# Time Series Analysis:

## Introduction to time series and forecasting

## Handout 1: Introduction

Laura Mayoral

IAE and Barcelona GSE

Barcelona, Winter 2018

# Outline

0. Introduction

1. Types of data in Econometrics. Time Series data.

2. How do macroeconomic data looks like?

3. Stochastic processes

4. The lag operator

5. The autocovariance function

6. Stationarity

7. Examples of stochastic processes

8. Ergodicity

9. Limit theorems for stochastic processes

10. Partial autocorrelations

# Introduction

## What does a Macroeconometrian do?

"*Macroeconometricians do four things: describe and summarize macro-economic data, make macroeconomic forecasts, quantify what we do or do not know about the true structure of the macroeconomy, and advise (and sometimes become) macroeconomic policymakers*" .

Stock and Watson, JEP (2001)

- Except advising policy makers, this is what we aim to do in this course.

# Three types of data in Econometrics

- *Cross-sectional data*: data collected by observing many subjects (such as individuals, firms, countries/regions, etc.) at the same point in time.

**Example**: investment in I+D of a group of firms

- *Time series data*: data collected for a single entity at multiple points in time.

**Example**: Yearly investment of country X.

- *Panel data*: data collected observing many individuals who are followed over time.

**Example**: Yearly investment of OECD countries.

- Time series data can answer questions for which cross-sectional data might be inadequate.

- What is the *dynamic* causal effect on a variable of interest, Y, of a change in another variable, X over time?

- What is the best forecast of the value of some variable at a future date?

- Examples:

  - What is the effect of a change of monetary policy in output and inflation, both initially and subsequently?

  - What is the best forecast of inflation and output for the next three terms?

# Time Series Data

- A time series is a set of observations

$$y_1, y_2, \ldots, y_t, \ldots, y_T,$$

where $t$ is the time index.

- Time series data come with a natural temporal ordering

- In cross-sectional analysis: observations $(y_i, x_i)$, $i = \{1, ..., N\}$ are randomly drawn from a fixed population. N observations from the same distribution. No ordering.

- Random sampling implies that observations from different units are independently distributed.

- In time series, we only have one observation from each variable (at each moment in time $t$). We don't have random sampling.

- Thus time series observations $(y_t, x_t)$, $t = \{1, ...T\}$ are in general **non-independent.**

$\longrightarrow$ Dependence among observations is a key feature in time series variables.

# How do economic time series look like?

- Time series are widely employed in empirical macroeconomics and in finance. Some examples are GNP, aggregate consumption, unemployment, population, exchange rates, inflation, stock returns...

Key features of economic time series are: trended behavior, seasonality and cyclical movements around trends.

In these links you can find many examples of economic time series data:

