COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1020

An Introduction to Econometric Applications of Functional Limit Theory for Dependent Random Variables

by

Donald W.K. Andrews

May 1992

AN INTRODUCTION TO ECONOMETRIC APPLICATIONS OF FUNCTIONAL LIMIT THEORY FOR DEPENDENT RANDOM VARIABLES

Donald W. K. Andrews[1]

Cowles Foundation for Research in Economics
Yale University

May 1992

ABSTRACT

This paper discusses some uses in econometrics of functional limit theory for dependent random variables. Attention is focussed on empirical process-type results rather than partial sum results that are prevalent in unit root econometrics. Examples considered include non-standard parametric hypotheses tests and semiparametric estimation. The application of bracketing functional limit results is discussed in some detail.

1. INTRODUCTION

The purpose of this paper is to illustrate how functional limit theory for dependent random variables can be usefully exploited in a number of econometric applications. Several econometric examples are discussed. Bracketing functional limit results are surveyed and their application to the econometric examples is described. The examples include several cases of tests when a nuisance parameter is present only under the alternative and semiparametric estimation based on a criterion function that depends on an estimator of a preliminary infinite-dimensional nuisance parameter.

Section 2 presents some basics concerning functional limit theory and introduces notation.

Section 3 introduces the general problem of testing in a parametric model when a nuisance parameter is present only under the alternative. Asymptotically optimal tests, due to Andrews and Ploberger (1991), are defined. Four examples are introduced. The first is a test of a threshold effect in a threshold autoregressive model. The second is a test for relevance of a regressor variable/vector in a nonlinear regression model. The third is a test for autocorrelation in the errors of a linear regression model against autoregressive-moving average (ARMA) alternatives of order (1, 1). The fourth is a test for conditional heteroskedasticity in the errors of a regression (or univariate time series) model against generalized autoregressive conditional heteroskedasticity (GARCH) alternatives of order (1, 1).

Sufficient conditions for obtaining the asymptotic null distributions of the Andrews-Ploberger tests are given in Section 3 for the general testing problem. A key high-level condition is the fulfillment of a functional limit result for a sequence of stochastic processes indexed by the parameter that appears under the alternative but not under the null. The precise form of the requisite functional limit result is given for each of the four examples mentioned above.

Section 4 initiates a discussion of bracketing functional limit/stochastic equicontinuity results. The section begins by introducing *bracketing cover numbers*. It then outlines what bracketing functional limit/stochastic equicontinuity results are available in the literature, with the emphasis being on results for dependent random variables. These bracketing results all employ a bracketing cover number assumption. An *L^p-continuity condition* is introduced that is sufficient for the bracketing cover number assumption. The verification of the L^p-continuity condition in each of the four hypothesis test examples is discussed. Alternative approaches to functional limit theory, including the Vapnik/Cervonenkis (VC)/symmetrization approach, are discussed briefly in relation to the examples.

Section 5 considers semiparametric estimation problems in which functional limit theory can be exploited. It gives a heuristic general discussion of how *stochastic equicontinuity* can be used in establishing the asymptotic distribution of semiparametric estimators. (As discussed below, stochastic equicontinuity forms an essential building block for functional limit results.) Section 5 describes an example of the method proposed. The example is the estimation of regression parameters in a time series regression model with conditional heteroskedasticity of unknown form.

The bracketing results outlined in Section 4 can be used to obtain the stochastic equicontinuity properties needed in the semiparametric estimation problems of Section 5. To use these bracketing results, however, one needs to be able to verify the bracketing cover number assumptions for infinite-dimensional classes of functions. Section 6 addresses this problem. It gives a sufficient condition in terms of smoothness of the functions. Extensions of the given smoothness conditions are discussed. Section 6 also describes results that allow one to manipulate bracketing cover number bounds. Such results allow one to amalgamate known cover number bounds to obtain bounds for new situations. They are quite handy in verifying bracketing cover number assumptions in applications. The results introduced in this section are applied to the example of estimation in a regression model with conditional heteroskedasticity of unknown form.

The last section of this paper, Section 7, briefly describes several possible areas of future research on functional limit theory for dependent random variables. These areas are chosen with the applications described in the paper in mind.

2. BASICS/NOTATION

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let Π denote an index set for various stochastic processes that will be considered. For example, Π might be a subset of Euclidean space or a set of functions. Let ρ be a pseudometric metric on Π (i.e., a metric without the condition that $\rho(\pi_1, \pi_2) = 0$ implies $\pi_1 = \pi_2$). We consider a sequence of stochastic processes $\{\nu_T(\cdot) : T \geq 1\}$ on Ω indexed by $\pi \in \Pi$. The rv $\nu_T(\pi)$ is R^v-valued. In applications T is the sample size. For example, $\nu_T(\pi)$ may be of the empirical process form

$$\nu_T(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(m(W_t, \pi) - Em(W_t, \pi)) , \tag{2.1}$$

where $\{W_t : t \geq 1\}$ is a sequence of rv's, or of the more general form

$$\nu_T(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(m_t(\pi) - Em_t(\pi)) , \tag{2.2}$$

where $m_t(\pi)$ is a rv for each $\pi \in \Pi$, $t \geq 1$, but $m_t(\pi)$ is not a function of a single rv W_t for all π. Examples of both types are considered below. Let $\nu(\cdot)$ be another stochastic process indexed by Π (that may or may not be defined on the same probability space $(\Omega, \mathcal{A}, P)$ as $\nu_T(\cdot)$ $\forall T \geq 1$).

Let $\Rightarrow$ denote weak convergence of stochastic processes. Let $\xrightarrow{d}$ denote convergence in distribution of rv's. All limits are taken as $T \to \infty$. Let $\|\cdot\|$ denote the Euclidean norm.

Definition of Weak Convergence:

$$\nu_T(\cdot) \Rightarrow \nu(\cdot) \quad \text{if} \quad E^*f(\nu_T(\cdot)) \to Ef(\nu(\cdot)) \quad \forall f \in \mathcal{U}(B(\Pi)) ,$$

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

In the definition, E^* denotes outer probability. (It is used because it is desirable not to require $\nu_T(\cdot)$ to be a measurable random element of the metric space $(B(\Pi), d)$ with its Borel σ-field. The limit stochastic process $\nu(\cdot)$, on the other hand, will be sufficiently well-behaved in applications that it is assumed to be measurable in the definition.)

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

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).

PROPOSITION: *If* (i) (Π, ρ) *is a totally bounded pseudo-metric space,*

(ii) *fidi convergence holds:* $\forall$ *finite subsets* $(\pi_1, ..., \pi_J)$ *of* Π, $(\nu_T(\pi_1)', ..., \nu_T(\pi_J)')'$ *converges in distribution,*

(iii) *stochastic equicontinuity holds:* $\forall \varepsilon > 0$, $\forall \delta > 0$, $\exists \delta > 0$ *such that*

$$\overline{\lim_{T\to\infty}} P^*\Big(\sup_{\substack{\pi_1,\pi_2\in\Pi \\ \rho(\pi_1,\pi_2)<\delta}} \|\nu_T(\pi_1) - \nu_T(\pi_2)\| > \eta\Big) < \varepsilon ,$$

then there exists a (Borel-measurable with respect to d) B(Π)-valued stochastic process v(·) whose sample paths are uniformly ρ continuous with probability one such that $\nu_T(\cdot) \Rightarrow \nu(\cdot)$.

Conversely, if $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ *for* $\nu(\cdot)$ *with the properties above and* (i) *holds, then* (ii) *and* (iii) *hold.*

In some applications, such as the semiparametric ones considered below, we do not need a full functional limit (i.e., weak convergence) result. We only need to establish stochastic equicontinuity. In other applications, such as the non-standard testing problems considered below, we do need a full functional limit result. For these cases, too, we focus attention in this paper on establishing stochastic equicontinuity, because fidi convergence typically can be obtained straightforwardly by applying a standard multivariate central limit theorem (CLT) and total boundedness of (Π, ρ) is typically obtained as a by-product of establishing stochastic equicontinuity. The above proposition then yields the desired functional CLT.

Since stochastic equicontinuity is a key property for the problems considered here, we provide a brief discussion of it. Basically, a stochastic process $\nu_T(\cdot)$ is stochastically equicontinuous if $\nu_T(\pi)$ is continuous in π uniformly over Π at least with high probability and for T large. Thus, stochastic equicontinuity is a probabilistic and asymptotic generalization of uniform continuity of a function.

Two equivalent definitions of stochastic equicontinuity are: (i) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous if for every sequence of constants $\{\delta_T : T \geq 1\}$ that converges to zero, $\sup_{\pi_1, \pi_2 \in \Pi : \rho(\pi_1, \pi_2) < \delta_T} |\nu_T(\pi_1) - \nu_T(\pi_2)| \xrightarrow{p} 0$. (ii) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous if for all sequences of random elements $\{\hat{\pi}_{1T} : T \geq 1\}$ and $\{\hat{\pi}_{2T} : T \geq 1\}$ in Π that satisfy $\rho(\hat{\pi}_{1T}, \hat{\pi}_{2T}) \xrightarrow{p} 0$, we have $\nu_T(\hat{\pi}_{1T}) - \nu_T(\hat{\pi}_{2T}) \xrightarrow{p} 0$. The latter characterization of stochastic equicontinuity reflects its use in the semiparametric examples below.

To demonstrate the plausibility of the stochastic equicontinuity property, suppose $\nu_T(\pi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (W_t'\pi - EW_t'\pi)$ for $\pi \in \Pi \subset R^s$ and ρ is the Euclidean metric. In this simple linear case,

$$
\begin{aligned}
&\overline{\lim_{T\to\infty}}\, P^*\Big(\sup_{\pi_1, \pi_2 \in \Pi : \rho(\pi_1, \pi_2) < \delta} \|\nu_T(\pi_1) - \nu_T(\pi_2)\| > \eta\Big) \\
&= \overline{\lim_{T\to\infty}}\, P^*\Big(\sup_{\pi_1, \pi_2 \in \Pi : \rho(\pi_1, \pi_2) < \delta} \Big\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (W_t - EW_t)'(\pi_1 - \pi_2)\Big\| > \eta\Big) \\
&\leq \overline{\lim_{T\to\infty}}\, P\Big(\Big\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (W_t - EW_t)\Big\| > \eta/\delta\Big) \\
&< \varepsilon ,
\end{aligned}
\tag{2.3}
$$

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

When $\nu_T(\pi)$ is not linear in π, the stochastic equicontinuity property is substantially more difficult to verify than in the above linear case. In addition, stochastic equicontinuity need not hold. For example, if $\nu_T(\pi)$ is of the empirical process form $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (m(W_t, \pi) - Em(W_t, \pi))$, some restrictions are needed on the functions $m(\cdot,\cdot)$ and on the index set Π for stochastic equicontinuity to hold. Sections 4 and 6 below present some sufficient conditions.

We now turn to the first set of examples.

