COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 208281 Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1059

Empirical Process Methods in Econometrics

by

Donald W.K. Andrews

September 1993

EMPIRICAL PROCESS METHODS
IN ECONOMETRICS

Donald W. K. Andrews[1]

Cowles Foundation
Yale University

May 1993

ABSTRACT

This paper provides an introduction to the use of empirical process methods in econometrics. These methods can be used to establish the large sample properties of econometric estimators and test statistics. In the first part of the paper, key terminology and results are introduced and discussed heuristically. Applications in the econometrics literature are briefly reviewed. A select set of three classes of applications is discussed in more detail.

The second part of the paper shows how one can verify a key property called stochastic equicontinuity. The paper takes several stochastic equicontinuity results from the probability literature, which rely on entropy conditions of one sort or another, and provides primitive sufficient conditions under which the entropy conditions hold. This yields stochastic equicontinuity results that are readily applicable in a variety of contexts. Examples are provided.

1. INTRODUCTION

This paper discusses the use of empirical process methods in econometrics. It begins by defining, and discussing heuristically, empirical processes, weak convergence, and stochastic equicontinuity. The paper then provides a brief review of the use of empirical process methods in the econometric literature. Their use is primarily in the establishment of the asymptotic distributions of various estimators and test statistics.

Next, the paper discusses three classes of applications of empirical process methods in more detail. The first is the establishment of asymptotic normality of parametric M-estimators that are based on non-differentiable criterion functions. This includes least absolute deviations and method of simulated moments estimators, among others. The second is the establishment of asymptotic normality of semiparametric estimators that depend on preliminary nonparametric estimators. This includes weighted least squares estimators of partially linear regression models and semiparametric generalized method of moments estimators of parameters defined by conditional moment restrictions, among others. The third is the establishment of the asymptotic null distributions of several test statistics that apply in the non-standard testing scenario in which a nuisance parameter appears under the alternative hypothesis, but not under the null. Examples of such testing problems include tests of variable relevance in certain nonlinear models, such as models with Box-Cox transformed variables, and tests of cross-sectional constancy in regression models.

As shown in the first part of the paper, the verification of stochastic equicontinuity in a given application is the key step in utilizing empirical process results. The second part of the paper provides methods for verifying stochastic equicontinuity. Numerous results are available in the probability literature concerning sufficient conditions for stochastic equicontinuity (references are given below). Most of these results rely on some sort of entropy condition. For application to specific estimation and testing problems, such entropy conditions are not sufficiently primitive. The second part of the paper provides an array of primitive conditions under which such entropy conditions hold, and hence, under which stochastic equicontinuity obtains. The

primitive conditions considered here include: differentiability conditions, Lipschitz conditions, L^p continuity conditions, Vapnik-Červonenkis conditions, and combinations thereof. Applications discussed in the first part of the paper are employed to exemplify the use of these primitive conditions.

The empirical process results discussed here apply only to random variables (rv's) that are independent or m-dependent (i.e., independent beyond lags of length m). There is a growing literature on empirical processes with more general forms of temporal dependence. See Andrews (1993a) for a review of this literature.

The remainder of this paper is organized as follows: Section 2 defines and discusses empirical processes, weak convergence, and stochastic equicontinuity. Section 3 gives a brief review of the use of empirical process methods in the econometrics literature and discusses three classes of applications in more detail. Sections 4 and 5 provide stochastic equicontinuity results of the paper. Section 6 provides a brief conclusion. An Appendix contains proofs of results stated in Sections 4 and 5.

2. WEAK CONVERGENCE AND STOCHASTIC EQUICONTINUITY

We begin by introducing some notation. Let $\{W_{Tt} : t \le T, T \ge 1\}$ be a triangular array of $\mathcal{W}$-valued rv's defined on a probability space $(\Omega, \mathcal{A}, P)$, where $\mathcal{W}$ is a (Borel measurable) subset of R^k. For notational simplicity, we abbreviate W_{Tt} by W_t below. Let $\mathcal{T}$ be a *pseudometric* space with pseudometric ρ.[2] Let

(2.1) $\mathcal{M} = \{m(\cdot, \tau) : \tau \in \mathcal{T}\}$

be a class of R^s-valued functions defined on $\mathcal{W}$ and indexed by $\tau \in \mathcal{T}$. Define an *empirical process* $\nu_T(\cdot)$ by

(2.2) $\nu_T(\tau) = \dfrac{1}{\sqrt{T}} \Sigma_1^T (m(W_t, \tau) - Em(W_t, \tau))$ for $\tau \in \mathcal{T}$,

where Σ_1^T abbreviates $\Sigma_{t=1}^T$. The empirical process $\nu_T(\cdot)$ is a particular type of stochastic process. If $\mathcal{T} = [0, 1]$, then $\nu_T(\cdot)$ is a stochastic process on $[0, 1]$. For parametric applications of

empirical process theory, $\mathcal{T}$ is usually a subset of R^p. For semiparametric and nonparametric applications, $\mathcal{T}$ is often a class of functions. In some other applications, such as chi-square diagnostic test applications, $\mathcal{T}$ is a class of subsets of R^p.

We now define *weak convergence* of the sequence of empirical processes $\{v_T(\cdot) : T \geq 1\}$ to some stochastic process $v(\cdot)$ indexed by elements τ of $\mathcal{T}$. ($v(\cdot)$ may or may not be defined on the same probability space $(\Omega, \mathcal{A}, P)$ as $v_T(\cdot) \ \forall T \geq 1$.) Let $\rightarrow$ denote weak convergence of stochastic processes, as defined below. Let $\xrightarrow{d}$ denote convergence in distribution of some sequence of rv's. Let $|\cdot|$ denote the Euclidean norm. All limits below are taken as $T \rightarrow \infty$.

DEFINITION OF WEAK CONVERGENCE:

$$v_T(\cdot) \rightarrow v(\cdot) \quad \text{if} \quad E^* f(v_T(\cdot)) \rightarrow Ef(v(\cdot)) \quad \forall f \in \mathcal{U}(B(\mathcal{T})) ,$$

where $B(\mathcal{T})$ is the class of bounded R^ς-valued functions on $\mathcal{T}$ (which includes all realizations of $v_T(\cdot)$ and $v(\cdot)$ by assumption), d is the uniform metric on $B(\mathcal{T})$ (i.e., $d(b_1, b_2) = \sup_{\tau \in \mathcal{T}} \|b_1(\tau) - b_2(\tau)\|$), and $\mathcal{U}(B(\mathcal{T}))$ is the class of all bounded uniformly continuous (with respect to the metric d) real functions on $B(\mathcal{T})$.

In the definition, E^* denotes outer expectation. Correspondingly, P^* denotes outer probability below. (It is used because it is desirable not to require $v_T(\cdot)$ to be a measurable random element of the metric space $(B(\mathcal{T}), d)$ with its Borel σ-field, since measurability in this context can be too restrictive. For example, if $(B(\mathcal{T}), d)$ is the space of functions $D[0, 1]$ with the uniform metric, then the standard empirical distribution function is not measurable with respect to its Borel σ-field. The limit stochastic process $v(\cdot)$, on the other hand, is sufficiently well-behaved in applications that it is assumed to be measurable in the definition.)

The above definition is due to Hoffman-Jorgensen. It is widely used in the recent probability literature, e.g., see Pollard (1990, Sec. 9).

Weak convergence is a useful concept for econometrics, because it can be used to establish the asymptotic distributions of estimators and test statistics. Section 3 below illustrates how.

For now, we consider sufficient conditions for weak convergence. In many applications of interest, the limit process $v(\cdot)$ is (uniformly ρ) continuous in τ with probability one. In such

cases, a property of the sequence of empirical processes $\{v_T(\cdot) : T \geq 1\}$, called *stochastic equicontinuity*, is a key member of a set of sufficient conditions for weak convergence. It also is implied by weak convergence (if the limit process $v(\cdot)$ is as above).

DEFINITION OF STOCHASTIC EQUICONTINUITY: $\{v_T(\cdot) : T \geq 1\}$ is *stochastically equicontinuous* if $\forall \varepsilon > 0$ and $\eta > 0$, $\exists \delta > 0$ such that

$$(2.3) \quad \overline{\lim_{T \to \infty}} \, P(\sup_{\tau_1, \tau_2 \in \mathcal{T}: \rho(\tau_1, \tau_2) < \delta} |v_T(\tau_1) - v_T(\tau_2)| > \eta) < \varepsilon .$$

Basically, a sequence of empirical processes $\{v_T(\cdot) : T \geq 1\}$ is stochastic equicontinuous if $v_T(\cdot)$ is continuous in τ uniformly over $\mathcal{T}$ at least with high probability and for T large. Thus, stochastic equicontinuity is a probabilistic and asymptotic generalization of uniform continuity of a function.

The concept of stochastic equicontinuity is quite old and appears in the literature under various guises. For example, it appears in Theorem 8.2 of Billingsley (1968, p. 55), which is attributed to Prohorov (1956), for the case of $\mathcal{T} = [0, 1]$. Moreover, a non-asymptotic analogue of stochastic equicontinuity arises in the even older literature on the existence of stochastic processes with continuous sample paths.

The concept of stochastic equicontinuity is important for two reasons. First, as mentioned above, stochastic equicontinuity is a key member of a set of sufficient conditions for weak convergence. These conditions are specified immediately below. Second, in many applications it is not necessary to establish a full functional limit (i.e., weak convergence) result to obtain the desired result -- it suffices to establish just stochastic equicontinuity. Examples of this are given in Section 3 below.

Sufficient conditions for weak convergence are given in the following widely used result. A proof of the result can be found in Pollard (1990, Sec. 10) (but the basic result has been around for some time). Recall that a pseudometric space is said to be *totally bounded* if it can be covered by a finite number of ε-balls $\forall \varepsilon > 0$. (For example, a subset of Euclidean space is totally bounded if and only if it is bounded.)

PROPOSITION: *If* (i) $(\mathcal{T}, \rho)$ *is a totally bounded pseudometric space,*

(ii) *finite dimensional* (fidi) *convergence holds:* $\forall$ *finite subsets* $(\tau_1, ..., \tau_J)$ *of* $\mathcal{T}$, $(v_T(\tau_1)', ..., v_T(\tau_J)')'$ *converges in distribution, and*

(iii) $\{v_T(\cdot) : T \geq 1\}$ *is stochastically equicontinuous,*

then there exists a (Borel-measurable with respect to d) $B(\mathcal{T})$-valued stochastic process $v(\cdot)$, *whose sample paths are uniformly* ρ *continuous with probability one, such that* $v_T(\cdot) \rightarrow v(\cdot)$.

Conversely, if $v_T(\cdot) \rightarrow v(\cdot)$ *for* $v(\cdot)$ *with the properties above and* (i) *holds, then* (ii) *and* (iii) *hold.*

Condition (ii) of the Proposition typically is verified by applying a multivariate central limit theorem (CLT) (or a univariate CLT coupled with the Cramer-Wold device, see Billingsley (1968)). There are numerous CLTs in the literature that cover different configurations of non-identical distributions and temporal dependence.

Condition (i) of the Proposition is straightforward to verify if $\mathcal{T}$ is a subset of Euclidean space and is typically a by-product of the verification of stochastic equicontinuity in other cases. In consequence, the verification of stochastic equicontinuity is the key step in verifying weak convergence (and, as mentioned above, is often the desired end in its own right). For these reasons, we provide further discussion of the stochastic equicontinuity condition here and we provide methods for verifying it in several sections below.

Two equivalent definitions of stochastic equicontinuity are the following: (i) $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous if for every sequence of constants $\{\delta_T\}$ that converges to zero, we have $\sup_{\rho(\tau_1, \tau_2) \leq \delta_T} |v_T(\tau_1) - v_T(\tau_2)| \xrightarrow{P} 0$ where "$\xrightarrow{P}$" denotes convergence in probability, and (ii) $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous if for all sequences of random elements $\{\hat{\tau}_{1T}\}$ and $\{\hat{\tau}_{2T}\}$ that satisfy $\rho(\hat{\tau}_{1T}, \hat{\tau}_{2T}) \xrightarrow{P} 0$, we have $v_T(\hat{\tau}_{1T}) - v_T(\hat{\tau}_{2T}) \xrightarrow{P} 0$. The latter characterization of stochastic equicontinuity reflects its use in the semiparametric examples below. Allowing $\{\hat{\tau}_{1T}\}$ and $\{\hat{\tau}_{2T}\}$ to be random in the latter characterization is crucial. If only fixed sequences were considered, then the property would be substantially weaker -- it would not deliver the result that $v_T(\hat{\tau}_{1T}) - v_T(\hat{\tau}_{2T}) \xrightarrow{P} 0$ -- and its proof would be substantially simpler -- the property would follow directly from Chebyshev's inequality.

To demonstrate the plausibility of the stochastic equicontinuity property, suppose $\mathcal{M}$ contains only linear functions, i.e., $\mathcal{M} = \{g : g(w) = w'\tau \text{ for some } \tau \in R^k\}$ and ρ is the Euclidean metric. In this simple linear case,

$$\overline{\lim}_{T \to \infty} P^*(\sup_{\tau_1,\tau_2:\rho(\tau_1,\tau_2)<\delta} |v_T(\tau_1) - v_T(\tau_2)| > \eta)$$

$$(2.4) \quad = \overline{\lim}_{T \to \infty} P^*(\sup_{\tau_1,\tau_2:\rho(\tau_1,\tau_2)<\delta} |\frac{1}{\sqrt{T}}\Sigma_1^T(W_t - EW_t)'(\tau_1 - \tau_2)| > \eta)$$

$$\leq \overline{\lim}_{T \to \infty} P(|\frac{1}{\sqrt{T}}\Sigma_1^T(W_t - EW_t)| > \eta/\delta)$$

$$< \varepsilon \, ,$$

where the first inequality holds by the Cauchy-Schwarz inequality and the second inequality holds for δ sufficiently small provided $\frac{1}{\sqrt{T}}\Sigma_1^T(W_t - EW_t) = O_p(1)$. Thus, $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous in this case if the rv's $\{W_t - EW_t : t \leq T, T \geq 1\}$ satisfy an ordinary CLT.

For classes of nonlinear functions, the stochastic equicontinuity property is substantially more difficult to verify than for linear functions. Indeed, it is not difficult to demonstrate that it does not hold for all classes of functions $\mathcal{M}$. Some restrictions on $\mathcal{M}$ are necessary -- $\mathcal{M}$ cannot be too complex/large.

To see this, suppose $\{W_t : t \leq T, T \geq 1\}$ are iid with distribution P_1 that is absolutely continuous with respect to Lebesgue measure and $\mathcal{M}$ is the class of indicator functions of all Borel sets in $\mathcal{W}$. Let τ denote a Borel set in $\mathcal{W}$ and let $\mathcal{T}$ denote the collection of all such sets. Then, $m(w, \tau) = 1(w \in \tau)$. Take $\rho(\tau_1, \tau_2) = \left(\int (m(w, \tau_1) - m(w, \tau_2))^2 dP_1(w)\right)^{1/2}$. For any two sets τ_1, τ_2 in $\mathcal{T}$ that have finite numbers of elements, $v_T(\tau_j) = \frac{1}{\sqrt{T}}\Sigma_1^T 1(W_t \in \tau_j)$ and $\rho(\tau_1, \tau_2) = 0$, since $P_1(W_t \in \tau_j) = 0$ for $j = 1, 2$. Given any $T \geq 1$ and any realization $\omega \in \Omega$, there exist finite sets $\tau_{1T\omega}$ and $\tau_{2T\omega}$ in $\mathcal{T}$ such that $W_t(\omega) \in \tau_{1T\omega}$ and $W_t(\omega) \notin \tau_{2T\omega}$ $\forall t \leq T$, where $W_t(\omega)$ denotes the value of W_t when ω is realized. This yields $v_T(\tau_{1T\omega}) = \sqrt{T}$, $v_T(\tau_{2T\omega}) = 0$, and $\sup_{\rho(\tau_1,\tau_2)<\delta} |v_T(\tau_1) - v_T(\tau_2)| \geq \sqrt{T}$. In consequence, $\{v_T(\cdot) : T \geq 1\}$ is not stochastically equicontinuous. The class of functions $\mathcal{M}$ is too large.

In Sections 4 and 5 below, we discuss various *entropy* conditions that restrict the complexity/size of the class of functions $\mathcal{M}$ sufficiently that stochastic equicontinuity holds. Before

doing so, however, we illustrate how weak convergence and stochastic equicontinuity results can be fruitfully employed in various econometric applications.

3. APPLICATIONS

3.1. Review of Applications

In this subsection, we briefly describe a number of applications of empirical process theory that appear in the econometrics literature. There are numerous others that appear in the statistics literature, see Shorack and Wellner (1986) for some references.

The applications and use of empirical process methods in econometrics are fairly diverse. Some applications use a full weak convergence result; others just use a stochastic equicontinuity result. Most applications use empirical process theory for normalized sums of rv's, but some use the corresponding theory for U-processes, see Kim and Pollard (1990) and Sherman (1992). The applications include estimation problems and testing problems. Here we categorize the applications not by the type of empirical process method used, but by area of application. We consider estimation first, then testing.

Empirical process methods are useful in obtaining the asymptotic normality of parametric optimization estimators when the criterion function that defines the estimator is not differentiable. Estimators that fit into this category include robust M-estimators (see Huber (1973)), regressions quantiles (see Koenker and Bassett (1978)), censored regression quantiles (see Powell (1984, 1986a)), trimmed LAD estimators (see Honore (1992)), and method of simulated moments estimators (see McFadden (1989) and Pakes and Pollard (1989)). Huber (1967) gave some asymptotic normality results for a class of M-estimators of the above sort using empirical process-like methods. His results have been utilized by numerous econometricians, e.g., see Powell (1984). Empirical process methods were utilized explicitly in several subsequent papers that treat parametric estimation with non-differentiable criterion functions, see Pollard (1984, 1985), McFadden (1989), Pakes and Pollard (1989), and Andrews (1988a). Also, see Newey and McFadden (1993) in this *Handbook*. In Section 3.2 below, we illustrate one way in which empirical process methods can be exploited for problems of this sort.

Empirical process methods also have been utilized in the semiparametric econometrics literature. They have been used to establish the asymptotic normality (and, in a few cases, other limiting distributions) of various estimators. References include Horowitz (1988, 1992), Kim and Pollard (1990), Andrews (1993b), Newey (1989), White and Stinchcombe (1991), Olley and Pakes (1991), Pakes and Olley (1991), Aït-Sahalia (1992a, b), Sherman (1993a, b), and Cavanagh and Sherman (1992). Kim and Pollard (1990) establish the asymptotic (non-normal) distribution of Manski's (1975) maximum score estimator for binary choice models using empirical process theory for U-statistics. Horowitz (1992) establishes the asymptotic normal distribution of a smoothed version of the maximum score estimator. Andrews (1993b), Newey (1989), Pakes and Olley (1991), and Aït-Sahalia (1992b) all use empirical process theory to establish the asymptotic normality of classes of semiparametric estimators that employ nonparametric estimators in their definition. Andrews (1993b), Newey (1989), and Pakes and Olley (1991) use stochastic equicontinuity results, whereas Aït-Sahalia (1992b) utilizes a full weak convergence result. Sherman (1993a, b) and Cavanagh and Sherman (1992) establish asymptotic normality of a number of semiparametric estimators using empirical process theory of U-statistics. Section 3.3 below gives a heuristic description of one way in which empirical process methods can be used for semiparametric estimation problems.

A third area of application of empirical process methods to estimation problems is that of nonparametrics. Gallant (1989) and Gallant and Souza (1989) use these methods to establish the asymptotic normality of certain seminonparametric (i.e., nonparametric series) estimators. In their proof, empirical process methods are used to establish that a law of large numbers holds uniformly over a class of functions that expands with the sample size. Andrews (1990) uses empirical process methods to show that nonparametric kernel density and regression estimators are consistent when the dependent variable or the regressor variables are residuals from some preliminary estimation procedure (as often occurs in semiparametric applications).

Empirical process methods also have been utilized very effectively in justifying the use of bootstrap confidence intervals. References include Gine and Zinn (1990), Arcones and Gine (1992), and Hahn (1992).