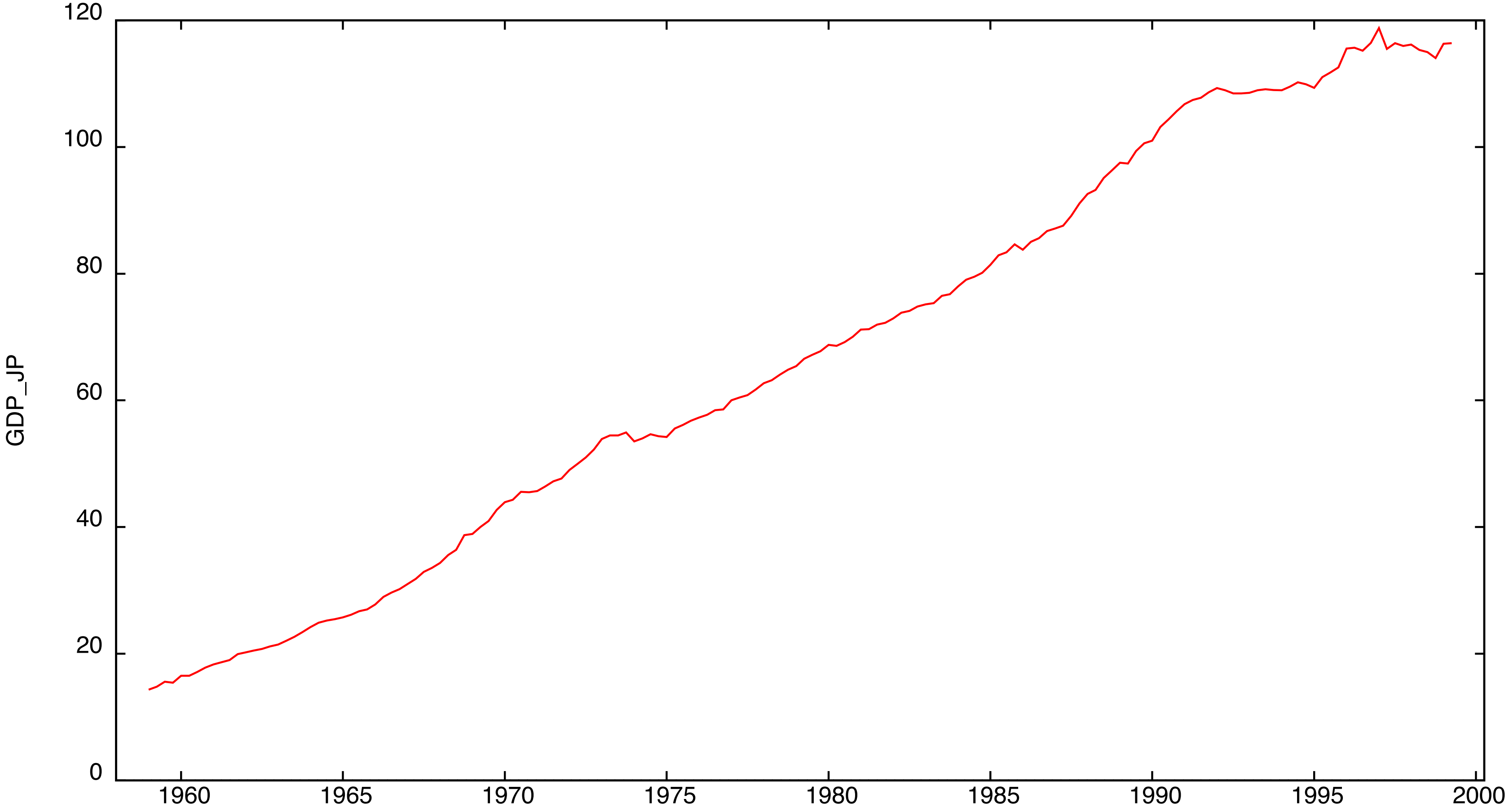
https://fred.stlouisfed.org

http://www.economagic.com

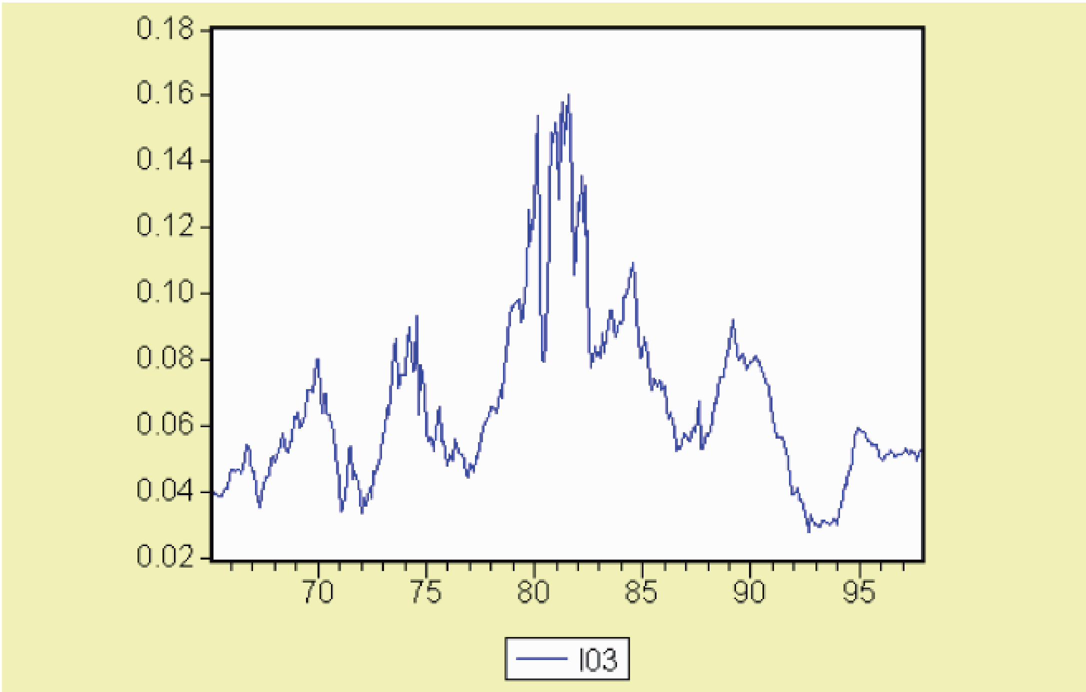
http://www.nber.org/links/data.html

# Examples

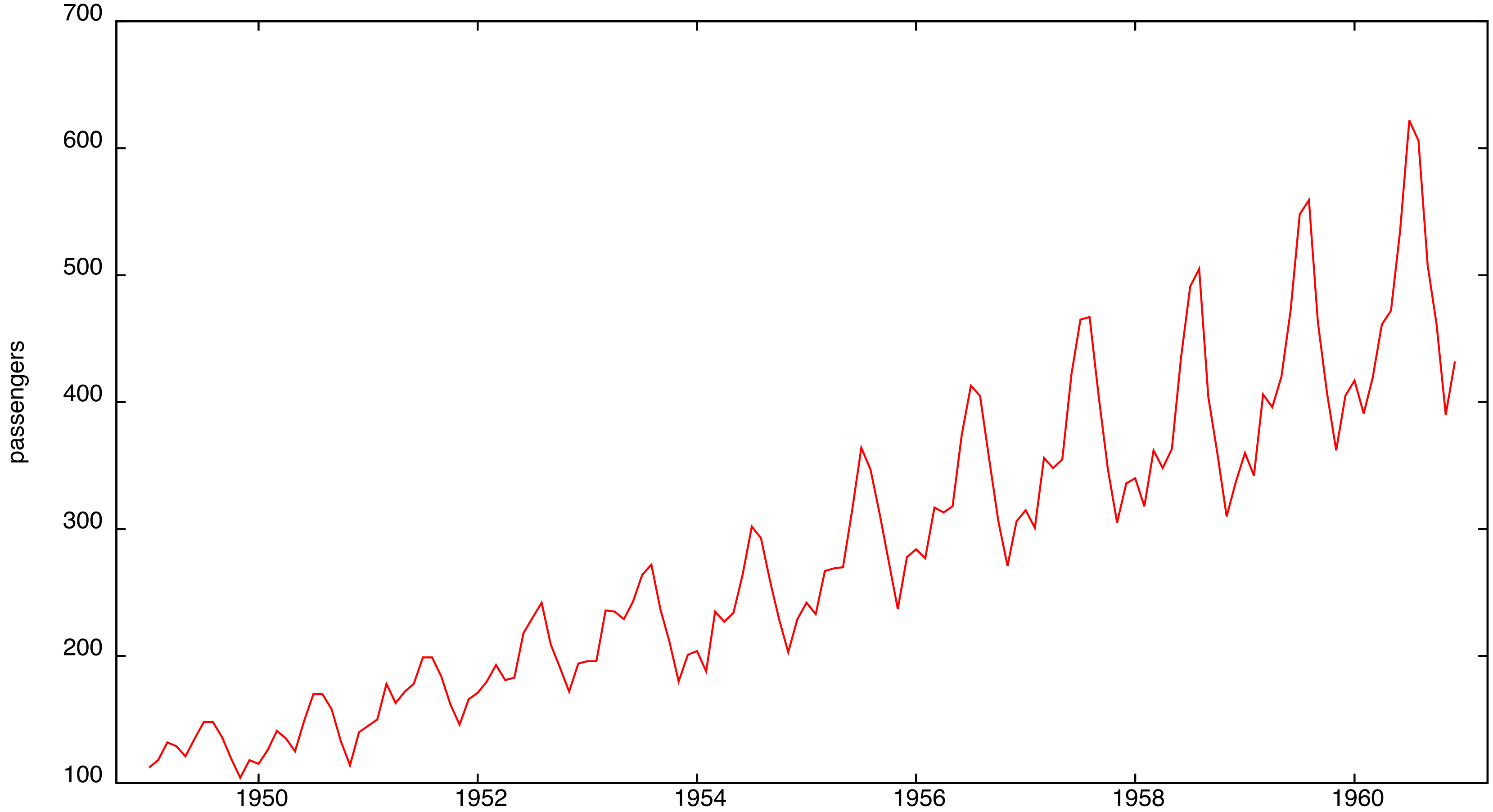
GDP Japan.



