Time Series

Author: Wenxiao Yang

Institute: Haas School of Business, University of California Berkeley

Date: 2024

All models are wrong, but some are useful.

Contents

Chapter 1 Univariate Stationary Time Series Analysis

1.1 Goals and Challenge

Data in time series is denoted by

$$\{\underbrace{y_t}_{n\times 1} : 1 \le t \le T\}$$

Assumption 1.1

Each y_t is the realization of some random vector Y_t.

The **objective** is to provide data-based answers to questions about the distribution of $\{Y_t : 1 \le t \le T\}$.

The **challenge** we face is $Y_1, Y_2, ..., Y_T$ are *not necessarily independent*. Time series analysis gives the models and methods that can accommodate dependence.

1.2 Stochastic Processes

Some terminologies we need to know:

Definition 1.1 (Stochastic Process)

A **stochastic process** is a collection $\{Y_t : t \in \mathcal{T}\}$ of random variables/vectors (defined on the same probability space).

1. $\{Y_t : t \in \mathcal{T}\}$ is **discrete time process** if $\mathcal{T} = \{1, ..., T\}$ or $\mathcal{T} = \mathbb{N} = \{1, 2, ...\}$ or $\mathcal{T} = \mathbb{Z} = \{..., -1, 0, 1, ...\}$.
2. $\{Y_t : t \in \mathcal{T}\}$ is **continuous time process** if $\mathcal{T} = [0, 1]$ or $\mathcal{T} = \mathbb{R}_+$ or $\mathcal{T} = \mathbb{R}$.

Observed data Y_t is a realization of a discrete time process with $\mathcal{T} = \{1, ..., T\}$.

1.2.1 Strictly Stationary

Definition 1.2 (Strictly Stationary (Discrete and Scalar Process))

A scalar[a] process $\{Y_t : t \in \mathbb{Z}\}$ is **strictly stationary** *if and only if*

$$(Y_t, ..., Y_{t+k}) \underbrace{\sim}_{\text{"is distributed as"}} (Y_0, ..., Y_k), \ \forall t \in \mathbb{Z}, k \ge 0$$

[a] i.e., Y_t is 1×1

Note

1. *If $Y_t \sim i.i.d.$, then $\{Y_t : t \in \mathbb{Z}\}$ is strictly stationary.*
2. *If $\{Y_t : t \in \mathbb{Z}\}$ is strictly stationary, then Y_t are identically distributed (i.e., "marginal stationary").*

Example 1.1 Strictly Stationary and Dependent

A constant process that ... $= Y_{-1} = Y_0 = Y_1 =$ ... is strictly stationary.

All these above hold for strictly stationary vector process.

Lemma 1.1 (Property of Strictly Stationary)

If $\{Y_t : t \in \mathbb{Z}\}$ is strictly stationary with $\mathbb{E}[Y_t^2] < \infty$ $(\forall t)$, then

1. Same Expectation:
$$\mathbb{E}[Y_t] = \mu, \ \forall t \text{ (for some constant } \mu) \tag{*}$$
2. Covariance only depends on time length:
$$\text{Cov}(Y_t, Y_{t-j}) = \gamma(j), \ \forall t, j \text{ (for some function } \gamma(\cdot)) \tag{**}$$
Note $\gamma(0) = \text{Var}(Y_t), \forall t$.

1.2.2 Covariance Stationary

A subset of strictly stationary processes that has second moment (i.e., $\mathbb{E}[Y_t^2] < \infty$) can be defined as **covariance stationary**.

Definition 1.3 (Covariance Stationary)

A process $\{Y_t : t \in \mathbb{Z}\}$ is **covariance stationary** *iff* $\mathbb{E}[Y_t^2] < \infty$ $(\forall t)$ and it satisfies (*) and (**).

Note *Not every strictly stationary process is covariance stationary. (e.g., if it does not have second moment).*

1.2.3 Autocovariance and Autocorrelation Functions

Definition 1.4 (Autocovariance and Autocorrelation Functions)

$\gamma(\cdot)$ in (**) is called **autocovariance function** of $\{Y_t : t \in \mathbb{Z}\}$.

The **autocorrelation function** is $\rho(j) = \text{Corr}(Y_t, Y_{t-j}) = \frac{Cov(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-j})}} = \frac{\gamma(j)}{\gamma(0)}$.

Lemma 1.2 (ACF Property)

The autocovariance function satisfies the following properties:

1. $\gamma(\cdot)$ is **even** i.e., $\gamma(j) = \gamma(-j)$.

2. $\gamma(\cdot)$ is **positive semi-definite** (psd) i.e., for any $n \in \mathbb{N}$ and any $a_1, ..., a_n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j) = \text{Var}(\sum_{i=1}^{n} a_i Y_i) \geq 0$$

1.3 Moving-Average (MA) Process

Definition 1.5 (White Noise)

A process $\{\epsilon_t : t \in \mathbb{Z}\}$ is a **white noise** process iff it is covariance stationary with $\mathbb{E}[\epsilon_t] = 0$ and

$$\text{Cov}(\epsilon_t, \epsilon_{t-j}) = \begin{cases} \sigma^2, & \text{if } j = 0 \\ 0, & \text{otherwise} \end{cases}$$

We use *notation* $\epsilon_t \sim \text{WN}(0, \sigma^2)$.

Note

1. *If* $\epsilon_t \sim \text{i.i.d.}(0, \sigma^2)$*, then* $\{\epsilon_t : t \in \mathbb{Z}\}$ *is white noise, i.e.,* $\epsilon_t \sim \text{WN}(0, \sigma^2)$.
2. *Gauss-Markov theorem assumes WN errors.*
3. *WN terms are used as "building blocks": often a variable can be generated as*

$$Y_t = h(\epsilon_t, \epsilon_{t-1}, ...) \text{ for some function } h(\cdot) \text{ and some } \epsilon_t \sim \text{WN}(0, \sigma^2).$$

1.3.1 Moving-Average Process

Definition 1.6 (MA(1))

First-order moving average process: $Y_t \sim \text{MA}(1)$ iff

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

where $\epsilon_t \sim \text{WN}(0, \sigma^2)$.

Claim 1.1 (ACF of MA(1))

$\{Y_t\}$ is covariance stationary: $\mathbb{E}[Y_t] = \mu$ and its autocovariance function is

$$\gamma(j) := \text{Cov}(Y_t, Y_{t-j}) = \begin{cases} (1+\theta^2)\sigma^2, & j = 0 \\ \theta\sigma^2, & j = 1 \\ 0, & j \geq 2 \end{cases}$$

Definition 1.7 (MA(p))

$Y_t \sim \mathrm{MA}(q)$ (for some $q \in \mathbb{N}$) iff

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where $\epsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

Claim 1.2 (ACF of MA(p))

$\{Y_t\}$ is covariance stationary: $\mathbb{E}[Y_t] = \mu$ and its autocovariance function is

$$\gamma(j) := \mathrm{Cov}(Y_t, Y_{t-j}) = \begin{cases} \left(\sum_{i=0}^{q-j} \theta_i \theta_{i+j}\right) \sigma^2, & j \leq q \\ 0, & j \geq q+1 \end{cases}$$

where $\theta_0 = 1$.

Definition 1.8 (Infinite Moving-Average Process)

$Y_t \sim \mathrm{MA}(\infty)$ iff

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

where $\epsilon_t \sim \mathrm{WN}(0, \sigma^2)$ and $\underline{\sum_{i=0}^{\infty} \psi_i^2 < \infty}$

1.3.2 Conditions for Infinite Moving-Average Process

Note *Conjecture:*

1. $\{Y_t\}$ *is covariance stationary;*
2. $\mathbb{E}[Y_t] = \mu$ *and*
3. *its autocovariance function is*

$$\gamma(j) := \mathrm{Cov}(Y_t, Y_{t-j}) = (\sum_{i=0}^{\infty} \psi_i \psi_{i+j})\sigma^2, \forall j \geq 0.$$

The necessary condition to make these conjectures correct is

$$\begin{aligned} \mathbb{E}[Y_t^2] &= (\mathbb{E}[Y_t])^2 + \Gamma(0) \\ &= \mu^2 + (\sum_{i=0}^{\infty} \psi_i^2)\sigma^2 < \infty \\ \Leftrightarrow \sum_{i=0}^{\infty} \psi_i^2 < \infty \end{aligned}$$

Claim 1.3

With the `right' definition of ``$\sum_{i=0}^{\infty}$'', the conjecture is correct.

Remark

1. If $X_0, X_1, \ldots$ are i.i.d. with $X_0 = 0$, then $\sum_{i=0}^{\infty} X_i$ denote $\lim_{n\to\infty} \sum_{i=0}^{n} X_i$ (assuming the limit exists).
2. $\exists$ various models of stochastic convergence.
3. There: convergence in mean square.

Definition 1.9 (Stochastic Convergence in Mean Square)

If $X_0, X_1, \ldots$ are random (with $\mathbb{E}[X_i^2] < \infty, \forall i$), then $\sum_{i=0}^{\infty} X_i$ denotes any S such that $\lim_{n\to\infty} \mathbb{E}[(S - \sum_{i=0}^{n} X_i)^2] = 0$.

Lemma 1.3

The properties of the S are

1. S is ``essentially unique.''
2. $\mathbb{E}[S] = \sum_{i=0}^{\infty} \mathbb{E}[X_i] = \lim_{n\to\infty} \sum_{i=0}^{n} \mathbb{E}[X_i]$
3. $\text{Var}[S] = \ldots = \lim_{n\to\infty} \text{Var}[\sum_{i=0}^{n} X_i]$
4. (Higher order moments of S are similar) $\cdots$

Theorem 1.1 (Cauchy Criterion)

$\sum_{i=0}^{\infty} X_i$ exists iff

$$\lim_{n\to\infty} \sup_{m>n} \mathbb{E}[(S_m - S_n)^2] = 0,$$

where $S_n = \sum_{i=0}^{n} X_i$.

In the $MA(\infty)$ context: The condition that can make

$$\lim_{n\to\infty} \sup_{m>n} \mathbb{E}[(Y_{t,m} - Y_{t,n})^2] = 0$$

where $Y_{t,n} = \mu + \sum_{i=0}^{n} \psi_i \epsilon_{t-i}$.

This condition is given as: If $m > n$,

$$\begin{aligned}
Y_{t,m} - Y_{t,n} &= \sum_{i=n+1}^{m} \psi_i \epsilon_{t-i} \\
\Rightarrow \mathbb{E}[(Y_{t,m} - Y_{t,n})^2] &= \mathbb{E}\left[\left(\sum_{i=n+1}^{m} \psi_i \epsilon_{t-i}\right)^2\right] = \left(\sum_{i=n+1}^{m} \psi_i^2\right)\sigma^2 \\
\Rightarrow \sup_{m>n} \mathbb{E}[(Y_{t,m} - Y_{t,n})^2] &= \left(\sum_{i=n+1}^{\infty} \psi_i^2\right)\sigma^2 \\
\Rightarrow \lim_{n\to\infty} \sup_{m>n} \mathbb{E}[(Y_{t,m} - Y_{t,n})^2] &= \left(\lim_{n\to\infty} \sum_{i=n+1}^{\infty} \psi_i^2\right)\sigma^2
\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\sup_{m>n}\mathbb{E}[(Y_{t,m}-Y_{t,n})^2]=0 \text{ iff } \lim_{n\to\infty}\sum_{i=n+1}^{\infty}\psi_i^2=0$$
$$\text{iff } \sum_{i=0}^{\infty}\psi_i^2<\infty$$

1.3.3 Remarks about $MA(\infty)$ models

1. $MA(\infty)$ models are useful in theoretical work.
2. The $MA(\infty)$ class is "large": Wold decomposition (theorem).
3. Parametric $MA(\infty)$ models are useful in inference.