3. TESTS WHEN A NUISANCE PARAMETER IS PRESENT ONLY UNDER THE ALTERNATIVE

3.1. Introduction and Definition of Optimal Tests

In this section we consider a class of testing problems for which functional limit theory can be usefully exploited. The testing problems considered are ones for which a nuisance parameter is present only under the alternative. Such testing problems are non-standard. In consequence, the usual asymptotic distributional and optimality properties of likelihood ratio (LR), Lagrange multiplier (LM), and Wald (W) tests do not apply.

Consider a parametric model with parameters θ and π, where $\theta \in \Theta \subset R^s$, $\pi \in \Pi \subset R^v$. 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

$$(3.1) \qquad \begin{aligned} H_0 &: \beta = \underset{\sim}{0} \\ H_1 &: \beta \neq \underset{\sim}{0} \; . \end{aligned}$$

Under the null hypothesis, the distribution of the data does not depend on the parameter π by assumption. Under the alternative hypothesis, it does. A simple example of this type of testing problem is the changepoint problem with unknown changepoint. In this problem, π indexes the changepoint as a fraction of the sample size. Observations for $t \leq T\pi$ have a distribution that depends on a parameter δ and observations for $t > T\pi$ have a distribution that depends on a parameter $\delta+\beta$. Under the null hypothesis the parameter π is not present because there is no changepoint.

Problems of the sort considered above were first treated in a general way by Davies (1977, 1987). Davies proposed using the LR test. Let LR(π) denote the LR test statistic (i.e., minus two times the log likelihood ratio) when π is specified under the alternative. For given π, LR(π) has standard asymptotic properties (under standard regularity conditions). In particular, it has a χ^2_p null distribution. When π is not given, but is allowed to take any value in Π, the LR statistic is

$$\sup_{\pi \in \Pi} \mathrm{LR}(\pi) \ . \tag{3.2}$$

Using functional limit results, the asymptotic null distribution of $\sup_{\pi \in \Pi} \mathrm{LR}(\pi)$ can be shown to be that of the supremum of a chi-square process.

Hansen (1991a) 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:

$$\text{Exp-LR} = (1+c)^{-p/2} \int \exp\left(\frac{1}{2}\,\frac{c}{1+c}\,\mathrm{LR}(\pi)\right) dJ(\pi) \ , \tag{3.3}$$

where $J(\cdot)$ is a specified weight function over $\pi \in \Pi$ 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 (3.3), but with $\mathrm{LR}(\pi)$ replaced by $\mathrm{LM}(\pi)$ and $\mathrm{W}(\pi)$, respectively, where the latter are defined analogously to $\mathrm{LR}(\pi)$. The three statistics Exp-LR, Exp-LM, and Exp-W each have asymptotic optimality properties. Using functional limit results, each can be shown to have an asymptotic null distribution that is a function of some stochastic process. In the remainder of this section, we focus on this use of functional limit theory.

3.2. Examples

3.2.1. Test of a Threshold Effect

The first example is a test of a threshold effect in a simple threshold autoregressive (AR) model. The model is

$$(3.4) \quad Y_t = \begin{cases} \delta_1 + \delta_2 Y_{t-1} + U_t & \text{for } Y_{t-d} > \pi \\ \delta_1 + (\delta_2 + \beta) Y_{t-1} + U_t & \text{for } Y_{t-d} \le \pi \end{cases}, \quad U_t \sim \text{iid } N(0, \delta_3), \quad t = 0, \dots, T,$$

where $\delta_2 \in (-1, 1)$ and the delay parameter d is a positive integer. This model and generalizations of it, including AR(p) threshold models for $p > 1$ and smooth transition AR models, have been applied in the physical and biological sciences, e.g., see Tong (1983, 1990), as well as in economics, e.g., see Potter (1989), Teräsvirta and Anderson (1991), and Hansen (1991a). A test of the threshold effect, viz., $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, exhibits the feature that the threshold parameter π is present only under the alternative. Note that in smooth transition AR models the parameter π is not necessarily scalar. It is given by the vector of parameters that index the transition function. LR tests of H_0 of the Davies "sup" form have been considered by Chan (1990) and Chan and Tong (1991).

3.2.2. Test for Variable Relevance

This example is a test of whether a regressor variable/vector Z_t belongs in a nonlinear regression model. The model is

$$(3.5) \qquad Y_t = g(X_t, \delta_1) + \beta h(Z_t, \pi) + U_t, \quad U_t \sim N(0, \delta_2), \quad t = 1, \dots, T.$$

The functions g and h are assumed known. The parameters $(\beta, \delta_1, \delta_2, \pi)$ 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 $h(Z_t, \pi)$ might be of the Box-Cox form $(Z_t^\pi - 1)/\pi$.

Under the null hypothesis $H_0 : \beta = 0$, Z_t does not enter the regression function and the parameter π is not present.

3.2.3. Test for Autocorrelation

In this example, we consider testing for autocorrelation in the errors of a regression model against autoregressive-moving average (ARMA) alternatives of order (1, 1). The model is

$$(3.6)\qquad \begin{aligned} Y_t &= X_t'\delta_1 + U_t\ , \\ U_t &= (\pi+\beta)U_{t-1} + \varepsilon_t - \pi\varepsilon_{t-1}\ ,\quad \varepsilon_t \sim \text{iid } N(0,\ \delta_2)\ ,\quad t = 0,\ \dots,\ T\ , \end{aligned}$$

where Θ and Π are such that the ARMA parameters $(\pi+\beta,\ \pi)$ lie in the stationary region.

When $\beta = 0$, the errors satisfy $U_t - \pi U_{t-1} = \varepsilon_t - \pi\varepsilon_{t-1}$, which is equivalent to $U_t = \varepsilon_t$ (because of the existence of a common factor in the AR and MA lag polynomials). Thus, under the null hypothesis $H_0 : \beta = 0$, the errors are uncorrelated and the distribution of the data does not depend on π. Under the alternative, on the other hand, the errors are of ARMA(1, 1) form with ARMA parameters $(\pi+\beta,\ \pi)$.

It is natural to consider tests against ARMA(1, 1) alternatives, since ARMA(1, 1) models have been found to yield parsimonious approximations to the distributions of many stationary time series. To date, tests against such alternatives have not been considered because of the non-standard feature of the resulting testing problem.

3.2.4. Test for Conditional Heteroskedasticity

As a fourth example, we consider testing for conditional heteroskedasticity in linear regression models against GARCH(1, 1) errors. (See Engle (1982) and Bollerslev (1986) regarding ARCH and GARCH models of conditional heteroskedasticity.) The model is

$$(3.7)\qquad \begin{aligned} &Y_t = X_t'\delta_1 + U_t\ ,\ U_t|\psi_{t-1} \sim N(0,\ h_t) \\ &h_t = \delta_2(1-\pi) + \beta U_{t-1}^2 + \pi h_{t-1}\ ,\ \psi_t = \sigma(\dots,\ U_{t-1},\ U_t,\ \dots,\ X_t,\ X_{t+1})\ ,\ t = 1,\ \dots,\ T\ , \end{aligned}$$

where Θ and Π are such that $\beta \geq 0$, $\pi \geq 0$, and $\beta+\pi < 1$. Under the null hypothesis $H_0 : \beta = 0$ with start-up variance $h_0 = \delta_2$, the conditional variance h_t is a constant for all t, viz., δ_2, and the parameter π drops out of the expression for h_t.

The GARCH(1, 1) model has been found to perform well for various financial and international financial variables such as stock returns and exchange rate changes. Thus, it is a natural choice for parametric alternative when testing for conditional heteroskedasticity. Tests of the Davies "sup" form for GARCH(1, 1) errors have been discussed briefly by Hansen (1991a) and for a nonlinear ARCH form of conditional heteroskedasticity by Bera and Higgins (1990).

3.3. Sufficient Conditions for Asymptotic Distributions and Optimality Results

In this section we provide sufficient conditions for the average exponential test statistics introduced above to have a particular asymptotic null distribution. In likelihood scenarios, these conditions are also sufficient for asymptotic optimality of these test statistics according to a weighted average power criterion function, see Andrews and Ploberger (1991). The sufficient conditions that are given are high level conditions that need to be verified in any given application. For example, this is done in Andrews and Ploberger (1991) for changepoint tests. Most of the sufficient conditions are straightforward to verify. The one condition that is potentially troublesome in some examples is one that requires a functional limit theorem to hold. To verify this condition one needs to apply an appropriate weak convergence result for a sequence of stochastic processes. Details are given below.

First, we introduce some notation. Let $\ell_T(\theta, \pi)$ denote a criterion function that is used to estimate the parameters θ and π. The leading case is when $\ell_T(\theta, \pi)$ is the log likelihood function for the sample of size T. Let $D\ell_T(\theta, \pi)$ denote the s-vector of partial derivatives of $\ell_T(\theta, \pi)$ with respect to θ. Let $D^2\ell_T(\theta, \pi)$ denote the $s \times s$ matrix of second partial derivatives of $\ell_T(\theta, \pi)$ with respect to θ. (Note that $D\ell_T(\theta_0, \pi)$ and $D^2\ell_T(\theta_0, \pi)$ depend on π in general even though $\ell_T(\theta_0, \pi)$ does not.) Let θ_0 denote the true value of θ under the null H_0, i.e., $\theta_0 = (0', \delta_0')'$.

All limits below are taken "as $T \to \infty$" unless stated otherwise. We say that a statement holds "under θ_0" if it holds when the null hypothesis is true. Let $\lambda_{\min}(A)$ denote the smallest eigenvalue of a matrix A.

The criterion function/parametric model is assumed to satisfy:

ASSUMPTION 1: (a) $\ell_T(\theta, \pi)$ *does not depend on* π *for all* θ *in the null hypothesis.*

(b) θ_0 *is an interior point of* Θ.

(c) $\ell_T(\theta, \pi)$ *is twice continuously partially differentiable in* θ *for all* $\theta \in \Theta_0$ *and all* $\pi \in \Pi$ *with probability one under* θ_0, *where* Θ_0 *is some neighborhood of* θ_0.

(d) $-\frac{1}{T}D^2\ell_T(\theta, \pi) \xrightarrow{p} \mathcal{I}(\theta, \pi)$ *uniformly over* $\pi \in \Pi$ *and* $\theta \in \Theta$ *under* θ_0 *for some non-random* $s \times s$ *matrix function* $\mathcal{I}(\theta, \pi)$.

(e) $\mathcal{I}(\theta, \pi)$ *is uniformly continuous in* (θ, π) over $\Theta_0 \times \Pi$.

(f) $\inf_{\pi \in \Pi} \lambda_{\min}(\mathcal{I}(\theta_0, \pi)) > 0$.

In likelihood scenarios, the matrix function $\mathcal{I}(\theta, \pi)$ introduced in Assumption 1 is the asymptotic *information matrix* for θ for given π, which depends on both θ and π.