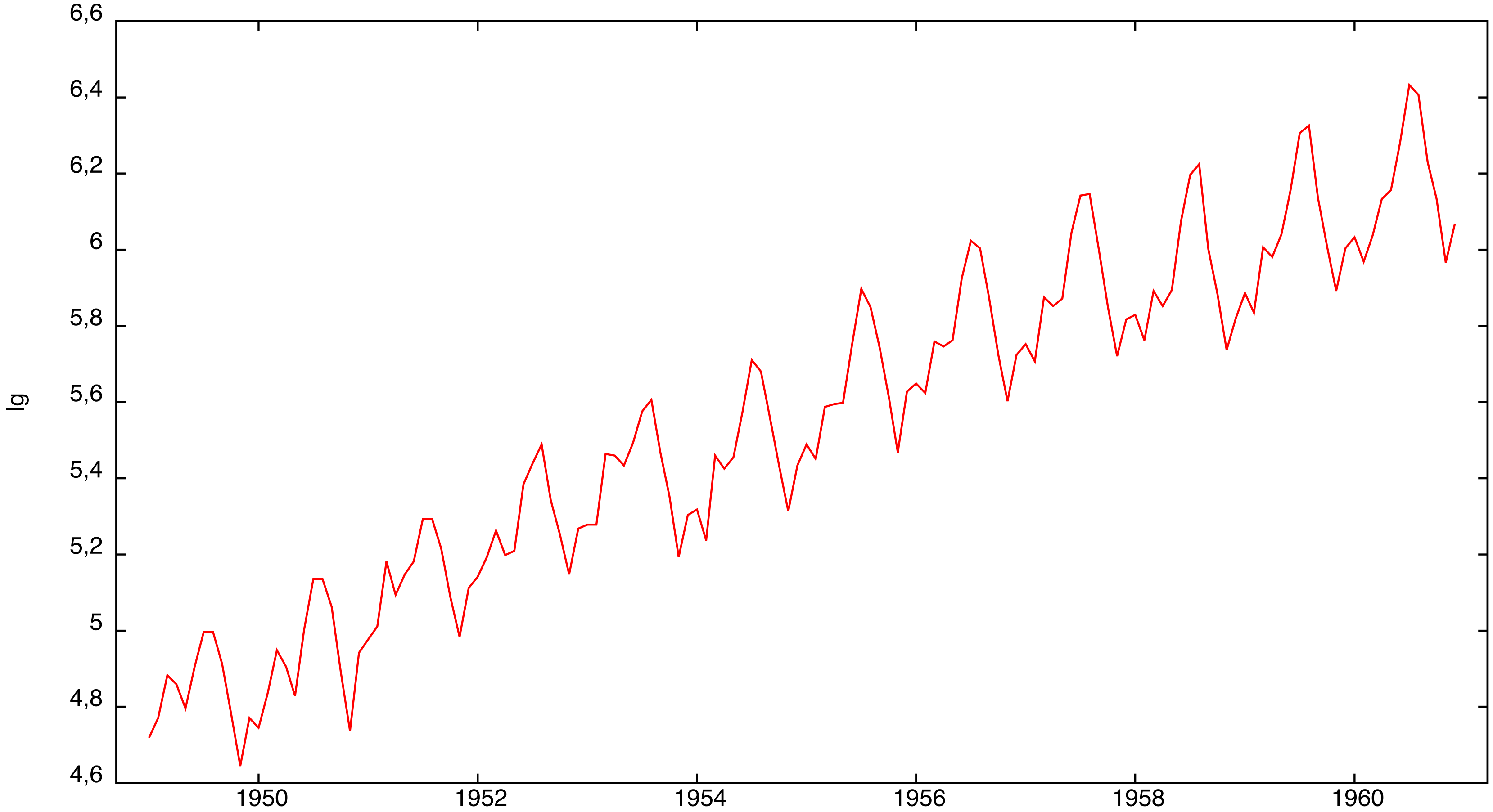
# Interest rate USA



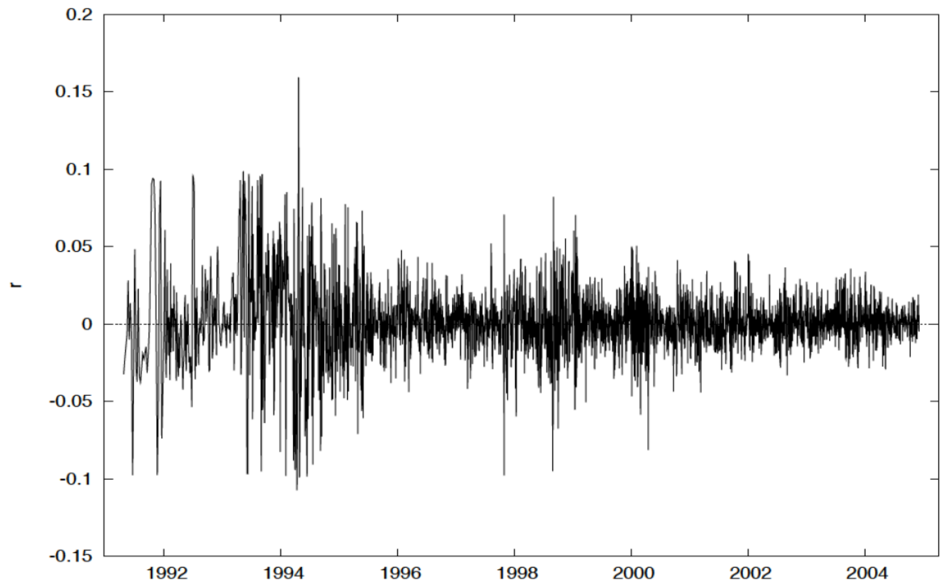
# Airline passengers



## Log airline passengers




## Stock returns


0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
r
1992
1994
1996
1998
2000
2002
2004

# Stochastic processes

■ Note: See the "Basic probability concepts" Section in Handout 0 (Preliminaries).