Next, we consider testing problems. Empirical process methods have been used in the literature to obtain the asymptotic null (and local alternative) distributions of a wide variety of test statistics. These include test statistics for chi-square diagnostic tests (see Andrews (1988b, c)), consistent model specification tests (see Bierens (1990), Yatchew (1992), Hansen (1992), DeJong (1992), and Stinchcombe and White (1993)), tests of nonlinear restrictions in semiparametric models (see Andrews (1988a)), tests of specification of semiparametric models (see Whang and Andrews (1993) and White and Hong (1992)), tests of stochastic dominance (see Klecan, McFadden, and McFadden (1990), and tests of hypotheses for which a nuisance parameter appears only under the alternative (see Davies (1977, 1987), Bera and Higgins (1992, 1993), Hansen (1991, 1993), Andrews and Ploberger (1991), and Stinchcombe and White (1993). For tests of the latter sort, Section 3.4 below describes how empirical process methods are utilized.

Last, we note that stochastic equicontinuity can be used to obtain uniform laws of large numbers that can be employed in proofs of consistency of extremum estimators. For example, see Pollard (1984, Ch. 2), Newey (1991), and Andrews (1992).

3.2. Parametric M-estimators Based on Non-differentiable Criterion Functions

Here we give a heuristic description of one way in which empirical process theory can be used to establish the asymptotic normality of parametric M-estimators (or GMM estimators) that are based on criterion functions that are not differentiable with respect to the unknown parameter. This treatment follows that of Andrews (1988a) most closely (in which a formal statement of assumptions and results can be found). Other references are given in Section 3.1 above.

Suppose $\hat{\tau}$ is a consistent estimator of a parameter $\tau_0 \in R^p$ that satisfies a set of p first order conditions

$$(3.1) \qquad \bar{m}_T(\hat{\tau}) = 0$$

at least with probability that goes to one as $T \to \infty$, where

$$(3.2) \qquad \bar{m}_T(\tau) = \frac{1}{T}\Sigma_1^T m(W_t, \tau) .$$

Here, W_t is an observed vector of random variables and $m(\cdot, \cdot)$ is a known R^p-valued function.

Examples are given below.

If $m(W_t, \tau)$ is differentiable in τ, one can establish the asymptotic normality of $\hat{\tau}$ by expanding $\sqrt{T}\bar{m}_T(\hat{\tau})$ about τ_0 using element by element mean value expansions. This is the standard way of establishing asymptotic normality of $\hat{\tau}$ (or, more precisely, of $\sqrt{T}(\hat{\tau} - \tau_0)$). In a variety of applications, however, the function $m(W_t, \tau)$ is not differentiable in τ, or not even continuous, due to the appearance of a sign function, an indicator function, a kinked function, etc. Examples are listed above and below. In such cases, one can still establish asymptotic normality of $\hat{\tau}$ provided $Em(W_t, \tau)$ is differentiable in τ. Since the expectation operator is a smoothing operator, $Em(W_t, \tau)$ is often differentiable in τ even though $m(W_t, \tau)$ is not.

One method is as follows. Let

$$(3.3) \qquad \bar{m}_T^*(\tau) = \frac{1}{T}\Sigma_1^T Em(W_t, \tau) \ .$$

To establish asymptotic normality of $\hat{\tau}$, one can replace (element by element) mean value expansions of $\bar{m}_T(\hat{\tau})$ about τ_0 by corresponding mean value expansions of $\bar{m}_T^*(\tau_0)$ about $\hat{\tau}$ and then use empirical process methods to establish the limit distribution of the expansion. In particular, such mean value expansions yield

$$(3.4) \qquad 0 = \sqrt{T}\bar{m}_T^*(\tau_0) = \sqrt{T}\bar{m}_T^*(\hat{\tau}) - \frac{\partial}{\partial\tau'}\bar{m}_T^*(\bar{\tau})\sqrt{T}(\hat{\tau} - \tau_0) \ ,$$

where the first equality holds by the population orthogonality conditions (by assumption) and $\bar{\tau}$ lies on the line segment joining $\hat{\tau}$ and τ_0 (and takes different values in each row of $\frac{\partial}{\partial\tau'}\bar{m}_T(\bar{\tau})$). Under suitable assumptions on $\{m(W_t, \tau) : t \le T, T \ge 1\}$, one obtains $\frac{\partial}{\partial\tau'}\bar{m}_T^*(\bar{\tau}) \xrightarrow{p} M$ $= \lim_{T\to\infty}\frac{1}{T}\Sigma_1^T\frac{\partial}{\partial\tau'}Em(W_t, \tau_0)$. (For example, if the rv's W_t are identically distributed, it suffices to have $\frac{\partial}{\partial\tau'}Em(W_t, \tau)$ continuous in τ at τ_0.) Thus, provided M is nonsingular, one has

$$(3.5) \qquad \sqrt{T}(\hat{\tau} - \tau_0) = (M^{-1} + o_p(1))\sqrt{T}\bar{m}_T^*(\hat{\tau}) \ .$$

(Here, $o_p(1)$ denotes a term that converges in probability to zero as $T \to \infty$.)

Now, the asymptotic distribution of $\sqrt{T}(\hat{\tau} - \tau_0)$ is obtained by using empirical process methods to determine the asymptotic distribution of $\sqrt{T}\bar{m}_T^*(\hat{\tau})$. We write

$$(3.6) \quad -\sqrt{T}\bar{m}_T^*(\hat{\tau}) = (\sqrt{T}\bar{m}_T(\hat{\tau}) - \sqrt{T}\bar{m}_T^*(\hat{\tau})) - \sqrt{T}\bar{m}_T(\hat{\tau})$$

$$= (v_T(\hat{\tau}) - v_T(\tau_0)) + v_T(\tau_0) - \sqrt{T}\bar{m}_T(\hat{\tau}) .$$

The third term on the right-hand side (rhs) of (3.6) is $o_p(1)$ by (3.1). The second term on the rhs of (3.6) is asymptotically normal by an ordinary CLT under suitable moment and temporal dependence assumptions, since $v_T(\tau_0)$ is a normalized sum of mean zero rv's. That is, we have

$$(3.7) \quad v_T(\tau_0) = \frac{1}{\sqrt{T}}\Sigma_1^T(m(W_t, \tau_0) - Em(W_t, \tau_0)) \xrightarrow{d} N(0, S) \quad \text{as} \quad T \to \infty,$$

where $S = \lim_{T\to\infty} \text{Var}(\frac{1}{\sqrt{T}}\Sigma_1^T m(W_t, \tau_0))$. (For example, if the rv's W_t are independent and identically distributed (iid), it suffices to have $S = Em(W_t, \tau_0)m(W_t, \tau_0)'$ well-defined.)

Next, the first term on the rhs of (3.6) is $o_p(1)$ provided $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous and $\hat{\tau} \xrightarrow{P} \tau_0$. This follows because given any $\eta > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$(3.8)$$
$$\overline{\lim_{T\to\infty}} P(|v_T(\hat{\tau}) - v_T(\tau_0)| > \eta)$$

$$\leq \overline{\lim_{T\to\infty}} P(|v_T(\hat{\tau}) - v_T(\tau_0)| > \eta, \rho(\hat{\tau}, \tau_0) \leq \delta) + \overline{\lim_{T\to\infty}} P(\rho(\hat{\tau}, \tau_0) > \delta)$$

$$\leq \overline{\lim_{T\to\infty}} P(\sup_{\tau\in\mathcal{T}:\rho(\tau,\tau_0)\leq\delta} |v_T(\tau) - v_T(\tau_0)| > \eta)$$

$$< \varepsilon,$$

where the second inequality uses $\hat{\tau} \xrightarrow{P} \tau_0$ and the third uses stochastic equicontinuity.

Combining (3.5)-(3.8) yields the desired result that

$$(3.9) \quad \sqrt{T}(\hat{\tau} - \tau_0) \xrightarrow{d} N\left(0, M^{-1}S(M^{-1})'\right) \quad \text{as} \quad T \to \infty.$$

It remains to show how one can verify the stochastic equicontinuity of $\{v_T(\cdot) : T \geq 1\}$. This is done in Sections 4 and 5 below. Before doing so, we consider several examples.

EXAMPLE 1: M-estimators for standard, censored, and truncated linear regressions models. In the models considered here, $\{(Y_t, X_t) : t \leq T\}$ are observed rv's and $\{(Y_t^*, X_t^*) : t \leq T\}$ are latent rv's. The models are defined by

$$Y_t^* = X_t^{*'}\theta_0 + U_t, \quad t = 1, \ldots, T,$$

(3.10)

          linear regression (LR): $(Y_t, X_t) = (Y_t^*, X_t^*)$,

          censored regression (CR): $(Y_t, X_t) = (Y_t^* 1(Y_t^* \geq 0), X_t^*)$,

          truncated regression (TR): $(Y_t, X_t) = (Y_t^* 1(Y_t^* \geq 0), X_t^* 1(Y_t^* \geq 0))$.

Depending upon the context, the errors $\{U_t\}$ may satisfy any one of a number of assumptions such as constant conditional mean or quantile for all t or symmetry about zero for all t. We need not be specific for present purposes.

We consider M-estimators $\hat{\tau}$ of τ_0 that satisfy the equations

(3.11) $0 = \Sigma_1^T \psi_1(Y_t - X_t'\hat{\tau})\psi_2(W_t, \hat{\tau})X_t$

with probability -1 as $T \to \infty$, where $W_t = (Y_t, X_t')'$. Such estimators fit the general framework of (3.1)-(3.2) with

(3.12) $m(w, \tau) = \psi_1(y - x'\tau)\psi_2(w, \tau)x$,   where $w = (y, x')'$.

Examples of such M-estimators in the literature include the following:

(a) LR model: Let $\psi_1(z) = \text{sgn}(z)$ and $\psi_2 = 1$ to obtain the least absolute deviations (LAD) estimator. Let $\psi_1(z) = q - 1(y - x'\tau < 0)$ and $\psi_2 = 1$ to obtain Koenker and Bassett's (1978) regression quantile estimator for quantile $q \in (0, 1)$. Let $\psi_1(z) = (z \wedge c) \vee (-c)$ (where $\wedge$ and $\vee$ are the min and max operators respectively) and $\psi_2 = 1$ to obtain Huber's (1973) M-estimator with truncation at $\pm c$. Let $\psi_1(z) = |q - 1(y - x'\tau < 0)|$ and $\psi_2(w, \tau) = y - x'\tau$ to obtain Newey and Powell's (1987) asymmetric LS estimator.

(b) CR model: Let $\psi_1(z) = q - 1(y - x'\tau < 0)$ and $\psi_2(w, \tau) = 1(x'\tau > 0)$ to obtain Powell's (1984, 1986a) censored regression quantile estimator for quantile $q \in (0, 1)$. Let $\psi_1 = 1$ and $\psi_2(w, \tau) = 1(x'\tau > 0)[(y - x'\tau) \wedge x'\tau]$ to obtain Powell's (1986b) symmetrically trimmed LS estimator.

(c) TR model: Let $\psi_1 = 1$ and $\psi_2(w, \tau) = 1(y < 2x'\tau)(y - x'\tau)$ to obtain Powell's (1986b) symmetrically trimmed LS estimator.

(Note that for the Huber M-estimator of the LR model one would usually simultaneously estimate a scale parameter for the errors U_t. For brevity, we omit this above.)

EXAMPLE 2: Method of simulated moments (MSM) estimator for multinomial probit. The model and estimator considered here are as in McFadden (1989) and Pakes and Pollard (1989). We consider a discrete response model with r possible responses. Let D_t be an observed response vector that takes values in $\{e_i : i = 1, ..., r\}$, where $e_i = (0, ..., 0, 1, 0, ..., 0)'$ is the i-th elementary r-vector. Let Z_{ti} denote an observed b-vector of covariates -- one for each possible response $i = 1, ..., r$. Let $Z_t = [Z'_{t1}, Z'_{t2} \cdots Z'_{tr}]'$. The model is defined such that

$$(3.13) \quad D_t = e_i \quad \text{if} \quad (Z_{ti} - Z_{t\ell})'(\beta(\tau_0) + A(\tau_0)U_t) \geq 0 \quad \forall \ell = 1, ..., r \, ,$$

where $U_t \sim N(0, I_c)$ is an unobserved normal rv, $\beta(\cdot)$ and $A(\cdot)$ are known $R^{b \times 1}$ and $R^{b \times c}$ valued functions of an unknown parameter $\tau_0 \in \mathcal{T} \subset R^p$.

McFadden's MSM estimator of τ_0 is constructed using s independent simulated $N(0, I_c)$ rv's $(Y_{t1}, ..., Y_{ts})'$ and a matrix of instruments $g(Z_t, \tau)$, where $g(\cdot, \cdot)$ is a known $R^{v \times b}$-valued function. The MSM estimator is an example of the estimator of (3.1)-(3.2) with $W_t = (D_t, Z_t, Y_{t1}, ..., Y_{ts})$ and

$$(3.14) \quad m(w, \tau) = g(z, \tau)\left[d - \frac{1}{s}\sum_{j=1}^{s} H[z(\beta(\tau) + A(\tau)y_j)]\right],$$

where $w = (d, z, y_1, ..., y_s)$. Here, $H[\cdot]$ is a $\{0, 1\}^r$-valued function whose i-th element is of the form

$$(3.15) \quad \prod_{\ell=1}^{r} 1((z_i - z_\ell)'(\beta(\tau) + A(\tau)y_j) \geq 0) \, .$$

3.3. Tests When a Nuisance Parameter Is Present Only Under the Alternative

In this section we consider a class of testing problems for which empirical process limit theory can be usefully exploited. The testing problems considered are ones for which a nuisance parameter is present under the alternative hypothesis, but not under the null hypothesis. Such testing problems are non-standard. In consequence, the usual asymptotic distributional and opti-

mality properties of likelihood ratio (LR), Lagrange multiplier (LM), and Wald (W) tests do not apply.

Consider a parametric model with parameters θ and τ, where $\theta \in \Theta \subset R^s$, $\tau \in \mathcal{T} \subset R^v$. Let $\theta = (\beta', \delta')'$, where $\beta \in R^p$, and $\delta \in R^q$, and $s = p+q$. The null and alternative hypotheses of interest are

(3.16)
$$H_0 : \beta = 0 \quad \text{and}$$
$$H_1 : \beta \neq 0 .$$

Under the null hypothesis, the distribution of the data does not depend on the parameter τ by assumption. Under the alternative hypothesis, it does. Two examples are the following.

EXAMPLE 3: This example is a test for variable relevance. We want to test whether a regressor variable/vector Z_t belongs in a nonlinear regression model. This model is

(3.17) $Y_t = g(X_t, \delta_1) + \beta h(Z_t, \tau) + U_t , \quad U_t \sim N(0, \delta_2) , \quad t = 1, ..., T .$

The functions g and h are assumed known. The parameters $(\beta, \delta_1, \delta_2, \tau)$ are unknown. The regressors (X_t, Z_t) and/or the errors U_t are presumed to exhibit some sort of asymptotically weak temporal dependence. As an example, the term $h(Z_t, \tau)$ might be of the Box-Cox form $(Z_t^\tau - 1)/\tau$. Under the null hypothesis $H_0 : \beta = 0$, Z_t does not enter the regression function and the parameter τ is not present.

EXAMPLE 4: This example is a test of cross-sectional constancy in a nonlinear regression model. A parameter τ ($\in R^b$) partitions the sample space of some observed variable Z_t ($\in R^t$) into two regions. In one region the regression parameter is δ_1 ($\in R^p$) and in the other region it is $\delta_1 + \beta$. A test of cross-sectional constancy of the regression parameters corresponds to a test of the null hypothesis $H_0 : \beta = 0$. The parameter τ is present only under the alternative.

To be concrete, the model is

(3.18) $Y_t = \begin{cases} g(X_t, \delta_1) + U_t & \text{for } h(Z_t, \tau) > 0 \\ g(X_t, \delta_1 + \beta) + U_t & \text{for } h(Z_t, \tau) \leq 0 \end{cases} \quad \text{for } t = 1, ..., T ,$

where the errors $U_t \sim$ iid $N(0, \delta_2)$, the regressors X_t and the rv Z_t are m-dependent and identically distributed, and $g(\cdot, \cdot)$ and $h(\cdot, \cdot)$ are known real functions. For example, $h(Z_t, \tau)$ could equal $Z_t - \tau$, where the real rv Z_t is an element of X_t, an element of X_{t-d} for some integer $d \geq 1$, or Y_{t-d} for some integer $d \geq 1$. The model could be generalized to allow for more regions than two.

Problems of the sort considered above were first treated in a general way by Davies (1977, 1987). Davies proposed using the LR test. Let $\mathrm{LR}(\tau)$ denote the LR test statistic (i.e., minus two times the log likelihood ratio) when τ is specified under the alternative. For given τ, $\mathrm{LR}(\tau)$ has standard asymptotic properties (under standard regularity conditions). In particular, it converges in distribution under the null to a random variable $X^2(\tau)$ that has a χ_p^2 distribution. When τ is not given, but is allowed to take any value in $\mathcal{T}$, the LR statistic is

$$(3.19) \qquad \sup_{\tau \in \mathcal{T}} \mathrm{LR}(\tau) \ .$$

This statistic has power against a much wider variety of alternatives than the statistic $\mathrm{LR}(\tau)$ for some fixed value of τ.

To mount a test based on $\sup_{\tau \in \mathcal{T}} \mathrm{LR}(\tau)$, one needs to determine its asymptotic null distribution. This can be achieved by establishing that the stochastic process $\mathrm{LR}(\tau)$, viewed as a random function indexed by τ, converges weakly to a stochastic process $X^2(\tau)$. Then, it is easy to show that the asymptotic null distribution of $\sup_{\tau \in \mathcal{T}} \mathrm{LR}(\tau)$ is that of the supremum of the chi-square process $X^2(\tau)$. The methods discussed below can be used to provide a rigorous justification of this type of argument.

Hansen (1991) extended Davies' results to non-likelihood testing scenarios, considered LM versions of the test, and pointed out a variety of applications of such tests in econometrics.

A drawback of the sup LR test statistic is that it does not possess standard asymptotic optimality properties. Andrews and Ploberger (1991) derived a class of tests that do. They considered a weighted average power criterion that is similar to that considered by Wald (1943). Optimal test statistics turn out to be of an average exponential form:

$$(3.20) \qquad \mathrm{Exp\text{-}LR} = (1+c)^{-P/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c} \mathrm{LR}(\tau)\right) dJ(\tau) \ ,$$

where $J(\cdot)$ is a specified weight function over $\tau \in \mathcal{T}$ and c is a scalar parameter that indexes whether one is directing power against close or distant alternatives (i.e., against β small or β large). Let Exp-LM and Exp-W denote the test statistic defined as in (3.20), but with $LR(\tau)$ replaced by $LM(\tau)$ and $W(\tau)$, respectively, where the latter are defined analogously to $LR(\tau)$. The three statistics Exp-LR, Exp-LM, and Exp-W each have asymptotic optimality properties. Using empirical process results, each can be shown to have an asymptotic null distribution that is a function of the stochastic process $X^2(\tau)$ discussed above.

First, we introduce some notation. Let $\ell_T(\theta, \tau)$ denote a criterion function that is used to estimate the parameters θ and τ. The leading case is when $\ell_T(\theta, \tau)$ is the log likelihood function for the sample size T. Let $D\ell_T(\theta, \tau)$ denote the s-vector of partial derivatives of $\ell_T(\theta, \tau)$ with respect to θ. Let θ_0 denote the true value of θ under the null hypothesis H_0, i.e. $\theta_0 = (0, \delta_0')'$. (Note that $D\ell_T(\theta_0, \tau)$ depends on τ in general even though $\ell_T(\theta_0, \tau)$ does not.)

By some manipulations (e.g., see Andrews and Ploberger (1991)), one can show that the test statistics $\sup_{\tau \in \mathcal{T}} LR(\tau)$, Exp-LR, Exp-LM, and Exp-W equal a continuous real function of the normalized score process $\{D\ell_T(\theta_0, \tau)/\sqrt{T} : \tau \in \mathcal{T}\}$ plus an $o_p(1)$ term under H_0. In view of the continuous mapping theorem (e.g., see Pollard (1984, Ch. III.2)), the asymptotic null distributions of these statistics are given by the same functions of the limit process as $T \to \infty$ of $\{D\ell_T(\theta_0, \tau)/\sqrt{T} : \tau \in \mathcal{T}\}$.