1.4 Autoregressive (AR) Model

1.4.1 Autoregressive Model as a Special Case of $MA(\infty)$

Autoregressive model is an example of well-defined $MA(\infty)$ model.

Example 1.2 (Autoregressive model)

Suppose

$$Y_t=\sum_{i=0}^{\infty}\psi_i\epsilon_{t-i},\ \forall t$$

where

- $\epsilon_t\sim \mathrm{WN}(0,\sigma^2)$;
- $\psi_i=\phi^i$ $(\forall i\geq 0)$ for some $|\phi|<1$.

Checking the condition: $\lim_{n\to\infty}\sum_{i=0}^{n}\psi_i^2=\lim_{n\to\infty}\sum_{i=0}^{n}\phi^{2i}=\lim_{n\to\infty}\frac{1-\phi^{2(n+1)}}{1-\phi^2}=\frac{1}{1-\phi^2}<\infty$.

Lemma 1.4 (Property of ACF of Autoregressive Model)

For $j\geq 0$, the autocovariance function is

$$\gamma(j)=\phi^j\gamma(0)$$

Note

1. *$\gamma(j)\neq 0,\forall j$ if $\phi\neq 0$.*
2. *$\gamma(j)\propto\phi^j$ decays exponentially.*

Proof 1.1

$\gamma(j):=\mathrm{Cov}(Y_t,Y_{t-j})=(\sum_{i=0}^{\infty}\psi_i\psi_{i+j})\sigma^2=\phi^j(\sum_{i=0}^{\infty}\phi^{2i})\sigma^2=\phi^j\frac{\sigma^2}{1-\phi^2}=\phi^j\gamma(0)$

1.4.2 Alternative Representation of AR Model

Definition 1.10 (Alternative Representation of AR)

Alternatively, the AR model can be represented as

$$Y_t = \phi Y_{t-1} + \epsilon_t, \ \forall t$$

Proof 1.2

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} = \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i} = \epsilon_t + \sum_{i=0}^{\infty} \psi_{i+1} \epsilon_{t-i-1} = \epsilon_t + \phi \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i-1} = \epsilon_t + \phi Y_{t-1}$$

The natural estimator of ϕ (OLS) is

$$\hat{\phi} = \frac{\sum_{t=2}^{T} Y_{t-1} Y_t}{\sum_{t=2}^{T} Y_{t-1}^2}$$

Definition 1.11 (Model for Finite AR)

$$Y_t = \phi Y_{t-1} + \epsilon_t, \ 2 \leq t \leq T$$

where

- $\epsilon_t \sim \mathrm{WN}(0, \sigma^2)$;
- $|\phi| < 1$;
- $Y_1 = \sum_{i=0}^{\infty} \phi^i \epsilon_{1-i}$

More generally, consider an AR with a drift,

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \ \forall t$$

where $c = \mu(1 - \phi)$.

1.4.3 AR(1)

Definition 1.12 ($AR(1)$)

$\{Y_t : 1 \leq t \leq T\}$ is an **autoregreessive process** of order 1, $Y_t \sim AR(1)$, if

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \ 2 \leq t \leq T$$

where $\epsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

Note *$|\phi| < 1$ is not assumed (yet) and $Y_1 = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{1-i}$ is not assumed.*

We call the AR(1) model is **stable** iff $|\phi| < 1$.

- If $|\phi| < 1$ and $Y_1 = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{1-i}$,

$$Y_t = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i},$$

  where $\mu = \frac{c}{1-\phi}$.

- OLS “works” when $|\phi| < 1$.
- The $AR(1)$ model admits and $MA(\infty)$ solution

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

  iff $|\phi| < 1$.

- The AR(1) model admits a covariance stationary solution iff $|\phi| \neq 1$.

Note *Consider the case that $\phi > 1$, the intuition is*

$$Y_t = \phi Y_{t-1} + \epsilon_t \Leftrightarrow Y_{t-1} = \phi^{-1}(Y_t - \epsilon_t)$$

1.4.4 AR(p)

Definition 1.13 (AR(p))

$\{Y_t : t \in \mathbb{N}\}$ is a p^{th}**-order autoregressive process**, $Y_t \sim AR(p)$, iff

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t,\ t \geq p+1$$

where $\epsilon_t \sim \text{WN}(0, \sigma^2)$.

In vector notation, we can write

$$Y_t = \beta' X_t + \epsilon_t,\ t \geq p+1$$

where $\beta = (c, \phi_1, \phi_2, \cdots, \phi_p)'$ and $X_t = (1, Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})'$.

Claim 1.4

OLS ``works'' when the $AR(p)$ model is stable. Then the *OLS estimator* is given by

$$\hat{\beta} = (\sum_{t=p+1}^{T} X_t' X_t)^{-1} (\sum_{t=p+1}^{T} X_t' Y_t)$$

Lag Operator Notation There is an alternative way to write the $AR(p)$ model.

Definition 1.14 (Lag Operator)

The **lag operator** (L) operates on an element of a time series to produce the previous element. That is, For a time series $\{X_t\}$,

$$\begin{aligned} LX_t &= X_{t-1} \\ &\vdots \\ L^k X_t &= X_{t-k},\ \forall t \in \mathbb{Z} \end{aligned}$$

Then, in this notation, the $AR(p)$ model can be written as

$$\phi(L)Y_t = c + q_t,\ t \geq p+1$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$.

Definition 1.15 (Stability of $AR(p)$)

The $AR(p)$ model is **stable** if

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \Rightarrow |z| > 1$$

(All solutions are greater than 1).

- The $AR(p)$ model admits an $MA(\infty)$ solution

  $$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

  iff it is *stable*. The $MA(\infty)$ solution has

  $$\mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p} = \frac{c}{\phi(1)}$$

  and (computable) ψ_i's satisfy

  $$|\psi_i| \leq M\lambda^i,\ \forall i,$$

  where $M < \infty$ and $|\lambda| < 1$.

1.5 More On $MA(q)$

1.5.1 Lag Operator Notation and Invertible $MA(q)$

$MA(q)$ **model in lag operator notation** :

$$Y_t = \mu + \underbrace{\epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}}_{:=\theta(L)\epsilon_t}$$

$$= \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

Definition 1.16 (Invertibility of $MA(q)$)

The $MA(q)$ model is **invertible** if

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0 \Rightarrow |z| > 1$$

(All solutions are greater than 1).

Note *If the $MA(q)$ model is invertible, then*

$$\epsilon_t = \Pi(L)(Y_t - \mu),$$

where $\Pi(L) = \sum_{i=0}^{\infty} \pi_i L^i$ *with* $\sum_{i=0}^{\infty} |\pi_i| < \infty$.

Technicalities

- If $\sum_{i=0}^{\infty} |\pi_i| < \infty$, then $\sum_{i=0}^{\infty} \pi_i^2 < \infty$.
- If

$$|\pi_i| \leq M\lambda^i,\ \forall i \text{ (some } M < \infty \text{ and } |\lambda| < 1), \tag{*}$$

  then

$$\sum_{i=0}^{\infty} i^r |\pi_i|^s < \infty,\ \forall r \geq 0, s > 0$$

- Invertibility $\Rightarrow$ (*).
- If $X_0, X_1, ...$ are random variables with $\sup_i \mathbb{E}X_i^2 < \infty$, then $\sum_{i=0}^{\infty} \pi_i X_i$ exists (as a limit in mean squared) if $\sum_{i=0}^{\infty} |\pi_i| < \infty$.

1.5.2 $MA(q)$ is the only covariance stationary process with $\gamma(j) = 0, \forall j > q$

Proposition 1.1 ($MA(q) \Leftrightarrow$ covariance stationary and $\gamma(j) = 0, \forall j > q$)

If $\{Y_t\}$ is covariance stationary, then $\gamma(j) = 0, \forall j > q$ iff $Y_t \sim MA(q)$.

Question: Is there a "$q = \infty$" analog?

Example 1.3

Suppose $Y_t = Z \sim \mathcal{N}(0, 1), \forall t$. Then, Cov$(Y_t, Y_{t-1}) = 1, \forall j$.

1. Y_t is covariance stationary.
2. It is not a $MA(\infty)$.
3. Y_t can be predicted without error using $\{Y_s : s \leq t-1\}$.
4. Y_t is "deterministic".

1.5.3 Deterministic covariance stationary process

Definition 1.17 (Deterministic)

A mean zero covariance stationary process $\{v_t\}$ is **deterministic** iff $\exists p$ and $\{\phi_i : 1 \leq i \leq p\}$ such that

$$\mathbb{E}[(v_t - \phi v_{t-1} - \cdots - \phi_p v_{t-p})^2] \leq \epsilon^2, \ \forall t$$

Claim 1.5

If v_t is deterministic, then v_t is not a $MA(\infty)$.

1.6 Spectral Representation

Definition 1.18 (Wold Decomposition)

If $\{Y_t\}$ is a mean zero covariance stationary process, then

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} + v_t, \forall t,$$

where

1. $\epsilon_t \sim \text{WN}(0, \sigma^2)$
2. $\psi_0 = 1$ and $\sum_{i=0}^{\infty} \psi_i^2 < \infty$
3. $\mathbb{E}[\epsilon_t v_s] = 0, \forall t, s$
4. $\{v_t\}$ is deterministic

Question: When is a function $\gamma(\cdot)$ the autocovariance function (ACF) of a covariance stationary process?

Recall that, if $\gamma(\cdot)$ is an ACF, it is given by

$$\gamma(j) = \mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)]$$

and satisfies the following properties by Lemma 1.2.

1. Even: $\gamma(j) = \gamma(-j), \forall j \in \mathbb{N}$.
2. Positive semi-definite (PSD) i.e., for any $n \in \mathbb{N}$ and any $a_1, ..., a_n$,
$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j) = \text{Var}(\sum_{i=1}^{n} a_i Y_i) \geq 0$$

1.6.1 ACF ⇔ Even and PSD

Proposition 1.2 (ACF ⇔ Even and PSD)

A function $\gamma(\cdot)$ is an ACF iff it is even and positive semi-definite.

Theorem 1.2 (Herglotz's Theorem)

A function $\gamma : \mathbb{Z} \to \mathbb{R}$ is *even* and *positive semi-definite* iff

$$\gamma(j) = \int_{-\pi}^{\pi} \exp{(ij\lambda)}\, dF(\lambda)$$

for some $F : [-\pi, \pi] \to \mathbb{R}_+$ that is bounded, non-decreasing, and right-continuous (and has $F(-\pi) = 0$).