**Definition 1** *A stochastic process is a family of random variables* $\{X_t(\omega),\, t \in T,\, \omega \in \Omega\}$ *defined on a probability space* $\{\Omega, \mathcal{F},\, P\}$*.*

In particular, for a fixed $\omega$, $X_{.}(\omega)$ is a function from $T$ to $\mathbb{R}$. On the other hand, for fixed $t$, $X_t(.)$ is a function from $\Omega$ to $\mathbb{R}$.

- A stochastic process can be discrete or continuous according to whether $T$ is continuous, $e.g., T = \mathbb{R}$, or discrete, e.g., $T = \mathbb{Z}$.

In time series, the index $T$ is a set of time points, very often $T = \mathbb{Z}$, $T = \mathbb{N}$, the sets or integer and natural numbers, respectively.

# Realization of a Stochastic Process

**Definition 2** *(Realization of a Stochastic Process). The functions* $\{X_{\cdot}(\omega), \omega \in \Omega\}$ *on* $T$ *are known as the realizations or sample paths of the process* $\{X_t(\omega)\}$, $t \in T$, $\omega \in \Omega\}$.

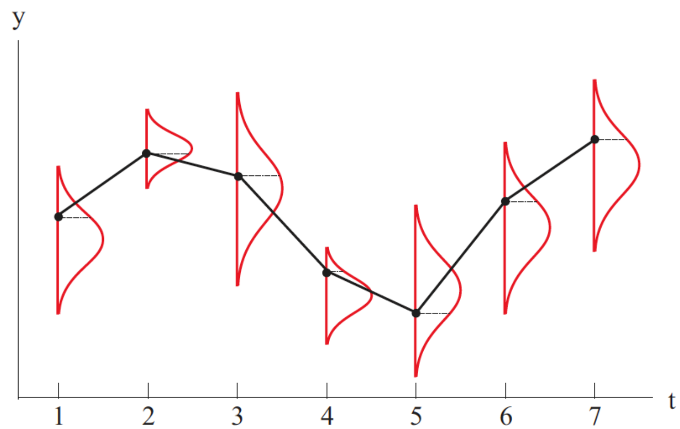
**Example 1** *Let the index set be T=*$\{$*1, 2, 3*$\}$ *and let the space of outcomes* ($\Omega$) *be the possible outcomes associated with tossing one dice:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

*Define* $Z_t(\omega)$= *t* + *[value on dice]*$^2$ *t. Therefore, for a particular* $\omega$*, say* $\omega$=$\{$*3*$\}$*, the realization or path would be (10, 20, 30), and this stochastic process has 6 different possible realizations (associated to each of the values of the dice).Remark: It is customary to use the term time series to denote both the observed data and the stochastic process of which it is a realization.*

- In time series: only one realization per random variable!

- Our goal is to draw inferences about the stochastic process based upon the realization we have observed.



# The Lag Operator

- The lag operator L maps a sequence $\{x_t\}$ into a sequence $\{y_t\}$ such that

$$y_t = Lx_t = x_{t-1}, \text{ for all } t.$$

- If we apply L repeatedly on a process, for instance $L\left(L\left(Lx_t\right)\right)$, we will use the convention

$$L\left(L\left(Lx_t\right)\right) = L^3 x_t = x_{t-3}.$$

- We can also form polynomials, $a_p\left(L\right) = 1 + a_1 L + a_2 L^2 + ... + a_p L^p$, such that

$$a_p\left(L\right) x_t = x_t + a_1 x_{t-1} + ... a_p x_{t-p}.$$

# Inversion

- L$^{-1}$ is the inverse of L, such that $L^{-1}(L)\,x_t = x_t$.

- Lag polynomials can also be inverted.

- For a polynomial $\phi_p(L)$, we are looking for the values of the coefficients $\alpha_i$ of $\alpha(L)$ such that
  - $\alpha(L) = \phi(L)^{-1} = 1 + \alpha_1 L + \alpha_2 L^2 + ...$, and
  - $\phi_p(L)\,\phi(L)^{-1} = 1$.

**Example 2** *Let p=1. Find the inverse of* $\phi_1(L) = (1 - \phi L)$.

*This amounts to finding the* $\alpha_i' s$ *that verify*

$$(1 - \phi L)\left(1 + \alpha_1 L + \alpha_2 L^2 + ...\right) = 1.$$

*Matching terms in* $L^j$*,it follows that*

$$\begin{aligned} -\phi + \alpha_1 &= 0 \Longrightarrow \alpha_1 = \phi, \\ -\phi\alpha_1 + \alpha_2 &= 0 \Longrightarrow \alpha_2 = \phi^2. \end{aligned}$$

*and*

$$(1 - \phi L)^{-1} = \left(1 + \sum_{j=1}^{\infty} \phi^j L^j\right), \text{ provided } |\phi| < 1.$$

*It is easy to check that* $\left(1+\sum_{j=1}^{\infty}\phi^j L^j\right)$ *is the inverse of* $(1-\phi L)$ *since:*

$$(1-\phi L)\left(1+\sum_{j=1}^{k}\phi^j L^j\right) = 1-\phi^{k+1}L^{k+1} \rightarrow 1 \text{ as } k \rightarrow \infty$$

**Example 3** *Let p=2. Find the inverse of* $\phi_1(L) = (1-\phi L) - \phi_2 L^2)$.

*If* $p > 1$, *we can invert the polynomial* $\phi_p(L)$ *by first factoring it and, then, use the formula for p=1. For example, let* $\lambda_1$ *and* $\lambda_2$ *be the roots of* $\phi_2(L)$. *Then,*

$$1+\phi_1 L+\phi_2 L^2 = (1-\lambda_1 L)(1-\lambda_2 L)$$

*Provided* $|\lambda_1|\,,\ |\lambda_2| < 1,$

$$\begin{aligned}
\left(1+\phi_1 L+\phi_2 L^2\right)^{-1} &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1} \\
&= \left(\sum_{j=0}^{\infty} \lambda_1^j L^j\right)\left(\sum_{j=0}^{\infty} \lambda_2^j L^j\right) \\
& \left(\sum_{j=0}^{\infty} L^j \left(\sum_{k=0}^{j} \lambda_1^k \lambda_2^{j-k}\right)\right)
\end{aligned}$$

# The autocovariance funtion

- This function

- is a measure of linear dependence between elements of the sequence $\{X_t,\ t \in \mathbb{Z}\}$

- extends the concept of covariance matrix (computed with a finite number of random variables) to the case where there is an infinite collection of random variables.

