# Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays

Bruce E. Hansen[*]

Boston College[†]

February 1993

Revised: November 1994

**Abstract**

This paper establishes stochastic equicontinuity for classes of mixingales. Attention is restricted to Lipschitz-continuous parametric functions. Unlike some other empirical process theory for dependent data, our results do not require bounded functions, stationary processes, or restrictive dependence conditions. Applications are given to martingale difference arrays, strong mixing arrays, and near epoch dependent arrays.

# 1  Introduction

In the past few years we have seen many applications of empirical process theory in econometrics and statistics. For recent reviews of this literature see Andrews (1993) and Wellner (1992). The origin of empirical process theory was in the study of the empirical distribution function, where the assumption that the summands were bounded and independent across observations was natural, and therefore the theory of empirical processes was built around these assumptions. For many recent applications, however, both boundedness and independence can be overly restrictive. As a result, we have seen in recent research an effort to generalize the existing empirical process theory to handle both dependent and unbounded functions.

Several authors have demonstrated results for bounded functions of strong mixing random variables. These include Philipp (1982), Massart (1988), and Andrews and Pollard (1994). In a recent contribution, de Jong (1993) provides a result for unbounded strong mixing processes.[1]

Other authors have used alternative dependence conditions. Leventhal (1988) introduced an empirical process theorem valid for bounded martingale differences.[2] Andrews (1991) provided results for smooth classes of near-epoch dependent random functions. Arcones and Yu (1994) studied bounded V-C classes of functions of stationary absolutely regular ($\beta-$mixing) processes. Perhaps the most impressive results are those of Doukhan, Massart and Rio (in press), whose results apply to unbounded functions of stationary absolutely regular processes under weak bracketing conditions.

This paper extends this literature by presenting a proof of stochastic equicontinuity for classes of mixingale arrays. This is the first paper to do so. The results are shown to apply to martingale difference arrays, strong mixing arrays, and near epoch dependent arrays.

---

[1] While de Jong's (1993) Theorem 4 allows for unbounded arrays and weak bracketing conditions, one of his conditions effectively requires that the mixing coefficients decay exponentially, which is quite restrictive.

[2] As pointed out by a referee, Leventhals's result can be extended to handle unbounded martingale differences.

For each of these applications, the restrictions on moments and mixing decay rates are mild. The allowable function classes, however, are restrictive, only applying to Lipschitz-continuous functions. In consequence, these results are complementary to existing results, and are not a strict improvement.

Interestingly, the proof is not particularly demanding, combining a simple moment inequality for mixingales based on Hansen (1991) with the proof technique of Andrews and Pollard (1994). The paper is organized as follows. Section 2 presents the main results. The function space of interest and the concept of mixingale classes are defined. The new results are a moment inequality and stochastic equicontinuity for mixingale classes. Section 3 contains applications to three special cases: martingale difference arrays, mixing arrays, and near epoch dependent arrays. Section 4 contains a brief conclusion. The proofs are left to the Appendix.

## 2    Main Theory

### 2.1    Function Class and Stochastic Equicontinuity

Let $\{X_{ni} : i \leq n;\ n = 1, 2, ...\}$ be a triangular array of $\mathcal{X}-$valued random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_{ni}\}$ be an array of sub $\sigma$-fields of $\mathcal{F}$, such that for each $n$, $\{\mathcal{F}_{ni}\}$ is nondecreasing in $i$.

Let $G$ denote the class of real functions on $\mathcal{X}$. Let $F \subset G$ be a class of parametric functions $f(x, \theta)$, where $x \in \mathcal{X}$, $\theta \in \Theta$, and $\Theta$ is a bounded subset of $R^a$. The elements $f \in F$ satisfy the Lipschitz condition

$$|f(x, \theta) - f(x, \theta')| \leq b(x) |\theta - \theta'|^{\lambda}. \tag{1}$$

for some function $b(\cdot) : \mathcal{X} \to R$ and some $\lambda > 0$. We will sometimes index the class of functions as $f \in F$ and sometimes by $\theta \in \Theta$, depending on which is more convenient.

Define the empirical process operator $\nu_n$ by

$$\nu_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_{ni}) - E f(X_{ni})).$$

It is well understood that the empirical process $\nu_n f$ converges weakly to a Gaussian process over $f \in F$ if the finite dimensional distributions satisfy a central limit theorem, the functions $f$ are totally bounded under an appropriate seminorm, and a stochastic equicontinuity condition is satisfied. (For an elegant statement of this result, see Theorem 10.6 of Pollard (1990).) We are concerned with stochastic equicontinuity over the class $F$. Let the $L^r$ norm for a random matrix $Z$ be denoted $\|Z\|_r = (E\,|Z|^r)^{1/r}$.

