CHAPTER 13

Bracketing methods in statistics and econometrics

David Pollard

1 Introduction

Many asymptotic results depend upon uniform limit theorems for sequences of stochastic processes indexed by a parameter set Θ. The best-known theorems of this type give conditions under which an average,

$$G_n(\omega,\theta) = n^{-1} \sum_{i \le n} g_i(\omega,\theta),$$

lies uniformly close to its expected value,

$$M_n(\theta) = \mathbb{P}G_n(\cdot,\theta) = n^{-1} \sum_{i \le n} \mathbb{P}g_i(\cdot,\theta).$$

For example, a uniform strong law of large numbers (SLLN) is an assertion that

$$\sup_\theta |G_n(\omega,\theta) - M_n(\theta)| \to 0 \text{ almost surely.} \tag{1}$$

This chapter explores one method, which is known in the empirical process literature as *bracketing,* for proving uniform SLLN's and other, sharper, uniform convergence theorems. Emphasis is placed on those one-sided results that are useful for establishing convergence properties of estimators defined by maximization of G_n over Θ. Section 2 explains why only one-sided bounds are needed.

The basic idea of (two-sided) bracketing is that uniform bounds for G_n can be obtained by sandwiching it between two simpler processes,

$$A_n(\omega,\theta) \le G_n(\omega,\theta) \le B_n(\omega,\theta) \quad \text{for all } \omega,\theta,$$

whose limit behaviors are easy to establish. Typically the bracketing processes are tractable because they are *π-simple* for a fixed finite partition

I am grateful to the audience at the meeting, and in particular to Professor Bierens, for most stimulating comments, questions, and criticisms.
Research supported in part by NSF Grants DMS-8503347 and DMS-8806900.

Reprinted from *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics*, edited by William A. Barnett, James Powell and George E. Tauchen.

$\pi = \{\Theta(1), \ldots, \Theta(m)\}$ of Θ. (A process $X(\omega, \theta)$ will be called π-simple if it has the representation

$$X(\omega, \theta) = \sum_j x_j(\omega)\{\theta \in \Theta(j)\}$$

for random variables $x_1, \ldots, x_m$ that do not depend on θ.) Uniform limit theorems for the bracketing process then follow from results for sequences of real random variables applied finitely many times.

Section 3 summarizes how one-sided bracketing can be used to prove one-sided uniform SLLN's, with particular attention to the problems caused by non-identically distributed summands g_i.

Section 4 discusses in detail how a recursive model, which is known in the empirical process literature as *chaining,* can be applied under bracketing assumptions to derive uniform bounds useful for establishing $n^{-1/2}$ rates of convergence. The main idea is that complexity of the bracketing process, as measured by the number of regions in the underlying finite partition of Θ, should not increase too rapidly as the quality of the bracketing approximation improves. The discussion reinterprets several techniques from the recent empirical process literature.

2 Consistency and $n^{-1/2}$ consistency

Consider an estimator T_n defined by maximization of G_n:

$$G_n(\omega, T_n(\omega)) = \sup_\theta G_n(\omega, \theta). \tag{2}$$

That is, T_n is the argument that maximizes G_n (the *argmax*). For convenience I ignore all difficulties related to existence, uniqueness, or measurability of T_n, because their resolution has only a minor effect upon the details of the bracketing argument. Also, for notational convenience, I assume Θ is a metric space.

Suppose we need to show that T_n lies close to some point θ_n in Θ. Often θ_n will be a fixed point in Θ, especially when the $\{g_i(\cdot, \theta)\}$ are assumed stationary. But, as emphasized by Domowitz and White (1982) and Bates and White (1985), there is virtue in allowing θ_n to wander. To force T_n into the ball $B(n)$ of radius r and center θ_n, we need G_n smaller outside $B(n)$ than inside it. More precisely, it suffices to find a process H_n for which, with probability 1, it is eventually true that

$$G_n(\omega, \theta) \le H_n(\omega, \theta) < G_n(\omega, \theta_n) \quad \text{for all } \theta \text{ in } B(n)^c. \tag{3}$$

Because $G_n(\omega, T_n(\omega)) \ge G_n(\omega, \theta_n)$ these uniform inequalities exclude values outside $B(n)$ as possibilities for T_n.

Usually H_n is a deterministic function constructed from the expected value function M_n. If each M_n has a clean maximum at θ_n, in the uniform sense that there exists an $\epsilon > 0$ for which

$$M_n(\theta_n) \geq \epsilon + M_n(\theta) \quad \text{for all } \theta \text{ in } B(n)^c, \text{ all } n,$$

then it is enough to show that $G_n(\theta_n) - M_n(\theta_n)$ converges to zero almost surely and also that a one-sided analogue of (1) holds:

$$\sup_\theta [G_n(\theta) - M_n(\theta)] \to 0 \text{ almost surely.} \tag{4}$$

The choice $H_n(\theta) = M_n(\theta) + \epsilon/3$ then gives (3). (Existence of a uniformly clean maximum for $M_n(\theta)$ corresponds to the property Domowitz and White 1982 call "identifiable uniqueness.") As the next example shows, this is not always the best way to construct H_n.

Example 1

Let $\{\xi_i\}$ be a sequence of independent random variables, each taking the value 1 with probability θ_0 and the value 0 with probability $1-\theta_0$, for some θ_0 in the open interval $(0, 1)$. The maximum likelihood estimator is chosen to maximize the G_n corresponding to

$$g_i(\omega, \theta) = \xi_i(\omega) \log \theta + (1 - \xi_i(\omega)) \log(1-\theta).$$

That is,

$$G_n(\omega, \theta) = X_n(\omega) \log \theta + (1 - X_n(\omega)) \log(1-\theta),$$

where X_n denotes the average of the first n observations. The expected value,

$$M_n(\theta) = \theta_0 \log \theta + (1-\theta_0) \log(1-\theta),$$

has a unique maximum at θ_0. The uniform SLNN fails because of bad behavior of $\log \theta$ at 0 and of $\log(1-\theta)$ at 1. Nevertheless, the argument based on (3) works for

$$H_n(\omega, \theta) = X_n(\omega) \log X_n(\omega) + [1 - X_n(\omega)] \log[1 - X_n(\omega)] - [X_n(\omega) - \theta]^2$$

because

$$G_n(\omega, \theta) \leq H_n(\omega, \theta) \quad \text{for all } \theta$$

and $H_n(\omega, \theta)$ converges with probability 1 to

$$\theta_0 \log \theta_0 + (1-\theta_0) \log(1-\theta_0) - (\theta - \theta_0)^2,$$

uniformly in θ. ■

A consistency result of the type just described is often merely the first step toward a limit theorem for the argmax of G_n. Subsequent steps require only bounds that hold in probability, rather than almost surely, but otherwise the argument is the same as the one based on (3).

