Time Series: Given joint ergodic stationarity between Y and X, when we have serially correlated errors, our consistency results do not change. However, our asymptotic distribution results do - we need to generalize CLT's to obtain asymptotic normality for serially correlated processes with the "right" variance. (The generalization is possible under certain conditions by restricting the degree of serial correlation, and the condition is transparent for linear stochastic processes)

Probability Measure, Sigma Algebra (Sigma Algebra is what we define our measures on), Borel Fields

Def: The set S of all possible outcomes of a particular experiment is called the **sample space** of the experiment.

Def: A collection of subsets of S, denoted *B*, is called a **sigma field** or **sigma algebra** if it satisfies the following:

1. $Empty\ Set: \varnothing \in \beta$
2. $Complements: If\ \ A \in \beta,\ then\ S \setminus A = A^c \in \beta$
3. $Unions: If\ \ A_1, A_2, ... \in \beta,\ then\ \left(\bigcup_{i=1} A_i\right) \in \beta$

$$Pf: If\ \ A_1, A_2, ... \in \beta,\ then\ \ clearly\ \left(\bigcap_{i=1}^{\infty} A_i\right) \in \beta \Rightarrow \left(\bigcap_{i=1}^{\infty} A_i\right)^c \in \beta\ by\ 2 \Rightarrow \left(\bigcup_{i=1}^{\infty} A_i^c\right) \in \beta\ by\ De\ Morgan's\ Laws$$

Def: P is a **probability measure** on the pair (S,*B*), if P satisfies:

1. $P(A) \ge 0\ \ for\ all\ \ A \in \beta$
2. $P(S) = 1$
3. $If\ \ A_1, A_2, ... \in \beta\ are\ pairwise\ disjoint,\ P\left(\bigcup_{i=1} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Def: Let X: (Ω,*F*) → (*R*, *B*) be *F* measurable. A Borel field is the smallest σ-field that makes X measurable, given by:

$$\sigma(X) \equiv \left\{G \subseteq \Omega : G = X^{-1}(B)\ \ for\ \ some\ \ B \in \beta\right\}$$

(Think of this is the only sets in the universe that the random variables gives us information about – since they are the sets that are preimages of all the possible outcomes of the r.v. So, the random variables **X is informative about members of σ(X) but not more than that!**)

Probability Space, Random Variables, and Measurability

Def: The triple (Ω,*F*,P) is called a **probability space**,
where Ω is the "universe" (or the whole set of outcomes, like S), *F* is the σ-field on Ω (like *B*), and *P* is the underlying probability measure that governs all random variables, i.e. a probability measure on (Ω,*F*)

Def: A **random variable** is a function from the sample space into the real numbers, or a measurable mapping from (Ω,*F*) into (*R*, *B*)
(So, for a random variable X: (Ω,*F*) → (*R*, *B*), the **sample space** for X is *R*)

Def : A random variable X: (Ω,*F*) → (*R*, *B*) is ***F*-measurable** if the preimage $\{\omega \in \Omega : X(\omega) \in \beta\} \in F\ \ \ for\ all\ \ B \in \beta$
(all the events in *B* can be mapped back to *F* and be measured there)

Note: X(ω) is a random variable that induces a probability measure P_X on (*R*, *B*), ω ε Ω (the universe)
P_X is defined from P (a probability measure on (Ω,*F*)) by

$$\Pr\ X\ takes\ on\ values\ in\ B : P_X(B) \equiv P(X \in B) = P\left(\{\omega \in \Omega : X(\omega) \in B\}\right)\ for\ some\ B \in \beta$$

Def: A random variable Y = g(X): (R_X, B_X)→ (R_Y, B_Y) induces the probability measure P_Y on the sample space R_Y as follows:

$$for\ some\ A \in B_Y,\ P_Y(A) \equiv P(Y \in A) = P\left(X \in \{x \in R_X : Y = g(x) \in A\}\right) = P_X\left(\{x \in R_X : Y = g(x) \in A\}\right)$$

Prop: Let *F* and *G* be 2 σ-fields s.t. *G* c *F* (all the sets in *G* are also in *F*). If a random variable X is *G*-measurable, then X is *F*-measurable[1].

Conditional Expectations and Law of Iterated Expectations

Def: Let X and Y be real-valued random variables on (Ω,*F*,P) and let G = σ(X). Suppose E|Y| finite. The conditional expected value of Y given X is a random variable (function of X) that satisfies the following 3 conditions[2]:

1. $E\left|E(Y \mid X)\right| < \infty$
2. $E(Y \mid X)\ is\ G-measurable: i.e.\ \forall B \in \beta,\ \{\omega \in \Omega : E(Y \mid X)(\omega)\} \in \sigma(X)\ \ (E(Y \mid X)\ is\ as\ informative\ as\ X\ but\ no\ more\ sophisticated\ than\ X)$
3. $For\ all\ \ g \in G,\ \int_g E(Y \mid X)(\omega)\ dP(\omega) = \int_g Y(\omega)\ dP(\omega)$

---

[1] $Pf: \forall B \in \beta, \{\omega \in \Omega : X(\omega) \in B\} \in G \subseteq F$

[2] Y always satisfies 1 and 3. But Y will only satisfy 2 if σ(Y) c σ(X) i.e. Y is no more informative than X. So typically not possible to use Y as E(Y|X).

Alternative representation of E(Y|X) and usefulness:

E(Y|X) = E(Y|σ(X)) = E(Y|G)

→ We do this bc when X takes on certain values, it maps to values in the preimage or equivalently the Borel field.

Example: Let E(Y|X) = E(Y|σ(X)) = E(Y|G) and E(Y|X,Z) = E(Y|σ(X,Z)) = E(Y|H)

Since G ⊆ H, then E(E(Y|X,Z)) = E(E(Y|H)) = E(E(Y|G) | H) = E(Y|G) = E(Y|X)

Law of Iterated Expectations: $E(Y) = E\left[E_X(Y \mid X)\right]$ [3]

Generalized Law of Iterated Expectations: For G ⊆ H (G is a less fine partition than H),

$$E(Y \mid G) = E\left[E(Y \mid H) \mid G\right] = E\left[E(Y \mid G) \mid H\right]$$ [4]

Property of Conditional Expectation: For real-valued random variables, Y and X, we have E(YX|X) = E(Y|X)X

Basic Time Series / Stochastic process Concepts:

Stochastic process : a sequence of random variables $\{z_i\}$

Realization/Sample Path of a stochastic process: a realization of $\{z_i\}$ → sequence of real numbers

Time Series: If the sequence of rv's is indexed by time, the stochastic process is called a time series. (we often will use "time series" to refer to the realization and the stochastic process)

Ensemble Mean: $\{\bar{z}_i\}$ ("true" mean of each of the rv's in the sequence)

Important Point: We can always think of our data $\{z_i\}$ as a sample path of an infinite sequence of random variables $\{\ldots, Z_{-1}, Z_0, Z_1, Z_2, \ldots\}$, where the data is a realization of a particular subsequence of the infinite sequence.

What can we say about the underlying random variables that generate our data/time series:

Strictly Stationary Processes[5]: A stochastic process/sequence of RV's (e.g. a time series) $\{z_i\}$ $(i = 1,2...)$ is (strictly) stationary if

$$F(Z_t, \ldots, Z_{t+K}) \text{ does not depend on } t \text{ for all } K = \{0,1,2,\ldots\} \Leftrightarrow F(Z_1,\ldots,Z_K) = F(Z_{t+1},\ldots,Z_{t+K}) \text{ for all } t \text{ and } K$$

→ **F(Z_t) does not depend on t! (all observations come from same distribution. Though this says identical but not necessarily independent)**

Prop: If $\{z_i\}$ $(i = 1,2...)$ (strictly) stationary, then $\{g(z_i)\}$ $(i = 1,2...)$ (strictly) stationary for some cont. function g[6].

Weakly Stationary Process: A stochastic process/sequence of RV's $\{z_i\}$ $(i = 1,2...)$ is weakly (or covariance) stationary if:

i. E($\mathbf{z}_i$) does not depend on *i* and
ii. Cov($\mathbf{z}_i$,$\mathbf{z}_{i-j}$) exists, is finite, and depends only on *j* but not on *i* (e.g. Cov($\mathbf{z}_1$,$\mathbf{z}_5$) =Cov($\mathbf{z}_{12}$,$\mathbf{z}_{16}$))

White Noise Processes[7]: A covariance stationary process {$\mathbf{z}_i$} is a **white noise** if

i. E($\mathbf{z}_i$) = 0
ii. Cov($\mathbf{z}_i$,$\mathbf{z}_{i-j}$) = 0 for j != 0

Note: This is an important class of weakly stationary processes. It is a process with **0-mean and no serial correlation**.

---

[3] By condition 3 in the definition of conditional expectation, since E(Y|X) is clearly Ω-measurable,

$$\text{for } \Omega \in \Omega,\ E(E(Y \mid X)) = \int_{\Omega} E(Y \mid X)(\omega)\ dP(\omega) = \int_{\Omega} Y(\omega)\ dP(\omega) = E(Y)$$

[4] So the usual law of iterated expectations is a special case where G = $\{\Omega, \varnothing\}$ because E(Y|G) = E(Y) in this case. Remember, E(Y) is just taking expectation over the trivial sigma field.

[5] Note: So, the joint distribution of $(z_i, z_{i_1}, z_{i_2}, \ldots, z_{i_r})$ depends only on (for any given finite integer r and any set of subscripts $i_1,\ldots,i_r$) i_1-*i*, i_2-*i*, …,i_r-*i* but not *i*. (e.g. joint distribution of ($\mathbf{z}_1$,$\mathbf{z}_5$) is same as ($\mathbf{z}_{12}$,$\mathbf{z}_{16}$) → what matters is the relative position in the sequence not the absolute position!)

[7] A white noise process that is not strictly stationary: Let *w* ~ unif(0,2π) and define z_i = cos(i*w*) for i = 1,2,… Then, E(z_i) = 0, Var(z_i) = ½, and Cov(z_i, z_j) = 0. So {z_i} is white noise process, but it is clearly not an independent (since each depends on w) white noise process or strictly stationary (since it depends on i).
(WHY????)

$$F_{Zi}(z) = P(Z_i \ge z) = P(Cos(iw) \ge z) = P(w \ge [Cos^{-1}z]/i) = \frac{2\pi - [Cos^{-1}z]/i}{2\pi} = 1 - [Cos^{-1}z]/(i2\pi) \text{ by uniform}$$

$$f_{Zi}(z) = \frac{\partial}{\partial z}\left[1 - [Cos^{-1}z]/(i2\pi)\right] = \frac{1}{(i2\pi)\sqrt{1-z^2}} \quad \text{since } \frac{\partial}{\partial z}\left[Cos^{-1}z\right] = \frac{1}{\sqrt{1-z^2}}$$

$$\text{For a fixed } i,\ E(Z_i) = \int_{z=-1}^{1} z f_Z(z)\,dz = \frac{1}{(i2\pi)}\int_{z=-1}^{1}\frac{z}{\sqrt{1-z^2}}dz = -\frac{1}{2(i2\pi)}\int_{z=-1}^{1}\frac{-2z}{\sqrt{1-z^2}}dz = -\frac{1}{2(i2\pi)}\left[\sqrt{1-z^2}\right]_{-1}^{1} = 0$$

Independent White Noise Process: An independently and identically distributed (iid) sequence with 0-mean and finite variance is a special case of a white noise process. (but generally, elements of white noise process may not come from the same distribution and may not be independent from each other. Independence and identical distribution are much stronger assumptions!)

Ergodic Process[8]: A stationary process $\{z_i\}$ is said to be ergodic if, for any two bounded functions: $f:\Re^K \to \Re,\ g:\Re^L \to \Re$

$$\lim_{n\to\infty}\left|E\left(f(z_i,...,z_{i+k})g(z_{i+n},...,z_{i+n+l})\right)\right| = \left|E\left(f(z_i,...,z_{i+k})\right)\right|\left|E\left(g(z_{i+n},...,z_{i+n+l})\right)\right|$$

(This requires independence between elements sufficiently far apart→ i.e. ($z_i,\ldots,z_{i+k}$) ind of ($z_{i+n},\ldots,z_{i+n+l}$) for n large)
(Intuitively, ergodicity says that the process is not "too persistent")

Heuristically, **a stationary process is ergodic (i.e. ergodic stationarity) if it is asymptotically independent** → i.e. any 2 rv's or positioned far apart in the sequence are almost independently distributed.