**Condition 1** *(Stochastic Equicontinuity) For some $q \geq 1$, some seminorm $\rho\left(\cdot\right)$, and each $\epsilon > 0$ there exists a $\delta > 0$ such that*

$$\limsup_{n \to \infty} \left\| \sup_{\rho(f' - f'') < \delta,\ f' \in F,\ f'' \in F} |v_n f' - v_n f''| \right\|_q < \epsilon. \tag{2}$$

## 2.2  $L^q$ Mixingale Classes

The concept of $L^2$ mixingales was introduced by McLeish (1975), and generalized to $L^q$ mixingales by Andrews (1988). We generalize their concept to *classes* of random variables.

**Definition 1** $\{f(X_{ni}), \mathcal{F}_{ni}, Q\}$ *is an $L^q-$mixingale class if there exist non-negative finite functions $\{c_{ni}(f) : i \leq n\}$ and constants $\{\psi_m : m \geq 0\}$ such that for all $n \geq 1$, $i \leq n$, all $f \in Q$, and all $m \geq 0$*

$$\|E\left(f(X_{ni})|\mathcal{F}_{ni-m}\right) - Ef(X_{ni})\|_q \leq \psi_m c_{ni}(f), \tag{3}$$

*and*

$$\|f(X_{ni}) - E\left(f(X_{ni})|\mathcal{F}_{ni+m}\right)\|_q \leq \psi_{m+1} c_{ni}(f). \tag{4}$$

Note that (4) is satisfied trivially whenever $X_{ni}$ is $\mathcal{F}_{ni}$ measurable (which is true in many applications). The "mixingale numbers" $\psi_m$ control the temporal dependence of the random variables $f(X_{ni})$, and the "mixingale norms" $c_{ni}(f)$ control their magnitude.

Let $\overline{F}$ denote the class of functions formed by linear combinations of the elements of $F$ :

$$\overline{F} = \{g : g = a_1 f_1 + a_2 f_2;\ \ f_1 \in F,\ f_2 \in F,\ a_1 \in R,\ a_2 \in R\}. \tag{5}$$

**Assumption 1** *For some $q \geq 2$, $\{f(X_{ni}), \mathcal{F}_{ni}, \overline{F}\}$ is an $L^q-$mixingale class with*

4

1. $c_{ni}(f) = \|f(X_{ni})\|_s^{\gamma}$ ,

   *and*

2. $\Psi = \sum_{m=1}^{\infty} \psi_m < \infty,$

*where $s \geq q$ and $1/2 \leq \gamma \leq 1$.*

Assumption 1 specifies that linear combinations of elements of $F$ constitute a mixingale class. Part 1 states that the mixingale norms equal a power of the $L^s$ norm. The use of this particular norm is not essential to the theory which follows, but it simplifies the analysis and all our applications satisfy this restriction. Part 2 of Assumption 1 is a standard summability restriction on the mixingale numbers.

Now define an asymptotic average of the mixingale norms:

$$\rho_r(f, \beta) = \limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|f(X_{ni})\|_r^{\beta} \right)^{1/\beta}. \tag{6}$$

where $r \geq 1$ and $\beta \geq 1$. In particular, consider

$$\rho_q(f) = \rho_s(f, 2\gamma) \tag{7}$$

where $s$ and $\gamma$ are defined in Assumption 1. Note that $\rho_q(f)$ is defined for all $f \in G$. Let $F^*$ denote the class of functions for which $\rho_q(f)$ is finite:

$$F^* = \{f \in G : \rho_q(f) < \infty\}.$$

We assume that $F \subset F^*$ and $b \in F^*$:

**Assumption 2** *For all $f \in F$, $\rho_q(f) < \infty$. In addition, $\rho_q(b) < \infty$.*

It is easy to verify that $\rho_q(\cdot)$ is a seminorm on $F^*$, which is important for our theory as we later set the seminorm $\rho$ of Condition 1 equal to $\rho_q$.

## 2.3 Results

Stochastic equicontinuity (Condition 1) depends on the choice of seminorm. We find $\rho_q$ convenient as it arises in the following moment bound, whose proof is quite similar to that of Lemma 2 of Hansen (1991).

**Lemma 2** *Under Assumptions 1 and 2, for all $f \in \overline{F}$,*

$$\|\nu_n f\|_q \leq K_q \rho_q(f)^\gamma, \tag{8}$$

*where $K_q = 72q\Psi$.*

We now state our main result. The proof is analogous to that of Andrews and Pollard (1994), except that the moment inequality (8) is used instead of their Lemma 3.1.

