Review of fundamental concepts in time series analysis

Constantino Hevia
Universidad Torcuato Di Tella

Introduction

1. Definitions and basic building blocks of linear stochastic processes
2. Linear least squares and recursive projections
3. Wold representation theorem
4. Brief review of limit theorems

Time series and stochastic processes

- A time series $\{x_1, x_2, ..., x_T\}$ is a set of repeated observations of a variable over time $t = 1, 2, .., T$.
- A stochastic process x_t is a collection of random variables

$$\mathbf{x} = \{x_t\}_{t=-\infty}^{\infty} = \{...x_{-2}, x_{-1}, x_0, x_1, x_2, ...\} .$$

- On each drawing of the stochastic process, we draw an entire *sequence* $\{x_t\}_{t=-\infty}^{\infty}$.
- We assume that each $x_t \in L^2$ (Hilbert space of squared integrable r.v.). That is,

$$E[x_t^2] < \infty.$$

- Hilbert space: complete normed linear space where the norm is defined in terms of an *inner product.*
- Given a pair $x, y \in L^2$, the inner product is defined as

$$\langle x, y \rangle = E[xy].$$

- The norm associated with this inner product is

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{E[x^2]}$$

- **Lemma 1** *(Cauchy-Schwarz inequality)* Let $x, y \in L^2$. Then,

$$|E[xy]| \leq \sqrt{E[x^2]}\sqrt{E[y^2]}. \tag{1}$$

Proof:

$$0 \leq \left(\frac{|x|}{\sqrt{E[x^2]}} - \frac{|y|}{\sqrt{E[y^2]}} \right)^2 = \frac{|x|^2}{E\left[x^2\right]} + \frac{|y|^2}{E\left[y^2\right]} - 2\frac{|x|\,|y|}{\sqrt{E[x^2]}\sqrt{E[y^2]}}$$

Rearranging,

$$\frac{|x|\,|y|}{\sqrt{E[x^2]}\sqrt{E[y^2]}} \leq \frac{1}{2}\left[\frac{|x|^2}{E[x^2]} + \frac{|y|^2}{E[y^2]}\right]$$

Taking expectations

$$E\,|x|\,|y| \leq \sqrt{E[x^2]}\sqrt{E[y^2]}.$$

But $|E\left[xy\right]| \leq E\left[|x|\,|y|\right]$, which leads to

$$|E\left[xy\right]| \leq \sqrt{E[x^2]}\sqrt{E[y^2]}. \blacksquare$$

- Let the mean and covariances of the process $\{x_t\}$ be

$$\begin{aligned} \mu_t &= E[x_t] \\ \sigma_{t,t-\tau} &= E\left[(x_t - \mu_t)(x_{t-\tau} - \mu_{t-\tau})\right]. \end{aligned}$$

- The process $\{x_t\}$ is *covariance stationary* if

$$\mu_t = \mu \text{ for all } t$$

$$E\left[(x_t - \mu)(x_{t-\tau} - \mu)\right] = E\left[(x_{t+s} - \mu)(x_{t+s-\tau} - \mu)\right] \text{ for all } t, s, \tau$$

- *Autocovariance function* is the sequence

$$\gamma(\tau) = E\left[(x_t - \mu)(x_{t-\tau} - \mu)\right]$$

- The autocovariance function is symmetric:

$$\gamma(\tau) = \gamma(-\tau)$$

- Cauchy-Schwarz inequality implies

$$\begin{aligned} \left|E\left[(x_t - \mu)(x_{t-\tau} - \mu)\right]\right| &\leq \sqrt{E[(x_t - \mu)^2]}\sqrt{E[(x_{t-\tau} - \mu)^2]} \\ |\gamma(\tau)| &\leq \gamma(0) \text{ for all } \tau \end{aligned}$$

- It is usual to construct x_t through linear combinations a of serially uncorrelated white noise shocks ε_t,

$$\begin{aligned} E\left[\varepsilon_t\right] &= 0 \text{ for all } t, \\ E\left[\varepsilon_t^2\right] &= \sigma^2 \text{ for all } t, \\ E\left[\varepsilon_t \varepsilon_{t-\tau}\right] &= 0 \text{ for all } t \text{ and } \tau \neq 0. \end{aligned}$$

- Consider the $MA(\infty)$ stochastic process

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \text{ where } \sum_{j=0}^{\infty} \theta_j^2 < \infty \tag{2}$$

- The Wold Representation Theorem implies that (2) is, for most purposes, "sufficiently general."