Ergodic stationarity is important in developing large sample theory because of the ergodic theorem. **Ergodic Theorem**:

Let $\{z_i\}_{i=1}^n$ be a stationary and ergodic process with $E(z_i)=\mu$. Then, $\boxed{\bar{z}_n \equiv \frac{1}{n}\sum_{i=1}^n z_i \xrightarrow{a.s.} \mu}$

Idea: This generalizes Kolmogorov's LLN[9]. **Ergodic theorem allows for serial dependence** (whereas Kol's rules it out by iid assumption), **provided that the serial dependence disappears in the "long-run" (by stationary ergodic), i.e. asymptotically in large samples.**
Implication: **Any moment of a stationary and ergodic process (if exists and finite) is consistently estimated by the sample moment.** [10] **(iid is a special case!)**

Martingale & Martingale Difference

Martingales Vector Process: A vector process $\{g_i\}$ is called a martingale if $E(g_i \mid g_{i-1},...,g_1) = g_{i-1}\ \ for\ \ i\ge 2$
Note: The conditioning set $g_{i-1},...,g_1$ is often called the information set, and $\{g_i\}$ is called a martingale since its information set is its own past values.

Martingale Difference Sequence: A vector process $\{g_i\}$ with $E(g_i)=0$ is called a margingale difference sequence (m.d.s.) or martingale difference if the expectation conditional on its past values is also 0: $\boxed{E(g_i \mid g_{i-1},...,g_1)=0\ \ for\ \ i\ge 2}$
Implication: A martingale difference sequence has no serial correlation Cov(g_i,g_{i-j}) = 0 for *i* and $j\neq 0$ [11]

Ergodic Stationary Martingale Differences CLT: Let $\{g_i\}$ be a vector martingale difference sequence that is stationary and ergodic with $E(g_i g_i') = \Sigma$ [12], and let $\bar{g} \equiv \frac{1}{n}\sum_{i=1}^n g_i$. Then, $\boxed{\sqrt{n}\bar{g} = \frac{1}{\sqrt{n}}\sum_{i=1}^n g_i \xrightarrow{D} N(0,\Sigma)}$

1c. **Autocorrelation, Ergodic Stationarity, and Martingale difference sequence**
It should be noted that autocorrelation in {Z_i} does not affect ergodicity or stationarity. Just check the definition. However, a MDS implies no autocovariance! This is why the model in Ch2 (specifically assumption 2.5) is changed by the assumption that there is autocorrelation. In which case, consistency results change, we need to invoke a different CLT to get asymptotic normality, and the asymptotic variance is different.

Pf: Suppose Z_i is MDS. Cov(Z_i,Z_{i-k}) = E(Z_iZ_{i-k})-E(Z_i)E(Z_{i-k}) = E[Z_{i-k} E(Z_i | Z_{i-1},…)] – E[E(Z_i | Z_{i-1},…)]E(Z_{i-k}) = 0 by MDS assumption.

---

[8] **Need for Ergodic Stationarity**: The fundamental problem in time-series analysis is that we observe the realization of the process only once. (i.e. we get only 1 sample and 1 observation of realized $\{z_i\}$!) Ideally, we would like to observe history many times over to obtain more samples, but clearly this is not feasible. But if each of the z_i's come from the same distribution (**stationarity**) , then we can view each realization of $\{z_i\}$ as n realizations from the same distribution. Furthermore, if the process is not too persistent (**ergodicity**), then each element of $\{z_i\}$ will contain some information not available from the other elements. In this case, the time average over the elements of $\{z_i\}$ will be **consistent for the ensemble mean**!

[9] Kolmogorov's Second Strong Law of Large Numbers: Let $\{z_i\}_{i=1}^n$ be iid with $E(z_i)=\mu$ .Then, $\bar{z}_n \equiv \frac{1}{n}\sum_{i=1}^n z_i \xrightarrow{a.s.} \mu$

[10] Since if $\{z_i\}$ ergodic stationary, then $\{f(z_i)\}$ also ergodic stationary for any measurable function *f(.)*.

[11] $Cov(g_i,g_{i-j}) = E(g_i g_{i-j}') - E(g_i)E(g_{i-j}) = E(g_i g_{i-j}') = E\left[E(g_i g_{i-j}' \mid g_{i-j})\right] = E\left[E(g_i \mid g_{i-j})g_{i-j}'\right] = E\left[E\left(E(g_i \mid g_{i-1},...,g_1)\mid g_{i-j}\right)g_{i-j}'\right] = 0 \text{ since } E(g_i \mid g_{i-1},...,g_1)=0$

[12] Since $\{g_i\}$ stationary, the matrix of cross moments does not depend on *i*. Also, we implicitly assume that all the cross moments exist and are finite.

Linear Processes: An important class of covariance-stationary processes, which are created by taking a moving average of a **white noise process**[13].

1. **MA(q):** A process $\{y_t\}$ is called the **q-th** order moving average process (MA(q)) if it can be written as a weighted average of the current and **most recent q values of a white noise process**:

$$y_t = \mu + \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q} \ \ with\ \theta_0 = 1 \ \text{or}\ y_t = \mu + \sum_{j=0}^{q}\theta_0\varepsilon_{t-j} \ \ with\ \theta_0 = 1$$

MA(q) process is Covariance-Stationary mean μ with j-th order autocovariance[14]:

$$\gamma_j = Cov(y_t, y_{t-j}) = \left(\theta_j\theta_0 + \theta_{j+1}\theta_1 + ... + \theta_q\theta_{q-j}\right)\sigma^2 = \sigma^2\sum_{k=0}^{q-j}\theta_{j+k}\theta_k \ \ for\ \ j = 0,1,...,q$$

$$\gamma_j = 0 \ \ for\ \ j > q$$

MA(inf) as a Mean Square Limit: MA(q) process dies out after q lags ($\gamma_j = 0 \ for\ j > q$). Though some time series have this property, we want to learn to model serial correlation that depends on the infinite past MA(inf)

$$MA(\infty):\ \ y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} \ \ where\ \psi_j\ is\ a\ sequence\ of\ reals,\ for\ \sum_{j=0}^{\infty}\left|\psi_j\right| < \infty\ (nec\ but\ not\ suff)$$ [15]

Prop 1: MA(inf) with absolutely summable coefficients

Let $\{\varepsilon_t\}$ be white noise and $\{\phi_j\}$ be a sequence of real numbers that is absolutely summable. Then

$$(a)\ For\ each\ t,\ \ y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\ \ converges\ in\ mean\ square\ (L2\ convergence)$$

$$i.e.\ so\ that\ \ y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\ \ is\ a\ meaningful\ \lim it\ of\ \ y_t = \mu + \sum_{j=0}^{n}\psi_j\varepsilon_{t-j}$$

$$(b)\ \{y_t\}\ is\ \text{cov}-stationary\ with\ E(y_t) = \mu\ and\ auto\,\text{cov}\,ariance\ \{\gamma_t\}\ are\ given\ by\ \gamma_j = \left(\theta_j\theta_0 + \theta_{j+1}\theta_1 + ...\right)\sigma^2 = \sigma^2\sum_{k=0}^{\infty}\theta_{j+k}\theta_k$$

$$(c)\ Auto\,\text{cov}\,ariances\ are\ absolutely\ summable:\ \sum_{j=0}^{\infty}\left|\gamma_j\right| < \infty$$

$$(d)\ If\ \{\varepsilon_t\}\ is\ iid\ white\ noise,\ then\ \{y_t\}\ is\ (strictly)\ stationary\ and\ ergodic$$

(Note: this includes MA(q) as a special case since absolute summability is trivially met for q finite)

---

[13] White noise process $\{e_t\}$ is a 0-mean covariance stationary (cov between terms does not depend on t) process with no serial correlation: $E(e_t) = 0$, $E(e_t^2) = s^2 > 0$, $E(e_te_{t-j}) = Cov(e_te_{t-j}) = 0$ for j != 0

[14]

$$E(y_t) = \mu + E(\varepsilon_t) + \theta_1E(\varepsilon_{t-1}) + ... + \theta_qE(\varepsilon_{t-q}) = \mu\ \ (does\ not\ depend\ on\ t)$$

$$For\ 0 < j < q,$$

$$Cov(y_t, y_{t-j}) = Cov(\mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}, \mu + \varepsilon_{t-j} + \theta_1\varepsilon_{t-j-1} + ... + \theta_q\varepsilon_{t-j-q}) = Cov(\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}, \varepsilon_{t-j} + \theta_1\varepsilon_{t-j-1} + ... + \theta_q\varepsilon_{t-j-q})$$

$$= Cov\left(\sum_{k=0}^{j}\theta_k\varepsilon_{t-k} + \sum_{k=1}^{q-j}\theta_{j+k}\varepsilon_{t-j-k}, \sum_{k=1}^{q-j}\theta_k\varepsilon_{t-j-k} + \sum_{k=q-j+1}^{q}\theta_{t-j-k}\varepsilon_{t-j-k}\right) = Cov\left(\sum_{k=1}^{q-j}\theta_{j+k}\varepsilon_{t-j-k}, \sum_{k=1}^{q-j}\theta_k\varepsilon_{t-j-k}\right)\ these\ are\ the\ overlapping\ terms$$

$$= \sum_{k=1}^{q-j}\theta_k\theta_{j+k}Var\left(\varepsilon_{t-j-k}\right)\ \ \sin ce\ white\ noise\ so\ no\ \text{cov}\ bt\ cross\ terms$$

$$= \sigma^2\sum_{k=0}^{q-j}\theta_{j+k}\theta_k\ \ \ \ by\ \text{cov}-stationarity\ of\ white\ noise$$

[15] However, for this to work we need the infinite series to converge to a random variables. A **necessary** condition is absolute summability.

Prop 2: Filtering covariance stationary processes[16] (This generalizes above theorem to not just white noise but cov-stat)
Let $\{x_t\}$ be white noise and $\{h_i\}$ be a sequence of real numbers that is absolutely summable. Then

$$(a)\ For\ each\ t,\ \ y_t = \sum_{j=0}^{\infty} h_j x_{t-j}\ converges\ in\ mean\ square\ (L2\ convergence)$$

$$i.e.\ \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} \xrightarrow{L_2} 0\ so\ that\ y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\ is\ a\ meaningful\ \lim it\ of\ y_t = \mu + \sum_{j=0}^{n}\psi_j\varepsilon_{t-j}$$

$$\{y_t\}\ is\ \text{cov}-stationary$$

$$(b)\ If\ the\ auto\,\text{cov}\,ariances\ of\ \{x_t\}\ are\ absolutely\ summable,\ then\ so\ are\ the\ auto\,\text{cov}\,ariances\ of\ \{y_t\}$$

[16] Recall the following results about mean square convergence:

$$(i)\{z_n\}\ converges\ in\ mean\ square\ iff\ E\left[\left(z_m - z_n\right)^2\right] \to 0\ as\ m,n \to \infty$$

$$(ii)\ If\ x_n \to_{ms} x\ and\ z_n \to_{ms} z,\ then\ \lim_{n\to\infty} E(x_n) = E(x)\ and\ \lim_{n\to\infty} E(x_n z_n) = E(xz)$$

Proof of Prop 1 is as follows:

2. **Filters Background:** The operation of taking a weighted average of (possibly infinitely many) successes values of a process, i.e. $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, is called **filtering**, and can be expressed compactly using the **lag-operator:** $L^j x_t = x_{t-j}$.

Def: For a given arbitrary sequence of reals, $\{\alpha_0, \alpha_1, ...\}$ define a **filter** by $\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_1 L^2 + ...$

Therefore, applying to a process $\{x_t\}$, $\alpha(L)x_t = \alpha_0 x_t + \alpha_1 L\, x_t + \alpha_2 L^2 x_t + ... = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... = \sum_{k=0}^{\infty} \alpha_k x_{t-k}$

Note: $\boxed{If\ \ x_t = c\ (cons),\ \alpha(L)c = \alpha(1)c = (\alpha_0 + \alpha_1 + \alpha_2 + ...)c = c\sum_{j=0}^{\infty} a_j .}$

Def: If $\alpha_j \neq 0\ for\ \ j = p\ \ and\ \ \alpha_j = 0\ \ for\ \ j > p$,then the filter reduces to a **p-th degree lag polynomial**:

$$\boxed{\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ... + \alpha_p L^p}$$

[Note: We can treat these like polynomials in terms of operations: multiplication, inverting, roots, etc…]

Def: A filter with absolutely summable coefficients is called **absolutely summable filter**. It is a mapping from the set of covariance stationary processes to itself.

