Lecture Notes for Time Series: Dependence and Stochastic Limit Theory

Jonathan B. Hill
Dept. of Economics
University of North Carolina - Chapel Hill

January 27, 2010

The following is a necessarily brief outline of dependence concepts and limit theory for dependent, heterogeneous processes. The purpose of these notes is to give an abridged source of limit theorems to help established probability and distribution limits for NLLS, QML and GMM estimators. This is not required reading, although some assignments problems are greatly expedited by citing the appropriate theorem or corollary, below. Consult the bibliography and course syllabus for excellent source material.

Throughout $K > 0$ is a finite constant whose value may change from line to line. Proofs of selected results are presented in the appendix, and a bibliograhy follows.

Throughout the L_p-norm for scalars is

$$\|x_t\|_p := (E|x_t|^p)^{1/p}$$

and for $m \times n$ matrices identically

$$\|x_t\|_p := \left(\sum_{i=1}^{m}\sum_{j=1}^{n} E|x_{t,i,j}|^p\right)^{1/p}.$$

1. Probability Spaces, Borel Functions, Stochastic Processes

1.1 σ-Fields, Borel Fields, Borel Measurability

Let Υ be the event space, $\Im$ the σ-field of events $\omega \in \Upsilon$, and P a probability measure. The *probability space* is denoted

$$(\Upsilon, \Im, P).$$

A *sigma-field* $\Im$ (σ-field, or σ-algebra) is a subset of Υ that contains $\varnothing$ and Υ, and is closed under complementation and countable unions. If particular, $\Im$ is a σ-field *if and only if*

$$\Im \neq \varnothing,\ \varnothing \in \Im,\ \ \Upsilon \in \Im$$

$$\text{If } A \subseteq \Im \text{ then } A^C \subseteq \Im$$

$$\{A_i\}_{i=1}^{\infty} : \text{if } A_i \subseteq \Im\ \forall i \text{ then } \cup_{i=1}^{\infty} A_i \subseteq \Im.$$

A random variable $x : \Upsilon \rightarrow \mathbb{R}$ is $\Im$-*measurable* (an $\Im$-measurable function, or $\Im/\mathbb{R}$ measurable) *if and only if*

$$x^{-1}(A) \in \Im \text{ for every } A \in \mathbb{R}.$$

That is, if a realization of x on the real line is associated with subsets of events ω in $\Im$, then probabilities of events involving x can be computed from the likelihood of events $\omega \in \Im$. Put more succinctly: all realizations of x are associate with events in $\Im$.

The *σ-algebra induced (or generated) by* x, denoted $\sigma(x)$, is the intersection of all σ-algebras $\Im$ on which x is measurable. That is, if x is only measurable on the sequences $\{\Im_i\}_{i=1}^{\infty}$ of σ-algebras, then

$$\sigma(x) = \bigcap_{i=1}^{\infty} \Im_i.$$

This is the smallest set of events associated with all realizations of x.

EXAMPLE 1.1.1 Let $\Upsilon = \{\varnothing, 0, 1, 2\}$. Let $x(0) = x(1) = 5$ and $x(2) = 10$. Construct

$$\Im = \{\{\varnothing\}, \{\Upsilon\}, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

It is easy to verify that $\Im$ is closed under countable unions and complementation. For example, $\{0, 1\}^c = \{2\} \in \Im$. Then

$$x^{-1}(5) = \{0, 1\} \in \Im \ \text{ and } \ x^{-1}(10) = \{2\} \in \Im.$$

Therefore $x(\omega)$ is $\Im/\{0, 1, 2\}$-measurable. Moreover, $\sigma(x) = \Im$ since any other σ-field on which x is measurable will be "larger" than $\Im$. In other words, $\Im$ is the smallest σ-field on which x is measurable.

A *Borel σ-field* β is a σ-field generated from subsets $(-\infty, x]$ of the real line $(-\infty, \infty)$:

$$\varnothing \in \beta,\ \mathbb{R} \subseteq \beta$$

$$(-\infty, x] \subseteq \beta \text{ and } (-\infty, x]^c \subseteq \beta\ \forall x \in \mathbb{R}$$

$$\bigcup_{q \in \mathbb{Q}} (-\infty, q] \subseteq \beta.$$

The union is over rational endpoints since the rationals are a countable space (Doob 1953).

1.2 Borel Function

A *Borel function* $g : \Upsilon \to \mathbb{R}$, or *Borel measurable function* (or $\beta/\mathbb{R}$-measurable) satisfies

$$x^{-1}(A) \in \beta \text{ for every } A \in \mathbb{R}.$$

It is very difficult to construct a function that is not Borel measurable. In general we are merely agreeing that we can assign real intervals $(-\infty, w]$ to otherwise abstract events ω, and "measure" the likelihood of $x(\omega)$ by "measuring" the interval $(-\infty, w]$.

That we talk at all about this abstraction is not an empty exercise. That are many econometric results that require very little of the functions under inspection. Whereas some results require functions to be continuos, others require only Borel measurability. In fact, continuity ensures Borel measurability.

LEMMA 1.2.1 *Any $g : \mathbb{R} \to \mathbb{R}$ that is almost everywhere continuous is Borel measurable.*

1.3 Stochastic Processes

1.3.1 Definition A *stochastic process* is a generalization of a measurable mapping $\{x\} : \Upsilon \to \mathbb{R}^{\mathbb{Z}}$:

$$\{x(\omega)\} = \{x_t(\omega)\}_{t=-\infty}^{\infty} = \{x_t(\omega) : t \in \mathbb{Z}\}.$$

The entire sequence $\{...x_t, x_{t+1}, ...\}$ must be jointly measurable: we must be able to assign a joint distribution function to any subsequence $\{x_{t_1}, ..., x_{t_l}\}$ for any non-redundant l-tuple $\{t_1, ..., t_l\} \in \mathbb{Z}^l$.

Notice that the entire sequence $\{x_t(\omega)\}_{t=-\infty}^{\infty}$ is a function of one random draw $\omega \in \Upsilon$. We must think of time-sequenced events as being one possible *realization* of a random experiment $\omega \in \Upsilon$.

We say "generalization" because the index need not be integer-valued, nor represent a natural linear ordering. In practice the coordinate t is time, and $x_{t+1}(\omega)$ occurs after $x_t(\omega)$, and never the other way around.

A *sample path* is an observable sequence based on one (not necessarily observable) random draw $\omega \in \Upsilon$:

$$\{x_t(\omega)\}_{t=1}^{n}.$$

1.3.2 Properties A process $\{x(\omega)\}$ is *uniformly bounded in probability* if for every $\varepsilon > 0$ there exists a $K_\varepsilon < \infty$ such that

$$\sup_{t \in \mathbb{Z}} P(|x_t| > K_\varepsilon) \leq \varepsilon.$$

Uniform boundedness is extremely important since many processes will be time-dependent, and may increase or decrease without bound. Consider a simple *linear trend*

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \overset{iiid}{\sim} (0, \sigma^2),$$

where ϵ_t has a strictly positive, continuous density (e.g. normal). Clearly x_t not uniformly bounded in probability since for any finite K_ε

$$\sup_{t \in \mathbb{Z}} P(|\beta_0 + \beta_1 t + \epsilon_t| > K_\varepsilon) = 1.$$

A process is *uniformly L_p-bounded*, $p > 0$, when

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p < \infty.$$

Notice uniform L_p-boundedness implies uniform boundedness in probability by Markov's inequality:

$$\sup_{t \in \mathbb{Z}} P\left(|x_t| > K_\varepsilon\right) \leq \frac{1}{K_\varepsilon^p} \sup_{t \in \mathbb{Z}} \|x_t\|_p^p < \infty.$$

Therefore for any $\varepsilon > 0$ choose

$$\frac{1}{\varepsilon^{1/p}} \sup_{t \in \mathbb{Z}} \|x_t\|_p = K_\varepsilon.$$

If $p = \infty$ the L_p-norm reduces to the *sup-norm*:

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p = \sup_{t \in \mathbb{Z}} |x_t| < \infty \quad a.s.$$

hence x_t is *uniformly almost surely bounded* (e.g. $P(a \leq x_t \leq b) = 1$ for any finite $a \leq b$: this is true when data are truncated or trimmed (e.g. $x_t \times I(|x_t| \leq a) \in [-a, a]$ $a.s.$ uniformly in t for any finite $a > 0$).

A random variable x_t is *uniformly integrable* when

$$\lim_{M \to \infty} E\left[|x_t| \times I\left(|x_t| \geq M\right)\right] = 0.$$

This means extreme-values (i.e. the tail support of the distribution) do not dominate because they occur too rarely. A process $\{x(\omega)\}$ is *uniformly integrable* when

$$\lim_{M \to \infty} \sup_{t \in \mathbb{Z}} E\left[|x_t| \times I\left(|x_t| \geq M\right)\right] = 0.$$

THEOREM 1.3.1.1 *$L_{1+\delta}$-boundedness for any $\delta > 0$ implies uniform integrability.*

2. Stationarity and Ergodicity

2.1 Stationarity

A process $\{x_t(\omega)\}$ is *strictly stationary* when *shift transformations* are *measure preserving*: the joint distribution of $\{x_{t-m_1}, x_{t-m_2}, ..., x_{t-m_k}\}$ depends only on $\{m_1, ..., m_k\}$ and not t. This means the *finite dimensional distributions* (the joint distribution of $\{x_{t-m_1}, x_{t-m_2}, ..., x_{t-m_k}\}$) depend only displacement and not the specific points in time.

A process $\{x_t(\omega)\}$ is *covariance stationary* if

$$E[x_t] = \mu \ \forall t$$

$$E\left[(x_t - E[x_t])^2\right] = \sigma^2 < \infty \ \forall t$$

$$\gamma(s,t) := E[(x_s - E[x_s])(x_t - E[x_t])] = \gamma(h) \text{ where } h = |s-t|.$$

That is, the process is stationarity with respect to its mean, variance and autocovariances.

If $\{x_t(\omega)\}$ is strictly stationary and L_p-bounded, then $\forall r \leq p$

$$E|x_t|^r = E|x_1|^r.$$

2.2 Ergodicity: Ergodic Theorem as a LLN

2.2.1 Ergodicity: Definition and Example Let $\{x_t(\omega)\}$ exist on $(\Upsilon, \Im, P)$. Consider a sample path over time for some ω_0, $\{x_t(\omega_0)\}_{t=1}^n$, and an ensemble sequence over Υ for some time t_0, $\{x_{t_0}(\omega_i)\}_{i=1}^N$. We want to know when the ensemble average for any t_0

$$\frac{1}{N}\sum_{i=1}^N x_{t_0}(\omega_i)$$

has the same limit as the time average for any event ω_0

$$\frac{1}{n}\sum_{t=1}^n x_t(\omega_0).$$

If $\{x_t(\omega)\}$ is stationary $E[x_t] = E[x_s] \ \forall s,t$, in which case we want to know when

$$\frac{1}{n}\sum_{t=1}^n x_t(\omega_0) \to E[x_{t_0}] \text{ for arbitrary } t_0.$$

EXAMPLE 2.2.1 Person A collects rainfall data $x_{t_0}(\omega_i)$ from N locations $\{\omega_i\}_{i=1}^N$ in North Carolina at time t_1. Person B collects rainfall data $x_t(\omega_1)$ from t time periods $t \in \{1, ..., n\}$ in North Carolina at place ω_1. Clearly the data collected by each person need not have the same population characteristics. Point ω_1 may be very arid, and time t_1 may have been very wet, such that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N x_{t_1}(\omega_i) > \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n x_t(\omega_1).$$

A stationary process $\{x_t(\omega)\}$ is *ergodic for the mean* when

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N x_t(\omega_i) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n x_t(\omega) = E[x_t] \ \forall t.$$

There is no "systematic bias" in the sample space (e.g. no dominant arid regions in North Carolina: fluctuation over time at location ω is no different, on average, than fluctuation over locations).