**Definition 3** *The autocovariance function. If $\{X_t,\ t \in T\}$ is a process such that $Var(X_t) < \infty$ for each $t \in T$, then the autocovariance function $\gamma_X(.,.)$ of $X_t$ is defined by*

$$\begin{aligned}\gamma_X(r,s) &= Cov(X_r, X_s)\\ &= E[(X_r - E(X_r))(X_s - E(X_s))],\ \ r,s \in T.\end{aligned}$$

# Weak Stationarity and Strict stationarity

Stationarity is a crucial concept. There are two basic definitions of stationarity: strict and weak (or second-order) stationarity.

## Strict stationarity

**Definition 4** *(Finite Dimensional distributions). Let* $\mathcal{T}$ *be the set of all vectors* $\{t = (t_1, ..., t_n)' \in T^n : t_1 < t_2 ... < t_n,$ *n=1, 2,...*$\}$*. Then the finite-dimensional distribution functions of* $\{X_t,\ t \in T\}$ *are the functions* $\{F_t(.)\,, t \in \mathcal{T}\}$ *defined for* $t = (t_1, ..., t_n)'$ *by*

$$F_t(x) = P(X_{t_1} \leq x_1, ..., X_{t_n} \leq x_n),\ \ x = (x_1, ..., x_n)' \in \mathbb{R}^n$$

**Definition 5** *(First, Second and $n$-th order stationary). The time series $\{X_t,\ t \in \mathbb{Z}\}$ is said to be first-order, second-order and n-th order stationary, respectively if*

$$F_t\left(x_{t_1}\right) = F_t\left(x_{t_1+h}\right), \text{ for any } t_1, h;$$

$$F_t\left(x_{t_1}, x_{t_2}\right) = F_t\left(x_{t_1+h}, x_{t_2+h}\right), \text{ for any } t_1,\ t_2,\ h;$$

$$\begin{aligned} F_t\left(x_{t_1}, x_{t_2}, ..., x_{t_n}\right) &= F_t\left(x_{t_1+h}, x_{t_2+h}, ..., x_{t_n+h}\right), \\ &\quad \text{for any } t_1,\ t_2, ..., t_h,\ h; \end{aligned}$$

**Definition 6** *(Strict Stationarity) The time series $\{X_t,\ t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, ....X_{t_k})'$ and $(X_{t_{1+h}}, ....X_{t_{k+h}})'$ are the same for all positive integers $k$ and for all $t_1, ..., t_k,\ h \in \mathbb{Z}$. In other words, $\{X_t,\ t \in \mathbb{Z}\}$ is strictly stationary if it is n-order stationary for any n.*

- Interpretation:

  - This means that the graphs over two equal-length time intervals of a realization of the time series should exhibit similar statistical characteristics.

- Joint finite dimensional distributions are difficult to work with. The following concept introduces a notion of stationarity that can be characterized by only looking at first and second moments.

# Weak Stationarity

**Definition 7** *(Weak Stationarity) The time series $\{X_t,\ t \in \mathbb{Z}\}$ is said to be weakly stationary if*

*i)* $E\left|X_t^2\right| < \infty$ *for all* $t \in \mathbb{Z}$

*ii)* $E(X_t) = m$ *for all* $t$

*iii)* $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ *for all* $r,\ s,\ t \in \mathbb{Z}$.

- This concept of stationarity is usually referred in the literature as second-order stationarity, weak stationarity or covariance stationarity.

- Notice that stationarity requires also the variance of $X_t$ to be constant. If $X_t$ is stationary, then

$$
\begin{aligned}
Var(X_r) &= \gamma_X(r,r) = \gamma_X(r+t, r+t) = Var(X_{r+t}), \\
\text{for all } r,\ t &\in \mathbb{Z}.
\end{aligned}
$$

- Stationarity basically means that the mean, the variance are finite and constant and that the autocovariance function only depends on $h$, **the distance between observations.**

# Stationarity and the autocovariance funtion

- If $\{X_t, t \in \mathbb{Z}\}$ is stationary, then $\gamma_X(r,s) = \gamma_X(r-s,0)$ for all $r$, $s \in \mathbb{Z}$. Then, for stationary processes one can define the autocovariance as a function of only one parameter, that is

$$\gamma_X(h) = Cov(X_{t+h}, X_t) \text{ for all } t, h \in \mathbb{Z}.$$

- The function $\gamma_X(.)$ will be referred to as the **autocovariance function** of the process $\{X_t\}$ and $\gamma_X(h)$ is the value of this function at lag $h$.

- Notation: $\gamma_X(h)$ or simply $\gamma_h$ will denote the $h$-th autocovariance of $X_t$.

- If $\gamma(.)$ is the autocovariance function of a stationary process, then it verifies

$$
\begin{aligned}
i)\ \gamma(0) &\geq 0 \\
ii)\ |\gamma(h)| &\leq \gamma(0) \text{ for all } h \in \mathbb{Z} \\
iii)\ \gamma(-h) &= \gamma(h) \text{ for all } h \in \mathbb{Z}
\end{aligned}
$$

# The autocorrelation function

**Definition 8** *(Autocorrelation function, ACF) For a stationary process $\{X_t\}$, the autocorrelation function at lag $h$ is defined as*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_{t+h}, X_t), \text{ for all } t, h \in \mathbb{Z}.$$

# The relation between Stationary and Strict Stationarity

- Strict stationarity implies weak stationarity, provided the first and second moments of the variables exist, but the converse of this statement is not true in general.

- Taking $k = 1$ in Definition 11, it is clear that all the variables have the same distribution, which implies that the mean and the variance are the same for all variables (provided they exist).

