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

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

Assumption *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 $\{Y_t : 1 \leq t \leq T\}$.

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*

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

[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 $\ldots = Y_{-1} = Y_0 = Y_1 = \ldots$ 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 $a_1, \ldots, a_n$,

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

1.3 Moving-Average 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*

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

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))

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

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

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

Claim 1.2 (ACF of MA(p))

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

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

where $\theta_0 = 1$.

Definition 1.8 (Infinite Moving-Average Process)

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

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

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

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

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. $\mathrm{Var}[S] = \ldots = \lim_{n\to\infty} \mathrm{Var}[\sum_{i=0}^{n} X_i]$
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,

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

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 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

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

More generally, consider an AR with a drift,

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

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

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

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 $X_0, X_1, \ldots$ 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, $\text{Cov}(Y_t, Y_{t-1}) = 1, \forall j$.

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".

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 $\Leftrightarrow$ Even and PSD

Proposition 1.2 (ACF $\Leftrightarrow$ 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 $\epsilon_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2)$ and if $(\epsilon_2, \epsilon_3, \ldots) \perp Y_1$, then $\hat{\phi}_{OLS}$ is the (conditional) MLE of ϕ.

The (conditional) MLE of (ϕ, σ^2) is

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

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

Definition 2.1 (Prediction-error Decomposition)

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

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

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

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

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

Therefore,

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

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, $\sqrt{n} \left(\hat{\beta}_{OLS} - \beta \right) \xrightarrow{d} N\left(0, \Omega\right)$, 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. $\sum_{i=0}^{\infty} i|\psi_i| < \infty$.

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

Note

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

2. $\epsilon_t \sim i.i.d.(0, \sigma^2) \Rightarrow \epsilon_t \sim \mathrm{WN}(0, \sigma^2)$ *(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**: $\bar{Y} \xrightarrow{P} \mu$ 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

   $\text{Var}(\bar{Y}) \leq \frac{1}{T}\sum_{j=-\infty}^{\infty} |\gamma(j)|$, 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.