<u>Product of Filters</u>

Def: Let $\{\alpha_j\},\{\beta_j\}$ be two arbitrary sequences of reals s.t. $\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...,\ \ \beta(L) \equiv \beta_0 + \beta_1 L + \beta_2 L^2 + ...$

Then, the **product of the filters** $\alpha(L), \beta(L)$ is: $\delta(L) = \alpha(L)\beta(L)$ where $\{\delta_j\}$ be defined as a **convolution** of $\{\alpha_j\},\{\beta_j\}$:

$\delta_0 = \alpha_0\beta_0,\ \delta_1 = \alpha_0\beta_1 + \alpha_1\beta_0,\ \ \delta_2 = \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0, ... ,\ \delta_j = \alpha_0\beta_j + \alpha_1\beta_{j-1} + \alpha_2\beta_{j-2} + ... + \alpha_{j-1}\beta_1 + \alpha_0\beta_0$

Example: for $\alpha(L) = 1 + \alpha_1 L,\ \ \beta(L) = 1 + \beta_1 L \Rightarrow \delta(L) = \alpha(L)\beta(L) = (1+\alpha_1 L)(1+\beta_1 L) = 1 + (\alpha_1 + \beta_1)L + \alpha_1\beta_1 L^2$

Property: Filter mult. is **commutative**: $\alpha(L)\beta(L) = \beta(L)\alpha(L)$ (Just like matrix multiplication, **will apply for matrix filters**)

Prop: If $\alpha(L), \beta(L)$ absolutely summable and if $\{x_t\}$ covariance stationary, then by Prop2 $\alpha(L)\beta(L)x_t$ is a well-defined variable, equal to $\delta(L)x_t$: $"\alpha(L)\beta(L) = \delta(L)" \Rightarrow "\alpha(L)\beta(L)x_t = \delta(L)x_t"$

<u>Inverses of Filters</u>

Def: We say that $\beta(L)$ is the **inverse** of $\alpha(L)$ if $\delta(L) = \alpha(L)\beta(L) = 1$, we denote $\beta(L) = \alpha(L)^{-1}$.
(i.e. The inverse $\alpha(L)^{-1}$ is a filter satisfying $\alpha(L)\alpha(L)^{-1} = 1$)

Note: As long as $\alpha_0 \neq 0$, the inverse $\alpha(L)^{-1}$ of $\alpha(L)$ can be defined for any arbitrary sequence $\{a_j\}$ because

$\delta(L) = \alpha(L)\beta(L) = 1$

$\delta(L) = 1 \Rightarrow \delta_0 = 1\ and\ \ \delta_j = 0\ \ for\ \ j > 0$. Then we can solve the following successively:

$\delta_0 = \alpha_0\beta_0,\ \delta_1 = \alpha_0\beta_1 + \alpha_1\beta_0,\ \ \delta_2 = \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0, ... ,\ \delta_j = \alpha_0\beta_j + \alpha_1\beta_{j-1} + \alpha_2\beta_{j-2} + ... + \alpha_{j-1}\beta_1 + \alpha_0\beta_0$

$\Rightarrow \beta_0 = \frac{1}{\alpha_0},\ \beta_1 = -\frac{\alpha_1}{\alpha_0^{\ 2}}, ...$

Example: Consider the filter 1-L. Then, $(1-L)^{-1} = 1 + L + L^2 + L^3 + ...$

Why? $(1-L)x_t = x_t - x_{t-1};\ (1-L)^{-1}(x_t - x_{t-1}) = 1 = x_t - x_{t-1} + x_{t-1} - x_{t-2} + x_{t-2} - x_{t-3} + ... = (1 + L + L^2 + ...)(x_t - x_{t-1})$

Prop: **Inverses are commutative**: $\alpha(L)^{-1}\alpha(L) = \alpha(L)\alpha(L)^{-1}$, and

$provided\ \ \alpha_0 \neq 0, \beta_0 \neq 0,\ \ \alpha(L)\beta(L) = \delta(L) \Leftrightarrow \beta(L) = \alpha(L)^{-1}\delta(L) \Leftrightarrow \alpha(L) = \delta(L)\beta(L)^{-1}$ (does not apply to matrix filters)

Inverting Lag Polynomials ("finite" filter)

Let $\phi(L)$ *be a* $p-th$ *degree lag polynomial s.t.* $\phi(L)=1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p$

$\phi_0=1 \Rightarrow$ *Inverse can be defined as follows*

Re*call convolution of* $\phi(L)$ *and* $\beta(L)$:

$\delta_0=\phi_0\beta_0,\ \delta_1=\phi_0\beta_1+\phi_1\beta_0,\ \delta_2=\phi_0\beta_2+\phi_1\beta_1+\phi_2\beta_0, .\ ...\ ,\ \delta_j=\phi_0\beta_j+\phi_1\beta_{j-1}+\phi_2\beta_{j-2}+...+\phi_{j-1}\beta_1+\phi_0\beta_0$

and if $\beta(L)$ *is the inverse* $\Rightarrow \delta(L)=1 \Rightarrow \delta_0=1$ *and* $\delta_j=0$ *for* $j>0$

Thus,

$$\beta_0=\frac{1}{\phi_0}=1\ (\text{since } \phi_0=1),\ \beta_1=\phi_1, \beta_1=\phi_2+\phi_1^{\,2},...$$

Takeaway: for any p-th degree lag polynomial $\phi(L)$ with $\phi_0=1$, its inverse is defined/exosts!
(though it may not be absolutely summable unless the invertibility condition is satisfied)

Prop 3: **Absolutely Summable Inverses of Lag Polynomials**

Consider a p-th degree lag polynomial $\phi(L)=1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p$ and let $\psi(L)=\phi(L)^{-1}$. If the associated p-th degree polynomial equation $\phi(z)=0$ satisfies the following **stability condition**:

a) All roots (λ_i) of the p-th degree polynomial equation in z
$\phi(z)=0$ *where* $\phi(z)\equiv 1-\phi_1 z-\phi_2 z^2-...-\phi_p z^p=(1-\lambda_1 z)(1-\lambda_2 z)...(1-\lambda_p z)$ are s.t. $|\lambda_i|>1$ *for all* i

Then the coefficient sequence $\{\psi_i\}$ is bounded in absolute value is bounded by a geometrically declining sequence and hence is absolutely summable. (This will help us express AR processes as "well-defined" MA processes).

- **Example**: Consider a lag polynomial of degree 1: $\phi(L)=1-\phi L$

  The root of the associated polynomial equation $1-\phi z=0$ is $z=\frac{1}{\phi}$.

  The stability condition is $\left|\frac{1}{\phi}\right|>1$ *or* $|\phi|<1$. Then, the inverse filter

  $\psi(L)=\phi(L)^{-1}$ *which is evidently* $=1+\phi L+\phi^2 L^2+...=\left(\sum_{j=0}^{\infty}\phi^j L^j\right)$ has absolutely summable coefficients.

  <u>We will use this to show that AR(1) can be expressed as MA(inf)</u>

Note: Why do we need stability condition? Stability condition guarantees that the inverse is absolutely summable, and therefore we can multiply both sides of the equation by the inverse filter. (It wouldn't make sense to multiply infinity to both sides)

3. **ARMA Processes**: Important class of linear processes (cov-stat gen from white noise), which are a parameterization of coefficients of MA(inf) process (the idea here is to be able to represent covariance stationary processes in a nice parametization of coefficients)

Def: A **First-Order Autoregressive Process AR(1)** satisfies the following stochastic difference equation:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \Leftrightarrow y_t - \phi y_{t-1} = \mu + \varepsilon_t \Leftrightarrow (1-\phi L) y_t = \mu + \varepsilon_t \quad where\ \{\varepsilon_t\}\ is\ white\ noise$$

I. **MA(inf) representation of AR(1) as long as $|\phi| < 1$ (stability condition to guarantee absolute summability)**[17]

$$(1-\phi L) y_t = \mu + \varepsilon_t \Rightarrow y_t = (1-\phi L)^{-1}(\mu + \varepsilon_t) = \sum_{j=0}^{\infty} \phi^j L^j (\mu + \varepsilon_t) = \mu \sum_{j=0}^{\infty} \phi^j + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} = \frac{\mu}{1-\phi} + \phi(L)\varepsilon_t = \tilde{\mu} + \phi(L)\varepsilon_t$$

Note: If y_t has mean $\tilde{\mu} = \frac{\mu}{1-\phi}$, then it is covariance stationary and has the above MA(inf) representation

(Here we have white noise and absolute summability. So all we need is for expectation to be right).

Note: Intuitively, this result says that the process, if it started a long time ago, "settles down", provided that $|\phi| < 1$ so that the effect of the past dies out geometrically as time progresses. This is called the **stationarity condition**.

II. **Autocovariances of AR(1) processes**

2 Methods of Calculating autocov of AR(1) (for case of $|\phi| < 1$):

a) $y_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_j \Rightarrow \gamma_j = \sigma^2(\phi^j \phi^0 + \phi^{j+1}\phi^1 + \phi^{j+2}\phi^2 + ...) = \phi^j (1 + \phi^2 + \phi^4 + ...)\sigma^2 = \left(\frac{\phi^j}{1-\phi^2}\right)\sigma^2$

and $\rho_j \equiv \frac{\gamma_j}{\gamma_0} = \left(\frac{\phi^j}{1-\phi^2}\right)\sigma^2 / \left(\frac{1}{1-\phi^2}\right)\sigma^2 = \phi^j$

b) Via "Yule-Walker" Equations (See 271B HW2)

III. **AR(p) and its MA(inf) Representation**: Above results for AR(1) can be generalized for AR(p) – p-th order autoreg. process

Def: A **P-th Order Autoregressive Process AR(P)** satisfies the following stochastic difference equation:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_2 y_{t-p} + \varepsilon_t$$
$$or\ \ y_t - \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_2 y_{t-p} = \mu + \varepsilon_t$$
$$or\ \ (1-\phi L) y_t = \mu + \varepsilon_t \quad where\ \{\varepsilon_t\}\ is\ white\ noise$$

The generalization of AR(p) of what we derived for AR(1) is as follows:

Prop 4: AR(p) as a MA(inf) with Absolutely Summable Coefficients

Suppose the p-th deree polynomial satisfies the stationarity/stability condition (so that the inverse is well defined) i.e. $\phi(z) = 0 \Rightarrow |z| > 1$. Then,

(a) The unique covariance-stationary solution to the p-th order stochastic difference equation has a MA(inf) representation :

$$y_t = \mu + (\phi_1 + \phi_2 L + ... + \phi_p L^p) y_{t-1} + \varepsilon_t \Rightarrow (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = \mu + \varepsilon_t \Rightarrow (1-\lambda_1 L)(1-\lambda_2 L)...(1-\lambda_p L) y_t = \mu + \varepsilon_t$$

$$\Rightarrow y_t = (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}...(1-\lambda_p L)^{-1}\mu + (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}...(1-\lambda_p L)^{-1}\varepsilon_t = \tilde{\mu} + \left(\sum_{j=0}^{\infty} \lambda_1{}^j L^j\right)\left(\sum_{j=0}^{\infty} \lambda_2{}^j L^j\right)...\left(\sum_{j=0}^{\infty} \lambda_p{}^j L^j\right)\varepsilon_t$$

$$\Rightarrow y_t = \tilde{\mu} + \psi(L)\varepsilon_t,\ \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + ... \ where\ \psi(L) = \phi(L)^{-1},\ \tilde{\mu} = \phi(1)^{-1}\mu$$

The coefficients $\{\psi_j\}$ are bounded in absolute value by a sequence by a geometrically declining sequence and hence is absolutely summable. (Thus the MA representation is well-defined. RHS converges.)

Note: The MA(inf) representation comes from the fact that the inverse is represented as **product of infinite sums**!

---

[17] Recall MA processes are well defined only when the lag coefficients are absolute summable. We can multiply the inverse on both sides only if absolute summability is met, and then RHS becomes a meaningful/well-defined limit. The $|\phi| < 1$ is equivalent to the stability condition because the root of $1 - \phi z = 0$ is $z = \frac{1}{\phi}$, and the stability condition is $\left|\frac{1}{\phi}\right| > 1\ or\ |\phi| < 1$.

(b) The mean of the process is given by: $\tilde{\mu} = \phi(1)^{-1}\mu$

(c) $\{\gamma_j\}$ is bounded in absolute value by a sequence that declines geometrically with j. Therefore autocovariances are absolutely summable.

Note: We refer to "an AR(p) process" as the unique covariance stationary solution to an AR(p) equation that satisfies the stationarity condition. The absolute summability of the inverse in part (a) follows from prop 3, since stability conditions are satisfied.

IV. ARMA(p,q): An AR(p,q) process combines AR(p) and MA(q)

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}$$

$$or\ \phi(L)y_t = \mu + \theta(L)\varepsilon_t \quad where\ \ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p,\ \ \theta(L) = 1 + \theta_1 L + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}$$

Where $\{\varepsilon_t\}$is white noise.

