}{\epsilon}\, \overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n \Big[EY_{t\delta}1(Y_{t\delta} \leq \epsilon^2/6) + EY_{t\delta}1(\epsilon^2/6 < Y_{t\delta} \leq M) + EY_{t\delta}1(Y_{t\delta} > M)\Big] \qquad (3.1) \\
&\leq \frac{2}{\epsilon}\Big[\epsilon^2/6 + \overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n M P(Y_{t\delta} > \epsilon^2/6) + \overline{\lim}_{n\to\infty} \frac{2}{n}\Sigma_1^n E d_t 1(2d_t > M)\Big] \leq \epsilon .
\end{aligned}$$

Thus, SE holds. Since $\overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n EY_{t\delta} < \epsilon^2/2$, CTY also holds. □

Theorem 1 and Lemma 3 combine to give the following U–WLLN result:

THEOREM 4: BD, P–WLLN, DM, & TSE $\Rightarrow$ U–WLLN & CTY.

We now introduce a series of assumptions that are primitive and that imply TSE. These assumptions will be shown to satisfy

$$\begin{array}{ccccccc}
 & & & \nearrow \text{TSE–1A} \searrow & & & \\
\text{TSE–1D} & \rightarrow & \text{TSE–1C} & & \text{TSE–1} & & \\
 & & & \searrow \text{TSE–1B} \nearrow & & \searrow & \\
 & & & \text{TSE–2A} \rightarrow & \text{TSE–2} & \rightarrow & \text{TSE} \\
 & & & & \text{W–LIP} & \nearrow &
\end{array} \qquad (3.2)$$

when Θ is totally bounded.

Let μ_t denote the probability distribution or distribution function of Z_t for $t \geq 1$. Let $\bar{\mu}_n = \frac{1}{n}\Sigma_1^n \mu_t$ for $n \geq 1$.

ASSUMPTION TSE–1: (a) $q_t(z,\theta)$ *is continuous in* θ *uniformly over* $\theta \in \Theta$ *and* $t \geq 1$ $\forall z \in \mathcal{Z}$.

(b) *For every sequence of measurable sets* $\{A_m \subset \mathcal{Z} : m \geq 1\}$ *such that* $A_m \downarrow \phi$ *as* $m \to \infty$, $\overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n P(Z_t \in A_m) \to 0$ *as* $m \to \infty$.

ASSUMPTION TSE–1A: (a) $q_t(z,\theta)$ *is continuous in* θ *uniformly over* $\theta \in \Theta$ *and* $t \geq 1$ $\forall z \in \mathcal{Z}$.

(b) μ_t *is absolutely continuous with respect to some* σ*–finite measure* μ $\forall t \geq 1$ *and* $\int \left[\sup_{n\geq 1} \frac{1}{n}\Sigma_1^n f_t(z)\right] d\mu(z) < \infty$, *where* $f_t(z)$ *denotes the density of* μ_t *with respect to* μ.

ASSUMPTION TSE–1B: (a) $q_t(z,\theta)$ *is continuous in* θ *uniformly over* $\theta \in \Theta$ *and* $t \geq 1$ $\forall z \in \mathcal{Z}$.

(b) $\bar{\mu}_n \to \mu$ *properly setwise for some probability measure* μ, *i.e., for all measurable sets* $A \subset \mathcal{Z}$, $\frac{1}{n}\Sigma_1^n P(Z_t \in A) \to \mu(A)$.

ASSUMPTION TSE–1C: (a) $q_t(z,\theta)$ *is continuous in* θ *uniformly over* $\theta \in \Theta$ *and* $t \geq 1$ $\forall z \in \mathcal{Z}$.

(b) $\{Z_t : t \geq 1\}$ *are identically distributed.*

ASSUMPTION TSE–1D: (a) $q_t(z,\theta) = q(z,\theta)$ $\forall t \geq 1$ *and* $q(z,\theta)$ *is continuous in* θ $\forall \theta \in \Theta$, $\forall z \in \mathcal{Z}$.