More specifically, let

$$(3.21) \qquad v_T(\tau) = \frac{1}{\sqrt{T}} D\ell_T(\theta_0, \tau) \ .$$

(Note that $ED\ell_T(\theta_0, \tau) = 0$ under H_0, since these are the population first order conditions for the estimator.) Then, for some continuous function g of $v_T(\cdot)$, we have

$$(3.22) \qquad \sup_{\tau \in \mathcal{T}} LR(\tau) = g(v_T(\cdot)) + o_p(1) \quad \text{under } H_0 \ .$$

(Here, continuity is defined with respect to the uniform metric d on the space of bounded R^s-valued functions on $\mathcal{T}$, i.e., $B(\mathcal{T})$.) If $v_T(\cdot) \Rightarrow v(\cdot)$, then

$$(3.23) \qquad \sup_{\tau \in \mathcal{T}} LR(\tau) \xrightarrow{d} g(v(\cdot)) \quad \text{under } H_0,$$

which is the desired result. The distribution of $g(v(\cdot))$ yields asymptotic critical values for the test statistic $\sup_{\tau \in \mathcal{T}} LR(\tau)$. The results are analogous for Exp-LR, Exp-LM, and Exp-W.

In conclusion, if one can establish the weak convergence result, $v_T(\cdot) \Rightarrow v(\cdot)$ as $T \to \infty$, then one can obtain the asymptotic distribution of the test statistics of interest. As discussed in Section 2, the key condition for weak convergence is stochastic equicontinuity. The verification of stochastic equicontinuity for Examples 3 and 4 is discussed in Sections 4 and 5 below. Here, we specify the form of $v_T(\tau)$ in these examples.

EXAMPLE 3 (cont.): In this example, $\ell_T(\theta, \tau)$ is the log likelihood function under the assumption of iid normal errors:

$$\ell_T(\theta, \tau) = -\frac{T}{2}\log 2\pi\delta_2 - \frac{1}{2\delta_2}\Sigma_1^T (Y_t - g(X_t, \delta_1) - \beta h(Z_t, \tau))^2 \quad \text{and}$$

(3.24)

$$v_T(\tau) = \frac{1}{\sqrt{T}}D\ell_T(\theta_0, \tau) = \begin{pmatrix} \dfrac{1}{\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T U_t h(Z_t, \tau) \\[2mm] \dfrac{1}{\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T U_t \dfrac{\partial}{\partial \delta_1} g(X_t, \delta_{10}) \\[2mm] \dfrac{1}{2\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T (U_t^2 - 1) \end{pmatrix}.$$

Since τ only appears in the first term, it suffices to show that $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T U_t h(Z_t, \cdot) : T \geq 1\right\}$ is stochastically equicontinuous.

EXAMPLE 4 (cont.): In this cross-sectional constancy example, $\ell(\theta, \tau)$ is the log likelihood function under the assumption of iid normal innovations:

$$\ell_T(\theta, \tau) = -\frac{T}{2}\log 2\pi\delta_2 - \frac{1}{2\delta_2}\Sigma_1^T(Y_t - g(X_t, \delta_1)1(h(Z_t, \tau) > 0)$$

$$- g(X_t, \delta_1 + \beta)1(h(Z_t, \tau) \leq 0))^2 \quad \text{and}$$

$$(3.25)$$

$$\frac{1}{\sqrt{T}}D\ell_T(\theta_0, \tau) = \begin{pmatrix} \dfrac{1}{\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T U_t \dfrac{\partial}{\partial\delta_1}g(X_t, \delta_{10})1(h(Z_t, \tau) \leq 0) \\[2ex] \dfrac{1}{\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T U_t \dfrac{\partial}{\partial\delta_1}g(X_t, \delta_{10}) \\[2ex] \dfrac{1}{2\delta_{20}} \cdot \dfrac{1}{\sqrt{T}}\Sigma_1^T(U_t^2/\delta_{20} - 1) \end{pmatrix}.$$

Since τ only appears in the first term, it suffices to show that $\Big\{\frac{1}{\sqrt{T}}\Sigma_1^T U_t\frac{\partial}{\partial\delta_1}g(X_t, \delta_{10})1(h(Z_t, \cdot)$

$\leq 0) : T \geq 1\Big\}$ is stochastically equicontinuous.

3.4. Semiparametric Estimation

We now consider the application of stochastic equicontinuity results to semiparametric estimation problems. The approach that is discussed below is given in more detail in Andrews (1993b). Other approaches are referenced in Section 3.1 above.

Consider a two-stage estimator $\hat{\theta}$ of a finite dimensional parameter $\theta_0 \in \Theta \subset R^p$. In the first stage, an infinite dimensional parameter estimator $\hat{\tau}$ is computed, such as a nonparametric regression or density estimator or its derivative. In the second stage, the estimator $\hat{\theta}$ of θ_0 is obtained from a set of estimating equations that depend on the preliminary estimator $\hat{\tau}$. Many semiparametric estimators in the literature can be defined in this way.

By linearizing the estimating equations, one can show that the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$ depends on an empirical process $v_T(\tau)$, evaluated at the preliminary estimator $\hat{\tau}$. That is, it depends on $v_T(\hat{\tau})$. To obtain the asymptotic distribution of $\hat{\theta}$, then, one needs to obtain that of $v_T(\hat{\tau})$. If $\hat{\tau}$ converges in probability to some τ_0 (under a suitable pseudometric) and $v_T(\tau)$ is stochastically equicontinuous, then one can show that $v_T(\hat{\tau}) - v_T(\tau_0) \xrightarrow{P} 0$ and the asymptotic behavior of $\sqrt{T}(\hat{\theta} - \theta_0)$ depends on that of $v_T(\tau_0)$, which is obtained straightforwardly from an ordinary CLT. Thus, one can effectively utilize empirical process stochastic equicontinuity results in establishing the asymptotic distributions of semiparametric estimators.

We now provide some more details of the argument sketched above. Let the data consist of $\{W_t : t \le T\}$. Consider a system of p estimating equations

$$(3.26) \quad \bar{m}_T(\theta, \hat{\tau}) = \frac{1}{T}\Sigma_1^T m(\theta, \hat{\tau}) \ ,$$

where $m(\cdot, \cdot, \cdot)$ is an R^p-valued known function. Suppose the estimator $\hat{\theta}$ solves the equations

$$(3.27) \quad \sqrt{T}\bar{m}_T(\hat{\theta}, \hat{\tau}) = 0$$

(at least with probability that goes to one as $T \to \infty$). These equations might be the first order conditions from some minimization problem.

We suppose consistency of $\hat{\theta}$ has already been established, i.e., $\hat{\theta} \xrightarrow{P} \theta_0$ (see Andrews (1993b) for sufficient conditions). We wish to determine the asymptotic distribution of $\hat{\theta}$. When $m(W_t, \theta, \tau)$ is a smooth function of θ, the following approach can be used. Element by element mean value expansions stacked yield

$$(3.28) \quad o_p(1) = \bar{m}_T(\hat{\theta}, \hat{\tau}) = \sqrt{T}\bar{m}_T(\theta_0, \hat{\tau}) + \frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\tau})\sqrt{T}(\hat{\theta} - \theta_0) \ ,$$

where θ^* lies between $\hat{\theta}$ and θ_0 (and θ^* may differ from row to row in $\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\tau})$). Under suitable conditions,

$$(3.29) \quad \frac{\partial}{\partial\theta}\bar{m}_T(\theta^*, \hat{\tau}) \xrightarrow{P} M = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E \frac{\partial}{\partial\theta'}\bar{m}_T(\theta_0, \tau_0) \ .$$

Thus,

$$(3.30) \quad \begin{aligned} \sqrt{T}(\hat{\theta} - \theta_0) &= -(M^{-1} + o_p(1))\sqrt{T}\bar{m}_T(\theta_0, \hat{\tau}) \\ &= -(M^{-1} + o_p(1))\left[\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - \bar{m}_T^*(\theta_0, \hat{\tau})) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\tau})\right] \ , \end{aligned}$$

where $\bar{m}_T^*(\theta, \tau) = \frac{1}{T}\Sigma_1^T Em(W_t, \theta, \tau)$.

Again under suitable conditions, either

$$(3.31) \quad \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\tau}) \xrightarrow{P} 0 \quad \text{or} \quad \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\tau}) \xrightarrow{d} N(0, A) \ ,$$

for some covariance matrix A, see Andrews (1993b).

Let

$$(3.32) \qquad v_T(\tau) = \sqrt{T}(\bar{m}_T(\theta_0, \tau) - \bar{m}_T^*(\theta_0, \tau)) \ .$$

Note that $v_T(\cdot)$ is a stochastic process indexed by an infinite dimensional parameter in this case. This differs from the other examples in this section for which τ is finite dimensional.

Under standard conditions, one can establish that

$$(3.33) \qquad v_T(\tau_0) \xrightarrow{d} N(0, S)$$

for some covariance matrix S, by applying an ordinary CLT. If, in addition, one can show that

$$(3.34) \qquad v_T(\hat{\tau}) - v_T(\tau_0) \xrightarrow{p} 0 \ ,$$

then we obtain

$$
\begin{aligned}
\sqrt{T}(\hat{\theta} - \theta_0) &= -(M^{-1} + o_p(1))[v_T(\hat{\tau}) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\tau})] \\
(3.35) \qquad &= -M^{-1}[v_T(\tau_0) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\tau})] + o_p(1) \\
&\xrightarrow{d} N(0, M^{-1}(S+A)(M^{-1})') \ ,
\end{aligned}
$$

which is the desired result.

To prove (3.34), we can use the stochastic equicontinuity property. Suppose

$$(3.36)$$

(i) $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous

for some choice of $\mathcal{T}$ and pseudometric ρ on $\mathcal{T}$,

(ii) $P(\hat{\tau} \in \mathcal{T}) \to 1$, and

(iii) $\rho(\hat{\tau}, \tau_0) \xrightarrow{p} 0$,

then (3.34) holds (as shown below).

Note that there exist tradeoffs between conditions (i), (ii), and (iii) of (3.36) in terms of the difficulty of verification and the strength of the regularity conditions needed. For example, a larger set $\mathcal{T}$ makes it more difficult to verify (i), but easier to verify (ii). A stronger pseudometric ρ makes it easier to verify (i), but more difficult to verify (iii).

Since the sufficiency of (3.36) for (3.34) is the key to the approach considered here, we pro-

vide a proof of this simple result. We have: $\forall \varepsilon > 0$, $\forall \eta > 0$, $\exists \delta > 0$ such that

$$\overline{\lim_{T \to \infty}} \, P(|v_T(\hat{\tau}) - v_T(\tau_0)| > \eta)$$

$$\leq \overline{\lim_{T \to \infty}} \, P(|v_T(\hat{\tau}) - v_T(\tau_0)| > \eta, \, \hat{\tau} \in \mathcal{T}, \, \rho(\hat{\tau}, \tau_0) \leq \delta)$$

$$(3.37) \qquad + \overline{\lim_{T \to \infty}} \, P(\hat{\tau} \notin \mathcal{T} \text{ or } \rho(\hat{\tau}, \tau_0) > \delta)$$

$$\leq \overline{\lim_{T \to \infty}} \, P^*(\sup_{\tau \in \mathcal{T}, \rho(\tau, \tau_0) < \delta} |v_T(\tau) - v_T(\tau_0)| > \eta)$$

$$< \varepsilon \, ,$$

where the term on the third line of (3.37) is zero by (ii) and (iii) and the last inequality holds by (i). Since $\varepsilon > 0$ is arbitrary, (3.34) follows.

To conclude, one can establish the $\sqrt{T}$-consistency and asymptotic normality of the semiparametric estimator $\hat{\theta}$ if one can establish, among other things, that $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous. Next, we consider the application of this approach to two examples and illustrate the form of $v_T(\cdot)$ in these examples. In Sections 4 and 5, we discuss the verification of stochastic equicontinuity when $\mathcal{M} = \{m(\cdot, \tau) : \tau \in \mathcal{T}\}$ is an infinite dimensional class of functions.

EXAMPLE 5: This example considers a weighted least squares (WLS) estimator of the partially linear regression (PLR) model. The PLR model is given by

$$(3.38) \qquad Y_t = X_t'\theta_0 + g(Z_t) + U_t \quad \text{and} \quad E(U_t|X_t, Z_t) = 0 \text{ a.s.}$$

for $t = 1, \ldots, T$, where the real function $g(\cdot)$ is unknown, $W_t = (Y_t, X_t', Z_t')'$ is iid or m-dependent and identically distributed, $Y_t, U_t \in R$, $X_t, \theta_0 \in R^p$, and $Z_t \in R^{k_a}$. This model is also discussed by Härdle and Linton (1993) in this *Handbook*. The WLS estimator is defined for the case where the conditional variance of U_t given (X_t, Z_t) depends only on Z_t. This estimator is a weighted version of Robinson's (1988) semiparametric LS estimator. The PLR model with heteroskedasticity of the above form can be generated by a sample selection model with non-parametric selection equation (e.g., see Andrews (1993b)). Let $\tau_{10}(Z_t) = E(Y_t|Z_t)$, $\tau_{20}(Z_t)$

$= E(X_t|Z_t)$, $\tau_{30}(Z_t) = E(U_t^2|Z_t)$, and $\tau_0 = (\tau_{10}, \tau'_{20}, \tau_{30})'$. Let $\hat{\tau}_j(\cdot)$ be an estimator of $\tau_{j0}(\cdot)$ for $j = 1, 2, 3$. The semiparametric WLS estimator of the PLR model is given by

$$
\begin{aligned}
(3.39) \quad \hat{\theta} = {}& [\Sigma_1^T \xi(W_t)(X_t - \hat{\tau}_2(Z_t))(X_t - \hat{\tau}_2(Z_t))'/\hat{\tau}_3(Z_t)]^{-1} \\
& \Sigma_1^T \xi(W_t)(X_t - \hat{\tau}_2(Z_t))(Y_t - \hat{\tau}_1(Z_t))/\hat{\tau}_3(Z_t) ,
\end{aligned}
$$

where $\xi(W_t) = 1(Z_t \in \mathcal{Z}^*)$ is a trimming function and $\mathcal{Z}^*$ is a bounded subset of R^{k_a}. This estimator is of the form (3.16)-(3.17) with

$$
(3.40) \quad m(W_t, \hat{\theta}, \hat{\tau}) = \xi(W_t)[Y_t - \hat{\tau}_1(Z_t) - (X_t - \hat{\tau}_2(Z_t))'\theta][X_t - \hat{\tau}_2(Z_t)]/\hat{\tau}_3(Z_t) .
$$

To establish the asymptotic normality of $\hat{\tau}$ using the approach above, one needs to establish stochastic equicontinuity for the empirical process $\nu_T(\cdot)$ when the class of functions $\mathcal{M}$ is given by

$$
\begin{aligned}
(3.41) \quad & \mathcal{M} = \{m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T}\} , \quad \text{where} \\
& m(w, \theta_0, \tau) = \xi(w)[y - \tau_1(z) - (x - \tau_2(z))'\theta_0][x - \tau_2(z)]/\tau_3(z) ,
\end{aligned}
$$

$w = (y, x', z')'$, $\tau = (\tau_1, \tau'_2, \tau_3)'$, and $\mathcal{T}$ is as defined below. Here, the elements $\tau \in \mathcal{T}$ are possible realizations of the vector nonparametric estimator $\hat{\tau}$. By definition, $\mathcal{Z} \subset R^{k_a}$ is the domain of $\tau_j(z)$ for $j = 1, 2, 3$ and $\mathcal{Z}$ includes the support of Z_t $\forall t \geq 1$. By assumption, the trimming set $\mathcal{Z}^* \subset \mathcal{Z}$. If $\mathcal{Z}^* = \mathcal{Z}$, then no trimming occurs and $\xi(w)$ is redundant. If $\mathcal{Z}^*$ is a proper subset of $\mathcal{Z}$, then trimming occurs and the WLS estimator $\hat{\theta}$ is based on only non-trimmed observations.

EXAMPLE 6: This example considers generalized method of moments (GMM) estimators of parameters defined by conditional moments restrictions (CMR).

In this example, θ_0 is the unique parameter vector that solves the equations

$$
(3.42) \quad E(\psi(Z_t, \theta)|X_t) = 0 \quad \text{a.s.} \quad \forall t \geq 1
$$

for some specified R^α-valued function $\psi(\cdot, \cdot)$, where $X_t \in R^{k_a}$. Examples of this model in econometrics are quite numerous, see Chamberlain (1987) and Newey (1990).

Let $\Omega_0(X_t) = E(\psi(Z_t, \theta_0)\psi(Z_t, \theta_0)'|X_t)$, $\Delta_0(X_t) = E\left[\dfrac{\partial}{\partial \theta'}\psi(Z_t, \theta_0)|X_t\right]$ and $\tau_0(X_t) = \Delta_0(X_t)'$

$\times \; \Omega_0^{-1}(X_t)$. By assumption, $\Omega_0(\cdot)$, $\Delta_0(\cdot)$, and $\tau_0(\cdot)$ do not depend on t. Let $\hat\Omega(\cdot)$ and $\hat\Delta(\cdot)$ be nonparametric estimators of $\Omega_0(\cdot)$ and $\Delta_0(\cdot)$. Let $\hat\tau(\cdot) = \hat\Delta(\cdot)'\hat\Omega^{-1}(\cdot)$. Let $W_t = (Z_t', X_t')'$.

A GMM estimator $\hat\theta$ of θ_0 minimizes

$$(3.43) \quad \left[\Sigma_1^T \hat\tau(X_t)\psi(Z_t, \theta)\right]' \hat\gamma \Sigma_1^T \hat\tau(X_t)\psi(Z_t, \theta) \quad \text{over} \quad \theta \in \Theta \subset R^{p^*},$$

where $\hat\gamma$ is a data-dependent weight matrix. To obtain the asymptotic distribution of this estimator using the approach above, we need to establish a stochastic equicontinuity result for the empirical process $v_T(\cdot)$ when the class of functions $\mathcal{M}$ is given by

$$(3.44) \quad \begin{aligned} \mathcal{M} &= \{m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T}\} \,, \quad \text{where} \\ m(w, \theta_0, \tau) &= \tau(x)\psi(z, \theta_0) = \Delta(x)'\Omega^{-1}(x)\psi(z, \theta_0) \,, \end{aligned}$$

$w = (z', x')'$ and $\mathcal{T}$ is defined below.

4. STOCHASTIC EQUICONTINUITY VIA SYMMETRIZATION

4.1. Primitive Conditions for Stochastic Equicontinuity

In this section we provide primitive conditions for stochastic equicontinuity. These conditions are applied to some of the examples of Section 3 in Section 4.2 below. We utilize an empirical process result of Pollard (1990) altered to encompass m-dependent rather than independent rv's and reduced in generality somewhat to achieve a simplification of the conditions. This result depends on a condition, which we refer to as *Pollard's entropy condition*, that is based on how well the functions in $\mathcal{M}$ can be approximated by a finite number of functions, where the distance between functions is measured by the largest $L^2(Q)$ distance over all distributions Q that have finite support. The main purpose of this section is to establish primitive conditions under which the entropy condition holds. Following this, a number of examples are provided to illustrate the ease of verification of the entropy condition.

First, we note that stochastic equicontinuity of a vector-valued empirical process (i.e., $s > 1$) follows from the stochastic equicontinuity of each element of the empirical process. In consequence, we focus attention on real-valued empirical processes ($s = 1$).

The pseudometric ρ on $\mathcal{T}$ is defined in this section by

$$(4.1) \qquad \rho(\tau_1, \tau_2) = \sup_{N \geq 1}\left(\frac{1}{N}\Sigma_1^N E(m(W_t, \tau_1) - m(W_t, \tau_2))^2\right)^{1/2} .[3]$$

Let Q denote a probability measure on $\mathcal{W}$. For a real function f on $\mathcal{W}$, let $Qf^2 = \int_{\mathcal{W}} f^2(w)dQ(w)$. Let $\mathcal{F}$ be a class of functions in $L^2(Q)$. The $L^2(Q)$ cover numbers of $\mathcal{F}$ are defined as follows:

DEFINITION: For any $\varepsilon > 0$, the *cover number* $N_2(\varepsilon, Q, \mathcal{F})$ is the smallest value of n for which there exist functions $f_1, ..., f_n$ in $\mathcal{F}$ such that $\min_{j \leq n}\left(Q(f - f_j)^2\right)^{1/2} \leq \varepsilon \quad \forall f \in \mathcal{F}$. $N_2(\varepsilon, Q, \mathcal{F}) = \infty$ if no such n exists.