Prop 4: AR(p) as a MA(inf) with Absolutely Summable Coefficients

Suppose the p-th deree polynomial satisfies the stationarity/stability condition (so that the inverse is well defined) i.e. $\phi(z) = 0 \Rightarrow |z| > 1$. Then,

(a) The unique covariance-stationary solution to the p-th order stochastic difference equation has a MA(inf) representation :

$$\phi(L)y_t = \mu + \theta(L)\varepsilon_t \quad where\ \ \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

$$\theta(L) = 1 + \theta_1 L + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q} y_t = \mu + (\phi_1 + \phi_2 L + ... + \phi_p L^p)y_{t-1} + \varepsilon_t$$

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)y_t = \mu + \varepsilon_t \Rightarrow (1-\lambda_1 L)(1-\lambda_2 L)...(1-\lambda_p L)y_t = \mu + \varepsilon_t$$

$$\Rightarrow y_t = (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}...(1-\lambda_p L)^{-1}\mu + (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}...(1-\lambda_p L)^{-1}(1-\eta_1)(1-\eta_2)...(1-\eta_q)\varepsilon_t$$

$$= \tilde{\mu} + \left(\sum_{j=0}^{\infty}\lambda_1{}^j L^j\right)\left(\sum_{j=0}^{\infty}\lambda_2{}^j L^j\right)...\left(\sum_{j=0}^{\infty}\lambda_p{}^j L^j\right)(1-\eta_1)(1-\eta_2)...(1-\eta_q)\varepsilon_t$$

$$\Rightarrow y_t = \tilde{\mu} + \psi(L)\varepsilon_t,\ \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + ... \ \ where\ \ \psi(L) = \phi(L)^{-1}\theta(L),\ \tilde{\mu} = \phi(1)^{-1}\mu$$

The coefficients $\{\psi_j\}$ **are bounded in absolute value** by a sequence by a geometrically declining sequence and **hence is absolutely summable**. (Thus the MA representation is well-defined. RHS converges.)

Note: The MA(inf) representation comes from the fact that the inverse is represented as **product of infinite sums**!

(b) The mean of the process is given by: $\tilde{\mu} = \phi(1)^{-1}\mu$

(c) $\{\gamma_j\}$ is bounded in absolute value by a sequence that declines geometrically with j. Therefore autocovariances are absolutely summable.

Note: We refer to "an ARMA(p,q) process" as the unique covariance stationary solution to an ARMA equation that satisfies the stationarity condition. The absolute summability of the inverse in part (a) follows from prop 3, since stability conditions are satisfied.

<u>Invertibility of ARMA(p,q) and its MA(inf) and AR(inf) Representations</u>

If an ARMA(p,q) process satisfies the stability condition $\phi(z) = 0 \Rightarrow |z| > 1$, and $\theta(z) = 0 \Rightarrow |z| > 1$ (stability condition on theta is called **invertibility condition**), and if in addition $\phi(z) = 0\ and\ \theta(z) = 0$ have no common roots. Then, the ARMA process is invertible. And **the process as an AR(inf) and MA(inf) representation**!

If $\theta(z) = 0 \Rightarrow |z| > 1$, then $\theta(L)^{-1}$ is absolutely summable and we can multiply both sides of ARMA by $\theta(L)^{-1}$ to obtain its AR(inf) representation[18]:

$$\phi(L)y_t = \mu + \theta(L)\varepsilon_t \Rightarrow \theta(L)^{-1}\phi(L)y_t = \theta(L)^{-1}\mu + \varepsilon_t \Rightarrow \theta(L)^{-1}\phi(L)y_t = \theta(1)^{-1}\mu + \varepsilon_t = \frac{\mu}{\theta(1)} + \varepsilon_t$$

$$\therefore \theta(L)^{-1}\phi(L)y_t = c + \varepsilon_t \quad (AR(\infty))$$

---

[18] Recall, $\theta(L)^{-1}$ exists and is defined if θ_0 =1, as is the case here. However, it may not be absolutely summable (and therefore we cannot multiply both sides of the equation by it) unless the stability condition $\theta(z) = 0 \Rightarrow |z| > 1$ (aka the invertibility condition) is satisfied.

4. Estimating Autoregressions: Autoregressive processes (autoregressions) are popular in econometrics because they have a natural interpretation and also because they are easy to estimate.

Prop 7: (Estimation of AR coefficients)
Let $\{y_t\}$ be an AR(p) process that satisfies the stationarity condition. Suppose further that $\{e_t\}$ is iid white noise. Then the OLS estimator of $\{c, \phi_1, \phi_2, ..., \phi_p\}$ obtained by regressing y_t on a constant and p lagged values of y is **consistent and asymptotically normal**, with

$$A\text{var}(\hat{\beta}) = \sigma^2 E(x_t x_t')^{-1} \qquad \textit{where } x_t = (1, y_{t-1}, ..., y_{t-p})'$$

which is consistently estimated by

$$\hat{A}\text{var}(\hat{\beta}) = s^2 \left(\frac{1}{n} \sum_{t=1}^{n} x_t x_t' \right)^{-1} \qquad \textit{where } s^2 = \frac{1}{n-p-1} \sum_{t=1}^{n} \left(y - \hat{c} - \hat{\phi}_1 y_{t-1} - ... - \hat{\phi}_p y_{t-p} \right)^2$$

Note: This follows from the fact that y_t can be written as a MA(inf), so if $\{e_t\}$ is iid white noise, then y_t is ergodic stationary from Prop 1(d), and it satisfies orthogonality to e_t by its independence. Therefore, assumptions for OLS are met. Though, we check that $E(x_t x_t')$ is nonsingular, or equivalently checking that the autocovariance matrix $Var(y_t, \ldots, t_{t-p})$ is nonsingular.
It can be shown that the autocovariance of a covariance-stationary process is nonsingular for any p if $\gamma_0 > 0$ *and* $\gamma_j \to 0$ *as* $j \to \infty$.
(which is satisified here since we have AR(p) as an MA(inf) and $\{e_t\}$ is iid)
Note: An AR(p) process with non-iid white noise can also be estimated via OLS… WHY?!

Estimating ARMA
To estimate AR part, we can use GMM/IV/2SLS to deal with endogeneity.

Autocovariance-Generating Function
AGF is a useful way to summarize the whole profile of autocovariances $\{\gamma_j\}$ of a covariance stationary process $\{y_t\}$.

$$g_Y(z) \equiv \sum_{j=-\infty}^{\infty} \gamma_j z^j = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (z^j + z^{-j}) \quad (\text{since } \gamma_{-j} = \gamma_j \text{ by cov stat})$$

There's a question about whether the summation exists, but for z = 1 and for $\{\gamma_j\}$ absolutely summable

$$g_Y(1) \equiv \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j$$

<u>Prop 6</u>: <u>Autocovariance generating function for MA(inf)</u>
Let $\{e_t\}$ be white noise and let $\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + ...$ be an absolutely summable filter (i.e. with $\{\psi_j\}$ absolutely summable. Then the autocovariance-generating function of the MA(inf) process $\{y_t\}$ where $y_t = \mu + \psi(L)\varepsilon_t$ is:

$$g_Y(z) = \sigma^2 \psi(z) \psi(z^{-1})$$ [19]

Therefore, for AR(p) and ARMA(p,q), since they have MA(inf) representations:

AR(p): $$g_Y(z) \equiv \sigma^2 \frac{1}{\phi(z)\phi(z^{-1})}$$

ARMA(p): $$g_Y(z) \equiv \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}$$

Note: These are useful because long-run variance can be expressed as $g_y(1)$

19

$$\text{Recall, for } MA(q) \text{ process, } \gamma_j = \sigma^2 \sum_{k=0}^{q-j} \theta_{j+k}\theta_k \text{ for } j = 0,1,...,q \text{ and } \gamma_j = 0 \text{ for } j > q$$

Take MA(1).

$$\text{Then, } g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (z^j + z^{-j}) = \gamma_0 + \gamma_1 (z + z^{-1}) = \sigma^2 (\theta_0^2 + \theta_1^2 + \theta_1\theta_0 z + \theta_1\theta_0 z^{-1}) = \sigma^2 (1 + \theta_1^2 + \theta_1 z + \theta_1 z^{-1}) = \sigma^2 (1 + \theta_1 z)(1 + \theta_1 z^{-1})$$

$$\text{Thefore, } g_Y(z) = \sigma^2 \theta(z)\theta(z^{-1})$$

5. **Asymptotics for Sample Means of Serially Correlated Processes**
Here we develop LLN and CLT for serially correlated processes. Previously, we used Billingsly (ergodic stationary MDS) CLT, which rules out serial correlation since the process is assumed to be a martingales difference sequence[20].

1a) <u>LLN for Covariance Stationary Processes with vanishing Autocovariances</u>[21]:
Let $\{y_t\}$ be covariance-stationary with mean μ and $\{\gamma_j\}$ be the autocovariances of $\{y_t\}$. Then,

$$(a)\ \ \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{m.s./L2} \mu \ \ if\ \ \lim_{j\to\infty}\gamma_j = 0$$

$$(b)\ \lim_{j\to\infty} Var\left(\sqrt{n}\bar{y}\right) = \sum_{j=-\infty}^{\infty}\gamma_j < \infty \ \ if$$

(Note: we also call this the **long-run variance** of the covariance stationary process[22], it can be expressed from AGF $g_Y(1)$).

1b) <u>LLN for Vector Covariance-Stationary Processes with vanishing Autocovariances (diag element of</u> $\{\Gamma_j\}$ <u>)</u>:
Let $\{\mathbf{y}_t\}$ be a vector covariance-stationary with mean $\bar{\mu}$ and $\{\Gamma_j\}$ be the autocovariances[23] of $\{\mathbf{y}_t\}$. Then,

$$(a)\ \ \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{m.s./L2} \mu \ \ if\ \ diagonal\ elements\ of\ \Gamma_j \to_{m.s.} 0\ as\ j\to\infty$$

$$(b)\ \lim_{j\to\infty} Var\left(\sqrt{n}\bar{y}\right) = \sum_{j=-\infty}^{\infty}\Gamma_j < \infty \ \ if\ \{\ \Gamma_j\ \}\ is\ summable\ (i.e.\ each\ component\ of\ \Gamma_j\ \ summable)$$

(Note: we also call this the **long-run covariance variance matrix** of the vector covariance stationary process[24], it can be expressed from Multivariate AGF: $G_Y(1) = \sum_{j=-\infty}^{\infty}\Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j'))$.

---

[20] g_t MDS → $E(g_t \mid g_{t-1},\ldots,)=0$. Therefore, $Cov(g_t,g_{t-1})=E(g_tg_{t-1}) = E(E(g_t \mid g_{t-1},\ldots)) = 0$.

[21]

$$Var\left(\sqrt{n}\bar{y}\right) = Var\left(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}y_i\right) = \frac{1}{n}Var\left(\sum_{i=1}^{n}y_i\right) = \frac{1}{n}\left(Cov\left(\sum_{i=1}^{n}y_i, \sum_{i=1}^{n}y_i\right)\right) = \frac{1}{n}\left(Cov\left(y_1,\sum_{i=1}^{n}y_i\right)+\ldots+Cov\left(y_n,\sum_{i=1}^{n}y_i\right)\right)$$

$$= \frac{1}{n}\left((\gamma_0+\gamma_1+\ldots+\gamma_{n-1})+(\gamma_{-1}+\gamma_0+\gamma_1\ldots+\gamma_{n-2})+\ldots+(\gamma_{-n+1}+\ldots+\gamma_0)\right)$$

$$= \frac{1}{n}\left(n\gamma_0+(n-1)\gamma_1+(n-2)\gamma_2+\ldots+\gamma_{n-1}+(n-1)\gamma_{-1}+(n-2)\gamma_{-2}+\ldots+\gamma_{-n+1}\right)$$

$$= \frac{1}{n}\left(n\gamma_0+2(n-1)\gamma_1+2(n-2)\gamma_2+\ldots+2\gamma_{n-1}\right)\ by\ \text{cov}-stationarity$$

$$= \gamma_0 + 2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\gamma_j = \gamma_0 + 2\sum_{j=1}^{n-1}\gamma_j - \frac{2}{n}\sum_{j=1}^{n-1}j\gamma_j$$

$$\gamma_j\ summable \Rightarrow \frac{1}{n}\sum_{j=1}^{n}j\gamma_j \to 0$$

$$Therefore,\ \lim_{n\to\infty}Var\left(\sqrt{n}\bar{y}\right) = \lim_{n\to\infty}\gamma_0 + 2\sum_{j=1}^{n-1}\gamma_j = \sum_{j=-\infty}^{\infty}\gamma_j$$

[22] We can think of the sample as being generated from an infinite sequence of random variables (which is cov. Stationary). So, the "long-run" variance is the sum of covariances from any 1 element in the sequence to all the other elements.