**Theorem 3** *Under Assumptions 1 and 2 with $q > a/(\lambda\gamma)$, Condition 1 holds with $\rho(\cdot) = \rho_q(\cdot)$.*

Theorem 1 establishes stochastic equicontinuity for the Lipschitz class $F$. This is the first empirical process stochastic equicontinuity result for mixingales. The requirement $q > a/(\lambda\gamma)$ implies a trade-off between the smoothness of the functions $f(x, \theta)$ with respect to $\theta$, the dimensionality of $\theta$, and the strength of the norm $\rho_q$.

# 3   Applications

## 3.1   Martingale Difference Arrays

**Assumption 3** *For some $q \geq 2$ and each $f \in F$, $\{f(X_{ni}), \mathcal{F}_{ni}\}$ is a martingale difference array (MDA),*

$$\limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|f(X_{ni})\|_q^2 \right)^{1/2} < \infty, \tag{9}$$

*and*

$$\limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|b(X_{ni})\|_q^2 \right)^{1/2} < \infty. \tag{10}$$

Note that linear combinations $f(X_{ni}, \theta) - f(X_{ni}, \theta')$ are still MDAs, so $\{f(X_{ni}), \mathcal{F}_{ni}, \overline{F}\}$ is a MDA class. (3) holds with $\psi_m = 0$ for $m \geq 1$ and $c_{ni}(f) = \|f(X_{ni})\|_q$, and (4) holds trivially since $X_{ni}$ is $\mathcal{F}_{ni}$ measurable, so $\{f(X_{ni}), \mathcal{F}_{ni}, \overline{F}\}$ is a mixingale class. Indeed, Assumption 1 holds with $s = q$, $\gamma = 1$, and $\Psi = 1$. Assumption 2 with $\rho_q(f) = \rho_q(f, 2)$ is equivalent to (9) and (10), respectively. Hence Assumption 3 implies Assumptions 1 and 2.

We state our findings formally.

**Theorem 4** *Assumption 3 with $q > a/\lambda$ implies Condition 1 with $\rho(\cdot) = \rho_q(\cdot, 2)$.*

One application of Theorem 2 is to the asymptotic distribution of test statistics when nuisance parameters are not present under the null hypothesis. The score functions considered as a function of the unidentified nuisance parameter (in the context of maximum likelihood estimation) and regression scores (in non-linear regression), constitute classes of martingale difference arrays. For recent discussions of this testing problem, see Andrews (1993), Andrews and Ploberger (1994), and Hansen (1994).

## 3.2    Mixing Arrays

Set $\mathcal{F}_{ni} = \sigma(X_{nj} : j \leq i)$ and $\mathcal{F}_n^i = \sigma(X_{nj} : j > i)$. The strong mixing coefficients are defined as

$$\alpha_m = \sup_{i,n} \sup_{A \in \mathcal{F}_{ni}, B \in \mathcal{F}_n^{i+m}} |P(A \cap B) - P(A)P(B)|.$$

The array $\{X_{ni}\}$ is said to be strong mixing if $\alpha_m \to 0$ as $m \to \infty$.

**Assumption 4** *For some $r > q \geq 2$,*

$$\sum_{m=1}^{\infty} \alpha_m^{1/q - 1/r} < \infty, \tag{11}$$

*and all $f \in F$*

$$\limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|f(X_{ni})\|_r^2 \right)^{1/2} < \infty. \tag{12}$$

*In addition,*

$$\limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|b(X_{ni})\|_r^2 \right)^{1/2} < \infty. \tag{13}$$

7

Since functions of mixing processes are mixing with the same decay rates, $f(X_{ni}, \theta)$ is $\mathcal{F}_{ni}$ measurable with strong mixing coefficients $\alpha_m$. The same holds for linear combinations. Hence $\overline{F}$ is a strong mixing class with mixing coefficients $\alpha_m$. By McLeish's strong mixing inequality (McLeish (1975) Lemma 2.1),

$$\left\| E\left(f(X_{ni})|\mathcal{F}_{ni-m}\right) - Ef(X_{ni}) \right\|_q \leq 6\alpha_m^{1/q-1/r} \left\| f(X_{ni}) \right\|_r ,$$

so (3) is satisfied with $\psi_m = 6\alpha_m^{1/q-1/r}$ and $c_{ni}(f) = \left\| f(X_{ni}) \right\|_r$, and (4) is trivial since $X_{ni}$ is $\mathcal{F}_{ni}$ measurable. Note that (11) implies that $\Psi = 6\sum_{m=1}^{\infty} \alpha_m^{1/q-1/r} < \infty$ so Assumption 1 holds with $s = r$ and $\gamma = 1$. Assumption 2 with $\rho_q(f) = \rho_r(f, 2)$ is equivalent to (12) and (13). Hence Assumption 4 implies Assumptions 1 and 2. This allows us to state the following result.