The log of $N_2(\varepsilon, Q, \mathcal{F})$ is referred to as the $L^2(Q)$ ε-*entropy* of $\mathcal{F}$. Let $\mathcal{Q}$ denote the class of all probability measures Q on $\mathcal{W}$ that concentrate on a finite set. The following entropy/cover number condition was introduced in Pollard (1982):

DEFINITION: A class $\mathcal{F}$ of real functions defined on $\mathcal{W}$ satisfies *Pollard's entropy condition* if

$$(4.2) \qquad \int_0^1 \sup_{Q \in \mathcal{Q}}\left(\log N_2\left(\varepsilon\left(QF^2\right)^{1/2}, Q, \mathcal{F}\right)\right)^{1/2} d\varepsilon < \infty,$$

where F is some envelope function for $\mathcal{F}$, i.e. F is a real function on $\mathcal{W}$ for which $|f(\cdot)| \leq F(\cdot)$ $\forall f \in \mathcal{F}$.

As $\varepsilon \downarrow 0$, the cover number $N_2\left(\varepsilon\left(QF^2\right)^{1/2}, Q, \mathcal{F}\right)$ increases. Pollard's entropy condition requires that it cannot increase too quickly as $\varepsilon \downarrow 0$. This restricts the complexity/size of $\mathcal{F}$ and does so in a way that is sufficient for stochastic equicontinuity given suitable moment and temporal dependence assumptions. In particular, the following three assumptions are sufficient for stochastic equicontinuity.

ASSUMPTION A: $\mathcal{M}$ satisfies Pollard's entropy condition with some envelope $\bar{M}$.

ASSUMPTION B: $\overline{\lim}_{T-\infty} \frac{1}{T}\Sigma_1^T E\bar{M}^{2+\delta}(W_t) < \infty$ for some $\delta > 0$, where $\bar{M}$ is as in Assumption A.

ASSUMPTION C: $\{W_t : t \le T, T \ge 1\}$ *is an m-dependent triangular array of* rv's.

THEOREM 1 (Pollard): *Under Assumptions* A-C, $\{v_T(\cdot) : T \ge 1\}$ *is stochastically equicontinuous with* ρ *given by* (4.1).

COMMENTS: 1. Theorem 1 is proved using a symmetrization argument. In particular, one obtains a maximal inequality for $v_T(\tau)$ by showing that $\sup_{\tau \in \mathcal{T}} |v_T(\tau)|$ is less variable than $\sup_{\tau \in \mathcal{T}} |\frac{1}{\sqrt{T}} \Sigma_1^T \sigma_t m(W_t, \tau)|$, where $\{\sigma_t : t \le T\}$ are iid rv's that are independent of $\{W_t : t \le T\}$ and have Rudemacher distribution (i.e., σ_t equals $+1$ or -1, each with probability 1/2). Conditional on $\{W_t\}$ one performs a chaining argument that relies on Hoeffding's inequality for tail probabilities of sums of bounded, mean zero, independent rv's. The bound in this case is small when the average sum of squares of the bounds on the individual rv's is small. In the present case, the latter is just $\frac{1}{T} \Sigma_1^T m^2(W_t, \tau)$. The maximal inequality ultimately is applied to the empirical measure constructed from differences of the form $m(W_t, \tau_1) - m(W_t, \tau_2)$ rather than to just $m(W_t, \tau)$. In consequence, the measure of distance between $m(\cdot, \tau_1)$ and $m(\cdot, \tau_2)$ that makes the bound effective is an $L^2(P_T)$ pseudometric, where P_T denotes the empirical distribution of $\{W_t : t \le T\}$. This pseudometric is random and depends on T, but is conveniently dominated by the largest $L^2(Q)$ pseudometric over all distributions Q with finite support. This explains the appearance of the latter in the definition of Pollard's entropy condition. To see why Pollard's entropy condition takes the precise form given above, one has to inspect the details of the chaining argument. The interested reader can do so, see Pollard (1990, Sec. 3).

2. When Assumptions A-C hold, $\mathcal{T}$ is totally bounded under the pseudometric ρ provided ρ is equivalent to the pseudometric ρ^* defined by $\rho^*(\tau_1, \tau_2) = \underline{\lim}_{N \to \infty} \left[\frac{1}{N} \Sigma_1^N E(m(W_t, \tau_1) - m(W_t, \tau_2))^2 \right]^{1/2}$. By equivalent, we mean that $\rho^*(\tau_1, \tau_2) \ge C\rho(\tau_1, \tau_2)$ $\forall \tau_1, \tau_2 \in \mathcal{T}$ for some $C > 0$. $(\rho^*(\tau_1, \tau_2) \le \rho(\tau_1, \tau_2)$ holds automatically.) Of course, ρ equals ρ^* if the rv's W_t are identically distributed. The proof of total boundedness is analogous to that given in the proof of Theorem 10.7 in Pollard (1990).

Combinatorial arguments have been used to establish that certain classes of functions, often referred to as Vapnik-Červonenkis (VC) classes of one sort or another, satisfy Pollard's entropy

condition, see Pollard (1984, Ch. 2; 1990, Sec. 4) and Dudley (1987). Here we consider the most important of these VC classes for applications (type I classes below) and we show that several other classes of functions satisfy Pollard's entropy condition. These include Lipschitz functions indexed by finite dimensional parameters (type II classes) and infinite dimensional classes of smooth functions (type III classes). The latter are important for applications to semiparametric and nonparametric problems because they cover realizations of nonparametric estimators (under suitable assumptions).

Having established that Pollard's entropy condition holds for several useful classes of functions, we proceed below to show that functions from these classes can be "mixed and matched," e.g., by addition, multiplication, and division, to obtain new classes that satisfy Pollard's entropy condition. In consequence, one can routinely build up fairly complicated classes of functions that satisfy Pollard's entropy condition. In particular, one can build up classes of functions that are suitable for use in the examples above.

The first class of functions we consider are applicable in the non-differentiable M-estimator Examples 1 and 2, see Section 3.2 above.

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type I class* if it is of the form (a) $\mathcal{F} = \{f : f(w) = w'\xi \quad \forall w \in \mathcal{W}$ for some $\xi \in \Psi \subset R^k\}$ or (b) $\mathcal{F} = \{f : f(w) = h(w'\xi)$ $\forall w \in \mathcal{F}$ for some $\xi \in \Psi \subset R^k, h \in V_K\}$, where V_K is some set of functions from R to R each with total variation less than or equal to $K < \infty$.

Common choices for h in (b) include the indicator function, the sign function, and Huber ψ-functions, among others.

For the more knowledgeable reader (concerning empirical processes), we note that it is sometimes useful to extend the definition of Type I classes of functions to include various classes of functions called VC classes. By definition, such classes include (i) classes of indicator functions of VC sets, (ii) VC major classes of uniformly bounded functions, (iii) VC hull classes, (iv) VC subgraph classes, and (v) VC subgraph hull classes, where each of these classes is as

The second class of functions we consider contains functions that are indexed by a finite dimensional parameter and are Lipschitz with respect to that parameter:

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type II class* if each function f in $\mathcal{F}$ satisfies: $f(\cdot) = f(\cdot, \tau)$ for some $\tau \in \mathcal{T}$, where $\mathcal{T}$ is some bounded subset of Euclidean space and $f(\cdot, \tau)$ is Lipschitz in τ, i.e.,

(4.3) $|f(\cdot, \tau_1) - f(\cdot, \tau_2)| \leq B(\cdot)|\tau_1 - \tau_2| \quad \forall \tau_1, \tau_2 \in \mathcal{T}$

for some function $B(\cdot) : \mathcal{W} \to R$.

The third class of functions we consider is an infinite dimensional class of functions that is useful for semiparametric and nonparametric applications such as Examples 5 and 6. This class is more complicated to define than Type I and II classes. The reader may wish to skip this section on first reading and move ahead to Theorem 2.

The third class of functions contains functions that depend on $w = (w_a', w_b')'$ only through a subvector w_a that has dimension $k_a \leq k$. The functions are smooth on a restricted subset of $\mathcal{W}$ and are equal to a constant elsewhere. Define $\mathcal{W}_a = \{w_a \in R^{k_a} : \exists w_b \text{ s.t. } (w_a', w_b')' \in \mathcal{W}\}$. For $w, h \in R^k$, we write $w = (w_a', w_b')'$ and $h = (h_a', h_b')'$.

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type III class* if

(i) each f in $\mathcal{F}$ depends on w only through a subvector w_a of dimension $k_a \leq k$,

(ii) for some real number $q > k_a/2$, some constant $C < \infty$, and some set $\mathcal{W}_a^*$, which is a subset of $\mathcal{W}_a$ and is a connected compact subset of R^{k_a}, each $f \in \mathcal{F}$ satisfies the smoothness condition: $\forall w \in \mathcal{W}$ and $w+h \in \mathcal{W}$,

(4.4)
$$f(w+h) = \sum_{v=0}^{[q]} \frac{1}{v!} B_v(h_a, w_a) + R(h_a, w_a) \quad \text{and}$$

$$R(h_a, w_a) \leq C|h_a|^q ,$$

where $B_v(h_a, w_a)$ is homogeneous of degree v in h_a and $(q, C, \mathcal{W}_a^*)$ do not depend on f, w, or h,

(iii) for some constant K and all $f \in \mathcal{F}$, $f(w) = K \quad \forall w \in \mathcal{W}$ such that $w_a \in \mathcal{W}_a - \mathcal{W}_a^*$.

where $B_\nu(h_a, w_a)$ is homogeneous of degree ν in h_a and $(q, C, \mathcal{W}_a^*)$ do not depend on f, w, or h,

(iii) for some constant K and all $f \in \mathcal{F}$, $f(w) = K$ $\forall w \in \mathcal{W}$ such that $w_a \in \mathcal{W}_a - \mathcal{W}_a^*$.

Typically the expansion of $f(w+h)$ in (4.4) is a Taylor expansion of order $[q]$ and the function $B_\nu(h_a, w_a)$ is the ν-th differential of f at w, i.e., $B_\nu(h_a, w_a) = \Sigma_\nu \dfrac{\nu!}{\nu_1! \cdots \nu_{k_a}!}$

$\dfrac{\partial^\nu f(w)}{\partial w_{a1}^{\nu_1} \cdots \partial w_{ak_a}^{\nu_{k_a}}} h_{a1}^{\nu_1} \cdots h_{ak_a}^{\nu_{k_a}}$, where Σ_ν denotes the sum over all ordered k_a-tuples $(\nu_1, ..., \nu_{k_a})$ of nonnegative integers such that $\nu_1 + \cdots + \nu_{k_a} = \nu$, $w_a = (w_{a1}, ..., w_{ak_a})'$, and $h = (h_{a1}, ..., h_{ak_a})'$.

Sufficient conditions for condition (ii) above are: (a) for some real number $q > k_a/2$, $f \in \mathcal{F}$ has partial derivatives of order $[q]$ on $\mathcal{W}^* = \{w \in \mathcal{W}: w_a \in \mathcal{W}_a^*\}$, (b) the $[q]$-th order partial derivatives of f satisfy a Lipschitz condition with exponent $q - [q]$ and some Lipschitz constant C^* that does not depend on f $\forall f \in \mathcal{F}$, and (c) $\mathcal{W}_a^*$ is a convex compact set.

The envelope of a type III class $\mathcal{F}$ can be taken to be a constant function, since the functions in $\mathcal{F}$ are uniformly bounded in absolute value over $w \in \mathcal{W}$ and $f \in \mathcal{F}$.

Type III classes can be extended to allow $\mathcal{W}_a$ to be a finite union of connected compact subsets of R^{k_a}. In this case, (4.4) only needs to hold $\forall w \in \mathcal{W}$ and $w+h \in \mathcal{W}$ such that w_a and $w_a + h_a$ are in the same connected set in $\mathcal{W}_a^*$.

In applications, type III classes of functions typically are classes of realizations of nonparametric function estimates. Since these realizations usually depend on only a subvector W_{at} of $W_t = (W_{at}', W_{bt}')'$, it is advantageous to define type III classes to contain functions that may depend on only part of W_t. By "mixing and matching" functions of type III with functions of types I and II (see below), classes of functions are obtained that depend on all of w.

In applications where the subvector W_{at} of W_t is a bounded rv, one may have $\mathcal{W}_a^* = W_a$. In applications where W_{at} is an unbounded rv, $\mathcal{W}_a^*$ must be a proper subset of $\mathcal{W}_a$ for $\mathcal{F}$ to be a type III class. A common case where the latter arises in the examples of Andrews (1993b) is when W_{at} is an unbounded rv, all the observations are used to estimate a nonparametric function $\tau_0(w_a)$ for $w_a \in \mathcal{W}_a$, and the semiparametric estimator only uses observations W_t such that W_{at} is in a bounded set $\mathcal{W}_a^*$. In this case, one sets the nonparametric estimator of $\tau_0(w_a)$ equal to

THEOREM 2: *If $\mathcal{F}$ is a class of functions of type I, II, or III, then Pollard's entropy condition* (4.2) *(i.e. Assumption A) holds with envelope $F(\cdot)$ given by* $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)|$, $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)| \vee B(\cdot)$, *or* $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)|$, *respectively, where $\vee$ is the maximum operator.*

COMMENT: For type I classes, the result of Theorem 2 follows from results in the literature such as Pollard (1984, Ch. II) and Dudley (1987) (see the Appendix for details). For type II classes, Theorem 2 is established directly. It is similar to Lemma 2.13 of Pakes and Pollard (1989). For type III classes, Theorem 2 is established using uniform metric entropy results of Kolmogorov and Tihomirov (1961).

We now show how one can "mix and match" functions of types I, II, and III to obtain a wide variety of classes that satisfy Pollard's entropy condition (Assumption A). Let $\mathcal{G}$ and $\mathcal{G}^*$ be classes of $r \times s$ matrix-valued functions defined on $\mathcal{W}$ with scalar envelopes G and G^* respectively (i.e., $G : \mathcal{W} \to R$ and $|g_{ij}(\cdot)| \leq G(\cdot)$ $\forall i = 1, ..., r$, $\forall j = 1, ..., s$, $\forall g \in \mathcal{G}$). Let g and g^* denote generic elements of $\mathcal{G}$ and $\mathcal{G}^*$. Let $\mathcal{H}$ be defined as $\mathcal{G}$ is, but with $s \times u$-valued functions. Let h denote a generic element of $\mathcal{H}$. We say that a class of matrix-valued functions $\mathcal{G}$, $\mathcal{G}^*$, or $\mathcal{H}$ satisfies Pollard's entropy condition or is of type I, II, or III if it is element by element for each of the rs or su elements of its functions.

Let $\mathcal{G} \oplus \mathcal{G}^* = \{g + g^*\}$ $(= \{g + g^* : g \in \mathcal{G}, g^* \in \mathcal{G}^*\})$, $\mathcal{G}\mathcal{H} = \{gh\}$, $\mathcal{G} \vee \mathcal{G}^* = \{g \vee g^*\}$, $\mathcal{G} \wedge \mathcal{G}^* = \{g \wedge g^*\}$, and $|\mathcal{G}| = \{|g|\}$, where $\vee$, $\wedge$, and $|\cdot|$ denote the element by element maximum, minimum, and absolute value operators respectively. If $r = s$ and $g(w)$ is non-singular $\forall w \in \mathcal{W}$ and $\forall g \in \mathcal{G}$, let $\mathcal{G}^{-1} = \{g^{-1}\}$. Let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue of the matrix $\cdot$.

THEOREM 3: *If $\mathcal{G}$, $\mathcal{G}^*$, and $\mathcal{H}$ satisfy Pollard's entropy condition with envelopes G, G^*, and H, respectively, then so do each of the following classes (with envelopes given in parentheses): $\mathcal{G} \cup \mathcal{G}^*$ ($G \vee G^*$), $\mathcal{G} \oplus \mathcal{G}^*$ ($G + G^*$), $\mathcal{G}\mathcal{H}$ ($(G \vee 1)(H \vee 1)$), $\mathcal{G} \vee \mathcal{G}^*$ ($G \vee G^*$), $\mathcal{G} \wedge \mathcal{G}^*$ ($G \vee G^*$), and $|\mathcal{G}|$ (G). If in addition $r = s$ and $\mathcal{G}^{-1}$ has a finite envelope $\tilde{G}$, then $\mathcal{G}^{-1}$ also satisfies Pollard's entropy condition (with envelope $(G \vee 1)^2 \tilde{G}^2$).*

$|\mathcal{G}|$ (G). *If in addition* $r = s$ *and* $\mathcal{G}^{-1}$ *has a finite envelope* $\tilde{G}$, *then* $\mathcal{G}^{-1}$ *also satisfies Pollard's entropy condition (with envelope* $(G \vee 1)^2 \tilde{G}^2$ *).*

COMMENTS: 1. The stability properties of Pollard's entropy condition given in Theorem 3 are quite similar to stability properties of packing numbers considered in Pollard (1990).

2. If $r = s$ and $\inf_{g \in \mathcal{G}} \inf_{w \in \mathcal{W}} \lambda_{\min}(g(w)) > 0$, then $\mathcal{G}^{-1}$ has an envelope that is uniformly bounded by a finite constant.

4.2. Examples

We now show how Theorems 1-3 can be applied in the Examples of Section 3 to obtain stochastic equicontinuity of $v_T(\cdot)$.

EXAMPLE 1 (cont.): By Theorems 1-3, the following conditions are sufficient for stochastic equicontinuity of $v_T(\cdot)$ in this example:

(i) $\{(Y_t, X_t) : t \geq 1\}$ is an m-dependent sequence of rv's.

(ii) $\overline{\lim}_{T \to \infty} \frac{1}{T} \Sigma_1^T E \|X_t\|^{2+\delta} < \infty$ for some $\delta > 0$.

(4.5)

(iii) $\{\psi_2(\cdot, \tau) : \tau \in \mathcal{T}\}$ satisfies Pollard's entropy condition with envelope

$$\sup_{\tau \in \mathcal{T}} |\psi_2(\cdot, \tau)| \quad \text{and} \quad \overline{\lim}_{T \to \infty} \frac{1}{T} \Sigma_1^T E\left[(\|X_t\|^{2+\delta} + 1) \sup_{\tau \in \mathcal{T}} |\psi_2(W_t, \tau)|^{2+\delta} \right] < \infty$$

for some $\delta > 0$.

(iv) $\psi_1(\cdot)$ is a function of bounded variation.

Sufficiency of conditions (i)-(iv) for stochastic equicontinuity of $v_T(\cdot)$ is established as follows: The sets $\{g : g(w) = \psi_1(y - x'\tau)$ for some $\tau \in \mathcal{T}\}$ and $\{h : h(w) = x\}$ are type I classes with envelopes C_1 and $|x|$, respectively, for some constant $C_1 < \infty$, and hence satisfy Pollard's entropy condition by Theorem 2. This result, condition (iii), and the $\mathcal{GH}$ result of Theorem 3 show that $\mathcal{H}$ satisfies Pollard's entropy condition with envelope $(|x| \vee 1)(\sup_{\tau \in \mathcal{T}} |\psi_2(w, \tau)| \vee 1)$. Stochastic equicontinuity now follows from Theorem 1, since Assumption B is implied by conditions (ii) and (iii).

For the particular M-estimators considered in Example 1 above, condition (iv) is always satisfied and condition (iii) is automatically satisfied given (ii) whenever $\psi_2 = 1$ or $\psi_2(w, \tau)$ $= 1(x'\tau > 0)$. When $\psi_2(w, \tau) = y - x'\tau$, $\psi_2(w, \tau) = 1(x'\tau > 0)[(y - x'\tau) \wedge x'\tau]$, or $\psi_2(w, \tau) = 1(y < 2x'\tau)(y - x'\tau)$, condition (iii) is satisfied provided $\mathcal{T}$ is bounded and $\overline{\lim}_{T \to \infty} \frac{1}{T}\Sigma_1^T[E|U_t|^{2+\delta} + E|X_t|^{4+\delta} + E|U_t X_t|^{2+\delta}] < \infty$ for some $\delta > 0$. This follows from Theorem 3, since $\{1(x'\tau > 0) : \tau \in \mathcal{T}\}$, $\{y - x'\tau : \tau \in \mathcal{T}\}$, $\{x'\tau : \tau \in \mathcal{T}\}$, and $\{1(y < 2x'\tau) : \tau \in \mathcal{T}\}$ are type I classes with envelopes 1, $|u| + |x|\sup_{\tau \in \mathcal{T}}|\tau - \tau_0|$, $|x|\sup_{\tau \in \mathcal{T}}|\tau|$, and 1, respectively, where $u = y - x'\tau_0$.