We briefly comment on Assumption 1. Assumption 1(a) specifies the crucial feature of the testing problem under consideration. Assumptions 1(b) and (c) are standard assumptions (though differentiability with respect to θ can be circumvented). For a fixed value of π, Assumption 1(d) can be verified under standard assumptions using a suitable weak law of large numbers (WLLN). Uniform convergence over $\pi \in \Pi$ can then be obtained, e.g., by using the generic uniform convergence results in Newey (1991) or Andrews (1992). Assumptions 1(e) and (f) also are standard assumptions except that they are required to hold uniformly over values of the nuisance parameter $\pi \in \Pi$. Nevertheless, Assumption 1(f) does not hold even for a fixed value of π in mixture models or, more generally, in regime switching models with unobserved regimes. Assumption 1(f) does hold for each fixed value of π, however, in the examples of Section 3.2 above. The uniformity requirements in Assumptions 1(d)-(f) may restrict the class Π that can be considered. For example, in the one-time changepoint case, uniformity requires that the closure of Π is bounded away from 0 and 1. That is, one cannot consider changepoints that are arbitrarily close to the beginning or end of the sample.

Let $\hat{\theta}(\pi)$ $(= \hat{\theta}_T(\pi))$ be the *unrestricted estimator* of θ for fixed $\pi \in \Pi$. (That is, $\hat{\theta}(\pi)$ satisfies $\ell_T(\hat{\theta}(\pi), \pi) = \max_{\theta \in \Theta} \ell_T(\theta, \pi)$ $\forall \pi \in \Pi$ with probability that goes to one as $T \to \infty$ under θ_0.) Let $\tilde{\theta}$ be the *restricted estimator* of θ. (That is, $\tilde{\theta}$ satisfies $\tilde{\theta} \in \tilde{\Theta} = \{\theta \in \Theta : \theta = (0', \delta')'$ for some $\delta \in R^q\}$ and $\ell_T(\tilde{\theta}, \pi) = \max_{\theta \in \tilde{\Theta}} \ell_T(\theta, \pi)$ with probability that goes to one as $T \to \infty$ under θ_0.) Note that since $\ell_T(\theta, \pi)$ does not depend on π when θ is in the null hypothesis, $\tilde{\theta}$ does not depend on π.

We assume that uniform consistency of $\hat{\theta}(\pi)$ and consistency of $\tilde{\theta}$ has already been established.

ASSUMPTION 2: (a) $\sup_{\pi \in \Pi} \|\hat{\theta}(\pi) - \theta_0\| \xrightarrow{p} 0$ *under* θ_0.

(b) $\tilde{\theta} \xrightarrow{p} \theta_0$ *under* θ_0.

For example, sufficient conditions for Assumption 2 (see Lemma A-1 of Andrews (1989c)) are:

(3.8) (i) $\sup_{\theta \in \Theta, \pi \in \Pi} \left| \frac{1}{T} \ell_T(\theta, \pi) - \ell(\theta, \pi) \right| \xrightarrow{p} 0$ under θ_0 for some real function ℓ on $\Theta \times \Pi$.

(ii) For every neighborhood Θ_0 ($\subset \Theta$) of θ_0, $\sup_{\pi \in \Pi} (\sup_{\theta \in \Theta / \Theta_0} \ell(\theta, \pi) - \ell(\theta_0, \pi)) < 0$.

Now, we have introduced sufficient notation to define the standard LR, LM, and W statistics for testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ for given π:

$$
\begin{aligned}
LR_T(\pi) &= -2(\ell_T(\tilde{\theta}, \pi) - \ell_T(\hat{\theta}(\pi), \pi)) , \\
LM_T(\pi) &= \left(\frac{1}{\sqrt{T}} D\ell_T(\tilde{\theta}, \pi) \right)' \mathcal{I}_T^{-1}(\tilde{\theta}, \pi) \frac{1}{\sqrt{T}} D\ell_T(\tilde{\theta}, \pi) \text{ , and} \\
W_T(\pi) &= T(H\hat{\theta}(\pi))' \left[H \mathcal{I}_T^{-1}(\hat{\theta}(\pi), \pi) H' \right]^{-1} H\hat{\theta}(\pi) \text{ , where} \\
H &= [I_p \; \vdots \; 0] \subset R^{p \times s} \text{ and } \mathcal{I}_T(\theta, \pi) = -\frac{1}{\sqrt{T}} D^2 \ell_T(\theta, \pi) .
\end{aligned} \tag{3.9}
$$

Note that only the first p elements of $\frac{1}{\sqrt{T}} D\ell_T(\tilde{\theta}, \pi)$ are non-zero in the definition of $LM_T(\pi)$, because $\frac{\partial}{\partial \delta} \ell_T(\tilde{\theta}, \pi) = 0$ by the first order conditions for the restricted estimator $\tilde{\theta}$. Also note that the $LM_T(\pi)$ statistic is constructed using only the restricted estimator $\tilde{\theta}$ and, hence, only requires estimation of the model one time. This has considerable computational advantages, especially in nonlinear models.

The exponential LR, LM, and W statistics, Exp-LR_T, Exp-LM_T, and Exp-W_T, respectively, are defined by combining (3.3) and (3.9).

Let $\nu_T(\pi) = \frac{1}{\sqrt{T}} D\ell_T(\theta_0, \pi)$. We assume that the normalized score function $\nu_T(\pi)$ satisfies:

ASSUMPTION 3: $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ *under* θ_0 *(as processes indexed by* $\pi \in \Pi$*) for some* R^s*-valued stochastic process* $\{\nu(\pi) : \pi \in \Pi\}$ *that has bounded uniformly* ρ *continuous sample paths with probability one.*

Typically, $\nu(\cdot)$ is a mean zero Gaussian stochastic process.

The verification of Assumption 3 requires a functional CLT. This is the aspect of the testing problems outlined above that is of particular interest in the present paper. As is discussed below, Assumption 3 can be verified in the examples of Section 3.2 by applying an empirical process or related CLT. In the changepoint example, it can be verified by applying a functional CLT for a partial sum process, as in Andrews and Ploberger (1991).

Given Assumptions 1-3, the asymptotic null distribution of Exp-LR$_T$, Exp-LM$_T$, and Exp-W$_T$ is shown in the following theorem to equal that of the random variable

$$\begin{aligned}\chi(\theta_0,c) = (1+c)^{-p/2}\int \exp\bigg(\frac{1}{2}\,\frac{c}{1+c}\Big(H\mathcal{I}^{-1}(\theta_0,\pi)\nu(\pi)\Big)' \\ \times \Big(H\mathcal{I}^{-1}(\theta_0,\pi)H'\Big)^{-1}H\mathcal{I}^{-1}(\theta_0,\pi)\nu(\pi)\bigg)dJ(\pi)\ .\end{aligned} \tag{3.11}$$

In likelihood scenarios, where $\nu(\cdot)$ is a mean zero Gaussian process with $E\nu(\pi)\nu(\pi)' = \mathcal{I}(\theta_0, \pi)$ $\forall\pi \in \Pi$, the quadratic form in the exponent has a χ^2_p distribution $\forall\pi \in \Pi$. In this case, $\chi(\theta_0, c)$ is a function of a chi-squared stochastic process.

THEOREM (Andrews and Ploberger): *Under the null hypothesis and Assumptions* 1-3, (a) Exp-LR$_T \xrightarrow{d} \chi(\theta_0, c)$, (b) Exp-LM$_T \xrightarrow{d} \chi(\theta_0, c)$, and (c) Exp-W$_T \xrightarrow{d} \chi(\theta_0, c)$.

COMMENT: In some applications, e.g., changepoint applications and the variable relevance, ARMA(1, 1), and GARCH(1, 1) applications of Section 3.2, the limit distribution $\chi(\theta_0, c)$ does not depend on θ_0. Hence, one can obtain critical values for the exponential LR, LM, and Wald tests by simulating the distribution $\chi(\theta_0, c)$. In other applications, e.g., the threshold AR application of Section 3.2, $\chi(\theta_0, c)$ does depend on θ_0. In such cases, one can obtain asymptotically valid critical values by simulating $\chi(\theta^*, c)$, where θ^* is some estimator of θ that is consistent under the null, provided $\nu(\pi) = \nu(\theta_0, \pi)$ is continuous at θ_0 uniformly over $\pi \in \Pi$. See Hansen (1991a, Sec. 7) for a method of simulating a single realization of $\chi(\theta^*, c)$ (which is a function of the stochastic process $\nu(\theta^*, \cdot)$).

3.4. Examples (cont.)

We return now to the three examples introduced in Section 3.2 and provide expressions for $\ell_T(\theta, \pi)$ and $\frac{1}{\sqrt{T}} D\ell_T(\theta_0, \pi)$ for the case where $\ell_T(\theta, \pi)$ is the log likelihood (or pseudo-likelihood) function. In Section 4 we discuss verification of the key assumption that $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ in each of these examples.

3.4.1. Test of a Threshold Effect (cont.)

For the threshold AR model, we have

$$\ell_T(\theta,\pi) = -\frac{T}{2}\log \delta_3 - \frac{1}{2\delta_3}\sum_{t=1}^{T}(Y_t - \delta_1 - \delta_2 Y_{t-1} - \beta Y_{t-1}1(Y_{t-d} \le \pi))^2 \quad \text{and}$$

$$\frac{1}{\sqrt{T}} D\ell_T(\theta_0, \pi) = \begin{pmatrix} \frac{1}{\delta_{30}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t Y_{t-1} 1(Y_{t-d} \le \pi) \\ \frac{1}{2\delta_{30}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t \\ \frac{1}{2\delta_{30}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t Y_{t-1} \\ \frac{1}{2\delta_{30}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} (U_t^2/\delta_{30} - 1) \end{pmatrix}. \tag{3.12}$$

(As given, $\ell_T(\theta, \pi)$ is the log likelihood function conditional on the first observation Y_0.) To verify Assumption 3 in this example, we need to verify that $\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t Y_{t-1} 1(Y_{t-d} \le \pi) : \pi \in \Pi \subset R\}$ converges weakly to a Gaussian process under the null hypothesis. Note that $\{Y_t : t \le T\}$ is an AR(1) process under the null. Although the test statistic is designed for the case where the innovations $\{U_t : t \le T\}$ are normally distributed, we would like the test statistic to have the same asymptotic null distribution for all innovations that are iid, mean zero, and have some small number of moments finite. To obtain such a result, we need, among other things, that the weak convergence result referred to above holds for a broad class of innovation distributions.

The index set Π is one-dimensional in this example and the function $m(W_t, \pi) = U_t Y_{t-1} 1(Y_{t-d} \leq \pi)$ is unbounded. The latter feature is relevant when verifying Assumption 3.

3.4.2. Test for Variable Relevance (cont.)

In this example, $\ell_T(\theta, \pi)$ is the pseudo-likelihood function under the assumption of iid normal errors:

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

$$\frac{1}{\sqrt{T}} D\ell_T(\theta_0, \pi) = \begin{pmatrix} \frac{1}{\delta_{20}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t h(Z_t, \pi) \\ \frac{1}{\delta_{20}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t \frac{\partial}{\partial \delta_1} g(X_t, \delta_{10}) \\ \frac{1}{2\delta_{20}} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} (U_t^2 - 1) \end{pmatrix}. \tag{3.13}$$

To verify Assumption 3, we need to show that $\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t h(Z_t, \pi) : \pi \in \Pi \subset R^v\}$ converges weakly to a Gaussian process. As above, we would like this result to hold for a wide variety of mean zero variance δ_{20} error distributions not just the normal. The function $m(W_t, \pi) = U_t h(Z_t, \pi)$ is unbounded in this case. It is typically differentiable in both π and W_t.

