



# 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

<b>Chapter 1 Stationary Time Series</b>	<b>1</b>
1.1 Goals and Challenge . . . . .	1
1.2 Stochastic Process . . . . .	1
1.3 Strictly Stationary . . . . .	1
1.4 Covariance Stationary . . . . .	2
1.5 Autocovariance Function . . . . .	3
1.6 White Noise . . . . .	4
<b>Chapter 2 Moving-Average (MA) Process</b>	<b>5</b>
2.1 Finite Moving-Average Process . . . . .	5
2.2 Infinite Moving-Average Process $MA(\infty), VMA(\infty)$ . . . . .	6
2.3 Lag Operator Notation and Invertible $MA(q)$ . . . . .	7
2.4 $MA(q) \Leftrightarrow$ covariance stationary process with $\gamma(j) = 0, \forall j > q$ . . . . .	8
2.5 Spectral Representation . . . . .	9
2.5.1 ACF $\Leftrightarrow$ Even and PSD $\Leftrightarrow \gamma(j) = \int_{-\pi}^{\pi} \exp(ij\lambda) dF(\lambda)$ . . . . .	10
2.5.2 Spectral Density Function of $\gamma(\cdot)$ . . . . .	11
2.5.3 Spectral Analysis for Vector Time Series . . . . .	12
<b>Chapter 3 Autoregressive (AR) Model</b>	<b>14</b>
3.1 Autoregressive Model as a Special Case of $MA(\infty)$ . . . . .	14
3.2 AR Model . . . . .	14
3.2.1 AR(1) . . . . .	15
3.2.2 AR(p) . . . . .	16
3.3 Vector AR model . . . . .	17
3.3.1 Vector $AR(1)$ . . . . .	17
3.3.2 $VAR(p)$ Models . . . . .	18
<b>Chapter 4 Estimation and Inference</b>	<b>20</b>
4.1 Properties of OLS Estimators . . . . .	20
4.1.1 Consistency . . . . .	20

4.1.2	Asymptotic Normality . . . . .	21
4.2	OLS for $MA(\infty)$ . . . . .	22
4.2.1	Estimator of $\mu$ : $\bar{Y} := \frac{1}{T} \sum_{t=1}^T Y_t$ . . . . .	23
4.2.2	Estimator of $\sigma^2$ : $S^2 = \frac{1}{T-1} \sum_{t=1}^T (Y_t - \bar{Y})^2$ . . . . .	25
4.3	OLS for $AR(1)$ . . . . .	26
4.3.1	OLS Estimator $\hat{\phi}_{OLS} = \frac{\sum_{t=2}^T Y_{t-1}Y_t}{\sum_{t=2}^T Y_{t-1}^2}$ is MLE . . . . .	26
4.3.2	OLS Estimator is Biased . . . . .	27
4.3.3	OLS Estimator is Consistent . . . . .	27
4.3.4	$L^1$ -mixingale . . . . .	29
4.4	OLS for $VAR(1)$ . . . . .	33
4.5	GMM for Time Series . . . . .	34
<b>Chapter 5 Non-stationary Time Series</b>		<b>39</b>
5.1	. . . . .	39
5.1.1	Unit Root Testing . . . . .	39
<b>Appendix A Proof</b>		<b>45</b>
A.1	Proof of Lemma 2.1 . . . . .	45

# Chapter 1 Stationary Time Series

## 1.1 Goals and Challenge

**Data** in time series is denoted by

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

### Assumption 1.1

Each  $y_t$  is the realization of some random vector  $Y_t$ . In a vector form,  $Y_t = (Y_{t,1}, \dots, Y_{t,n})' \in \mathbb{R}^{n \times 1}$ .

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

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

## 1.2 Stochastic Process

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  $\mathcal{T} = \{1, \dots, T\}$  or  $\mathcal{T} = \mathbb{N} = \{1, 2, \dots\}$  or  $\mathcal{T} = \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ .
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  $\mathcal{T} = \{1, \dots, T\}$ .

## 1.3 Strictly Stationary

The definition of strict stationary is the same for both scalar and vector time series.

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

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

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

**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  $\dots = Y_{-1} = Y_0 = Y_1 = \dots$  is strictly stationary.

All these above hold for strictly stationary vector process.

**Lemma 1.1 (Property of Strictly Stationary)**

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

1. Same Expectation:

$$\mathbb{E}[Y_t] = \mu, \quad \forall t \text{ (for some constant } \mu) \quad (*)$$

2. Covariance only depends on time length:

$$\text{Cov}(Y_t, Y_{t-j}) = \gamma(j), \quad \forall t, j \text{ (for some function } \gamma(\cdot)) \quad (**)$$

Note  $\gamma(0) = \text{Var}(Y_t), \forall t$ .

## 1.4 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 scalar 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).