2.2.2 Ergodicity: Memory in σ-Fields Let $\{x_t(\omega)\}$ be stationary. Stationarity is typically unrealistic in practice due to event time trend in mean or variance, or seasonal effects. Nevertheless, ergodicity of a stationary process permits substantial persistence, and is therefore interesting because it sets a certain horizon by which to compare the persistence of other dependence concepts.

Let $\Im = \sigma(x)$ and $A, B \in \Im$. Let T be a measure-preserving shift transformation: $P(TB) = P(B)$ and TB shifts the events that make up A. Thus $T^k B$ performs k shift operations.

We say a *measuring preserving shift transformation T is ergodic* if the resulting average memory vanishes.

THEOREM 2.2.2.1 *A measuring preserving shift transformation T is* ergodic *if and only if $\forall A, B \in \Im$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[P\left(A \cap T^i B\right) - P\left(A\right)P\left(B\right)\right] = 0. \tag{1}$$

That is, on average events are independent.

A process $\{x_t(\omega)\}$ is *ergodic* if for all t

$$x_t(T^k\omega) = x_{t+k}(\omega)$$

where T is measure-preserving and ergodic. This is a key, albeit abstract, point. A process is *ergodic* when moving ("shifting") over the event space is equivalent to moving over time. Thus, random draws at one point in time will generate the same information, asymptotically, as drawing observations over time. That way an ensemble at time t (which we typically do not have) will have the same probabilistic structure as a sequence over time (which we do have).

EXAMPLE 2.2.2.2 Let $\Upsilon = \{0,1\}$ so that $\omega = 0$ or 1, and $\Im = \{\{\varnothing\}, \{\Upsilon\}, \{0\}, \{1\}, \{0,1\}\}$, $P(0) = P(1) = 1/2$. Let $T0 = 1$ and $T1 = 0$. Clearly T is measuring preserving: $P(T\omega) = P(\omega) = 1/2$. If $A = \{0,1\}$ and $B = \{0\}$ then for all even k

$$P\left(A \cap T^k B\right) - P\left(A\right)P\left(B\right) = P\left(0\right) - P\left(\{0,1\}\right)P\left(0\right) = 0,$$

and for all odd k

$$P\left(A \cap T^k B\right) - P\left(A\right)P\left(B\right) = P\left(1\right) - P\left(\{0,1\}\right)P\left(0\right) = 0.$$

Similarly, if $A = \{0\}$ and $B = \{1\}$ then for all even k

$$P\left(A \cap T^k B\right) - P\left(A\right)P\left(B\right) = P\left(\varnothing\right) - P\left(0\right)P\left(1\right) = -1/4,$$

and for all odd k

$$P\left(A \cap T^k B\right) - P\left(A\right)P\left(B\right) = P\left(0\right) - P\left(0\right)P\left(1\right) = 1/4,$$

which averages to zero. And so on. T is ergodic. Now define $x_1(\omega) := \omega$ and $x_t(\omega) := x_1(T^{t-1}\omega)$ for $t = 2, 3, ...$ Then the sequence starts off randomly as 0 or 1, and alternates thereafter, say $0, 1, 0, 1, ...$ Over time the average is obviously $1/2$, and by construction the ensemble average is $1/2$: $E[x_t(\omega)] = 0 \times .5 + 1 \times .5 = .5$. Therefore $\{x_t(\omega)\}$ is ergodic.

EXAMPLE 2.2.2.3 Let $\Im = \{\{\varnothing\}, \{\Upsilon\}, \omega_1, ..., \omega_6\}$, where $P(\omega_i) = 1/6$. Consider the subsets $A = \{\omega_1, \omega_2, \omega_3\}$, $B = \{\omega_1, \omega_2, \omega_4\}$. The shift transformation $T^k B = \{\omega_{1+2k}, \omega_{2+2k}, \omega_{4+2k}\}$ is measure preserving since $P(T^k B) = P(B) = 1/2$, where $\omega_b = \omega_{b-6}$ for all $b > 6$. Then

$$P(A \cap TB) - P(A)P(B) = P(\omega_3) - 1/4 = 1/6 - 1/4 = -1/12$$

$$P\left(A \cap T^2 B\right) - P(A)P(B) = P(\omega_2) - 1/4 = 1/6 - 1/4 = -1/12$$

$$P\left(A \cap T^3 B\right) - 1/4 = P(\omega_1, \omega_2) - 1/4 = 1/3 - 1/4 = 1/12$$

$$P\left(A \cap T^4 B\right) - 1/4 = P(\omega_3) - 1/4 = 1/6 - 1/4 = -1/12,$$

and so on. Thus

$$\frac{1}{n}\sum_{i=1}^{n}\left[P\left(A \cap T^i B\right) - P(A)P(B)\right] = \frac{1}{n}\left[-1/12 - 1/12 + 1/12 + \cdots\right] \to -1/3,$$

hence T is not ergodic.

There are two way to summarize the memory of an ergodic process. The first merely re-states (1) in terms of time distance, rather than an abstract "shift".

THEOREM 2.2.2.4 Let $\Im_t$ *denote the* σ-field induced by $\{x_t, x_{t-1}, ...\}$. *The process* $\{x_t\}$ *is ergodic if and only if for all* $A_t, B_t \subseteq \Im_t$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[P(A_t \cap B_{t-i}) - P(A_t)P(B_{t-i})\right] = 0. \quad (2)$$

This states x_t is ergodic if its serial memory (i.e. dependence between x_t and x_{t-i}) vanishes *on average*. That does not mean that distant events are independent, which implies typically that ergodicity permits too much persistence to be of use in theory (central limit theory requires more restrictions). We invoke ergodicity, along with other structural assumptions, to ensure that time averages have a well-defined probability limit.

THEOREM 2.2.2.5 *If* $\{x_t(\omega)\}$ *is stationary, ergodic, and* L_2*-bounded then*

$$\frac{1}{n}\sum_{i=1}^{n} cov(x_i, x_1) \to 0.$$

The following theorem presents a simple sufficient condition for mean ergodicity.

THEOREM 2.2.2.6 *If* $\{x_t(\omega)\}$ *is covariance stationary and* $\sum_{i=0}^{\infty}|cov(x_{i+1}, x_1)| < \infty$ *, then it is ergodic for the mean*:

$$\frac{1}{n}\sum_{t=1}^{n} x_t(\omega) \to E\left[x_t(\omega)\right].$$

EXAMPLE 2.2.2.7 *Let* $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$, $\sigma^2 < \infty$. *Then the MA(1) process define by*

$$x_t = \theta\epsilon_{t-1} + \epsilon_t,\ |\theta| < \theta$$

has very short memory. Clearly for all $A_t, B_t \subseteq \Im_t = \sigma(x_\tau : \tau \leq t)$ *and all* $i \geq 2$, $P(A_t \cap B_{t-i}) = P(A_t)\,P(B_{t-i})$. *Therefore* $1/n \sum_{i=1}^{n}[P(A_t \cap B_{t-i}) - P(A_t)\,P(B_{t-i})]$ $\to 0$, *and* $1/n \sum_{t=1}^{n} x_t(\omega) \to 0$. *Of course* $\sum_{i=0}^{\infty} |cov(x_{i+1}, x_1)| = \sigma^2(1+\theta^2) + \sigma^2|\theta|$ $< \infty$.

EXAMPLE 2.2.2.8 *Let* $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$ *and*

$$x_t = a + bt + \epsilon_t.$$

Since $E[x_t] = a + bt$ *is non-stationary (and non-covariance stationary) Theorem 2.2.2.6 does not apply.*

LEMMA 2.2.2.9 *Consider an AR(1) data generating process*

$$x_t = \phi x_{t-1} + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2) \text{ and } |\phi| < 1.$$

Then $\sum_{i=0}^{\infty} |cov(x_{i+1}, x_1)| = \sigma^2/[(1-\phi^2)(1-|\phi|)] < \infty$. *Therefore* x_t *is ergodic for the mean.*

3. Dependent, Heterogeneous Arrays

3.1 Mixing

3.1.1 Mixing Transformation Let $\{x_t(\omega)\}$ be a stochastic process. Ergodicity is a fairly weak memory property that applies to time sequences *on average*, but does not imply asymptotically distant events are independent. In economics we typically assume infinitely distant events (e.g. x_t and x_{t-N} as $N \to \infty$) are independent, or central limit theory requires that they be dependent. A *mixing* condition implies just that.

Let $\Im = \sigma(x)$ and $A, B \in \Im$. An ergodic, measure-preserving transformation T is said to be mixing when for any $A, B \in \Im$

$$\lim_{k\to\infty} P\left(T^k A \cap B\right) = P(A) P(B).$$

A process $\{x_t(\omega)\}$ is mixing when

$$x_t(T^k\omega) = x_{t+k}(\omega).$$

where T is an *ergodic, measure-preserving mixing transformation.*

3.1.2 Strong and Uniform Mixing More concrete representations of mixing is available. Let $\{x_t\}$ be a stochastic process with σ-field

$$\Im_a^b := \sigma\left(x_t : a \le t \le b\right).$$

Define sequences α_m and ϕ_m as follows:

$$\alpha_m = \sup_{t\in\mathbb{Z}} \sup_{G\in\Im_{-\infty}^t, H\in\Im_{t+m}^\infty} \left|P\left(G\cap H\right) - P(G)P(H)\right|$$

and

$$\phi_m = \sup_{t\in\mathbb{Z}} \sup_{G\in\Im_{-\infty}^t, H\in\Im_{t+m}^\infty} \left|P\left(H|G\right) - P(H)\right|.$$

The process $\{x_t\}$ is *uniform mixing* if

$$\phi_m \to 0 \text{ as } m \to \infty$$

and *strong mixing* if

$$\alpha_m \to 0 \text{ as } m \to \infty.$$

Mixing simply states the asymptotically distant events $G \in \Im_{-\infty}^t$ and $H \in \Im_{t+m}^\infty$ are independent. Therefore the random variables x_{t-m} and x_t are independent as $m \to \infty$ for any t.

It is not difficult to prove

$$\alpha_m \le \phi_m,$$

hence "strong mixing" is actually weaker than "uniform mixing", and therefore of greater interest in theory.

3.1.3 Mixing Size The rate at which x_t and x_{t-k} become independent as $k \to \infty$ is measured by "size". There are several representations of size, but the most popular is hyperbolic boundedness.

We say x_t is strong or uniform mixing of *size* $\lambda > 0$ if

$$\alpha_m = o(m^{-\lambda}) \text{ or } \phi_m = o(m^{-\lambda}).$$

Hyperbolic boundedness is typically the weakest condition allowed in theory since it represents a "*long*" form of memory: x_t and x_{t-k} become independent very slowly.

Notice what size means. If x_t is strong mixing of size 2 then $\forall \delta \leq 2$

$$m^{\delta} \times \sup_{t \in \mathbb{Z}} \sup_{G \in \Im_{-\infty}^{t}, H \in \Im_{t+m}^{\infty}} |P(G \cap H) - P(G)P(H)| \to 0$$

as $m \to \infty$. Thus, x_t and x_{t-k} become independent at a rate m^2. The larger the size, the faster the rate of convergence, hence the weaker is dependence (less persistence).

By contrast, x_t is *geometrically* strong or uniform mixing if there exists a $\rho \in (0, 1)$ such that

$$\alpha_m = o(\rho^m) \text{ or } \phi_m = (\rho^m).$$

It is easy to prove that geometric mixing implies mixing of any size $\lambda > 0$, hence geometric mixing is a strong condition, and it represents a "*short*" form of memory: x_t and x_{t-k} become independent comparatively quickly.