Remark

1. $F(\cdot)$ is called the spectral distribution function (of $\gamma(\cdot)$).
2. If $\exists f : [-\pi, \pi] \to \mathbb{R}$ such that
$$F(\lambda) = \int_{-\pi}^{\lambda} f(r)dr, \forall \lambda \in [-\pi, \pi],$$
then $f(\cdot)$ is called a spectral density function (of $\gamma(\cdot)$) and
$$\gamma(j) = \int_{-\pi}^{\pi} \exp{(ij\lambda)}\, f(\lambda)d\lambda = \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda)d\lambda$$

Symmetry Suppose $\gamma(j) = \int_{-\pi}^{\pi} \exp{(ij\lambda)}\, dF(\lambda), j \in \mathbb{Z}$, where

$$\begin{aligned}\int_{-\pi}^{\pi} \exp{(ij\lambda)}\, dF(\lambda) &= \int_{-\pi}^{\pi} (\cos(j\lambda) + i\sin(j\lambda))\, dF(\lambda) \\ &= \int_{-\pi}^{\pi} \cos(j\lambda)dF(\lambda) + i\int_{-\pi}^{\pi} \sin(j\lambda)dF(\lambda)\end{aligned}$$

Given $\gamma(j) \in \mathbb{R}, \forall j$, we must have $\int_{-\pi}^{\pi} \sin(j\lambda)dF(\lambda) = 0$. Therefore,

$$\gamma(j) = \int_{-\pi}^{\pi} \cos(j\lambda)dF(\lambda),$$

which is even by the property of $\cos(\cdot)$.

Then, $\frac{F(\cdot)}{F(\pi)}$ is the CDF of a symmetric distribution on $[-\pi, \pi]$.

Example 1.4

Suppose $\epsilon_t \sim \text{WN}(0, \sigma^2)$. Then,

$$\begin{aligned}\gamma(j) &= \begin{cases} 1, & \text{if } j = 0 \\ 0, & \text{otherwise} \end{cases} \\ &= \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda) d\lambda \\ \Rightarrow f(\lambda) &= \frac{1}{2\pi}\end{aligned}$$

Example 1.5

Suppose $Y_t = Z \sim \mathcal{N}(0, 1)$ for all t. Then,

$$\begin{aligned}\gamma(j) &= 1 \\ &= \int_{-\pi}^{\pi} \cos(j\lambda) dF(\lambda) \\ \Rightarrow F(\lambda) &= \begin{cases} 1, & \lambda \geq 0 \\ 0, & \lambda < 0 \end{cases}\end{aligned}$$

Question: When does an ACF $\gamma(\cdot)$ admits a spectral density function?

Partial Answer: An even function $\gamma : \mathbb{Z} \to \mathbb{R}$ with "$\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$" is psd iff

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \cos(j\lambda)\, \gamma(j) \geq 0, \ \forall \lambda \in [-\pi, \pi], \tag{1.1}$$

in which case $f(\cdot)$ is a spectral density function of $\gamma(\cdot)$.

Remark A covariance stationary process with an ACF $\gamma(\cdot)$ has **short memory** if "$\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$".

Proposition 1.3 (Implication of Short Memory)

Given the covariance stationary process has **short memory** ($\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$), we have

1. $f(\cdot)$ exists (given as (1.1)) and is bounded.
2. $\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$.
3. $f(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j)$.

$MA(\infty)$ **Case:** Suppose

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \ \forall t,$$

where

- $\epsilon_t \sim \text{WN}(0, \sigma^2)$
- $\sum_{i=0}^{\infty} |\psi_i| < \infty$

Then,

- $\gamma(\cdot)$ has short memory
- $\gamma(\cdot)$ has spectral density function given by

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \cos{(j\lambda)}\, \gamma(j) \\ &= \frac{\sigma^2}{2\pi} |\psi(e^{i\lambda})|^2 \end{aligned}$$

  where $\gamma(j) = (\sum_{i=0}^{\infty} \psi_i \psi_{i+j})\sigma^2$ and $\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i$.

- $f(0) = \frac{\sigma^2}{2\pi} \psi(1)^2$

Chapter 2 Estimation and Inference

2.1 OLS Estimation in $AR(1)$ Model

Suppose

$$Y_t = \phi Y_{t-1} + \epsilon_t,\ \forall t \geq 2,$$

where $\epsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

The **OLS Estimator of** ϕ is

$$\hat{\phi}_{OLS} = \frac{\sum_{t=2}^{T} Y_{t-1} Y_t}{\sum_{t=2}^{T} Y_{t-1}^2}$$

Claim 2.1 (OLS Estimator is MLE)

If $\epsilon_t \sim$ i.i.d.$\mathcal{N}(0, \sigma^2)$ and if $(\epsilon_2, \epsilon_3, ...) \perp Y_1$, then $\hat{\phi}_{OLS}$ is the (conditional) MLE of ϕ.

The (conditional) MLE of (ϕ, σ^2) is

$$(\hat{\phi}_{ML}, \hat{\sigma}^2_{ML}) = \underset{(\phi,\sigma^2)}{\operatorname{argmax}}\, f_{2:T}\left(Y_2, ...Y_T \mid Y_1; \phi, \sigma^2\right),$$

where $f_{2:T}(\cdot \mid Y_1; \phi, \sigma^2)$ is the (conditional) pdf of $(Y_2, ..., Y_T)$ given Y_1.

Definition 2.1 (Prediction-error Decomposition)

The objective function (conditional likelihood function) can be written as

$$f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \phi, \sigma^2\right) = \prod_{t=2}^{T} f_t\left(Y_t \mid Y_1, ..., Y_{t-1}; \phi, \sigma^2\right),$$

where $f_t\left(Y_t \mid Y_1, ..., Y_{t-1}; \phi, \sigma^2\right)$ is the conditional pdf of Y_t given $Y_1, ..., Y_{t-1}$.

By the definition that $Y_t = \phi Y_{t-1} + \epsilon_t,\ \forall t \geq 2$ and $\epsilon_t \mid Y_1, ..., Y_{t-1} \sim \mathcal{N}(0, \sigma^2)$, we have

$$\begin{aligned}
&Y_t \mid Y_1, ..., Y_{t-1} \sim \mathcal{N}(\phi Y_{t-1}, \sigma^2)\\
&\Rightarrow f_t\left(Y_t \mid Y_1, ..., Y_{t-1}; \phi, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(Y_t - \phi Y_{t-1}\right)^2\right)\\
&\Rightarrow f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \phi, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{T-1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\left(Y_t - \phi Y_{t-1}\right)^2\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\phi}_{ML} &= \underset{\phi}{\operatorname{argmin}}\, f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \phi, \sigma^2\right) = \frac{\sum_{t=2}^{T} Y_{t-1} Y_t}{\sum_{t=2}^{T} Y_{t-1}^2} = \hat{\phi}_{OLS}\\
\hat{\sigma}^2_{ML} &= \underset{\sigma^2}{\operatorname{argmin}}\, f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \phi, \sigma^2\right) = \frac{1}{T-1}\sum_{t=2}^{T}(Y_t - \hat{\phi}_{ML} Y_{t-1})^2
\end{aligned}$$

2.2 Properties of OLS Estimators (in time series)

2.2.1 OLS Review

The OLS model can be written as

$$y_i = \beta' x_i + \epsilon_i,\ i = 1, ..., n$$

Iff $\sum_{i=1}^n x_i x_i'$ is positive definite ($\sum_{i=1}^n x_i x_i' \succ 0$), the OLS estimator (of β) is given by

$$\begin{aligned}\hat{\beta}_{OLS} &= \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta' x_i)^2 \right\} \\ &= \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \left(\sum_{i=1}^n x_i y_i \right) = \beta + \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \left(\sum_{i=1}^n x_i \epsilon_i \right)\end{aligned}$$

Lemma 2.1 (Unbiasedness)

Suppose that

(i). $\Pr[\sum_{i=1}^n x_i x_i' \succ 0] = 1$ and $\mathbb{E}[\hat{\beta}_{OLS}]$ exists.

(ii). Strict exogeneity: $\mathbb{E}[\epsilon_i \mid x_1, ..., x_n] = 0, \forall i$.

Then, $\mathbb{E}[\hat{\beta}_{OLS}] = \beta$.

Remark

1. If $(x_i, \epsilon_i) \sim i.i.d.$, then the "strictly exogeneity" holds iff $\mathbb{E}[\epsilon_i \mid x_i] = 0$.
2. The first assumption (i.e., $\Pr[\sum_{i=1}^n x_i x_i' \succ 0] = 1$ and $\mathbb{E}[\hat{\beta}_{OLS}]$ exists) is necessary and cannot be reduced in i.i.d. case, we need additional assumptions.

Lemma 2.2 (Consistency)

Suppose that

(i). $\frac{1}{n} \sum_{i=1}^n x_i x_i' \xrightarrow{P} Q$ for some $Q \succ 0$.

(ii). $\frac{1}{n} \sum_{i=1}^n x_i \epsilon_i \xrightarrow{P} 0$.

Then, $\hat{\beta}_{OLS} \xrightarrow{P} \beta$.

Proof 2.1

With probability approaching one (as $n \to \infty$),

$$\hat{\beta} = \beta + \left(\underbrace{\sum_{i=1}^n x_i x_i'}_{\xrightarrow{P} Q} \right)^{-1} \underbrace{\left(\sum_{i=1}^n x_i \epsilon_i \right)}_{\xrightarrow{P} 0} \xrightarrow{P} \beta + Q^{-1} \cdot 0 = \beta$$

by the continuity theorem (for $\xrightarrow{P}$).

Remark If $\begin{bmatrix} x_i \\ \epsilon_i \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & 0 \\ 0 & \sigma^2 \end{bmatrix} \right)$, then

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{P} \mu_x \mu_x' + \Sigma_{xx} = \mathbb{E}[x_i x_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{P} 0 = \mathbb{E}[x_i \epsilon_i]$$

by LLN.

Lemma 2.3 (Asymptotic Normality)

Suppose that

(i). $\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{P} Q$ for some $Q \succ 0$.

(ii). $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, V)$ for some $V \succ 0$.

Then, $\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) \xrightarrow{d} N(0, \Omega)$, where $\Omega := Q^{-1} V Q^{-1}$

Proof 2.2

With probability approaching one (as $n \to \infty$),

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\underbrace{\sum_{i=1}^{n} x_i x_i'}_{\xrightarrow{P} Q} \right)^{-1} \left(\underbrace{\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right)}_{\xrightarrow{d} \mathcal{N}(0,V)} \right) \xrightarrow{d} Q^{-1}\mathcal{N}(0, V) = \mathcal{N}(0, Q^{-1} V Q^{-1})$$

by the continuous mapping theorem (CMT).

Remark If $\begin{bmatrix} x_i \\ x_i \epsilon_i \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & C' \\ C & V \end{bmatrix} \right)$, then

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{P} \mu_x \mu_x' + \Sigma_{xx} = \mathbb{E}[x_i x_i']$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{P} 0 = \mathbb{E}[x_i \epsilon_i]$$

by LLN. Moreover,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, V)$$

by CLT.

Proposition 2.1 (Variance Estimation)

Suppose that

(i). $\hat{Q} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{P} Q \succ 0.$

(ii). $\hat{V} \xrightarrow{P} V.$

Then, $\hat{\Omega} := \hat{Q}^{-1}\hat{V}\hat{Q}^{-1} \xrightarrow{P} Q^{-1}VQ^{-1} := \Omega$ (by the continuity theorem for $\xrightarrow{P}$).