(b) Θ *is compact.*

(c) $\{Z_t : t \geq 1\}$ *are identically distributed.*

ASSUMPTION TSE–2: (a) $q_t(z,\theta) = \Sigma_{k=1}^K r_{kt}(z) s_{kt}(z,\theta)$ *and* $s_{kt}(z,\theta)$ *is continuous in* θ *uniformly over* $\theta \in \Theta$, $z \in C_j$, *and* $t \geq 1$, $\forall 1 \leq k \leq K$, $\forall j \geq 1$, *where* $\{C_j \subseteq \mathcal{Z} : j \geq 1\}$ *is some nondecreasing sequence of compact sets whose union is* $\mathcal{Z}$.

(b) $\sup_{n\geq 1} \frac{1}{n}\Sigma_1^n E|r_{kt}(Z_t)| < \infty$ $\forall 1 \leq k \leq K$.

(c) $\{\bar{\mu}_n : n \geq 1\}$ *is tight, i.e.,* $\lim_{j\to\infty} \sup_{n\geq 1} \frac{1}{n}\Sigma_1^n P(Z_t \in \tilde{C}_j) = 1$ *for some sequence of compact sets* $\{\tilde{C}_j \subseteq \mathcal{Z} : j \geq 1\}$.

ASSUMPTION TSE–2A: (a) $q_t(z,\theta) = \Sigma_{k=1}^{K} r_{kt}(z) s_{kt}(z,\theta)$ *and* $\{s_{kt}(z,\theta) : t \geq 1\}$ *is equicontinuous on* $\mathcal{Z} \times \Theta$ $\forall 1 \leq k \leq K$, *i.e.,* $\sup_{t \geq 1} |s_{kt}(z', \theta') - s_{kt}(z,\theta)| \to 0$ *as* $(z', \theta') \to (z,\theta)$ $\forall (z,\theta) \in \mathcal{Z} \times \Theta$, $\forall 1 \leq k \leq K$.

(b) Θ *is compact.*

(c) TSE–2(b) *and* (c) *hold.*

The condition TSE–1 has not been considered previously in the literature. We view it as a third alternative to the Lipschitz condition of Andrews [1] (i.e., W–LIP or S–LIP) and the equicontinuity condition of Pötscher and Prucha [12] (i.e., TSE–2 or TSE–2A). Roughly speaking, TSE–1(b) requires that the standard continuity property of the distribution of a rv Z_t (i.e., $A_m \downarrow \phi \Rightarrow P(Z_t \in A_m) \downarrow 0$ as $m \to \infty$) holds uniformly over $t \geq 1$. Of course, with identically distributed rv's, this condition holds automatically. Also, in the case where $\{Z_t : t \geq 1\}$ takes on at most a finite number of different distributions (e.g., when "fixed in repeated samples" asymptotics are used), this condition automatically holds. As mentioned in the Introduction, an attribute of TSE–1 is that it applies with dnid rv's, yet when the rv's are identically distributed, it yields the best available sufficient conditions for a generic U–WLLN.

TSE–1D corresponds to Jennrich's [8, Thm. 2] conditions for a uniform LLN for iid rv's. TSE–1C generalizes the Jennrich–type conditions to allow for a totally bounded parameter space, functions $q_t(z,\theta)$ that depend on t, and rv's $\{Z_t : t \geq 1\}$ that are non-Euclidean–valued. A nice feature of TSE is that the sufficiency of TSE–1C and TSE–1D for TSE follows immediately by the monotone convergence theorem. TSE–1B corresponds to Bieren's [6, Lem. 3] conditions for a uniform LLN (but generalized in the same way that TSE–1C generalizes TSE–1D).[6] TSE–1A is a new condition. It does not require convergence of the average distributions of $\{Z_t\}$ as in TSE–1B.

TSE–2A corresponds to Pötscher and Prucha's [12] conditions for a generic uniform LLN. TSE–2 generalizes these conditions to allow for a totally bounded parameter space.

LEMMA 4: *Suppose* BD *holds. Then,*

$$\text{(a) TSE–1D} \longrightarrow \text{TSE–1C} \begin{array}{c}\nearrow \text{TSE–1A} \searrow \\ \searrow \text{TSE–1B} \nearrow\end{array} \text{TSE–1} \longrightarrow \text{TSE.}$$

(b) TSE–2A ⟶ TSE–2 ⟶ TSE.

(c) W–LIP ⟶ TSE.