To simplify the notation, from now on I will omit ω, writing $g_i(\theta)$ instead of $g_i(\omega, \theta)$, and so on.

Assume that consistency has already been established for T_n. Then there exists a sequence $\{\delta_n\}$ of positive numbers converging to zero for which

$$\mathbb{P}\{d(T_n, \theta_n) > \delta_n\} \to 0.$$

For $n^{-1/2}$ consistency one must also exclude T_n from an annular region

$$A(n) = \{\theta : Kn^{-1/2} < d(\theta, \theta_n) \leq \delta_n\}.$$

More precisely, for each $\epsilon > 0$ one needs to determine a constant K for which

$$\mathbb{P}\{T_n \in A(n)\} < \epsilon \ \text{ eventually.}$$

For this it is enough to find processes H_n for which it is eventually true with probability at least $1-\epsilon$ that

$$G_n(\theta) \leq H_n(\theta) < G_n(\theta_n) \quad \text{for all } \theta \text{ in } A(n). \tag{5}$$

These inequalities force T_n outside $A(n)$.

Construction of an appropriate H_n usually depends upon a strengthening of the existence of a uniformly clean maximum for M_n, made possible by a Taylor expansion of M_n about its maximizing value. Suppose that there exists a constant $c > 0$ such that, for all n,

$$M_n(\theta) \leq M_n(\theta_n) - cd(\theta, \theta_n)^2 \quad \text{for } \theta \text{ in } A(n). \tag{6}$$

Define

$$h_i(\theta) = \frac{g_i(\theta) - g_i(\theta_n)}{d(\theta, \theta_n)}.$$

Then, at least for θ in $A(n)$,

$$\begin{aligned} &G_n(\theta) - G_n(\theta_n) \\ &\quad = M_n(\theta) - M_n(\theta_n) + n^{-1} \sum_{i \leq n} [g_i(\theta) - g_i(\theta_n) - \mathbb{P}g_i(\theta) + \mathbb{P}g_i(\theta_n)] \\ &\quad \leq -cd(\theta, \theta_n)^2 + d(\theta, \theta_n) n^{-1} \sum_{i \leq n} [h_i(\theta) - \mathbb{P}h_i(\theta)]. \end{aligned}$$

That suggests we define

$$H_n(\theta) = G_n(\theta_n) - cd(\theta, \theta_n)^2 + d(\theta, \theta_n) n^{-1} \sum_{i \leq n} [h_i(\theta) - \mathbb{P}h_i(\theta)].$$

Then it remains only to find a constant K for which, eventually,

$$\mathbb{P}\{H_n(\theta) < G_n(\theta_n) \text{ for all } \theta \text{ in } A(n)\} > 1 - \epsilon.$$

Equivalently, a constant K (used to define $A(n)$) is needed for which, eventually,

$$\mathbb{P}\left\{\exists\theta\in A(n): n^{-1}\sum_{i\le n}[h_i(\theta)-\mathbb{P}h_i(\theta)]>cd(\theta,\theta_n)\right\}<\epsilon. \tag{7}$$

Because $d(\theta,\theta_n)\ge Kn^{-1/2}$ on $A(n)$, it would be enough to find constants C and K for which, eventually,

$$\mathbb{P}\left\{\exists\theta\in A(n): n^{-1/2}\sum_{i\le n}[g_i(\theta)-g_i(\theta_n)-\mathbb{P}g_i(\theta)+\mathbb{P}g_i(\theta_n)] > Cd(\theta,\theta_n)\right\}<\epsilon. \tag{8}$$

Increasing K, if necessary, to ensure that $cK>C$, we recover (7). The form (8) is well suited to the methods that will be described in Section 4.

Requirement (7) looks like a weak uniform law of large numbers for the random process

$$n^{-1}\sum_{i\le n}\frac{h_i(\theta)-\mathbb{P}h_i(\theta)}{d(\theta,\theta_n)}.$$

Indeed, when h_i is sufficiently smooth (7) does succumb to an application of such a uniform law. However, there are nontrivial problems where this approach fails.

Example 2

Suppose $y_i=x_i\beta_0+e_i$ with the $\{x_i\}$ a deterministic sequence of row vectors and the $\{e_i\}$ independent, identically distributed with a unique median at zero. Reparameterize by setting $\theta=\beta-\beta_0$. The θ_n of interest is the zero vector. The least absolute deviations (LAD) estimator corresponds to the argmax for the G_n constructed from

$$g_i(\theta)=|e_i|-|e_i-x_i\theta|.$$

The assumption about the median of e_i implies that the function

$$M(t)=\mathbb{P}(|e_1|-|e_1-t|)$$

has a unique maximum of zero at $t=0$. Suppose the distribution of e_1 has a continuous, positive density at the origin. Then M has a nonzero second derivative at $t=0$; and for each finite K there is an $\epsilon>0$ for which

$$M(t)\le-\epsilon|t|^2 \quad\text{for}\quad |t|\le K.$$

Condition (6) will be satisfied by

$$M_n(\theta)=n^{-1}\sum_{i\le n}M(x_i\theta)$$

if there exists a constant K such that the matrices

$$n^{-1}\sum_{i\le n} x_i'x_i\{|x_i|\le K\}$$

are uniformly positive definite, in the sense that the sequence of smallest eigenvalues is bounded away from zero. Such a requirement would be satisfied if the $\{x_i\}$ were a realization of an independent sample from a distribution with a nonsingular covariance matrix.

If (6) holds, the problem of establishing an $n^{-1/2}$ rate of convergence for the LAD estimator can be reduced to a uniformity result for the process

$$n^{-1}\sum_{i\le n}\frac{|e_i|-|e_i-x_i\theta|-M(x_i\theta)}{|\theta|^2}.$$

The absolute value function is not smooth enough at the origin to allow a two-term Taylor expansion, which might have reduced analysis of this process to a simple application of a uniform law of large numbers. However, it is possible to exploit what smoothness the absolute value function does have to establish an inequality like (8).