Remark To achieve these properties we need, except for $\begin{bmatrix} x_i \\ x_i\epsilon_i \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & C' \\ C & V \end{bmatrix} \right)$, we need more conditions:

1. If also $\mathbb{E}[(x_i' x_i)^r] < \infty$ for some $r > 1$, then
$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{\epsilon}_i^2 \xrightarrow{P} \mathbb{E}[x_i x_i' \epsilon_i^2] = V, \text{ where } \hat{\epsilon}_i = y_i - \hat{\beta}_{OLS}' x_i$$
2. If also $\mathbb{E}[\epsilon_i^2 \mid x_i] = \sigma^2$ (aka "homoskedasticity"), then
$$V = \mathbb{E}[x_i x_i' \epsilon_i^2] = \ldots \underbrace{=}_{LIE} \sigma^2 \mathbb{E}[x_i x_i'] = \sigma^2 Q$$
and
$$\hat{V} = \hat{\sigma}^2 \hat{Q}, \text{ where } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{OLS}' x_i\right)^2$$

2.2.2 OLS for $MA(\infty)$: $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$

Consider the $MA(\infty)$ model:
$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},\ t \geq 1$$
where

1. $\epsilon_t \sim i.i.d.(0, \sigma^2)$,
2. $\sum_{i=0}^{\infty} i|\psi_i| < \infty$.

Mean Estimation Consider the estimator (for μ):
$$\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

Note

1. $\bar{Y} = \text{argmin}_m \sum_{t=1}^{T} (Y_t - m)^2$.

2. $\epsilon_t \sim i.i.d.(0, \sigma^2) \Rightarrow \epsilon_t \sim \text{WN}(0, \sigma^2)$ *(i.e., a stronger assumption than white noise).*

3. $\sum_{i=0}^{\infty} i|\psi_i| < \infty \Rightarrow \sum_{i=0}^{\infty} |\psi_i| < \infty \Rightarrow \sum_{i=0}^{\infty} \psi_i^2 < \infty$ *(also a stronger assumption)*

The properties of $\bar{Y}$ can be checked in the following:

1. **Unbiasedness**: Recall that $\mathbb{E}(Y_t) = \mu, \forall t$ because $\epsilon_t \sim \text{WN}(0, \sigma^2)$ and $\sum_{i=0}^{\infty} \psi_i^2 < \infty$. Then,

$$\mathbb{E}[\bar{Y}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} Y_i] = \mu$$

2. **Consistency**:

$$\bar{Y} \xrightarrow{P} \mu$$

which can be proven by $P(|\bar{Y} - \mu| > \eta) \xrightarrow{T\to\infty} 0$ for all $\eta > 0$. This can be given by Chebyshev's inequality: $P(|\bar{Y} - \mu| > \eta) \leq \frac{\text{Var}(\bar{Y})}{\eta^2}$ for all $\eta > 0$.

Claim 2.2 (Bounded Variance)

$$\begin{aligned}
\text{Var}(\bar{Y}) &= \text{Cov}\left(\frac{1}{T}\sum_t Y_t, \frac{1}{T}\sum_s Y_s\right) \\
&= \frac{1}{T^2}\sum_t \sum_s \text{Cov}\left(Y_t, Y_s\right) \\
&= \frac{1}{T^2}\sum_t \sum_s \gamma(t-s) \\
&= \frac{1}{T^2}\sum_{j=1-T}^{T-1} (T - |j|)\gamma(j) \\
&= \frac{1}{T}\sum_{j=1-T}^{T-1} (1 - \frac{|j|}{T})\gamma(j) \\
&\leq \frac{1}{T}\sum_{j=1-T}^{T-1} |\gamma(j)| \\
&\leq \frac{1}{T}\sum_{j=-\infty}^{\infty} |\gamma(j)|
\end{aligned}$$

where $\gamma(j) := \text{Cov}(Y_t, Y_{t-j})$ is the autocovariance function.

Recall that if $\epsilon_t \sim \text{WN}(0, \sigma^2)$ and if $\sum_{i=0}^{\infty} |\psi_i| < \infty$, then $\sum_{i=0}^{\infty} |\gamma(i)| < \infty$ (aka "short memory"). Therefore, we have $\bar{Y} \xrightarrow{P} \mu$.

3. **Asymptotic Consistency**:

$$\sqrt{T}\left(\bar{Y} - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \omega^2\right)$$

where $\omega^2 \neq \text{Var}\left(Y_t\right)$ (in general).

Idea of proof:

$$\sqrt{T}\left(\bar{Y} - \mu\right) = \underbrace{\psi(1)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \epsilon_t}_{\xrightarrow{d}\psi(1)\mathcal{N}(0,\sigma^2)=\mathcal{N}(0,\omega^2)} + \underbrace{o_p(1)}_{\xrightarrow{p}0\text{, by definition}}$$

where $\psi(1) = \sum_{i=0}^{\infty} \psi_i$ and $\omega^2 = \psi(1)^2\sigma^2$. This is given by BN decomposition.

Theorem 2.1 (BN Decomposition)

If $\psi(L)=\sum_{i=0}^{\infty}\psi_i L^i$ is a lag polynomial with $\sum_{i=0}^{\infty} i|\psi_i|<\infty$, then

$$\psi(L)=\psi(1)+\tilde{\psi}(L)(1-L) \tag{2.1}$$

where

- $\tilde{\psi}(L)=\sum_{i=0}^{\infty}\tilde{\psi}_i L^i$, $\tilde{\psi}_i=-\sum_{j=i+1}^{\infty}\psi_j$.
- $\sum_{i=0}^{\infty}|\tilde{\psi}_i|<\infty$.

Proof 2.3

By the definition of $\tilde{\psi}(L)=\sum_{i=0}^{\infty}\tilde{\psi}_i L^i$, the RHS of (2.1) can be written as

$$\psi(1)+\tilde{\psi}(L)(1-L)=\psi(1)+\sum_{i=0}^{\infty}\tilde{\psi}_i L^i-\sum_{i=1}^{\infty}\tilde{\psi}_{i-1}L^i$$

Let's check the coefficients of L^i:

(a). $i=0$: $\psi(1)+\tilde{\psi}_0=\psi_0$

(b). $i\geq 1$: $\tilde{\psi}_i-\tilde{\psi}_{i-1}=\psi_i$

The (2.1) is proved. Moreover,

$$\sum_{i=0}^{\infty}|\tilde{\psi}_i|\leq\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}|\psi_j|=\sum_{i=0}^{\infty}i|\psi_i|<\infty$$

Given the BN decomposition, we have

$$\begin{aligned}
\psi(L)&=\psi(1)+\tilde{\psi}(L)(1-L)\\
\psi(L)\epsilon_t&=\psi(1)\epsilon_t+\tilde{\psi}(L)(\epsilon_t-\epsilon_{t-1})\\
\sum_{t=1}^{T}\psi(L)\epsilon_t&=\psi(1)\sum_{t=1}^{T}\epsilon_t+\tilde{\psi}(L)(\epsilon_T-\epsilon_0)
\end{aligned}$$

Thus,

$$\sqrt{T}\left(\bar{Y}-\mu\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi(L)\epsilon_t=\psi(1)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t+\frac{1}{\sqrt{T}}\tilde{\psi}(L)(\epsilon_T-\epsilon_0)$$

where $\frac{1}{\sqrt{T}}\tilde{\psi}(L)(\epsilon_T-\epsilon_0)\xrightarrow{p}0$ is proved by

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{T}}\tilde{\psi}(L)(\epsilon_T-\epsilon_0)\right]&=0\\
\text{Var}\left[\frac{1}{\sqrt{T}}\tilde{\psi}(L)(\epsilon_T-\epsilon_0)\right]&=\frac{1}{T}\text{Var}\left[\tilde{\psi}(L)\epsilon_T-\tilde{\psi}(L)\epsilon_0\right]\\
&\leq\frac{2}{T}\left[\text{Var}\left(\tilde{\psi}(L)\epsilon_T\right)+\text{Var}\left(\tilde{\psi}(L)\epsilon_0\right)\right]\\
&=\frac{4}{T}\text{Var}\left(\tilde{\psi}(L)\epsilon_T\right)=\frac{4\sigma^2}{T}\underbrace{\sum_{i=0}^{\infty}\tilde{\psi}_i^2}_{<\infty}\to 0
\end{aligned}$$

Remark

(a). If $\sum_{i=0}^{\infty} i|\psi_i| < \infty$, then $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and $\sum_{i=0}^{\infty} |\tilde{\psi}_i| < \infty$. Note: we only need $\sum_{i=0}^{\infty} \tilde{\psi}_i^2 < \infty$, so we can only require $\sum_{i=0}^{\infty} \sqrt{i}|\psi_i| < \infty$.

(b). If $\epsilon_t \sim i.i.d.\left(0, \sigma^2\right)$, then $\epsilon_t \sim \mathrm{WN}\left(0, \sigma^2\right)$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$. (These two properties may hold even if $\epsilon_t \nsim i.i.d.\left(0, \sigma^2\right)$, i.e., there is a weaker condition can be used.)

(c). $\omega^2 = \psi(1)^2 \sigma^2 \neq \left(\sum_{i=0}^{\infty} \psi_i^2\right) \sigma^2 = \mathrm{Var}\left(Y_t\right)$ (in general.)

(d). ω^2 is called the **"long-run variance"** of Y_t:

$$\omega^2 = \lim_{T \to \infty} T \mathrm{Var}\left(\bar{Y}\right) = \lim_{T \to \infty} \frac{1}{T} \sum_{j=1-T}^{T-1} (1 - \frac{|j|}{T}) \gamma(j) = \sum_{j=0}^{\infty} \gamma(j)$$

Variance Estimation The OLS (variance) estimator is

$$S^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left(Y_t - \bar{Y}\right)^2$$

Claim 2.3

$S^2 \xrightarrow{p} \mathrm{Var}(Y_t)$.

Recall that

$$\omega^2 = \sigma^2 \psi(1)^2 = \sum_{j=-\infty}^{\infty} \gamma(j) = 2\pi f(0),$$

where $f(\cdot)$ is the spectral density function of $\gamma(\cdot)$:

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \cos\left(j\lambda\right) \gamma(j) \\ &= \frac{\sigma^2}{2\pi} |\psi(e^{i\lambda})|^2 \end{aligned}$$

where $\gamma(j) = (\sum_{i=0}^{\infty} \psi_i \psi_{i+j}) \sigma^2$ and $\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i$.

The variance estimator can be given by

$$\hat{\omega}^2 = 2\pi \hat{f}(0),$$

where $\hat{f}$ is an estimator of f.

Example 2.1 (Newey-West, 1987)

$\hat{\omega}^2 = \hat{\gamma}(0) + 2 \sum_{j=1}^{b} \left(1 - \frac{j}{b}\right) \hat{\gamma}(j)$, where $\hat{\gamma}(j) = \frac{1}{T} \sum_{t=1}^{T} (Y_t = \bar{Y}) \left(Y_{t-j} - \bar{Y}\right)$ and b is a "turning" parameter.

Remark If $\epsilon_t \sim i.i.d.(0, \sigma^2)$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$, then

$$\hat{\omega}^2 \xrightarrow{p} \omega^2$$

provided $b \to \infty$ and $\frac{b}{\sqrt{T}} \to 0$ as $T \to \infty$.