3.4.3. Test for Autocorrelation (cont.)

For the case of testing for autocorrelation against ARMA(1, 1) errors, we have

$$\begin{aligned} \ell_T(\theta, \pi) &= -\frac{T}{2}\log \delta_2 - \frac{1}{2\delta_2}\sum_{t=1}^{T} \varepsilon_t(\theta, \pi)^2 , \quad \text{where} \\ \varepsilon_t(\theta,\pi) &= Y_t - X_t'\delta_1 - (\pi+\beta)(Y_{t-1} - X_{t-1}'\delta_1) + \pi\varepsilon_{t-1}(\theta,\pi) \\ &\quad \text{for } t = 1, ..., T, \text{ and } \varepsilon_0(\theta,\pi) = 0 , \text{ and} \end{aligned} \tag{3.14}$$

$$\frac{1}{\sqrt{T}} D\ell_T(\theta_0, \pi) = \begin{pmatrix} \frac{1}{\delta_{20}} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t (\sum_{s=0}^{t-2} \pi^s U_{t-s-1}) \\ \frac{1}{\delta_{20}} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t (X_t - \pi X_{t-1}) \\ \frac{1}{2\delta_{20}} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (U_t^2/\delta_{20} - 1) \end{pmatrix} .$$

(As above, $\ell_T(\theta, \pi)$ is the log likelihood function conditional on the first observation (Y_0, X_0).) To verify Assumption 3 in this example, we need to verify that $\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t (\sum_{s=0}^{t-2} \pi^s U_{t-s-1}) : \pi \in \Pi \subset (-1, 1)\}$ converges weakly to a Gaussian process under the null hypothesis. As in the previous examples, for robustness of the test results, we would like this convergence to hold not just for iid normal random variables $\{U_t : t \le T\}$, but for all iid sequences of mean zero random variables with several moments finite. Note that the required functional CLT is not of the standard empirical process form, because the stochastic process under consideration is not of the form $\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (m(W_t, \pi) - Em(W_t, \pi)) : \pi \in \Pi\}$.

3.4.4. Test for Conditional Heteroskedasticity (cont.)

In the example of testing for GARCH(1, 1) conditional heteroskedasticity, we have

(3.15)
$$\begin{aligned} &\ell_T(\theta, \pi) = \sum_{t=1}^{T} \ell_t(\theta, \pi), \quad \text{where} \quad \ell_t(\theta, \pi) = -\frac{1}{2}\log h_t(\theta, \pi) - \frac{1}{2} U_t(\theta)^2 / h_t(\theta, \pi) , \\ &h_t(\theta, \pi) = \delta_2(1 - \pi) + \beta U_{t-1}(\theta)^2 + \pi h_{t-1}(\theta, \pi) , \quad U_t(\theta) = Y_t - X_t'\delta_1 , \\ &h_0(\theta, \pi) = \text{const}, \quad \text{and} \quad U_0(\theta) = 0 \quad \text{for} \quad t = 1, ..., T . \end{aligned}$$

In addition, we have

(3.16)
$$\frac{1}{\sqrt{T}} \ell_T(\theta_0, \pi) = \begin{pmatrix} \frac{1}{2\delta_{20}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (U_t^2/\delta_{20} - 1) \sum_{s=0}^{t-2} \pi^s U_{t-s-1}^2 \\ \frac{1}{\delta_{20}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t X_t \\ \frac{1}{2\delta_{20}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1 - \pi^t)(U_t^2/\delta_{20} - 1) \end{pmatrix} .$$

The verification of Assumption 3 in this example is quite similar to that in the previous example. The main difference is that the innovations to the AR(1) process $Z_t(\pi) = \sum_{s=0}^{t-2} \pi^s U_{t-s-1}^2$ are squared errors rather than the errors themselves.

4. STOCHASTIC EQUICONTINUITY VIA BRACKETING

This section discusses the establishment of stochastic equicontinuity using bracketing conditions. Pioneering work on this method was done by Dudley (1978). Here we discuss more recent results that are available in the literature with emphasis on results for dependent random variables. We discuss the bracketing conditions themselves, how they can be verified in applications, and their restrictiveness in applications. We do not discuss the methods of proof. These can be found in the references listed below.

4.1. Bracketing Results for Dependent Random Variables

Let $\{W_{Tt} : t \le T, T \ge 1\}$ denote a triangular array of $\mathcal{W}$-valued rv's for $\mathcal{W} \subset R^k$. Consider $\nu_T(\cdot)$ of the form:

(4.1) $$\nu_T(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(m(W_{Tt}, \pi) - Em(W_{Tt}, \pi)) , \quad m(\cdot,\pi) : R^k \to R^s .$$

Define the pseudometric ρ on Π to be

(4.2) $$\rho(\pi_1, \pi_2) = \sup_{t \le T, T \ge 1} \left(E\|m(W_{Tt}, \pi_1) - m(W_{Tt}, \pi_2)\|^2\right)^{1/2} .$$

Stochastic equicontinuity results depend on a measure of the size/complexity of the set of functions

(4.3) $$\mathcal{M} = \{m(\cdot, \pi) : \pi \in \Pi\} .$$

L^p-bracketing cover numbers are one such measure.

In the following definition, let $\mathcal{M}$ be a class of real functions on $\mathcal{W}$.

DEFINITION: For $\varepsilon > 0$ and $p \in [2, \infty]$, the *L^p-bracketing cover number* $N_p^B(\varepsilon, \mathcal{M})$ is the smallest value of n for which there exist real functions $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ on $\mathcal{W}$ such that $\forall m \in \mathcal{M}$ $|m(w) - a_j(w)| \leq b_j(w)$ $\forall w \in \mathcal{W}$ for some $j \leq n$ and

$$\max_{j \leq n} \sup_{t \leq T, T \geq 1} \left(Eb_j^p(W_{Tt})\right)^{1/p} \leq \varepsilon \ .$$

For $p = \infty$, the latter condition is replaced by $\max_{j \leq n} \sup_{w \in \mathcal{W}} b_j(w) \leq \varepsilon$.

The log of $N_p^B(\varepsilon, \mathcal{M})$ is called the *L^p-bracketing ε-entropy* of $\mathcal{M}$. The pairs of functions referred to in the literature as lower and upper bracketing functions are $(\ell_1, u_1) = (a_1 - b_1, a_1 + b_1), \ldots, (\ell_n, u_n) = (a_n - b_n, a_n + b_n)$.

For a class $\mathcal{M}$ of R^s-valued functions, let

$$N_p^B(\varepsilon, \mathcal{M}) = \max_{i \leq s} N_p^B(\varepsilon, \mathcal{M}_i) \ , \tag{4.4}$$

where $\mathcal{M}_i = \{m_i(\cdot, \pi) : \pi \in \Pi\}$ and $m(\cdot, \pi) = (m_1(\cdot, \pi)', \ldots, m_s(\cdot, \pi)')'$. That is, for vector-valued functions, the bracketing cover numbers equal the maximum of the cover numbers for different elements of the functions.

A real function $\bar{M}$ on $\mathcal{W}$ is an *envelope* of $\mathcal{M}$ if

$$\max_{i \leq s} \sup_{\pi \in \Pi} |m_i(w, \pi)| \leq \bar{M}(w) \quad \forall w \in \mathcal{W} \, . \tag{4.5}$$

Stochastic equicontinuity results via bracketing typically require conditions on (i) the temporal dependence of $\{W_{Tt} : t \leq T, T \geq 1\}$, (ii) the moments of the envelope $\{\bar{M}(W_{Tt}) : t \leq T, T \geq 1\}$, and (iii) the rate of increase of the L^2 (or L^1) bracketing cover numbers as $\varepsilon \downarrow 0$. We now describe a number of bracketing stochastic equicontinuity results that are available in the literature for dependent rv's. In each case the given conditions (i)-(iii) are sufficient for $\{\nu_T(\cdot) : T \geq 1\}$ to be stochastically equicontinuous and for Π to be totally bounded under ρ.

Below, C denotes a generic constant. (It is not meant to be equal in any two places it appears.)

1. Independence or M-dependence (Ossiander (1987) and Pollard (1989)):

   (i) $\{W_{Tt}\}$ are M-dependent for some $M < \infty$,

   (ii) $\overline{\lim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\bar{M}^{2+\delta}(W_{Tt}) < \infty$ for some $\delta > 0$,

   (iii) $\log N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$ for some $B < 1/2$, $\forall \varepsilon \in (0, 1]$.

2. Bounded functions with mixing conditions (Philipp (1982), Massart (1988), Andrews and Pollard (1992), Doukhan, Massart, and Rio (1992)):

   (a) Philipp (1982):

   (i) $\{W_t : t \geq 1\}$ is a stationary strong mixing sequence with strong mixing numbers $\alpha(s)$ that satisfy $\alpha(s) \leq Cs^{-A}$,

   (ii) $\mathcal{M}$ contains indicator functions of sets and, in consequence, $\bar{M} = 1$,

   (iii) $N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$,

   where A and B are positive constants that must satisfy certain constraints (and are positively related, i.e., a larger B requires a larger A).

   (b) Massart (1988) and Andrews and Pollard (1992):

   (i) $\{W_{Tt} : t \leq T, T \geq 1\}$ is a strong mixing triangular array with $\alpha(s) \leq Cs^{-A}$,

   (ii) $\bar{M} = 1$,

   (iii) $N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$,

   where A and B are positive constants that must satisfy certain constraints.

   (c) Massart (1988):

   (i) $\{W_t : t \geq 1\}$ is a stationary strong mixing sequence with $\alpha(s) \leq C\theta^s$ for some $\theta \in (0, 1)$,

   (ii) $\bar{M} = 1$,

   (iii) $\log N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$ for some $0 < B < 1/4$.

   (d) Doukhan, Massart, and Rio (1992):

   (i) $\{W_t : t \geq 1\}$ is a stationary absolutely regular (i.e., β-mixing) sequence with $\beta(s) \leq Cs^{-A}$,

   (ii) $\bar{M} = 1$,

   (iii) $\log N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$, where A and B are positive constants that satisfy $B < 1/2$ and $A > 1/(1-2B)$.

3. Unbounded functions with mixing conditions (Doukhan, Massart, and Rio (1992), Andrews and Pollard (1991)):

   (a) Doukhan, Massart, and Rio (1992):

      (i) $\{W_t : t \geq 1\}$ is a stationary absolutely regular sequence with $\beta(s) \leq Cs^{-A}$ for some $A > p/(p-2)$ and some $p > 2$,

      (ii) $E\bar{M}^p(W_t) < \infty$ for p as in part (i),

      (iii) $\log N_p^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$ for some $B < 1/2$ and for p as in part (i).

      In this case, the pseudometric ρ on Π is the L^p-pseudometric rather than the L^2-pseudometric given in (4.2).