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

**Definition 1.4 (Covariance Stationary of Vector Process)**


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:

$$\mathbb{E}[Y_t] = (\mathbb{E}[Y_{t,1}], \dots, \mathbb{E}[Y_{t,n}])' = \mu, \forall t \text{ (for some } \mu \in \mathbb{R}^{n \times 1}) \quad (*)$$

2. Covariance only depends on time length:

$$\text{Cov}(Y_t, Y_{t-j}) = \underbrace{\mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)']}_{n \times n} = \Gamma(j), \forall t, j \text{ (for some } \Gamma(\cdot) : \mathbb{Z} \rightarrow \mathbb{R}^{n \times n}) \quad (**)$$

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

## 1.5 Autocovariance Function

**Definition 1.5 (Autocovariance Function)**

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

**Lemma 1.2 (ACF Property)**

The autocovariance function satisfies the following properties:

For a scalar process:

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  $a_1, \dots, a_n$ ,

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

For a vector process: 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

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

**Definition 1.6 (Autocorrelation Function for Scalar Process)**

The **autocorrelation function** is

$$\rho(j) = \text{Corr}(Y_t, Y_{t-j}) = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t-j})}} = \frac{\gamma(j)}{\gamma(0)}$$

## 1.6 White Noise

### Definition 1.7 (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.,  $\epsilon_t \sim \text{WN}(0, \sigma^2)$ . That is,  $\epsilon_t \sim \text{i.i.d.}(0, \sigma^2)$  is a ‘stronger’ assumption than  $\epsilon_t \sim \text{WN}(0, \sigma^2)$ .
2. Gauss-Markov theorem assumes WN errors.
3. WN terms are used as “building blocks”: often a variable can be generated as

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

In the vector form, we have

### Definition 1.8 (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  $\epsilon_t \sim \text{WN}(\underbrace{0}_{n \times 1}, \underbrace{\Sigma}_{n \times n})$ .

## Chapter 2 Moving-Average (MA) Process

### 2.1 Finite Moving-Average Process

Each data is related to white noises in previous periods.

#### Definition 2.1 (MA(1))

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

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

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

The MA(1) process,  $\{Y_t\}$ , is *covariance stationary*:

1.  $\mathbb{E}[Y_t] = \mu$  and
2. the 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 2.2 (MA(p))

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

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

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

The MA(p) process,  $\{Y_t\}$ , is *covariance stationary*:

1.  $\mathbb{E}[Y_t] = \mu$  and
2. the autocovariance function is

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

where  $\theta_0 = 1$ .



## 2.2 Infinite Moving-Average Process $MA(\infty), VMA(\infty)$

### Definition 2.3 ( $MA(\infty)$ )

Infinite Moving-Average Process:  $Y_t \sim MA(\infty)$  iff

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

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

### Lemma 2.1 ( $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ is required for covariance stationarity)

For the  $MA(\infty)$  process defined above,  $\{Y_t\}$ , it is *covariance stationary*: i.e.,

1.  $\mathbb{E}[Y_t] = \mu$  and
2. the autocovariance function is

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

if and only if

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty$$

### Proof 2.1

See A.1.

### Definition 2.4 (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

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



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

### Lemma 2.2 (Properties of Vector $MA(\infty)$ )

For  $Y_t \sim VMA(\infty)$ , the following properties hold:  $\{Y_t\}$  is covariance stationary,

1.  $\mathbb{E}[Y_t] = \mu$  and
2. the autocovariance function is

$$\Gamma(j) = \text{Cov}(Y_t, Y_{t-j}) = \sum_{i=0}^{\infty} \psi_{i+j} \Sigma \psi_i'$$

Note that the existence requirement here is  $\sum_{i=0}^{\infty} \|\psi_i\|^2 < \infty$ .

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

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

where  $\psi_{ijk}$  is element  $(j, k)$  of  $\psi_i$ . Equivalent Formulations:

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

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

#### Remark

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.

## 2.3 Lag Operator Notation and Invertible $MA(q)$

### Definition 2.5 (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  $\{X_t\}$ ,

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

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

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

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

**Definition 2.6 (Invertibility of  $MA(q)$ )**

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

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

(All solutions are greater than 1).

**Lemma 2.3 (Invertible  $\Leftrightarrow \exists \Pi(L)$ )**

If the  $MA(q)$  model is invertible, then there exists a  $\Pi(L) = \sum_{i=0}^{\infty} \pi_i L^i$  with  $\sum_{i=0}^{\infty} |\pi_i| < \infty$  such that

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

**Proof 2.2**

The equation is equivalent to  $\epsilon_t = \Pi(L)\theta(L)\epsilon_t \Leftrightarrow 1 = \Pi(L)\theta(L)$ .

**Technicalities**

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

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

then

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

- Invertibility  $\Rightarrow (*)$ .
- If  $X_0, X_1, \dots$  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  $\sum_{i=0}^{\infty} |\pi_i| < \infty$ .

## 2.4 $MA(q) \Leftrightarrow$ covariance stationary process with $\gamma(j) = 0, \forall j > q$

**Proposition 2.1 ( $MA(q) \Leftrightarrow$  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? That is, if a covariance stationary process has  $\gamma(j) > 0, \forall j$ , is it an  $MA(\infty)$ ? No.

**Example 2.1 (Counterexample)**

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

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”.

**Definition 2.7 (Deterministic)**

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

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

**Claim 2.1**

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

## 2.5 Spectral Representation

**Definition 2.8 (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 of a covariance stationary process, 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  $a_1, \dots, a_n$ ,

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

**2.5.1 ACF  $\Leftrightarrow$  Even and PSD  $\Leftrightarrow \gamma(j) = \int_{-\pi}^{\pi} \exp(ij\lambda) dF(\lambda)$**

**Proposition 2.2 (ACF  $\gamma(\cdot) \Leftrightarrow$  Even and PSD)**

A function  $\gamma(\cdot) : \mathbb{Z} \rightarrow \mathbb{R}$  is an ACF iff it is even and positive semi-definite.