**Theorem 5** *Assumption 4 with $q > a/\lambda$ implies Condition 1 with $\rho(\cdot) = \rho_r(\cdot, 2)$.*

## 3.3 Near-Epoch Dependent Arrays

Near epoch dependence (NED) in $L^2$ (also called functions of mixing processes) was introduced by Ibragimov (1962), and has gained popularity in the recent econometrics literature. It is often easier to verify that a function of an infinite history of a process is near epoch dependent than to verify that the function is strong mixing. Andrews (1988) introduced the concept of $L^q$ NED arrays.

Let $\{Y_{ni} : i \leq n, n \geq 1\}$ be an $\alpha$-mixing array of random variables with $\alpha$-mixing coefficients $\alpha_m$. Set $\mathcal{F}_{ni}^j = \sigma(Y_{nk} : i \leq k \leq j)$ and $\mathcal{F}_{ni} = \sigma(Y_{nk} : k \leq i)$.

**Assumption 5** *For some $r > q \geq 2$,*

$$\sum_{m=1}^{\infty} \alpha_m^{1/q-1/r} < \infty, \tag{14}$$

$$\sup_n \sup_{1 \leq i \leq n} \left\| b(X_{ni}) \right\|_r < \infty, \tag{15}$$

*and*

$$C = \sup_{\theta \in \Theta} \sup_n \sup_{1 \leq i \leq n} \left\| f(X_{ni}, \theta) \right\|_r < \infty. \tag{16}$$

(14) is a standard summability condition on the underlying mixing numbers. (15) and (16) are uniform moment bounds on the bounding functions $b(X_{ni})$ and the functions $f(X_{ni}, \theta)$.

We will consider both high-level and primitive conditions regarding the near-epoch dependence.

**Assumption 6** *For all $i, n$, there exist finite constants $\eta_m$ such that for all $f \in \overline{F}$,*

$$\left\| f(X_{ni}) - E\left(f(X_{ni})|\mathcal{F}_{ni-m}^{i+m}\right) \right\|_q \leq \eta_m, \tag{17}$$

*and for some $1/2 \leq \gamma < 1$,*

$$\sum_{m=1}^{\infty} \eta_m^{1-\gamma} < \infty, \tag{18}$$

*where $q$ is given in Assumption 5.*

(17) states that the class $f(X_{ni})$ is uniformly $L^q$-NED with respect to $Y_{ni}$. (18) is a standard summability condition on the NED and mixing numbers. We can show that these assumptions are sufficient for the application of Theorem 1.

**Theorem 6** *Assumptions 5 and 6 with $q > a/(\gamma\lambda)$ imply Condition 1 with $\rho(\cdot) = \rho_r(\cdot, 2\gamma)$.*

In some cases, it might be possible to verify Assumption 6 directly. In other cases it may not be straightforward. It is of interest to find a more primitive condition. We can do so when $f(x, \theta)$ is differentiable with respect to $x$. Let

$$D(\theta, x', x'') = \sup_{x \in R(x', x'')} \left| \frac{\partial}{\partial x} f(x, \theta) \right|, \tag{19}$$

where $R(a, b)$ is the cube containing all points between $a$ and $b$. Let $X_{ni}^m = E\left(X_{ni}|\mathcal{F}_{ni-m}^{i+m}\right)$.

**Assumption 7** *For all $i, n$, there exist finite constants $\varphi_m$ such that for some $1/2 \leq \gamma < 1$,*

$$\|X_{ni} - X_{ni}^m\|_{2q} \leq \varphi_m, \tag{20}$$

$$\sum_{m=1}^{\infty} \varphi_m^{1-\gamma} < \infty, \tag{21}$$

*and*

$$D = \sup_{\theta \in \Theta} \sup_{m} \sup_{n} \sup_{1 \leq i \leq n} \|D(\theta, X_{ni}, X_{ni}^m)\|_{2q} < \infty, \tag{22}$$

*where $q$ is given in Assumption 5.*

Assumption (20) states that $X_{ni}$ is $L^{2q}$ NED with respect to $Y_{ni}$ with NED numbers $\varphi_m$. This replaces (17) which directly assumed that the functions $f(X_{ni})$ were $L^q$ NED with respect to $Y_{ni}$. (22) is a uniform bound on the derivative function $\frac{\partial}{\partial x} f(x, \theta)$. The following theorem is proved by showing that (16), (20), and (22) imply (17) with $\eta_m = 2D\varphi_m$, and thus $\overline{F}$ is an $L^q$ NED class.