- Family of ARMA processes is an example of (2)

Examples of ARMA processes

- AR(p):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \varepsilon_t$$

- MA(q):

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

- ARMA(p,q):

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

The "Lag" operator

- The lag operator takes a sequence x_t and moves the index back one period

$$Lx_t = x_{t-1}.$$

therefore,

$$L^p x_t = x_{t-p}.$$

- The "forward" operator is the inverse of the lag operator

$$L^{-p} x_t = x_{t+p}$$

- The lag polynomials $\theta(L)$ is defined as

$$\theta(L) = \theta_0 + \theta_1 L + \theta_2 L^2 + ... = \sum_{j=1}^{\infty} \theta_j L^j$$

- The process (2) can be written as

$$y_t = \theta(L)\varepsilon_t = \left(\sum_{j=0}^{\infty} \theta_j L^j\right)\varepsilon_t$$

ARMA processes written in terms of the lag operator

- AR(p):

$$\left(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p\right) x_t=\varepsilon_t$$

- MA(q):

$$x_t=\left(1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q\right) \varepsilon_t$$

- ARMA(p,q):

$$\left(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p\right) x_t=\left(1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q\right) \varepsilon_t$$

- We manipulate lag polynomials as if they were regular polynomials.
- AR(1) example by repeated substitution

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_t \\ &\vdots \\ &= \phi^{s+1} x_{t-s-1} + \phi^s \varepsilon_{t-s} + \phi^{s-1} \varepsilon_{t-s+1} + \ldots + \phi \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

if $|\phi| < 1$, $\phi^{s+1} x_{t-s-1}$ tends to zero in the mean-squared sense

$$\lim_{s \to \infty} E(\phi^{s+1} x_{t-s-1})^2 = \lim_{s \to \infty} \phi^{2(s+1)} E(x^2) = 0.$$

- Taking the limit as $s \to \infty$ gives the $MA(\infty)$ representation

$$x_t = \sum_{s=0}^{\infty} \phi_s \varepsilon_{t-s} = \left[1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \ldots \right] \varepsilon_t$$

- Obtain the same expression using the lag operator.
- Write the AR(1) as

$$(1 - \phi L)\, x_t = \varepsilon_t .$$

- *Invert* the lag polynomial $(1 - \phi L) \Rightarrow (1 - \phi L)^{-1}$.
- Recall the geometric series expansion for $|c| < 1$,

$$\frac{1}{1 - c} = 1 + c + c^2 + c^3 + \ldots.$$

  Treat ϕL like a number with the hope that $|\phi| < 1$ implies $|\phi L| < 1$ in some sense. Hence,

$$(1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \ldots.$$

- Therefore

$$x_t = \frac{\varepsilon_t}{1 - \phi L} = \left[1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \ldots\right] \varepsilon_t .$$

Multiplication of lag polynomials: let

$$a(L) = a_0 + a_1 L$$

$$b(L) = b_0 + b_1 L$$

then,

$$\begin{aligned} a(L)\, b(L) &= (a_0 + a_1 L)(b_0 + b_1 L) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) L + b_1 a_1 L^2 \end{aligned}$$

A trick for lag operators

- Write the AR(2) model,

$$\left(1 - \phi_1 L - \phi_2 L^2\right) x_t = \varepsilon_t,$$

  in terms of a $MA(\infty)$ representation.
- Inverting the second order lag polynomial is difficult. Write instead

$$\begin{aligned} 1 - \phi_1 L - \phi_2 L^2 &= (1 - \lambda_1 L)(1 - \lambda_2 L) \\ &= 1 - (\lambda_1 + \lambda_2) L + \lambda_1 \lambda_2 L^2 \end{aligned}$$

  λ_1 and λ_2 solve $\lambda_1 + \lambda_2 = \phi_1$ and $\lambda_1 \lambda_2 = -\phi_2$.
- Therefore,

$$(1 - \lambda_1 L)(1 - \lambda_2 L) x_t = \varepsilon_t.$$

- Polynomials are invertible if $|\lambda_1| < 1$ and $|\lambda_2| < 1$,

$$\begin{aligned} x_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\varepsilon_t \\ &= \left(\sum_{j=0}^{\infty}\lambda_1^j L^j\right)\left(\sum_{i=0}^{\infty}\lambda_2^i L^i\right)\varepsilon_t. \end{aligned}$$