Define D_i as $\{e_i\ge 0\}-\{e_i<0\}$, a difference of indicator functions. Notice that $\mathbb{P}D_i=0$. It plays the role that would normally belong to a first derivative at zero. Define

$$r_i(\theta)=|e_i|-|e_i-x_i\theta|-D_ix_i\theta.$$

It is enough to find a constant C for which, eventually,

$$\mathbb{P}\left\{\exists\theta\in A(n):n^{-1/2}\sum_{i\le n} D_ix_i\theta+r_i(\theta)-\mathbb{P}r_i(\theta)>C|\theta|\right\}<\epsilon.$$

Mild constraints on the $\{x_i\}$ are enough to control the contribution from the $\{D_i\}$. For example, if

$$\limsup n^{-1}\sum_{i\le n}|x_i|^\alpha<\infty \quad \text{for some } \alpha>2, \tag{9}$$

then $n^{-1/2}\sum D_ix_i$ will be of order $O_p(1)$. Section 4 presents methods to handle the contribution from the $\{r_i(\theta)\}$ and continues the discussion in Example 3. ■

3 Uniform SLLN

A one-sided uniform SLLN as in (4) can often be established by means of a one-sided bracketing based on a finite partition $\pi=\{\Theta(1),\dots,\Theta(m)\}$ of Θ. The natural π-simple upper bracket, $B_n(\omega,\theta)$, would be defined by

$$B_n(\omega,\theta)=n^{-1}\sum_{i\le n} b_i(\omega,\theta)$$

with

$$b_i(\omega,\theta)=\sum_j \sup_{\Theta(j)} g_i(\omega,\theta')\{\theta\in\Theta(j)\}.$$

For fixed θ, one needs either to assume, or to deduce from lower-level assumptions (such as independence, stationarity plus ergodicity, or some sort of mixing), that $B_n(\cdot,\theta)$ satisfies a SLLN: for each θ,

$$B_n(\omega,\theta)-\mathbb{P}B_n(\cdot,\theta)\to 0 \text{ almost surely.}$$

Applied finitely many times, to a representative set of values $\theta_1,\ldots,\theta_m$ from each region, this immediately implies a uniform SLLN for B_n:

$$\sup_\theta |B_n(\omega,\theta)-\mathbb{P}B_n(\cdot,\theta)|\to 0 \text{ almost surely.}$$

To complete the proof of (4) it is then enough to show that for each $\epsilon>0$ there is a partition π for which

$$\mathbb{P}B_n(\cdot,\theta_j)\le\epsilon+\inf_{\Theta(j)} M_n(\theta) \text{ eventually}$$

for each j. Or, written out in terms of the g_i,

$$n^{-1}\sum_{i\le n}\mathbb{P}\sup_{\Theta(j)}[g_i(\cdot,\theta)-g_i(\cdot,\theta_j)]\le\epsilon+\inf_{\Theta(j)} n^{-1}\sum_{i\le n}\mathbb{P}[g_i(\cdot,\theta)-g_i(\cdot,\theta_j)].$$

Clearly the left-hand side requires more delicate handling than the right-hand side, because the supremum is inside the expectation and the infimum is outside. However, in the literature it is common to have both quantities handled by means of two-sided bounds on the increments of g_i. Such an approach concedes nothing if the ultimate aim is a two-sided uniform SLLN. In this chapter I discuss only the arguments needed to make the left-hand side small.

When Θ is compact, an appropriate partition can be constructed from a finite covering of Θ by open sets if, for each θ^* in Θ, there is a neighborhood N for which, eventually,

$$n^{-1}\sum_{i\le n}\mathbb{P}\sup_N[g_i(\cdot,\theta)-g_i(\cdot,\theta^*)]<\epsilon. \tag{10}$$

The basic idea for finding such an N, due to Wald (1949), is to combine pointwise upper semicontinuity at θ^* with a domination condition. Let $\{N(k)\}$ be a sequence of neighborhoods that shrinks to θ^*. Define

$$\Delta_{ik}(\omega)=\sup_{N(k)}[g_i(\omega,\theta)-g_i(\omega,\theta^*)].$$

For each i, almost sure upper semicontinuity would give $\Delta_{ik}(\omega)\to 0$ as $k\to\infty$, almost surely. If the g_i all had the same distribution, and if $\mathbb{P}\Delta_{1k}$ were finite for some k, then a dominated convergence argument would give, for all i,

$$\mathbb{P}\Delta_{ik} = \mathbb{P}\Delta_{1k} \to 0 \quad \text{as } k \to \infty.$$

An $N(k)$ with k large enough would suffice for (10).

For nonidentically distributed g_i some sort of uniformity in the upper semicontinuity is needed to ensure that

$$\limsup_n n^{-1} \sum_{i \le n} \mathbb{P}\Delta_{ik} < \epsilon \quad \text{for some } k. \tag{11}$$

Hoadley (1971) gave conditions, which, for a one-sided bound, correspond to

(i) $\limsup_k \sup_i \Delta_{ik}(\omega) = 0$ almost surely;

(ii) $\sup_i \mathbb{P}\Delta_{i0}^{1+\delta} < \infty$ for some $\delta > 0$.

If we write Ω_k for the set $\{\sup_i \Delta_{ik} > \frac{1}{2}\epsilon\}$ then condition (i) implies that $\mathbb{P}\Omega_k \to 0$. Condition (ii) makes the $\{\Delta_{i0}\}$ uniformly integrable, and hence the sequence of their averages is also uniformly integrable. Inequality (11) then follows from the bound

$$n^{-1} \sum_{i \le n} \mathbb{P}\Delta_{ik} \le \frac{1}{2}\epsilon + \mathbb{P}\left(\Omega_k n^{-1} \sum_{i \le n} \Delta_{i0}\right).$$

Both Andrews (1987) and Potscher and Prucha (1987) have pointed out that requirement (i) excludes such simple applications as ordinary least squares, except under restrictive assumptions. For if $y_i = x_i\theta_0 + e_i$, then for a neighborhood $N(k)$ of radius τ_k about θ_0,

$$2|x_i| \cdot |e_i| \tau_k \ge \Delta_{ik} \ge \min(|e_i|^2, |x_i| \cdot |e_i| \tau_k).$$

If $\limsup \min(|x_i|, |e_i|) = \infty$ with positive probability, then requirement (i) is violated.

The remedy suggested by Andrews (1987) corresponds, for one-sided results, to existence of nonnegative random variables $\{D_i\}$ for which

(i) $\Delta_{ik}(\omega) \le D_i(\omega) h_k$ for some sequence of constants $\{h_k\}$ that converges to zero;

(ii) $\limsup n^{-1} \sum_{i \le n} \mathbb{P}D_i < \infty$.