- Taking $k = 2$ in Definition 11, it follows that Cov$(X_{t_1},\ X_{t_2}) = Cov(X_{t_1+h}, X_{t_2+h})$ for all $t_1$, $t_2$, $h \in \mathbb{Z}$, implying condition iii) in the stationarity definition.

- However, it is easy to find counterexamples where a process is stationary but not strictly stationary.

**Example 4** *Consider a sequence of independent random variables such that if $t < T_1$, $X_t$ follows an exponential distribution with mean and variance equal to 1 and if $t \geq T_1$, $X_t$ is normally distributed with mean and variance equal to 1. $\{X_t\}$ is stationary but it is not strictly stationary because $X_t$ and $X_{t^*}$ have different distributions if $t < T_1$ and $t^* \geq T_1$.*

## The relation between Stationary and Strict Stationarity, II

There is an one important case where both concepts of stationary are equivalent.

**Definition 9** *(Gaussian Time series) The process* $\{X_t\}$ *is a Gaussian time series if and only if the distribution functions of* $\{X_t\}$ *are all multivariate normal.*

If $\{X_t,\ t \in \mathbb{Z}\}$ is a stationary Gaussian time series, then it is also strictly stationary, since for all $n \in \{1, 2, ...\}$ and for all $h$, $t_1$ , $t_2, ...$ $\in \mathbb{Z}$, the random vectors $(X_{t_1}, ..., X_{t_n})'$, and $(X_{t_1+h}, ..., X_{t_n+h})'$ have the same mean, and covariance matrix, and hence they have the same distribution.

# Some examples of stationary processes

**Example 5** *iid sequences.*

■ *The sequence* $\{\varepsilon_t\}$ *is* $i.i.d$ *(independent and identically distributed) if all the variables are independent and share the same univariate distribution.*

■ Clearly, an $iid$ sequence is strictly stationary and provided the first and second order moments exist, it is also weak stationary.

**Example 6** *White noise process.*

■ *The process* $\{\varepsilon_t\}$ *is called white noise if it is weakly stationary with* $E(\varepsilon_t) = 0$ *and autovariance function*

$$\gamma_\varepsilon(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

- The white noise process is important because is used as a building block for more general processes, as can be seen in some of the examples below.

- An $i.i.d$ sequence with zero mean and variance $\sigma^2$ is also white noise. The converse is not true in general. Furthermore, a white noise process might not be strictly stationary.

**Example 7** *Martingale difference sequence, m.d.s.*

- *A process* $\{\varepsilon_t\}$, *with* $E(\varepsilon_t) = 0$ *is called a martingale difference sequence if*

$$E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, ...) = 0, \text{ for } t \geq 2.$$

- It can be shown that if $E(\varepsilon_t) = 0$ and the second order moments exist then,

$$\varepsilon_t \text{ is } i.i.d. \Rightarrow \varepsilon_t \text{ is m.d.s} \Rightarrow \varepsilon_t \text{ is white noise}$$

but the converse implication is not true in general.

**Example 8** *Moving average of order one.*

■ *The process $\{X_t\}$ is called a moving average of order 1, or MA(1), if $\{X_t\}$ is defined as*

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1},$$

where $\{\varepsilon_t\}$ is a white noise process and is stationary for any value of $\theta$.

**Example 9** *Autoregression of order 1.*

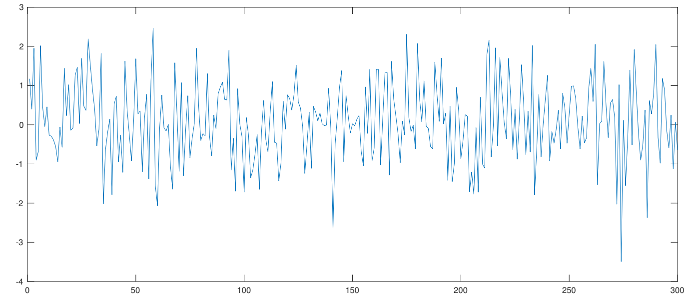
■ *The process $\{X_t\}$ is called an autoregressive process of order 1 if $\{X_t\}$ is defined as*

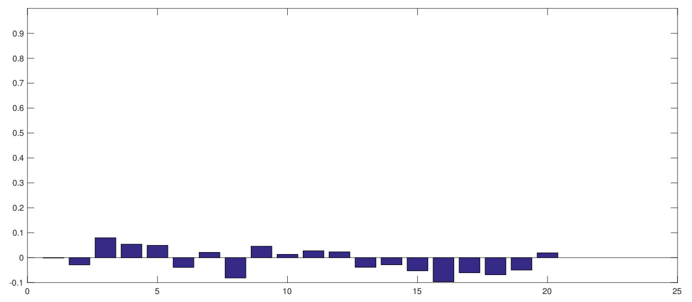
$$X_t = \phi X_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a white noise process. $X_t$ is stationary provided $|\phi| < 1$.

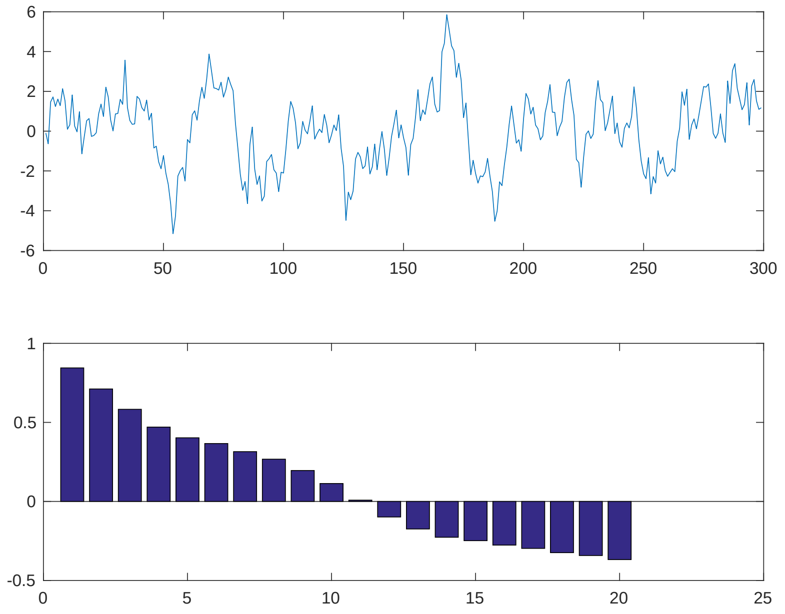
# Some graphs

- The graphs below correspond to simulated data.
- IID process: plot and autocorrelation function