2.2.3 OLS of AR(1) Model

Consider an AR(1) model

$$Y_t = \phi Y_{t-1} + \epsilon_t, \ \forall t \geq 2,$$

where

1. $|\phi| < 1$
2. $\epsilon_t \sim i.i.d.(0, \sigma^2)$

The **OLS Estimator of ϕ** is

$$\hat{\phi} = \frac{\sum_{t=2}^T Y_{t-1} Y_t}{\sum_{t=2}^T Y_{t-1}^2} = \phi + \frac{\sum_{t=2}^T Y_{t-1}\epsilon_t}{\sum_{t=2}^T Y_{t-1}^2}$$

Unbiasedness

Usual template ("strict exogeneity"): $\mathbb{E}[\epsilon_t \mid Y_1, ..., Y_{T-1}] = 0, \ t \geq 2$. However, it doesn't hold here: $\epsilon_t = Y_t - \phi Y_{t-1}, t \geq 2 \Rightarrow \mathbb{E}[\epsilon_t \mid Y_1, ..., Y_{T-1}] = \epsilon_t \neq 0 \ (2 \leq t \leq T-1)$.

Claim 2.4

The OLS estimator of ϕ, $\hat{\phi}$, is biased (in general.)

Consistency

Usual template: If

(i). $\frac{1}{T-1}\sum_{t=2}^T Y_{t-1}^2 \xrightarrow{p} Q > 0$,

(ii). $\frac{1}{T-1}\sum_{t=2}^T Y_{t-1}\epsilon_t \xrightarrow{p} 0$,

then $\hat{\phi} \xrightarrow{p} \phi$.

Claim 2.5

These two conditions (i) and (ii) hold.

Let $\tilde{Y}_t = \sum_{i=0}^\infty \phi^i \epsilon_{t-i}$, which equals to Y_t iff $Y_1 = \sum_{i=0}^\infty \phi^i \epsilon_{1-i}$. By assuming $Y_1 = \sum_{i=0}^\infty \phi^i \epsilon_{1-i}$, we have

1. $\sum_{t=2}^T Y_{t-1}^2 = \sum_{t=2}^T \tilde{Y}_{t-1}^2 + O_P(1)$.
2. $\sum_{t=2}^T Y_{t-1}\epsilon_t = \sum_{t=2}^T \tilde{Y}_{t-1}\epsilon_t + O_P(1)$.

(Proof by heuristics: $Y_{t-1} = \tilde{Y}_{t-1} + \phi^{t-2}(Y_1 - \tilde{Y}_1) \approx \tilde{Y}_{t-1}$ when t is large and $|\phi| < 1$.)

Recall that if $\{X_t\}$ is non-random and bonded and if $r_t \to \infty$, $\frac{X_t}{r_t} \to 0$.

1. If $X_t = O(1)$ and if $r_t \to \infty$, then $\frac{X_t}{r_t} = o(1)$ ("$\to 0$").
2. If $\{X_t\}$ is random with $X_t = O_P(1)$ and if $r_t \to \infty$, then $\frac{X_t}{r_t} = o_P(1)$ ("$\xrightarrow{P} 0$").

Definition 2.2 (Stochastically Bounded)

A random sequence $\{X_t\}$ is **stochastically bounded**, $X_t = O_P(1)$, iff $\lim_{M\to\infty}\sup_{T\geq 1} P(|X_T| > M) = 0$.

Then, we can prove the consistency:

Proof 2.4

$$\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}^2 = \frac{1}{T}\sum_{t=2}^{T}\tilde{Y}_{t-1}^2 + \underbrace{\frac{O_P(1)}{T}}_{=o_P(1)}$$

$$\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}\epsilon_t = \frac{1}{T}\sum_{t=2}^{T}\tilde{Y}_{t-1}\epsilon_t + \underbrace{\frac{O_P(1)}{T}}_{=o_P(1)}$$

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T} Y_{t-1}\epsilon_t = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\tilde{Y}_{t-1}\epsilon_t + \underbrace{\frac{O_P(1)}{\sqrt{T}}}_{=o_P(1)}$$

If $\mathbb{E}[\epsilon_t^4] < \infty$, we have

$$\text{Var}(\frac{1}{T}\sum_{t=2}^{\infty}\tilde{Y}_{t-1}^2) \to 0 \ \& \ \text{Var}(\frac{1}{T}\sum_{t=2}^{\infty}\tilde{Y}_{t-1}\epsilon_t) \to 0$$

so,

1. $\frac{1}{T}\sum_{t=2}^{T}\tilde{Y}_{t-1}^2 \xrightarrow{p} \mathbb{E}[\tilde{Y}_{t-1}^2] = \frac{\sigma^2}{1-\phi^2} > 0$
2. $\frac{1}{T}\sum_{t=2}^{T}\tilde{Y}_{t-1}\epsilon_t \xrightarrow{p} \mathbb{E}[\tilde{Y}_{t-1}\epsilon_t] = 0$

Note *If $\mathbb{E}[|\epsilon_t|^r] < \infty$ for some $r > 2$, then the consistency can hold by Mixingale LLN.*

Theorem 2.2 (Mixingale LLN)

If $\{X_t\}$ is a uniformly integrable L^1-mixingale with the upper bound of limitation

$$\underbrace{\overline{\lim_{T\to\infty}}}_{\text{“}\limsup T\to\infty\text{''}} \frac{1}{T}\sum_{t=1}^{T} C_t < \infty,$$

then

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} 0$$

L^1-mixingale

Definition 2.3 (L^1-mixingale)

A sequence $\{X_t\}$ is an L^1-**mixingale** iff $\exists \{Z_t\}, \{c_t\}, \{\xi_m\}$ s.t.

$$\mathbb{E}[X_t \mid Z_t, Z_{t-1}, ...] = X_t, \forall t \tag{2.2}$$

$$\mathbb{E}\left(|\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, ...]|\right) \leq c_t \xi_m, \forall t, m \geq 1 \tag{2.3}$$

where $\lim_{m\to\infty} \xi_m = 0$.

Lemma 2.4 (Some Properties of L^1-mixingale)

If $X_t \sim i.i.d$ with $\mathbb{E}[X_t] = 0$, then

(i). $\{X_t\}$ is an L^1-mixingale (with $Z_t = X_t$, $c_t = 0$, $\xi_m = 0$).

(ii). $\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} 0$.

If $X_t = Z \sim \mathcal{N}(0,1), \forall t$, then

(i). $\{X_t\}$ is not an L^1-mixingale,

(ii). $\frac{1}{T}\sum_{t=1}^{T} X_t = Z \overset{p}{\nrightarrow} 0$.

If $\{X_t\}$ is an L^1-mixingale,

$$\mathbb{E}[X_t] = \mathbb{E}\left(\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, ...]\right) = 0$$

Remark

1. If $Z_t = X_t$, then 2.2 holds.
2. If 2.2 and 2.3 hold, then they hold with $Z_t = X_t$.
3. If $X_t = g\left(\epsilon_t, \epsilon_{t-1}, ...\right)$, then 2.2 holds with $Z_t = \epsilon_t$.

In AR(1) examples:

1. $\{\underbrace{\tilde{Y}_{t-1}^2 - \mathbb{E}[\tilde{Y}_{t-1}^2]}_{X_t}\}$ is an L^1-mixingale (with $Z_t = \epsilon_{t-1}, c_t \equiv 1$).
2. $\{\underbrace{\tilde{Y}_{t-1}\epsilon_t}_{X_t}\}$ is an L^1-mixingale (with $Z_t = \epsilon_t, \xi_1 = 0$).

Example 2.2 (Important Case)

If $\{X_t\}$ is an L^1-mixingale with $\xi_1 = 0$, then

$$\begin{aligned}\mathbb{E}\left[X_t \mid Z_{t-1}, Z_{t-2}, ...\right] = 0 \overset{LIE}{\Rightarrow} &\mathbb{E}\left[X_t \mid Z_{t-m}, Z_{t-m-1}, ...\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[X_t \mid Z_{t-1}, Z_{t-2}, ...\right] \mid Z_{t-m}, Z_{t-m-1}, ...\right] = 0, \forall m \\ &\Rightarrow \xi_m = 0, \forall m \geq 1 \\ &\Rightarrow \text{we can have } c_t \equiv 1\end{aligned}$$

$\mathbb{E}\left[X_t \mid Z_{t-1}, Z_{t-2}, ...\right] = 0 \overset{LIE}{\Rightarrow} \mathbb{E}\left[X_t \mid X_{t-1}, X_{t-2}, ...\right] = 0.$

Terminology: $\{X_t\}$ is a **martingale difference sequence (MDS)** if $\mathbb{E}\left[X_t \mid X_{t-1}, X_{t-2}, ...\right] = 0.$

Definition 2.4 (Martingale Difference Sequence (MDS))

$\{X_t\}$ is an MDS iff it is an L^1-mixingale with $\xi_m = 0$.

$\{\tilde{Y}_{t-1}\epsilon_t\}$ is an MDS because

$$\mathbb{E}[\tilde{Y}_{t-1}\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, ...] = \tilde{Y}_{t-1}\mathbb{E}[\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, ...] = 0$$

Thus, $\mathbb{E}[\tilde{Y}_{t-1}\epsilon_t \mid \tilde{Y}_{t-2}\epsilon_{t-1}, \tilde{Y}_{t-3}\epsilon_{t-2}, ...] = 0$

Uniformly Integrality

Definition 2.5 (Uniformly Integrable)

A sequence $\{X_t\}$ is **uniformly integrable** iff

$$\lim_{m\to\infty} \sup_t \mathbb{E}\left[|X_t| \mathbf{1}\left(|X_t| > M\right)\right] = 0$$

Remark

1. If $X_T \xrightarrow{d}_{T\to\infty} \mathcal{N}(0,1)$ and if $\{X_T\}$ is uniformly integrable, then $\mathbb{E}[X_T] \to_{T\to\infty} 0$.
2. Integrality: $\mathbb{E}[|X_T|] < \infty$ iff $\lim_{m\to\infty} \mathbb{E}\left[|X_T| \mathbf{1}\left(|X_T| > M\right)\right] = 0$.
3. If $\{X_t\}$ is uniformly integrable, then $\sup_t \mathbb{E}[|X_t|] < \infty$.
4. If $\sup_t \mathbb{E}[|X_t|^r] < \infty$ for some $r > 1$, then $\{X_t\}$ is uniformly integrable.

   AR(1) example: If $\mathbb{E}[|\epsilon_t|^r] < \infty$ for some $r > 2$, then $\sup_t \mathbb{E}[|\tilde{Y}_{t-1}^2 - \mathbb{E}[\tilde{Y}_{t-1}^2]|^{\frac{r}{2}}] < \infty$. So, $\{\tilde{Y}_{t-1}^2 - \mathbb{E}[\tilde{Y}_{t-1}^2]\}$ is uniformly integrable.
5. If $\{X_t\}$ is strictly (marginally) stationary, then $\{X_t\}$ is uniformly integrable iff $\mathbb{E}[|X_T|] < \infty, \forall T$.

Example 2.3 (AR(1) Example)

If $\mathbb{E}[\epsilon_t^2] < \infty$, then $\{\tilde{Y}_{t-1}^2 - \mathbb{E}[\tilde{Y}_{t-1}^2]\}$ and $\{\tilde{Y}_{t-1}\epsilon_t\}$ are uniformly integrable L^1-mixingales with $c_t \equiv 1$.

Then,

1. $\frac{1}{T}\sum_{t=2}^{T} \tilde{Y}_{t-1}^2 \xrightarrow{P} \mathbb{E}[\tilde{Y}_{t-1}^2] = \frac{\sigma^2}{1-\phi^2}$.
2. $\frac{1}{T}\sum_{t=2}^{T} \tilde{Y}_{t-1}\epsilon_t \xrightarrow{P} \mathbb{E}[\tilde{Y}_{t-1}\epsilon_t] = 0$.