Inequality (11) follows easily from these assumptions. Potscher and Prucha (1987) suggested a more delicate modification of the Hoadley conditions. For one-sided results their conditions correspond to existence, for each $\epsilon > 0$, of subsets $A_i(\epsilon)$ of Ω for which

(i) $\limsup_k \sup_i \Delta_{ik}(\omega)\{\omega \in A_i(\epsilon)\} = 0$ almost surely;

(ii) $\limsup_n n^{-1} \sum_{i \le n} \mathbb{P}\Delta_{i0}(\omega)\{\omega \in A_i(\epsilon)^c\} < \frac{1}{3}\epsilon$;

(iii) the sequence $Z_n(\omega, \epsilon) = n^{-1}\sum_{i\le n} \Delta_{i0}(\omega)\{\omega \in A_i(\epsilon)\}$ is uniformly integrable.

If we write $\Omega_k(\epsilon)$ for the set $\{\sup_i \Delta_{ik} A_i(\epsilon) > \frac{1}{3}\epsilon\}$ then inequality (11) follows from these assumptions via the bound

$$n^{-1} \sum_{i\le n} \mathbb{P}\Delta_{ik} \le \frac{1}{3}\epsilon + \mathbb{P}Z_n(\cdot, \epsilon)\Omega_k(\epsilon) + n^{-1} \sum_{i\le n} \mathbb{P}\Delta_{i0} A_i(\epsilon)^c.$$

For the least squares example, weak moment assumptions would allow us to satisfy these requirements by choosing

$$A_i(\epsilon) = \{|x_i| \le M, |e_i| \le M\}$$

for a suitably large constant M.

4 Chaining

This section explains how repeated application of a bracketing argument can provide bounds for

$$\mathbb{P}\left\{\sup_\theta n^{-1/2} \sum_{i\le n} [f_i(\theta) - \mathbb{P}f_i(\theta)] > \eta\right\}. \tag{12}$$

The processes $f_i(\theta)$, or $f_i(\omega, \theta)$ with the dependence on the underlying probability space made explicit, are assumed mutually independent. As will be shown by Example 3, such a bound can be useful for establishing $n^{-1/2}$ rates of convergence for argmax estimators.

The argument will be based on the methods of Ossiander (1987) and Andersen, Giné, Ossiander, and Zinn (1988). The improvement over the results from Section 3 will be made possible by an exponential inequality for sums of bounded, independent random variables. The inequality is easily deduced from Bernstein's inequality or the more general Bennett inequality (Shorack and Wellner 1986, p. 852).

Inequality. *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables, all bounded above by a constant M. If $V \ge \mathbb{P}\xi_1^2 + \cdots + \mathbb{P}\xi_n^2$ then*

$$\mathbb{P}\left\{\sum_i \xi_i \ge t + \sum_i \mathbb{P}\xi_i\right\} \le \exp\left(\frac{-\frac{1}{4}t^2}{V}\right) \quad \text{for } 0 \le t \le \frac{V}{M}. \tag{13}$$

■

The bracketing will be expressed in a form slightly different from the one in Section 3. The upper bound B will be decomposed into an approximating process Φ plus a bounding process U. Instead of just one bracketing approximation, there will be a sequence of approximations linked together using a chaining argument.

To bound the contributions made by successive approximations we will need to control the rate at which the number of regions in a bracketing increases as the error of approximation decreases.

Definition 1. *For each $\epsilon > 0$ let $N(\epsilon)$ be the smallest number of regions in a partition π of Θ for which there exist π-simple processes $\phi_i(\theta)$ and $u_i(\theta) > 0$ with*

(i) $f_i(\theta) \le \phi_i(\theta) + u_i(\theta)$ *for all* θ;
(ii) $n^{-1}\sum_{i\le n} \mathbb{P}|f_i(\theta) - \phi_i(\theta)|^2 \le \epsilon^2$ *for all* θ;
(iii) $n^{-1}\sum_{i\le n} \mathbb{P}u_i(\theta)^2 \le \epsilon^2$ *for all* θ;
(iv) *the processes* $(\phi_i(\theta), u_i(\theta))$ *for* $i = 1, \dots, N(\epsilon)$ *are independent.*

Call $N(\epsilon)$ the covering number for the upper bracketing. ■

Notice that (ii) does not follow from (iii), because (i) only sets an upper bound on the difference $f_i(\theta) - \phi_i(\theta)$.

The argument will require not only that $N(\epsilon)$ be finite for each $\epsilon > 0$, but also that

$$\int_0^1 \sqrt{\log N(x)}\, dx < \infty. \tag{14}$$

Notice that $N(\epsilon) \uparrow \infty$ as $\epsilon \downarrow 0$. The integrability condition stops $N(\cdot)$ from growing too rapidly.

Our bound will involve a second positive, decreasing function $\lambda(\cdot)$. We will require that it grows fast enough to ensure

$$\int_0^1 x^{-1} \exp(-\lambda(x))\, dx < \infty, \tag{15}$$

but not so fast as to violate

$$\int_0^1 \sqrt{\lambda(x)}\, dx < \infty. \tag{16}$$

Quite often one would choose $\lambda(\cdot)$ as a multiple of $\log N(\cdot)$; a more rapidly increasing λ will correspond to a larger η in (12).

Theorem 1. *Suppose $N(\cdot)$ and $\lambda(\cdot)$ satisfy the integrability conditions (14)–(16). Then there exist universal constants C_1, C_2, and a function $\epsilon(\cdot)$ such that*

$$\mathbb{P}\left\{\sup_\theta n^{-1/2} \sum_{i\le n} [f_i(\theta) - \mathbb{P}f_i(\theta)] > C_1 J(\delta)\right\}$$
$$\le C_2 B(\delta) + \sum_{i\le n} \mathbb{P}\{F_i > n^{1/2}\epsilon(\delta)\},$$

where

$$F_i \geq \sup_\theta f_i(\theta)^+,$$

$$\delta^2 \geq \max\left[n^{-1}\sum_{i\leq n} \mathbb{P}F_i^2, \sup_\theta n^{-1}\sum_{i\leq n} \mathbb{P}f_i(\theta)^2\right],$$