Geometric mixing specifically, and mixing with size $\lambda > 1$ in general, imply summability of the mixing coefficients:

$$\sum_{m=0}^{\infty} \alpha_m < \infty \text{ or } \sum_{m=0}^{\infty} \phi_m < \infty.$$

3.1.4 Mixing Time Series Consider the following example to get an idea about what mixing refers to.

EXAMPLE 3.1.4.1 Consider placing one red and one blue floating ball into an infinitely large swimming pool, and let x_t denote the Euclidean distance between them. Consider installing a sequence of shocks over time: we stick a large stick in the water and stir once. If we stir the pool with a large stick once, the resulting distance is x_{t+1}. Clearly x_{t+1} will depend on x_t (if $x_t = 10$ then x_{t+1} is more likely to be 12 than, say, 20000). But after repeatedly stirring, or *mixing* the pool k-times, the resulting distance x_{t+k} as $k \to \infty$ will be unrelated to x_t since eventually x_{t+k} can take on any value in $\mathbb{R}_+$, irrespective of the original distance x_t. Thus x_t is said to "mix."

This example is not obscure. If random shocks ϵ_t enter into a data generating process for some daily exchange rate x_t, say, according to

$$x_t = f(x_{t-1}, ..., x_{t-p}; \phi) + \epsilon_t$$

where $f : \mathbb{R}^p \times \Phi \to \mathbb{R}$ is unknown, $\Phi \subseteq \mathbb{R}^p$, then how does the contemporary rate x_t relate to past rates x_{t-k} as $k \to \infty$? The pool example implied past ϵ_{t-i} shocks (old stirs, as it were) had a diminishing impact on the present distance x_t, and recent shocks (new stirs) were more important. In general this turns out to be very difficult to answer from a mixing perspective, and only fairly recent results in the literature have clarified what kinds of nonlinear data generating processes mix.

The following is a special case of a far more general proof for nonlinear autoregressions. See An and Huang (1996) and Leibscher (2005).

LEMMA 3.1.4.2 *Consider an AR(1) data generating process*

$$x_t = \phi x_{t-1} + \epsilon_t$$

where $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$ *has a strictly positive, continuous density function, and* $|\phi| < \infty$*. Then* x_t *is geometrically strong mixing.*

The condition that the density of ϵ_t be continuous and positive over its entire support is very restrictive, although Gaussian, exponential, etc., densities satisfy the condition. It is not difficult to find a simple stationary, ergodic AR(1) process that is not strong mixing because the density of ϵ_t is not continuous (e.g. Bernoulli). See Andrews (1984).

3.1.5 Mixing Properties The following are useful for extending ergodicity, and laws of large numbers, to mixing random variables.

LEMMA 3.1.5.1 *For $p > 1$ and $r \geq p/(p-1)$*

$$|Cov(x_t, x_{t-m})| \leq 2\left(2^{1/p}+1\right)\alpha_m^{1-1/p-1/r}\left\|x_t\right\|_p\left\|x_t\right\|_r$$

where x_t is $L_{\max\{p,r\}}$-bounded.

LEMMA 3.1.5.2 *For $r \geq 1$*

$$|Cov(x_t, x_{t-m})| \leq 2\phi_m^{1/r}\left\|x_t\right\|_r\left\|x_t\right\|_{r/(r-1)}$$

where x_t is $L_{\max\{r,,r/(r-1)\}}$-bounded.

Both results instantly imply ergodicity for the mean for mixing processes. The geometric case is trivial.

THEOREM 3.1.5.3 *If x_t is covariance stationary, $L_{2+\delta}$-bounded, and geometrically strong mixing then it is ergodic for the mean: $1/n\sum_{t=1}^n x_t \xrightarrow{p} E[x_t]$.*

One of the most useful characteristics of mixing random variables is the fact that the mixing property extends to finite lag functions of mixing random variables.

LEMMA 3.1.5.4 *Let $y_t := g(x_t, x_{t-1}, ..., x_{t-k})$ be a measurable function for finite $k > 0$. If x_t is strong or uniform mixing of size $\lambda > 0$ then y_t is too.*

EXAMPLE 3.1.5.5 If x_t is strong mixing of size $\lambda > 0$ then so is

$$\sum_{i=0}^k \psi_i x_{t-i}$$

for any finite $k > 0$, and any $\psi_i \in \mathbb{R}$.

EXAMPLE 3.1.5.6 If ϵ_t is strong mixing of size $\lambda > 0$ and $\{x_t\}$ is a stochastic process, then

$$E\left[x_t|\epsilon_t, ..., \epsilon_{t-k}\right]$$

is strong mixing of size $\lambda > 0$.

3.2 Martingales and Martingale Difference Sequences

Consider a process $\{x_t\}$ with σ-field $\Im_t := \sigma(x_\tau : \tau \leq t)$. A *martingale* $\{x_t, \Im_t\}_{-\infty}^{\infty}$ satisfies

$$E[x_t|\Im_{t-1}] = x_{t-1}.$$

Evidently "martingale" originally referred to a betting scheme, and the concept in probability theory dates at least to P. Lévy (1925, 1954) and Doob (1953).

EXAMPLE 3.2.1 Consider an AR(1) with slope of unity:

$$x_t = x_{t-1} + \epsilon_t, \;\; \epsilon_t \overset{iid}{\sim} (0, \sigma^2).$$

The autocovariances are not summable, and x_t exhibits too much persistence to be mixing. It is, however, a martingale since independence in the shocks implies $E[x_t|\Im_{t-1}] = x_{t-1}$. The Efficient Market Hypothesis posits asset prices are martingales: the price of an asset at time $t-1$ in an efficient market contains all relevant information to make the best (minimum mean-squared-error) forecast for time t.

Although the mathematical structure of a martingale has important applications in economics and finance, it permits more memory than is typically allowed for laws of large numbers and central limit theorems.

Conversely, the first difference of a martingale x_t,

$$\epsilon_t := x_t - x_{t-1},$$

has so little structure, it easily permits a LLN and CLT. The process $\{\epsilon_t, \Im_t\}_{-\infty}^{\infty}$ is referred to a *martingale difference sequence* [mds] with the prominent characteristics

$$E[\epsilon_t|\Im_{t-1}] = 0 \tag{3}$$

and

$$E\left[\epsilon_s \epsilon_t\right] = E\left(\epsilon_s E\left[\epsilon_t|\Im_{t-1}\right]\right) = E\left(\epsilon_s \times 0\right) = 0 \tag{4}$$

In general a martingale difference sequence is simply defined by (3) without reference to a martingale. Trivially a mean-zero iid process $\{\epsilon_t, \Im_t\}_{-\infty}^{\infty}$, $\Im_t := \sigma(\epsilon_\tau : \tau \leq t)$ forms a martingale difference sequence $\{\epsilon_t, \Im_t\}_{-\infty}^{\infty}$ since by construction $E[\epsilon_t|\Im_{t-1}] = E[\epsilon_t] = 0$.

Since all autocovariances are zero, an *mds* $\{\epsilon_t\}$ is ergodic for the mean (see Theorem 2.2.2.6).

3.2.1 Linear Distributed Lags Linear distributed lags of *mds*'s represent a useful time series process that surfaces in many contexts (e.g. covariance stationary ARIMA models, random walks).

Let $\{\epsilon_t, \Im_t\}_{-\infty}^{\infty}$ be a martingale difference sequence, $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = \sigma^2$ and define

$$x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_t.$$

Then the autocovariances $\gamma(h) := E[x_t x_{t-h}]$ are (use (4))

$$\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}.$$

The autocovariances $\{\gamma(h)\}_{h\in\mathbb{N}}$ are absolutely summable sufficiently if $\{\psi_i\}_{i=0}^{\infty}$ are absolutely summable:

$$\sum_{h=0}^{\infty} |\gamma(h)| = \sigma^2 \sum_{h=0}^{\infty} \left| \sum_{i=0}^{\infty} \psi_i \psi_{i+h} \right| \leq \sigma^2 \sum_{h=0}^{\infty} \sum_{i=0}^{\infty} \left(|\psi_i| \times |\psi_{i+h}| \right) \leq \sigma^2 \left(\sum_{i=0}^{\infty} |\psi_i| \right)^2 < \infty.$$

LEMMA 3.2.1.1 *Let $\{\epsilon_t, \Im_t\}_{-\infty}^{\infty}$ be a martingale difference sequence, $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = \sigma^2 < \infty$, and define $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_t$. If $\{\psi_i\}$ are absolutely summable then x_t is ergodic for the mean: $1/n \sum_{t=1}^{n} x_t \overset{p}{\to} E[x_t]$.*

3.3 Near Epoch Dependence

Martingales permit too much persistence to allow for LLN's and CLT's in general (without serious restrictions), and *mds*'s permit far too little structure to be able to model the kind of memory economic and financial time series exhibit. Nevertheless, many processes with substantial degrees of persistence and heterogeneity can be approximated by *mds*'s, making *mds*'s indispensable for LLN's and CLT's.

Mixing properties are one solution, but they are laborious to verify, and often apply only under rather abstract, or restrictive conditions (e.g. positive, continuous density f on the entire support of f).

The *Near Epoch Dependence* [NED] property, by comparison, trivially covers mixing processes, and characterize a plethora of time series models. In particular, the property does not explicitly require any restriction on the continuity or positiveness of a density, and is based on expectations making it relatively easy to verify in practice.

3.3.1 NED Definition Let $\{\mathcal{F}_t\}$ be a sequence of σ-algebras ostensibly induced by a sequence of random variables $\{v_t\}$,

$$\mathcal{F}_t := \sigma\left(v_\tau : \tau \leq t\right).$$

Notice v_t may be vector valued (e.g. the shocks in a VAR model).

We say $\{x_t\}$ is L_p-NED on $\{\mathcal{F}_t\}$ (or on $\{v_t\}$) with size $\lambda > 0$ if

$$\left\|x_t - E\left[x_t|\mathcal{F}_{t-m}^{t+m}\right]\right\|_p \leq d_t \times \varphi_m$$

where $d_t > 0$ and $\varphi_m = o(m^{-\lambda})$. The "*constants*" $\{d_t\}$ capture trending moments and therefore $d_t \to \infty$ as $t \to \infty$ is possible (e.g. heteroscedasticity a la linear time trend $E[x_t^2] = a + bt$, $a, b > 0$). The "*coefficients*" simply capture persistence. The random variable $\{v_t\}$ is the NED "*base*".

The two-sidedness of $\mathcal{F}_{t-m}^{t+m}$ simply admits two-sided time-series, like a two-sided linear distributed lag

$$x_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i}. \quad (5)$$

In general, this will not be encountered in economic settings, in which case substitute $\mathcal{F}_{t-m}^{t}$ for $\mathcal{F}_{t-m}^{t+m}$.

In words, $\{x_t\}$ is L_p-NED on $\{\mathcal{F}_t\}$ when a sequence of random variables $\{v_\tau\}_{\tau=t-m}^{t+m}$ can be used to predict x_t with zero error in L_p-norm as $m \to \infty$. As with mixing, $\{x_t\}$ is L_p-NED on $\{\mathcal{F}_t\}$ with size 2, for example, if $\forall\delta \leq 2$

$$m^\delta \times \left\|x_t - E\left[x_t|\mathcal{F}_{t-m}^{t+m}\right]\right\|_p \to 0 \text{ as } m \to \infty.$$

3.3.2 NED Examples Trivially iid random variables are NED on themselves: if $\Im_t := \sigma(x_\tau : \tau \leq t)$ then $E[x_t|\Im_{t-m}^{t+m}] = E[x_t|\Im_{t-m}^{t}] = x_t$ $\forall m \geq 1$ hence $\varphi_m = 0$ $\forall m \geq 1$.