[23] In a vector process, the diagonal elements of $\{\Gamma_j\}$ are the autocovariances and the off diagonal are the covariances between the lagged values of the elements of the vector.

$$Let\ \ y_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}.$$

For example: $Then,\ \Gamma_j = \text{cov}(y_t, y_{t-j}) = E(y_t y_{t-j}') - E(y_t)E(y_{t-j}) = E\left(\begin{bmatrix} x_t \\ z_t \end{bmatrix}\begin{bmatrix} x_{t-j} & z_{t-j} \end{bmatrix}\right) - E\begin{bmatrix} x_t \\ z_t \end{bmatrix}E\begin{bmatrix} x_{t-j} & z_{t-j} \end{bmatrix}$

$$= \begin{bmatrix} E(x_t x_{t-j}) - E(x_t)E(x_{t-j}) & E(x_t z_{t-j}) - E(x_t)E(z_{t-j}) \\ E(x_{t-j} z_t) - E(x_{t-j})E(z_t) & E(z_t z_{t-j}) - E(z_t)E(z_{t-j}) \end{bmatrix} = \begin{bmatrix} Cov(x_t, x_{t-j}) & Cov(x_t, z_{t-j}) \\ Cov(x_{t-j}, z_t) & Cov(z_t, z_{t-j}) \end{bmatrix}$$

2a) **CLT for MA(inf)** (Billingsley generalizes Lindberg-Levy[25] to stationary and ergodic mds, now we generalize for serial corr)

Let $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ where $\{\varepsilon_t\}$ is iid white noise and $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty$. Then,

$$\boxed{\sqrt{n}\left(\bar{y} - \mu\right) \xrightarrow{D} N\left(0, \sum\nolimits_{j=-\infty}^{\infty} \gamma_j\right)}$$

2b) **MV CLT for MA(inf)**

Let $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ where $\{\varepsilon_t\}$ is vector iid white noise (i.e. jointly covariance stationary) and $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty$. Then,

$$\boxed{\sqrt{n}\left(\bar{y} - \mu\right) \xrightarrow{D} N\left(0, \sum\nolimits_{j=-\infty}^{\infty} \Gamma_j\right)}$$

---

[24] We can think of the sample as being generated from an infinite sequence of random variables (which is cov. Stationary). So, the "long-run" variance is the sum of covariances from any 1 element in the sequence to all the other elements.

[25] Let $\{z_i\}$ be iid with $E(z_i) =$ $E(z_i) = \mu$ *and* $Var(z_i) = \Sigma$. *Then* $\sqrt{n}\left(\bar{z} - \mu\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (z_i - \mu) \xrightarrow{D} N(\mu, \Sigma)$ (an iid random vector converges to a random vector that is normally distributed)

3a) <u>Gordin's CLT for Ergodic + Stationary + Gordin[26] (conditions for restricting degree of autocorrelation) Processes</u>

Suppose $\{y_t\}$ is stationary and ergodic and suppose Gordin's conditions are satisfied:

a) $E(y_t^2) < \infty$ (this is a restriction on [strictly] stationary processes which might not have finite 2nd moments)

b) $E(y_t \mid y_{t-j}, y_{t-j-1}, \ldots) \xrightarrow{m.s.} 0$ *as* $j \to \infty$ *or equivalently* (since *stationary*) $E(y_0 \mid y_{-j}, y_{-j-1}, \ldots) \xrightarrow{m.s.} 0$ [27]

c) $\sum_{k=0}^{\infty} E\left(r_{t,t-k}{}^2\right)^{1/2} < \infty$ *where* $r_{t,t-k} \equiv E(y_t \mid I_{t-k}) - E(y_t \mid I_{t-k-1})$ *and Info Set* $I_{t-k} \equiv (y_{t-k}, y_{t-k-1}, y_{t-k-2}, \ldots)$

(**Telescoping sum condition: restricts degree of autocorrelation**)[28]

Then,

1. The autocovariances $\{\gamma_j\}$ are absolutely summable
2. $\boxed{\sqrt{n}\,(\bar{y} - \mu) \xrightarrow{D} N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right)}$

3b) <u>Gordin's MV CLT for Ergodic + Stationary + Gordin[29] (conditions for restricting degree of autocorrelation) Processes</u>

Suppose $\{\mathbf{y}_t\}$ is stationary and ergodic and suppose Gordin's conditions are satisfied for the vector ergodic stationary process:

a) $E(\underline{y}_t \underline{y}_t') < \infty$ (this is a restriction on [strictly] stationary processes which might not have finite 2nd moments)

b) $E(\underline{y}_t \mid \underline{y}_{t-j}, \underline{y}_{t-j-1}, \ldots) \xrightarrow{m.s.} 0$ *as* $j \to \infty$ *or equivalently* (since *stationary*) $E(\underline{y}_0 \mid \underline{y}_{-j}, \underline{y}_{-j-1}, \ldots) \xrightarrow{m.s.} 0$

c) $\sum_{k=0}^{\infty} E\left(\underline{r}_{t,t-k}{}^2\right)^{1/2} < \infty$ *where* $r_{t,t-k} \equiv E(\underline{y}_t \mid I_{t-k}) - E(\underline{y}_t \mid I_{t-k-1})$ *and Info Set* $I_{t-k} \equiv (\underline{y}_{t-k}, \underline{y}_{t-k-1}, \underline{y}_{t-k-2}, \ldots)$

(**Telescoping sum condition: restricts degree of autocorrelation**)

Then,

1. The covariance matrices $\{\Gamma_j\}$ are absolutely summable
2. $\boxed{\sqrt{n}\,(\bar{y} - \mu) \xrightarrow{D} N\left(0, \sum_{j=-\infty}^{\infty} \Gamma_j\right)}$

Note: This is a generalization of MDS CLT, since if $\{y_t\}$is MDS sequence, then Gordin's condition is satisfied.

---

[26] We should be able to check that with these condition we satisfy Billingsly's general requirements for CLT to hold (Peter's notes slide 16).

[27] This condition implies that the unconditional mean is 0: $E(y_t) = 0$. So, as j → infinity, that is, as the forecast (the conditional expectation) is based on less and less information, it should approach the forecast based on no information (i.e. the unconditional expectation). (See White text for proof)

[28] We can write y_t as a telescoping sum as follows:

$$y_t = y_t - \left(E(y_t \mid I_{t-1}) + E(y_t \mid I_{t-1})\right) - \left(E(y_t \mid I_{t-2}) + E(y_t \mid I_{t-2})\right) - \ldots - \left(E(y_t \mid I_{t-j}) + E(y_t \mid I_{t-j})\right)$$

$$= \left(y_t - E(y_t \mid I_{t-1})\right) + \left(E(y_t \mid I_{t-1}) - E(y_t \mid I_{t-2})\right) + \ldots + \left(E(y_t \mid I_{t-j+1}) - E(y_t \mid I_{t-j})\right) + E(y_t \mid I_{t-j})$$

$$= \left(r_{t,t} + r_{t,t-1} + r_{t,t-2} + \ldots r_{t,t-j}\right) + E(y_t \mid I_{t-j})$$

By (*b*), $E(y_t \mid I_{t-j}) \to_{ms} 0$ *as* $j \to \infty$, *therefore*

$$y_t = \sum_{k=0}^{\infty} r_{t,t-k}$$

This telescoping sum indicates how the "shocks" represented by $(r_{t,t}, r_{t,t-1}, \ldots)$ influence the current value of y. Condition (c) says, roughly, that shocks that occurred a long time ago do not have disproportionately large influence. As such, this condition restricts the degree of serial correlation in {y}.

[29] We should be able to check that with these condition we satisfy Billingsly's general requirements for CLT to hold (Peter's notes slide 16).

4) Examples: Asymptotic Distributions of AR(p), ARMA(p,q)

Let $\{y_t\}$ be ARMA(1,1). To check whether LLN applies, we need to check whether $\{\gamma_j\}$ vanishes as j → inf.

ARMA(1,1) has a MA(inf) representation, and from previous, we know autocovariance is obtained by

$$\gamma_j = Cov(y_t, y_{t-j}) = \left(\theta_j\theta_0 + \theta_{j+1}\theta_1 + \ldots + \theta_q\theta_{q-j}\right)\sigma^2 = \sigma^2\sum_{k=0}^{q-j}\theta_{j+k}\theta_k \quad for \quad j = 0,1,\ldots,q$$

$$\gamma_j = 0 \quad for \quad j > q$$

So for j >q, verify from here that the autocovariance. Then, to get asymptotic variance,

$$\sum_{j=-\infty}^{\infty}\gamma_j = g_y(1) = \sigma^2\frac{\theta(1)\theta(1^{-1})}{\phi(1)\phi(1^{-1})} = \sigma^2\left(\frac{1-\theta_1}{1-\phi_1}\right)^2$$

In general we have for ARMA(p,q):

$$\sqrt{n}(\bar{y}-\mu) \to_D N\left(0, \sum_{j=-\infty}^{\infty}\gamma_j\right) = N\left(0, \left(\frac{1-\phi_1-\ldots-\phi_q}{1-\psi_1-\ldots-\psi_q}\right)^2\right)$$ by same reasoning as above.

6. **Incorporating Serial Correlation in GMM**

Background: Recall in our GMM assumptions, from assumption 3.3 (orthogonality condition) $E(g_t) = 0$, and in 3.5 we assumed it was a MDS → no serial correlation in $\mathbf{g}_t$ is allowed under this assumption, and as such, the matrix **S** (from which we obtain the asymptotic variance) is defined to be the variance of $\mathbf{g}_t$. Now we can relax assumption 3.5 a follows.

Assumptions

3.1 **Linearity**: The data we observe comes from underlying RV's { $y_{t\,(1x1)}$, $\mathbf{x}_{t\,(1xd)}$}with
$y_t = \mathbf{x}_t'\delta + \varepsilon_t$ (t = 1,2,…,n) (this is the equation we want to estimate)

3.2 **Ergodic Stationarity**:
Let $\mathbf{z}_t$ be a M-dimensional vector of instruments, and let $\mathbf{w}_t$ be the unique and nonconstant elements of (y_t,$\mathbf{x}_t$, $\mathbf{z}_t$).
{$\mathbf{w}_t$} is jointly stationary and ergodic.

3.3 **Orthogonality Condition**
All the M variables in $\mathbf{z}_t$ are predetermined in the sense that they are all orthogonal to the current error term:
$E(z_{tm}\varepsilon_t) = 0 \ \forall t, \forall m \Leftrightarrow E[(y_t - \mathbf{x}_t'\beta) . \mathbf{z}_t] = 0 \Leftrightarrow E(\mathbf{g}_t) = \mathbf{0}$ where $\mathbf{g}_t = \mathbf{z}_t . (y_t - \mathbf{x}_t'\beta) = \mathbf{z}_t . \varepsilon_t$

3.4 **Rank Condition for Identification : Guarantees there's a unique solution to the system of equations**
The m x d matrix $E(\mathbf{z}_t\, \mathbf{x}_t')_{mxd}$ is of full column rank (or $E(\mathbf{x}_t\, \mathbf{z}_t')$ is of full row rank), **M $\geq$ d (# of equations $\geq$ # of unknowns)**.
We denote this matrix by $\mathbf{\Sigma_{ZX}}$.

3.5 **Gordin's Condition Restricting the Degree of Serial Correlation**: Assumption for Asymptotic Normality
The vector process $\{\mathbf{g}_t\} = \mathbf{x}_t . \varepsilon_t$ satisfies the Gordin conditions:

a) $E(g_t g_t') = E(\varepsilon_t' x_t' x_t \varepsilon_t) < \infty$ (this is a restriction on [strictly] stationary processes which might not have finite 2nd moments)

b) $E(g_t \mid g_{t-j}, g_{t-j-1}, ...) \xrightarrow{m.s.} 0 \ \ as \ \ j \to \infty \ \ or \ \ equivalently \ \ (\text{since stationary}) \ \ E(g_0 \mid g_{-j}, g_{-j-1}, ...) \xrightarrow{m.s.} 0$

c) $\sum_{k=0}^{\infty} E\left(r_{t,t-k}^{\ 2}\right)^{1/2} < \infty \ \ where \ \ r_{t,t-k} \equiv E(g_t \mid I_{t-k}) - E(g_t \mid I_{t-k-1}) \ \ and \ \ Info\ Set\ \ I_{t-k} \equiv (g_{t-k}, g_{t-k-1}, g_{t-k-2}, ...)$

(**Telescoping sum condition: restricts degree of autocorrelation**)

{$\mathbf{g}_t$} has nonsingular long-run covariance.

Then, by 3(b),

$$\sqrt{n}\bar{g} \to_D N(0,S), \ \ S = \sum_{j=-\infty}^{\infty} \Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j'\right) \ \ where \ \Gamma_j = E\left(g_t g_{t-j}'\right) \ \ for \ \ (j = 0, \pm 1, \pm 2, ...) = j-th \ \ auto\text{cov} \ \ matrix$$

Note: THIS IS THE ONLY DIFFERENCE BETWEEN THE GMM MODEL WITH AND WITHOUT SERIAL CORRELATION! Under the previous assumption 3.5 (martingale differences sequence), $S = E(g_i g_i') = \Gamma_0$
All the results developed previously carry over, using the new S (long run variance) matrix.