   (b) Doukhan, Massart, and Rio (1992):

      (i) $\{W_t : t \geq 1\}$ is a stationary absolutely regular sequence with $\beta(s) \leq C\theta^s$ for some $\theta \in (0, 1)$,

      (ii) $E\bar{M}^2(W_t)\log \bar{M}(W_t) < \infty$,

      (iii) $\log N_p^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$ for some $B < 1/2$ and some $p > 2$.

      In this case, the pseudometric ρ on Π is the L^p-pseudometric rather than the L^2-pseudometric given in (4.2).

   (c) Andrews and Pollard (1991):

      (i) $\{W_{Tt} : t \leq T, T \geq 1\}$ is a ϕ-mixing triangular array with $\phi(s) \leq C\theta^s$ for some $\theta \in (0, 1)$,

      (ii) $\overline{\lim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\bar{M}^{2+\delta}(W_{Tt}) < \infty$,

      (iii) $\log N_2^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^B$,

      where (θ, δ, B) must satisfy certain constraints.

See Bradley (1986) and Doukhan (1992) for surveys of results concerning the mixing concepts referred to above. In particular, they discuss a variety of examples of processes that are known to satisfy various mixing conditions.

4.2. Verification of Bracketing Conditions

To apply any of the stochastic equicontinuity results listed in Section 4.1, one needs to be able to bound the cover numbers of $\mathcal{M} = \{m(\cdot, \pi) : \pi \in \Pi\}$. When $\mathcal{M}$ is a finite dimensional class of functions, one of the most useful ways of bounding its cover numbers is via an L^p-continuity condition.

L^p-Continuity Condition: If Π is a bounded subset of R^d and $\forall \pi \in \Pi$

$$\sup_{t \le T, T \ge 1} \Big(E \sup_{\substack{\pi_1 \in \Pi: \\ |\pi_1 - \pi| < \delta}} |m(W_{Tt}, \pi_1) - m(W_{Tt}, \pi_2)|^p\Big)^{1/p} \le C\delta^{\psi} \tag{4.6}$$

$\forall \delta > 0$ small, for some positive constants ψ and C (that do not depend on π), then

$$N_2^B(\varepsilon, \mathcal{M}) \le C^*(1/\varepsilon)^{d/\psi} . \tag{4.7}$$

To prove this result, consider disjoint cubes in Π of diameter $\delta = (\varepsilon/C)^{1/\psi}$. The number $N(\varepsilon)$ of such cubes satisfies $N(\varepsilon) \le C^* \varepsilon^{-d/\psi}$ for some $C^* < \infty$. Let π_j be some element of the j-th cube in Π. For $\mathcal{M}_i = \{m_i(\cdot, \pi) : \pi \in \Pi\}$ (where $m(\cdot, \pi) = (m_1(\cdot, \pi)', ..., m_s(\cdot, \pi)')'$), let $a_j(\cdot) = m_i(\cdot, \pi_j)$ and $b_j(\cdot) = \sup_{\pi \in \Pi: |\pi - \pi_j| < \delta} |m_i(\cdot, \pi) - m_i(\cdot, \pi_j)|$. By (4.6), $\sup_{t \le T, T \ge 1} (Eb_j^p(W_{Tt}))^{1/p} \le C\delta^{\psi} = \varepsilon$. Thus, $N_p^B(\varepsilon, \mathcal{M}_i) \le N(\varepsilon)$. This holds $\forall i = 1, ..., s$, which yields (4.7).

We now use the L^p-continuity condition in the examples of Section 3.

1. Test of a threshold effect: In this example, $m(W_t, \pi) = U_t Y_{t-1} 1(Y_{t-d} \le \pi)$ and we have: $\forall \pi \in \Pi$,

$$\begin{aligned} & E \sup_{\pi_1 \in \Pi: |\pi_1 - \pi| < \delta} |U_t Y_{t-1} 1(Y_{t-d} \le \pi) - U_t Y_{t-1} 1(Y_{t-d} \le \pi)|^p \\ & = E|U_t Y_{t-1}|^p 1(Y_{t-d} \in (\pi - \delta, \pi + \delta)) \\ & \le C_1 P(Y_{t-d} \in (\pi - \delta, \pi + \delta)) \\ & \le C\delta \end{aligned} \tag{4.8}$$

provided Π ($\subset R$) is bounded, $\sup_{t \ge 1} E|U_t|^p < \infty$, U_t and $(Y_{t-1}, ..., Y_{t-d})$ are independent and

Y_{t-d} has bounded density with respect to Lebesgue measure. Thus, under these conditions,

$$N_p^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^p \ . \tag{4.9}$$

In the leading case, $U_t \sim$ iid $N(0, \sigma^2)$ and all of the above conditions are satisfied. Alternatively, if $U_t \sim$ iid$(0, \sigma^2)$ with bounded density, then the above conditions hold. (Note that with a more complicated argument, it would be possible to weaken the bounded density assumption and still obtain usable bounds on the bracketing cover numbers.)

Since the innovations U_t typically are unbounded rv's, the only bracketing results of Section 4.1 that are applicable are those of Doukhan *et al.* (1992) and Andrews and Pollard (1991). The latter results impose ϕ-mixing, which is known to be violated by AR(1) processes with normal innovations (see Ibragimov and Linnik (1971)), and hence are not applicable. The bracketing results of Doukhan *et al.* (1992) impose the dependence condition of absolute regularity. As shown by Mokkadem (1986, 1987), AR(1) processes are absolutely regular with $\beta(s) \leq C\theta^s$ for some $\theta \in (0, 1)$ provided the innovations have a density with respect to Lebesgue measure and $E|U_t|^r < \infty$ for some $r > 0$. Thus, we can verify stochastic equicontinuity in this example provided the innovations $\{U_t : t \leq 1\}$ are iid with mean zero, variance σ^2, and bounded density with respect to Lebesgue measure.

An alternative method of verifying stochastic equicontinuity in this example is to use a VC/symmetrization approach rather than a bracketing approach. Stochastic equicontinuity and functional CLTs are now available for absolutely regular processes using the VC/symmetrization approach, see Yu (1990) and Arcones and Yu (1991). With this approach, one also has to verify a cover number condition, but the cover numbers are defined differently. It is possible to verify the requisite VC-type cover number condition without the requirement that the innovations have a density. To verify that the AR(1) process $\{Y_t : t \geq 1\}$ is absolutely regular, however, one still needs the assumption of a density, though it need not be bounded. Both the VC/symmetrization approach and the bracketing approach have the disadvantage in this example that they require the (probably superfluous) assumption that the innovations have a density with respect to Lebesgue measure. (The latter condition would be unnecessary for a functional CLT if $\{Y_{t-1}\ t \geq 1\}$ was an iid sequence and it is unnecessary for a CLT for any fixed π by using a CLT for near epoch dependent rv's, as in McLeish (1975), or a CLT for linear processes, as in Phillips and Solo (1992).)

Next, we note that the L^p-continuity condition is particularly easy to verify if $m(w, \pi)$ is differentiable in π. In this case, it holds with $\psi = 1$ (by element by element mean-value expansions) provided

(4.10) $$\sup_{t \le T, T \ge 1} E \sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \pi'} m(W_{Tt}, \pi) \right\|^p < \infty \, .$$

This condition can be applied in the following example.

2. Test of Variable Relevance: In this example, $m(W_t, \pi) = U_t \, h(Z_t, \pi)$. The L^p-continuity condition holds with $\psi = 1$ provided

(4.11) $$\sup_{t \ge 1} E \sup_{\pi \in \Pi} \left\| U_t \frac{\partial}{\partial \pi} h(Z_t, \pi) \right\|^p < \infty \, .$$

In this case,

(4.12) $$N_p^B(\varepsilon, \mathcal{M}) \le C(1/\varepsilon) \, .$$

As in the previous example, the errors $\{U_t : t \ge 1\}$ are usually unbounded, so the only bracketing results that are applicable are those of Doukhan *et al.* (1992) for absolutely regular processes or those of Andrews and Pollard (1991) for ϕ-mixing processes. If one assumes that $\{(U_t, X_t, Z_t) : t \ge 1\}$ is absolutely regular with β-mixing numbers that decline sufficiently quickly and U_t has sufficient moments, then the results of Doukhan *et al.* (1992) provide the desired functional CLT.

An alternative approach to establishing a functional CLT is to use the VC/symmetrization approach for absolutely regular processes mentioned above. One can exploit the differentiability of $m(W_t, \pi)$ in π to obtain sufficient bounds on the VC-type cover numbers (e.g., see the result for type II classes of functions in Andrews (1989b, Thm. 2)). This approach yields quite similar regularity conditions to those generated by the bracketing approach.

A third approach is to use the series expansion method considered in Andrews (1991a). This approach yields a functional CLT under a weaker dependence condition, viz., strong mixing, than the other two, but requires that Z_t is a bounded rv and $h(Z_t, \pi)$ is smooth in Z_t. It does not impose any conditions on how $h(Z_t, \pi)$ depends on π.

3. Test of Autocorrelation and Test of Conditional Heteroskedasticity: These two examples exhibit the feature that the summands of $\nu_T(\pi)$ are not of the empirical process form. That is, $m_t(\pi) = U_t(\sum_{s=0}^{t-2} \pi^s U_{t-s-1})$ cannot be written as $m(W_t, \pi)$ for some rv W_t that does not depend on π. In consequence, none of the bracketing results considered in Section 4.1 are applicable. In addition, none of the VC/symmetrization results of Yu (1990) and Arcones and Yu (1991) are applicable.

A more general form of bracketing, however, has been considered in Pollard (1989) and Andrews and Pollard (1991). This form of bracketing covers summands of the type that appear in these two examples. In addition, their bracketing conditions can be verified using the differentiability of $m_t(\pi)$ in π. Unfortunately, there is still a problem with using these results. In their current state, these results require ϕ-mixing rv's. As mentioned above, AR(1) processes with normal innovations are not ϕ-mixing. This suggests that it would be useful to extend the Andrews and Pollard (1991) bracketing results to more general mixing processes.

There is an alternative approach that can be used successfully to establish a functional CLT for these examples. This approach applies to summands $m_t(\pi)$ that are given by the t-th observation of a linear process (multiplied by some other rv) indexed by its moving average coefficients (i.e., π is an infinite dimensional vector of moving average coefficients). The proof of stochastic equicontinuity for such summands is similar to that used in the series expansion approach of Andrews (1991a). (Results of the above type are under preparation by the author.)

Last, we mention two additional examples where a nuisance parameter is present only under the alternative. The first is a test of regime switching with unobserved regimes, e.g., the Markov switching model of Hamilton (1989), see Hansen (1991b). The second is a test of functional form in a regression model against nonparametric alternatives, see Hansen (1992). In each case, functional limit theory is needed to obtain the limit distribution of a suitable test statistic.