The NED property covers mixing random variables since v_t can be anything. If $x_t = v_t$ is strong mixing then x_t is trivially NED on itself (i.e. on $\{\mathcal{F}_t\} = \{\Im_t\}$) with $\varphi_m = 0$ $\forall m \geq 1$.

Verifying the NED property is often very easy. Not all covariance stationary, ergodic linear distributed lags are strong mixing (e.g. Andrews 1984, Guegan and Ladoucette 2001), but all are NED.

LEMMA 3.3.2.1 *Let ϵ_t be zero-mean, uniformly L_p-bounded with $||\epsilon_t||_p = s_t$. Define $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ where $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Then $\{x_t\}$ is L_p-NED on $\{\epsilon_t\}$. The size is $\lambda > 0$ if $\psi_i = O(i^{-1-\lambda+\iota})$ for some infinitessimal $\iota > 0$.*

Any time series with an infinite order distributed lag representation based on L_p-bounded zero-mean innovations, and absolutely summable coefficients, is therefore L_p-NED. This includes any stationary, ergodic AR(p), MA(q) and ARMA(p, q) for finite p, q.

3.3.3 NED Properties Evidently not every function of NED random variables is necessarily NED. Compare this statement to what we know about mixing: finite lag measurable functions of mixing random variables are always mixing (Lemma 3.1.5.4). Nevertheless, some basic properties give some leeway when deriving asymptotic distributions of sample means or sample covariances of NED vectors.

LEMMA 3.3.3.1 (Linear Combinations are NED) *If $\{x_{i,t} : i = 1...k\}$ are L_p-bounded and L_p-NED on some $\{\digamma_t\}$ with constants $\{d_{i,t}\}$ and coefficients $\{\varphi_{i,m}\}$ of size $\{\lambda_i\}$ then $\sum_{i=1}^{k} \psi_i x_{i,t}$ is L_p-bounded and L_p-NED on $\{\digamma_t\}$ with constants $d_t := \sum_{i=1}^{k} |\psi_i| d_{i,t}$ and coefficients $\varphi_m := \sum_{i=1}^{k} \varphi_{i,m}$ of size $\min_{1 \le i \le k}\{\lambda_i\}$.*

LEMMA 3.3.3.2 (Products are NED) *If $\{x_t, y_t\}$ are L_p-bounded and L_p-NED on some $\{\digamma_t\}$ with constants $\{d_{x,t}, d_{y,t}\}$ and coefficients $\{\varphi_{x,m}, \varphi_{y,m}\}$ of size $\{\lambda_x, \lambda_y\}$ then $x_t \times y_t$ is $L_{p/2}$-NED on $\{\digamma_t\}$ with constants $d_t := d_{x,t} d_{y,t}$ and coefficients $\varphi_m := \varphi_{x,m} \varphi_{y,m}$ of size $\min_{1 \le i \le k}\{\lambda_i\}$.*

4. Laws of Large Number for Stochastic Processes

Let $\{x_t\}$ be a mean zero stochastic process $\mu_t := E(x_t)$, $\sigma_t^2 := E(x_t - \mu_t)^2$, and autocovariances $\gamma(|s-t|) = E(x_s - \mu_s)(x_t - \mu_t)$. We implicitly allow for non-stationary time series (e.g. trend in mean; heteroscedasticity), but for brevity we assume the autocovariances depend only on displacement $h = |s-t|$. For example, if x_t is simply iid noise about a time trend

$$x_t = \beta_0 + \beta_1 t + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2)$$

then

$$\mu_t = \beta_0 + \beta_1 t$$

$$\sigma_t^2 = \sigma^2$$

$$E\left(x_s - \mu_s\right)\left(x_t - \mu_t\right) = 0\ \forall s \neq t.$$

Decompose the sample mean variance into

$$E\left(\frac{1}{n}\sum_{t=1}^{n}\left(x_t - \mu_t\right)\right)^2 = \frac{1}{n^2}\sum_{t=1}^{n}\sigma_t^2 + \frac{2}{n}\sum_{i=1}^{n-1}(1 - i/n) \times \gamma(i). \tag{6}$$

If we are to use Chebyshev's inequality

$$P\left(\left|\frac{1}{n}\sum_{t=1}^{n}\left(x_t - \mu_t\right)\right| > \varepsilon\right) \leq \varepsilon^{-2} E\left(\frac{1}{n}\sum_{t=1}^{n}\left(x_t - \mu_t\right)\right)^2$$

to prove a LLN, we must be sure

$$\frac{1}{n^2}\sum_{t=1}^{n}\sigma_t^2 + \frac{2}{n}\sum_{i=1}^{n-1}(1 - i/n) \times \gamma(i) \to 0. \tag{7}$$

Certainly there are other methods for proving LLN's. We will focus on the above simple trick for the sake of brevity and to build intuition on what various dependence properties imply about (7).

The martingale difference case is the simplest, with iid as a special case.

4.1 Martingale Difference Sequences

THEOREM 4.1.1 *If $\{x_t, \Im_t\}$ is a zero mean martingale difference sequence with variances $\{\sigma_t^2\}$, $1/n\sum_{t=1}^{n}\sigma_t^2 = o(n)$ then $1/n\sum_{t=1}^{n} x_t \overset{p}{\to} 0$.*

Remark 1: As long as the variances are uniformly bounded, $\sup_{t\in\mathbb{N}}\sigma_t^2 \leq K < \infty$, clearly $1/n\sum_{t=1}^{n}\sigma_t^2 \leq K$ hence $1/n^2\sum_{t=1}^{n}\sigma_t^2 = o(1)$ since $K/n \to 0$. The MDS assumption of course ensures $\sum_{i=1}^{n-1}(1 - i/n)\gamma(i) = 0$.

Remark 2: Note that the scale $1/n$ is simply a special case. There are many interesting cases involving non-stationary data in which the required scale is something other than $1/n$.

4.2 Heterogeneous Martingale Difference Sequences

More enlightened results are available.

LEMMA 4.2.1 *If $\{S_n, \Im_n\}$ is a martingale, then for any $p > 1$*

$$P\left(\max_{1\leq k\leq n} |S_k| > \varepsilon\right) \leq \frac{E\,|S_n|^p}{\varepsilon^p}.$$

Remark: This result is a bit subtle since Chebyshev's inequality already implies

$$P\left(\max_{1\leq k\leq n} |S_k| > \varepsilon\right) \leq \frac{E\left[\max_{1\leq k\leq n} |S_k|\right]^p}{\varepsilon^p}.$$

The result importantly abstracts from the max operation, a hugely useful result.

LEMMA 4.2.2 *Let $\{x_t\}$ be a stochastic process and $S_n = \sum_{t=1}^n x_t$. Suppose there exists a sequence of positive numbers $\{c_t\}$ such that for some $p > 0$, every $m \geq 0$ and $n > m$, and every $\varepsilon > 0$*

$$P\left(\max_{m<j\leq n} |S_j - S_m| > \varepsilon\right) \leq \frac{K}{\varepsilon^2} \sum_{t=m+1}^{n} c_t^p.$$

If $\sum_{t=1}^\infty c_t^p < \infty$ then $S_n \stackrel{a.s.}{\to} S$ (the sequence converges).

THEOREM 4.2.3 *If $\{x_t, \Im_t\}$ is a zero mean mds with variance sequence $\{\sigma_t^2\}$ that satisfies $\sum_{t=1}^\infty \sigma_t^2/a_t^2 < \infty$ for positive sequence $\{a_t\}$, $a_t \uparrow \infty$, then $1/a_n \sum_{t=1}^n x_t \stackrel{p}{\to} 0$.*

Simple examples abound. If $\{x_t\}$ is a covariance stationary *mds* then $\sigma^2 = \sigma_t^2 < \infty$, and $\sum_{t=1}^\infty 1/t^2 < \infty$ is trivial, so $1/n \sum_{t=1}^n x_t \stackrel{p}{\to} 0$. Generally, if $\sup_{t\in\mathbb{N}} \sigma_t^2 \leq K < \infty$ then $\sum_{t=1}^\infty \sigma_t^2/t^2 \leq K \sum_{t=1}^\infty 1/t^2 < \infty$.

The finite variance assumption can be relaxed

THEOREM 4.2.4 *If $\{x_t, \Im_t\}$ is a zero mean where $\sum_{t=1}^\infty E|x_t|^p/a_t$ for some sequence of positive numbers $\{a_t\}$, $a_t \uparrow \infty$, $1 \leq p \leq 2$, then $1/a_n \sum_{t=1}^\infty x_t \stackrel{a.s.}{\to} 0$.*

4.3 Strong Mixing Processes

Assume variances σ_t^2 and covariances $\gamma(t, t-h)$ are uniformly bounded

$$B \;:=\; \sup_{t\in\mathbb{N}} \sigma_t^2 < K \;\text{ and }\; B_h \;:=\; \sup_{t\in\mathbb{N}} |\gamma(t, t-h)| < K$$

and write

$$E\left(\frac{1}{n}\sum_{t=1}^n (x_t - \mu_t)\right)^2 = \frac{1}{n^2}\sum_{t=1}^n \sigma_t^2 + \frac{1}{n^2}\sum_{i=1}^{n-1} (n-1)\,\gamma(i) \leq \frac{B}{n} + \frac{1}{n}\sum_{i=1}^n B_i.$$

LEMMA 4.3.1 *If $\{x_t\}$ uniformly L_2-bounded, and $\sum_{i=1}^\infty i^{-1}B_i < \infty$ then $1/n \sum_{t=1}^n x_t \stackrel{p}{\to} 0$.*

Applications of Lemma 4.3.1 are immediate. We need only establish bounded memory in the sense of $\sum_{i=1}^\infty i^{-1}B_i < \infty$. Although not very intuitive on the surface, this condition is easily satisfied for a massive array of stochastic processes. Consider an easy result.

THEOREM 4.3.2 *Let $\{x_t\}$ be mean zero, uniformly $L_{2+\delta}$-bounded, and strong mixing with coefficients $\sum_{m=1}^\infty m^{-1}\alpha_m^{\delta/(2+\delta)} < \infty$. Then $1/n \sum_{t=1}^n x_t \stackrel{p}{\to} 0$.*

4.4 Near Epoch Dependent Processes

The beauty of NED random variables x_t is they are *mixingales*, and mixingales under minimal additional assumptions have $O(1/n)$-bounded sample mean variances $E(1/n \sum_{t=1}^{n} [x_t - \mu_t])^2$. In other words, under the appropriate conditions

$$E\left(\frac{1}{n}\sum_{t=1}^{n}(x_t - \mu_t)\right)^2 \to 0$$

and Chebyshev's inequality again suffices to ensure a LLN. See McLeish (1974, 1975) and Davidson (1994) for a formal definition of mixingale. The following is partially adapted from work due to de Jong (1997).

We need the notion of a double (or stochastic) array $\{x_{n,t}\}$. In the following $\{x_{n,t}\}$ should be thought of as a triangular array: $\{\{x_{n,t}\}_{t=1}^{n}\}_{n=1}^{\infty}$. The classic example is

$$x_{n,t} := \frac{1}{\sqrt{n}}\left(\frac{x_t - E[x_t]}{\sqrt{E(x_t - E[x_t])^2}}\right) \text{ where } x_t \sim iid.$$

Define the σ-algebra induced by the NED base: $\mathcal{F}_t := \sigma(\epsilon_\tau : \tau \leq t)$.