Note2: If we know the form of the autocorrelation, we could always use a GLS estimator.
However, GLS requires us to know $\Omega = \text{var}(\varepsilon \mid X)$ **, and is typically infeasible. Futhermore, when we have to estimate** $\Omega = \text{var}(\varepsilon \mid X)$ **, GLS may not dominate OLS!**
Thus, if we were to run OLS/GMM, then the consistent var-cov matrix is given here!

GMM Estimator with Serial Correlation

Results:

1. Consistency[30]: Under assumptions 3.1 – 3.4, $\mathbf{b}_{GMM}(\mathbf{W}_n) \to_p \delta$

2. Asymptotic Normality[31]: Under assumptions 3.1 – 3.5,
$$\sqrt{n}\left(b_{GMM}(W_n) - \delta\right) \xrightarrow{D} N(0, V)$$
$$where\ V = A\,\text{var}\left(b_{GMM}(W_n)\right) = \left(\Sigma'_{XZ}\ W\ \Sigma_{XZ}\right)^{-1} \Sigma'_{XZ}\ W\ S\ W\ \Sigma_{XZ}\left(\Sigma'_{XZ}\ W\ \Sigma_{XZ}\right)^{-1}$$
$$= \left(E(x_i z_i')' W\ E(x_i z_i')\right)^{-1} E(x_i z_i')'\ W\ E(g_i g_i') W\ E(x_i z_i')\left(E(x_i z_i')' W\ E(x_i z_i')\right)^{-1} \quad where\ W = p\lim W_n$$

3. Consistent Estimate of Avar($\mathbf{b}_{GMM}(\mathbf{W}_n)$)[32]:
Suppose there exists a consistent estimator **S*** of $\mathbf{S}_{mxm}$ = E($\mathbf{g}_i\mathbf{g}_i$'). Then, under 3.2, Avar($\mathbf{b}_{GMM}(\mathbf{W}_n)$) is consistently estimated by
$$\hat{A}\,\text{var}(\hat{b}_{GMM}) \equiv \left(S_{ZX}' W_n S_{ZX}\right)^{-1} S_{ZX}' W\ S^* W' S_{ZX}\left(S_{ZX}' W_n S_{ZX}\right)^{-1}$$

4. Consistency of s^2 (estimation of variance of "true" error is consistent)[33]:
For any consistent estimator $\hat{b}(W_n)$ *of* δ, *define* $\hat{\varepsilon}_i \equiv y_i - x_i'\hat{b}(W_n)$. *Under* 3.1, 3.2, *and assume* $E(x_i x_i')$ *exists and is finite,*
$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \xrightarrow{P} E\left(\varepsilon_i^2\right)\ \ provided\ E\left(\varepsilon_i^2\right)\ exists\ and\ is\ finite$$

5. Consistent estimation of S (long-run variance matrix): (see next)

---

[30] From above, since
$$\left(\frac{1}{n}\sum_{i=1}^{n} z_i \cdot \varepsilon_i\right) = \bar{g}_n \xrightarrow{P} E\left(z_i \cdot \varepsilon_i\right) = E(g_i) = 0\ by\ ergodic\ theorem\ and\ 3.3$$
$$\therefore\ \hat{b}_{GMM} \xrightarrow{P} \delta$$

[31] Continuing from above,
$$\hat{b}_{GMM} - \delta = \left(S_{ZX}' W_n S_{ZX}\right)^{-1} S_{ZX}' W_n s_{x\varepsilon} \Rightarrow \sqrt{n}\left(\hat{b}_{GMM} - \delta\right) = \left(S_{ZX}' W_n S_{ZX}\right)^{-1} S_{ZX}' W_n \sqrt{n} s_{z\varepsilon}$$
$$\sqrt{n} s_{x\varepsilon} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i \cdot \varepsilon_i\right) \xrightarrow{D} N(0, E(g_i g_i'))\ by\ Ergodic\ Martingale\ Differences\ CLT,\ S_{ZX} \xrightarrow{P} \Sigma_{XZ}\ by\ Ergodic\ Theorem,\ W_n \xrightarrow{P} W\ by\ construction$$
$$\therefore \sqrt{n}\left(\hat{b}_{GMM} - \delta\right) = \left(S_{ZX}' W_n S_{ZX}\right)^{-1} S_{ZX}' W_n \sqrt{n} s_{z\varepsilon} \xrightarrow{D} N\left(0, \left(\Sigma_{ZX}' W_n \Sigma_{ZX}\right)^{-1} \Sigma_{ZX}' W\ E(g_i g_i') W' \Sigma_{ZX}\left(\Sigma_{ZX}' W_n \Sigma_{ZX}\right)^{-1}\right)\ by\ CMT\ and\ Slutsky's$$

[32] This follows from above. Standard asymptotic tools.

[33] This proof is very similar to 3D from previous notes
$$\hat{\varepsilon}_i = y_i - x_i'\hat{b} = y_i - x_i'\hat{b} + x_i'\delta - x_i'\delta = \varepsilon_i + x_i'(\hat{b} - \delta) \Rightarrow e_i^2 = \varepsilon_i^2 + 2\varepsilon_i x_i'(\hat{b} - \delta) + (\hat{b} - \delta)' x_i x_i'(\hat{b} - \delta)$$
$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 + 2\varepsilon_i x_i'(\hat{b} - \delta) + (\hat{b} - \delta)' x_i x_i'(\hat{b} - \delta) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 + 2(\hat{b} - \delta)\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i x_i'\right) + (\hat{b} - \delta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\hat{b} - \delta) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 + 2(\hat{b} - \beta)' s_{x\varepsilon} + (\hat{b} - \beta)' S_{XX}(\hat{b} - \beta)$$
$$2(\hat{b} - \beta)' s_{x\varepsilon} \xrightarrow{P} 0\ \text{since}\ s_{x\varepsilon} \xrightarrow{P} some\ finite\ vector\ and\ \hat{b} \xrightarrow{P} \beta$$
$$(\hat{b} - \beta)' S_{XX}(\hat{b} - \beta) \xrightarrow{P} 0\ \text{since}\ \hat{b} - \beta \xrightarrow{P} 0\ and\ S_{XX} \xrightarrow{P} \Sigma_{XX}\ finite\ by\ assumption$$

<u>Methods of Estimating S</u>: This is difficult, particularly when autocovariances do not vanish after finite lags

1. Estimating **S** when Autocovariances Vanish after Finite Lags
   We can estimate j-th order autocovariance for our vector process $\{g_t\}$ of moment conditions as follows:

$$Let\ \hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} \hat{g}_t \hat{g}_{t-j}' = \frac{1}{n}\sum_{t=j+1}^{n} x_t \hat{\varepsilon}_t^2 x_{t-j}' \ \ where\ \hat{g}_t = x_t \hat{\varepsilon}_t, \hat{\varepsilon}_t = y_t - z_t'\hat{\delta}\ \ for\ \ \hat{\delta}\ consistent\ \ for\ \delta$$ [34]

   If we know a priori that for a known and finite q, $\Gamma_j = 0\ \ for\ \ j > q$ (i.e. a finite lag after which there is no autocorrelation), then

   S is consistently estimated by: $\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{q} \hat{\Gamma}_j + \hat{\Gamma}_j'$

   If we do not know q, then the following are methods for estimating LR variance matrix:

2. Using Kernels to Estimate **S**
   Kernel-based (or non-parametric) estimators can be expressed as a weighted average of estimated autocovariances:

$$\boxed{\hat{S} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{q(n)}\right)\cdot\hat{\Gamma}_j}$$

   where the "kernel", k(.), is a function that gives weights to autocovariances, and q(n) is the "bandwidth"

   Truncated Kernel

   The above estimator in 1 is a special kernel-based estimator with q(n) = q and $k(x) = \begin{cases} 1 & for\ |x| \le 1 \\ 0 & for\ |x| > 1 \end{cases}$ a "truncated" kernel.

   Problems
   In this case, with unknown q, we can use the truncated kernel with a bandwidth q that increases with sample size. As q(n) increases to infinity, more and more (and ultimately all) autocovariances can be included in the calculation of **S***. **However, the truncated kernel-based S* is not guaranteed to be positive semi-definite in finite samples. This creates a problem because then the estimated asymptotic variance of the GMM estimator (which includes S) may not be positive semi-definite.**)

   Bartlett Kernel and Newey-West Estimator: Newey and West found that the kernel-based estimator can be made nonnegative definite in finite samples using the **Bartlett kernel**

$$\boxed{k(x) = \begin{cases} 1-|x| & for\ |x| \le 1 \\ 0 & for\ |x| > 1 \end{cases}}$$

   (looks like a bell-shaped curve bt [-1,1] that peaks at 0)

   Bartlett kernel based estimators of **S** is called in econometrics the Newey-West estimator.

$$\sum_{j=-n+1}^{n-1}\left(1-\frac{|j|}{q(n)}\right)\mathbb{I}_{\{|j|\le q(n)\}}\hat{\Gamma}_j = \hat{\Gamma}_0 + \sum_{j=1}^{q(n)}\left(\frac{q(n)-j}{q(n)}\right)\left(\hat{\Gamma}_j + \hat{\Gamma}_j'\right)$$

   Example: For q(n) = 3, the kernel-based estimator includes autocovariances up to two (not three) lags:

$$\boxed{\hat{S} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{3}\right)\cdot\hat{\Gamma}_j = k\left(\frac{0}{3}\right)\cdot\hat{\Gamma}_0 + k\left(\frac{1}{3}\right)\cdot(\hat{\Gamma}_1+\hat{\Gamma}_1') + k\left(\frac{2}{3}\right)\cdot(\hat{\Gamma}_2+\hat{\Gamma}_2') + k(1)\cdot(\hat{\Gamma}_3+\hat{\Gamma}_3') = \hat{\Gamma}_0 + \frac{2}{3}(\hat{\Gamma}_1+\hat{\Gamma}_1') + \frac{1}{3}\cdot(\hat{\Gamma}_2+\hat{\Gamma}_2')}$$

3. VARHAC

---

[34] Note that here g_t is no longer ergodic stationary because we need to estimate the error term. For consistency we need to impose a fourth-moment condition (similar to that in prop 3.4). See Hansen's notes p.56

Background: In time series regressions, {y_t, $\mathbf{x}_t$} not always stationary as we have assumed thus far. There are 2 cases in which OLS is still consistent: **1. Time Regressions (Ch 6) and 2. Cointegrating Regressions (Ch 10)**

Time Regressions

Setup:

Population Model of Interest:

$$y_t = \alpha + \delta \cdot t + \varepsilon_t \quad \textit{where } \{\varepsilon_t\} \textit{ is independent white noise } (\{\varepsilon_t\} \textit{ iid with } E(\varepsilon_t) = 0, Var(\varepsilon_t) \textit{ finite})$$

$$\textit{or } \; y_t = x_t^{'}\beta + \varepsilon_t \;\; \textit{with } x_t^{'} = (1, t), \; \beta = (\alpha, \delta)'$$

Trend Stationarity: Here, the random variable $\mathbf{x}_t$ = (1,t)' is not stationary, since the second element (t) has mean t, which increases with time. Similarly, y_t is not stationary since the **time trend**, α + δ.t is (trivially) non-stationary. However, y_t is **trend stationary**.

Def: A process/sequence of rv's is **trend stationary** if it can be written as the sum of a time trend and a stationary process (here the stationary component is the independent white noise)

Consistency of OLS estimates in Time Regressions: It turns out, OLS estimates $\hat{\alpha}, \hat{\delta}$ are consistent but have different rates of convergence.

Prop: Consider the time regression where ε_t is independent white noise with $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t^4) <$ inf and let $\hat{\alpha}, \hat{\beta}$ be the OLS estimates of α and δ. Then,

$$\begin{bmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\delta} - \delta) \end{bmatrix} \xrightarrow{D} N\left(0, \sigma^2 \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}^{-1}\right)$$

Just as in the stationary case, $\hat{\alpha}$ is $\sqrt{n}$ -consistent because $\sqrt{n}(\hat{\alpha} - \alpha)$ converges to a (normal) R.V. On the other hand, $\hat{\delta}$ is $n^{3/2}$- consistent in that $n^{3/2}(\hat{\delta} - \delta)$ converges to a R.V. Therefore, $\hat{\delta}$ is hyperconsistent[35].

Proof: See next page.