$$J(\delta) = \int_0^\delta [\lambda(x) + \log N(x)]^{1/2}\,dx,$$

$$B(\delta) = \int_0^\delta x^{-1}\exp[-\lambda(x)]\,dx,$$

$$\epsilon(\delta) = \frac{\frac{1}{2}\delta^2}{J(\delta)}.$$

Proof: Define $\delta_k = \delta/2^k$ for $k = 0, 1, \ldots$. Let π_k be a partition of Θ into $N_k = N(\delta_k)$ regions. Add an extra subscript k to identify the processes corresponding to $\epsilon = \delta_k$ in Definition 1. That is, for each k let the π_k-simple processes $\phi_{ik}(\theta)$ and $u_{ik}(\theta) \geq 0$ satisfy the following conditions:

$$f_i(\theta) \leq \phi_{ik}(\theta) + u_{ik}(\theta) \quad \text{for all } \theta; \tag{17}$$

$$n^{-1}\sum_{i\leq n} \mathbb{P}|f_i(\theta) - \phi_{ik}(\theta)|^2 \leq \delta_k^2 \quad \text{for all } \theta; \tag{18}$$

$$n^{-1}\sum_{i\leq n} \mathbb{P}u_{ik}(\theta)^2 \leq \delta_k^2 \quad \text{for all } \theta; \tag{19}$$

the $(\phi_{ik}(\theta), u_{ik}(\theta))$ processes for $i = 1, \ldots, N_k$ are independent. (20)

We may assume that π_0 consists of a single region (the whole of Θ), that $\phi_{i0}(\theta) \equiv 0$, and that

$$u_{i0}(\theta) = F_i = \sup_\theta f_i(\theta)^+. \tag{21}$$

Inequality (13) and Minkowski's inequality imply that the approximating ϕ_{ik} processes satisfy

$$n^{-1}\sum_{i\leq n} \mathbb{P}|\phi_{ik}(\theta) - \phi_{i,k+1}(\theta)|^2 \leq 4\delta_k^2 \quad \text{for all } \theta. \tag{22}$$

Soon we will also need an upper bound for $\phi_{i,k+1} - \phi_{ik}$. If the π_k partitions are nested – that is, if π_{k+1} is a refinement of π_k, for each k – this can be arranged by substituting the minimum of $\phi_{i,k+1}(\theta)$ and $\phi_{ik}(\theta) + u_{ik}(\theta)$ for $\phi_{i,k+1}(\theta)$, without disturbing any of (17) to (20) or the fact that $\phi_{i,k+1}$ is π_{k+1}-simple. The substitution ensures that

$$\phi_{i,k+1}(\theta) - \phi_{i,k}(\theta) \leq u_{ik}(\theta) \quad \text{for all } \theta. \tag{23}$$

The added complications when the partitions are not nested will be disposed of later. In the meantime assume they are nested.

We will obtain the asserted maximal inequality by means of a recursive argument, which will relate successive approximations for the centered sum of the $f_i(\theta)$ processes. To bound the difference between the kth and $(k+1)$th approximation we introduce a truncation of the summands – the key idea behind the method of Ossiander (1987) – so that inequality (13) can apply. The truncations will be determined by decreasing sequences of truncating sets,

$$D_{ik}(\theta) = \{u_{ij}(\theta) \le n^{1/2}\alpha_j \text{ for all } j \le k\},$$

for a sequence $\{\alpha_k\}$ of positive constants that will be specified soon. Notice that $D_{i0}(\theta) = \{F_i \le n^{1/2}\alpha_0\}$.

First we will argue recursively to obtain a probabilistic upper bound for the centered, truncated sums

$$\sum_{i \le n} [f_i(\theta)D_{i0}(\theta) - \mathbb{P}f_i(\theta)D_{i0}(\theta)].$$

Then we will handle the difference between this sum and the centered sum in (12).

The recursive step will depend upon a decomposition of the errors of approximation for the truncated sums into three parts. If we define

$$F_k(\theta) = \sum_{i \le n} f_i(\theta)D_{ik}(\theta) - \mathbb{P}f_i(\theta)D_{ik}(\theta),$$

$$\Phi_k(\theta) = \sum_{i \le n} \phi_{ik}(\theta)D_{ik}(\theta) - \mathbb{P}\phi_{ik}D_{ik}(\theta),$$

$$\Psi_{k+1}(\theta) = \sum_{i \le n} \phi_{ik}(\theta)D_{i,k+1}(\theta) - \mathbb{P}\phi_{ik}(\theta)D_{i,k+1}(\theta),$$

then

$$\begin{aligned} F_k(\theta) - \Phi_k(\theta) = {} & (F_{k+1}(\theta) - \Phi_{k+1}(\theta)) + (\Phi_{k+1}(\theta) - \Psi_{k+1}(\theta)) \\ & + (F_k(\theta) - F_{k+1}(\theta) - \Phi_k(\theta) + \Psi_{k+1}(\theta)). \end{aligned} \tag{24}$$

Notice that $\Phi_0(\theta) \equiv 0$. Define probabilities

$$P_k = \mathbb{P}\{\sup_\theta F_k(\theta) - \Phi_k(\theta) > n^{1/2}R_k\},$$

where $\{R_k\}$ is a decreasing sequence of constants satisfying

$$R_k \to R_\infty = J(\delta) \text{ as } k \to \infty,$$

$$R_k = R_{k+1} + \beta_k + \gamma_k.$$

The positive constants β_k and γ_k will be specified soon.

We will get a recursive bound for P_k by apportioning the three components of R_{k+1} among the three terms of the right-hand side of (24):

$$P_k \le P_{k+1} + \mathbb{P}\{\sup_\theta \Phi_{k+1}(\theta) - \Psi_{k+1}(\theta) > n^{1/2}\beta_k\}$$
$$+ \mathbb{P}\{\sup_\theta F_k(\theta) - F_{k+1}(\theta) - \Phi_k(\theta) + \Psi_{k+1}(\theta) > n^{1/2}\gamma_k\}. \tag{25}$$

Consider first the middle term on the right-hand side. The difference $\Phi_{k+1} - \Psi_{k+1}$ is a sum of independent π_{k+1}-simple random processes,

$$[\phi_{i,k+1}(\theta) - \phi_{i,k}(\theta)]D_{i,k+1}(\theta),$$

minus their expected values. From (26), these processes are bounded above by $u_{ik}(\theta)$, which is less than $n^{1/2}\alpha_k$ on $D_{i,k+1}(\theta)$; and (25) bounds the sum of second moments by $4n\delta_k^2$. Inequality (13) will apply if $n^{1/2}\beta_k$ is less than $4n\delta_k^2/n^{1/2}\alpha_k$ - that is, if $\alpha_k\beta_k \le 4\delta_k^2$. Adding up the exponential bounds over a set of N_{k+1} representative θ values, we get, if the last inequality is satisfied,

$$\mathbb{P}\{\sup_\theta \Phi_{k+1}(\theta) - \Psi_{k+1}(\theta) > n^{1/2}\beta_k\} \le N_{k+1}\exp\left(\frac{-\frac{1}{4}\beta_k^2}{4\delta_k^2}\right).$$