- Still ugly. When $\lambda_1 \neq \lambda_2$ use partial fraction expansions

$$\begin{aligned} \frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} &= \frac{a}{1-\lambda_1 L} + \frac{b}{1-\lambda_2 L} \\ &= \frac{a(1-\lambda_2 L) + b(1-\lambda_1 L)}{(1-\lambda_1 L)(1-\lambda_2 L)} \\ &= \frac{a+b-(a\lambda_2 + b\lambda_1)L}{(1-\lambda_1 L)(1-\lambda_2 L)} \end{aligned}$$

which implies

$$a = \frac{\lambda_1}{\lambda_1-\lambda_2},\ \ b = \frac{\lambda_2}{\lambda_2-\lambda_1}.$$

- Therefore,

$$x_t = \sum_{j=0}^{\infty}\left[\frac{\lambda_1}{\lambda_1-\lambda_2}\lambda_1^j + \frac{\lambda_2}{\lambda_2-\lambda_1}\lambda_2^j\right]\varepsilon_{t-j}.$$

- In general, if we have an $AR(p)$ process we need to find the p roots of the polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0.$$

- The λ's are the reciprocal of these roots.
- The $AR(p)$ is invertible if all roots of the polynomial are greater than 1 in absolute value (all λ's, are *less* than 1 in absolute value)
- In this case we can write the $AR(p)$ model as

$$y_t = \left[(1 - \lambda_1 L)(1 - \lambda_2 L) \ldots (1 - \lambda_p L)\right]^{-1} \varepsilon_t$$

- If all λ's are different, the partial fractions expansions implies

$$x_t = \sum_{j=0}^{\infty} \left(\sum_{i=1}^{p} a_i \lambda_i^j \right) \varepsilon_{t-j}$$

  where

$$a_i = \frac{\lambda_i}{\prod_{j \neq i} (\lambda_i - \lambda_j)} \text{ for all } i,$$

Stationarity of a MA process

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$$

- $MA(\infty)$ is stationary if and only if $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.
- The unconditional mean does not depend on time

$$E\left[x_t\right] = \sum_{j=0}^{\infty} \theta_j E\left[\varepsilon_t\right] = 0$$

- The variance of x_t is given by

$$E\left[x_t^2\right] = E\left[\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}\right]^2 = \sum_{j=0}^{\infty} \theta_j^2 E\left[\varepsilon_{t-j}^2\right] = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 < \infty$$

- The autocovariance depends only on τ :

$$\gamma\left(\tau\right) = E\left[x_t x_{t-\tau}\right] = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j-\tau},$$

Linear projections

- Let y, x_1, x_2, ... , x_n be random variables in L^2.
- Consider estimating y on the basis of knowing x_1, x_2,..., x_n.
- Compute the '*linear projection*' that best approximates y,

$$\hat{y} = a_0 + a_1 x_1 + ... + a_n x_n$$

- The problem is

$$\min_{\{a_i\}} E\left[(y - a_0 x_0 - a_1 x_1 - ... - a_n x_n)^2\right], \qquad (3)$$

  where $x_0 \equiv 1$.
- Like OLS but using population moments.

Orthogonality principle: $a_0, a_1, a_2 ..., a_n$ minimize (3) if and only if

$$E\left[\left(y - a_0 x_0 - a_1 x_1 - ... - a_n x_n\right) x_i\right] = 0 \text{ for } i = 0, 1, 2, ..., n. \qquad (4)$$

Proof: Let $a = (a_0, a_1, ..., a_n)'$ and consider

$$\min_a J(a) = \min_a \frac{1}{2} E[(y - \sum_{j=0}^{n} a_j x_j)^2]$$

Necessity: the FOC is

$$\frac{\partial J(a)}{\partial a_i} = -E[(y - \sum_{j=0}^{n} a_j x_j) x_i] = 0 \text{ for } i = 0, 1, 2, ..., n.$$

or

$$\nabla_a J(a) = -\left(E[xy] - E\left[xx'\right] a\right) = \mathbf{0}_{n+1 \times 1}.$$

Sufficiency: differentiate FOC with respect to a

$$\nabla_{aa'} J(a) = E\left[xx'\right]$$

which is positive definite because $E\left[xx'\right]$ is a covariance matrix.$\square$

- The constants a satisfy the normal equations

$$a = E\left[xx'\right]^{-1} E\left[xy\right].$$

- The "prediction" error $y - \sum_{j=0}^{n} a_j x_j$ is *orthogonal* to each of the x_i's. Therefore, can write

$$y = \sum_{j=0}^{n} a_j x_j + \varepsilon \tag{5}$$

  where $E\left[\varepsilon \sum_{i=0}^{n} \phi_i x_i\right] = 0$ for any $\{\phi_i\}$.