EXAMPLE 2 (cont.): In the method of simulated moments example, the following conditions are sufficient for stochastic equicontinuity of $v_T(\cdot)$:

(i) $\{(D_t, Z_t, Y_{t1}, ..., Y_{ts}) : t \geq 1\}$ is an m-dependent sequence of rv's.

(4.6)

(ii) $\{g(\cdot, \tau) : \tau \in \mathcal{T}\}$ is a type II class of functions with Lipschitz function $B(\cdot)$ that satisfies $\overline{\lim}_{T \to \infty} \frac{1}{T}\Sigma_1^T\left(EB^{2+\delta}(Z_t) + E\sup_{\tau \in \mathcal{T}}|g(Z_t, \tau)|^{2+\delta}\right) < \infty$

for some $\delta > 0$.

Note that condition (ii) holds if $g(w, \tau)$ is differentiable in τ $\forall w \in \mathcal{W}$, $\forall \tau \in \mathcal{T}$, $\mathcal{T}$ is open, and $\overline{\lim}_{T \to \infty} \frac{1}{T}\Sigma_1^T\left(E \sup_{\tau \in \mathcal{T}}|\frac{\partial}{\partial \tau}g(Z_t, \tau)|^{2+\delta} + E \sup_{\tau \in \mathcal{T}}|g(Z_t, \tau)^{2+\delta}\right) < \infty$ for some $\delta > 0$.

Sufficiency is established as follows: Classes of functions of the form $\{1((z_i - z_t)'(\beta(\tau) + A(\tau)y_j) \geq 0) : \tau \in \mathcal{T} \subset R^p\}$ are type I classes with envelopes equal to 1 (by including products $z_i y_j$ and $z_t y_j$ as additional elements of w) and hence satisfy Pollard's entropy condition by Theorem 2. $\{g(\cdot, \tau) : \tau \in \mathcal{T}\}$ also satisfies Pollard's entropy condition with envelope $1 \vee \sup_{\tau \in \mathcal{T}}|g(\cdot, \tau)| \vee B(\cdot)$ by condition (ii) and Theorem 2. The $\mathcal{GH}$ result of Theorem 3 now implies that $\mathcal{H}$ satisfies Pollard's entropy condition with envelope $1 \vee \sup_{\tau \in \mathcal{T}}|g(\cdot, \tau)| \vee B(\cdot)$. Stochastic equicontinuity now follows by Theorem 1.

EXAMPLE 5 (cont.): By applying Theorems 1-3, we find the following conditions are sufficient for stochastic equicontinuity of $v_T(\cdot)$ in the WLS/PLR example. With some abuse of notation, let $\tau_j(w)$ denote a function on $\mathcal{W}$ that depends on w only through the k_a-subvector z and equals $\tau_j(z)$ above for $j = 1, 2, 3$. The sufficient conditions are:

(4.7)

(i) $\{(Y_t, X_t, Z_t) : t \geq 1\}$ is an m-dependent identically distributed sequence of rv's.

(ii) $E|Y_t - X_t'\theta_0|^{2+\delta} + E|X_t|^{2+\delta} + E|(Y_t - X_t'\theta_0)X_t|^{2+\delta} < \infty$ for some $\delta > 0$.

(iii) $\mathcal{T} = \{\tau : \tau = (\tau_1, \tau_2, \tau_3), \tau_j \in \mathcal{T}_j \text{ for } j = 1, 2, 3\}$, $\mathcal{T}_j$ is a type III class of

R^{p_j}-valued functions on $\mathcal{W} \subset R^k$ that depend on $w = (y, x', z')'$ only through the k_a-vector z for $j = 1, 2, 3$, where $p_1 = 1$, $p_2 = p$, and $p_3 = 1$,

and $\mathcal{T}_3 \subset \{\tau_3 : \inf_{w \in \mathcal{W}} |\tau_3(w)| \geq \varepsilon\}$ for some $\varepsilon > 0$.

The set $\mathcal{W}_a^*$ in the definition of the type III class $\mathcal{T}_j$ equals Z^* in this example for $j = 1, 2, 3$. Since Z^* is bounded by condition (iii), conditions (i)-(iii) can be satisfied without trimming only if the rv's $\{Z_t : t \geq 1\}$ are bounded.

Sufficiency of conditions (i)-(iii) for stochastic equicontinuity is established as follows: Let $h_1(w) = y - x'\theta_0$ and $h_2(w) = x$. By Theorem 2, $\{\xi\}$, $\{h_1\}$, $\{h_2\}$, and $\mathcal{T}_j$ satisfy Pollard's entropy condition with envelopes 1, $|h_1|$, $|h_2|$, and C_j, respectively, for some constant $C_j \in [1, \infty)$, for $j = 1, 2, 3$. By the $\mathcal{G}^{-1}$ result of Theorem 3, so does $\{1/\tau_3 : \tau_3 \in \mathcal{T}_3\}$ with envelope C_3^2/ε^2. By the $\mathcal{G}\mathcal{H}$ and $\mathcal{G} \oplus \mathcal{G}^*$ results of Theorem 3 applied several times, $\mathcal{M}$ satisfies Pollard's entropy condition with envelope $(|h_1| \vee 1)C_4 + (|h_2| \vee 1)C_5 + (|h_1| \vee 1)(|h_2| \vee 1)C_6$ for some finite constants C_4, C_5, and C_6. Hence, Theorem 1 yields the stochastic equicontinuity of $v_T(\cdot)$, since (ii) suffices for Assumption B.

Next, we consider the conditions $P(\hat{\tau} \in \mathcal{T}) \to 1$ and $\hat{\tau} \xrightarrow{P} \tau_0$ of (3.36). Suppose (i) $\hat{\tau}_j(z)$ is a nonparametric estimator of $\tau_{j0}(z)$ that is trimmed outside Z^* to equal zero for $j = 1, 2$ and one for $j = 3$, (ii) Z^* is a finite union of convex compact subsets of R^{k_a}, (iii) $\hat{\tau}_j(z)$ and its partial derivatives of order $\leq [q] + 1$ are uniformly consistent over $z \in Z^*$ for $\tau_{j0}(z)$ and its corresponding partial derivatives, for $j = 1, 2, 3$, for some $q > k_a/2$, and (iv) the partial derivatives of order $[q] + 1$ of $\tau_{j0}(z)$ are uniformly bounded over $z \in Z^*$ and $\inf_{z \in Z^*} \lambda_{min}(\tau_{30}(z)) > 0$. Then, the realizations of $\hat{\tau}_j(z)$, viewed as functions of w, lie in a type III class of functions with probability $\to 1$ for $j = 1, 2, 3$ and $\hat{\tau} \xrightarrow{P} \tau_0$ uniformly over Z (where $\tau_{j0}(z)$ is defined for $z \in Z - Z^*$ to equal zero for $j = 1, 2$ and one for $j = 3$). Hence, the above conditions plus (i) and (ii) of (4.7) imply that conditions (i)-(iii) of (3.36) hold. If $\hat{\tau}_j(z)$ is a kernel regression estimator for

$j = 1, 2, 3$, then sufficient conditions for the above uniform consistency properties are given in Andrews (1990).

5. STOCHASTIC EQUICONTINUITY VIA BRACKETING

This section provides an alternative set of sufficient conditions for stochastic equicontinuity to those considered in Section 4. We utilize a bracketing result of Ossiander (1987) for iid rv's altered to encompass m-dependent rather than independent rv's and extended as in Pollard (1989) to allow for non-identically distributed rv's. This result depends on a condition, that we refer to as *Ossiander's entropy condition*, that is based on how well the functions in $\mathcal{M}$ can be approximated by a finite number of functions that "bracket" each of the functions in $\mathcal{M}$. The bracketing error is measured by the largest $L^2(P_t)$ distance over all distributions P_t of W_t for $t \le T$, $T \ge 1$. The main purpose of this section is to give primitive conditions under which Ossiander's entropy condition holds.

The results given here are particularly useful in three contexts. The first context is when τ is finite dimensional and $m(W_t, \tau)$ is a non-smooth function of some *nonlinear* function of τ and W_t. For example, the $m(W_t, \tau)$ function for the LAD estimator of a nonlinear regression model is of this form. In this case, it is difficult to verify Pollard's entropy condition, so Theorems 1-3 are difficult to apply. The second context is semiparametric and nonparametric applications in which the parameter τ is infinite dimensional and is a bounded smooth function with an unbounded domain. Realizations of smooth nonparametric estimators are sometimes of this form. Theorem 2 above does not apply in this case. The third context is semiparametric and nonparametric applications in which τ is infinite dimensional; is a bounded smooth function on one set out of a countable collection of sets; and is constant outside this set. For example, realizations of trimmed nonparametric estimators with variable trimming sets are sometimes of this form.

The pseudometric ρ on $\mathcal{T}$ that is used in this section is defined by

$$(5.1) \qquad \rho(\tau_1, \tau_2) = \sup_{t \leq N, N \geq 1} \left(E(m(W_t, \tau_1) - m(W_t, \tau_2))^2 \right)^{1/2} .$$

We adopt the following notational convention: For any real function f on $\mathcal{W}, \left(E |f(W_t)|^p \right)^{1/p}$

$= \sup_{w \in \mathcal{W}} |f(w)|$ if $p = \infty$.

An entropy condition analogous to Pollard's is defined using the following bracketing cover numbers:

DEFINITION: For any $\varepsilon > 0$ and $p \in [2, \infty]$, the L^p *bracketing cover number* $N_p^B(\varepsilon, P, \mathcal{F})$ is the smallest value of n for which there exist real functions $a_1, ..., a_n$ and $b_1, ..., b_n$ on $\mathcal{W}$ such that for each $f \in \mathcal{F}$ one has $|f - a_j| \leq b_j$ for some $j \leq n$ and $\max_{j \leq n} \sup_{t \leq T, T \geq 1} \left(E b_j^p(W_t) \right)^{1/p} \leq \varepsilon$, where $\{W_t : t \leq T, T \geq 1\}$ has distribution determined by P.

The log of $N_p^B(\varepsilon, P, \mathcal{F})$ is referred to as the L^p *bracketing ε-entropy* of $\mathcal{F}$. The following entropy condition was introduced by Ossiander (1987) (for the case $p = 2$):

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ satisfies *Ossiander's L^p entropy condition* for some $p \in [2, \infty]$ if

$$(5.2) \qquad \int_0^1 \left(\log N_p^B(\varepsilon, P, \mathcal{F}) \right)^{1/2} d\varepsilon < \infty .$$

As with Pollard's entropy condition, Ossiander's entropy condition restricts the complexity/size of $\mathcal{F}$ by restricting the rate of increase of the cover numbers as $\varepsilon \downarrow 0$.

Often our interest in Ossiander's L^p entropy condition is limited to the case where $p = 2$, as in Ossiander (1987) and Pollard (1989). To show that Ossiander's L^p entropy condition holds for $p = 2$ for a class of products of functions $\mathcal{G}\mathcal{H}$, however, we need to consider the case $p > 2$. The latter situation arises quite frequently in applications of interest.

ASSUMPTION D: $\mathcal{H}$ *satisfies Ossiander's L^p entropy condition with $p = 2$ and has envelope $\bar{M}$.*

THEOREM 4: *Under Assumptions B-D (with $\bar{M}$ in Assumption B given by Assumption D rather than Assumption A), $\{v_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous with ρ given by (5.1) and $\mathcal{T}$ is totally bounded under ρ.*

COMMENTS: 1. The proof of this Theorem follows easily from Theorem 2 of Pollard (1989) (as shown in the Appendix). Pollard's result is based on methods introduced by Ossiander (1987). Ossiander's result, in turn, is an extension of work by Dudley (1978).

2. As in Section 4, one establishes stochastic equicontinuity here via maximal inequalities. With the bracketing approach, however, one applies a chaining argument directly to the empirical measure rather than to a symmetrized version of it. The chaining argument relies on the Bernstein inequality for the tail probabilities of a sum of mean zero, independent rv's. The upper bound in Bernstein's inequality is small when the $L^2(P_t)$ norms of the underlying rv's are small, where P_t denotes the distribution of the t-th underlying rv. The bound ultimately is applied with the underlying rv's given by the centered difference between an arbitrary function in $\mathcal{M}$ and one of the functions from a finite set of approximating functions, each evaluated at W_t. In consequence, these functions need to be close in an $L^2(P_t)$ sense for all $t \leq T$ for the bound to be effective, where P_t denotes the distribution of W_t. This explains the appearance of the supremum $L^2(P_t)$ norm as the measure of approximation error in Ossiander's L^2 entropy condition.

We now provide primitive conditions under which Ossiander's entropy condition is satisfied. The method is analogous to that used for Pollard's entropy condition. First, we show that several useful classes of functions satisfy the condition. Then, we show how functions from these classes can be mixed and matched to obtain more general classes that satisfy the condition.

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type IV class under* P *with index* $p \in [2, \infty]$ if each function f in $\mathcal{F}$ satisfies: $f(\cdot) = f(\cdot, \tau)$ for some $\tau \in \mathcal{T}$, where $\mathcal{T}$ is some bounded subset of Euclidean space, and

$$(5.3) \qquad \sup_{t \leq T, T \geq 1} \left(E \sup_{\tau_1 : |\tau_1 - \tau| < \delta} |f(W_t, \tau_1) - f(W_t, \tau)|^p \right)^{1/p} \leq C \delta^\psi$$

$\forall \tau \in \mathcal{T}$ and $\forall \delta > 0$ in a neighborhood of 0, for some finite positive constants C and ψ, where $\{W_t : t \leq T, T \geq 1\}$ has distribution determined by P.[4]

Condition (5.3) is an L^p continuity condition that weakens the Lipschitz condition (4.3) of type II classes (provided $\sup_{t \leq T, T \geq 1}\left(EB^p(W_t)\right)^{1/p} < \infty$). The L^p continuity condition allows for discontinuous functions such as sign and indicator functions. For example, for the LAD estimator of a nonlinear regression model one takes $f(W_t, \tau) = \text{sgn}(Y_t - g(X_t, \tau))\frac{\partial}{\partial \tau_j}g(X_t, \tau)$ for different elements τ_j of τ. Under appropriate conditions on (Y_t, X_t) and on the regression function $g(\cdot, \cdot)$, the resultant class of functions can be shown to be of type IV under P with index p.

EXAMPLE 3 (cont.): In this test of variable relevance example, $\mathcal{M}$ is a type IV class with $p = 2$ under the following condition:

$$(5.4) \qquad \sup_{t \geq 1} EU_t^2 \sup_{\tau_1:|\tau_1 - \tau| < \delta} |h(Z_t, \tau_1) - h(Z_t, \tau)|^2 \leq C\delta^\psi$$

for all $\tau \in \mathcal{T}$, for all $\delta > 0$, and for some finite positive constants C and ψ. Condition (5.4) is easy to verify if $h(Z_t, \tau)$ is differentiable in τ. By a mean value expansion, (5.4) holds if $\sup_{t \geq 1} E\|U_t \sup_{\tau \in \mathcal{T}} \frac{\partial}{\partial \tau}h(Z_t, \tau)|^2 < \infty$ and $\mathcal{T}$ is bounded. On the other hand, condition (5.4) can be verified even if $h(Z_t, \tau)$ is discontinuous in τ. For example, suppose $h(Z_t, \tau) = 1(h^*(Z_t, \tau) \leq 0)$ for some real differentiable function $h^*(Z_t, \tau)$. In this case, it can be shown that condition (5.4) holds if $\sup_{t \geq 1} E|U_t|^{2+\delta} < \infty$ for some $\delta > 0$, $\sup_{t \geq 1} \sup_{\tau \in \mathcal{T}}\|\frac{\partial}{\partial \tau}h^*(Z_t, \tau)\| \leq C_1 < \infty$ for some constant C_1, and $h^*(Z_t, \tau)$ has a (Lebesgue) density that is bounded above uniformly over $\tau \in \mathcal{T}$.

EXAMPLE 4 (cont.): $\mathcal{M}$ is a type IV class with $p = 2$ in this cross-sectional constancy example under the same conditions as in Example 3 with U_t of Example 3 replaced by $U_t\frac{\partial}{\partial \delta_1}g(X_t, \delta_{10})$ and with $h(Z_t, \tau)$ taken to be of the non-differentiable form $1(h^*(Z_t, \tau) \leq 0)$ discussed above.

Note that the conditions placed on a type IV class of functions are weaker in several respects than those placed on the functions in Huber's (1967, Lemma 3, p. 227) stochastic equicontinuity result. (Huber's conditions N-2, N-3(i), and N-3(ii) are not used here, nor is his independence assumption on $\{W_t\}$.) Huber's result has been used extensively in the literature on M-estimators.

Next we consider an analogue of type III classes that allows for uniformly bounded functions that are smooth on an unbounded domain. (Recall that the functions of type III are smooth

only on a bounded domain and equal a constant elsewhere.) The class considered here can be applied to the WLS/PLR Example 5 or the GMM/CMR Example 6. Define $\mathcal{W}_a$ as in Section 4 and let $w = (w_a', w_b')'$, $h = (h_a', h_b')'$, and $W_t = (W_a', W_b')'$.

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type V class* under **P** *with index* $p \in [2, \infty]$, if

(i) each f in $\mathcal{F}$ depends on w only through a subvector w_a of dimension $k_a \leq k$,

(ii) $\mathcal{W}_a$ is such that $\mathcal{W}_a \cap \{w_a \in R^{k_a} : |w_a| \leq r\}$ is a connected compact set $\forall r > 0$,

(iii) for some real number $q > k_a/2$ and some finite constants $C_0, ..., C_{[q]}, C_q$, each $f \in \mathcal{F}$ satisfies the smoothness condition: $\forall w \in \mathcal{W}$ and $w+h \in \mathcal{W}$,

$$(5.5) \quad f(w+h) = \sum_{v=0}^{[q]} \frac{1}{v!} B_v(h_a, w_a) + R(h_a, w_a) ,$$

$$R(h_a, w_a) \leq C_q |h_a|^q , \quad \text{and} \quad |B_v(h_a, w_a)| \leq C_v |h_a|^v \quad \text{for} \quad v=0, ..., [q] ,$$

where $B_v(h_a, w_a)$ is homogeneous of degree v in h_a and $(q, C_0, ..., C_q)$ do not depend on f, w, or h,

(iv) $\sup_{t \leq T, T \geq 1} E\|W_{at}\|^\zeta < \infty$ for some $\zeta > pqk_a/(2q - k_a)$ under **P**.

In condition (iv) above, the condition $\zeta > \infty$, which arises when $p = \infty$, is taken to hold if $\zeta = \infty$.

Condition (ii) above holds, for example, if $\mathcal{W}_a = R^{k_a}$.

As with type III classes, the expansion of $f(w+h)$ in (5.5) typically is a Taylor expansion and $B_v(h_a, w_a)$ is usually the v-th differential of f at w. In this case, the third condition of (5.5) holds if the partial derivatives of f of order $\leq [q]$ are uniformly bounded.

Sufficient conditions for condition (iii) above are: (a) for some real number $q > k_a/2$, each $f \in \mathcal{F}$ has partial derivatives of order $[q]$ on $\mathcal{W}$ that are bounded uniformly over $w \in \mathcal{W}$ and $f \in \mathcal{F}$, (b) the $[q]$-th order partial derivatives of f satisfy a Lipschitz condition with exponent $q - [q]$ and some Lipschitz constant C_q that does not depend on f, and (c) $\mathcal{W}_a$ is a convex set.

The envelope of a type V class $\mathcal{F}$ can be taken to be a constant function, since the functions in $\mathcal{F}$ are uniformly bounded over $w \in \mathcal{W}$ and $f \in \mathcal{F}$.