For the last term on the right-hand side of (26), write the process as a sum of independent random processes,

$$[f_i(\theta) - \phi_{ik}(\theta)]D_{ik}(\theta)D_{i,k+1}(\theta)^c,$$

minus their expected values. These processes are bounded above by $u_{ik}(\theta)D_{ik}(\theta)D_{i,k+1}(\theta)^c$, which is less than $n^{1/2}\alpha_k$. Their expected values can be bounded below by means of the Cauchy–Schwarz inequality. The argument works for any random variables $\{Z_i\}$ with finite second moments. It depends only on (19) and the fact that $u_{im}(\theta) > n^{1/2}\alpha_m$ on $D_{im}(\theta)^c$:

$$\left|n^{-1}\sum_{i\le n} \mathbb{P}Z_i D_{im}(\theta)^c\right| \le \left(n^{-1}\sum_{i\le n}\mathbb{P}Z_i^2\right)^{1/2}\left(n^{-1}\sum_{i\le n}\mathbb{P}u_{im}(\theta)^2\right)^{1/2}\Big/(n^{1/2}\alpha_m)$$
$$\le \left(n^{-1/2}\frac{\delta_m}{\alpha_m}\right)\left(n^{-1}\sum_{i\le n}\mathbb{P}Z_i^2\right)^{1/2}. \tag{26}$$

Applied with $m = k+1$ and $Z_i = |f_i(\theta) - \phi_{ik}(\theta)|D_{ik}(\theta)$, inequality (26), together with (18), gives

$$n^{-1}\sum_{i\le n}\mathbb{P}|f_i(\theta) - \phi_{ik}(\theta)|D_{ik}(\theta)D_{i,k+1}(\theta)^c \le n^{-1/2}\frac{\delta_k\delta_{k+1}}{\alpha_{k+1}}.$$

Applied with $m = k+1$ and $Z_i = u_{ik}(\theta)D_{ik}(\theta)$, inequality (26), together with (19), gives

$$n^{-1}\sum_{i\le n}\mathbb{P}u_{ik}(\theta)D_{ik}(\theta)D_{i,k+1}(\theta)^c \le n^{-1/2}\frac{\delta_k\delta_{k+1}}{\alpha_{k+1}}.$$

Replace $\delta_k\delta_{k+1}$ by $2\delta_{k+1}^2$. Then

$$
\begin{aligned}
&\mathbb{P}\{\sup_\theta F_k(\theta)-F_{k+1}(\theta)-\Phi_k(\theta)-\Phi_{k+1}(\theta)>n^{1/2}\gamma_k\}\\
&\quad\le \mathbb{P}\left\{\sup_\theta \sum_{i\le n}[f_i(\theta)-\phi_{ik}(\theta)]D_{ik}(\theta)D_{i,k+1}(\theta)^c>n^{1/2}\gamma_k-2n^{1/2}\frac{\delta_{k+1}^2}{\alpha_{k+1}}\right\}\\
&\quad\le \mathbb{P}\left\{\sup_\theta \sum_{i\le n}[u_{ik}(\theta)D_{ik}(\theta)D_{i,k+1}(\theta)^c-\mathbb{P}u_{ik}(\theta)D_{ik}(\theta)D_{i,k+1}(\theta)^c]\right.\\
&\qquad\qquad\qquad\qquad\left.>n^{1/2}\left(\gamma_k-\frac{4\delta_{k+1}^2}{\alpha_{k+1}}\right)\right\}.
\end{aligned}
$$

Inequality (13) applies to each of the N_{k+1} representative θ values if the constants satisfy the inequality

$$n^{1/2}\left(\gamma_k-\frac{4\delta_{k+1}^2}{\alpha_{k+1}}\right)\le\frac{n\delta_k^2}{n^{1/2}\alpha_k}.$$

The constraints on the $\alpha_k, \beta_k, \gamma_k$ are simpler if expressed in terms of $\eta_k=\delta_k^2/\alpha_k$. We need

$$
\begin{aligned}
&\alpha_k=\frac{\delta_k^2}{\eta_k},\\
&\beta_k\le 4\eta_k,\\
&\gamma_k\le 4\eta_{k+1}+\eta_k.
\end{aligned}
$$

That suggests we define $\beta_k=4\eta_k$ and $\gamma_k=4\eta_{k+1}+\eta_k$. With these choices, repeated substitution gives

$$R_k=R_{k+1}+4\eta_{k+1}+5\eta_k\le J(\delta)+9\sum_{i=k}^{\infty}\eta_i,$$

and, from (25),

$$P_k\le P_{k+1}+2N_{k+1}\exp\left(\frac{-\frac{1}{4}\eta_k^2}{\delta_k^2}\right).$$

This recursive inequality holds for all k. Repeated substitution leads to

$$P_0\le P_{k+1}+\sum_{j\le k}2\exp\left[\log N_{j+1}-\frac{\frac{1}{4}\eta_j^2}{\delta_j^2}\right].$$

The η_k must be large enough to cancel out the $\log N_{k+1}$ term with enough left over to make the resulting series convergent. Take

$$\eta_k=2\delta_k\sqrt{\lambda(\delta_{k+1})+\log N_{k+1}}\,.$$

Notice that $\alpha_k=\delta_k^2/\eta_k\to 0$ as $k\to\infty$. Because R_k stays bounded away from zero, this ensures that P_k is zero for k large enough:

$$
\begin{aligned}
F_k(\theta)-\Phi_k(\theta) &\le \sum_{i\le n}[u_{ik}(\theta)D_{ik}(\theta)+\mathbb{P}|f_i(\theta)-\Phi_{ik}(\theta)|]\\
&\le n^{3/2}\alpha_k+n\delta_k\\
&\to 0 \text{ as } k\to\infty.
\end{aligned}
$$