- (5) decomposes y into two orthogonal components

$$E\left[y^2\right] = E\left[\left(\sum_{j=0}^{n} a_j x_j\right)^2\right] + E\left[\varepsilon^2\right].$$

- *Projection* of y on $\mathbf{x} \equiv \{1, x_1, x_2, ..., x_n\}$.

$$P\left[y|\mathbf{x}\right] \equiv x'a = \sum_{j=0}^{n} a_j x_j.$$

Lemma 2: The *projection is a linear operator,*

$$P\left[\alpha y + \beta z | \mathbf{x}\right] = \alpha P\left[y | \mathbf{x}\right] + \beta P\left[z | \mathbf{x}\right].$$

<u>*Proof*</u>: Let $P\left[y|\mathbf{x}\right] = \sum_{j=0}^{n} a_j x_j$ and $P\left[z|\mathbf{x}\right] = \sum_{j=0}^{n} b_j x_j$.

- The orthogonality principle implies

$$\begin{aligned} E\left(y - \sum_{j=0}^{n} a_j x_j\right) x_i &= 0 \text{ for all } i \\ E\left(z - \sum_{j=0}^{n} b_j x_j\right) x_i &= 0 \text{ for all } i \end{aligned}$$

Multiplying the first condition by α and the second by β gives

$$\begin{aligned} E\left(\alpha y - \alpha \sum_{j=0}^{n} a_j x_j\right) x_i &= 0 \text{ for all } i \\ E\left(\beta z - \beta \sum_{j=0}^{n} b_j x_j\right) x_i &= 0 \text{ for all } i \end{aligned}$$

- Adding these equations gives

$$E\left[\alpha y+\beta z-\sum_{j=0}^{n}\left(\alpha a_j+\beta b_j\right)x_j\right]x_i=0 \text{ for all } i.$$

- $\left(\alpha a_j+\beta b_j\right)$ for all j satisfy the orthogonality principle of a projection of $\alpha y+\beta z$ on $\mathbf{x}=\{1,x_1,x_2,\dots,x_n\}$

- Therefore,

$$P\left[\alpha y+\beta z|\mathbf{x}\right]=\alpha P\left[y|\mathbf{x}\right]+\beta P\left[z|\mathbf{x}\right].\blacksquare$$

Recursive projections

- **Problem**: Update a projection when new information arrives.
- Given $\Omega = \{1, x_1, x_2, ..., x_n\}$ we have the projection $P[y|\Omega]$.
- Observe $\mathbf{z} = (z_1, z_2, ..., z_m)'$ and want to compute $P[y|\Omega, z]$ given $P[y|\Omega]$
- Decomposition (5) for the updated projection:

$$y = P[y|\Omega, \mathbf{z}] + \varepsilon = \sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s z_s + \varepsilon \tag{6}$$

  $E(\varepsilon) = 0,\ E(\varepsilon x_j) = 0,\ E(\varepsilon z_s) = 0.$
- Project both sides on the smaller set Ω, so that

$$\begin{aligned} P[y|\Omega] &= P\left[\sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s z_s + \varepsilon|\Omega\right] \\ &= \sum_{j=0}^{n} a_j P[x_j|\Omega] + \sum_{s=1}^{m} \delta_s P[z_s|\Omega] + P[\varepsilon|\Omega]. \end{aligned}$$

  But $P[x_j|\Omega] = x_j$ and $P[\varepsilon|\Omega] = 0$. (Why?).

- Therefore,

$$P\left[y|\Omega\right] = \sum_{j=0}^{n} a_j x_j + \sum_{s=1}^{m} \delta_s P\left[z_s|\Omega\right]. \tag{7}$$

- Subtracting (7) from (6)

$$y - P\left[y|\Omega\right] = \sum_{s=1}^{m} \delta_s \left(z_s - P\left[z_s|\Omega\right]\right) + \varepsilon. \tag{8}$$

- This looks like a projection of the prediction error $y - P\left[y|\Omega\right]$ on the prediction errors $z_s - P\left[z_s|\Omega\right]$.
- To prove it, we need to show that ε is orthogonal to $z_s - P\left[z_s|\Omega\right]$ for all s (Orthogonality Principle).
- But this is obvious because $\varepsilon \perp z_s$, $\varepsilon \perp x_j$ and $P\left[z_s|\Omega\right]$ is a linear function of $\left\{x_j\right\}$.