Type V classes can be extended to allow $\mathcal{W}_a$ to be such that $\mathcal{W}_a \cap \{w_a \in R^{k_a} : \|w_a\| \le r\}$ is a finite union of connected sets $\forall r > 0$. In this case, (5.5) only needs to hold $\forall w \in \mathcal{W}$ and $w+h \in \mathcal{W}$ such that w_a and h_a are in the same connected set in $\mathcal{W}_a \cap \{w_a : \|w_a\| \le r\}$ for some $r > 0$.

In applications, the functions in type V classes usually are the realizations of nonparametric function estimates. For example, nonparametric kernel density estimates for bounded and unbounded rv's satisfy the uniform smoothness conditions of type V classes under suitable assumptions. In addition, kernel regression estimates for bounded and unbounded regressor variables satisfy the uniform smoothness conditions if they are trimmed to equal a constant outside a suitable bounded set and then smoothed (e.g., by convolution with another kernel). The bounded set in this case may depend on T.

In some cases one may wish to consider nonparametric estimates that are trimmed (i.e., set equal to a constant outside some set), but not subsequently smoothed. Realizations of such estimates do not comprise a type V class because the trimming procedure creates a discontinuity. The following class of functions is designed for this scenario. It can be used with the WLS/PLR Example 5 and the GMM/CMR Example 6. The trimming sets are restricted to come from a countably infinite number of sets $\{\mathcal{W}_{aj} : j \ge 1\}$. (This can be restrictive in practice.)

DEFINITION: A class $\mathcal{F}$ of real functions on $\mathcal{W}$ is called a *type VI class under* P *with index* $p \in [2, \infty]$, if

(i) each f in $\mathcal{F}$ depends on w only through a subvector w_a of w of dimension $k_a \le k$,

(ii) for some real number $q > k_a/2$, some sequence $\{\mathcal{W}_{aj} : j \ge 1\}$ of connected compact subsets of R^{k_a} that lie in $\mathcal{W}_a$, some sequence $\{K_j : j \ge 1\}$ of constants that satisfy $\sup_{j \ge 1} |K_j| < \infty$, and some finite constants $C_0, ..., C_{[q]}, C_q$, each $f \in \mathcal{F}$ satisfies the smoothness condition: For some integer J,

(a) $f(w) = K_J \;\; \forall w \in \mathcal{W}$ for which $w_a \notin \mathcal{W}_{aJ}$ and

(b) $\forall w \in \mathcal{W}$ and $w+h \in \mathcal{W}$ for which $w_a \in \mathcal{W}_{aJ}$ and $w_a + h_a \in \mathcal{W}_{aJ}$,

$$f(w+h) = \sum_{v=0}^{[q]} \frac{1}{v!} B_v(h_a, w_a) + R(h_a, w_a) \ ,$$

(5.6)

$$R(h_a, w_a) \le C_q |h_a|^q \ , \quad \text{and} \quad |B_v(h_a, w_a)| \le C_v |h_a|^v \quad \text{for} \quad v = 0, \ ..., \ [q] \ ,$$

where $B_v(h_a, w_a)$ is homogeneous of degree v in h_a and $(q, \{ \mathcal{W}_{aj} : j \ge 1 \}, C_0, \ ..., \ C_q)$ do not depend on f, w, or h,

(iii) $\sup_{t \le T, T \ge 1} E |W_{at}|^\zeta < \infty$ for some $\zeta > pqk_a/(2q - k_a)$ under P,

(iv) $n(r) \le K_1 \exp(K_2 r^\xi)$ for some $\xi < 2\zeta/p$ and some finite constants K_1, K_2, where $n(r)$ is the number of sets $\mathcal{W}_{aj}$ in the sequence $\{ \mathcal{W}_{aj} : j \ge 1 \}$ that do not include $\{ w_a \in \mathcal{W}_a : |w_a| \le r \}$.

Conditions (i)-(iii) in the definition of a type VI class are quite similar to conditions used above to define type III and type V classes. The difference is that with a type VI class, the set on which the functions are smooth is not a single set, but may vary from one function to the next among a countably infinite number of sets.

Condition (iv) restricts the number of $\mathcal{W}_{aj}$ sets that may be of a given radius or less. Sufficient conditions for condition (iv) are the following: Suppose $\mathcal{W}_{aj} \supset \{ w_a \in \mathcal{W}_a : |w_a| \le \eta(j) \}$ for all j sufficiently large, where $\eta(\cdot)$ is a nondecreasing real function on the positive integers that diverges to infinity as $j \to \infty$. For example, $\{ \mathcal{W}_{aj} : j \ge 1 \}$ could contain spheres, ellipses, and/or rectangles whose "radii" are large for large j. If

(5.7) $\eta(j) \ge D^*(\log j)^{1/\xi}$

for some positive finite constant D^*, then condition (iv) holds. Thus, the "radii" of the sets $\{ \mathcal{W}_{aj} : j \ge 1 \}$ are only required to increase logarithmically for condition (iv). This condition is not too restrictive, given that the number of trimming sets $\{ \mathcal{W}_{aj} \}$ is countable. More restrictive is the latter condition that the number of trimming sets $\{ \mathcal{W}_{aj} \}$ is countable.

As with type III and type V classes, the envelope of a type VI class of functions can be taken to be a constant function.

The trimmed kernel regression estimators discussed in Andrews (1990) provide examples of nonparametric function estimates for which type VI classes are applicable. For suitable trimming sets $\{ \mathcal{W}_{aj} : j \ge 1 \}$ and suitable smoothness conditions on the true regression function, one can

specify a type VI class that contains all of the realizations of such kernel estimators in a set whose probability ~ 1.

The following result establishes Ossiander's L^p entropy condition for classes of type II-VI:

THEOREM 5: *Let* $p \in [2, \infty]$. *If* $\mathcal{F}$ *is a class of functions of type II with* $\sup_{t \leq T, T \geq 1} \left(EB^p(W_t) \right)^{1/p} < \infty$, *of type III, or of type IV, V, or VI under* P *with index* p, *then Ossiander's* L^p *entropy condition* (5.2) *holds (with envelope* $F(\cdot)$ *given by* $\sup_{f \in \mathcal{F}} |f(\cdot)|$).

COMMENTS: 1. To obtain Assumption D for any of the classes of functions considered above, one only needs to consider $p = 2$ in Theorem 5. To obtain Assumption D for a class of the form $\mathcal{GH}$, where $\mathcal{G}$ and $\mathcal{H}$ are classes of types II, III, IV, V, or VI, however, one needs to apply Theorem 5 to $\mathcal{G}$ and $\mathcal{H}$ for values of p greater than 2, see Theorem 6 below.

2. Theorem 5 covers classes containing a finite number of functions, because such functions are of type IV under any distribution P and for any index $p \in [2, \infty]$. In particular, this is true for classes containing a single function. This observation is useful when establishing Ossiander's L^p entropy condition for classes of functions that can be obtained by mixing and matching functions from several classes, see below.

We now show how one can "mix and match" functions of types II-VI. Let $\mathcal{G}$, $\mathcal{G}^*$, $\mathcal{H}$, $\mathcal{G} \oplus \mathcal{G}^*$, etc., be as defined in Section 4. We say that a class of matrix-valued functions $\mathcal{G}$, $\mathcal{G}^*$, or $\mathcal{H}$ satisfies Ossiander's L^p entropy condition or is of type II, III, IV, V, or VI if it does so, or if it is, element by element for each of the rs or su elements of its functions. We adopt the convention that $\lambda\mu/(\lambda+\mu) = \mu \in (0, \infty]$ if $\lambda = \infty$ and vice versa.

THEOREM 6: (a) *If* $\mathcal{G}$ *and* $\mathcal{G}^*$ *satisfy Ossiander's* L^p *entropy condition for some* $p \in [2, \infty]$, *with envelopes* G *and* G^*, *respectively, then so do each of the following classes (with envelopes given in parentheses):* $\mathcal{G} \cup \mathcal{G}^*$ $(G \vee G^*)$, $\mathcal{G} \oplus \mathcal{G}^*$ $(G + G^*)$, $\mathcal{G} \vee \mathcal{G}^*$ $(G \vee G^*)$, $\mathcal{G} \wedge \mathcal{G}^*$ $(G \vee G^*)$, *and* $|\mathcal{G}|$ (G). *If in addition* $r = s$ *and* $\inf_{g \in \mathcal{G}} \inf_{w \in \mathcal{W}} \lambda_{\min}(g(w)) = \lambda_*$ *for some* $\lambda_* > 0$, *then* $\mathcal{G}^{-1}$ *also satisfies Ossiander's* L^p *entropy condition (with envelope* r/λ_*).

(b) *The class* $\mathcal{GH}$ *satisfies Ossiander's* L^p *entropy condition with* p *equal to* $\alpha \in [2, \infty]$ *and envelope*

sGH, if (i) $\mathcal{G}$ *and* $\mathcal{H}$ *satisfy Ossiander's L^p entropy condition with p equal to* $\lambda \in (\alpha, \infty]$ *and p equal to* $\mu \in (\alpha, \infty]$, *respectively*, (ii) $\lambda\mu/(\lambda+\mu) \geq \alpha$, *and* (iii) *the envelopes G and H of* $\mathcal{G}$ *and* $\mathcal{H}$ *satisfy*

$$\sup_{t \leq T, T \geq 1}\left(EG^{\lambda}(W_t)\right)^{1/\lambda} < \infty \text{ and } \sup_{t \leq T, T \geq 1}\left(EH^{\mu}(W_t)\right)^{1/\mu} < \infty.$$

EXAMPLE 6 (cont.): Theorems 4-6 can be used to verify stochastic equicontinuity of $v_T(\cdot)$ and total boundedness of $\mathcal{T}$ in the GMM/CMR example. With some abuse of notation, let $\Delta(w)$ and $\Omega(w)$ denote functions on $\mathcal{W}$ whose values depend on w only through the k_a-vector x and equal $\Delta(x)$ and $\Omega(x)$ respectively. Similarly, let $\psi(w, \theta_0)$ denote the function on $\mathcal{W}$ that depends on w only through z and equals $\psi(z, \theta_0)$. The following conditions are sufficient:

(5.8)

    (i) $\{(Z_t, X_t) : t \geq 1\}$ is an m-dependent sequence of rv's.

    (ii) $\sup_{t \geq 1} E|\psi(Z_t, \theta_0)|^6 < \infty$.

    (iii) $\mathcal{T} = \{\tau : \tau = \Delta'\Omega^{-1}$ for some $\Delta' \in \mathcal{D}$ and $\Omega \in \mathcal{A}\}$, where $\mathcal{D}$ and $\mathcal{A}$ are type V or type VI classes of functions on $\mathcal{W} \subset R^k$ with index $p = 6$ whose functions depend on w only through the k_a-vector x, and $\mathcal{A} \subset \{\Omega : \inf_{w \in \mathcal{W}} \lambda_{min}(\Omega(w)) \geq \varepsilon\}$ for some $\varepsilon > 0$.

Note that condition (iii) of (5.8) includes a moment condition on X_t : $\sup_{t \geq 1} E|X_t|^\zeta < \infty$ for some $\zeta > 6qk_a/(2q - k_a)$.

Sufficiency of conditions (i)-(iii) for stochastic equicontinuity and total boundedness is established as follows: By Theorem 5, $\{\psi(\cdot, \theta_0)\}, \mathcal{D}$, and $\mathcal{A}$ satisfy Ossiander's L^p entropy condition with $p = 6$ and with envelopes $|\psi(\cdot, \theta_0)|$, C_1, and C_2, respectively, for some finite constants C_1, C_2. By the $\mathcal{G}^{-1}$ result of Theorem 6, so does $\mathcal{A}^{-1}$ with some constant envelope $C_3 < \infty$. By the $\mathcal{GH}$ result of Theorem 6 applied with $\alpha = 3$ and $\lambda = \mu = 6$, $\mathcal{DA}^{-1}$ satisfies Ossiander's L^p entropy condition with $p = 3$ and some constant envelope $C_4 < \infty$. By this result, condition (ii), and the $\mathcal{GH}$ result of Theorem 6 applied with $\alpha = 2$, $\lambda = 3$, $\mu = 6$, $\mathcal{G} = \mathcal{DA}^{-1}$, and $\mathcal{H} = \{\psi(\cdot, \theta_0)\}$, $\mathcal{M}$ satisfies Ossiander's L^p entropy condition with $p = 2$ and envelope $C_5|\psi(\cdot, \theta_0)|$ for some constant $C_5 < \infty$. Theorem 4 now yields stochastic equicontinuity, since condition (ii) is sufficient for Assumption B.

Condition (iii) above covers the case where the domain of the nonparametric functions is unbounded and the nonparametric estimators $\hat{\Delta}$ and $\hat{\Omega}$ are not trimmed to equal zero outside a single fixed bounded set, as is required when the symmetrization results of Section 4 are applied. As discussed above, nonparametric kernel regression estimators that are trimmed and smoothed or trimmed on variable sets provide examples where condition (iii) holds under suitable assumptions for realizations of the estimators that lie in a set whose probability $\rightarrow 1$. For example, Andrews (1990) provides uniform consistency on expanding sets and L^Q consistency results for such estimators, as are required to establish that $P(\hat{\tau} \in \mathcal{T}) \rightarrow 1$ and $\hat{\tau} \xrightarrow{P} \tau_0$ (the first and second parts of (3.36)) when stochastic equicontinuity is established using conditions (i)-(iii) above.

6. CONCLUSION

This paper illustrates how empirical process methods can be utilized to find the asymptotic distributions of econometric estimators and test statistics. The concepts of empirical processes, weak convergence, and stochastic equicontinuity are introduced. Primitive sufficient conditions for the key stochastic equicontinuity property are outlined. Applications of empirical process methods in the econometrics literature are reviewed briefly. More detailed discussion is given for three classes of applications: M-estimators based on non-differentiable criterion functions, tests of hypotheses for which a nuisance parameter is present only under the alternative hypothesis, and semiparametric estimators that utilize preliminary nonparametric estimators.

APPENDIX

PROOF OF THEOREM 1: Write $v_T(\cdot)$ as the sum of m empirical processes $\{v_{Tj}(\cdot) : T \geq 1\}$ for $j = 1, ..., m$, where $v_{Tj}(\cdot)$ is based on the independent summands $\{m(W_t, \cdot) : t = j + sm, s = 1, 2, ...\}$. By standard inequalities it suffices to prove the stochastic equicontinuity of $\{v_{Tj}(\cdot) : T \geq 1\}$ for each j.

The latter can be proved using Pollard's (1990) proof of stochastic equicontinuity for his functional CLT (Theorem 10.7). We take his functions $f_{ni}(\omega, t)$ to be of the form $m(W_t, \tau)/\sqrt{T}$. We alter his pseudometric from $\lim_{N \to \infty} \left(\frac{1}{N}\Sigma_1^N E|m(W_t, \tau_1) - m(W_t, \tau_2)|^2\right)^{1/2}$ to that given in (3.1). Pollard's proof of stochastic equicontinuity relies on conditions (i) and (iii)-(v) of his Theorem 10.7. Condition (ii) of Theorem 10.7 is used only for obtaining convergence of the finite dimensional distributions, which we do not need, and for ensuring that his pseudometric is well-defined. Our pseudometric does not rely on this condition. Inspection of Pollard's proof shows that any pseudometric can be used for his stochastic equicontinuity result (although not for his total boundedness result) provided his condition (v) holds. Thus, it suffices to verify his conditions (i) and (iii)-(v).

Condition (i) requires that the functions $\{m(W_t, \tau)/\sqrt{T} : t \leq T, T \geq 1\}$ are "manageable." This holds under Assumption A because Pollard's packing numbers satisfy

$$(A.1) \qquad \sup_{\omega \in \Omega, n \geq 1, \alpha \in R_+^n} D(\varepsilon |\alpha \odot \mathcal{F}_n(\omega)|, \alpha \odot \mathcal{F}_{n\omega}) \leq \sup_{Q \in \mathcal{Q}} N_2(\varepsilon/2, Q, \mathcal{M}) .$$

Conditions (iii) and (iv) are implied by Assumption B. Condition (v) holds automatically given our choice of pseudometric. $\square$

PROOF OF THEOREM 2: Type I classes of form (a) satisfy Pollard's entropy condition by Lemmas II.28 and II.36(ii) of Pollard (1984, pp. 30 and 34). Type I classes of form (b) satisfy Pollard's entropy condition because (i) they are contained in VC hull classes by the proof of Proposition 4.4 of Dudley (1987) and the fact that $\{f : f(w) = w'\xi \ \forall w \in \mathcal{W}, \xi \in R^k\}$ is a VC major class, see Pollard (1984, Lemma II.18, p. 20), (ii) VC hull classes are contained in VC

subgraph hull classes, and (iii) VC subgraph hull classes satisfy Pollard's entropy condition by Corollary 5.8 of Dudley (1987).

For classes of type II, consider the functions $f(\cdot, \tau_1), ..., f(\cdot, \tau_n)$, where $\tau_1, ..., \tau_n$ are points at the centers of disjoint cubes of diameter $\varepsilon(QF^2)^{1/2}/(QB^2)^{1/2}$ whose union covers $\mathcal{T}(\subset R^s$ for some $s \geq 1)$. Since

$$(A.2) \quad \min_{j \leq n}\left(Q(f(\cdot, \tau) - f(\cdot,\tau_j))^2\right)^{1/2} \leq \min_{j \leq n}\left(QB^2\right)^{1/2}|\tau - \tau_j| \leq \varepsilon\left(QF^2\right)^{1/2} ,$$

$N_2\left(\varepsilon(QF^2)^{1/2}, Q, \mathcal{F}\right)$ is $\leq$ the number of cubes above. By choice of the envelope $F(\cdot) = 1 \vee$ $\sup_{f \in \mathcal{F}}|f(\cdot)| \vee B(\cdot)$, $\varepsilon(QF^2)^{1/2}/(QB^2)^{1/2} \geq \varepsilon$, so the number of cubes is $\leq C\varepsilon^{-s}$ for some $C > 0$ and all $Q \in \mathcal{Q}$. Thus, Pollard's entropy condition holds with envelope $F(\cdot)$.

For classes of type III, Pollard's entropy condition holds because

$$(A.3) \quad \sup_{Q \in \mathcal{Q}} N_2\left(\varepsilon(QF^2)^{1/2}, Q, \mathcal{F}\right) \leq C \exp\left(\varepsilon^{-k_a/q}\right) \quad \forall \varepsilon \in (0, 1]$$

for some $C < \infty$ by Kolmogorov and Tihomirov (1961, Thm. XIII, p. 308). Since $q > k_a/2$ by assumption, Pollard's entropy condition holds. $\square$

PROOF OF THEOREM 3: For $\mathcal{G} \cup \mathcal{G}^*$, we have

$$N_2(\varepsilon, Q, \mathcal{G} \cup \mathcal{G}^*) \leq N_2(\varepsilon, Q, \mathcal{G}) + N_2(\varepsilon, Q, \mathcal{G}^*) , \quad \text{and so,}$$

$(A.4)$

$$N_2\left(\varepsilon(Q(G \vee G^*)^2)^{1/2}, Q, \mathcal{G} \cup \mathcal{G}^*\right) \leq N_2\left(\varepsilon(QG^2)^{1/2}, Q, \mathcal{G}\right) + N_2\left(\varepsilon(QG^{*2})^{1/2}, Q, \mathcal{G}^*\right) ,$$

where the second inequality uses the facts that $N_2(\varepsilon, Q, \mathcal{F})$ is nonincreasing in ε, $Q(G \vee G^*)^2 \geq QG^2$, and $Q(G \vee G^*)^2 \geq QG^{*2}$. Pollard's entropy condition follows from the second inequality of (A.4).

For $\mathcal{G} \oplus \mathcal{G}^*$, it suffices to suppose that $r = s = 1$. As above, Pollard's entropy condition follows from the inequalities

$$
(A.5) \quad N_2(\varepsilon, Q, \mathcal{G} \oplus \mathcal{G}^*) \leq N_2(\varepsilon/2, Q, \mathcal{G})N_2(\varepsilon/2, Q, \mathcal{G}^*) ,
$$

$$
Q(G + G^*)^2 \geq QG^2 , \quad \text{and} \quad Q(G + G^*)^2 \geq QG^{*2} ,
$$

where the first inequality holds because $\min_{j,k \leq n}\left(\int(g + g^* - g_j - g_k^*)^2 dQ\right)^{1/2} \leq \min_{j \leq n}\left(\int(g - g_j)^2 dQ\right)^{1/2}$
$+ \min_{k \leq n}\left(\int(g^* - g_k^*)^2 dQ\right)^{1/2}$.