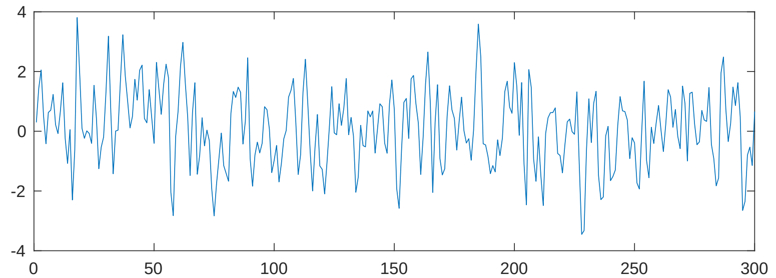


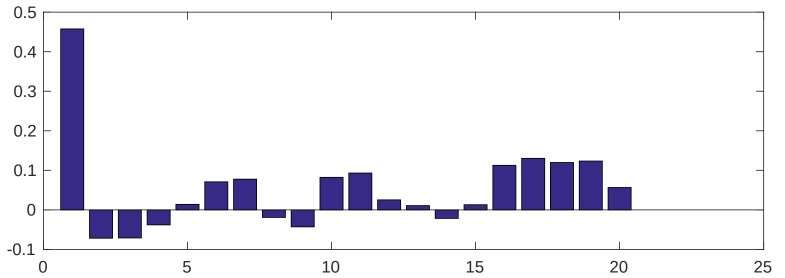



- AR(1) process, $\phi = 0.8$: plot and autocorrelation function



- MA(1) process, $\theta = 0.8$: plot and autocorrelation function



# Some examples of non-stationary processes

**Example 10** *A trended process.*

$$X_t = \beta t + \varepsilon_t,$$

*where* $t = 1, ...., T$ *is a deterministic time trend.*

**Example 11** *A random walk process*

$$X_t = X_{t-1} + \varepsilon_t, \ t \geq 0$$

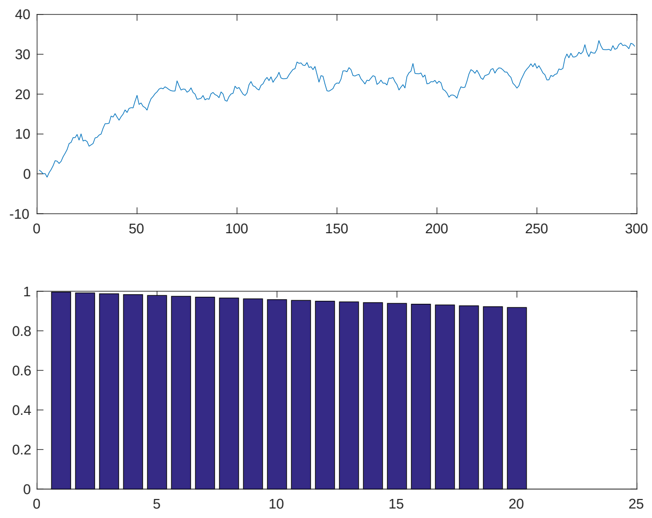
**Example 12** *A process with a break*

$$\begin{aligned} X_t &= \varepsilon_t, t < k \\ X_t &= \mu + \varepsilon_t, \ t \geq k \end{aligned}$$

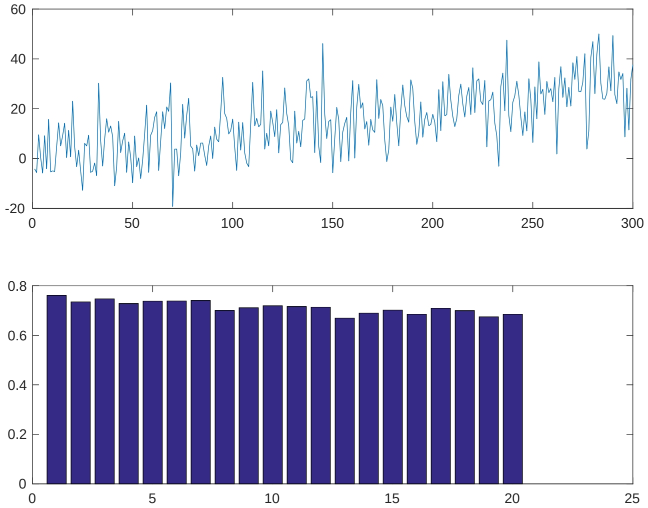
*where* $\mu \neq 0$.

# And some more graphs

- Random walk: plot and autocorrelation funciton



- Trend-stationary process: plot and autocorrelation funciton



# Ergodicity

- Note: Please review the section "Asymptotic theory" in Handout 0 to refresh the definition of convergence of random variables, limit theorems, etc.

- Ergodicity is a condition that restricts the memory of the process.

- A loose definition of ergodicity is that the process is asymptotically independent. That is, for sufficiently large n, $X_i$ and $X_{i+n}$ are nearly independent. A more formal definition is provided below. All of these definitions essentially say that the effect of current events eventually disappears.

- Stationary and ergodic processes are temporally dependent but with sufficiently weak memory for learning to take place as new observations are revealed.

- Consider a stationary stochastic proces $\{X_t\}$, with $E(X_t) = \mu$ for all $t$. Assume that we are interested in estimating $\mu$.

- The standard approach for estimating the mean of a single random variable consists of computing its sample mean

$$\bar{X} = N^{-1} \sum_{i=1}^{N} X_t^{(i)}, \; i = 1, ...N, \tag{1}$$

where $X_t^{(i)}$ are different realizations of the variable $X_t$. We will referred to expression $(1)$ as the 'ensemble average'.

- When working in a laboratory (or when dealing with cross sectional data), one could generate different observations for the variable $X_t$ under identical conditions.

- When analyzing economic variables over time, we can only observe a unique realization of each of the random variable $X_t$ so that it is not possible to estimate $\mu$ by computing the ensemble average $(1)$.