**Theorem 2.1 (Herglotz's Theorem:  $\gamma(j) = \int_{-\pi}^{\pi} \exp(ij\lambda) dF(\lambda) \Leftrightarrow$  Even and PSD)**

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

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

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

**Definition 2.9 (Spectral Distribution/Density Function)**

If  $\exists f : [-\pi, \pi] \rightarrow \mathbb{R}$  such that

$$\begin{aligned} \gamma(j) &= \int_{-\pi}^{\pi} \exp(ij\lambda) dF(\lambda), \\ F(\lambda) &= \int_{-\pi}^{\lambda} f(r) dr, \forall \lambda \in [-\pi, \pi], \end{aligned}$$

then  $F(\cdot)$  is called the spectral distribution function and  $f(\cdot)$  is called a spectral density function (of  $\gamma(\cdot)$ ).

**Lemma 2.4 ( $\int_{-\pi}^{\pi} \exp(ij\lambda) f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda) d\lambda$ )**

The spectral representation can be written as

$$\gamma(j) = \int_{-\pi}^{\pi} \exp(ij\lambda) f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda) d\lambda$$

**Proof 2.3**

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,

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

By the property of  $\cos(\cdot)$ ,  $\gamma(j)$  is even.

### Example 2.2

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

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

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

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

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

## 2.5.2 Spectral Density Function of $\gamma(\cdot)$

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

*Partial Answer:*

### Proposition 2.3 (Spectral Density Function of $\gamma(\cdot)$ )

An even function  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$  with " $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$ " is psd if and only if

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \cos(j\lambda) \gamma(j) \geq 0, \quad \forall \lambda \in [-\pi, \pi], \quad (2.1)$$

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

### Definition 2.10 (Short Memory: $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$ )

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

### Corollary 2.1 (Formally, Spectral Density Function of $\gamma(\cdot)$ )

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

1.  $f(\cdot)$  exists (given as (2.1)) and is bounded.
2.  $\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$ .
3.  $f(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j)$ .

### Example 2.3 ( $MA(\infty)$ Case)

Suppose

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

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

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

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

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

### 2.5.3 Spectral Analysis for Vector Time Series

#### Definition 2.11 ((Vector Form) Spectral Density Function)

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

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

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 2.5 (Short Memory)

If the covariance stationary process has **short memory** ( $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$ ), then the spectral density function  $f$  exists and

$$\underbrace{f(\lambda)}_{n \times n} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \exp(-ij\lambda) \underbrace{\Gamma(j)}_{n \times n}, \quad \lambda \in [-\pi, \pi], \quad (2.2)$$

Then, given (2.2), we have the following properties:

$$f(\lambda) = f(-\lambda)^T$$

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

**Example 2.4 ( $VMA(\infty)$  Case)**

Suppose

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

where

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

Then,

- $\Gamma$  has short memory ( $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$ );
- $\Gamma$  has spectral density function given by

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

where  $\Gamma(j) = \sum_{k=0}^{\infty} \psi_{k+j} \Sigma \psi_k^T$ . Alternatively, it can be rewritten as

$$f(\lambda) = \frac{1}{2\pi} \Psi(\exp(-i\lambda)) \Sigma \Psi(\exp(-i\lambda))^T$$

where  $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ . Then,

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



## Chapter 3 Autoregressive (AR) Model

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

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

#### Example 3.1 (Autoregressive model)

Suppose

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

where

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

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

#### Lemma 3.1 (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.  $\gamma(j) \propto \phi^j$ , i.e., decays exponentially.

#### Proof 3.1

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

### 3.2 AR Model

#### Definition 3.1 (Alternative Representation of AR)

Alternatively, the AR model can be represented as

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

**Proof 3.2**

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

The natural estimator of  $\phi$  (OLS) is

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

**Definition 3.2 (Model for Finite AR)**

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

where

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

More generally, consider an AR with a drift,

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

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

**3.2.1 AR(1)****Definition 3.3 (AR(1))**

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

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

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



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

**Definition 3.4 (Stability of AR(1))**

The AR(1) model is **stable** iff  $|\phi| < 1$ .

- If stable ( $|\phi| < 1$ ) 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

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

if and only if  $|\phi| < 1$ .

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



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

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

### 3.2.2 AR(p)

#### Definition 3.5 (AR(p))

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

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

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

In vector notation, we can write

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

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

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

#### Definition 3.6 (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  $\{X_t\}$ ,

$$LX_t = X_{t-1}$$

$$\vdots$$

$$L^k X_t = X_{t-k}, \quad \forall t \in \mathbb{Z}$$

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

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

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

**Definition 3.7 (Stability of  $AR(p)$ )**

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

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

(All solutions are greater than 1).