For $\mathcal{G}\mathcal{H}$, each element of gh is a finite union of products of scalar functions, and so, using the result for $\mathcal{G} \oplus \mathcal{G}^*$, it suffices to suppose that $r = s = u = 1$. For notational simplicity, assume $G = G \vee 1$ and $H = H \vee 1$. Let $Q_G(\cdot) = Q(\cdot G^2)/QG^2$ and $Q_H(\cdot) = Q(\cdot H^2)/QH^2$. Note that $Q_G, Q_H \in Q$. Let $n = N_2\left(\varepsilon\left(Q_H G^2\right)^{1/2}, Q_H, \mathcal{G}\right)$ and $n^* = N_2\left(\varepsilon\left(Q_G H^2\right)^{1/2}, Q_G, \mathcal{H}\right)$. Let $g_1, ..., g_n$ and $h_1, ..., h_{n^*}$ denote approximating functions in $\mathcal{G}$ and $\mathcal{H}$, respectively, that correspond to the cover numbers n and n^*. We use $g_j h_k$ to approximate gh for $g \in \mathcal{G}$ and $h \in \mathcal{H}$:

$$
\min_{j \leq n, k \leq n^*}\left(\int(gh - g_j h_k)^2 dQ\right)^{1/2}
$$

$$
(A.6) \quad \leq \min_{j \leq n}\left(QH^2\int(g - g_j)^2 d\left[\frac{Q(\cdot H^2)}{QH^2}\right]\right)^{1/2} + \min_{k \leq n^*}\left(QG^2\int(h - h_k)^2 d\left[\frac{Q(\cdot G^2)}{QG^2}\right]\right)^{1/2}
$$

$$
\leq \left(QG^2 H^2\right)^{1/2}\varepsilon .
$$

Thus, we get

$$
N_2\left(\varepsilon\left(QG^2 H^2\right)^{1/2}, Q, \mathcal{G}\mathcal{H}\right) \leq N_2\left(\frac{\varepsilon}{2}\left(Q_H G^2\right)^{1/2}, Q_H, \mathcal{G}\right)N_2\left(\frac{\varepsilon}{2}\left(Q_G H^2\right)^{1/2}, Q_G, \mathcal{H}\right) \quad \text{and}
$$

$$
(A.7) \quad \sup_{Q \in Q} N_2\left(\varepsilon\left(QG^2 H^2\right)^{1/2}, Q, \mathcal{G}\mathcal{H}\right)
$$

$$
\leq \sup_{Q_H \in Q} N_2\left(\frac{\varepsilon}{2}\left(Q_H G^2\right)^{1/2}, Q_H, \mathcal{G}\right)\sup_{Q_G \in Q} N_2\left(\frac{\varepsilon}{2}\left(Q_G H^2\right)^{1/2}, Q_G, \mathcal{H}\right)
$$

$$
= \sup_{Q \in Q} N_2\left(\frac{\varepsilon}{2}\left(QG^2\right)^{1/2}, Q, \mathcal{G}\right)\sup_{Q \in Q} N_2\left(\frac{\varepsilon}{2}\left(QH^2\right)^{1/2}, Q, \mathcal{H}\right) .
$$

Pollard's entropy condition follows from the latter inequality.

For $\mathcal{G} \vee \mathcal{G}^*$, it suffices to suppose $r = s = 1$. Pollard's entropy condition follows from the inequalities

$$N_2(\varepsilon, Q, \mathcal{G} \vee \mathcal{G}^*) \leq N_2(\varepsilon/2, Q, \mathcal{G}) N_2(\varepsilon/2, Q, \mathcal{G}^*) ,$$

(A.8)

$$Q(G \vee G^*)^2 \geq QG^2 , \quad \text{and} \quad Q(G \vee G^*)^2 \geq QG^{*2} ,$$

where the first inequality uses $|g \vee g^* - g_j \vee g_k^*| \leq |g - g_j| + |g^* - g_k^*|$. The proof for $\mathcal{G} \wedge \mathcal{G}^*$ is analogous (with the envelope still given by $G \vee G^*$ rather than $G \wedge G^*$). The result for $|\mathcal{G}|$ follows because $||g| - |a_j|| \leq |g - a_j|$.

Lastly, consider $\mathcal{G}^{-1}$. For $g \in \mathcal{G}$, let g^ℓ denote the ℓ-th element of g, where $\ell = 1, ..., L$ and $L = r^2$. Let $\mathcal{G}_\ell = \{g^\ell : g \in \mathcal{G}\}$ and $n_\ell = N_2(\varepsilon/2, Q, \mathcal{G}_\ell)$ for some $Q \in \mathcal{Q}$. We claim that given any $\varepsilon > 0$ and $Q \in \mathcal{Q}$, there exist functions $g_1, ..., g_n$ in $\mathcal{G}$ with $n \leq \Pi_{\ell=1}^L n_\ell$ such that for all $g \in \mathcal{G}$

(A.9) $\quad \min_{j \leq n} \max_{\ell \leq L} \left(Q\left(g^\ell - g_j^\ell\right)^2 \right)^{1/2} \leq \varepsilon .$

To see this, note that by the assumption that $\mathcal{G}$ satisfies Pollard's entropy condition, for each ℓ there exist real functions $g_{\ell 1}, ..., g_{\ell n_\ell}$ in $\mathcal{G}_\ell$ such that for all $g \in \mathcal{G}$ $\min_{j \leq n_\ell} \left(Q\left(g^\ell - g_{\ell j}\right)^2 \right)^{1/2} < \varepsilon/2$. Form the set $\mathcal{G}^+$ of all R^L-valued functions whose ℓ-th element is $g_{\ell j}$ for some $j = 1, ..., n_\ell$ for $\ell = 1, ..., L$. The number of such functions is $n^+ = \Pi_{\ell=1}^L n_\ell$. The functions in $\mathcal{G}^+$ are not necessarily in $\mathcal{G}$. For each function g^+ in $\mathcal{G}^+$ consider the $L^2(Q)$ $\varepsilon/2$-ball in $\mathcal{G}$ centered at g^+. Take one function from each non-empty ball and let $g_1, ..., g_n$ denote the chosen functions. These functions satisfy the claim above.

If $\mathcal{G}$ satisfies Pollard's entropy condition with envelope G, it also does so with envelope $G \vee 1$. For notational simplicity, suppose $G = G \vee 1$. Given $Q \in \mathcal{Q}$, let $\tilde{Q}(\cdot) = Q(\cdot \tilde{G}^4)/Q\tilde{G}^4$ $(\in \mathcal{Q})$, where $\tilde{G}$ is the envelope of $\mathcal{G}^{-1}$. Take ε and Q in the claim above to equal $\varepsilon \left(\tilde{Q} G^4\right)^{1/2}/r^4$ and $\tilde{Q}$ respectively. Then, there exist functions $g_1, ..., g_n$ in $\mathcal{G}$ such that $\min_{j \leq n} \max_{\ell \leq L} \left(\tilde{Q}\left(g^\ell - g_j^\ell\right)^2 \right)^{1/2} \leq \varepsilon \left(\tilde{Q} G^4\right)^{1/2}/r^4$ and $n \leq \Pi_{\ell=1}^L N_2\left(\frac{\varepsilon}{2}\left(\tilde{Q} G^4\right)^{1/2}/r^4, \tilde{Q}, \mathcal{G}_\ell\right)$. Let $1_r = (1, ..., 1)'$ $(\in R^r)$ and let $|\cdot|$ denote the matrix of absolute values of the matrix $\cdot$. For arbitrary unit vectors $b, c \in R^r$, we have

$$\min_{j \leq n} Q\left(b'g^{-1}c - b'g_j^{-1}c\right)^2 = \min_{j \leq n} Q\left(b'g^{-1}(g_j - g)g_j^{-1}c\right)^2$$

(A.10)
$$\leq \min_{j \leq n} r^4 Q\left(\tilde{G}^2 1_r' |g_j - g| 1_r\right)^2 = \min_{j \leq n} r^4 Q\tilde{G}^4 \sum_{\ell=1}^{L} \sum_{m=1}^{L} \tilde{Q} |g^\ell - g_j^\ell| |g^m - g_j^m|$$

$$\leq r^8 Q\tilde{G}^4 \min_{j \leq n} \max_{\ell \leq L} \tilde{Q}\left(g^\ell - g_j^\ell\right)^2 \leq r^8 Q\tilde{G}^4 \varepsilon^2 \tilde{Q} G^4/r^8 = \varepsilon^2 Q G^4 \tilde{G}^4 .$$

Thus, $N_2\left(\varepsilon\left(QG^4\tilde{G}^4\right)^{1/2}, Q, \mathcal{G}^{-1}\right) \leq n \leq \Pi_{\ell=1}^{L} N_2\left(\frac{\varepsilon}{2}\left(\tilde{Q} G^4\right)^{1/2}/r^4, \tilde{Q}, \mathcal{G}_\ell\right)$ and

(A.11)
$$\sup_{Q \in \mathcal{Q}} N_2\left(\varepsilon\left(QG^4\tilde{G}^4\right)^{1/2}, Q, \mathcal{G}^{-1}\right) \leq \sup_{\tilde{Q} \in \mathcal{Q}} \prod_{\ell=1}^{L} N_2\left(\frac{\varepsilon}{2}\left(\tilde{Q} G^4\right)^{1/2}/r^4, \tilde{Q}, \mathcal{G}_\ell\right)$$

$$= \sup_{Q \in \mathcal{Q}} \prod_{\ell=1}^{L} N_2\left(\frac{\varepsilon}{2}\left(QG^4\right)^{1/2}/r^4, Q, \mathcal{G}_\ell\right) \leq \sup_{Q \in \mathcal{Q}} \prod_{\ell=1}^{L} N_2\left(\frac{\varepsilon}{2}\left(QG^2\right)^{1/2}/r^4, Q, \mathcal{G}_\ell\right) .$$

The integral over $\varepsilon \in [0, 1]$ of the square root of the logarithm of the right-hand side (rhs) of (A.11) is finite since $\mathcal{G}$ satisfies Pollard's entropy condition with envelope $G = G \vee 1$. Thus, $\mathcal{G}^{-1}$ satisfies Pollard's entropy condition with envelope $(G \vee 1)^2 \tilde{G}^2$. $\square$

PROOF OF THEOREM 4: Total boundedness of $\mathcal{T}$ under ρ follows straightforwardly from $N_2^B(\varepsilon, P, \mathcal{M}) < \infty \ \forall \varepsilon > 0$. For stochastic equicontinuity of $\{v_T(\cdot) : T \geq 1\}$, by the same argument as in the proof of Theorem 1, it suffices to prove the result when $\{W_t : t \leq T\}$ are independent rv's. By Markov's inequality and Theorem 2 of Pollard (1989), we have

$$\varlimsup_{T \to \infty} P^*\left(\sup_{\rho(\tau_1, \tau_2) < \delta} |v_T(\tau_1) - q_T(\tau_2)| > \eta\right)$$

(A.12)
$$\leq \varlimsup_{T \to \infty} E^* \sup_{\rho(\tau_1, \tau_2) < \delta} |v_T(\tau_1) - v_T(\tau_2)|/\eta$$

$$\leq \varlimsup_{T \to \infty} \frac{1}{\sqrt{T}} \Sigma_1^T E\bar{M}(W_t) 1(\bar{M}(W_t) > \sqrt{T}\xi_\delta)/\eta + C\int_0^\delta \left(\log N_2^B(\varepsilon, P, \mathcal{F})\right)^{1/2} d\varepsilon/\eta$$

for some constant $C < \infty$, where $\xi_\delta > 0$ is a constant that does not depend on T. The second term on the right-hand side of (A.12) can be made arbitrarily small by choice of δ using Assumption D. The first term is less than or equal to

(A.13) $\quad 4 \varlimsup_{T \to \infty} T^{-\delta/2} \frac{1}{T} \Sigma_1^T E \bar{M}^{2+\delta}(W_t) 1(\bar{M}(W_t) > \sqrt{T}\xi_\delta)/\xi_\delta^{1+\delta} = 0$

using Assumption B. Stochastic equicontinuity follows. □

PROOF OF THEOREM 5: It suffices to prove the result for classes of type III-VI, because a type II class with $\sup_{t \leq t, T \geq 1}(EB^p(W_t))^{1/p} < \infty$ is a type IV class under P with index p.

First, we consider classes of type III. For given $\varepsilon > 0$, define the functions a_j, b_j, $j = 1, ..., n_\varepsilon$ of the definition of L^p bracketing cover numbers as follows: (a) $\forall w \in \mathcal{W}$ such that $w_a \in \mathcal{W}_a - \mathcal{W}_a^*$, let $a_j(w) = K$ and $b_j(w) = 0$ $\forall j$ and (b) $\forall w \in \mathcal{W}$ such that $w_a \in \mathcal{W}_a^*$, let $\{a_j(w) : j = 1, ..., n_\varepsilon\}$ be the functions constructed by Kolmogorov and Tihomirov (1961, pp. 312-314) in their proof of Theorem XIV and let $b_j(w) = \varepsilon$ $\forall j$. These functions satisfy the conditions for L^p bracketing cover numbers for all $p \in [2, \infty]$. Hence, $N_p^B(\varepsilon, P, \mathcal{F}) \leq n_\varepsilon$ $\forall \varepsilon \in (0, 1]$, $\forall p \in [2, \infty]$. The number n_ε of such functions is $\leq C \exp(\varepsilon^{-k_a/q})$ $\forall \varepsilon \in (0, 1]$ for some $C < \infty$ by Kolmogorov and Tihomirov (1961, Thm. XIV). Since $q > k_a/2$ by assumption, Ossiander's entropy condition holds for all $p \in [0, \infty]$.

For a type IV class with index p, consider disjoint cubes in $\mathcal{T}$ of diameter $\delta = (\varepsilon/C)^{1/\psi}$. The number $N(\varepsilon)$ of such cubes satisfies $N(\varepsilon) \leq C^* \varepsilon^{-d/\psi}$ for some $C^* < \infty$, where d is the dimension of $\mathcal{T}$. Let τ_j be some element of the j-th cube in $\mathcal{T}$. Let $a_j(\cdot) = f(\cdot, \tau_j)$ and $b_j(\cdot) = \sup_{|\tau - \tau_j| < \delta} |f(\cdot, \tau) - a_j(\cdot)|$. By (4.3), $\sup_{t \leq T, T \geq 1}\left[Eb_j^p(W_t)\right]^{1/p} \leq C\delta^\psi = \varepsilon$. Thus, $N_p^B(\varepsilon, P, \mathcal{F}) \leq N(\varepsilon)$. Since $\int_0^1 (\log N(\varepsilon))^{1/2} d\varepsilon < \infty$, Ossiander's L^p entropy condition holds.

For a type V class with index p, let $\mathcal{W}_r = \mathcal{W} \cap \{w \in R^k : \|w_a\| \leq r\}$, let $\mathcal{F}_r$ denote the class of functions $\mathcal{F}$ restricted to $\mathcal{W}_r$ and let $N_\infty(\varepsilon, \mathcal{W}_r, \mathcal{F}_r)$ be the minimal number n of real functions $f_1, ..., f_n$ on $\mathcal{W}_r$ such that $\min_{j \leq n} \sup_{w \in \mathcal{W}_r} |f(w) - f_j(w)| < \varepsilon$ for each $f \in \mathcal{F}_r$. We claim that

(A.14) $\quad N_p^B(\varepsilon, P, \mathcal{F}) \leq N_\infty(\varepsilon/2, \mathcal{W}_{r(\varepsilon)}, \mathcal{F}_{r(\varepsilon)})$,

where $r(\varepsilon) = C\varepsilon^{-p/\zeta}$ for some constant $C < \infty$ when $p < \infty$ and $r(\varepsilon) = \sup\{\|w_a\| : w \in \mathcal{W}\}$ $(< \infty)$ when $p = \infty$.

Using the proof of Theorem XIV of Kolmogorov and Tihomirov (1961, pp. 312-314), it can be seen that

$$(A.15) \quad \log N_\infty(\varepsilon, \mathcal{W}_{r(\varepsilon)}, \mathcal{F}_{r(\varepsilon)}) \le Dr(\varepsilon)^{k_a}\varepsilon^{-k_a/q} \le D^*\varepsilon^{-k_a\left(\frac{p}{\zeta} + \frac{1}{q}\right)}$$

for some constants $D, D^* < \infty$, where the second inequality holds only when $p < \infty$. When $p < \infty$, (A.14) and (A.15) combine to yield Ossiander's L^p entropy condition for $\mathcal{F}$ if $k_a(p/\zeta + 1/q)/2 < 1$, or equivalently, if $\zeta > pqk_a/(2q - k_a)$ and $q > k_a/2$, as is assumed. When $p = \infty$, (A.14) and the first inequality of (A.15) combine to yield Ossiander's L^p entropy condition for $\mathcal{F}$ provided $q > k_a/2$, as is assumed.

It remains to show (A.14). For $p = \infty$, (A.14) follows immediately from the definition of $N_\infty^B(\cdot)$ and $N_\infty(\cdot)$, since $\mathcal{W}_{r(\varepsilon)} = \mathcal{W}$ and $\mathcal{F}_{r(\varepsilon)} = \mathcal{F}$ when $p = \infty$. Next, suppose $p < \infty$. For $n = N_\infty(\varepsilon/2, \mathcal{W}_r, \mathcal{F}_r)$, define real functions $a_j, b_j, j = 1, ..., n$ on $\mathcal{W}$ as follows: On $\mathcal{W}_r$, take $\{a_j(\cdot) : j = 1, ..., n\}$ to be the functions constructed by Kolmogorov and Tihomirov (1961, pp. 312-214) in their proof of Theorem XIV and let $b_j(\cdot) = \varepsilon/2$ for $j = 1, ..., n$. On $\mathcal{W} - \mathcal{W}_r$, take $a_j(\cdot) = 0$ and take $b_j(\cdot) = F$ for $j = 1, ..., n$, where F is a constant for which $\sup_{w \in \mathcal{W}}|f(w)| \le F$ $\forall f \in \mathcal{F}$. Then, for each $f \in \mathcal{F}$, $\min_{j \le n}|f - a_j| \le b_j$ and

$$(A.16) \quad \begin{aligned} \sup_{t \le T, T \ge 1} Eb_j^p(W_t) &= \sup_{t \le T, T \ge 1} Eb_j^p(W_t)1(W_t \in \mathcal{W}_r) + \sup_{t \le T, T \ge 1} Eb_j^p(W_t)1(W_t \in \mathcal{W} - \mathcal{W}_r) \\ &\le (\varepsilon/2)^p + F^p r^{-\zeta} \sup_{t \le T, T \ge 1} E\|W_{at}\|^\zeta = (\varepsilon/2)^p + C^* r^{-\zeta}, \end{aligned}$$

where C^* is defined implicitly. If we let $r = r(\varepsilon) = \left(2^p C^*/(2^p - 1)\right)^{1/\zeta}\varepsilon^{-p/\zeta}$, then $\sup_{t \le T, T \ge 1} Eb_j^p(W_t) \le \varepsilon^p$ and (A.14) holds.