COMMENTS: 1. TSE–1 is weaker than TSE–2 with regard to the assumptions on $\{q_t(z,\theta) : t \geq 1\}$. It is stronger than TSE–2, however, regarding the assumptions on $\{Z_t : t \geq 1\}$. TSE–2 uses the additional assumptions on $\{q_t(z,\theta) : t \geq 1\}$ to identify a particular sequence of sets $\{A_m : m \geq 1\}$ that satisfies $A_m \downarrow \phi$ as $m \to \infty$ and for which $\overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n P(Z_t \in A_m) \to 0$ as $m \to \infty$ implies TSE (see the proof of Lemma 4).

2. Part (a) of TSE–1B, TSE–1C, and TSE–1D only needs to hold a.s. $[\mu_s]$ $\forall s \geq 1$, rather than $\forall z \in \mathcal{Z}$, in order for each to imply TSE. (Although Lemma 4(a) no longer holds with this change.)

PROOF OF LEMMA 4: TSE–1D ⇒ TSE–1C, TSE–1C ⇒ TSE–1B, and TSE–1C ⇒ TSE–1A are immediate. TSE–1B ⇒ TSE–1 follows from

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\Sigma_1^n P(Z_t \in A_m) = \mu(A_m) \to 0 \text{ as } m \to \infty \tag{3.3}$$

whenever $A_m \in \mathcal{B}$ $\forall m \geq 1$ and $A_m \downarrow \phi$ as $m \to \infty$. TSE–1A ⇒ TSE–1 follows from

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\Sigma_1^n P(Z_t \in A_m) \leq \int 1(z \in A_m)\left[\sup_{n\geq 1} \frac{1}{n}\Sigma_1^n f_t(z)\right] d\mu(z) \to 0 \text{ as } m \to \infty \tag{3.4}$$

whenever $A_m \in \mathcal{B}$ $\forall m \geq 1$ and $A_m \downarrow \phi$ as $m \to \infty$ by the dominated convergence theorem. TSE–1 ⇒ TSE is obtained by setting $A_m = \{z \in \mathcal{Z} : \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,1/m)} \sup_{t\geq 1} |q_t(z,\theta') - q_t(z,\theta)| > \epsilon\}$. By BD and TSE–1(a), A_m is measurable $\forall m \geq 1$ (this is the only place BD is used in Lemma 4) and by TSE–1(a), $A_m \downarrow \phi$ as $m \to \infty$. Hence, TSE–1(b) gives the desired result.

TSE–2A ⇒ TSE–2, since continuous functions on compact sets are uniformly continuous. To show TSE–2 ⇒ TSE, take $K = 1$ and $C_j = \tilde{C}_j$ in TSE–2 wlog. Let $Y_{t\delta}(z) = \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |s_{kt}(z, \theta') - s_{kt}(z,\theta)|$. Given $\epsilon > 0$, take j and $\delta > 0$ such that $\overline{\lim}_{n\to\infty} \frac{1}{n}\Sigma_1^n P(Z_t \notin C_j) < \epsilon/2$ and $\sup_{z\in C_j} \sup_{t\geq 1} Y_{t\delta}(z) < \epsilon^2/(4R)$, where $R = \sup_{n\geq 1} \frac{1}{n}\Sigma_1^n E|r_{kt}(Z_t)|$. Then, we have

$$
\begin{aligned}
&\overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n P\Big[\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |q_t(Z_t,\theta') - q_t(Z_t,\theta)| > \epsilon\Big] \\
&\leq \overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n P(|r_{kt}(Z_t)|Y_{t\delta}(Z_t)1(Z_t \in C_j) > \epsilon/2) \\
&\quad + \overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n P(|r_{kt}(Z_t)|Y_{t\delta}(Z_k)1(Z_t \notin C_j) > \epsilon/2) \qquad (3.5) \\
&\leq \overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n P(|r_{kt}(Z_t)|\epsilon^2/(4R) > \epsilon/2) + \overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n P(Z_t \notin C_j) \\
&\leq \overline{\lim_{n\to\infty}} \tfrac{1}{n}\Sigma_1^n E|r_{kt}(Z_t)|\epsilon/(2R) + \epsilon/2 \leq \epsilon .
\end{aligned}
$$

Lastly, W–LIP ⇒ TSE is immediate by substitution. □

Theorem 1 and Lemmas 3 and 4 combine to give the following U–WLLN result:

THEOREM 5: BD, P–WLLN, DM, & *any one of the* TSE *conditions* ⇒ U–WLLN & CTY.