Strict Stationary: If $\{(X_t, Z_t)\}$ is strictly stationary, then

- $\mathbb{E}\left[|X_t|\mathbf{1}\left(|X_t| > M\right)\right]$ does not depend on t. Then, $\{X_t\}$ is uniformly integrable iff $\mathbb{E}\left[|X_t|\right] < \infty, \forall t$.
- $\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, ...]$ does not depend on t. Then, if $\{X_t\}$ is uniformly integrable, then $\{X_t \mid Z_{t-m}, Z_{t-m-1}, ...\}$ is an L^1-mixingale, then $c_t \equiv 1$ "works".

Corollary 2.1 (to Mixingale LLN)

If $\{X_t\}$ is a strictly stationary L^1-mixingale, then

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{P} \mathbb{E}[X_t] = 0$$

Asymptotic Normality:

Suppose

(i). $\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}^2 \xrightarrow{p} Q$ (some $Q \succ 0$);

(ii). $\frac{1}{\sqrt{T}}\sum_{t=2}^{T} Y_{t-1}\epsilon_t \xrightarrow{d} \mathcal{N}(0, V)$ (some $V \succ 0$).

Then, $\sqrt{T}\left(\hat{\phi} - \phi\right) \xrightarrow{d} \mathcal{N}\left(0, Q^{-1}VQ^{-1}\right)$.

Claim 2.6

(i) and (ii) hold with $Q = \frac{\sigma^2}{1-\phi^2}$ and $V = \sigma^2 Q$. Thus,

$$\sqrt{T}\left(\hat{\phi} - \phi\right) \xrightarrow{d} \mathcal{N}\left(0, 1-\phi^2\right).$$

Remark Recall that

1. We can assume $Y_{t-1} = \tilde{Y}_{t-1} = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-1-i}$.
2. (Definition 2.3) $\{X_t\}$ is an L^1-mixingale iff $\exists\{Z_t\}, \{c_t\}, \{\xi_m\}$ s.t.

$$\mathbb{E}[X_t \mid Z_t, Z_{t-1}, ...] = X_t, \forall t$$

$$\mathbb{E}\left(|\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, ...]|\right) \leq c_t \xi_m, \forall t, m \geq 1$$

where $\lim_{m\to\infty} \xi_m = 0$.

3. $\{X_t\}$ is an MDS iff it is an L^1-mixingale with $\xi_m = 0$.

Theorem 2.3 (Martingale CLT, (Brown, 1971))

If $\{X_t\}$ is an MDS with $\{(X_t, Z_t)\}$ strictly stationary and if

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[X_t^2 \mid Z_{t-1}, Z_{t-2}, \ldots\right] \xrightarrow{P} \mathbb{E}[X_1^2]\ (<\infty)$$

(conditional second moment condition). Then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[X_1^2]\right)$$

For the AR(1) example, we have

- $X_t = \tilde{Y}_{t-1}\epsilon_t, Z_t = \epsilon_t$.
- MDS property:

$$\mathbb{E}\left[\tilde{Y}_{t-1}\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right] = \tilde{Y}_{t-1}\mathbb{E}\left[\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right] = 0$$

- (Conditional) second moment condition:

$$\mathbb{E}\left[\tilde{Y}_{t-1}^2\epsilon_t^2 \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right] = \tilde{Y}_{t-1}^2\mathbb{E}\left[\epsilon_t^2 \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right] = \tilde{Y}_{t-1}^2\sigma^2 \xrightarrow{P} \mathbb{E}\left[\tilde{Y}_{t-1}^2\epsilon_t^2\right]$$

Proof 2.5

The Convergence of $\tilde{Y}_{t-1}^2\sigma^2$:

$$\frac{1}{T}\sum_{t=2}^{T}\left[\sigma^2\tilde{Y}_{t-1}^2\right] = \sigma^2\frac{1}{T}\sum_{t=2}^{T}\tilde{Y}_{t-1}^2 \xrightarrow{P} \frac{\sigma^4}{1-\phi^2}$$

and the expectation of $\tilde{Y}_{t-1}^2\epsilon_t^2$

$$\mathbb{E}\left[\tilde{Y}_{t-1}^2\epsilon_t^2\right] = \mathbb{E}[\tilde{Y}_{t-1}^2]\mathbb{E}[\epsilon_t^2] = \frac{\sigma^4}{1-\phi^2}$$

Therefore, by the Martingale CLT, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{Y}_{t-1}\epsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{1-\phi^2}\right)$$

Then, by the template of asymptotic normality, we have

$$\sqrt{T}\left(\hat{\phi}-\phi\right) \xrightarrow{d} \mathcal{N}\left(0, 1-\phi^2\right).$$

Variance Estimation

To be estimated:

$$1-\phi^2 = \sigma^2 Q^{-1};\ \sigma^2 = \mathbb{E}[\epsilon_t^2], Q = \mathbb{E}[\tilde{Y}_{t-1}^2]$$

Consistent estimators:

(i). $1-\hat{\phi}^2$

(ii). $\hat{\sigma}^2\hat{Q}^{-1}$, where $\hat{\sigma}^2=\frac{1}{T-1}\sum_{t=2}^{T}\left(Y_t-\hat{\phi}Y_{t-1}\right)^2$ and $\hat{Q}=\frac{1}{T-1}\sum_{t=2}^{T}\tilde{Y}_{t-1}^2$.

Remark

1. (ii) is proportional to the "homoskedasticity-only" OLS variance estimator;
2. (ii)/OLS variance estimator also works in AR(p) models.

Chapter 3 Vector Time Series

3.1 Generalized Definitions

Notation: $Y_t = (Y_{t,1}, ..., Y_{t,n})' \in \mathbb{R}^{n\times 1}$.

The definition of strict stationarity and covariance stationary can be generalized to vector time series.

Definition 3.1 (Strict Stationarity)

A process $\{Y_t : t \in \mathbb{Z}\}$ is **strictly stationary** *if and only if*

$$(Y_t, ..., Y_{t+k}) \underbrace{\sim}_{\text{"is distributed as"}} (Y_0, ..., Y_k)\,,\ \forall t \in \mathbb{Z}, k \geq 0$$

Definition 3.2 (Covariance Stationary)

A process $\{Y_t : t \in \mathbb{Z}\}$ is **covariance stationary** *iff* $\mathbb{E}[Y_{t,i}^2] < \infty$ $(\forall t, i)$ and it satisfies (*) and (**).

1. Same Expectation:
$$\mathbb{E}[Y_t] = (\mathbb{E}[Y_{t,1}], ..., \mathbb{E}[Y_{t,n}])' = \mu, \qquad (*)$$
$$\forall t \text{ (for some } \mu \in \mathbb{R}^{n\times 1})$$
2. Covariance only depends on time length:
$$\text{Cov}(Y_t, Y_{t-j}) = \mathbb{E}[\underbrace{(Y_t-\mu)(Y_{t-j}-\mu)'}_{n\times n}] = \Gamma(j), \qquad (**)$$
$$\forall t, j \text{ (for some } \Gamma(\cdot) : \mathbb{Z} \to \mathbb{R}^{n\times n})$$

Note *Matrix multiplication is not commutative. Thus,* $\Gamma(j) = \text{Cov}(Y_t, Y_{t-j}) \neq \text{Cov}(Y_{t-j}, Y_t) = \Gamma(-j)$. *However, we have*

$$\Gamma(j) = \text{Cov}(Y_t, Y_{t-j}) = \text{Cov}(Y_{t-j}, Y_t)' = \Gamma(-j)'$$

Note $\mathbb{E}[Y_{t,i}^2], \forall t, i < \infty \Leftrightarrow \sum_{i=1}^n \mathbb{E}[Y_{t,i}^2] < \infty, \forall t \Leftrightarrow \mathbb{E}[\sum_{i=1}^n Y_{t,i}^2] < \infty, \forall t \Leftrightarrow \mathbb{E}\left[\|Y_t\|^2\right] < \infty, \forall t$, *where* $\|Y_t\|^2 = Y_t'Y_t$ *is the Euclidean norm.*

Definition 3.3 (White Noise)

A process $\{\epsilon_t : t \in \mathbb{Z}\}$ is a **white noise** process iff it is covariance stationary with $\mathbb{E}[\epsilon_t] = 0$ and

$$\text{Cov}(\epsilon_t, \epsilon_{t-j}) = \begin{cases} \Sigma, & \text{if } j = 0 \\ 0, & \text{otherwise} \end{cases}$$

We use *notation* $\epsilon_t \sim \text{WN}(\underbrace{0}_{n\times 1}, \underbrace{\Sigma}_{n\times n})$.

3.2 Vector $MA(\infty)$

Definition 3.4 (Vector $MA(\infty)$)

$Y_t \sim VMA(\infty)$ iff

$$\underbrace{Y_t}_{n\times 1} = \underbrace{\mu}_{n\times 1} + \sum_{i=0}^{\infty} \underbrace{\psi_i}_{n\times n} \underbrace{\epsilon_{t-i}}_{n\times 1}, \ \forall t,$$

where

- $\epsilon_t \sim \text{WN}(0, \Sigma)$.
- $\sum_{i=0}^{\infty} \|\psi_i\|^2 < \infty$.

Note *The white noise can have different dimension than Y_t: $\epsilon_t \in \mathbb{R}^{m\times 1}, \psi_i \in \mathbb{R}^{n\times m}$.*

Existence: $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ exists (element-by-element, as a limit in mean square) iff

$$\sum_{i=0}^{\infty} \psi_{ijk}^2 < \infty, \ j,k = 1, ..., n$$

where ψ_{ijk} is element (j,k) of ψ_i. Equivalent Formulations:

$$\begin{aligned}
&\sum_{i=0}^{\infty} \psi_{ijk}^2 < \infty, \ j,k = 1, ..., n \\
\Leftrightarrow &\sum_{j,k=1}^{n} \sum_{i=0}^{\infty} \psi_{ijk}^2 < \infty \\
\Leftrightarrow &\sum_{i=0}^{\infty} \sum_{j,k=1}^{n} \psi_{ijk}^2 < \infty \\
\Leftrightarrow &\sum_{i=0}^{\infty} \|\psi_i\|^2 < \infty
\end{aligned}$$

where $\|\psi_i\|^2 = \sum_{j,k=1}^{n} \psi_{ijk}^2 = Tr(\psi_i'\psi_i)$ is (the squared) Frobenius norm of ψ_i.

Lemma 3.1 (Properties of Vector $MA(\infty)$)

For $Y_t \sim VMA(\infty)$, the following properties hold:

1. $\{Y_t\}$ is covariance stationary.
2. $\mathbb{E}[Y_t] = \mu$.
3. $\text{Cov}[Y_t, Y_{t-j}] = \sum_{i=0}^{\infty} \psi_{i+j} \Sigma \psi_i'$.

3.3 Vector $AR(1)$

Definition 3.5 (Vector $AR(1)$)

$Y_t \sim VAR(1)$ iff

$$\underbrace{Y_t}_{n\times 1} = \underbrace{c}_{n\times 1} + \underbrace{\Phi}_{n\times n}\underbrace{Y_{t-1}}_{n\times 1} + \underbrace{\epsilon_t}_{n\times 1},\ t \geq 2$$

where $\epsilon_t \sim \text{WN}(0, \Sigma)$

Lemma 3.2

If $Y_t = \mu + \sum_{i=0}^{\infty} \Phi^i \epsilon_{t-i}$, then $Y_t = c + \Phi Y_{t-1} + \epsilon_t$, where $c = (I_n - \Phi)\mu$.