Also

$$
\begin{aligned}
R_0 &\le J(\delta)+9\sum_{k=0}^{\infty}\eta_k\\
&\le J(\delta)+9\sum_{k=0}^{\infty}8\int\{\delta_{k+2}<x\le\delta_{k+1}\}\sqrt{\lambda(x)+\log N(x)}\,dx\\
&\le 72J(\delta).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathbb{P}\left\{\sup_{\theta} n^{-1/2}\sum_{i\le n}[f_i(\theta)D_{i0}(\theta)-\mathbb{P}f_i(\theta)D_{i0}(\theta)]>72J(\delta)\right\}\\
&\quad\le\sum_{k=0}^{\infty}2N_{k+1}\exp[-\lambda(\delta_{k+1})-\log N_{k+1}]\\
&\quad\le 2\sum_{k=0}^{\infty}(\log 2)^{-1}\int\{\delta_{k+1}<x\le\delta_k\}x^{-1}\exp[-\lambda(x)]\,dx\\
&\quad\le\frac{2}{\log 2}\int_0^{\delta}x^{-1}\exp[-\lambda(x)]\,dx. \qquad (27)
\end{aligned}
$$

To get from (27) to (12) we have to account for the effect of the sum

$$
\sum_{i\le n}f_i(\theta)D_{i0}(\theta)^c=\sum_{i\le n}f_i(\theta)\{F_i>n^{1/2}\alpha_0\}
$$

and its expected value. First notice that

$$
\begin{aligned}
\alpha_0 &= \frac{\delta_0^2}{\eta_0}\\
&=\delta^2\left(\delta\sqrt{\lambda(\delta/2)+\log N(\delta/2)}\right)^{-1}\\
&\ge\frac{\tfrac{1}{2}\delta^2}{J(\delta)}\\
&=\epsilon(\delta).
\end{aligned}
$$

Inequality (26) with $m=0$ and $Z_i=f_i(\theta)$ gives

$$
\begin{aligned}
\left|\sum_{i\le n}\mathbb{P}f_i(\theta)D_{i0}(\theta)^c\right| &\le n\left(n^{-1/2}\frac{\delta}{\alpha_0}\right)\left(n^{-1}\sum_{i\le n}\mathbb{P}f_i(\theta)^2\right)^{1/2}\\
&\le n^{1/2}\frac{\delta^2}{\alpha_0}\\
&\le 2n^{1/2}J(\delta).
\end{aligned}
$$

It follows that

$$\mathbb{P}\left\{\sup_{\theta} n^{-1/2} \sum_{i\le n} [f_i(\theta)D_{i0}(\theta)^c - \mathbb{P}f_i(\theta)D_{i0}(\theta)^c] > 2J(\delta)\right\}$$
$$\le \mathbb{P}\{F_i > n^{1/2}\alpha_0 \text{ for some } i \le n\}$$
$$\le \sum_{i\le n} \mathbb{P}\{F_i > n^{1/2}\epsilon(\delta)\}.$$

To remove the restriction that the partitions π_k are nested, we have only to replace π_k by the common refinement of $\pi_0, \ldots, \pi_k$. This new partition consists of at most $N_1 \cdots N_k$ regions. Since

$$\sum_{k=0}^{\infty} \delta_k \sqrt{\lambda(\delta_{k+1}) + \log(N_1 \cdots N_{k+1})} \le \sum_{k=0}^{\infty} \delta_k \sum_{j\le k} \sqrt{\lambda(\delta_{j+1}) + \log N_{j+1}}$$
$$\le \sum_{j=0}^{\infty} \left(\sqrt{\lambda(\delta_{j+1}) + \log N_{j+1}} \sum_{k\ge j} \alpha_k \right)$$
$$\le 2 \sum_{j=0}^{\infty} \delta_j \sqrt{\lambda(\delta_{j+1}) + \log N_{j+1}},$$

this modification has at worst the effect of doubling the constant C_1.

The inequality asserted by Theorem 1 holds if $C_1 = 188$ and $C_2 = 2/\log 2$. ■

Anderson et al. (1988) have developed an improved form of Ossiander's argument, which allows one to relax slightly the second moment assumptions about the $u_i(\theta)$ in Definition 1. They have also replaced conditions involving covering numbers by assumptions about existence of "majorizing measures." Their method of proof is similar to the method used to establish Theorem 1, except that it replaces the constants η_k and α_{k+1} by π_{k+1}-simple processes.

Example 3

Now we can complete the discussion for the least absolute deviations regression estimator from Example 2. The application will not require the full power of Theorem 1; the necessary bracketing approximations will all be two-sided.

Assume that the $\{x_i\}$ satisfy the regularity condition (9), and that the distribution of e_1 has a bounded density. It will be convenient to treat separately the contributions with $x_i\theta \ge 0$ and the contributions with $x_i\theta < 0$. Define

$$f_i(\theta) = |e_i| - |e_i - (x_i\theta)^+| - D_i(x_i\theta)^+$$
$$= 2(e_i - (x_i\theta)^+)\{0 \le e_i \le (x_i\theta)^+\}.$$

The $r_i(\theta)$ process is a sum of $f_i(\theta)$ and a similar-looking process obtained by substituting $-(x_i\theta)^-$ for $(x_i\theta)^+$ in $f_i(\theta)$. As the arguments for both contributions to $r_i(\theta)$ would be similar, only $f_i(\theta)$ will be treated. It will be enough to show that

$$n^{-1/2}\sum_{i\le n} f_i(\theta)-\mathbb{P}f_i(\theta)=O_p(|\theta|)$$

uniformly over $A(n)=\{Kn^{-1/2}\le|\theta|\le\delta_n\}$, where $\{\delta_n\}$ is a sequence converging to zero.

The idea will be to decompose $A(n)$ into a sequence of annuli on each of which $|\theta|$ varies by at most a factor of 2. The expression needed to establish the uniform $O_p(|\theta|)$ inequality will be bounded by a sum of terms of the form

$$\mathbb{P}\left\{\exists\frac{R}{2}<|\theta|\le R: n^{-1/2}\sum_{i\le n} f_i(\theta)-\mathbb{P}f_i(\theta)>CR\right\}. \tag{28}$$

Theorem 1 will bound this probability by a function of R that decreases rapidly enough to produce a summable series.