3.4. Generic Uniform SLLN'*s Via Termwise Stochastic Equicontinuity*

To obtain a uniform *strong* LLN via the TSE condition the following additional condition is needed.

ASSUMPTION P–SLLN2: $\frac{1}{n}\Sigma_1^n(Y_{t1/j} - EY_{t1/j}) \to 0$ a.s. $\forall j \geq 1$, *where* $Y_{t\delta} = \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |q_t(Z_t, \theta') - q_t(Z_t, \theta)|$ *and* $Y_{t1/j}$ *is measurable* $\forall j \geq 1$.

LEMMA 5: DM,TSE, & P–SLLN2 ⇒ SSE & CTY.

COMMENT: The addition of P–SLLN2 to P–SLLN negates the simplicity of the pointwise convergence assumptions of Section 3.3 for uniform WLLN's in comparison to the assumptions of Andrews [1] and Pötscher and Prucha [12]. Nevertheless, P–SLLN and P–SLLN2 are no more restrictive than the assumptions of the latter two papers.

PROOF OF LEMMA 5: CTY holds by Lemma 3. To show SSE, given $\epsilon \in (0, 1/2)$, take δ as in the proof of Lemma 3 and such that $\delta = 1/j$ for some integer j. Then, we have

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} P\left[\sup_{\theta\in\Theta}\ \sup_{\theta'\in B(\theta,\delta)}\left|\sup_{m\geq n}|G_m(\theta')| - \sup_{m\geq n}|G_m(\theta)|\right| > \epsilon\right] \\
&\leq \overline{\lim_{n\to\infty}} P\left[\sup_{m\geq n}\tfrac{1}{m}\Sigma_1^m(Y_{t\delta} + EY_{t\delta}) > \epsilon\right] \\
&\leq \overline{\lim_{n\to\infty}} P\left[\sup_{m\geq n}\left|\tfrac{1}{m}\Sigma_1^m(Y_{t\delta} - EY_{t\delta})\right| > \epsilon/2\right] + \overline{\lim_{n\to\infty}} 1\left[\sup_{m\geq n}\tfrac{1}{m}\Sigma_1^m 2EY_{t\delta} > \epsilon/2\right] \\
&= 0 + 1\left[\overline{\lim_{n\to\infty}}\tfrac{1}{n}\Sigma_1^n EY_{t\delta} > \epsilon/4\right] = 0\,,
\end{aligned} \tag{3.6}$$

where the second last equality uses the characterization of a.s. convergence referred to in Section 2 and the last equality uses (3.1) of the proof of Lemma 3. □

Theorem 2 and Lemmas 4 and 5 yield the following U–SLLN result:

THEOREM 6: BD, P–SLLN, DM, *any one of the* TSE *conditions*, & P–SLLN2 ⇒ U–SLLN & CTY.

COMMENT: Theorem 6 adds the following to the literature on generic uniform strong LLN's. First, it gives generic U–SLLN's that do not require the parameter space to be compact. Second, it gives a third alternative sufficient condition (viz., TSE–1 and the conditions TSE–1A, ..., TSE–1D that imply it) to the Lipschitz and equicontinuity conditions of Andrews [1] and Pötscher and Prucha [12].

The results of this section have been concerned with *sequences* of rv's $\{Z_t : t \geq 1\}$ and functions $\{q_t(z,\theta) : t \geq 1\}$. They can be extended straightforwardly to *triangular arrays* $\{Z_{nt} : t \leq n, n \geq 1\}$ and $\{q_{nt}(z,\theta) : t \leq n, n \geq 1\}$ by adding the subscript n in the appropriate places. (Note, however, that pointwise *strong* LLN's hold for triangular arrays only under quite restrictive assumptions.) For example, uniform LLN's for triangular arrays are useful for establishing consistency and asymptotic normality results in the context of sequences of local alternatives.

4. CONSISTENCY RESULTS

This section shows how one can use the uniform convergence and uniform LLN results of Sections 2 and 3 to obtain consistency of extremum estimators when the parameter space is not compact but is totally bounded. The results of this section are quite similar to various results in the literature, and hence, they are intended to complement the results of Sections 2 and 3 rather than to be a significant new contribution.