Lemma 3.3

The existence of $\sum_{i=0}^{\infty} \Phi^i \epsilon_{t-i}$ can be given by one of the following *equivalent* formulations:

1. $\sum_{i=0}^{\infty} \|\Phi^i\|^2 < \infty$.
2. $|\lambda| < 1$, where λ is an eigenvalue of Φ.
3. $|I_n - \Phi z| = 0 \Rightarrow |z| > 1$. (Mostly used).

Definition 3.6 (Stability of $VAR(1)$)

The $VAR(1)$ model is **stable** iff $|I_n - \Phi z| = 0 \Rightarrow |z| > 1$.

Facts:

1. The $VAR(1)$ model admits a $VMA(\infty)$ solution of the form

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

   iff it is stable.

2. OLS "works" when the $VAR(1)$ is stable.

3.4 Spectral Analysis

Definition 3.7 (Spectral Density Function)

If $\exists f : [-\pi, \pi] \to \mathbb{C}^{n\times n}$ such that

$$\Gamma(j) = \int_{-\pi}^{\pi} \exp{(ij\lambda)}\, f(\lambda) d\lambda,\ \forall j \in \mathbb{Z},$$

then $f(\cdot)$ is called a **spectral density function**.

Given the existence of a spectral density function,

$$\Gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$$

Lemma 3.4 (Short Memory)

If $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$, then the spectral density function f exists and

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \exp(-ij\lambda) \Gamma(j), \ \lambda \in [-\pi, \pi],$$

Then,

$$f(\lambda) = f(-\lambda)^T$$
$$2\pi f(0) = \sum_{j=-\infty}^{\infty} \Gamma(j) = \Gamma(0) + \sum_{j=1}^{\infty} \{\Gamma(j) + \Gamma(j)^T\}$$

Example 3.1

$VMA(\infty)$ Case: Suppose

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \ \forall t,$$

where $\epsilon_t \sim \text{WN}(0, \Sigma)$ and $\sum_{i=0}^{\infty} \|\psi_i\|^2 < \infty$. Then, $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$ and

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \exp(-ij\lambda) \Gamma(j)$$
$$\Gamma(j) = \sum_{k=0}^{\infty} \psi_{k+j} \Sigma \psi_k^T$$

which can be rewritten as

$$f(\lambda) = \frac{1}{2\pi} \Psi(\exp(-i\lambda)) \Sigma \Psi(\exp(-i\lambda))^T$$
$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$$

Then,

$$2\pi f(0) = \Psi(1) \Sigma \Psi(1)^T$$

3.5 Estimation

Suppose

$$Y_t = \Phi Y_{t-1} + \epsilon_t, \ t \geq 2,$$

where

1. $\epsilon_t \sim i.i.d. \mathcal{N}(0, \Sigma)$.

2. $Y_1 \perp (\epsilon_2, ..., \epsilon_T)$.

Claim 3.1

$$\begin{aligned}\hat{\Phi}_{ML} = \cdots &= \left(\sum_{t=2}^{T} Y_t Y_{t-1}^T\right)\left(\sum_{t=2}^{T} Y_t Y_{t-1}^T\right)^{-1} \\ &= \underset{\Phi}{\operatorname{argmin}} \sum_{t=2}^{T} \left(Y_t - \Phi Y_{t-1}\right)^T \left(Y_t - \Phi Y_{t-1}\right) \\ &= \hat{\Phi}_{OLS}\end{aligned}$$

where

$$\left(\hat{\Phi}_{ML}, \hat{\Sigma}_{ML}\right) = \underset{(\Phi,\Sigma)}{\operatorname{argmax}} f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \Phi, \Sigma\right)$$

Lemma 3.5 (Prediction-error Decomposition)

$Y_t \mid Y_1, ..., Y_{t-1} \sim \mathcal{N}(\Phi Y_{t-1}, \Sigma)$ for $t \geq 2$. Then,

$$f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \Phi, \Sigma\right) = \prod_{t=2}^{T} f_t\left(Y_t \mid Y_1, ..., Y_{t-1}; \Phi, \Sigma\right),$$

where $f_t\left(Y_t \mid Y_1, ..., Y_{t-1}; \Phi, \Sigma\right) = \frac{1}{\sqrt{2\pi}}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(Y_t - \Phi Y_{t-1}\right)^T \Sigma^{-1}\left(Y_t - \Phi Y_{t-1}\right)\right)$

Then,

$$\underset{\Phi}{\operatorname{argmax}} f_{2:T}\left(Y_2, ..., Y_T \mid Y_1; \Phi, \Sigma\right) = \underset{\Phi}{\operatorname{argmin}} \sum_{t=2}^{T} \left(Y_t - \Phi Y_{t-1}\right)^T \Sigma^{-1}\left(Y_t - \Phi Y_{t-1}\right)$$

Lemma 3.6

$\operatorname{argmin}_\Phi \sum_{t=2}^{T} \left(Y_t - \Phi Y_{t-1}\right)^T \Sigma^{-1}\left(Y_t - \Phi Y_{t-1}\right)$ does not depend on Σ.

Thus,

$$\begin{aligned}\hat{\Phi}_{ML} &= \underset{\Phi}{\operatorname{argmin}} \sum_{t=2}^{T} \left(Y_t - \Phi Y_{t-1}\right)^T \Sigma^{-1}\left(Y_t - \Phi Y_{t-1}\right) \\ &= \underset{\Phi}{\operatorname{argmin}} \sum_{t=2}^{T} \left(Y_t - \Phi Y_{t-1}\right)^T \left(Y_t - \Phi Y_{t-1}\right) = \hat{\Phi}_{OLS}\end{aligned}$$

Proposition 3.1 (Hamilton, Prop 11.1)

Suppose

$$Y_t = \Phi Y_{t-1} + \epsilon_t,\ t \geq 2,$$

where

1. $|I_n - \Phi z| = 0 \Rightarrow |z| > 1$.
2. $\epsilon_t \sim i.i.d.(0, \Sigma)$ with $\mathbb{E}(\|\epsilon_t\|^4) < \infty$.

3. $Y_1 = \sum_{i=0}^{\infty} \Phi^i \epsilon_{1-i}$.

Then,

1. $\hat{\Phi}_{OLS}$ is consistent.
2. $\hat{\Phi}_{OLS}$ is asymptotically normal.
3. OLS variance estimator ``works."

3.6 $VAR(p)$ Models

Definition 3.8 ($VAR(p)$ Model)

$Y_t \sim VAR(p)$ iff

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \epsilon_t,\ t \geq p+1$$

where $\epsilon_t \sim WN(0, \Sigma)$.

Lemma 3.7

OLS ``works" if $\epsilon_t \sim i.i.d.(0, \Sigma)$ and if the $VAR(p)$ model is stable.

The OLS estimator is given by

$$\left(\hat{c}_{OLS}, \hat{\Phi}_{1,OLS}, \cdots, \hat{\Phi}_{p,OLS}\right) = \underset{(c,\Phi_1,\cdots,\Phi_p)}{\operatorname{argmin}} \sum_{t=p+1}^{T} \|Y_t - c - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p}\|^2$$

Using the Lag operator notation, the $VAR(p)$ model can be written as

$$\Phi(L) Y_t = c + \epsilon_t,\ t \geq p+1$$

where

$$\Phi(L) = I_n - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$

Definition 3.9 (Stability of $VAR(p)$)

The $VAR(p)$ is **stable** iff

$$|\Phi(z)| = 0 \Rightarrow |z| > 1$$

where

$$\Phi(z) = I_n - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p$$

Lemma 3.8

The $VAR(p)$ model admits an $MA(\infty)$ solution of the form

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},\ t \geq 1$$

iff the $VAR(p)$ model is stable.

Theorem 3.1 (Granger-Sims Causality)

Suppose $\underbrace{Z_t}_{n\times 1} = \left(Y_t^T, X_t^T\right)^T \sim VAR(p)$:

$$\begin{bmatrix} \underbrace{Y_t}_{m\times 1} \\ \underbrace{X_t}_{k\times 1} \end{bmatrix} = \begin{bmatrix} c_Y \\ c_X \end{bmatrix} + \begin{bmatrix} \underbrace{\Phi_{YY,1}}_{m\times m} & \underbrace{\Phi_{YX,1}}_{m\times k} \\ \underbrace{\Phi_{XY,1}}_{k\times m} & \underbrace{\Phi_{XX,1}}_{k\times k} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \cdots$$
$$+ \begin{bmatrix} \Phi_{YY,p} & \Phi_{YX,p} \\ \Phi_{XY,p} & \Phi_{XX,p} \end{bmatrix} \begin{bmatrix} Y_{t-p} \\ X_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_Y \\ \epsilon_X \end{bmatrix}$$

Then, X_t does not **Granger(-Sims) cause** Y_t if and only if

$$\Phi_{YX,1} = \cdots = \Phi_{YX,p} = 0$$

3.7 GMM for Time Series

Notation/Settings:

1. Data: $X_1, ..., X_T$
2. Parameters of interests: $\theta_0 \in \Theta \subseteq \mathbb{R}^k$ for some $k \in \mathbb{N}$.
3. Model: $\mathbb{E}[h(x_t, \theta)] = 0 \Leftrightarrow \theta = \theta_0$ for some known $\mathbb{R}^m$-valued function $h(\cdot)$, where $m \geq k$.
4. Estimator: $g_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} h(X_t, \theta) = 0$ at $\theta = \hat{\theta}_{GMM}$.

Definition 3.10 (GMM Estimator)

The GMM estimator is

$$\hat{\theta}_{GMM} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, g_T(\theta)' W_T g_T(\theta)$$

for some $m \times m$ matrix $W_T = W_T' \succeq 0$.

Example 3.2 (Sample Average)

1. $\{Y_t\}$ is covariance stationary.
2. Parameter of interest: $\mu = \mathbb{E}[Y_t], \forall t$.

3. Estimator $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$.

GMM interpretation: Let

1. $X_t = Y_t$
2. $\theta_0 = \mu \in \mathbb{R} = \Theta$ $(k = 1)$.
3. $h(x_t, \theta) = x_t - \theta$ $(m = 1)$.

Claim: $\hat{\theta}_{GMM} = \bar{Y}$ for all $W_T > 0$ (e.g. $W_T = 1$).

Example 3.3 (OLS estimator in AR(1) without intercept)

1. $Y_t = \phi Y_{t-1} + \epsilon_t$ where $\epsilon_t \sim WN(0, \sigma^2)$ and Y_0 is observed.
2. Parameter of interest: $\phi \in \mathbb{R}$.
3. OLS estimator: $\hat{\phi}_{OLS} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_{t-1}^2}$.

GMM interpretation: Let

1. $X_t = (Y_t, Y_{t-1})'$
2. $\theta_0 = \phi \in \mathbb{R} \supseteq \Theta$ $(k = 1)$.
3. $h(X_t, \theta) = Y_{t-1}(Y_t - \theta Y_{t-1})$ $(m = 1)$.

Claim: $\hat{\theta}_{GMM} = \hat{\phi}_{OLS}$ for all $W_T > 0$ (e.g. $W_T = 1$) (provided $\Theta = \mathbb{R}$).