5. 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 (1989a, b, 1990, 1991b) and Whang and Andrews (1992). Other uses of functional limit theory/stochastic equicontinuity in the semiparametric and nonparametric econometrics literature include Horowitz (1987, 1992), Newey (1989), Wooldridge (1989), Kim and Pollard (1990), Klecan, McFadden, and McFadden (1990), White and Stinchcombe (1990), Yatchew (1990), and Pakes and Olley (1991) among others.

5.1. A General Approach

Consider an estimator of a finite dimensional parameter $\theta \in \Theta \subset R^p$ that is based on a criterion function that depends on a preliminary estimator $\hat{\pi}$ of an infinite-dimensional nuisance parameter π_0. Typically, $\hat{\pi}$ is a nonparametric estimator of one or more nonparametric regression or density functions or their derivatives. Let the data consist of $\{W_t : t \leq T\}$. Consider a system of p estimating equations

$$\bar{m}_T(\theta, \hat{\pi}) = \frac{1}{T}\sum_{t=1}^{T} m(W_t, \theta, \hat{\pi}) , \tag{5.1}$$

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

$$\sqrt{T}\,\bar{m}_T(\hat{\theta}, \hat{\pi}) = 0 \tag{5.2}$$

(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 (1989a) for sufficient conditions). We wish to determine the asymptotic distribution of $\hat{\theta}$. When $m(W_t, \theta, \pi)$ is a smooth function of θ, the following approach can be used. (When $m(W_t, \theta, \pi)$ is not a smooth function of θ, but $Em(W_t, \theta, \pi)$ is, an alternative approach described in Andrews (1989a) can be used.) Element by element mean value expansions stacked yield

$$(5.3) \qquad o_p(1) = \bar{m}_T(\hat{\theta}, \hat{\pi}) = \sqrt{T}\bar{m}_T(\theta_0, \hat{\pi}) + \frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\pi})\sqrt{T}(\hat{\theta} - \theta_0) ,$$

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

$$(5.4) \qquad \frac{\partial}{\partial\theta}\bar{m}_T(\theta^*, \hat{\pi}) \xrightarrow{p} M = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\frac{\partial}{\partial\theta'}\bar{m}_T(\theta_0, \pi_0)$$

(e.g., see Andrews (1989a)). Thus,

$$(5.5) \qquad \begin{aligned} \sqrt{T}(\hat{\theta} - \theta_0) &= -(M^{-1} + o_p(1))\sqrt{T}\bar{m}_T(\theta_0, \hat{\pi}) \\ &= -(M^{-1} + o_p(1))\left[\sqrt{T}(\bar{m}_T(\theta_0, \hat{\pi}) - \bar{m}_T(\theta_0, \hat{\pi})) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\pi})\right] , \end{aligned}$$

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

Again under suitable conditions, either

$$(5.6) \qquad \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\pi}) \xrightarrow{p} 0 \quad \text{or} \quad \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\pi}) \xrightarrow{d} N(0, A) ,$$

for some covariance matrix A, see Andrews (1989a, 1991b).

Let

$$(5.7) \qquad \nu_T(\pi) = \sqrt{T}(\bar{m}_T(\theta_0, \pi) - \bar{m}_T^*(\theta_0, \pi)) .$$

Note that $\nu_T(\cdot)$ is a stochastic process indexed by an infinite dimensional parameter in this case. This differs from the examples of Section 3 for which π is finite dimensional.

Under standard conditions, one can establish that

$$(5.8) \qquad \nu_T(\pi_0) \xrightarrow{d} N(0, S)$$

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

$$(5.9) \qquad \nu_T(\hat{\pi}) - \nu_T(\pi_0) \xrightarrow{p} 0 ,$$

then we obtain

$$(5.10)\quad \begin{aligned}\sqrt{T}(\hat{\theta} - \theta_0) &= -(M^{-1} + o_p(1))[\nu_T(\hat{\pi}) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\pi})] \\ &= -M^{-1}[\nu_T(\pi_0) + \sqrt{T}\bar{m}_T^*(\theta_0, \hat{\pi})] + o_p(1) \\ &\xrightarrow{d} N(0, M^{-1}(S+A)(M^{-1})') ,\end{aligned}$$

which is the desired result.

To prove (5.9), we can use the stochastic equicontinuity property. Suppose

$$(5.11)\quad \begin{aligned}&(i)\ \left\{\nu_T(\cdot) : T \geq 1\right\} \text{ is stochastically equicontinuous (at } \pi_0) \\ &\qquad \text{for some choice of } \Pi \text{ and pseudometric } \rho \text{ on } \Pi, \\ &(ii)\ P(\hat{\pi} \in \Pi) \to 1, \text{ and} \\ &(iii)\ \rho(\hat{\pi}, \pi_0) \xrightarrow{p} 0 ,\end{aligned}$$

then (5.9) holds.

Note that there exist tradeoffs between conditions (i), (ii), and (iii) of (5.11) in terms of the difficulty of verification and the strength of the regularity conditions needed. For example, a larger set Π makes it more difficult to verify (i), but easier to verify (ii). A stronger metric ρ makes it easier to verify (i), but more difficult to verify (iii).

Since the sufficiency of (5.11) for (5.9) is the key to the approach considered here, we provide a proof of this simple result. We have: $\forall\varepsilon > 0$, $\forall\eta > 0$, $\exists\delta > 0$ such that

$$(5.12)\quad \begin{aligned}&\overline{\lim_{T\to\infty}}\, P(|\nu_T(\hat{\pi}) - \nu_T(\pi_0)| > \eta) \\ &\leq \overline{\lim_{T\to\infty}}\, P(|\nu_T(\hat{\pi}) - \nu_T(\pi_0)| > \eta,\ \hat{\pi} \in \Pi,\ \rho(\hat{\pi}, \pi_0) \leq \delta) \\ &\quad + \overline{\lim_{T\to\infty}}\, P(\hat{\pi} \notin \Pi \text{ or } \rho(\hat{\pi}, \pi_0) > \delta) \\ &\leq \overline{\lim_{T\to\infty}}\, P^*\big(\sup_{\pi\in\Pi:\rho(\pi,\pi_0)<\delta} |\nu_T(\pi) - \nu_T(\pi_0)| > \eta\big) \\ &< \varepsilon ,\end{aligned}$$

where the term on the third line of (5.12) is zero by (ii) and (iii) and the last inequality holds by (i). Since $\varepsilon > 0$ is arbitrary, (5.9) 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 $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous. In the next subsection, we consider the application of this approach to a particular example and illustrate the form of $\nu_T(\cdot)$ in this example. In Section 6, we discuss the verification of stochastic equicontinuity when $\mathcal{M} = \{m(\cdot, \pi) : \pi \in \Pi\}$ is an infinite dimensional class of functions.

5.2. Example

In this subsection we consider the estimation of a time series *nonlinear regression model with conditional heteroskedasticity of unknown form*. The model is

$$(5.13)\qquad \begin{aligned} &Y_t = g(Z_t, \theta_0) + U_t\,, \quad E(U_t|Z_t, \mathcal{F}_{t-1}) = 0 \text{ a.s.},\\ &E(U_t^2|Z_t, \mathcal{F}_{t-1}) = E(U_t^2|X_t) = \pi_0(X_t)\,, \quad \text{and } E(U_sU_t|Z_t, \mathcal{F}_t) = 0 \text{ a.s.,}\end{aligned}$$

where $\mathcal{F}_t$ is the σ-field generated by $(\ldots, (Y_{t-1}, Z_{t-1}, X_{t-1}, U_{t-1}), (Y_t, Z_t, X_t, U_t))$ and Z_t and/or X_t include lagged values of Y_t. Estimation of this model when the data are iid (so no lagged values of Y_t appear in Z_t or X_t) has been considered by Robinson (1987) and Newey (1987). Andrews (1989a) and White and Stinchcombe (1990) consider the model more or less as above.

We consider a weighted nonlinear least squares (LS) estimator of θ_0 with nonparametrically estimated weights. First, one estimates θ_0 by an unweighted nonlinear least squares regression of Y_t on $g(Z_t, \theta)$; call the estimator $\hat{\theta}_1$. Denote the residuals by $\hat{U}_t = Y_t - g(Z_t, \hat{\theta}_1)$. Next, one does a nonparametric regression of $\hat{U}_t^2$ (or $\hat{U}_t^2 - \frac{1}{T}\sum_{s=1}^{T}\hat{U}_s^2$) on X_t to get a preliminary estimator $\hat{\pi}(\cdot)$. For example, a kernel estimator can be used. One only needs the values of $\hat{\pi}(\cdot)$ evaluated at $X_1, \ldots, X_T$. Last, the weighted nonlinear LS estimator $\hat{\theta}$ is defined to minimize

$$(5.14)\qquad \frac{1}{T}\sum_{t=1}^{T}\xi(X_t)(Y_t - g(Z_t, \theta))^2/\hat{\pi}(X_t) \quad \text{over} \quad \theta \in \Theta \subset R^p\,,$$

where $\xi(X_t)$ is a trimming function such as $\xi(X_t) = 1(X_t \in \mathcal{X}^*)$, $\xi(X_t) = 1(X_t \in \mathcal{X}_T^*)$, $\xi(X_t) = 1(\hat{f}(X_t) \geq \varepsilon)$, or $\xi(X_t) = 1(\hat{f}(X_t) \geq \varepsilon_T,\ \hat{\pi}(X_t) \geq \varepsilon_T^*)$ where $\hat{f}(\cdot)$ is an estimator of the density of X_t.

The first-order conditions for $\hat{\theta}$ are given by

$$0 = \bar{m}_T(\hat{\theta}, \hat{\pi}) = \frac{1}{T}\sum_{t=1}^{T} \xi(X_t)(Y_t - g(Z_t, \hat{\theta}))\frac{\partial}{\partial\theta}g(Z_t, \hat{\theta})/\hat{\pi}(X_t) \ . \tag{5.15}$$

This yields

$$\nu_T(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi(X_t)U_t \frac{\partial}{\partial\theta}g(Z_t, \theta_0)/\pi(X_t) \ , \tag{5.16}$$

where $\pi(\cdot)$ is some potential realization of the nonparametric estimator $\hat{\pi}(\cdot)$.

The simplest choice of $\xi(X_t)$, Π, and ρ, although it is somewhat restrictive, is as follows: Take $\xi(X_t)$ to be zero outside some bounded set $\mathcal{X}^* \subset R^{k_*}$. Define Π to contain functions from R^{k_*} to R that are smooth on $\mathcal{X}^*$ and bounded away from zero on $\mathcal{X}^*$. Let ρ be the L^2 metric defined in (4.2) with $m(W_t, \pi) = \xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)/\pi(X_t)$. We choose Π to contain smooth functions in order to facilitate the verification of stochastic equicontinuity, see Section 6.

As a measure of smoothness, we can use the Sobolev norm. Let $\mu = (\mu_1, \ldots, \mu_k)'$ be a k-vector of non-negative integers. Let $|\mu| = \mu_1 + \cdots + \mu_k$. For any function $c(x)$ on R^k, let

$$D^{\mu}c(x) = \frac{\partial^{|\mu|}}{\partial x_1^{\mu_1} \cdots \partial x_k^{\mu_k}}c(x) \quad \text{and} \tag{5.17}$$

$$\|c(\cdot)\|_{q,\mathcal{X}^*} = \sum_{|\mu| \le q} |D^{\mu}c(x)| \ .$$

$\|\cdot\|_{q,\mathcal{X}^*}$ is the *supremum Sobolev norm* over $\mathcal{X}^*$ of integer order q.