THEOREM 4.4.1 *Assume the following:*

a. $\{x_{n,t}\}$ is a mean zero, L_r-bounded, $r > 2$, stochastic array.

b. $\{x_{n,t}\}$ is L_2-NED of size $1/2$ with constants $\{d_{n,t}\}$ and coefficients $\{\varphi_m\}$. The base ϵ_t is uniform mixing of size $r/[2(r-1)]$, or strong mixing of size $r/(r-2)$.

c. The constants $\{d_{n,t}\}$ satisfy

$$\sum_{t=1}^{n}\left(\max\left\{\|x_{n,t}\|_r, d_{n,t}\right\}\right)^2 = O(1).$$

Then there exists a sequence of positive constant numbers $\{a_n\}$, $a_n \nearrow \infty$, satisfying $1/a_n \sum_{t=1}^{n} x_{n,t} \xrightarrow{p} 0$.

The strict stationarity case involving a process $\{x_t\}$ rather than an array $\{x_{n,t}\}$ removes some messy notation.

COROLLARY 4.4.2 *Let $\{x_t\}$ be strictly stationary, have mean $E[x_t]$, be L_r-bounded for some $r > 2$. Moreover, assume $\{x_t\}$ is L_2-NED of size $1/2$ with constants $d_t = d$ and coefficients φ_m, and with NED base $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$. Then $\{x_t\}$ is ergodic for the mean: $1/n \sum_{t=1}^{n} x_t \to E[x_t]$.*

5. Central Limit Theorems for Dependent, Heterogeneous Arrays

Our goal is to produce a reasonably tight set of conditions that ensures some standardized triangular array $\{x_{n,t}\}$, $||\Sigma_{t=1}^{n} x_{n,t}||_2 = 1$, satisfies

$$\sum_{t=1}^{n} x_{n,t} \xrightarrow{d} N(0,1).$$

5.1 Martingale Differences

Despite the fact that *mds*'s have so little structure, unlike the *mds* LLN we will have to invoke a few assumptions regarding heterogeneity. The following result, presented in Davidson (1994: Theorem 24.3) exploits a general result due to McLeish (1974).

THEOREM 5.1.1 *Let $\{x_{n,t}, \Im_{n,t}\}$ be a martingale difference array with finite variances $\{\sigma^2_{n,t}\}$, and $\sum_{t=1}^{n} \sigma^2_{n,t} = 1$. If $\sum_{t=1}^{n} x^2_{n,t} \xrightarrow{p} 1$ and $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$, then $\sum_{t=1}^{n} x_{n,t} \xrightarrow{d} N(0,1)$.*

Consider the conventional case

$$x_{n,t} = \frac{1}{\sqrt{n}} \left(\frac{x_t - E[x_t]}{\sqrt{E\left(1/\sqrt{n} \sum_{t=1}^{n} (x_t - E[x_t])\right)^2}} \right)$$

where $\{x_t, \Im_t\}$ is a martingale difference sequence with variances $\{\sigma^2_t\}$, $\sigma^2_t > 0$ uniformly in t. Notice $\Im_{n,t} = \Im_t$ since the deterministic n does not inform a σ-field. The *mds* assumption implies

$$E\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (x_t - E[x_t])\right)^2 = \frac{1}{n} \sum_{t=1}^{n} \sigma^2_t$$

hence

$$\sigma^2_{n,t} = \frac{1}{n} \frac{\sigma^2_t}{\frac{1}{n} \sum_{t=1}^{n} \sigma^2_t} = \frac{\sigma^2_t}{\sum_{t=1}^{n} \sigma^2_t}.$$

so trivially $\sum_{t=1}^{n} \sigma^2_{n,t} = 1$.

The condition $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$ requires

$$\max_{1 \leq t \leq n} \left| \left(\frac{x_t - E[x_t]}{\sqrt{\sum_{t=1}^{n} \sigma^2_t}} \right) \right| \xrightarrow{p} 0$$

which is nearly trivial since $\sum_{t=1}^{n} \sigma^2_t \to \infty$ due to $\inf_t \sigma^2_t > 0$. We need only assume no one mean-differenced observation $|x_t - E[x_t]\,|$ dominates all others.

Finally, exploit the *mds* property to deduce

$$\sum_{t=1}^{n} x^2_{n,t} = \frac{\sum_{t=1}^{n} (x_t - E[x_t])^2}{E\left(\sum_{t=1}^{n} (x_t - E[x_t])\right)^2} = \frac{\sum_{t=1}^{n} (x_t - E[x_t])^2}{\sum_{t=1}^{n} E\left(x_t - E[x_t]\right)^2}.$$

The condition $\sum_{t=1}^{n} x^2_{n,t} \xrightarrow{p} 1$ therefore reduces to $\sum_{t=1}^{n} (x_t - E[x_t])^2$ satisfying a LLN. Sufficient conditions are $\{(x_t - E[x_t])^2 - \sigma^2_t, \Im_t\}$ is an *mds* where $E((x_t - E[x_t])^2 - \sigma^2_t)^{\ 2}$ is constant uniformly in t.

Although relatively simple to satisfy, the condition $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$ is slightly more than required. The following is presented in White (1984: Corollary 5.25). The idea is that while an mds $\{x_t, \Im_t\}$ may be heterogeneous (e.g. non-constant variances), as long as the heterogeneity is not explosive (i.e. no one observation dominates all others) then the central limit property is preserved.

THEOREM 5.1.2 *Let $\{x_t, \Im_t\}$ be an mds with finite variances $\{\sigma_t^2\}$, and $1/n \sum_{t=1}^n \sigma_t^2 \rightarrow \sigma^2$, $\sup_{t\in\mathbb{Z}} E|x_t|^r < \infty$ for some $r > 2$ and $1/n \sum_{t=1}^n x_t^2 \xrightarrow{p} \sigma^2$, then $1/\sqrt{n} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \sigma^2)$.*

5.2 Near Epoch Dependent Arrays

The following is a slightly trimmed version of Theorem 2 of de Jong (1997).

THEOREM 5.2.1 *Let $\{x_{n,t}\}$ be a zero mean triangular array, and assume the following:*

a. The exists a sequence of positive numbers $\{c_{n,t}\}$ such that $\{x_{n,t}/c_{n,t}\}$ is L_r-bounded for some $r > 2$, uniformly in t and n.

b. $\{x_{n,t}\}$ is L_2-NED of size 1/2 with constants $\{d_{n,t}\}$ and coefficients $\{\varphi_m\}$. The base ϵ_t is uniform mixing of size $r/[2(r-1)]$, or strong mixing of size $r/(r-2)$.

c. The numbers $\{c_{n,t}\}$ satisfy $\sum_{t=1}^n c_{n,t}^2 = O(1)$.

Then $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$.

Remark: Since Theorem 5.2.1 only requires $\{x_{n,t}\}$ to be NED on a uniform or strong mixing process, mixing processes with an appropriate size trivially satisfy.the above conditions

Consider a simplified corollary.

COROLLARY 5.2.2 *Let $\{x_t\}$ be covariance stationary process and L_r-bounded for some $r > 2$. Moreover, assume $\{x_t\}$ is L_2-NED of size 1/2 with constants $d_t = d$ and coefficients φ_m, and with NED base $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$. Finally, $\liminf_{n\geq 1} E(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\})^2 > 0$. Write*

$$x_{n,t} = \frac{1}{\sqrt{n}} \left(\frac{x_t - E[x_t]}{\sqrt{E\left(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\}\right)^2}} \right).$$

Then $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$ and $E(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\})^2 = O(1)$.

Remark: Notice we assume the standard deviation is non-degenerate $\liminf_{n\geq 1} E(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\})^2 > 0$ and the result claims it is also bounded: $E(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\})^2 \leq K$. Therefore $E(1/\sqrt{n} \sum_{t=1}^n \{x_t - E[x_t]\})^2 \rightarrow v > 0$ where $v < \infty$.

5.3 Cramer-Wold Device

Extending the above CLT's to the multivariate case $x_t \in \mathbb{R}^k$ is typically no more difficult than establishing a CLT on a linear combination $\sum_{i=1}^k r_i x_{i,t}$.

THEOREM 5.3.1 (Cramér-Wold Device) *If $\{x_{n,t}\}$ is an $\mathbb{R}^k$-valued stochastic array, and $\forall r \in \mathbb{R}^k$*

$$\sum_{t=1}^n r' x_{n,t} \xrightarrow{d} N(0, r'r)$$

then $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, I_k)$.

COROLLARY 5.3.2 *If* $\{x_t\}$ *is an* $\mathbb{R}^k$*-valued stochastic process and* $\forall r \in \mathbb{R}^k$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} r' x_t \stackrel{d}{\to} N(0, r'\Sigma r).$$

then $1/\sqrt{n} \sum_{t=1}^{n} x_t \stackrel{d}{\to} N(0, \Sigma)$ *where* $\Sigma = \lim_{n\to\infty} E\left(1/\sqrt{n} \sum_{t=1}^{n} x_t\right)^2$.

Linear combinations of mixing processes are mixing, and linear combinations of NED processes are NED. In general, then, Theorem 5.3.1 and Corollary 5.3.2 are immensely useful for the derivation of limit distributions of multivariate estimators (e.g. QMLE, OLS, GMM).

6. Weak Limit Theory and Brownian Motion

In this section we detail functional limits that are useful for characterizing the limit distribution of estimators associated with unit roots. We exemplify the main results by computing from first principles the limit distribution of the OLS slope estimator in an AR(1) with a unit root.

6.1 Random Walk with Drift: Example

Consider a random walk without drift:

$$x_t = x_{t-1} + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2).$$

Consider the OLS estimator $\hat{\rho}$ of the slope parameter ρ in an AR(1) model

$$x_t = \rho x_{t-1} + \epsilon_t \quad \text{where } \rho = 1.$$

Then

$$\hat{\rho} = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2} = 1 + \frac{\sum_{t=2}^{n} \epsilon_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}.$$

Thus $\hat{\rho} \xrightarrow{p} 1$ follows if a properly scaled $\sum_{t=2}^{n} \epsilon_t x_{t-1} \xrightarrow{p} 0$. But since

$$x_t = \sum_{i=0}^{\infty} \epsilon_{t-i}$$

has an infinite variance

$$E\left[x_t^2\right] = \sum_{i=0}^{\infty} \sigma^2 = \infty \text{ given } \epsilon_t \overset{iid}{\sim} (0, \sigma^2),$$

what *are* the probability and distribution limits of $\sum_{t=2}^{n} \epsilon_t x_{t-1}$? Indeed, since $E[x_t^2] = \infty$ can we use the scale $1/n$ in both components of the ratio (e.g. $1/n \sum_{t=2}^{n} x_{t-1}^2$?). The fast answers are: decidedly *non-Gaussian*; and *no*.

We decipher the limit law of $\hat{\rho}$ circuitously by first reviewing Functional Central Limit theory, and then tackling $\hat{\rho}$ head-on.

6.2 Functional Limit Theory

Like all topics covered in these notes, this is necessarily abbreviated. We only focus on what we need to characterize the distribution limit of $\hat{\rho}$ in Section 6.1.

Let $\{x_t\}$ be an arbitrary zero mean stochastic process. If $x_t \overset{iid}{\sim} (0, \sigma^2)$ then (e.g. Lindeberg-Lévy central limit theorem)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \xrightarrow{d} N\left(0, \sigma^2\right).$$

Notice "n" is completely arbitrary as long as the limit is with respect to $n \to \infty$. The result remains valid for any fraction $[\lambda n]$ of n, where $\lambda \in [0, 1]$ and $[z]$ denotes the integer part of z. A simple re-scaling and the fact that $[\lambda n]/\lambda n \to 1$ implies

$$\begin{aligned} Z_n(\lambda) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{[\lambda n]} x_t \\ &= \frac{\sqrt{[\lambda n]}}{\sqrt{n}} \times \frac{1}{\sqrt{[\lambda n]}} \sum_{t=1}^{[\lambda n]} x_t \xrightarrow{d} N(0, \lambda) = Z(\lambda), \end{aligned}$$

say, where the limit law $Z(\lambda)$ is a zero mean Gaussian distribution with variance λ That is, the apparent "function" $Z_n(\lambda)$ evaluated at a point $\lambda \in [0,1]$ converges to a zero mean Gaussian law $Z(\lambda)$ that is itself evidently a function of λ. That is,

$$Z_n(\lambda) \xrightarrow{d} Z(\lambda) \text{ pointwise on } [0,1].$$

Since Gaussian laws are completely characterized by their mean and variance (Doob 1953), λ completely characterizes this zero mean Gaussian law.