The parameter space of interest is a metric space (B, ρ_B). Extremum estimators $\{\hat{\beta}_n : n \geq 1\}$ are defined that approximately minimize a stochastic criterion function $\hat{Q}_n(\beta)$. We will consider two combinations of the following assumptions regarding the extremum estimator (EE) and the criterion function (one combination for convergence in probability of $\hat{\beta}_n$ and the other for a.s. convergence):

ASSUMPTION W–EE: $\hat{Q}_n(\hat{\beta}_n) \leq \inf_{\beta \in B} \hat{Q}_n(\beta) + \epsilon_n$ *with probability* $\longrightarrow 1$ *for some* $\{\epsilon_n : n \geq 1\}$ *such that* $\epsilon_n \xrightarrow{p} 0$.

ASSUMPTION S–EE: $\hat{Q}_n(\hat{\beta}_n) \leq \inf_{\beta \in B} \hat{Q}_n(\beta) + \epsilon_n$ *for some* $\{\epsilon_n : n \geq 1\}$ *such that* $\epsilon_n \longrightarrow 0$ a.s.

ASSUMPTION U–WCON: $\sup_{\beta \in B} |\hat{Q}_n(\beta) - Q_n(\beta)| \xrightarrow{p} 0$ *for some sequence of non-stochastic functions* $\{Q_n(\beta) : n \geq 1\}$ *on* B.

ASSUMPTION U–SCON: $\sup_{\beta \in B} |\hat{Q}_n(\beta) - Q_n(\beta)| \longrightarrow 0$ a.s. *for some sequence of non-stochastic functions* $\{Q_n(\beta) : n \geq 1\}$ *on* B.

The next assumption is that of *identifiable uniqueness* of some non–stochastic sequence $\{\beta_{n0} : n \geq 1\}$ of elements of B. It is to this sequence that the estimators

$\{\hat{\beta}_n : n \geq 1\}$ "converge" as $n \to \infty$. Let $B(\beta,\epsilon)$ denote a closed ball in B of radius ϵ centered at β.

ASSUMPTION ID: $\forall \epsilon > 0$, $\varliminf_{n\to\infty} \inf_{\beta \notin B(\beta_{n0},\epsilon)} (Q_n(\beta) - Q_n(\beta_{n0})) > 0$.

The first result we present is the same as Lemma 3.1 of Pötscher and Prucha [13], which in turn is similar to other results in the literature. See Pötscher and Prucha [13] for references.

LEMMA 6: (a) W–EE, ID, & U–WCON $\Rightarrow \rho_B(\hat{\beta}_n, \beta_{n0}) \xrightarrow{p} 0$.
(b) S–EE, ID, & U–SCON $\Rightarrow \rho_B(\hat{\beta}_n, \beta_{n0}) \longrightarrow 0$ a.s.

COMMENT: To utilize Lemma 6, one can apply the results of Sections 2 or 3 to verify U–WCON or U–SCON.

For completeness, we give a proof of Lemma 6. This proof differs from that of Pötscher and Prucha [13] and is somewhat simpler.

PROOF OF LEMMA 6: Given $\epsilon > 0$, $\exists \delta > 0$ such that $\beta \notin B(\beta_{n0}, \epsilon) \Rightarrow Q_n(\beta) - Q_n(\beta_{n0}) \geq \delta > 0$ $\forall n$ large. Thus, for n large,

$$\begin{aligned} P(\hat{\beta}_n \notin B(\beta_{n0}, \epsilon)) &\leq P(Q_n(\hat{\beta}_n) - \hat{Q}_n(\hat{\beta}_n) + \hat{Q}_n(\hat{\beta}_n) - Q_n(\beta_{n0}) \geq \delta) \\ &\leq P(Q_n(\hat{\beta}_n) - \hat{Q}_n(\hat{\beta}_n) + \hat{Q}_n(\beta_{n0}) + \epsilon_n - Q_n(\beta_{n0}) \geq \delta) \qquad (4.1) \\ &\leq P(2 \sup_{\beta \in B} |\hat{Q}_n(\beta) - Q_n(\beta)| + \epsilon_n \geq \delta) \xrightarrow{p} 0 . \end{aligned}$$

For part (b), add "i.o." in the probabilities above and replace $\xrightarrow{p} 0$ by $= 0$. □

Next, we present a result that places more structure on the criterion function (CF) $\hat{Q}_n(\beta)$, and hence, which allows one to replace U–WCON and U–SCON by more primitive conditions.