Example 3.4 (Additional Examples of GMM)

1. Any OLS estimator.
2. Any Method of Moments (MM) estimator.
3. Any 2SLS estimator.
4. Any ML estimator.

Lemma 3.9 (Properties of GMM Estimator)

Let

$$\underbrace{G_T(\theta)}_{m\times k} = \frac{\partial}{\partial\theta'}\underbrace{g_T(\theta)}_{m\times 1},\ \theta \in \mathbb{R}^k$$

Suppose

(i). $\sqrt{T}\left(\hat{\theta}_{GMM} - \theta_0\right) = -\left[G_T(\theta_0)'W_T G_T(\theta_0)\right]^{-1} G_T(\theta_0)'W_T\sqrt{T}g_T(\theta_0) + o_P(1)$.

(ii). $G_T(\theta_0) \xrightarrow{P} G$ for some $G \in \mathbb{R}^{m\times k}$ with rank k.

(iii). $\sqrt{T}g_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, V)$ for some $V \succ 0$.

(iv). $W_T \xrightarrow{P} W$ for some $W \in \mathbb{R}^{m\times m}$ with $G'WG \succ 0$.

Then, $\sqrt{T}\left(\hat{\theta}_{GMM}-\theta_0\right) \xrightarrow{d} \mathcal{N}(0,\Omega)$, where $\Omega := [G'WG]^{-1} G'WVWG\,[G'WG]^{-1}$,

$$\Omega(W) \geq \Omega(V^{-1}) = (G'V^{-1}G)^{-1}$$

Remark

1. (iv) is automatic when $W_T = W = I_m$ (and (ii) holds).
2. 2SLS has $W_T \neq I_m$.
3. "Optimal" matrix is choosing $W = V^{-1}$ such that Ω is minimized (when $m > k$).
4. $\sqrt{T} g_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h(X_t,\theta_0)$. Thus, if $h(X_t,\theta_0)$ satisfies CLT, then (iii) holds and "usually"
$$V = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[h(X_t,\theta_0)h(X_{t-j},\theta_0)'\right]$$
5. $G_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\theta'} h(X_t,\theta_0)$. Thus, if $\frac{\partial}{\partial\theta'} h(X_t,\theta_0)$ satisfies LLN, then (ii) holds and $G = \mathbb{E}[\frac{\partial}{\partial\theta'} h(X_t,\theta_0)]$.
6. Condition (i) requires additional work.
   (a). Condition (i) - Heuristic: GMM F.O.C. is
   $$\frac{1}{2}\frac{\partial}{\partial\theta}\left[g_T(\theta)'W_T g_T(\theta)\right]\Big|_{\theta=\hat{\theta}_{GMM}} = G_T(\hat{\theta}_{GMM})'W_T g_T(\hat{\theta}_{GMM}) = 0$$
   Suppose $\hat{\theta}_{GMM} \approx \theta_0$ ($\hat{\theta}_{GMM} \xrightarrow{P} \theta_0$) and $G_T(\cdot)$ exists and is "smooth" (continuous). Then,
      I. $G_T(\hat{\theta}_{GMM}) \approx G_T(\theta_0)$,
      II. $g_T(\hat{\theta}_{GMM}) \approx g_T(\theta_0) + G_T(\hat{\theta}_{GMM})\left(\hat{\theta}_{GMM}-\theta_0\right)$
   Thus, $\left(\hat{\theta}_{GMM}-\theta_0\right) \approx -\left[G_T(\theta_0)'W_T G_T(\theta_0)\right]^{-1} G_T(\theta_0)'W_T g_T(\theta_0)$.

   (b). Condition (i) - Special Case: Suppose $g_T(\cdot)$ is affine:
   $$g_T(\theta) = A_T + B_T\theta (\text{for some } A_T, B_T)$$
   Then, $G_T(\cdot) \equiv B_T$. Thus
      I. $G_T(\hat{\theta}_{GMM}) = B_T = G_T(\theta_0)$
      II. $g_T(\hat{\theta}_{GMM}) = g_T(\theta_0) + G_T(\hat{\theta}_{GMM})\left(\hat{\theta}_{GMM}-\theta_0\right)$
   Given $[G_T(\theta_0)'W_T G_T(\theta_0)]^{-1}$ exists, then
   $$\left(\hat{\theta}_{GMM}-\theta_0\right) = -\left[G_T(\theta_0)'W_T G_T(\theta_0)\right]^{-1} G_T(\theta_0)'W_T g_T(\theta_0)$$
   e.g. OLS, 2SLS.

Choosing W_T Steps:

1. Find W^* that minimizes $\Omega(W) = [G'WG]^{-1} G'WVWG\,[G'WG]^{-1}$.
2. Find W_T such that $W_T \xrightarrow{P} W^*$.

Claim 3.2

$W^* = V^{-1}$.

Proof 3.1

$$\Omega(W) - \Omega(V^{-1}) = [G'WG]^{-1} \underbrace{\left[G'WVWG - (G'WG)\left[G'V^{-1}G\right]^{-1}(G'WG)\right]}_{:=D} [G'WG]^{-1}$$

$\Omega(W) - \Omega(V^{-1}) \succeq 0$ iff $D \succeq 0$.

Let $Z \sim \mathcal{N}(0, V)$. Then,

$$\text{Var}\left(G'WZ \mid G'V^{-1}Z\right) = G'WVWG - G'WG\left[G'V^{-1}G\right]^{-1}(G'WG) \succeq 0$$

Then, we find $W_T = \hat{V}^{-1}$ such that $\hat{V} \xrightarrow{P} V$. By (iii), $V = \lim_{T\to\infty} \text{Var}[\sqrt{T}g_T(\theta_0)] = \Gamma_n(0) + \sum_{j=1}^{\infty}[\Gamma_n(j) + \Gamma_n(j)']$, where $\Gamma_n(j) = \mathbb{E}[h(X_t, \theta_0)h(X_{t-j}, \theta_0)']$.

Proposition 3.2 (Newey-West Estimator of V)

$$\hat{V} = \hat{\Gamma}_n(0) + \sum_{j=1}^{b}\left(1 - \frac{j}{b}\right)[\hat{\Gamma}_n(j) + \hat{\Gamma}_n(j)']$$

where $\hat{\Gamma}_n(j) = \frac{1}{T}\sum_{t=j+1}^{T} h(X_t, \hat{\theta})h(X_{t-j}, \hat{\theta})'$ and $\hat{\theta}$ is an estimator of θ_0.

b is a ``tuning'' parameters ($b \to \infty$ as $T \to \infty$).

Algorithm (Two-Step GMM):

1. Find $\hat{\theta}$. (e.g. $\hat{\theta}_{GMM}$ with $W_T = I_m$).
2. Using $\hat{\theta}$ to find $\hat{V}$.
3. Using $W = \hat{V}^{-1}$ to find $\hat{\theta}_{GMM}$.

Claim 3.3

Under ``regularity'' condition,

$$\sqrt{T}\left(\hat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} N(0, \Omega^*)$$

where $\Omega^* = (G'V^{-1}G)^{-1}$

Variance Estimation for Efficient GMM: The estimator's variance is $\Omega^* = (G'V^{-1}G)^{-1}$. Its estimator is given by

$$\hat{\Omega}^* = (\hat{G}'\hat{V}^{-1}\hat{G})^{-1}$$

where $\hat{G} = G_T(\hat{\theta}_{GMM})$.

Claim 3.4

Under ``regularity'' condition, $\hat{\Omega}^* \xrightarrow{P} \Omega^*$.

Variance Estimation for GMM: The estimator's variance is $\Omega := [G'WG]^{-1} G'WVWG [G'WG]^{-1}$. Its estimator is given by

$$\hat{\Omega} = \left[\hat{G}'\hat{W}\hat{G}\right]^{-1} \hat{G}'\hat{W}\hat{V}\hat{W}\hat{G} \left[\hat{G}'\hat{W}\hat{G}\right]^{-1}$$

where

1. $\hat{G} = G_T(\hat{\theta}_{GMM})$.
2. $\hat{W} = W_T$.
3. $\hat{V}$... (why not do efficient GMM).

Chapter 4 Non-stationary Time Series

4.1

Recall that a process $\{Y_t\}$ (with $\mathbb{E}[\|Y_t\|^2] < \infty$ for all t) is covariance stationary iff (*) and (**) hold:

(*): $\mathbb{E}[Y_t] = \mu, \forall t$ (some constant μ).

(**): $\text{Cov}(Y_t, Y_{t-j}) = \Gamma(j), \forall t, j$ (some function $\Gamma(\cdot)$).

Claim 4.1

Assumption (*) is implausible for most macroeconomic time series.

Solution:

1. Decomposition:

$$Y_t = \mu_t + u_t,$$

where $\mu_t = \mathbb{E}(Y_t)$ ($\Rightarrow \mathbb{E}(u_t) = 0$).

2. (Parametric) Model for μ_t:

Example 4.1 (Leading special case: ``linear trend")

$\mu_t = \mu + \delta t$ (for some constant μ, δ).

(Reading: Chapter 16 in Hamilton.)

Theorem 4.1 (Folk Theorem)

If $\{Y_t\}$ is a macroeconomic time series, then $\{\Delta Y_t\}$ satisfies (**), but $\{Y_t\}$ does not.

How do we test this folk theorem? – Unit root testing.

If rejected, how should we model macroeconomic time series? – Cointegration.

4.1.1 Unit Root Testing

Model: The observable variable is assumed to follow

$$y_t = \mu_t + u_t, \; t \geq 1$$

where $\mu_t = \mathbb{E}[y_t]$ and $u_t \sim ARMA(1, \infty)$.

In lag operator notation,

$$(1 - \rho L)u_t = \psi(L)\epsilon_t, \; t \geq 1$$

with

1. $\|\rho\| \leq 1$.
2. $\epsilon_t \sim i.i.d.(0, \sigma^2)$.
3. $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$ with $\sum_{i=0}^{\infty} i|\psi_i| < \infty$ and $\psi(1) = \sum_{i=0}^{\infty} \psi_i \neq 0$.

Remark

1. If $\rho = 1$, then $\Delta u_t \sim MA(\infty)$.
2. If $|\rho| < 1$, then $u_t \sim MA(\infty)$ iff $u_0 = \sum_{i=0}^{\infty} \rho^i \{\psi(L)\epsilon_{-i}\}$.

Thus, we can test folk theorem by testing

$$H_0 : \rho = 1 \text{ vs. } H_1 : |\rho| < 1$$

Three Cases:

1. "Canonical Model": $\mu_t = 0$, $\psi(L) = 1$. $(1 - \rho L)y_t = \epsilon_t$. Thus, $y_t \sim AR(1)$. It is a non-standard testing problem.
2. "Serial Correlation": $\mu_t = 0$, $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$. Test statistics must be modified in this case.
3. "Deterministic": $\mu_t = \mu$ or $\mu_t = \mu + \delta t$, $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$. Distribution theory must be modified in this case.

Canonical Model

$$y_t = \rho y_{t-1} + \epsilon_t, \ t \geq 1$$

where

1. $|\rho| \leq 1$.
2. $\epsilon_t \sim i.i.d.(0, \sigma^2)$.
3. y_0 (e.g. $y_0 = 0$, using it here).

Testing problem:

$$H_0 : \rho = 1 \text{ vs. } H_1 : |\rho| < 1$$

Testing procedure: Reject for small values of $t(1)$, where

$$t(\rho_0) = \frac{\hat{\rho} - \rho_0}{\hat{\sigma}/\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}$$

with

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}, \ \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\rho} y_{t-1})^2$$