Now, for each n treat

$$\{Z_n(\lambda)\} = \{Z_n(\lambda) : 0 \leq \lambda \leq 1\}$$

as a stochastic process indexed by λ. What are the properties of this process? what are its sample paths (i.e. what does $Z_n(\lambda)$ look for over λ for some random draw $\{x_t\}_{t=1}^n$)? to what does the process $\{Z_n(\lambda)\}$ converge, if anything?

If a process $\{Z_n(\lambda)\}$ has a limit, say $\{Z(\lambda)\}$, it is the *weak limit*[1], or *functional limit*. We denote the fact that $\{Z(\lambda)\}$ is the weak limit of $\{Z_n(\lambda)$ by

$$Z_n(\lambda) \Longrightarrow Z(\lambda).$$

Note this is very different from mere pointwise coivergence: $Z_n(\lambda) \xrightarrow{d} Z(\lambda)$ for each $\lambda \in [0,1]$. There are, unfortunately, cases where pointwise convergence holds, but the process $\{Z_n(\lambda)\}$, a random draw from a distribution of sample paths, does not converge.

It $Z(\lambda)$ has known properties we say $Z_n(\lambda)$ convergence on a space with these properties. For example, if $Z(\lambda)$ is a continuous function of λ then $Z_n(\lambda)$ convergence on $C[0,1]$, the space of continous real functions on $[0,1]$.

As it turns out, for iid, non-iid and degenerate functions of non-iid stochastic processes $\{x_t\}$ the partial sum process $\{Z_n(\lambda)\}$ converges to a limiting process $\{Z(\lambda)\}$ which is identically *Brownian motion*. This includes martingale difference, mixing and NED processes $\{x_t\}$ (Davidson 1994, de Jong and Davidson 2000, Hill 2009).

We need only focus on the iid case, but the reader should consult the literature for related results for martingale differences.

Consider first the definitions of Brownian motion and Wiener measure (Wiener 1923).

DEFINITION: Brownian Motion (Wiener Process) *Brownian motion* $\{Z(\lambda)\} = \{Z(\lambda) : 0 \leq \lambda \leq 1\}$ *is a continuous process on* $[0,1]$ *such that*

i. $Z(0) = 0$ *with probability one;*

ii. $[Z(\lambda_2) - Z(\lambda_1), ..., Z(\lambda_k) - Z(\lambda_{k-1})]$ *is for any* $k \in \mathbb{N}$ *multivariate normally distributed with diagonal covariance matrix for every* $0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_k \leq 1$, *and Gaussian increments* $Z(\lambda) - Z(\lambda') \sim N(0, \lambda - \lambda')\ \forall \lambda > \lambda'$*;*

iii. $Z(\lambda)$ *is continuous in* λ *with probability one.*

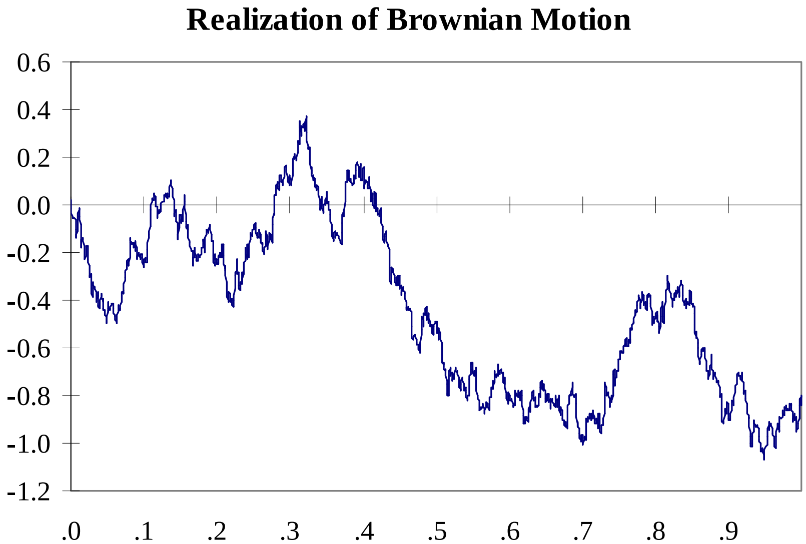
Remark 1: Since $Z(\lambda)$ is an *almost surely* continuous function on $[0,1]$ its sample paths exist in the space $C[0,1]$ with probability one. A sample path is a realization $\{Z(\lambda,\omega) : 0 \leq \lambda \leq 1\}$ where ω is some random event in Υ. See Figure 1 for a realization of Brownian motion.

Remark 2: A multivariate normal has a diagonal covariance matrix *if and only if* its scalar components are independent: the increment $Z(\lambda_i) - Z(\lambda_{i-1})$ is independent of the increment $Z(\lambda_j) - Z(\lambda_{j-1})\ \forall \lambda_{i-1} < \lambda_i < \lambda_{j-1} < \lambda_j$.

[1] Do not confuse this "weak" with consistency, which is often refered to as a "weak probability limit", or weak limit.

Remark 3: The restriction $\lambda \in [0,1]$ is a convention we make use of below. In general λ can be restricted to any compact $A \subset \mathbb{R}$, or $[0,\infty)$, and the representative processes exists on $C[A]$ or $C[0,\infty)$ respectively. Consult Billingsley (1999).

FIGURE 1



Remark: In fact, Figure 1 is only a piece-wise approximation of Brownian motion since what we have plotted is in fact not a truly continuous function (we have only connected the dots of a discrete set of points!). We have plotted

$$x_t = \frac{1}{\sqrt{n}} \sum_{i=0}^{t-1} \epsilon_{t-i} \text{ for } t = 1, ..., 1000, \text{ where } \epsilon_t \overset{iid}{\sim} N(0,1),$$

which can be written as

$$Z_n(\lambda) := \frac{1}{\sqrt{n}} \sum_{i=0}^{[n\lambda]} \epsilon_{t-i} \text{ for } \lambda \in \left\{ \frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n} \right\} = \{.001, .002, ..., 1\}.$$

Notice $Z_n(\lambda)$ satisfies the properties of Brownian motion, *except* continuity: we can only say it is *continuous except at countably many points on* $[0,1]$. As λ smoothly increases $Z_n(\lambda)$ exhibits small jumps as λ crosses the point $1/n$, $2/n$ etc. Asymptotically the jumps vanish with probability one. Further, although there are always these (zero in size with probability one) jumps asymptotically, there are only *countably* infinitely many. That may seem like a lot, but a set with countably many objects, finite or infinite, has measure zero. Asymptotically the probably that a jump occurs is zero. In other words, the sample paths are almost surely continuous, and therefore exist on $C[0,1]$ with probability one. This is enough to ensure a wide range of asymptotic results that we summarize and exploit below.

Wiener measure is simply the probability measure (probability distribution) of Brownian motion $\{Z(\lambda)\}$.

Wiener Measure *Wiener measure* W *is a probabilty measure on* $C[0,1]$ *with the attributes* *i.* $W(Z(0) = 0) = 1$;

ii. $W(Z(\lambda) \leq a) = \int_{-\infty}^{a} (\sqrt{2\pi\lambda})^{-1} \exp\{.5 \times Z(\lambda)^2/\lambda\}$;

iii. If $\{Z(\lambda)\} \sim W$ *then* $Z(\lambda_2) - Z(\lambda_1)$ *is independent of* $Z(\lambda_4) - Z(\lambda_3)$ *for all* $0 \leq \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \leq 1$.

Remark 1: Thus, Wiener measure W is the probablity measure associated with a Wiener process, or Brownian motion.

Remark 2: It is easy to show the Brownian motion definition is now superfluous: any $\{Z(\lambda)\} \sim W$ has exactly the properties defined under Brownian motion. However, the observse is hardly obvious: it is not a simple task proving that Wiener measure exists (i.e. that process $\{Z(\lambda)\}$ has a probability measure assigned to it).

The following functional central limit theorem (FCLT), or weak limit theorem, will allow us to characterize the asymptotic sample paths of $\{Z_n(\lambda)\}$ in Figure 1, and limit law of the $\hat{\rho}$ from Section 6.1.

THEOREM 6.2.1 *Let* $\{x_t\}$ *be iid with zero mean and* $E[x_t^2] = \sigma^2 < \infty$*, and define* $Z_n(\lambda) = 1/\sqrt{n}\sum_{t=1}^{[\lambda n]} x_t$, $\lambda \in [0,1]$*. Then*

$$Z_n(\lambda) \Longrightarrow Z(\lambda).$$

where $Z(\lambda)$ *has almost surely continuous sample paths. In particular,* $Z(\lambda)$ *is Brownian motion.*

Recall the *continuous mapping theorem*: if some $Z_n \xrightarrow{d} Z$ then continuous functions $g(Z_n) \xrightarrow{d} g(Z)$. In fact, the mapping theorem carries over the weak limits (Billingsley 1999):

$$g(Z_n(\lambda)) \Longrightarrow g(Z(\lambda)).$$

We have the following useful result based on Theorem 6.2.1 and a generalization of the mapping theorem

THEOREM 6.2.2 *Consider any function* $g : \mathbb{R} \to \mathbb{R}$ *that as continuous except possibly for countably many points in its support. Under the conditions of Theorem 6.2.1*

$$\int_0^1 g(Z_n(\lambda))\, d\lambda \xrightarrow{d} \int_0^1 g(Z(\lambda))\, d\lambda.$$

6.3 Random Walk with Drift: OLS Asymptotic Theory

Return now to the problem of OLS inference for a unit root process. Assume time begins at $t = 1$: $\epsilon_t = 0$ *a.s.* $\forall t \leq 0$. Simple backward substitution leads to

$$x_t = \sum_{i=0}^{t-1} \epsilon_{t-i} = \sum_{\tau=1}^{t} \epsilon_\tau.$$

We now have a well known result (Phillips 1987).

THEOREM 6.3.1 *Consider the random walk process of section 6.1.The OLS slope estimator satisfies*

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{(1/2) \times \left\{Z(\lambda)^2 - 1\right\}}{\int_0^1 Z(\lambda)^2\, d\lambda}$$

Remark 1: There is one remarkable and one noticeable result: the latter is the limit is obviously non-Gaussian; the former is convergence is now at rate n. Thus, $\hat{\rho} \xrightarrow{p} 1$ at an extraordinary rate[2].