ASSUMPTION W–CF: (a) $\hat{Q}_n(\beta) = d_n\left[\frac{1}{n}\Sigma_1^n q_t(Z_t, \beta, \hat{\tau}_n), \hat{\gamma}_n\right]$, *where* $d_n(\cdot,\cdot) : R^v \times \Gamma \to R^1$ *and* $q_t(\cdot,\cdot,\cdot) : \mathcal{Z} \times B \times \mathcal{T} \longrightarrow R^v$ *are non–stochastic functions,* $\hat{\tau}_n$ *and* $\hat{\gamma}_n$ *are stochastic preliminary nuisance parameter estimators that take values in the metric spaces* $(\mathcal{T}, \rho_{\mathcal{T}})$ *and* (Γ, ρ_Γ), *respectively,* $\{Z_t : t \geq 1\}$ *is a sequence of* $\mathcal{Z}$*–valued random variables, and* $q_t(z,\beta,\tau)$ *is measurable in* z *for each* $\beta \in B$ *and* $\tau \in \mathcal{T}_0 \subset \mathcal{T}$ *for* $\mathcal{T}_0$ *as defined below* $\forall t \geq 1$.

(b) $\rho_{\mathcal{T}}(\hat{\tau}_n, \tau_{n0}) \xrightarrow{p} 0$ *for some nonrandom sequence* $\{\tau_{n0} \in \mathcal{T} : n \geq 1\}$.

(c) $\rho_\Gamma(\hat{\gamma}_n, \gamma_{n0}) \xrightarrow{p} 0$ *for some nonrandom sequence* $\{\gamma_{n0} \in \Gamma : n \geq 1\}$.

(d) $d_n(m,\gamma)$ *is continuous in* (m,γ) *uniformly over* $\mathcal{M} \times \Gamma_0$ *and* $n \geq 1$, *where* $\mathcal{M} = \{m \in R^v : m = \frac{1}{n}\Sigma_1^n Eq_t(Z_t, \beta, \tau)$ *for some* $\beta \in B$, $\tau \in \mathcal{T}_0$, *and* $n \geq 1\}$ *and* $\mathcal{T}_0$ *and* Γ_0 *are some subsets of* $\mathcal{T}$ *and* Γ *that contain* ϵ*–neighborhoods of* $\{\tau_{n0} : n \geq 1\}$ *and* $\{\gamma_{n0} : n \geq 1\}$, *respectively, for some* $\epsilon > 0$.

ASSUMPTION S–CF: W–CF *holds with* $\xrightarrow{p} 0$ *replaced by* $\longrightarrow 0$ a.s.

In the standard case, where $d_n(m,\gamma) = m'\gamma m$ for $m \in R^v$ or $d_n(m,\gamma) = m$ for $m \in R^1$, W–CF(d) holds if $\sup_{n\geq 1} \sup_{\beta \in B} \left\|\frac{1}{n}\Sigma_1^n Eq_t(Z_t, \beta, \tau_{n0})\right\| < \infty$ and $\sup_{n\geq 1}\|\gamma_{n0}\| < \infty$.

Let $q_t(z,\beta,\tau) = (q_{1t}(z,\beta,\tau), \ldots, q_{vt}(z,\beta,\tau))'$. For $j = 1, \ldots, v$, we define:

ASSUMPTION U–WLLN$_j$: $\sup_{(\beta,\tau)\in B\times\mathcal{T}_0} \left|\frac{1}{n}\Sigma_1^n(q_{jt}(Z_t, \beta, \tau) - Eq_{jt}(Z_t, \beta, \tau))\right| \xrightarrow{p} 0$.

ASSUMPTION U–SLLN$_j$: $\sup_{(\beta,\tau)\in B\times\mathcal{T}_0} \left|\frac{1}{n}\Sigma_1^n(q_{jt}(Z_t, \beta, \tau) - Eq_{jt}(Z_t, \beta, \tau))\right| \longrightarrow 0$ a.s.

ASSUMPTION CTY$_j$: $\frac{1}{n}\Sigma_1^n Eq_{jt}(Z_t, \beta, \tau)$ *is continuous in* β *uniformly over* $B \times \mathcal{T}_0$ *and* $n \geq 1$.