**Theorem 7** *Assumptions 5 and 7 with $q > a/(\gamma\lambda)$ imply Condition 1 with $\rho(\cdot) = \rho_r(\cdot, 2\gamma)$.*

In Theorems 4 and 5, the choice of $\gamma$ implies a trade-off between the NED decay rate ((18) or (21)) and the moment requirement $q > a/(\gamma\lambda)$. If $\gamma = 1/2$, the decay rate for $\varphi_m$ is quite mild, but we need $q > 2a/\lambda$. On the other hand, if $\varphi_m \downarrow 0$ exponentially fast, then we can let $\gamma$ be arbitrarily close to 1, and only require that $q > a/\lambda$.

# 4  Conclusion

This paper has presented empirical process limit theory applicable to dependent random functions. The conditions are weak in two dimensions: the functions are not required to be bounded; and the dependence restrictions are mild. The cost is that the results are restricted to Lipschitz-continuous function classes. Many applications involve differentiable functions, and thus can satisfy the conditions. Other applications involve discontinuous functions (such as threshold models, see Hansen (1993)) for which the results of this paper are not appropriate.

# 5 Appendix

**Proof of Lemma 1.**

Take any $f \in \overline{F}$ and let $f_{ni} = f(X_{ni})$. Without loss of generality, assume that the $\psi_m$ are (weakly) decreasing, and assume that $Ef_{ni} = 0$. Following McLeish (1975), define

$$Y_{nmi} = E\left(f_{ni} \mid \mathcal{F}_{ni-m}\right) - E\left(f_{ni} \mid \mathcal{F}_{ni-m-1}\right)$$

so that for each $m$, $\{Y_{nmi}, \mathcal{F}_{ni-m}\}$ is a MDA. We can express $f_{ni}$ as the infinite sum

$$f_{ni} = \sum_{m=-\infty}^{\infty} Y_{nmi},$$

which converges almost surely since for each $i$, $\left\{E\left(f_{ni} \mid \mathcal{F}_{ni-m}\right), \mathcal{F}_{ni-m}\right\}$ is a reverse martingale in $m$ which converges to zero almost surely as $m \to \infty$, and $\left\{E\left(f_{ni} \mid \mathcal{F}_{ni+m}\right), \mathcal{F}_{ni+m}\right\}$ is a martingale which converges a.s. to $f_{ni} - Ef_{ni}$ as $m \to \infty$. Hence we can rewrite the empirical process as

$$\nu_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_{ni} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{m=0}^{\infty} Y_{nmi} = \sum_{m=0}^{\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{nmi}. \tag{23}$$

We now show that

$$\left\|Y_{nmi}\right\|_q \leq 2\psi_m \left\|f(X_{ni})\right\|_s^\gamma. \tag{24}$$

Indeed, for $m \geq 0$, using Minkowski's inequality, (3), and Assumption 1, part 1,

$$\left\|Y_{nmi}\right\|_q \leq \left\|E\left(f_{ni} \mid \mathcal{F}_{ni-m}\right)\right\|_q + \left\|E\left(f_{ni} \mid \mathcal{F}_{ni-m-1}\right)\right\|_q \leq 2\psi_m c_{ni}(f) = 2\psi_m \left\|f(X_{ni})\right\|_s^\gamma.$$

Similarly for $m < 0$, using (4) and Assumption 1,

$$\left\|Y_{nmi}\right\|_q \leq \left\|f_{ni} - E\left(f_{ni} \mid \mathcal{F}_{ni-m}\right)\right\|_q + \left\|f_{ni} - E\left(f_{ni} \mid \mathcal{F}_{ni-m-1}\right)\right\|_q \leq 2\psi_m \left\|f(X_{ni})\right\|_s^\gamma.$$