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

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

if and only if it is *stable*. The  $MA(\infty)$  solution has

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

and (computable)  $\psi_i$ 's satisfy

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

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

**Claim 3.1**

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

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

### 3.3 Vector AR model

#### 3.3.1 Vector $AR(1)$

**Definition 3.8 (Vector  $AR(1)$ )**

$Y_t \sim VAR(1)$  iff

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

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

**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  $c = (I_n - \Phi)\mu$ .

**Definition 3.9 (Stability of  $VAR(1)$ )**

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

**Lemma 3.3 (Equivalence of Stability)**

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

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

Facts:

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

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

if and only if it is stable.

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

**3.3.2  $VAR(p)$  Models****Definition 3.10 ( $VAR(p)$  Model)**

$Y_t \sim VAR(p)$  iff

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

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

**Lemma 3.4**

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

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

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

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

where

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

**Definition 3.11 (Stability of  $VAR(p)$ )**

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

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

where

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

**Lemma 3.5**

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

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

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

**Theorem 3.1 (Granger-Sims Causality)**

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

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

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

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

## Chapter 4 Estimation and Inference

### 4.1 Properties of OLS Estimators

The OLS model can be written as

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

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  $\beta$ ) is given by

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

#### Lemma 4.1 (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:  $\mathbb{E}[\epsilon_i \mid x_1, \dots, x_n] = 0, \forall i$ .

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.

#### 4.1.1 Consistency

#### Lemma 4.2 (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 4.1**

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

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

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.

**4.1.2 Asymptotic Normality****Lemma 4.3 (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,  $\sqrt{n} (\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, \Omega)$ , where  $\Omega := Q^{-1} V Q^{-1}$

**Proof 4.2**

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

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

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 4.1 (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,  $\hat{\Omega} := \hat{Q}^{-1} \hat{V} \hat{Q}^{-1} \xrightarrow{P} Q^{-1} V Q^{-1} := \Omega$  (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

$$V = \mathbb{E}[x_i x_i' \epsilon_i^2] = \underbrace{\dots}_{LIE} \sigma^2 \mathbb{E}[x_i x_i'] = \sigma^2 Q$$

and

$$\hat{V} = \hat{\sigma}^2 \hat{Q}, \text{ where } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}'_{OLS} x_i)^2$$

## 4.2 OLS for $MA(\infty)$

Consider the  $MA(\infty)$  model:

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

where

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

$$2. \sum_{i=0}^{\infty} i|\psi_i| < \infty.$$

#### 4.2.1 Estimator of $\mu$ : $\bar{Y} := \frac{1}{T} \sum_{t=1}^T Y_t$

Consider the estimator (for  $\mu$ ):

$$\bar{Y} := \frac{1}{T} \sum_{t=1}^T Y_t = \underset{m}{\operatorname{argmin}} \sum_{t=1}^T (Y_t - m)^2$$



**Note**

1.  $\epsilon_t \sim i.i.d.(0, \sigma^2) \Rightarrow \epsilon_t \sim \text{WN}(0, \sigma^2)$  (i.e., a stronger assumption than common assumption).
2.  $\sum_{i=0}^{\infty} i|\psi_i| < \infty \Rightarrow \sum_{i=0}^{\infty} |\psi_i| < \infty \Rightarrow \sum_{i=0}^{\infty} \psi_i^2 < \infty$  (i.e., a stronger assumption than common assumption)

#### Lemma 4.4 (Unbiasedness)

$\bar{Y}$  is an unbiased estimator of  $\mu$ .

#### Proof 4.3

Recall that  $\mathbb{E}(Y_t) = \mu$  because  $\epsilon_t \sim \text{WN}(0, \sigma^2)$  and  $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ . Then,  $\mathbb{E}[\bar{Y}] = \mathbb{E}[\frac{1}{n} \sum_{i=1}^n Y_i] = \mu$ .

#### Lemma 4.5 (Consistency)

$\bar{Y}$  is a consistent estimator of  $\mu$ , i.e.,  $\bar{Y} \xrightarrow{P} \mu$ .

#### Proof 4.4

It can be proven by  $P(|\bar{Y} - \mu| > \eta) \xrightarrow{T \rightarrow \infty} 0$  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  $\eta > 0$ . Then, we prove that the variance of  $\bar{Y}$  is bounded:

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

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  $\sum_{i=0}^{\infty} |\gamma(i)| = \sum_{j=0}^{\infty} |(\sum_{i=0}^{\infty} \psi_i \psi_{i+j}) \sigma^2| < \infty$  (aka “short memory”). Therefore, we have  $\bar{Y} \xrightarrow{P} \mu$ .

#### Lemma 4.6 (Asymptotic Normality)

$\bar{Y}$  is an asymptotic normal estimator of  $\mu$ , i.e.,  $\sqrt{T}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \omega^2)$ , where  $\omega^2 \neq \text{Var}(Y_t)$  (in general).