LEMMA 7: (a) W–EE, ID, W–CF, U–WLLN$_j$, & CTY$_j$ $\forall j \Rightarrow \rho_B(\hat{\beta}_n, \beta_{n0}) \xrightarrow{p} 0$.

(b) S–EE, ID, S–CF, U–SLLN$_j$, & CTY$_j$ $\forall j \Rightarrow \rho_B(\hat{\beta}_n, \beta_{n0}) \longrightarrow 0$ a.s.

COMMENT: The conditions U–WLLN$_j$, U–SLLN$_j$, and CTY$_j$ for $j = 1, \ldots, v$ can be obtained from primitive assumptions on $\{Z_t : t \geq 1\}$ and $\{q_t(z,\beta,\tau) : t \geq 1\}$ by applying the results of Section 3.

PROOF OF LEMMA 7: By Lemma 6, it suffices to show that W–CF, U–WLLN$_j$, & CTY$_j$ $\forall j \Rightarrow$ U–WCON and that S–CF, U–SLLN$_j$, & CTY$_j$ $\forall j \Rightarrow$ U–SCON, where $Q_n(\beta)$ of U–WCON and U–SCON is defined by $d_n\left[\frac{1}{n}\Sigma_1^n Eq_t(Z_t, \beta, \tau_{n0}), \gamma_{n0}\right]$. Let $Eq_t(Z_t, \beta, \hat{\tau}_n)$ denote $Eq_t(Z_t, \beta, \tau)\big|_{\tau=\hat{\tau}_n}$. U–WCON [U–SCON] follows from

$$\sup_{\beta\in B}|\hat{Q}_n(\beta) - Q_n(\beta)| \leq \sup_{\beta\in B}\left|d_n\left[\tfrac{1}{n}\Sigma_1^n q_t(Z_t,\beta,\hat{\tau}_n), \hat{\gamma}_n\right] - d_n\left[\tfrac{1}{n}\Sigma_1^n Eq_t(Z_t,\beta,\hat{\tau}_n), \hat{\gamma}_n\right]\right|$$

$$+ \sup_{\beta\in B}\left|d_n\left[\tfrac{1}{n}\Sigma_1^n Eq_t(Z_t,\beta,\hat{\tau}_n), \hat{\gamma}_n\right] - d_n\left[\tfrac{1}{n}\Sigma_1^n Eq_t(Z_t,\beta,\tau_{n0}), \gamma_{n0}\right]\right| \qquad (4.2)$$

$$\xrightarrow{p} 0 \quad [\longrightarrow 0 \text{ a.s.}] . \ \square$$

The results of this section all concern extremum estimators $\hat{\beta}_n$ that minimize a criterion function over a parameter space B that does not depend on n. For non-parametric sieve estimators, however, the parameter space depends on n. It is straightforward to generalize the results such that $\hat{\beta}_n$ maximizes a criterion function over a set B_n $\forall n \geq 1$, provided $B_n \uparrow B$ as $n \to \infty$, $\beta_{n0} \in B_n$ $\forall n \geq 1$, and $B = \lim_{n\to\infty} B_n$ is the set that appears in U–WCON, W–SCON, U–WLLN$_j$, U–SLLN$_j$, and CTY$_j$. If $\{B_n\}$ is chosen such that $\beta_{n0} \to \beta_0$ for some $\beta_0 \in B$, then Lemmas 6 and 7 yield $\hat{\beta}_n \xrightarrow{p} \beta_0$ [a.s.].

FOOTNOTES

[1]The author thanks W. K. Newey and Y. J. Whang for helpful comments and suggestions. Also, the author gratefully acknowledges the financial support of the National Science Foundation via grant number SES–8821021.

[2]The weakening of the compactness condition is not tied to the SE approach used here. A similar modification of the first moment continuity (FMC) condition used by Andrews [1] and Pötscher and Prucha [12] would allow their results to be applied to totally bounded parameter spaces. The modification required is that the FMC condition (see Assumption A3 of Andrews [1]) needs to hold uniformly over $\theta \in \Theta$ (i.e., one needs to add $\sup_{\theta \in \Theta}$ before $\lim_{\rho \to 0}$ in Assumption A3). Then, the same argument as in the proof of the Theorem of Andrews [1] can be used to obtain a uniform LLN, noting that $\rho(\theta)$ in (4) does not depend on θ, and hence, that the collection of ρ–balls $\{B(\theta,\rho) : \theta \in \Theta\}$ that covers Θ has a finite subcover under the total boundedness assumption.