- Therefore, to update a linear projection:
  - We take the initial projection
  - And add the projection of prediction errors on prediction errors

$$
\begin{aligned}
P\left[y|\Omega, \mathbf{z}\right] = & \underbrace{P\left[y|\Omega\right]}_{\text{Original projection}} \\
& + \underbrace{P\left[\left(y - P\left[y|\Omega\right]\right) | \left(\mathbf{z} - P\left[\mathbf{z}|\Omega\right]\right)\right]}_{\text{Projection of prediction errors on prediction errors}} .
\end{aligned} \tag{9}
$$

$(\mathbf{z} - P[\mathbf{z}|\Omega])$ is the "new information" contained in $\mathbf{z}$.

Wold Representation Theorem

- We constructed a covariance stationary process by combining white noise shocks according to

$$x_t = \theta(L)\,\varepsilon_t,$$

where $a(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots$ and $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.

- The Wold representation theorem reverses the procedure.

- Theorem: any covariance stationary process can be written as an infinite order moving average plus a (linearly) deterministic term.

Wold Representation Theorem

Any mean zero, covariance stationary process $\{x_t\}$ *can be represented as*

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \eta_t$$

where

1. $\varepsilon_t = x_t - P\left[x_t | x_{t-1}, x_{t-2}, ...\right]$ is the prediction error of the projection of x_t on its lagged values,
2. $P\left[\varepsilon_t | x_{t-1}, x_{t-2}, ...\right] = 0$, $E\left(\varepsilon_t x_{t-j}\right) = 0$ for all $j \geq 1$; $E\left(\varepsilon_t^2\right) = \sigma^2$ for all t; $E\left(\varepsilon_t\right) = 0$ for all t; $E\left(\varepsilon_t \varepsilon_s\right) = 0$ for all $s \neq t$,
3. $\theta_0 = 1$; $\sum_{j=0}^{\infty} \theta_j^2 < \infty$,
4. $\{\theta_j\}$ and $\{\varepsilon_t\}$ are unique,
5. η_t is linearly deterministic; that is, $\eta_t = P\left[\eta_t | x_{t-1}, x_{t-2}, x_{t-3}, ...\right]$.

See my notes for a sketch of the proof.

What the theorem says and what it does not say

1. The ε_t's are a white noise but need not be i.i.d. or normally distributed.
2. Although $E\left(\varepsilon_t x_{t-j}\right) = 0$, $E\left(\varepsilon_t | x_{t-j}\right)$ need not be zero. This is the difference between orthogonality and independence:
   2.1 $x \sim N\left(0, \sigma^2\right)$, $y = x^2$. Then $E\left(xy\right) = E\left(x^3\right) = 0$ but $E\left(y|x\right) = x^2$.
3. The Wold decomposition is a purely probabilistic decomposition. The innovations ε_t don't have any structural interpretation.
4. The Wold decomposition is *one representation* of the process $\{x_t\}$. There could be other non-linear representations.
5. Moreover, the Wold decomposition is not even the *unique linear* $MA\left(\infty\right)$ representation of the process
6. We usually ignore the term η_t.

Wold Representation Theorem for vector processes

Let $X_t = [x_{1t}, x_{2t}, ..., x_{nt}]'$ be a covariance stationary vector stochastic process with $E[X_t] = 0$ and $E[X_t X'_{t-\tau}] = \Gamma_\tau$. This process can be represented as

$$X_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} + \eta_t \tag{10}$$

where

1. $\varepsilon_t = X_t - P[X_t | X_{t-1}, X_{t-2}, X_{t-3}, ...]$,
2. $P[\varepsilon_t | X_{t-1}, X_{t-2}, X_{t-3}, ...] = 0$, $E(\varepsilon_t X_{t-j}) = 0$ for all $j \geq 1$; $E(\varepsilon_t^2) = \Sigma$; $E(\varepsilon_t) = 0$; $E(\varepsilon_t \varepsilon'_s) = 0$ for all $s \neq t$,
3. Θ_j are $n \times n$ matrices that satisfy $\Theta_0 = I$; $\sum_{j=0}^{\infty} \Theta_j \Theta'_j < \infty$,
4. $\{\Theta_j\}$ and $\{\varepsilon_t\}$ are unique,
5. η_t is linearly deterministic: $\eta_t = P[\eta_t | X_{t-1}, X_{t-2}, X_{t-3}, ...]$.