**Proof 4.5**

Idea of proof:

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

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

**Theorem 4.1 (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

$$\psi(L) = \psi(1) + \tilde{\psi}(L)(1 - L) \quad (4.1)$$

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 4.6**

By the definition of  $\tilde{\psi}(L) = \sum_{i=0}^{\infty} \tilde{\psi}_i L^i$ , the RHS of (4.1) 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  $L^i$ :

1.  $i = 0$ :  $\psi(1) + \tilde{\psi}_0 = \psi_0$
2.  $i \geq 1$ :  $\tilde{\psi}_i - \tilde{\psi}_{i-1} = \psi_i$

The (4.1) 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,

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

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

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

### Remark

1. 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  $\sum_{i=0}^{\infty} \sqrt{i}|\psi_i| < \infty$ .
2. If  $\epsilon_t \sim i.i.d. (0, \sigma^2)$ , then  $\epsilon_t \sim \text{WN}(0, \sigma^2)$  and  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ . (These two properties may hold even if  $\epsilon_t \sim i.i.d. (0, \sigma^2)$ , i.e., there is a weaker condition can be used.)
3.  $\omega^2 = \psi(1)^2 \sigma^2 \neq (\sum_{i=0}^{\infty} \psi_i^2) \sigma^2 = \text{Var}(Y_t)$  (in general.)
4.  $\omega^2$  is called the “**long-run variance**” of  $Y_t$ :

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

### 4.2.2 Estimator of $\sigma^2$ : $S^2 = \frac{1}{T-1} \sum_{t=1}^T (Y_t - \bar{Y})^2$

The OLS (variance) estimator is

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

#### Claim 4.1

$$S^2 \xrightarrow{P} \text{Var}(Y_t).$$

Recall that  $\sqrt{T}(\bar{Y} - \mu) \xrightarrow{d} \mathcal{N}(0, \omega^2)$ , where  $\omega^2 = \psi(1)^2 \sigma^2$  and  $\psi(1) = \sum_{i=0}^{\infty} \psi_i$ .

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

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

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

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 4.1 (Newey-West, 1987)

$\hat{\omega}^2 = \hat{\gamma}(0) + 2 \sum_{j=1}^b \left(1 - \frac{j}{b}\right) \hat{\gamma}(j)$ , where  $\hat{\gamma}(j) = \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})$  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  $b \rightarrow \infty$  and  $\frac{b}{\sqrt{T}} \rightarrow 0$  as  $T \rightarrow \infty$ .

### 4.3 OLS for AR(1)

Suppose

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

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

The OLS Estimator of  $\phi$  is

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

#### 4.3.1 OLS Estimator $\hat{\phi}_{OLS} = \frac{\sum_{t=2}^T Y_{t-1} Y_t}{\sum_{t=2}^T Y_{t-1}^2}$ is MLE

##### Claim 4.2 (OLS Estimator is MLE)

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

##### Proof 4.7

The (conditional) MLE of  $(\phi, \sigma^2)$  is

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

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

**Definition 4.1 (Prediction-error Decomposition)**

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

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

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

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

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

Therefore,

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

**4.3.2 OLS Estimator is Biased**

Usual template (“strict exogeneity”):  $\mathbb{E}[\epsilon_t | Y_1, \dots, Y_{T-1}] = 0$ ,  $t \geq 2$ . However, it doesn’t hold here.

**Claim 4.3 ( $\hat{\phi}_{OLS}$  is Biased)**

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

**Proof 4.8**

The OLS estimator can be written as

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

where  $\epsilon_t = Y_t - \phi Y_{t-1}$ ,  $t \geq 2$ . For every  $t$ ,  $\epsilon_t$  is independent of  $Y_{t-1}$  but is not independent of  $\sum_{i=2}^T Y_{i-1}^2$ .

If  $\phi$  is positive, then a positive shock to  $\epsilon_t$  raises all  $Y_i$  with  $i \geq t$ . This means there is negative correlation between  $\epsilon_t$  and  $\frac{Y_{t-1}}{\sum_{i=2}^T Y_{i-1}^2}$ , so  $\mathbb{E}[\hat{\phi}_{OLS}] < \phi$ .

**4.3.3 OLS Estimator is Consistent**

Usual template, i.e., the Lemma 4.2. The estimator  $\hat{\phi}$  is consistent 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 4.4 ( $\hat{\phi}_{OLS}$ is Consistent)

$\hat{\phi}_{OLS}$  is consistent. That is, 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  $r_t \rightarrow \infty$ ,  $\frac{X_t}{r_t} \rightarrow 0$ .

1. If  $X_t = O(1)$  and if  $r_t \rightarrow \infty$ , then  $\frac{X_t}{r_t} = o(1)$  (“ $\rightarrow 0$ ”).

2. If  $\{X_t\}$  is random with  $X_t = O_P(1)$  and if  $r_t \rightarrow \infty$ , then  $\frac{X_t}{r_t} = o_P(1)$  (“ $\xrightarrow{P} 0$ ”).

#### Definition 4.2 (Stochastically Bounded)

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

Then, we can prove the consistency:

#### Proof 4.9

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

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

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

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 4.2 (Mixingale LLN)

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

$$\underbrace{\lim_{T \rightarrow \infty}}_{\text{"lim sup } T \rightarrow \infty"} \frac{1}{T} \sum_{t=1}^T C_t < \infty,$$

then

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

To implement the Mixingale LLN, we need to define ' $L^1$ -mixingale' and 'uniformly integrable'.

#### 4.3.4 $L^1$ -mixingale

##### Definition 4.3 ( $L^1$ -mixingale)

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

$$\mathbb{E}[X_t \mid Z_t, Z_{t-1}, \dots] = X_t, \forall t \quad (4.2)$$

$$\mathbb{E}(|\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, \dots]|) \leq c_t \xi_m, \forall t, m \geq 1 \quad (4.3)$$

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

##### Lemma 4.7 ( $X_t \sim i.i.d$ with $\mathbb{E}[X_t] = 0 \Rightarrow L^1$ -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).  $\frac{1}{T} \sum_{t=1}^T X_t \xrightarrow{P} 0$ .