Now, for the simplest case referred to above, Π can be defined more precisely to be

$$\Pi = \left\{\pi : R^{k_*} \to R^s \,|\, \|\pi(\cdot) - \pi_0(\cdot)\|_{q,\mathcal{X}^*} \le C, \inf_{x \in \mathcal{X}^*} \pi(x) \ge \varepsilon\right\} \tag{5.18}$$

for some $C > 0$ and $\varepsilon > 0$, where q is chosen so that stochastic equicontinuity can be verified. In addition, we assume that $\|\pi_0(\cdot)\|_{q,\mathcal{X}^*} < \infty$. Now, condition (ii) above, viz., $P(\hat{\pi} \in \Pi) \to 1$, holds if $\hat{\pi}$ satisfies $\|\hat{\pi}(\cdot) - \pi_0(\cdot)\|_{q,\mathcal{X}^*} \xrightarrow{p} 0$. Note that this condition is restrictive in that it is not required in the papers by Robinson (1987) and Newey (1987).

6. STOCHASTIC EQUICONTINUITY VIA BRACKETING FOR INFINITE-DIMENSIONAL INDEX SETS II

In this section, we discuss the verification of the bracketing conditions of Section 4.1 when the index set Π is infinite-dimensional. Results of the type considered here are needed to apply the bracketing stochastic equicontinuity results of Section 4.1 to semiparametric and nonparametric problems, such as those of Section 5.

6.1. Bracketing Cover Number Bounds for Classes of Smooth Functions

A basic bracketing bound for infinite-dimensional classes of smooth functions is the following. Partition W_{Tt} as $(W_{Tta}', W_{Ttb}')'$, where $W_{Tta} \in R^{k_a}$. (In applications, W_{Tta} is the part of W_{Tt} that enters as the argument of a function that lies in an infinite-dimensional class. For example, in the semiparametric example of Section 5.2, W_{Tta} equals X_t, the argument of $\hat{\pi}(\cdot)$.) Let $\mathcal{W}_a \subset R^{k_a}$ include the supports of $\{W_{Tta} : t \leq T, T \geq 1\}$. If

$$\text{(6.1)} \quad \begin{array}{l} (i)\ m(w,\pi) \text{ depends on } w = (w_a', w_b')' \text{only through the subvector} \\ \quad w_a \text{ of dimension } k_a \leq k\ , \\ (ii)\ \sup_{\pi \in \Pi} \|m(\cdot,\pi)\|_{q,\mathcal{W}_a^*} < \infty \text{ for some integer } q > 0, \text{ where } \mathcal{W}_a^* \\ \quad \text{is some convex compact subset of } \mathcal{W}_a\ , \text{ and} \\ (iii)\ \text{for some constant } K \text{ and all } \pi \in \Pi\ ,\ m(w,\pi) = K \quad \forall w \in \mathcal{W} \\ \quad \text{such that } w_a \in \mathcal{W}_a \backslash \mathcal{W}_a^*\ , \end{array}$$

then

$$\text{(6.2)} \qquad \log N_p^B(\varepsilon, \mathcal{M}) \leq C(1/\varepsilon)^{k_a/q} \quad \forall p \in [2, \infty] \quad \forall \varepsilon \in (0, 1]\ .$$

In short, condition (6.1) requires the functions $m(w, \pi)$ in $\mathcal{M}$ to be sufficiently smooth when w_a is in a convex compact set $\mathcal{W}_a^*$ and to be constant otherwise. The proof of this result follows from Kolmogorov and Tihomirov (1961, Thm. XIV).

Note that the bound in (6.2) is on the *log* cover numbers, which is a much weaker bound than the bounds in Section 4.2 on the cover numbers themselves. The reason, of course, is that $\mathcal{M}$ is an infinite-dimensional, rather than a finite-dimensional, class of functions under

(6.1). Also note that the functions in $\mathcal{M}$ are bounded under (6.1), so their envelope $\bar{M}$ is a constant.

Given the cover number bound in (6.2) and the fact that $\bar{M}$ is a constant, we can apply the stochastic equicontinuity results of 2(c) or 2(d) of Section 4.2. In particular, the strong mixing results of Massart (1988) yield stochastic equicontinuity provided $q > k_a/4$ and $\alpha(s) \leq C\theta^s$ for some $\theta \in (0, 1)$. The absolutely regular results of Doukhan *et al.* (1992) yield stochastic equicontinuity provided $q > k_a/2$ and $\beta(s) \leq Cs^{-A}$ for some $A > q/(q - 2k_a)$.

Consider, now, the semiparametric conditional heteroskedasticity example of Section 5.2. In this example, $m(W_t, \pi) = \xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)/\pi(X_t)$, $W_t = (X_t', Z_t', U_t)'$, and $W_{ta} = X_t$. One cannot apply the results above directly to this example, because $m(W_t, \pi)$ is not constant for X_t outside a convex compact set $\mathcal{W}_a^*$. Instead, we proceed by obtaining bounds on the component parts of $m(W_t, \pi)$ and use results given in Section 6.2 below to manipulate these bounds to yield bounds on the functions of interest. In particular, suppose $\xi(X_t) = 1(X_t \in \mathcal{X}^*)$, where $\mathcal{X}^*$ is a convex compact subset of R^{k_a}, and write

$$\begin{aligned} m(W_t, \pi) &= m_1(W_t, \pi)/m_2(W_t, \pi) \quad \text{for} \\ m_1(W_t, \pi) &= \xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0) \quad \text{and} \\ m_2(W_t, \pi) &= \begin{cases} \pi(X_t) & \text{if } X_t \in \mathcal{X}^* \\ K & \text{otherwise} \end{cases} \quad \text{for some } K \neq 0 . \end{aligned} \tag{6.3}$$

Let

$$\mathcal{M}_1 = \{m_1(\cdot, \pi) : \pi \in \Pi\} \quad \text{and} \quad \mathcal{M}_2 = \{m_2(\cdot, \pi) : \pi \in \Pi\} . \tag{6.4}$$

Obviously, $N_p^B(\varepsilon, \mathcal{M}_1) = 1 \ \ \forall p \in [2, \infty]$ and $\mathcal{M}_1$ has envelope $\bar{M}_1(W_t) = \|U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)\|$. In addition, by the definition of Π in (5.18), condition (ii) of (6.1) holds and

$$\log N_p^B(\varepsilon, \mathcal{M}_2) \leq C(1/\varepsilon)^{k_a/q} \quad \forall p \in [2, \infty], \forall \varepsilon \in (0, 1] , \tag{6.5}$$

by applying the result of (6.1)-(6.2), where q is as in (5.18). Since the functions in $\mathcal{M}_2$ are bounded, the envelope $\bar{M}_2$ can be taken to be a constant. To obtain bounds on the cover numbers of $\mathcal{M}$, it remains to show how one can manipulate the bounds on $\mathcal{M}_1$ and $\mathcal{M}_2$ to yield bounds on $\mathcal{M}$.

6.2. Manipulating Bracketing Bounds

In this subsection we review some results for manipulating bracketing bounds, see Andrews (1989b, Sec. 4 and Appendix) for more details and proofs. Let $\mathcal{G}$, $\mathcal{G}^*$, and $\mathcal{H}$ be classes of $r \times s$, $r \times s$, and $s \times u$ matrix-valued functions defined on $\mathcal{W} \subset R^k$ with scalar envelopes G, G^*, and H respectively. Let g, g^*, and h denote generic elements of $\mathcal{G}$, $\mathcal{G}^*$, and $\mathcal{H}$ respectively. Let $\mathcal{G} \oplus \mathcal{G}^* = \{g + g^*\}$ ($= \{g + g^* : g \in \mathcal{G}, g^* \in G^*\}$), $\mathcal{G} \vee \mathcal{G}^* = \{g \vee g^*\}$ (i.e., $g \vee g^*$ equals the matrix of element by element maximums of g and g^*), $\mathcal{G} \wedge \mathcal{G}^* = \{g \wedge g^*\}$, $|\mathcal{G}| = \{|g|\}$ (i.e., $|g|$ equals the matrix of element by element absolute values of g), $\mathcal{G}^{-1} = \{g^{-1}\}$ provided $r = s$ and $g(w)$ is nonsingular $\forall w \in \mathcal{W}$, and $\mathcal{G}\mathcal{H} = \{gh\}$.

Some calculations yield: $\forall p \in [2, \infty]$, $\forall \varepsilon \in (0, 1]$,

$$
\begin{aligned}
(6.6) \qquad & N_p^B(\varepsilon, \mathcal{G} \cup \mathcal{G}*) \le N_p^B(\varepsilon, \mathcal{G}) + N_p^B(\varepsilon, \mathcal{G}*) && \text{with envelope } G \vee G* \,, \\
& N_p^B(\varepsilon, \mathcal{G} \oplus \mathcal{G}*) \le N_p^B(\varepsilon/2, \mathcal{G}) N_p^B(\varepsilon/2, \mathcal{G}*) && \text{with envelope } G + G* \,, \\
& N_p^B(\varepsilon, \mathcal{G} \vee \mathcal{G}*) \le N_p^B(\varepsilon/2, \mathcal{G}) N_p^B(\varepsilon/2, \mathcal{G}*) && \text{with envelope } G \vee G* \,, \\
& N_p^B(\varepsilon, \mathcal{G} \wedge \mathcal{G}*) \le N_p^B(\varepsilon/2, \mathcal{G}) N_p^B(\varepsilon/2, \mathcal{G}*) && \text{with envelope } G \vee G* \,, \\
& N_p^B(\varepsilon, |\mathcal{G}|) \le N_p^B(\varepsilon, \mathcal{G}) && \text{with envelope } G \,, \\
& N_p^B(\varepsilon, \mathcal{G}^{-1}) \le N_p^B\left(\frac{\lambda_*^2}{2r^6}\varepsilon, \mathcal{G}\right)^{r^2} && \text{with envelope } r/\lambda_* \,, \\
& N_p^B(\varepsilon, \mathcal{G}\mathcal{H}) \le N_\lambda^B(D\varepsilon, \mathcal{G}) N_\mu^B(D\varepsilon, \mathcal{H}) && \text{with envelope } sGH \,,
\end{aligned}
$$

where the bound on $N_p^B(\varepsilon, \mathcal{G}^{-1})$ holds provided $r = s$ and $\lambda_* = \inf_{g \in \mathcal{G}} \inf_{w \in \mathcal{W}} \lambda_{\min}(g(w)) > 0$ and the bound on $N_p^B(\varepsilon, \mathcal{G}\mathcal{H})$ holds for $\lambda \in (p, \infty]$ and $\mu \in (p, \infty]$ provided $\lambda\mu/(\lambda+\mu) \ge p$ and $D = 1/\sup_{t \le T, T \ge 1} ((EG^\lambda)^{1/\lambda} + (EH^\mu)^{1/\mu})$ is well-defined.