From (23), Minkowski's inequality, Burkholder's inequality (see, e.g., Hall and Heyde (1980) Theorem 2.10), again Minkowski's inequality, (24), (6), and (7), we have

$$\left\|\nu_n f\right\|_q \leq \sum_{m=0}^{\infty} \left\|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{nmi}\right\|_q \leq \sum_{m=0}^{\infty} \frac{18q^{3/2}}{\sqrt{q-1}} \frac{1}{\sqrt{n}} \left(E\left|\sum_{i=1}^{n} Y_{nmi}^2\right|^{q/2}\right)^{1/q}$$

$$\leq 36q \sum_{m=0}^{\infty} \left( \frac{1}{n} \sum_{i=1}^{n} \|Y_{nmi}\|_q^2 \right)^{1/2} \leq 36q \sum_{m=0}^{\infty} \left( \frac{1}{n} \sum_{i=1}^{n} 4\psi_m^2 \|f(X_{ni})\|_s^{2\gamma} \right)^{1/2}$$

$$= 72q \Psi \left( \frac{1}{n} \sum_{i=1}^{n} \|f(X_{ni})\|_s^{2\gamma} \right)^{1/2} \leq K_q \rho_s(f, 2\gamma)^{\gamma} = K_q \rho_q(f)^{\gamma},$$

as required.    □

## Proof of Theorem 1.

For each $k = 1, 2, ...,$ set $\delta_k = 2^{-k\lambda}$ and $N(k) = \delta_k^{-a/\lambda} = 2^{ka}$. Let $\Theta_k = \{\theta^1, ..., \theta^{N(k)}\}$ denote a set of elements of $\Theta$ with the property that for all $\theta \in \Theta$, there exists some $\theta^j \in \Theta_k$ such that $|\theta - \theta^j| \leq Q 2^{-k}$ where $Q < \infty$. This is possible since $\Theta$ is a bounded subset of $R^a$. Let $\theta_k$ be the function of $\theta$ which selects $\theta^j \in \Theta_k$ so that

$$|\theta - \theta_k| \leq \frac{Q}{2^k}. \tag{25}$$

By (1) and (25),

$$|f(X_{ni}, \theta) - f(X_{ni}, \theta_k)| \leq b(X_{ni}) |\theta - \theta_k|^{\lambda} \leq Q^{\lambda} b(X_{ni}) \delta_k,$$

and $\rho_q(Q^{\lambda} b(X_{ni}) \delta_k) = Q^{\lambda} \rho_q(b) \delta_k$, which show that $N(k)$ are proportional to the bracketing numbers for $F$ with respect to the metric $\rho_q(\cdot)$. Let $F_k$ denote the class $F_k = \{f(\cdot, \theta) : \theta \in \Theta_k\}$. Observe that $F_k$ has $N(k)$ distinct elements and the sets can be constructed in such a way that $F_{k-1} \subset F_k$. Let $f_k(\cdot, \theta) = f(\cdot, \theta_k)$, and note that $f_k \in F_k$.

Fix $\epsilon > 0$. Take a sequence of integers $k(n)$ which satisfies $\sqrt{n} 2^{-k(n)\lambda} \to 0$. Then

$$\left\| \sup_f |\nu_n f - \nu_n f_{k(n)}| \right\|_q = \left\| \sup_{\theta \in \Theta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \left( f(X_{ni}, \theta) - f(X_{ni}, \theta_{k(n)}) \right) - E \left( f(X_{ni}, \theta) - f(X_{ni}, \theta_{k(n)}) \right) \right] \right| \right\|_q$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \left\| \sup_{\theta \in \Theta} \left| f(X_{ni}, \theta) - f(X_{ni}, \theta_{k(n)}) \right| \right\|_q + \left\| \sup_{\theta \in \Theta} E \left| f(X_{ni}, \theta) - f(X_{ni}, \theta_{k(n)}) \right| \right\|_q \right]$$

$$\leq \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \|b(X_{ni})\|_q \sup_{\theta \in \Theta} \left| \theta - \theta_{k(n)} \right|^{\lambda} \leq 2\rho_q(b, 1) \sqrt{n} \frac{Q^{\lambda}}{2^{k(n)\lambda}}. \tag{26}$$

The first inequality is Minkowski's, the second uses (1), and the third uses definition (6) and property (25). Since $1 \leq 2\gamma$ and $q \leq s$ under Assumption 1, by the monotonicity properties of norms,

$$\rho_q(b, 1) \leq \rho_q(b, 2\gamma) \leq \rho_s(b, 2\gamma) < \infty$$

12

where the final inequality is from Assumption 2. Thus (26) tends to 0 as $n \to \infty$, and so for $n$ sufficiently large,

$$\left\| \sup_f \left| \nu_n f - \nu_n f_{k(n)} \right| \right\|_q \le \epsilon. \tag{27}$$