##### Lemma 4.8 ( $L^1$ -mixingale $\Rightarrow \mathbb{E}[X_t] = \mathbb{E}(\mathbb{E}[X_t \mid Z_{t-m}, Z_{t-m-1}, \dots]) = 0$ )

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

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

##### Example 4.2 ( $X_t = Z \sim \mathcal{N}(0, 1), \forall t$ is not an $L^1$ -mixingale)

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

- (i).  $\{X_t\}$  is not an  $L^1$ -mixingale,
- (ii).  $\frac{1}{T} \sum_{t=1}^T X_t = Z \not\xrightarrow{P} 0$ .



**Remark**

1. If  $Z_t = X_t$ , then (4.2) holds.
2. If (4.2) and (4.3) hold, then they hold with  $Z_t = X_t$ .
3. If  $X_t = g(\epsilon_t, \epsilon_{t-1}, \dots)$ , then (4.2) 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 4.3 (Important Case)**

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

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

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

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

**Definition 4.4 (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

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

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

**Uniformly Integrability****Definition 4.5 (Uniformly Integrable)**

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

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

**Remark**

1. If  $X_T \xrightarrow{d}_{T \rightarrow \infty} \mathcal{N}(0, 1)$  and if  $\{X_T\}$  is uniformly integrable, then  $\mathbb{E}[X_T] \rightarrow_{T \rightarrow \infty} 0$ .

2. Integrality:  $\mathbb{E}[|X_T|] < \infty$  iff  $\lim_{m \rightarrow \infty} \mathbb{E}[|X_T| \mathbf{1}(|X_T| > M)] = 0$ .

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 4.4 (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

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

#### Corollary 4.1 (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,  $\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, Q^{-1}VQ^{-1})$ .

#### Claim 4.5

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

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

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

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

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

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

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

#### Theorem 4.3 (Martingale CLT, (Brown, 1971))

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

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

(conditional second moment condition). Then,

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

For the AR(1) example, we have

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

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

- (Conditional) second moment condition:

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

#### Proof 4.10

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

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

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

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

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

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

### Variance Estimation

To be estimated:

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

Consistent estimators:

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

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

## 4.4 OLS for VAR(1)

Suppose

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

where

1.  $\epsilon_t \sim i.i.d. \mathcal{N}(0, \Sigma)$ .
2.  $Y_1 \perp (\epsilon_2, \dots, \epsilon_T)$ .

#### Claim 4.6 (OLS Estimator is MLE)

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

where

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

**Proof 4.11****Definition 4.6 (Prediction-error Decomposition)**

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

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

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

Then,

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

**Lemma 4.9**

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

Thus,

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

**Proposition 4.2 (Hamilton, Prop 11.1)**

Suppose

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

where

1. Stable:  $\sum_{i=0}^{\infty} \|\Phi^i\|^2 < \infty$  (by Lemma 3.3, it is equivalent to  $|I_n - \Phi z| = 0 \Rightarrow |z| > 1$ ).
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."

**4.5 GMM for Time Series**

Notation/Settings:

1. Data:  $X_1, \dots, X_T$
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 4.7 (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 4.5 (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 4.6 (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 4.7 (Additional Examples of GMM)

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

## 4. Any ML estimator.

**Lemma 4.10 (Properties of GMM Estimator)**

Let

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

Suppose

- (i).  $\sqrt{T}(\hat{\theta}_{GMM} - \theta_0) = -[G_T(\theta_0)'W_T G_T(\theta_0)]^{-1} G_T(\theta_0)'W_T \sqrt{T}g_T(\theta_0) + o_P(1)$ .
- (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,  $\sqrt{T}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$ , 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  $\Omega$  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”

$$V = \sum_{j=-\infty}^{\infty} \mathbb{E} [h(X_t, \theta_0) h(X_{t-j}, \theta_0)']$$

- 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

$$\frac{1}{2} \frac{\partial}{\partial \theta} [g_T(\theta)'W_T g_T(\theta)] \Big|_{\theta=\hat{\theta}_{GMM}} = G_T(\hat{\theta}_{GMM})'W_T g_T(\hat{\theta}_{GMM}) = 0$$

Suppose  $\hat{\theta}_{GMM} \approx \theta_0$  ( $\hat{\theta}_{GMM} \xrightarrow{P} \theta_0$ ) and  $G_T(\cdot)$  exists and is “smooth” (continuous). Then,

$$\text{I. } G_T(\hat{\theta}_{GMM}) \approx G_T(\theta_0),$$

$$\text{II. } g_T(\hat{\theta}_{GMM}) \approx g_T(\theta_0) + G_T(\hat{\theta}_{GMM}) (\hat{\theta}_{GMM} - \theta_0)$$

$$\text{Thus, } (\hat{\theta}_{GMM} - \theta_0) \approx -[G_T(\theta_0)'W_T G_T(\theta_0)]^{-1} G_T(\theta_0)'W_T g_T(\theta_0).$$

(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.  $g_T(\hat{\theta}_{GMM}) = g_T(\theta_0) + G_T(\hat{\theta}_{GMM}) (\hat{\theta}_{GMM} - \theta_0)$

Given  $[G_T(\theta_0)' W_T G_T(\theta_0)]^{-1}$  exists, then

$$(\hat{\theta}_{GMM} - \theta_0) = - [G_T(\theta_0)' W_T G_T(\theta_0)]^{-1} G_T(\theta_0)' W_T g_T(\theta_0)$$

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 4.7**

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

**Proof 4.12**

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

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

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

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

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

**Proposition 4.3 (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  $\theta_0$ .

$b$  is a "tuning" parameters ( $b \rightarrow \infty$  as  $T \rightarrow \infty$ ).