Consider again the semiparametric example. The $\mathcal{G}^{-1}$ and $\mathcal{G}\mathcal{H}$ results can be applied to the classes $\mathcal{M}_2$ and $\mathcal{M}_1$ ($\mathcal{M}_2^{-1}$) of Section 6.1. Taking $\lambda > p$ and $\mu = \infty$, we obtain from (6.5) and (6.6) that

$$\log N^B_\infty(\varepsilon, \mathcal{M}_2^{-1}) \le C(1/\varepsilon)^{k_*/q} ,$$

(6.7) $$\log N^B_p(\varepsilon, \mathcal{M}) = \log N^B_p(\varepsilon, \mathcal{M}_1\mathcal{M}_2^{-1}) \le C(1/\varepsilon)^{k_*/q} \ \forall p \in [2, \infty] , \text{ and}$$

$$\bar{M}(W_t) = \left|\xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)\right|$$

provided $\sup_{t\ge1} E\|\xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)\|^\lambda < \infty$ for some $\lambda > p$, using the fact that $\lambda_* = \inf_{\pi\in\Pi}\inf_{x\in\mathcal{X}_*} \pi(x) > 0$ by the definition of Π in (5.18).

Given the bounds on the cover numbers in (6.7), one can apply the stochastic equicontinuity results 3(a), 3(b), or 3(c) of Section 4.1 to verify the key condition (i) of (5.11) for the semiparametric example. In particular, if one uses the result 3(b) of Doukhan *et al.* (1992) for stationary absolutely regular processes, then it suffices to have $q > k_*/2$ (where q is as in (5.18)), $\beta(s) \le Cs^{-A}$, and $E\|\xi(X_t)U_t\frac{\partial}{\partial\theta}g(Z_t, \theta_0)\|^\lambda < \infty$ for some $A > p/(p-2)$ and $\lambda > p > 2$. The pseudometric ρ that is used in this case and that appears in condition (ii) of (5.11) (i.e., $\rho(\hat{\pi}, \pi_0) \xrightarrow{p} 0$) in this case is the L^p-pseudometric for $p > 2$ as above.

6.3. Alternatives to, and Extensions of, Bracketing Results

Two alternative approaches to verifying stochastic equicontinuity with an infinite-dimensional index set Π are the VC/symmetrization and series expansion approaches. Bounds on the VC-type cover numbers can be obtained for classes of functions satisfying (6.1), e.g., see Andrews (1989b, Thm. 2). In consequence, the VC/symmetrization stochastic equicontinuity results of Yu (1990) and Arcones and Yu (1991) for absolutely regular processes can be applied to such classes of functions.

Alternatively, series expansion results of Andrews (1991a) and Andrews (1989b, Sec. 5) yield stochastic equicontinuity results for functions that satisfy (6.1) (and for functions that are the products of an unbounded function with functions of the form (6.1)). The series expansion results allow for the most general form of dependence considered in this paper. They apply when the underlying rv's $\{W_{Tt}\}$ are near epoch dependent (i.e., functions of strong mixing processes), which includes AR and ARMA processes without the restriction that the innovations possess a density with respect to Lebesgue measure.

Next, we briefly mention some extensions of the bracketing bounds given in (6.1)-(6.2). The result of (6.1)-(6.2) requires the functions $m(w, \pi)$ to be smooth in w on a compact set and constant elsewhere. This can be restrictive in some applications. Two extensions are treated in Andrews (1989b, Thm. 5). First, a bound on the bracketing cover numbers is given there for a class of functions that are smooth on their entire domain, which may be unbounded. For this result some moment conditions on $\{W_{Tt} : t \leq T, T \geq 1\}$ are needed. The result gives bounds on the log cover numbers of the form $C(1/\varepsilon)^B$, where B exceeds the corresponding value k_a/q in (6.2) and depends on the moments of $\{W_{Tt}\}$ as well as on q and k_a. These results can handle smooth trimming in semiparametric examples.

Second, a bound on the cover numbers is given for a class of functions that are smooth on one of a countable number of bounded sets and constant elsewhere. These results can handle data-dependent (non-smooth) trimming of a restricted form in semiparametric examples.

7. SOME DIRECTIONS FOR FUTURE RESEARCH

We conclude this paper by mentioning several directions in which existing stochastic equicontinuity/functional limit results in the literature could be extended to ease their application in econometric problems. Some of the extensions are relatively straightforward given existing results and others are more challenging.

First, the examples of tests for autocorrelation and conditional heteroskedasticity given in Section 3 illustrate the usefulness of having available stochastic equicontinuity results that apply to more general stochastic processes than just empirical processes. Various bracketing and VC/symmetrization results in the literature could be extended in such directions along the lines of Pollard (1989)-Andrews and Pollard (1991) and Pollard (1990) respectively.

Second, it would be useful to obtain stochastic equicontinuity results for processes that allow for more general dependence than mixing of one form or another. For example, such results could be used to circumvent the assumption that the innovations in AR or ARMA processes have a density with respect to Lebesgue measure. Some results along these lines are given by Andrews (1991a) for near epoch dependent rv's.

Third, it is desirable to have more flexible stochastic equicontinuity results for infinite-dimensional classes of functions. In particular, results that accommodate a wider variety of trimming functions are desirable.

Fourth, numerous stochastic equicontinuity results in the literature apply to strictly stationary sequences. For applications, it is useful to have results available for nonstationary triangular arrays. Results for triangular arrays are needed for applications concerning the local power of test statistics. Nonstationary results are needed when inherent nonstationarity arises due to seasonal dummies, time trends, or other fixed regressors. Many results in the literature for stationary processes could be extended to cover such situations.

FOOTNOTE

[1]I gratefully acknowledge research support from the National Science Foundation through grant no. SES-9121914.

REFERENCES

Arcones, M. A. and B. Yu (1991): "Central Limit Theorems for Empirical and *U*-processes of Stationary Absolutely Regular Sequences," unpublished manuscript, Department of Statistics, University of Wisconsin, Madison.

Andrews, D. W. K. (1989a): "Asymptotics for Semiparametric Econometric Models: I. Estimation and Testing," Cowles Foundation Discussion Paper No. 908R, Yale University.

Andrews, D. W. K. (1989b): "Stochastic Equicontinuity for Semiparametric Models," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. (1989c): "Tests for Parameter Instability and Structural Change with Unknown Change point," Cowles Foundation Discussion Paper No. 943 (revised), Yale University.

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

Andrews, D. W. K. (1991a): "An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables," *Journal of Multivariate Analysis*, 38, 187-203.

Andrews, D. W. K. (1991b): "Asymptotics for Kernel-based Non-orthogonal Semiparametric Estimators," unpublished manuscript, Cowles Foundation, Yale University.

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

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.

Andrews, D. W. K. and D. Pollard (1991): "Maximal Inequalities for Sums of Unbounded Processes under Bracketing Conditions," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. and D. Pollard (1992): "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes," *International Statistical Review*, 60, forthcoming.

Bera, A. K. and M. L. Higgins (1990): "A Test for Conditional Heteroskedasticity in Time Series Models," BEBR Faculty Working Paper No. 90-1638, College of Commerce and Business Administration, University of Illinois, Urbana-Champaign.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bradley, R. C. (1986): "Basic Properties of Strong Mixing Conditions," in *Dependence in Probability and Statistics*, ed. by E. Eberlein and M. S. Taqqu. Progress in Probability and Statistics, Vol. 11, 165-192. Boston: Birkhauser.

Chan, K. S. (1990): "Testing for Threshold Autoregression," *Annals of Statistics*, 18, 1886-1894.

Chan, K. S. and H. Tong (1991): "On Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society*, Series B, 52, 469-476.

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.

Doukhan, P. (1992): *Mixing: Properties and Examples*. Book manuscript in preparation. Department of Mathematics, University of Paris-South, Orsay.

Doukhan, P., P. Massart, and E. Rio (1992): "Invariance Principles for Empirical Mixing Processes under Bracketing Conditions," unpublished manuscript, Department of Mathematics, University of Paris-South, Orsay, France.

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

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Hansen, B. E. (1991a): "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. (1991b): "The Likelihood Test under Non-standard Conditions: Testing the Markov Trend Model of GNP," Rochester Center for Economic Research Working Paper No. 279, 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.

Horowitz, J. (1987): "Semiparametric M-estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45-83.

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

Ibragimov, I. A. and Y. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

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.

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.

Massart, P. (1988): "Invariance Principles for Empirical Processes: The Weakly Dependent Case," Ch. 1 in *Quelques Problèmes de Vitesse de Convergence pour des Mesures Empiriques*. Ph.D. Thesis, University of Paris-South, Orsay.

McLeish, D. L. (1975): "Invariance Principles for Dependent Variables," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 32, 165-178.

Mokkadem, A. (1986): "Sur le Mélange d'un processus ARMA vectoriel," *Comptes Rendus de l'Academie des Sciences Paris*, 303, Series I, no. 11, 519-521.

Mokkadem, A. (1987): "Conditions Suffisantes de Mélange Géometrique pour des Processus Autorégressifs Polynomiaux; Application aux processus ARMA et aux processus bilinéaire," *Comptes Rendus de l'Academie des Sciences Paris*, 305, Series I, 477-480.

Newey, W. K. (1987): "Efficient Estimation of Models with Conditional Moment Restrictions," unpublished manuscript, Department of Economics, Princeton University.

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

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

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.

Philipp, W. (1982): "Invariance Principles for Sums of Mixing Random Elements and the Multivariate Empirical Process," *Colloquia Mathematica Societatis János Bolyai*, 36, 843-873.

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, forthcoming.

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.

Potter, S. M. (1989): "A Nonlinear Approach to U.S. GNP," unpublished manuscript, Department of Economics, UCLA.

Robinson, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875-891.

Teräsvirta, T. and H. M. Anderson (1991): "Modelling Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," Discussion Paper #91-24, Department of Economics, University of California, San Diego.

Tong, H. (1983): *Threshold Models in Non-linear Time Series Analysis. Springer Lecture Notes in Statistics*, Vol. 21. Oxford: Oxford University Press.

Tong, H. (1990): *Non-linear Time Series: A Dynamical Systems Approach*. Berlin: Springer.

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 (1992): "Tests of Model Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, forthcoming.

White, H. and M. Stinchcombe (1990): "Adaptive Efficient Weighted Least Squares with Dependent Observations," Discussion Paper 89-45R, Department of Economics, University of California, San Diego.

Wooldridge, J. M. (1989): "Some Results on Specification Testing against Nonparametric Alternatives," unpublished manuscript, Department of Economics, M.I.T.

Yatchew, A. (1990): "Nonparametric Regression Tests Based on Least Squares," unpublished manuscript, Department of Economics, University of Toronto.

Yu, B. (1990): "Rates of Convergence and Central Limit Theorems for Empirical Processes of Stationary Mixing Sequences," Technical Report 260, Department of Statistics, University of Berkeley.