A Remark on the Wold representation theorem

- Wold says that there is a unique representation of a stationary process as a $MA(\infty)$ satisfying 1-5.
- *It does not say that it is the unique MA representation*
- We can always write

$$X_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j} + \eta_t = \sum_{j=0}^{\infty} \Theta_j \Lambda \Lambda^{-1} \varepsilon_{t-j} + \eta_t = \sum_{j=0}^{\infty} \Phi_j \nu_{t-j} + \eta_t,$$

  where Λ is invertible, $\Phi_j = \Theta_j \Lambda$ and $\nu_{t-j} = \Lambda^{-1} \varepsilon_{t-j}$.

$$X_t = \sum_{j=0}^{\infty} \Phi_j \nu_{t-j} + \eta_t$$

  is *another* $MA(\infty)$ representation of the process X_t.
- The residual ν_t *is not* the forecast error of projecting X_t on its infinite history.
- *This non-uniqueness result will be used when discussing structural vector autoregressions later in the course.*

Brief review of limit theorems

- We use different versions of two limit theorems:
  1. Laws of Large Numbers (LLN)
  2. Central Limit Theorems (CLT).
- Both are concerned with the behavior of sample means under different assumptions.
  1. The LLN is about convergence—in probability, almost surely, in L^2—of the sample mean to the population mean.
  2. The CLT is about convergence in distribution of the sample mean. By appropriately weighting the sample mean by a function of the sample size (typically $\sqrt{T}$), the CLT provides a non-degenerate distribution theory for the sample mean

Properties of the sample mean of a vector process

- We have a sample of size T of a covariance stationary vector process $\{X_t\}$ with

$$\begin{aligned} E\left[X_t\right] &= \mu \\ E\left[\left(X_t - \mu\right)\left(X_{t-v} - \mu\right)'\right] &= \Gamma_v. \end{aligned}$$

  and

$$\sum_{v=-\infty}^{\infty} |\Gamma_v| < \infty,$$

- Consider the sample mean

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t.$$

- We want to obtain the mean and covariance matrix of the sample mean.

Properties of the sample mean of a vector process

Clearly,

$$E\left[\bar{X}_T\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[X_t\right] = \mu$$

The covariance matrix of the sample mean satisfies (see notes)

$$T \times E\left[\left(\bar{X}_T - \mu\right)\left(\bar{X}_T - \mu\right)'\right] = \sum_{v=-(T-1)}^{T-1} \Gamma_v - \sum_{v=-(T-1)}^{T-1} \frac{|v|}{T}\Gamma_v$$

The following is an important result:

Proposition:

$$\lim_{T\to\infty} T \times E\left[\left(\bar{X}_T - \mu\right)\left(\bar{X}_T - \mu\right)'\right] = \sum_{v=-\infty}^{\infty} \Gamma_v$$

Fundamental limit theorems with iid data

Suppose that $X_1, X_2, \ldots$ are iid random variables with $E(X_t) = \mu$ and $E(X_t - \mu)^2 = \sigma^2 < \infty$. Then,

Law of large numbers (LLN): $\frac{1}{T}\sum_{t=1}^{T} X_t \to \mu$ (converges in probability, a.s., in L^2)

Central limit theorem (CLT): $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu\right) \Rightarrow N\left(0, \sigma^2\right)$ (converges in distribution)

- These results use independence and give an idea of how quickly and in what sense the sample average $\frac{1}{T}\sum_{t=1}^{T} X_t$ tends to the population mean μ.
- In time series we don't have independence. There are, however, versions of these theorems for dependent data.

Fundamental limit theorems with dependent data

Suppose $\{X_t\}$ is covariance stationary with $E(X_t) = \mu$ and $Cov(X_s, X_t) = \gamma_{|s-t|}$ for all s, t and with absolutely summable autocovariances,

$$\sum_{j=-\infty}^{\infty} \left|\gamma_j\right| = c < \infty.$$

Then,
Law of large numbers (LLN): $\bar{X}_T \to \mu$

Central limit theorem (CLT): $\sqrt{T}\left(\bar{X}_T - \mu\right) \Rightarrow N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right)$