Next, let $g_k = f_k - f_{k-1}$ and let $G_k = \{ f(\cdot, \theta') - f(\cdot, \theta'') : \theta' \in \Theta_k, \theta'' \in \Theta_{k-1} \} \subset \overline{F}$. Note that $G_k$ has $N(k)$ distinct elements (the same as $F_k$) since the sets $F_k$ are nested. Lemma 1, (1), and (25) yield

$$\left\| \nu_n g_k \right\|_q \le K_q \rho_q(g_k)^\gamma = K_q \limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} \left\| f(X_{ni}, \theta_k) - f(X_{ni}, \theta_{k-1}) \right\|_s^{2\gamma} \right)^{1/2}$$

$$\le K_q \limsup_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} |\theta_k - \theta_{k-1}|^{2\gamma\lambda} \left\| b(X_{ni}) \right\|_s^{2\gamma} \right)^{1/2} \le A 2^{-k\lambda\gamma}, \tag{28}$$

where $A = (4Q)^{\lambda\gamma} \rho_q(b)^\gamma < \infty$ by Assumption 2. By Pisier's inequality (Pisier, 1983) and (28),

$$\left\| \sup_f |\nu_n g_k| \right\|_q = \left\| \max_{g_k \in G_k} |\nu_n g_k| \right\|_q \le N(k)^{1/q} \max_{g_k \in G_k} \left\| \nu_n g_k \right\|_q \le A\beta^k \tag{29}$$

where $\beta = 2^{(a/q - \lambda\gamma)}$. Note that $\beta < 1$ since $\lambda > a/(q\gamma)$ by assumption.

Now let $M = M(\epsilon)$ be an integer large enough to satisfy

$$A \sum_{k=M+1}^{\infty} \beta^k \le \epsilon. \tag{30}$$

Let $n$ be sufficiently large so that (27) holds and $k(n) > M$. We now apply a chaining argument. Since $f_{k(n)} - f_M = \sum_{k=M+1}^{k(n)} g_k$, we have by Minkowski's inequality, (29) and (30)

$$\left\| \sup_f \left| \nu_n f_{k(n)} - \nu_n f_M \right| \right\|_q \le \sum_{k=M+1}^{k(n)} \left\| \sup_f |\nu_n g_k| \right\|_q \le \sum_{k=M+1}^{\infty} A\beta^k \le \epsilon. \tag{31}$$

(27) and (31) together imply that

$$\left\| \sup_f |\nu_n f - \nu_n f_M| \right\|_q \le 2\epsilon. \tag{32}$$

The proof is completed by an argument identical to that of "Comparison of pairs" of Andrews and Pollard (1994) and is omitted. □

**Proof of Theorem 4.**

Minkowski's inequality and the Rao-Blackwell theorem yield

$$\left\| f(X_{ni}) - E\left(f(X_{ni})|\mathcal{F}_{ni-m}^{i+m}\right) \right\|_q \leq 2 \left\| f(X_{ni}) \right\|_q. \tag{33}$$

Inequalities (17) and (33) can be combined[3] to yield

$$\left\| f(X_{ni}) - E\left(f(X_{ni})|\mathcal{F}_{ni-m}^{i+m}\right) \right\|_q \leq 2^\gamma \eta_m^{1-\gamma} \left\| f(X_{ni}) \right\|_q^\gamma \leq 2^\gamma \eta_m^{1-\gamma} \left\| f(X_{ni}) \right\|_r^\gamma, \tag{34}$$

where the final inequality uses the assumption that $q < r$ and the monotonicity of the $L^r$ norm.

Andrews (1988, equation (2)) showed that under (34)

$$\left\| E\left(f(X_{ni})|\mathcal{F}_{ni-2m}\right) - Ef(X_{ni}) \right\|_q \leq 2^\gamma \eta_m^{1-\gamma} \left\| f(X_{ni}) \right\|_r^\gamma + 6\alpha_m^{1/q-1/r} \left\| f(X_{ni}) \right\|_r,$$

$$\leq \left( 2^\gamma \eta_m^{1-\gamma} + 6\alpha_m^{1/q-1/r} C^{1-\gamma} \right) \left\| f(X_{ni}) \right\|_r^\gamma,$$

where $C$ is defined in (16). He also showed that under (34)

$$\left\| f(X_{ni}) - E\left(f(X_{ni})|\mathcal{F}_{ni+m}\right) \right\|_q \leq 2^{1+\gamma} \eta_m^{1-\gamma} \left\| f(X_{ni}) \right\|_r^\gamma.$$

This shows that (3) and (4) are satisfied with

$$\psi_{2m} = 2^{1+\gamma} \eta_m^{1-\gamma} + 6\alpha_m^{1/q-1/r} C^{1-\gamma}$$

and

$$c_{ni}(f) = \left\| f(X_{ni}) \right\|_r^\gamma.$$

Thus $\{f(X_{ni}), \mathcal{F}_{ni}, \overline{F}\}$ is an $L^q$-mixingale class. Assumption 1 holds with $s = r$ and

$$\Psi = 2\left( 2^{1+\gamma} \sum_{m=1}^{\infty} \eta_m^{1-\gamma} + 6C^{1-\gamma} \sum_{m=1}^{\infty} \alpha_m^{1/q-1/r} \right) < \infty$$

under (18) and (14). Assumption 2 is trivially satisfied under (15) and (16). Thus Assumptions 5 and 6 are sufficient for Assumptions 1 and 2. Hence Theorem 1 holds. $\square$

---

[3]I owe this suggestion to an anonymous referee.