**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 4.8**

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 4.9**

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

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

where

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

# Chapter 5 Non-stationary Time Series

## 5.1

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  $\mu$ ).

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

### Claim 5.1

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  $\mu_t$ :

**Example 5.1 (Leading special case: ``linear trend'')**

$$\mu_t = \mu + \delta t \text{ (for some constant } \mu, \delta \text{)}.$$

(Reading: Chapter 16 in Hamilton.)

### Theorem 5.1 (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.

### 5.1.1 Unit Root Testing

Model: The observable variable is assumed to follow

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

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

In lag operator notation,

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

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**

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

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

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

Motivation: If  $|\rho_0| < 1$ , then

(i). The one-sided test based on  $t(\rho_0)$  is the asymptotically UMP level  $\alpha$  test of

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

where  $\epsilon_t \sim N(0, \sigma^2)$ .

(ii).  $t(\rho_0) \xrightarrow{d} \mathcal{N}(0, 1)$ .



**Note**

1. Both (i) and (ii) fail for  $\rho_0 = 1$ .
2. The one-sided test based on  $t(1)$  is ‘nearly’ optimal when  $\rho = 1$ .

Distribution Theory:

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

Recall that, if  $|\rho| < 1$ , we have

1.  $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2 \xrightarrow{P} \sigma^2$
2.  $\frac{1}{T} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{P} \frac{\sigma^2}{1-\rho^2}$
3.  $\frac{1}{\sqrt{T}} \sum_{t=1}^T y_{t-1} \epsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{1-\rho^2}\right)$ .

Thus,

$$t(\rho) = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^T y_{t-1} \epsilon_t}{\hat{\sigma} \sqrt{\frac{1}{T} \sum_{t=1}^T y_{t-1}^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$



**Note** If  $\rho = 1$ , are these three results robust? We have that

1.  $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$
2.  $\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{P} \text{something}$
3.  $\frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t \xrightarrow{d} \text{something else.}$

Thus,

$$t(1) = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t}{\hat{\sigma} \sqrt{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}} \xrightarrow{d} \text{something complicated}$$

Now, we want to find the asymptotic null distribution of

$$\left( \frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2, \frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t \right)$$

We can use the following approach:

Step 1. Approximate (joint) distribution of  $y_1, \dots, y_T$ .

Tool: Functional CLT

Task: Find 1-to-1 function of  $y_1, \dots, y_T$  that satisfies the Functional CLT.

Step 2. Approximate null distribution of  $t(1)$ .

Tool: Continuous mapping theorem (CMT)

Task: Express  $\left(\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2, \frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t\right)$  as a continuous transformation of the function from step 1.

Step 1: Heuristics.

1. When  $\rho = 1$ , the distribution of  $\{y_t : 0 \leq t \leq T\}$  is that of a discrete time random walk.
2. The  $\rho = 1$  distribution of  $\{y_t : 0 \leq t \leq T\}$  can be approximated by (mean of) that of a continuous time Gaussian random walk (i.e. Brownian motion).

#### Definition 5.1 (Random Walk)

A process  $\{y_t : t \geq 0\}$  is a random walk if

1.  $y_0 = 0$
2.  $y_t - y_{t-1} \sim i.i.d.(0, \sigma^2)$  (for  $t \geq 1$ )

If the model is

$$y_t = \rho y_{t-1} + \epsilon_t, \quad t \geq 1 \text{ with } y_0 = 0, \epsilon_t \sim i.i.d.(0, \sigma^2)$$

Then,  $y_t$  is a random walk when  $\rho = 1$ .

#### Lemma 5.1

Properties of the random walk

1.  $y_t = \sum_{i=1}^t \epsilon_i, \epsilon_i \sim i.i.d.(0, \sigma^2)$  for all  $t$ .
2. For any  $0 \leq t_1 \leq t_2$ ,

$$y_{t_2} - y_{t_1} = \sum_{i=t_1+1}^{t_2} \epsilon_i \perp \{y_t : 0 \leq t \leq t_1\}$$

and

$$y_{t_2} - y_{t_1} \sim y_{t_2-t_1} - \underbrace{y_0}_{=0}$$

#### Definition 5.2 (Brownian Motion)

A continuous time process  $\{Y(r) : 0 \leq r \leq 1\}$  is a Brownian motion (with variance  $\sigma^2$ )  $Y \sim BM(\sigma^2)$ , iff

1.  $Y(0) = 0$
2. For any  $0 \leq r_1 \leq r_2 \leq 1$ ,

$$Y(r_2) - Y(r_1) \perp \{Y(r) : 0 \leq r \leq r_1\}$$

and

$$Y(r_2) - Y(r_1) \sim Y(r_2 - r_1) \sim N(0, \sigma^2(r_2 - r_1))$$

3.  $Y(\cdot)$  is continuous.