[2] In fact this is not too surprizing: probabilty convergence of the OLS estimator is super-consistent

APPENDIX : SELECTED PROOFS

CHAPTER 1

Proof of Lemma 1.2.1. See Davidson (1994). ■

Proof of Theorem 1.3.1.1. Clearly

$$\begin{aligned} E\left|x_t\right|^{1+\delta} &= E\left|x_t\right|^{1+\delta} I\left(\left|x_t\right| \geq M\right) + E\left|x_t\right|^{1+\delta} I\left(\left|x_t\right| < M\right) \\ &\geq E\left|x_t\right|^{1+\delta} I\left(\left|x_t\right| \geq M\right) \geq M^{\delta} \times E\left|x_t\right| I\left(\left|x_t\right| \geq M\right). \end{aligned}$$

Therefore,

$$E\left|x_t\right| I\left(\left|x_t\right| \geq M\right) \leq E\left|x_t\right|^{1+\delta} / M^{\delta} \to 0$$

as $M \to \infty$ since $E\left|x_t\right|^{1+\delta} < \infty$ by $L_{1+\delta}$-boundedness. ■

CHAPTER 2

Proof of Theorem 2.2.2.1. See Theorem 13.13 of Davidson (1994). ■

Proof of Theorem 2.2.2.5. See Corollary 13.14 of Davidson (1994). ■

Proof of Theorem 2.2.2.6. Use covariance stationarity and Chebyshev's inequality to write to any $\varepsilon > 0$

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{t=1}^{n}\left\{x_t - E\left[x_t\right]\right\}\right| > \varepsilon\right) &\leq E\left(\frac{1}{n}\sum_{t=1}^{n}\left\{x_t - E\left[x_t\right]\right\}\right)^2 \varepsilon^{-2} \\ &= \varepsilon^{-2} \times \frac{1}{n^2}\sum_{s,t=1}^{n} cov(x_s, x_t) \\ &= \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n}\sum_{i=1}^{n-1}(1 - i/n) \times cov(x_{i+1}, x_1)\right) \\ &\leq \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n}\sum_{i=1}^{n}\left|cov(x_{i+1}, x_1)\right|\right), \end{aligned}$$

since $1 - i/n \in [0, 1]$. Covariance stationarity implies $\sigma^2 < \infty$, hence $1/n \sum_{i=1}^{n} |cov(x_{i+1}, x_1)| \to 0$ suffices for

$$P\left(\left|\frac{1}{n}\sum_{t=1}^{n}\left\{x_t - E\left[x_t\right]\right\}\right| > \varepsilon\right) \to 0.$$

Now, absolute summability $\sum_{i=0}^{\infty} |cov(x_{i+1}, x_1)| < \infty$ implies $cov(x_{N+1}, x_1) \to 0$ as $N \to \infty$. But if $\{cov(x_{i+1}, x_1)\}$ converges to zero, then the Cesàro sum satisfies $1/n \sum_{i=1}^{n} |cov(x_{i+1}, x_1)| \to 0$ (Davidson 1994: Theorem 2.26). ■

Proof of Lemma 2.2.2.9. Since $x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$ $a.s.$ has a zero mean and $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$, the autocovariances $\gamma(t-s) := E[x_s x_t]$, $t \geq s$, are

$$\begin{aligned} \sum_{i,j=0}^{\infty} \phi^i \phi^j E\left[\epsilon_{s-i}\epsilon_{t-j}\right] &= \sigma^2 \sum_{j=i+t-s}^{\infty} \phi^i \phi^j = \sigma^2 \sum_{i=0}^{\infty} \phi^i \phi^{i+t-s} \\ &= \sigma^2 \phi^{t-s} / \left(1 - \phi^2\right), \end{aligned}$$

(i.e. faster than $\sqrt{n}$) for a variety of infinite variance processes. The unit root process is simply one type of infinite variance process.

which only depends on displacement $h := t - s$, and not s and t specifically. Hence $\gamma(h) = \sigma^2 \phi^h / (1 - \phi^2)$, and

$$\sum_{i=1}^{\infty} |\gamma(h)| = \frac{\sigma^2}{1-\phi^2} \sum_{i=1}^{\infty} |\phi|^h = \frac{\sigma^2}{(1-\phi^2)(1-|\phi|)} < \infty.$$

Absolutely summable autocovariances implies ergodicity of the mean, cf. Theorem 2.2.2.6.

■

CHAPTER 3

Proof of Lemma 3.1.5.1. See Ibragimov (1962), and Theorem 14.2 and Corollary 14.3 of Davidson (1994). ■

Proof of Lemma 3.1.5.2. See Serfling (1968), and Theorem 14.4 and Corollary 14.5 of Davidson (1994). ■

Proof of Theorem 3.1.5.3. Geometric strong mixing and implies $L_{2+\delta}$-boundedness imply for some $K > 0$ and tiny $\iota > 0$

$$|Cov(x_t, x_{t-m})| \leq K \alpha_m^{\iota} \leq K \rho^{\iota m}.$$

Therefore

$$\sum_{m=0}^{\infty} |Cov(x_t, x_{t-m})| \leq K (1 - \rho^{\iota})^{-1} < \infty.$$

Now apply Theorem 2.2.2.6. ■

Proof of Lemma 3.1.5.4. See Theorem 14.1 of Davidson (1994). ■

Proof of Theorem 3.2.1.1. Apply Theorem 2.2.2.6. ■

Proof of Lemma 3.3.2.1. Define $\digamma_t := \sigma(\epsilon_\tau : \tau \leq t)$ and note

$$E\left[x_t | \digamma_{t-m}^{t+m}\right] = E\left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} | \digamma_{t-m}^{t}\right] = \sum_{i=0}^{m} \psi_i \epsilon_{t-i} + \sum_{i=m+1}^{\infty} \psi_i E\left[\epsilon_{t-i} | \digamma_{t-m}^{t+m}\right].$$

Use Minkowski's and the conditional Jensen's inequalities to deduce

$$\begin{aligned} \left\|x_t - E\left[x_t | \digamma_{t-m}^{t+m}\right]\right\|_2 &= \left\|\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} - \sum_{i=0}^{m} \psi_i \epsilon_{t-i} - \sum_{i=m+1}^{\infty} \psi_i E\left[\epsilon_{t-i} | \digamma_{t-m}^{t+m}\right]\right\|_p \\ &= \left\|\sum_{i=m+1}^{\infty} \psi_i \left(\epsilon_{t-i} - E\left[\epsilon_{t-i} | \digamma_{t-m}^{t+m}\right]\right)\right\|_p \\ &\leq \sum_{i=m+1}^{\infty} |\psi_i| \times \|\epsilon_{t-i}\|_p \leq \sup_{s \in \mathbb{Z}} \|\epsilon_s\|_p \times \sum_{i=m+1}^{\infty} |\psi_i| = d_t \times \varphi_m, \end{aligned}$$

say, where $d_t = \sup_{s \in \mathbb{Z}} ||\epsilon_s||_p < \infty \; \forall t$. Clearly $\sum_{i=0}^{\infty} |\psi_i| < \infty$ implies $\sum_{i=m+1}^{\infty} |\psi_i| \to 0$ as $m \to \infty$. If $\psi_i = O(i^{-1-\lambda-\iota})$ then

$$m^{\lambda} \varphi_m \leq K \sum_{i=m+1}^{\infty} i^{-1-\iota} < \infty$$

hence $m^\lambda \varphi_m \to 0$ which implies the NED size is $\lambda > 0$. ■

Proof of Lemma 3.3.3.2. See Theorem 17.9 of Davidson (1994). ■

Proof of Lemma 3.3.3.1. Minkowski's inequality and the NED property imply

$$\begin{aligned}\left\|\sum_{i=1}^k \psi_i x_{i,t} - E\left[\sum_{i=1}^k \psi_i x_{i,t}|\digamma_{t-m}^{t+m}\right]\right\|_p &= \left\|\sum_{i=1}^k \psi_i \left(x_{i,t} - E\left[x_{i,t}|\digamma_{t-m}^{t+m}\right]\right)\right\|_p \\ &\leq \sum_{i=1}^k |\psi_i| \left\|\left(x_{i,t} - E\left[x_{i,t}|\digamma_{t-m}^{t+m}\right]\right)\right\|_p \\ &\leq \sum_{i=1}^k |\psi_i| d_{i,t}\varphi_{i,m} \leq \left(\sum_{i=1}^k |\psi_i| d_{i,t}\right)\left(\sum_{i=1}^k \varphi_{i,m}\right) \\ &= d_t \times \varphi_m.\end{aligned}$$

■

CHAPTER 4

Proof of Theorem 4.1.1. Iterated expectations and the *mds* property imply for all $h \geq 1$

$$\begin{aligned}\gamma(h) &= E\left[x_t x_{t-h}\right] = E\left(x_{t-h} E\left[x_t|\Im_{t-h}\right]\right) \\ &= E\left(x_{t-h} E\left(E\left[x_t|\Im_{t-1}\right]|\Im_{t-h}\right)\right) = E\left(x_{t-h} E\left(0|\Im_{t-h}\right)\right) = 0\end{aligned}$$

Then by assumption

$$\frac{1}{n^2}\sum_{t=1}^n \sigma_t^2 + \frac{2}{n}\sum_{i=1}^{n-1}(1 - i/n) \times \gamma(i) = \frac{1}{n^2}\sum_{t=1}^{n-1}\sigma_t^2 = o(1).$$

■

Proof of Lemma 4.2.1. See Theorem 15.14 of Davidson (1994). ■

Proof of Lemma 4.2.2. See Corollary 20.2 of Davidson (1994). ■

Proof of Theorem 4.2.3. Define $T_m := \sum_{t=1}^m x_t/a_t$ and note $\{T_{n,m} := T_n - T_m, \Im_n\}$ is a martingale since

$$\begin{aligned}E\left[T_{n,m}|\Im_{n-1}\right] &= E\left(\sum_{t=m+1}^n x_t/a_t|\Im_{n-1}\right) \\ &= \sum_{t=m+1}^n E\left[x_t|\Im_{n-1}\right]/a_t = \sum_{t=m+1}^{n-1} x_t/a_t = T_{n-1,m}.\end{aligned}$$

Moreover, use the *mds* property and iterated expectations to get

$$E\left(T_n - T_m\right)^2 = \sum_{t=m+1}^n \sigma_t^2/a_t^2.$$

Lemma 4.2.1 implies

$$P\left(\max_{1\leq k\leq n}|T_k - T_m| > \varepsilon\right) \leq \frac{E\left(T_n - T_m\right)^2}{\varepsilon^2} = \varepsilon^{-2}\sum_{t=m+1}^{n}\sigma_t^2/a_t^2.$$

Now apply Lemma 4.2.2 with $c_t^2 = \sigma_t^2/a_t^2$ to deduce

$$T_n = \sum_{t=1}^{m} x_t/a_t \overset{a.s.}{\to} T,$$

for some T. Kronecker's lemma[3] completes the proof: if $\sum_{t=1}^{m} x_t/a_t \overset{a.s.}{\to} T$ then $\sum_{t=1}^{m} x_t/a_n \overset{a.s.}{\to} 0$. ■

Proof of Theorem 4.2.4. See Theorem 20.11 of Davidson (1994). ■

Proof of Lemma 4.3.1. We need only show $1/n\sum_{i=1}^{n} B_i \to 0$. But this follows from Kronecker's lemma: $\sum_{i=1}^{\infty} i^{-1}B_i < \infty \Longrightarrow 1/n\sum_{i=1}^{n} B_i \to 0$. ■

Proof of Theorem 4.3.2. The result follows from Lemmas 3.1.5.1 and 4.3.1. ■

Proof of Theorem 4.4.1. Under conditions (a) and (b), Theorem 17.5 of Davidson (1994) implies $\{x_{n,t}, \digamma_t\}$ forms an L_2-mixingale array of size $1/2$ with mixingale constants $c_{n,t} \leq \max\left\{\|x_{n,t}\|_r, d_{n,t}\right\}$. Therefore (see McLeish 1975)

$$E\left(\sum_{t=1}^{n} x_{n,t}\right)^2 = O\left(\sum_{t=1}^{n} c_{n,t}^2\right) = O\left(\sum_{t=1}^{n}\left(\max\left\{\|x_{n,t}\|_r, d_{n,t}\right\}\right)^2\right)$$

Under condition (c)

$$E\left(\sum_{t=1}^{n} x_{n,t}\right)^2 = O(1).$$

But this implies for any sequence of positive numbers $\{a_n\}$, $a_n \nearrow \infty$,

$$P\left(\left|\frac{1}{a_n}\sum_{t=1}^{n} x_{n,t}\right| > \varepsilon\right) \leq \varepsilon^{-2}E\left(\frac{1}{a_n}\sum_{t=1}^{n} x_{n,t}\right)^2 = O\left(1/a_n^2\right) = o(1).$$

■

Proof of Corollary 4.4.2. We need only verify the conditions of Theorem 4.4.1 are satisfied for

$$x_{n,t} := \frac{1}{\sqrt{n}}\left(x_t - E\left[x_t\right]\right).$$

Clearly $E\left[x_{n,t}\right] = 0$ and $x_{n,t}$ is L_r-bounded by Minkowski's and Liapanov's inequalities:

$$\|x_{n,t}\|_r \leq \frac{1}{\sqrt{n}}\left(\|x_t\|_r + \|x_t\|_1\right) \leq \frac{2}{\sqrt{n}}\|x_t\|_r < \infty.$$

Therefore condition (a) of Theorem 4.4.1 is satisfied.