**Proof of Theorem 5.**

We show below that (17) holds with $\eta_m = 2D\varphi_m$. Thus Assumption 6 holds and Theorem 4 yields the result. The differentiability of $f$ (or any linear combination of two elements of $F$) implies that

$$|f(x', \theta) - f(x'', \theta)| \leq D(\theta, x', x'') \cdot |x' - x''|. \tag{35}$$

By Minkowski's inequality, the fact that $f(X_{ni}^m, \theta)$ is $\mathcal{F}_{ni-m}^{i+m}$-measurable, the Rao-Blackwell Theorem, (35), Holder's inequality, (22) and (20),

$$\left\| f(X_{ni}) - E\left( f(X_{ni}) | \mathcal{F}_{ni-m}^{i+m} \right) \right\|_q \leq \left\| f(X_{ni}) - f(X_{ni}^m) \right\|_q + \left\| E\left( f(X_{ni}) | \mathcal{F}_{ni-m}^{i+m} \right) - f(X_{ni}^m) \right\|_q$$

$$\leq 2 \left\| f(X_{ni}, \theta) - f(X_{ni}^m, \theta) \right\|_q$$

$$\leq 2 \left\| D(\theta, X_{ni}, X_{ni}^m) \cdot |X_{ni} - X_{ni}^m| \right\|_q$$

$$\leq 2 \left\| D(\theta, X_{ni}, X_{ni}^m) \right\|_{2q} \left\| X_{ni} - X_{ni}^m \right\|_{2q} \leq 2D\eta_m,$$

as required. $\square$

15

# References

[1] Andrews, D.W.K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458-467.

[2] Andrews, D.W.K. (1991) An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis* 38, 187-203.

[3] Andrews, D.W.K. (1993) An introduction to econometric applications of empirical process theory for dependent random variables. *Econometric Reviews* 12, 183-216.

[4] Andrews, D.W.K. and W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383-1414.

[5] Andrews, D.W.K. and D. Pollard (1994) An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* 62, 119-132.

[6] Arcones, M.A. and Yu, B. (1994) Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability* 7, 47-71.

[7] Bierens, J.H. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443-1458.

[8] Doukhan, P., P. Massart and E. Rio (in press) Invariance principles for absolutely regular empirical processes. *Annales de l'Institut H. Poincare.*

[9] Gallant, A.R. and H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

[10] Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and Its Application.* New York: Academic Press.

[11] Hansen, B.E. (1991) Strong laws for dependent heterogeneous processes. *Econometric Theory* 7, 213-221, and (1992) Erratum. *Econometric Theory* 8, 421-422.

[12] Hansen, B.E. (1993) Inference in threshold models. Boston College.

[13] Hansen, B.E. (1994) Inference when a nuisance parameter is not identified under the null hypothesis. Boston College.

[14] Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349-382.

[15] de Jong, R.M. (1993) Stochastic equicontinuity for unbounded mixing processes. Free University Amsterdam.

[16] Leventhal, S. (1988) A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* 2, 271-287.

[17] McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829-839.

[18] Massart, P. (1988) *Invariance Principles for Empirical Processes: The Weakly Dependent Case.* Chapter 1 of Ph. D. dissertation, University of Paris.

[19] Philipp, W. (1982) Invariance principles for sums of mixing random elements and the multivariate empirical process. *Colloquia Mathematica Societatis János Bolyai* 36, 843-873. (Limit Theorems in Probability and Statistics Verzprém (Hungary).)

[20] Pisier, G. (1983) Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis, and Probability Theory. Lecture Notes in Mathematics* 995, 123-154. New York: Springer.

[21] Pollard, D. (1984) *Convergence of Stochastic Processes.* NY: Springer.

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

[23] Wellner, J. (1992) Empirical processes in action: A review. *International Statistical Review* 60, 247-269.

[24] White, H. (1989) An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *Proceedings of the International Joint Conference on Neural Networks*, II, 451-455. New York: IEEE Press.