**Note** A ‘‘standard’’ Brownian motion,  $BM(1)$ , is called a Wiener process.

Objective: Suppose  $\{y_t : t \geq 0\}$  is a discrete time random walk:  $y_t - y_{t-1} \sim i.i.d.(0, \sigma^2)$ . Find functions  $\{Y_T(\cdot) : T \geq 1\}$  such that

1. For each  $T \geq 1$ ,  $Y_T(\cdot)$  is a 1-to-1 function of  $\{y_t : 0 \leq t \leq T\}$ .
2.  $Y_T(\cdot) \xrightarrow{d} Y \sim BM(\sigma^2)$ .

Definition/Construction of  $Y_T(\cdot)$ : For  $T \geq 1$  and  $r \in [0, 1]$ , let

$$Y_T(r) = \frac{y_{\lfloor Tr \rfloor}}{\sqrt{T}},$$

where  $\lfloor \cdot \rfloor$  is the integer part of the argument. Thus,

$$Y_T(r) = \begin{cases} \frac{y_0}{\sqrt{T}} = 0 & \text{if } 0 \leq r < \frac{1}{T} \\ \frac{y_1}{\sqrt{T}} & \text{if } \frac{1}{T} \leq r < \frac{2}{T} \\ \vdots & \vdots \\ \frac{y_{T-1}}{\sqrt{T}} & \text{if } \frac{T-1}{T} \leq r < 1 \\ \frac{y_T}{\sqrt{T}} & \text{if } r = 1 \end{cases}$$

Distribution Theory: Suppose

1.  $Y_T(r) = \frac{y_{\lfloor Tr \rfloor}}{\sqrt{T}}$ ,  $y_t = \sum_{s=1}^t \epsilon_s$ ,  $\epsilon_s \sim i.i.d.(0, \sigma^2)$ .
2.  $Y \sim BM(\sigma^2)$ .

Then,

$$\begin{aligned} \mathbb{E}[Y_T(r)] &= 0 = \mathbb{E}[Y(r)], \forall r, T \\ \text{Var}[Y_T(r)] &= \frac{\text{Var}[y_{\lfloor Tr \rfloor}]}{T} = \frac{\sigma^2 \lfloor Tr \rfloor}{T} \xrightarrow{T \rightarrow \infty} \sigma^2 r = \text{Var}[Y(r)], \forall r \\ Y_T(r) &= \frac{y_{\lfloor Tr \rfloor}}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor Tr \rfloor} \epsilon_s = \underbrace{\frac{\sqrt{\lfloor Tr \rfloor}}{\sqrt{T}}}_{\xrightarrow{T \rightarrow \infty} \sqrt{r}} \underbrace{\frac{1}{\sqrt{\lfloor Tr \rfloor}} \sum_{s=1}^{\lfloor Tr \rfloor} \epsilon_s}_{\xrightarrow{d} \mathcal{N}(0, \sigma^2)} \xrightarrow{d} \mathcal{N}(0, \sigma^2 r) \sim Y(r), \forall r \end{aligned}$$

### Claim 5.2

$$Y_T \xrightarrow{d} Y$$

Probability theory of function spaces:

Sample space:  $D[0, 1]$  = The space of ‘CADLAG’ functions on  $[0, 1]$ .

- ‘CAD’: Continuous from the right.
- ‘LAG’: Has left limits.

$\sigma$ -algebra: Borel  $\sigma$ -algebra induced by metric on  $D[0, 1]$ .

Metric:

$$d_{\text{sup}}(f, g) = \sup \{|f(r) - g(r)| : 0 \leq r \leq 1\}, f, g \in D[0, 1]$$

Functional Limit Theory:

- Convergence in probability:

Scalar Case:  $X_T \xrightarrow{P} X$  iff

$$P(|X_T - X| > \eta) \xrightarrow{T \rightarrow \infty} 0, \forall \eta > 0$$

Functional Case:  $Y_T \xrightarrow{P} Y$  iff

$$P(d_{\text{sup}}(Y_T, Y) > \eta) \xrightarrow{T \rightarrow \infty} 0, \forall \eta > 0$$

- Convergence in distribution

# Appendix A Proof

## A.1 Proof of Lemma 2.1



**Note** Conjecture:

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

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

The necessary condition to make these conjectures correct is

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

which is sufficient given our definition of  $MA(\infty)$ .

### Claim A.1

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

### Remark

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

### Definition A.1 (Stochastic Convergence in Mean Square)

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

### Lemma A.1

The properties of the  $S$  are

1.  $S$  is "essentially unique."
2.  $\mathbb{E}[S] = \sum_{i=0}^{\infty} \mathbb{E}[X_i] = \lim_{n \rightarrow \infty} \sum_{i=0}^n \mathbb{E}[X_i]$
3.  $\text{Var}[S] = \dots = \lim_{n \rightarrow \infty} \text{Var}[\sum_{i=0}^n X_i]$



4. (Higher order moments of  $S$  are similar)  $\dots$ **Theorem A.1 (Cauchy Criterion)**

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

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

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

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

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

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

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

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

Thus,

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