Last, we consider type VI classes of functions. First, suppose $p < \infty$. We derive an upper bound on $N_p^B(\varepsilon, P, \mathcal{F})$ for arbitrary $\varepsilon > 0$. Let $r_\varepsilon = C\varepsilon^{-p/\zeta}$ for some $C < \infty$ and let F be a constant for which $\sup_{w \in \mathcal{W}}|f(w)| \le F$ $\forall f \in \mathcal{F}$. Let J be the index of a set $\mathcal{W}_{aJ}$ that does not include $\{w_a \in \mathcal{W}_a : \|w_a\| \le r_\varepsilon\}$. For functions $f \in \mathcal{F}$ whose corresponding integer of part (ii) (of the definition of type VI classes) is J, take the centering and ε-bracketing functions $\{(a_\ell, b_\ell) : \ell = 1, ..., n_{\varepsilon J}\}$ (of the definition of L^p bracketing cover numbers) as follows: (a) $\forall w \in \mathcal{W}$ such that $\|w_a\| > r_\varepsilon$, let $a_\ell(w) = 0$ and $b_\ell(w) = F$, (b) $\forall w \in \mathcal{W}$ such that $\|w_a\| \le r_\varepsilon$ and $w_a \notin \mathcal{W}_{aJ}$, let $a_\ell(w) = K_J$ and $b_\ell(w) = 0$, and (c) $\forall w \in \mathcal{W}$ such that $\|w_a\| \le r_\varepsilon$ and $w_a \in \mathcal{W}_{aJ}$, let $\{a_\ell(w) :$

$\ell = 1, ..., n_{\varepsilon J}\}$ be the functions constructed by Kolmogorov and Tihomirov (1961) in the proof of their Theorem XIV and let $b_\ell(w) = \varepsilon/2$ $\forall \ell$. The number $n_{\varepsilon J}$ of such functions is $\leq D_1 \exp\left[D_2 r_\varepsilon^{k_a} \varepsilon^{-k_a/q}\right]$ by Theorem XIV of Kolmogorov and Tihomirov (1961), since $\{w : \|w_a\| \leq r_\varepsilon, w_a \in \mathcal{W}_{aJ}\} \subset \{w : \|w_a\| \leq r_\varepsilon\}$.

Next, for all functions $f \in \mathcal{F}$ whose corresponding integer J of part (ii) is such that $\mathcal{W}_{aJ}$ contains $\{w_a \in \mathcal{W}_a : \|w_a\| \leq r\}$, take the centering and ε-bracketing functions $\{(a_\ell, b_\ell) : \ell = 1, ..., n_\varepsilon\}$ as follows: (a) $\forall w \in \mathcal{W}$ such that $\|w_a\| > r$, let $a_\ell(w) = 0$ and $b_\ell(w) = F$ $\forall \ell$ and (b) $\forall w \in \mathcal{W}$ such that $\|w_a\| \leq r$, let $\{a_\ell(w) : \ell = 1, ..., n_\varepsilon\}$ be the functions constructed by Kolmogorov and Tihomirov (1961) in the proof of their Theorem XIV and let $b_\ell(w) = \varepsilon/2$ $\forall \ell$. The number of such functions also is $\leq D_1 \exp\left[D_2 r_\varepsilon^{k_a} \varepsilon^{-k_a/q}\right]$.

Now, the number of indices J for which $\mathcal{W}_{aJ}$ does not include $\{w_a \in \mathcal{W}_a : \|w_a\| \leq r\}$ is $n(r_\varepsilon)$. Hence, the total number of centering/ε-bracketing functions considered above is $\leq (n(r_\varepsilon) + 1)$ $D_1 \exp\left[D_2 r_\varepsilon^{k_a} \varepsilon^{-k_a/q}\right]$. Also note that $\sup_{t \leq T, T \geq 1}\left(Eb_\ell^p(W_t)\right)^{1/p} < \varepsilon$ for all of the functions b_ℓ introduced above by the same calculations as in (A.16) provided C (of the definition of r_ε) is defined suitably. Hence,

$$(A.17) \quad \begin{aligned} N_p^B(\varepsilon, \mathrm{P}, \mathcal{F}) &\leq (n(r_\varepsilon) + 1)D_1 \exp\left[D_2 r_\varepsilon^{k_a} \varepsilon^{-k_a/q}\right] \\ &\leq \left(K_1 \exp[K_2 C^\xi \varepsilon^{-p\xi/\zeta}] + 1\right)D_1 \exp\left[D_2 C^{k_a} \varepsilon^{-k_a(p/\zeta + 1/q)}\right] . \end{aligned}$$

With this bound, Ossiander's L^p entropy condition holds provided $p\xi/(2\zeta) < 1$ and $k_a(p/\zeta + 1/q)/2 < 1$, or equivalently, $\xi < 2\zeta/p, q > k_a/2$, and $\zeta > pqk_a/(2q - k_a)$, as is assumed.

For the case where $p = \infty$, take $r(\varepsilon) = \sup\{\|w_a\| : w \in \mathcal{W}\} < \infty$ $\forall \varepsilon > 0$ in the argument above. Then, Ossiander's L^∞ entropy condition holds provided $q > k_a/2$, as is assumed. $\square$

PROOF OF THEOREM 6: For $\mathcal{G} \cup \mathcal{G}^*$, the result is obvious. For $\mathcal{G} \oplus \mathcal{G}^*$, it suffices to suppose that $r = s = 1$. Let (g, a_j, b_j) and $(g^*, a_\ell^*, b_\ell^*)$ for $g \in \mathcal{G}$ and $g^* \in \mathcal{G}^*$ be defined analogously to (f, a_j, b_j) given in the definition of the L^p bracketing cover numbers. We have

$$(A.18) \quad \left(E(b_j + b_\ell^*)^p\right)^{1/p} \leq \left(Eb_j^p\right)^{1/p} + \left(Eb_\ell^{*p}\right)^{1/p} \leq 2\varepsilon \ , \quad \text{and so} \ ,$$

$$N_p^B(2\varepsilon, \ \mathrm{P}, \ \mathcal{G} \oplus \mathcal{G}^*) \leq N_p^B(\varepsilon, \ \mathrm{P}, \ \mathcal{G})N_p^B(\varepsilon, \ \mathrm{P}, \ \mathcal{G}^*) \ .$$

The result follows.

For $\mathcal{G} \vee \mathcal{G}^*$, it also suffices to suppose that $r = s = 1$. We have

$$(A.19) \quad |g \vee g^* - a_j \vee a_\ell^*| \leq |g - a_j| + |g^* - a_\ell^*| \leq b_j + b_\ell^* \ , \quad \text{and so,}$$

$$N_p^B(2\varepsilon, \ \mathrm{P}, \ \mathcal{G} \vee \mathcal{G}^*) \leq N_p^B(\varepsilon, \ \mathrm{P}, \ \mathcal{G})N_p^B(\varepsilon, \ \mathrm{P}, \ \mathcal{G}^*) \ .$$

The result for $\mathcal{G} \wedge \mathcal{G}^*$ is analogous.

For $|\mathcal{G}|$, the result follows from the inequality $\big|\, |g| - |a_j| \,\big| \leq |g - a_j|$.

Next consider $\mathcal{G}^{-1}$. For $g \in \mathcal{G}$, let g^ℓ denote the ℓ-th element of g for $\ell = 1, ..., L$, where $L = r^2$. By the same argument as used to prove the claim in the proof of the $\mathcal{G}^{-1}$ result of Theorem 3, there exist $r \times r$ matrix functions $a_1, ..., a_n$ and $b_1, ..., b_n$ such that (i) $a_j \in \mathcal{G}$ for all $j \leq n$, (ii) for all $g \in \mathcal{G}$, $|g^\ell - a_j^\ell| \leq b_j^\ell$ for all $\ell = 1, ..., L$ for some $j \leq n$, (iii) $\left(E\left(b_{j}^{\ell}\right)^p\right)^{1/p} < \varepsilon$ $\forall \ell, \forall j$, and (iv) $n \leq \Pi_{\ell=1}^L N_2^B(\varepsilon/2, \ \mathrm{P}, \ \mathcal{G}_\ell)$.

By an eigenvector/eigenvalue decomposition, we get $|g^{-1}| \leq 1_{r \times r}(I_r/\lambda_*)1_{r \times x} = 1_{r \times r}r/\lambda_*$ element by element and $|a_j^{-1}| \leq 1_{r \times r}r/\lambda_*$. Thus, for arbitrary unit vectors $b, c \in R^r$, we have: For any $g \in \mathcal{G}$ there exists a_j and b_j for which

$$(A.20) \quad |b'g^{-1}c - b'a_j^{-1}c| \leq |b|'|g^{-1}||a_j - g||a_j^{-1}||c| \leq (r^4/\lambda_*^2)1_r'b_j 1_r \ \text{and}$$

$$\left(E\left[(r^4/\lambda_*^2)1_r'b_j 1_r\right]^p\right)^{1/p} \leq (r^6/\lambda_*^2)\varepsilon \ .$$

Thus, $N_p^B(r^6\varepsilon/\lambda_*^2, \ \mathrm{P}, \ \mathcal{G}^{-1}) \leq n \leq \Pi_{\ell=1}^L N_2^B(\varepsilon/2, \ \mathrm{P}, \ \mathcal{G}_\ell)$ and the result follows.

To prove part (b) of Theorem 6 concerning $\mathcal{GH}$, note that each element of gh (for $g \in \mathcal{G}$ and $h \in \mathcal{H}$) is a finite union of products of scalar functions, and so, using the result for $\mathcal{G} \oplus \mathcal{G}^*$ it suffices to suppose that $r = s = u = 1$. Let (g, a_j, b_j) and (h, a_ℓ^*, b_ℓ^*) be defined analogously to (f, a_j, b_j) given in the definition of the L^p bracketing cover numbers, with $p = \lambda$ and $p = \mu$ respectively. We have

$$|gh - a_j a_\ell^*| \leq |gh - ga_\ell^*| + |ga_\ell^* - a_j a_\ell^*|$$

(A.21)

$$\leq Gb_\ell^* + |a_\ell^* - h + h|b_j \leq Gb_\ell^* + Hb_j + b_j b_\ell^*$$

and

$$\left[E(Gb_\ell^* + Hb_j + b_j b_\ell^*)^\alpha\right]^{\frac{1}{\alpha}} \leq \left(EG^\alpha b_\ell^{*\alpha}\right)^{\frac{1}{\alpha}} + \left(EH^\alpha b_j^\alpha\right)^{\frac{1}{\alpha}} + \left(Eb_j^\alpha b_\ell^{*\alpha}\right)^{\frac{1}{\alpha}}$$

(A.22)

$$\leq \left(EG^{\frac{\alpha\mu}{\mu-\alpha}}\right)^{\frac{\mu-\alpha}{\alpha\mu}}\left(Eb_\ell^{*\mu}\right)^{\frac{1}{\mu}} + \left(EH^{\frac{\alpha\lambda}{\lambda-\alpha}}\right)^{\frac{\lambda-\alpha}{\alpha\lambda}}\left(Eb_j^\lambda\right)^{\frac{1}{\lambda}} + \left(Eb_j^{\frac{\alpha\mu}{\mu-\alpha}}\right)^{\frac{\mu-\alpha}{\alpha\mu}}\left(Eb_\ell^{*\mu}\right)^{\frac{1}{\mu}}$$

$$\leq \sup_{t \leq T, T \geq 1}\left(\left(EG^\lambda\right)^{1/\lambda} + \left(EH^\mu\right)^{1/\mu}\right)\varepsilon + \varepsilon^2$$

$$\leq C^* \varepsilon$$

for $\varepsilon \in (0, 1]$, where C^* is defined implicitly and the dependence of each of the functions G, b_ℓ^*, etc. on W_t is suppressed for notational simplicity. The second and third inequalities hold by Hölder's inequality and the fact that $\lambda\mu/(\lambda+\mu) \geq \alpha$ implies that $\alpha\mu/(\mu-\alpha) \leq \lambda$ and $\alpha\lambda/(\lambda-\alpha) \leq \mu$. Equations (A.21) and (A.22) imply that

(A.23) $\quad N_\alpha^B(C^*\varepsilon, P, \mathcal{GH}) \leq N_\lambda^B(\varepsilon, P, \mathcal{G})N_\mu^B(\varepsilon, P, \mathcal{H})$

and the desired result follows. Note that using the notational conventions stated in the text, (A.21)-(A.23) hold whether or not $\alpha = \infty$, $\lambda = \infty$, or $\mu = \infty$. $\quad\square$

Footnotes

[1]This paper is a substantial revision of the first part of the paper Andrews (1989). I thank D. McFadden for comments and suggestions concerning this revision. I gratefully acknowledge research support from the Alfred P. Sloan Foundation and the National Science Foundation through a Research Fellowship and grant nos. SES-8618617, SES-8821021, and SES-9121914 respectively.

[2]That is, $\mathcal{T}$ is a metric space except that $\rho(\tau_1, \tau_2) = 0$ does not necessarily imply that $\tau_1 = \tau_2$. For example, the class of square integrable functions on $[0, 1]$ with $\rho(\tau_1, \tau_2) = \left[\int_0^1 (\tau_1(w) - \tau_2(w))^2 dw\right]^{1/2}$ is a pseudometric space, but not a metric space. The reason is that if $\tau_1(w)$ equals $\tau_2(w)$ for all w except one point, say, then $\rho(\tau_1, \tau_2) = 0$, but $\tau_1(\cdot) \neq \tau_2(\cdot)$. In order to handle sets $\mathcal{T}$ that are function spaces of the above type, we allow $\mathcal{T}$ to be a pseudometric space rather than a (more restrictive) metric space.

[3]The pseudometric $\rho(\cdot, \cdot)$ is defined here using a dummy variable N (rather than T) to avoid confusion when we consider objects such as $\text{plim}_{T \to \infty} \rho(\hat{\tau}, \tau_0)$. Note that $\rho(\cdot, \cdot)$ is taken to be independent of the sample size T.

[4]If need be, the bound in (5) can be replaced by $C|\log \delta|^{-\lambda}$ for arbitrary constants $C \in (1, ')$ and $\lambda \geq 1$ and Theorem 5 still goes through.

References

Aït-Sahalia, Y. (1992a): "Nonparametric Pricing of Interest Rate Derivative Securities," unpublished manuscript, Department of Economics, M.I.T.

Aït-Sahalia, Y. (1992b): "The Delta and Bootstrap Methods for Nonparametric Kernel Functionals," unpublished manuscript, Department of Economics, M.I.T.

Andrews, D. W. K. (1988a): "Asymptotics for Semiparametric Econometric Models: I. Estimation and Testing," Cowles Foundation Discussion Paper No. 908R, Yale University.

Andrews, D. W. K. (1988b): "Chi-square Diagnostic Tests for Econometric Models: Introduction and Applications," *Journal of Econometrics*, 37, 135-156.

Andrews, D. W. K. (1988c): "Chi-square Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 56, 1419-1453.

Andrews, D. W. K. (1989): "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity and Nonparametric Kernel Estimation," Cowles Foundation Discussion Paper No. 909R, Yale University.

Andrews, D. W. K. (1990): "Nonparametric Kernel Estimation for Semiparametric Models," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

Andrews, D. W. K. (1993a): "An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Variables," *Econometric Reviews*, 12, 183-216.

Andrews, D. W. K. (1993b): "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," unpublished manuscript, Cowles Foundation.

Andrews, D. W. K. and W. Ploberger (1991): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," Cowles Foundation Discussion Paper No. 1015, Yale University.

Arcones, M. and E. Gine (1992): "On the Bootstrap of *M*-estimators and Other Statistical Functionals," in *Exploring the Limits of the Bootstrap*, ed. by R. LePage and L. Billard. New York: Wiley.

Bera, A. K. and M. L. Higgins (1992): "A Test for Conditional Heteroskedasticity in Time Series Models," *Journal of Time Series Analysis*, 13, 501-519.

Bierens, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

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

Cavanagh, C. and R. P. Sherman (1992): "Rank Estimators for Monotone Index Models," Bellcore Economics Discussion Paper No. 84, Bellcore, Morristown, NJ.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-324.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247-254.

Davies, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33-43.

DeJong, R. M. (1992): "The Bierens Test under Data Dependence," unpublished manuscript, Department of Econometrics, Free University, Amsterdam.

Dudley, R. M. (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899-929.

Dudley, R. M. (1987): "Universal Donsker Classes and Metric Entropy," *Annals of Probability*, 15, 1306-1326.

Gallant, A. R. (1989): "On Asymptotic Normality When the Number of Regressors Increases and the Minimum Eigenvalue of $X'X/n$ Decreases," Institute of Statistics Mimeograph Series No. 1955, North Carolina State University, Raleigh, NC.

Gallant, A. R. and G. Souza (1989): "On the Asymptotic Normality of Fourier Flexible Form Estimates," unpublished manuscript, Department of Statistics, North Carolina State University, Raleigh, NC.

Gine, E. and J. Zinn (1990): "Bootstrapping General Empirical Measures," *Annals of Probability*, 18, 851-869.

Hahn, J. (1992): "Bootstrapping *M*-estimators: The Asymptotic Theory," unpublished manuscript, Department of Economics, Harvard University.

Hansen, B. E. (1991): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," Working Paper No. 296, Rochester Center for Economic Research, University of Rochester.

Hansen, B. E. (1992): "Testing the Conditional Mean Specification in Parametric Regression Using the Empirical Score Process," unpublished manuscript, Department of Economics, University of Rochester.

Hansen, B. E. (1993): "The Likelihood Test under Non-standard Conditions: Testing the Markov Trend Model of GNP," *Journal of Applied Econometrics*, forthcoming.

Hardle, W. and O. Linton (1993): "Applied Nonparametric Methods," in *Handbook of Econometrics*, Volume 4. Amsterdam: North-Holland.

Honore, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

Horowitz, J. L. (1988): "Semiparametric *M*-estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45-83.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-531.

Horowitz, J. L. and G. R. Neumann (1992): "A Generalized Moments Specification Test of the Proportional Hazards Model," *Journal of the American Statistical Association*, 87, 234-240.

Huber, P. J. (1967): "The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, 221-233. Berkeley: University of California.

Huber, P. J. (1973): "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1, 799-821.

Kim, J. and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191-219.

Klecan, L., R. McFadden, and D. McFadden (1990): "A Robust Test for Stochastic Dominance," unpublished manuscript, Department of Economics, M.I.T.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33-50.

Kolmogorov, A. N. and V. M. Tihomirov (1961): "ε-entropy and ε-capacity of Sets in Functional Spaces," *American Mathematical Society Translations*, Ser. 2, Vol. 17, 277-364.

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-228.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

Newey, W. K. (1989): "The Asymptotic Variance of Semiparametric Estimators," unpublished manuscript, Department of Economics, Princeton University.

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809-837.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161-1167.

Newey, W. K. and D. McFadden (1993): "Estimation in Large Samples," in *Handbook of Econometrics*, Volume 4. Amsterdam: North-Holland.

Newey, W. K. and J. L. Powell (1987): "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819-847.

Olley, S. and A. Pakes (1991): "The Dynamics of Productivity in the Telecommunications Equipment Industry," unpublished manuscript, Department of Economics, Yale University.

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

Pakes, A. and S. Olley (1991): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," unpublished manuscript, Department of Economics, Yale University.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

Pollard, D. (1982): "A Central Limit Theorem for Empirical Processes," *Journal of the Australian Mathematical Society (Series A)*, 33, 235-248.

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

Pollard, D. (1985): "New Ways to Prove Central Limit Theorems," *Econometric Theory*, 1, 295-314.

Pollard, D. (1989): "A Maximal Inequality for Sums of Independent Processes under a Bracketing Condition," unpublished manuscript, Department of Statistics, Yale University.

Pollard, D. (1990): *Empirical Processes: Theory and Applications*. CBMS Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

Powell, J. L. (1986a): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.

Powell, J. L. (1986b): "Symmetrically Trimmed Least Squares Estimators for Tobit Models," *Econometrica*, 54, 1435-1460.

Prohorov, Yu. V. (1956): "Convergence of Random Processes and Limit Theorems in Probability Theory," *Theory of Probability and Its Applications*, 1, 157-214.

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

Sherman, R. P. (1992): "Maximal Inequalities for Degenerate U-processes with Applications to Optimization Estimators," unpublished manuscript, Bell Communications Research, Morristown, NJ.

Sherman, R. P. (1993a): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123-137.

Sherman, R. P. (1993b): "U-processes in the Analysis of a Generalized Semiparametric Regression Estimator," *Econometric Theory*, 9, forthcoming.

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

Stinchcombe, M. B. and H. White (1993): "Consistent Specification Testing with Unidentified Nuisance Parameters Using Duality and Banach Space Limit Theory," unpublished manuscript, Department of Economics, University of California, San Diego.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426-482.

Whang, Y.-J. and D. W. K. Andrews (1993): "Tests of Model Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277-318.

White, H. and Y. Hong (1992): "M-testing Using Finite and Infinite Dimensional Parameter Estimators," unpublished manuscript, Department of Economics, University of California, San Diego.

White, H. and M. Stinchcombe (1991): "Adaptive Efficient Weighted Least Squares with Dependent Observations," in *Directions in Robust Statistics and Diagnostics*, Part II, ed. by W. Stahel and S. Weisberg. Berlin: Springer.

Yatchew, A. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435-451.