Hypothesis Testing for time Regressions:

For $H_0 : \hat{\alpha} = \alpha_0$

$$t = \frac{\hat{\alpha} - \alpha}{\sqrt{s^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\sum_t x_t x_t'\right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \frac{\sqrt{n}\left(\hat{\alpha} - \alpha\right)}{\sqrt{s^2 \begin{bmatrix} \sqrt{n} & 0 \end{bmatrix} \left(\sum_t x_t x_t'\right)^{-1} \begin{bmatrix} \sqrt{n} \\ 0 \end{bmatrix}}} = \frac{\sqrt{n}\left(\hat{\alpha} - \alpha\right)}{\sqrt{s^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \Upsilon_n \left(\sum_t x_t x_t'\right)^{-1} \Upsilon_n \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \frac{\sqrt{n}\left(\hat{\alpha} - \alpha\right)}{\sqrt{s^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \Upsilon_n \left(\sum_t x_t x_t'\right)^{-1} \Upsilon_n \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \frac{\sqrt{n}\left(\hat{\alpha}\right.}{\sqrt{s^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\Upsilon_n^{-1} \left(\sum\right.\right.}}$$

$$= \frac{\sqrt{n}\left(\hat{\alpha} - \alpha\right)}{\sqrt{s^2 \begin{bmatrix} 1 & 0 \end{bmatrix} Q_n^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}}$$

For $H_0 : \hat{\delta} = \delta_0$

$$t = \frac{\hat{\delta} - \delta_0}{\sqrt{s^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left(\sum_t x_t x_t'\right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}} = \frac{n^{3/2}\left(\hat{\delta} - \delta_0\right)}{\sqrt{s^2 \begin{bmatrix} 0 & n^{3/2} \end{bmatrix} \left(\sum_t x_t x_t'\right)^{-1} \begin{bmatrix} 0 \\ n^{3/2} \end{bmatrix}}} = \frac{n^{3/2}\left(\hat{\delta} - \delta_0\right)}{\sqrt{s^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \Upsilon_n \left(\sum_t x_t x_t'\right)^{-1} \Upsilon_n \begin{bmatrix} 0 \\ 1 \end{bmatrix}}} = \frac{n^{3/2}\left(\hat{\delta} - \delta_0\right)}{\sqrt{s^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \Upsilon_n \left(\sum_t x_t x_t'\right)^{-1} \Upsilon_n \begin{bmatrix} 0 \\ 1 \end{bmatrix}}} = \frac{n^{3/2}\left(\hat{\delta} - \delta_0\right)}{\sqrt{s^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left(\Upsilon_n^{-1} \left(\sum_t x_t x_t'\right) \Upsilon_n^{-1} \right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}}$$

$$= \frac{n^{3/2}\left(\hat{\delta} - \delta_0\right)}{\sqrt{s^2 \begin{bmatrix} 0 & 1 \end{bmatrix} Q_n^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}}$$

[35] Estimators that are n-consistent are usually called **superconsistent**. (οτηερσ υσε τηε τερμ φορ εστιματορσ τηατ αρε νγ consistent for γ > ½.

$$\hat{b} \equiv \begin{bmatrix} \hat{\alpha} \\ \hat{\delta} \end{bmatrix} = (X_t ' X_t)^{-1} X_t y_t = \left(\sum_t x_t x_t '\right)^{-1} \sum_t x_t \cdot y_t = \left(\sum_t x_t x_t '\right)^{-1} \sum_t x_t \cdot (x_t ' \beta + \varepsilon_t)$$

$$\hat{b} - \beta = \begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\delta} - \delta \end{bmatrix} = \left(\sum_t x_t x_t '\right)^{-1} \left(\sum_t x_t \cdot \varepsilon_t\right) \quad \text{with} \sum_t x_t x_t ' = \sum_t \begin{bmatrix} 1 \\ t \end{bmatrix} \begin{bmatrix} 1 & t \end{bmatrix} = \begin{bmatrix} \sum_t 1 & \sum_t t \\ \sum_t t & \sum_t t^2 \end{bmatrix} = \begin{bmatrix} n & (n+1)n/2 \\ (n+1)n/2 & (2n+1)(n+1)n/6 \end{bmatrix}$$

To get consistency result, let $\Upsilon_n = \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{bmatrix}$. *Multiply both sides by* Υ_n *to get...*

$$\Upsilon_n(\hat{b} - \beta) = \begin{bmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\delta} - \delta) \end{bmatrix} = \Upsilon_n \left(\sum_t x_t x_t '\right)^{-1} \left(\sum_t x_t \cdot \varepsilon_t\right) = \Upsilon_n \left(\sum_t x_t x_t '\right)^{-1} \Upsilon_n \Upsilon_n^{-1} \left(\sum_t x_t \cdot \varepsilon_t\right) = \left(\Upsilon_n^{-1} \left(\sum_t x_t x_t '\right) \Upsilon_n^{-1}\right)^{-1} \left(\Upsilon_n^{-1} \left(\sum_t x_t \cdot \varepsilon_t\right)\right)$$

$$\equiv (Q_n)^{-1} v_n$$

$$\Upsilon_n^{-1} = \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{bmatrix}^{-1} = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix}$$

$$Q_n = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} n & (n+1)n/2 \\ (n+1)n/2 & (2n+1)(n+1)n/6 \end{bmatrix} \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} n^{1/2} & (n+1)n^{-1/2}/2 \\ (n+1)n^{1/2}/2 & (2n+1)(n+1)n^{-1/2}/6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & (n+1)/(2n) \\ (n+1)/(2n) & (2n+1)(n+1)/(6n^2) \end{bmatrix}$$

$$v_n = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \left(\sum_t \begin{bmatrix} 1 \\ t \end{bmatrix} \cdot \varepsilon_t\right) = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} \sum_t \varepsilon_t \\ \sum_t t\varepsilon_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_t \varepsilon_t \\ \frac{1}{n^{3/2}} \sum_t t\varepsilon_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_t \frac{t}{n} \varepsilon_t \end{bmatrix}$$

Clearly, $Q_n \xrightarrow{n\to\infty} Q = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \Rightarrow (Q_n)^{-1} \xrightarrow{n\to\infty} \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1}$

Based on a CLT for nonstationary processes,

$v_n \xrightarrow{D} N(0, \sigma^2 Q)$

$$V = E(u_n u_n ') - E(u_n)E(u_n ') = E\begin{pmatrix} \frac{1}{\sqrt{n}}\left(\sum_t \varepsilon_t\right)\frac{1}{\sqrt{n}}\left(\sum_t \varepsilon_t\right) & \frac{1}{\sqrt{n}}\left(\sum_t \varepsilon_t\right)\frac{1}{\sqrt{n}}\left(\sum_t \frac{t}{n}\varepsilon_t\right) \\ \frac{1}{\sqrt{n}}\left(\sum_t \varepsilon_t\right)\frac{1}{\sqrt{n}}\left(\sum_t \frac{t}{n}\varepsilon_t\right) & \frac{1}{\sqrt{n}}\left(\sum_t \frac{t}{n}\varepsilon_t\right)\frac{1}{\sqrt{n}}\left(\sum_t \frac{t}{n}\varepsilon_t\right) \end{pmatrix} - E\begin{bmatrix} \frac{1}{\sqrt{n}} \sum_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_t \frac{t}{n} \varepsilon_t \end{bmatrix} E\begin{bmatrix} \frac{1}{\sqrt{n}} \sum_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_t \frac{t}{n} \varepsilon_t \end{bmatrix}'$$

$$= \begin{pmatrix} \frac{1}{n}E\left(\sum_t \varepsilon_t \sum_t \varepsilon_t\right) & \frac{1}{n}E\left(\sum_t \varepsilon_t \sum_t \frac{t}{n}\varepsilon_t\right) \\ \frac{1}{n}E\left(\sum_t \varepsilon_t \sum_t \frac{t}{n}\varepsilon_t\right) & \frac{1}{n}E\left(\sum_t \frac{t}{n}\varepsilon_t \sum_t \frac{t}{n}\varepsilon_t\right) \end{pmatrix} - \begin{bmatrix} \frac{1}{n} nE(\varepsilon_t) \\ \frac{1}{n} E\sum_t \frac{t}{n}\varepsilon_t \end{bmatrix} \begin{bmatrix} \frac{1}{n} nE(\varepsilon_t) \\ \frac{1}{n} E\sum_t \frac{t}{n}\varepsilon_t \end{bmatrix}' \quad \text{with } E(\varepsilon_t) = 0 \text{ so } E\sum_t \frac{t}{n}\varepsilon_t = 0$$

$$= \begin{pmatrix} \frac{1}{n}E\left(\sum_t \varepsilon_t^2\right) & \frac{1}{n}E\left(\sum_t \frac{t}{n}\varepsilon_t^2\right) \\ \frac{1}{n}E\left(\sum_t \frac{t}{n}\varepsilon_t^2\right) & \frac{1}{n}E\left(\sum_t \frac{t^2}{n^2}\varepsilon_t^2\right) \end{pmatrix} \text{since } Cov(e_i, e_j) = 0,\ E(e_i) = 0 \Rightarrow E(e_i e_j) = 0 \text{ for } i \neq j$$

$$= \begin{pmatrix} \frac{1}{n} nE(\varepsilon_t^2) & \frac{1}{n}\left[\frac{1}{n}E(\varepsilon_1^2) + \frac{2}{n}E(\varepsilon_2^2) + \ldots + \frac{n}{n}E(\varepsilon_n^2)\right] \\ \frac{1}{n}\left[\frac{1}{n}E(\varepsilon_1^2) + \frac{2}{n}E(\varepsilon_2^2) + \ldots + \frac{n}{n}E(\varepsilon_n^2)\right] & \frac{1}{n}\left[\frac{1^2}{n^2}E(\varepsilon_1^2) + \frac{2^2}{n^2}E(\varepsilon_2^2) + \ldots + \frac{n^2}{n}E(\varepsilon_n^2)\right] \end{pmatrix} \text{ with } E(\varepsilon_t^2) = \sigma^2 \text{ for all } t$$

$$= \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{n^2}(1+2+\ldots n) \\ \frac{\sigma^2}{n^2}(1+2+\ldots n) & \frac{\sigma^2}{n^3}(1^2+2^2+\ldots n^2) \end{pmatrix} = \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{n^2}(n+1)n/2 \\ \frac{\sigma^2}{n^2}(n+1)n/2 & \frac{\sigma^2}{n^2}(2n+1)(n+1)n/6 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \sigma^2(n+1)/2n \\ \sigma^2(n+1)/2n & \sigma^2(2n+1)(n+1)/6n^2 \end{pmatrix}$$

$$\xrightarrow{n\to\infty} \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} = \sigma^2 Q$$

$\therefore$ *By Slutsky,* $\begin{bmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\delta} - \delta) \end{bmatrix} = (Q_n)^{-1} v_n \xrightarrow{D} N(0, Q^{-1}\sigma^2 Q Q^{-1}{}') = N(0, \sigma^2 Q^{-1})$ since Q^{-1} *symmetric*

Cointegration Regressions
This is the other case where we don't have stationarity, but we can still apply OLS to obtain consistent results.

1. **Background: Modeling Trends**
   The idea of this chapter is that economic time series can be represented as the sum of…
   - **i.** a **linear time trend,**
   - **ii.** a **stochastic trend,**
   - **iii.** and a **stationary process**

   Def: Deterministic Trend / Linear Time Trend is a trend in the mean
   Def: Stochastic Trend is a Martingale[36], a trend where the best predictor of future value is its current value (and thus every change in the trend seems to have a permanent effect on its future value).