[3]Note that it is not possible in general to obtain a uniform convergence or uniform LLN result for a totally bounded parameter space by simply extending the functions involved to the (compact) completion of the parameter space and then applying a uniform result to the larger parameter space. Even if an extension is available, the extended functions may exhibit different properties, such as continuity properties, than the original functions. For example, consider the parameter space $\Theta = [0,1) \cup (1,2]$ and the function $f(\theta) = 1(\theta \in [0,1)) + 2 \times 1(\theta \in (1,2])$. This function is continuous on Θ, but neither of its two natural extensions to $\bar{\Theta} = [0,2]$ is continuous.

[4]Measurability assumptions are only needed in this paper in those situations where the expectation of a sample average must be unique in order for an LLN to hold.

[5]To see this, suppose $q_t(z,\theta)$ is continuous in θ uniformly over $\theta \in \Theta$, $z \in \mathcal{Z}$, and $t \geq 1$. Then, given $\epsilon > 0$ $\exists \delta(\epsilon) > 0$ such that $\sup_{t \geq 1} \sup_{z \in \mathcal{Z}} |q_t(z,\theta') - q_t(z,\theta)| < \epsilon$ $\forall \theta' \in B(\theta, \delta(\epsilon))$, $\forall \theta \in \Theta$. We can take $\delta(\epsilon)$ such that $\delta(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$. Define the inverse function $\delta^{-1}(y) = \inf\{\epsilon > 0 : \delta(\epsilon) \geq y\}$. Note that $\epsilon \leq \delta^{-1}(\delta(\epsilon))$. Now, given any $\theta', \theta \in \Theta$, take $\epsilon > 0$ such that $d(\theta',\theta) \in [\delta(\epsilon/2), \delta(2\epsilon)]$. Then, $\sup_{t \geq 1} \sup_{z \in \mathcal{Z}} |q_t(z,\theta') - q_t(z,\theta)| < 2\epsilon \leq 4\delta^{-1}(\delta(\epsilon/2)) \leq 4\delta^{-1}(d(\theta',\theta))$. Hence, W–LIP and S–LIP hold with $B_t(Z_t) = 4$ and $h(y) = \delta^{-1}(y)$.

[6]As stated, Lemma 3 of Bierens [6] only assumes proper convergence of $\bar{\mu}_n$ to μ. As noted in Bierens [7, Remark 2 to Thm. 2.6], however, proper setwise convergence is actually needed, as in TSE–1B.

REFERENCES

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

[2] Andrews, D. W. K. (1988). "Laws of Large Numbers for Dependent non–identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

[3] Andrews, D. W. K. (1989). "Asymptotics for Semiparametric Econometric Models: I. Estimation," Cowles Foundation Discussion Paper No. 908, Yale University.

[4] Andrews, D. W. K. (1989). "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity," Cowles Foundation Discussion Paper No. 909, Yale University.

[5] Andrews, D. W. K. (1989). "Asymptotics for Semiparametric Econometric Models: III. Testing and Examples," Cowles Foundation Discussion Paper No. 910, Yale University.

[6] Bierens, H. J. (1984). "Model Specification Testing of Time Series Regressions," *Journal of Econometrics*, 26, 323–353.

[7] Bierens, H. J. (1989). "Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Heterogeneity," unpublished manuscript, Department of Econometrics, Free University, Amsterdam.

[8] Jennrich, R. I. (1969). "Asymptotic Properties of Non–linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

[9] McLeish, D. L. (1975). "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 829–839.

[10] Newey, W. K. (1989). "Uniform Convergence in Probability and Stochastic Equicontinuity," unpublished manuscript, Department of Economics, Princeton University.

[11] Pollard, D. (1989). *Empirical Processes: Theory and Applications.* Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics: Philadelphia, manuscript in preparation.

[12] Pötscher, B. M. and I. R. Prucha (1989). "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica*, 57, 675–683.

[13] Pötscher, B. M. and I. R. Prucha (1989). "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts," unpublished manuscript, Department of Economics, University of Maryland.