[3]See Davidson (1994: Lemma 2.35).

Notice $\{x_{n,t}\}$ is L_2-NED of size $1/2$ on $\{\epsilon_t\}$ with constants $d_{n,t} = d/\sqrt{n}$ and coefficients φ_m since

$$\left\|x_{n,t} - E\left[x_{n,t}|\mathcal{F}_{t-m}^{t+m}\right]\right\|_2 = \left(1/\sqrt{n}\right)\left\|x_t - E\left[x_t|\mathcal{F}_{t-m}^{t+m}\right]\right\|_2 \leq \left(d/\sqrt{n}\right)\varphi_m.$$

Therefore condition (b) of Theorem 4.4.1 is satisfied.

Strict stationarity and L_r-boundedness therefore imply

$$\begin{aligned}\sum_{t=1}^{n}\left(\max\left\{\|x_{n,t}\|_r, d_{n,t}\right\}\right)^2 &\leq \sum_{t=1}^{n}\left(\max\left\{\left\|\left(1/\sqrt{n}\right)\left(x_t - E\left[x_t\right]\right)\right\|_r, \left(d/\sqrt{n}\right)\right\}\right)^2 \\ &\leq K\max\left\{\|x_t\|_r, d\right\}\sum_{t=1}^{n} 1/n \\ &= K\max\left\{\|x_t\|_r, d\right\} < \infty,\end{aligned}$$

where the equality holds uniformly in t. Thus condition (c) of Theorem 4.4.1 is satisfied.

■

CHAPTER 5

Proof of Corollary 5.2.4. Borrowing arguments from the proofs of Theorem 4.4.1 and Corollary 4.4.2, $\{1/\sqrt{n}(x_t - E[x_t])\}$ is L_2-NED of size $1/2$ on $\{\mathcal{F}_t\}$ with constants $\tilde{d}_{n,t} = d/\sqrt{n}$ and coefficients φ_m. Since the base is iid it is trivially mixing, hence from Theorem 17.5 of Davidson (1994) $\{1/\sqrt{n}(x_t - E[x_t]), \mathcal{F}_t\}$ forms an L_2-mixingale array with constants

$$\tilde{c}_{n,t} \leq \max\left\{\|x_{n,t}\|_r, \tilde{d}_{n,t}\right\} \leq K/\sqrt{n}$$

uniformly in t. Therefore (see McLeish 1975)

$$E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(x_t - E\left[x_t\right]\right)\right)^2 = O\left(\sum_{t=1}^{n} c_{n,t}^2\right) = O\left(\sum_{t=1}^{n} 1/n\right) = O(1).$$

Along with $\liminf_{n\geq 1} E(1/\sqrt{n}\sum_{t=1}^{n}(x_t - E[x_t]))^2 > 0$ if follows that

$$x_{n,t} = \frac{1}{\sqrt{n}}\left(\frac{x_t - E\left[x_t\right]}{E\left(1/\sqrt{n}\sum_{t=1}^{n} x_t\right)^2}\right)$$

is a well defined triangular array.

All of the conditions of Theorem 5.2.3 are satisfied by putting $\tilde{c}_{n,t} = K/\sqrt{n}$ for any finite $K > 0$. For example, obviously $\{x_{n,t}/c_{n,t}\}$ is L_r-bounded since

$$\begin{aligned}\left\|\frac{x_{n,t}}{c_{n,t}}\right\|_r &= K\left\|\left(\frac{x_t - E\left[x_t\right]}{E\left(1/\sqrt{n}\sum_{t=1}^{n} x_t\right)^2}\right)\right\|_r \\ &\leq K\left\|x_t - E\left[x_t\right]\right\|_r \leq K\left\|x_t\right\|_r,\end{aligned}$$

where the first inequality follows from $\liminf_{n\geq 1} E(1/\sqrt{n}\sum_{t=1}^{n}(x_t - E[x_t]))^2 > 0$, and the second from Minkowski's and Liapanov's inequalities. ■

Proof of Theorem 5.3.1. See Theorems 25.5 and 25.6 of Davidson (1994). ■

CHAPTER 6

Proof of Theorem 6.3.1. Write

$$n\left(\hat{\rho}-1\right)=\frac{1/n\sum_{t=2}^{n}\epsilon_t x_{t-1}}{1/n^2\sum_{t=2}^{n}x_{t-1}^2}.$$

We need only prove

$$\frac{1}{n}\sum_{t=2}^{n}\epsilon_t x_{t-1}\xrightarrow{d}\sigma^2\frac{1}{2}\left\{Z\left(\lambda\right)^2-1\right\}\quad\text{and}\quad\frac{1}{n^2}\sum_{t=2}^{n}x_{t-1}^2\xrightarrow{d}\sigma^2\int_0^1 Z\left(\lambda\right)^2 d\lambda.$$

Step 1: Since $x_t^2=(x_{t-1}+\epsilon_t)^2=x_{t-1}^2+2x_{t-1}\epsilon_t+\epsilon_t^2$, we may write

$$x_{t-1}\epsilon_t=\frac{1}{2}\left(x_t^2-x_{t-1}^2\right)-\frac{1}{2}\epsilon_t^2$$

hence

$$\sum_{t=2}^{n}\epsilon_t x_{t-1}=\frac{1}{2}\sum_{t=2}^{n}\left(x_t^2-x_{t-1}^2\right)-\frac{1}{2}\sum_{t=1}^{n}\epsilon_t^2=\frac{1}{2}\left(x_n^2-\sum_{t=1}^{n}\epsilon_t^2\right).$$

Since $\epsilon_t\overset{iid}{\sim}(0,\sigma^2)$ was easily have $1/n\sum_{t=1}^{n}\epsilon_t^2\xrightarrow{p}\sigma^2$. Consider the remaining term

$$x_n^2=\left(\sum_{\tau=1}^{n}\epsilon_\tau\right)^2.$$

By the central limit theorem for iid random variances, and the continuous mapping theorem,

$$\frac{1}{n}x_n^2=\sigma^2\left(\frac{1}{\sqrt{n}}\sum_{\tau=1}^{n}\epsilon_\tau/\sigma\right)^2\xrightarrow{d}\sigma^2\times Z\left(1\right)^2,$$

where $Z\left(1\right)\sim N(0,1)$. Therefore

$$\frac{1}{n}\sum_{t=2}^{n}\epsilon_t x_{t-1}\xrightarrow{d}\frac{1}{2}\left\{\sigma^2\times Z\left(1\right)^2-\sigma^2\right\}=\sigma^2\frac{1}{2}\left\{Z\left(1\right)^2-1\right\}.$$

Step 2: In order to prove $1/n^2\sum_{t=2}^{n}x_{t-1}^2\xrightarrow{d}\sigma^2\int_0^1 Z\left(\lambda\right)^2 d\lambda$, where $\{Z\left(\lambda\right):0\leq\lambda\leq 1\}$ is Brownian motion, simply note $1/n\sum_{t=2}^{n}(x_{t-1}^2/n)$ is an approximate integral of x_{t-1}^2/n, and for any t by backward substitution

$$x_t/\sqrt{n}=\frac{1}{\sqrt{n}}\sum_{\tau=1}^{[n\times t/n]}\epsilon_\tau.$$

The limit of $1/n\sum_{t=2}^{n}(x_{t-1}^2/n)$ is easy to obtain if we define a piece-wise process

$$\begin{aligned}S_n\left(\lambda\right)&=0\ \text{ if }0\leq\lambda<1/n\\&=x_1^2/n\ \text{ if }\ 1/n\leq\lambda<2/n\\&\quad...\\&=x_{n-1}^2/n\ \text{ if }\ (n-1)/n\leq\lambda<n/n=1\\&=x_n^2/n\ \text{ if }\lambda=1.\end{aligned}$$

By construction, therefore,

$$\int_0^1 S_n\left(\lambda\right)d\lambda=\frac{1}{n}\sum_{t=2}^{n}x_{t-1}^2/n=\frac{1}{n^2}\sum_{t=2}^{n}x_{t-1}^2.$$

Further, by Theorems 6.2.1 and 6.2.2 we have weak convergence on $C[0,1]$:

$$S_n(\lambda) = \sigma^2 \times \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[n\lambda]} \epsilon_t / \sigma \right)^2 \Longrightarrow \sigma^2 \times Z(\lambda)^2 .$$

Now invoke the continuous mapping theorem to conclude

$$\int_0^1 S_n(\lambda)\, d\lambda \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2\, d\lambda.$$

■

REFERENCES

[1] An, H.Z. and F.C. Huang (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models, *Statistica Sinica* 6, 943-956.

[2] Andrews, D.W.K. (1984). Non-Strong Mixing Autoregressive Processes, J*ournal of Applied Probability* 21, 930-934.

[3] Billingsely, P.J. (1999). *Convergence of Probability Measure*. Wiley: New York.

[4] Davidson, J. (1994). *Stochastic Limit Theory.* Oxford Univ. Press: Oxford.

[5] de Jong, R.M. (1997). Central Limit Theorems for Dependent Heterogeneous Random Variables, *Econometric Theory* 13, 353-367.

[6] Davidson, J., and R.M. de Jong (2000). The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals, *Econometric Theory* 16, 621-642.

[7] Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.

[8] Guegan D., and S. Ladoucette (2001). Non-mixing Properties of Long Memory Processes, *Comptes Rendus de l'Academie des Sciences Series I Mathematics* 333, 373-376.

[9] Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogenous Arrays with Applications to Tail Index and Tail Dependence Estimation, *Journal of Statistical Planning and Inference*: in press.

[10] Ibragimov, I.A. (1962). Some Limit Theorems for Stationary Processes, *Theory of Probability and its Applications* 7, 349-382.

[11] Leibscher, E. (2005). Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes, *Journal of Time Series Analysis* 26, 669-689.

[12] Lévy, P. (1925). *Calcul de Probabilités.* Paris: Gauthier-Villars.

[13] Lévy, P. (1954). *Théorie de l'Addition des Variables Eléatoires.* Paris: Gauthier-Villars.

[14] McLeish, D.L. (1974). Dependent Central Limit Theorems, *Annals of Probability* 2, 620-628.

[15] McLeish, D.L. (1975). A Maximal Inequality and Dependent Strong Law, *Annals of Probability* 3, 329-339.

[16] Phillips, P.C.B. (1987). Time Series Regression with a Unit Root, *Econometrica* 55, 277-301.

[17] Serfling, R.J. (1968). Contributions to Central Limit Theory for Dependent Variables, *Annals of Mathematical Statistics* 39, 1158-1175.

[18] Wiener, N. (1923). Differential Space, *Journal of Mathematic Physics* 2, 131-174.

[19] White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando Fla.: Academic Press.