---

[36] Z_t is a martingale if $E(Z_t \mid Z_{t-1}, Z_{t-2}, \ldots) = Z_{t-1}$

Unit Root : Let $Y_t = Y_{t-1} + \varepsilon_t$ (where ε_t is iid white noise)

1. **Step Function (maps discrete time to cumulative shocks i.e Y_t):** $W_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \frac{\varepsilon_t}{\sigma_\varepsilon}$

2. **Asymptotic Results of $W_n(u)$**

- **CLT**[37]**:** For any $For\ any\ u \in [0,1],\ W_n(u) \to_D N(0,u)$
- **FCLT:** If ε_t is iid white noise, then $\forall\ u \in [0,1],\ W_n(u) \to_D W(u)$ (this is a convergence of a sequence of functions)
  Note: This means that for our step function converges to a Brownian motion at the limit, when the steps become finer and finer, or equivalently, when the shocks are accumulated in shorter and shorter periods of time until it becomes "continuous time"

2. **Useful Things to Keep in Mind in Deriving Asymptotic Convergence Results to Stochastic Integral:**

$$\sigma_e W_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t \to_D \sigma_e W(u)$$

$$\frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{t-1} \varepsilon_s \varepsilon_t \to_D \sigma_\varepsilon^2 \int_0^1 W(u)\ dW(u)$$

a) $W_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \frac{\varepsilon_t}{\sigma_\varepsilon}$

b) $W_n\left(\frac{t}{n}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{t} \frac{\varepsilon_t}{\sigma_\varepsilon} = \frac{1}{\sqrt{n}\sigma_\varepsilon} Y_t \quad OR \quad W_n\left(\frac{t-1}{n}\right) = \frac{1}{\sqrt{n}\sigma_\varepsilon} Y_{t-1}$

c) $W_n\left(\frac{t}{n} + \varepsilon\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{t} \frac{\varepsilon_t}{\sigma_\varepsilon} = \frac{1}{\sqrt{n}\sigma_\varepsilon} Y_t \quad for\ \varepsilon \in [0, \frac{1}{n})$

d) $\frac{1}{n} W_n\left(\frac{t}{n}\right) = \int_{\frac{t}{n}}^{\frac{t+1}{n}} W_n(u)\ du \quad this\ \ follows\ \ from\ (c)$

e) $\sum_{t=0}^{n-1} \frac{1}{n} W_n\left(\frac{t}{n}\right) = \sum_{t=0}^{n-1} \int_{\frac{t}{n}}^{\frac{t+1}{n}} W_n(u)\ du = \int_0^1 W_n(u)\ du \quad OR \quad \sum_{t=1}^{n} \frac{1}{n} W_n\left(\frac{t-1}{n}\right) = \sum_{t=1}^{n} \int_{\frac{t-1}{n}}^{\frac{t}{n}} W_n(u)\ du = \int_0^1 W_n(u)\ du$

f) $\frac{1}{n} Y_{t-1} = \frac{1}{n} \sum_{t=0}^{t-1} W_n\left(\frac{t-1}{n}\right) = \sum_{t=0}^{n} \int_{\frac{t-1}{n}}^{\frac{t}{n}} W_n(u)\ du = \int_0^1 W_n(u)\ du$

---

[37] $W_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \frac{\varepsilon_t}{\sigma_\varepsilon} = \sqrt{\frac{\lfloor un \rfloor}{n}} \frac{1}{\sqrt{\lfloor un \rfloor}} \sum_{t=1}^{\lfloor un \rfloor} \frac{\varepsilon_t}{\sigma_\varepsilon} \to_D N(0,u)$ since $\frac{1}{\sqrt{\lfloor un \rfloor}} \sum_{t=1}^{\lfloor un \rfloor} \frac{\varepsilon_t}{\sigma_\varepsilon} \to_D N(0,1)$ and $\sqrt{\frac{\lfloor un \rfloor}{n}} \to u$

Applications of Unit Root Tools

1. Unit Root and Regression Coefficient

Suppose we have the following unit root process, $Y_t = Y_{t-1} + \varepsilon_t = Y_0 + \sum_{s=1}^{t} \varepsilon_s = \sum_{s=1}^{t} \varepsilon_s$ (That is, let $Y_0 = 0$)

What is the distribution of sampling error when we regress Y_t on Y_{t-1}?

$$n(\hat{\phi}-1) = \frac{\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}\varepsilon_t}{\frac{1}{n}\frac{1}{n}\sum_{t=1}^{n}(Y_{t-1})^2} = \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t}{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right)^2} = \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t}{\frac{1}{n}\sum_{t=1}^{n}\left(\sigma_\varepsilon W_n\left(\frac{t-1}{n}\right)\right)^2} = \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t}{\sigma_\varepsilon^2\sum_{t=1}^{n}\frac{1}{n}\left(W_n\left(\frac{t-1}{n}\right)\right)^2}$$

$$= \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t}{\sigma_\varepsilon^2\sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}}(W_n(u))^2 du} = \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t}{\sigma_\varepsilon^2\int_0^1 (W_n(u))^2 du}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t \rightarrow_D \sigma_\varepsilon^2\int_0^1 W(u)\ dW(u) \ \ and \ \ \sigma_\varepsilon^2\int_0^1 (W_n(u))^2 du \rightarrow_D \sigma_\varepsilon^2\int_0^1 (W(u))^2 du \ \ by\ CMT$$

$$\Rightarrow n(\hat{\phi}-1) \rightarrow_D \frac{\int_0^1 W(u)\ dW(u)}{\int_0^1 (W(u))^2 du} \ \ by\ CMT \ \ (This\ is\ Dickie-Fuller)$$

2. AR(1) with Unit Root and Constant

Suppose we have the following unit root process: $Y_t = Y_{t-1} + \mu + \varepsilon_t$

(Here each period there are 2 terms in the shock)

What is the distribution of sampling error of the OLS estimate from regression Y_t on $(1, Y_{t-1})$

$$n\left(\hat{\phi}-\phi\right) = \frac{\sum_{t=1}^{n}\left(Y_{t-1}-\bar{Y}_-\right)\varepsilon_t}{\frac{1}{n}\sum_{t=1}^{n}\left(Y_{t-1}-\bar{Y}_-\right)^2} = \frac{\sum_{t=1}^{n}\left(Y_{t-1}\varepsilon_t\right)-\sum_{t=1}^{n}\bar{Y}_-\varepsilon_t}{\frac{1}{n}\sum_{t=1}^{n}\left(Y_{t-1}-\bar{Y}_-\right)^2}$$

$$Now,\ \bar{Y}_- = \frac{1}{n}\sum_{t=1}^{n}Y_{t-1} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\mu+\varepsilon_s) = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\varepsilon_s) + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\mu) = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\varepsilon_s) + \frac{n(n-1)}{2n}\mu = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\varepsilon_s) + \frac{n-1}{2}\mu$$

$$\Rightarrow\left(\bar{Y}_- - \frac{(n-1)}{2}\mu\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}(\varepsilon_s)\right) = \sigma_\varepsilon\sum_{t=1}^{n}W_n\left(\frac{t-1}{n}\right)$$

$$\Rightarrow\frac{1}{\sqrt{n}}\left(\bar{Y}_- - \frac{(n-1)}{2}\mu\right) = \sigma_\varepsilon\sum_{t=1}^{n}\frac{1}{n}W_n\left(\frac{t-1}{n}\right) = \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}}W_n\left(\frac{t-1}{n}\right) = \int_0^1 W_n(u)\ du \equiv \bar{W}_n$$

$$Similarly,\ Y_{t-1} = \sum_{s=1}^{t-1}(\varepsilon_s+\mu) = \sum_{s=1}^{t-1}(\varepsilon_s) + \sum_{s=1}^{t-1}(\mu) = \sum_{s=1}^{t-1}(\varepsilon_s) + (t-1)\mu = \sigma_\varepsilon\sqrt{n}W_n\left(\frac{t-1}{n}\right) + (t-1)\mu$$

$$\Rightarrow\left[Y_{t-1}-(t-1)\mu\right] = \sigma_\varepsilon\sqrt{n}W_n\left(\frac{t-1}{n}\right) \Rightarrow \frac{1}{\sqrt{n}}\left[Y_{t-1}-(t-1)\mu\right] = \sigma_\varepsilon W_n\left(\frac{t-1}{n}\right)$$

$$\therefore n\left(\hat{\phi}-\phi\right) = \frac{\frac{1}{n}\sum_{t=1}^{n}\left(\left[Y_{t-1}-(t-1)\mu\right]-\left[\bar{Y}_- - \frac{n-1}{2}\mu\right]\right)\varepsilon_t}{\frac{1}{n}\frac{1}{n}\sum_{t=1}^{n}\left(\left[Y_{t-1}-\frac{n(n-1)}{2}\mu\right]-\left[\bar{Y}_- - \frac{n(n-1)}{2}\mu\right]\right)^2} = \frac{\sigma_\varepsilon\sum_{t=1}^{n}\left(W_n\left(\frac{t-1}{n}\right)-\bar{W}_n\right)\frac{1}{\sqrt{n}}\varepsilon_t}{\frac{1}{n}\sigma_\varepsilon\sum_{t=1}^{n}\left(W_n\left(\frac{t-1}{n}\right)-\bar{W}_n\right)^2}\ldots$$

3. Spurious Regression Example

Suppose we're estimating

$$Y_t = \beta X_t + \varepsilon_t \ \ by\ OLS,\ where\ Y_t = \sum_{s=1}^{t} \varepsilon_s \quad X_t = \sum_{s=1}^{t} \eta_s\ ,\ W_n\left(\frac{t}{n}\right) \equiv \frac{1}{\sqrt{n}\sigma_\varepsilon} Y_{t+1} \quad V_n\left(\frac{t}{n}\right) \equiv \frac{1}{\sqrt{n}\sigma_\varepsilon} X_{t+1}$$

$$\hat{\beta} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2} = \frac{\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t}\varepsilon_s \sum_{s=1}^{t}\eta_s}{\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=1}^{t}\eta_s\right)^2} = \frac{\sum_{t=1}^{n}\frac{1}{\sqrt{n}}\sum_{s=1}^{t}\varepsilon_s \frac{1}{\sqrt{n}}\sum_{s=1}^{t}\eta_s}{\sum_{t=1}^{n}\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t}\eta_s\right)^2} = \frac{\sum_{t=1}^{n}\sigma_\varepsilon W_n\left(\frac{t-1}{n}\right)\sigma_\eta V_n\left(\frac{t-1}{n}\right)}{\sum_{t=1}^{n}\left(\sigma_\eta V_n\left(\frac{t-1}{n}\right)\right)^2} = \frac{\sigma_\varepsilon\sigma_\eta\sum_{t=1}^{n}\frac{1}{n}W_n\left(\frac{t-1}{n}\right)V_n\left(\frac{t-1}{n}\right)}{\sigma_\eta^2\sum_{t=1}^{n}\frac{1}{n}\left(V_n\left(\frac{t-1}{n}\right)\right)^2}$$

$$= \frac{\sigma_\varepsilon \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}} W_n\left(\frac{t-1}{n}\right)V_n\left(\frac{t-1}{n}\right)}{\sigma_\eta \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}}\left(V_n\left(\frac{t-1}{n}\right)\right)^2} = \frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du} \to_D \frac{\sigma_\varepsilon\int_0^1 W(u)V(u)\,du}{\sigma_\eta\int_0^1\left[V(u)\right]^2 du}$$

What is the distribution of variance of errors?

$$\hat{e}_t = Y_t - \hat{\beta}X_t \Rightarrow \hat{e}_t^2 = \left(Y_t - \hat{\beta}X_t\right)^2 = Y_t^2 - 2\hat{\beta}X_tY_t + \left(\hat{\beta}X_t\right)^2$$

$$\Rightarrow \sum_{t=1}^{n}\hat{e}_t^2 = \sum_{t=1}^{n}Y_t^2 - 2\hat{\beta}\sum_{t=1}^{n}X_tY_t + \hat{\beta}^2\sum_{t=1}^{n}(X_t)^2 = \sum_{t=1}^{n}\left(\sum_{s=1}^{t}\varepsilon_s\right)^2 - 2\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\sum_{t=1}^{n}X_tY_t + \left[\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\right]^2\sum_{t=1}^{n}(X_t)^2$$

$$\Rightarrow \frac{1}{n^2}\sum_{t=1}^{n}\hat{e}_t^2 = \sum_{t=1}^{n}\frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t}\varepsilon_s\right)^2 - 2\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{n}}X_t\frac{1}{\sqrt{n}}Y_t + \left[\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\right]^2\sum_{t=1}^{n}\frac{1}{n}\left(\frac{1}{\sqrt{n}}X_t\right)^2$$

$$\Rightarrow \frac{1}{n^2}\sum_{t=1}^{n}\hat{e}_t^2 = \sigma_\varepsilon\int_0^1\left[W_n(u)\right]^2 du - 2\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\sigma_\varepsilon\sigma_\eta\int_0^1 W_n(u)V_n(u)\,du + \left[\frac{\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}\right]^2\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du$$

$$= \sigma_\varepsilon\int_0^1\left[W_n(u)\right]^2 du - 2\frac{\left[\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du\right]^2}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du} + \frac{\left[\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du\right]^2}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}$$

$$= \sigma_\varepsilon\int_0^1\left[W_n(u)\right]^2 du - \frac{\left[\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du\right]^2}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}$$

$$\to_D \sigma_\varepsilon\int_0^1\left[W(u)\right]^2 du - \frac{\left[\sigma_\varepsilon\int_0^1 W(u)V(u)\,du\right]^2}{\sigma_\eta\int_0^1\left[V(u)\right]^2 du}$$

$$Since\ \hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{t=1}^{n}X_t^2} = \frac{\frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2}{\sum_{t=1}^{n}X_t^2} = \frac{\frac{1}{n^2}\sum_{t=1}^{n}\hat{e}_t^2}{\frac{1}{n}\sum_{t=1}^{n}X_t^2} = \frac{1}{n}\frac{\sigma_\varepsilon\int_0^1\left[W_n(u)\right]^2 du - \frac{\left[\sigma_\varepsilon\int_0^1 W_n(u)V_n(u)\,du\right]^2}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du}}{\sigma_\eta\int_0^1\left[V_n(u)\right]^2 du} \to 0$$