We need to calculate the covering numbers $N(\epsilon)$ for the region $\{|\theta|\le R\}$. For θ and θ' in this region,

$$|f_i(\theta)-f_i(\theta')|\le 2|x_i|\cdot|\theta|\{0\le e_i\le|x_i|R\}.$$

Because e_1 has a bounded density, there exists a constant K_1 for which

$$\mathbb{P}\sup_{|\theta|\le R}|f_i(\theta)|^2\le K_1|x_i|^2R^2\min(1,|x_i|R)$$

and

$$\mathbb{P}\sup|f_i(\theta)-f_i(\theta')|^2\le K_1|x_i|^2\tau^2\min(1,|x_i|R)$$

if the second supremum runs over pairs in a subregion of diameter τ. Let β be a constant with $2<\beta<\alpha$. Later it will become clear how close β must be to 2. For all x_i and R,

$$\min(|x_i|^2,|x_i|^3R)\le R^{\beta-2}|x_i|^\beta \quad \text{if } 2<\beta<3.$$

Using this inequality, averaging over i, then invoking (9), we arrive at a constant K_2 for which

$$n^{-1}\sum_{i\le n}\mathbb{P}\sup_{|\theta|\le R}|f_i(\theta)|^2\le K_2R^\beta \quad \text{and}$$
$$n^{-1}\sum_{i\le n}\mathbb{P}\sup|f_i(\theta)-f_i(\theta')|^2\le K_2\tau^2R^{\beta-2},$$

where the second supremum again runs over pairs in a subregion of diameter τ. The first inequality lets us take $\delta^2=K_2R^\beta$ for the application of Theorem 1. If θ is d-dimensional, the ball of radius R can be covered by $O((R/\tau)^d)$ subregions of diameter τ; the second inequality gives us a bound for the covering numbers,

$$N(\epsilon) \le K_3(R^{2-\beta/2}/\epsilon)^d$$

for some constant K_3 and ϵ small enough.

If we choose $\lambda(x)$ equal to the positive part of $\log(R^{2-\beta/2}/x)$ then the $J(\delta)$ for Theorem 1 decreases like $R^{\beta/2}(\log 1/R)^{1/2}$, which is of order $o(R)$, the $B(\delta)$ decreases like $R^{\beta-2}$, and the sum of tail probabilities is bounded on $A(n)$ by

$$\sum_{i\le n}\left\{|x_i|R > K_4\frac{n^{1/2}R^{\beta}}{R^{\beta/2}(\log 1/R)^{1/2}}\right\} = \sum_{i\le n}\left\{|x_i| > K_5\frac{n^{1-\beta/4}}{(\log n)^{1/2}}\right\}$$
$$\le \left(n^{-1}\sum_{i\le n}|x_i|^{\alpha}\right)n\left[K_5\frac{n^{1-\beta/4}}{(\log n)^{1/2}}\right]^{-\alpha}.$$

Choose β close enough to 2 to ensure that $1<\alpha(1-\beta/4)$. Then, by virtue of (9), the bound converges to zero as n increases. That forces the sum of indicator functions, which can only take integer values, to be zero eventually. When that happens, Theorem 1 bounds the probability in (28) by a quantity of order $O(R^{\beta-2})$, for all $R \ge \frac{1}{2}Kn^{-1/2}$. Thus

$$\mathbb{P}\left\{\exists\theta\in A(n): n^{-1/2}\sum_{i\le n} f_i(\theta)-\mathbb{P}f_i(\theta) > 2C|\theta|\right\}$$
$$\le \sum_{k=0}^{k(n)} \mathbb{P}\left\{\exists R_{k+1}<|\theta|\le R_k : n^{-1/2}\sum_{i\le n} f_i(\theta)-\mathbb{P}f_i(\theta) > CR_k\right\},$$

where $R_k=\delta_n/2^k$ and $k(n)$ is such that $\frac{1}{2}Kn^{-1/2}<R_{k(n)}\le Kn^{-1/2}$. The series is eventually bounded by

$$K_6\sum_{k=0}^{\infty} R_k^{\beta-2} = O(\delta_n^{\beta-2}) = o(1)$$

as required.

For further applications of Theorem 1 see Pollard (1984, chapter 7; 1985).

5 Postscript

Following the presentation of this paper at the meeting, the discussant and several members of the audience asked about the possible extension to the case of dependent variables for the results in Section 4. Inequality (13) is the most important step at which independence was used. Even though analogs of this inequality exist for dependent variables, I have had no success in applying them to get a bound as in (12). However, in joint work with Don Andrews, I have obtained a weaker result for α-mixing sequences, using moment inequalities instead of inequality (13).

The analysis of the least absolute deviations regression estimator, as in Examples 2 and 3, is now obsolete. Under weaker, more natural conditions, I have established (Pollard 1991) a central limit theorem using elementary methods. Nevertheless, the analysis in the present chapter does illustrate the mechanics of bracketing approximation.

The ideas in Section 4 have been carried further in Pollard (1989), where I have derived a two-sided maximal inequality by chaining with first moments. This approach greatly simplifies the argument.

REFERENCES

Anderson, N. T., E. Giné, M. Ossiander, and J. Zinn (1988), "The Central Limit Theorem and the Law of Iterated Logarithm for Empirical Processes under Local Conditions," *Probability Theory and Related Fields,* 77, 271–305.

Andrews, D. W. K. (1987), "Consistency in Non-linear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica,* 55, forthcoming.

Bates, C., and H. White (1985), "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory,* 1, 151–78.

Domowitz, I., and H. White (1982), "Misspecified Models with Dependent Observations," *Journal of Econometrics,* 20, 35–58.

Dudley, R. M. (1978), "Central Limit Theorems for Empirical Measure," *Annals of Probability,* 6, 899–929; Correction, (1979), 7, 909–11.

(1984), *A Course on Empirical Processes* (Lectures given at Ecole d'Eté de Probabilités de St. Flour, 1982), Springer Lecture Notes in Mathematics, Vol. 1097 (New York: Springer-Verlag), pp. 1–142.

Hoadley, B. (1971), "Asymptotic Properties for Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," *Annals of Mathematical Statistics,* 42, 1977–91.

Ossiander, M. (1987), "A Central Limit Theorem under Metric Entropy with Bracketing," *Annals of Probability,* 15, 897–919.

Pollard, D. (1984), *Convergence of Stochastic Processes* (New York: Springer-Verlag).

(1985), "New Ways to Prove Central Limit Theorems," *Econometric Theory,* 1, 295–314.

(1989), "A Maximal Inequality for Sums of Independent Processes under a Bracketing Condition" (Preprint).

(1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory,* 7, 186–99.

Potscher, B. M., and I. R. Prucha (1987), "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results," Working Paper 86-9, Department of Economics, University of Maryland.

Shorack, G. R., and J. A. Wellner (1986), *Empirical Processes with Applications to Statistics* (New York: Wiley).

Wald, A. (1949), "Note on the Consistency of the Maximum Likelihood Estimator," *Annals of Mathematical Statistics,* 20, 595–691.