- However, we can compute a time average

$$\overline{X} = T^{-1} \sum_{t=1}^{T} X_t \tag{2}$$

Whether time averages such as $(2)$ converge to the same limit as the ensemble average, $\mathsf{E}(X_t)$, has to do with the concept of Ergodicity.

**Definition 10** *(Ergodicity for the mean) A covariance stationary process* $X_t$ *is said to be ergodic for the mean if* $(2)$ *converges to* *E*$(X_t)$ .

**Definition 11** *(Ergodicity for the second moments) A covariance stationary process is said to be ergodic for the second moments if*

$$\frac{1}{(T-j)}\sum_{t=j+1}(Y_t-\mu)(Y_{t-j}-\mu)\xrightarrow{p}\gamma_j, \text{ for all } j.$$

In many applications, ergodicity and stationarity turn out to amount to the same requirements. However, we present now an example of a stationary process that is not ergodic.

**Example 13** *Consider the process $\{Y_t\}$, whose $i-th$ realization is given by*

$$Y_t^{(i)} = \mu^{(i)} + \varepsilon_t$$

*where $\mu$ is a realization of a random variable $N\left(0, \sigma_\mu^2\right)$ and $\varepsilon_t$ is a Gaussian white noise process, with distribution $N\left(0, \sigma_\varepsilon^2\right)$ and $\mu$ and $\varepsilon$ are independent random variables. $\{Y_t\}$ is stationary since*

*i)* $E(Y_t) = E\left(\mu + \varepsilon_t\right) = 0$

*ii)* $Var(Y_t) = Var(\mu + \varepsilon_t) = \sigma_\mu^2 + \sigma_\varepsilon^2$

*iii)* $\gamma_{jt} = E\left(\mu + \varepsilon_t\right)\left(\mu + \varepsilon_{t-j}\right) = E\left(\mu^2\right) = \sigma_\mu^2.$

*However, $\{Y_t^{(i)}\}$ is not ergodic.*

$$\frac{1}{T}\sum_{t=1}^{T} Y_t^{(i)} = \frac{1}{T}\sum_{t=1}^{T}(\mu^{(i)} + \varepsilon_t) = \mu^{(i)} + \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t) \xrightarrow{p} \mu^{(i)}.$$

## Sufficient conditions for ergodicity:

- if $\gamma_n \to 0$, then $\{X_t\}$ is ergodic for the mean.
- If $\sum_{j=0}^{\infty} \left|\gamma_j\right| < \infty$, then $\{X_t\}$ is ergodic for the second moments.
- Furthermore, if $\{\mathsf{X}_t\}$ is a stationary Gaussian process and $\sum_{j=0}^{\infty} \left|\gamma_j\right| < \infty$, then the process is ergodic of all moments.

# Limit Theorems

- The Law of Large Numbers and the Central Limit Theorem are the most important results for computing the limits of sequences of random variables.

- The basic statements of these theorems are only applicable to $i.i.d.$ sequences. However, both can be restated in a more general way, as will be shown below.

**Theorem 1** *(Weak law of large numbers for $iid$ sequences) If $\{X_t\}$ is an $i.i.d$ sequence of random variables with finite mean $\mu$ then*

$$\bar{X}_T = T^{-1} \sum_{t=1}^{T} X_t \xrightarrow{p} \mu$$

A very simple proof of this result can be provided if we further assume that $\text{var}(X_i) = \sigma^2 < \infty$. Then, by Chebychev's inequality:

$$
\begin{aligned}
P\left(\left|T^{-1}\sum_{t=1}^{T}X_t-\mu\right|>\varepsilon)\right) &\leq var(T^{-1}\sum_{t=1}^{T}X_t)/\varepsilon^2 \\
&= T^{-2}\sum_{t=1}^{T}var(X_t)/\varepsilon^2 \\
&= \frac{T\sigma^2}{T^2\varepsilon^2}\to 0.
\end{aligned}
$$

The $i.i.d$ can be easily relaxed provided $X_t$ is weakly stationary and ergodic for the mean, as we will see below.

**Theorem 2** *(Central limit theorem for $i.i.d.$ sequences) If $\{X_t\}$ is a sequence of $iid(\mu,\, \sigma^2)$ random variables then*

$$\sqrt{T}(\bar{X}_T - \mu)/\sigma \xrightarrow{d} N(0,1).$$

A more general version of this theorem can be stated as follows.

**Theorem 3** *(Central limit theorem for martingale difference sequences) Let $\{X_t\}$ be a martingale difference sequence. If $a)$ $E\left(X_t^2\right) = \sigma_t^2 > 0$ with $T^{-1}\sum_{t=1}^{T}\sigma_t^2 \to \sigma^2$, b) E($|X_t|)^r < \infty$ for some $r$ ¿2 and all $t$ and c) T$^{-1}\sum X_t^2 \xrightarrow{p} \sigma^2$, then $\sqrt{T}\bar{X}_T \xrightarrow{d} N\left(0, \sigma^2\right)$.*

# Moments of a stationary process

- An we have seen above, second-order stationarity is characterized by the first and the second moments (i.e, mean, variance and autocovariance functon).

- Thus, estimation of these moments from a sample of data $\{x_1$, $x_2$, ..., $x_T\}$ of an stationary time series $\{X_t\}$ plays a crucial role.

- we now present some estimators for these moments and examine some of their properties.

# The sample mean of a stationary process

The sample mean is the natural estimator for the expected value of a process.

**Definition 12** *Let* $\{X_t\}$ *be a process and* $\{x_1, x_2, ..., x_T\}$ *the observed realization of this process. The sample mean is defined as*

$$\bar{X}_T = T^{-1} \sum_{t=1}^{T} x_t.$$

**Theorem 4** *If* $\{X_t\}$ *is stationary with mean* $\mu$ *and autocovariance function* $\gamma(.)$, *then*

*i)* $E(\bar{X}_T) = \mu$

*ii)* $If\ \gamma(T) \to 0$ *as* $T \to \infty$, *Var*$(\bar{X}_T) = E(\bar{X}_T - \mu)^2 \to 0.$

*iii) If* $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, $TE(\bar{X}_T - \mu)^2 \to \sum_{h=-\infty}^{\infty} \gamma(h).$

*Proof. See Brockell and Davis, (1991), p. 219.*

- Notice that $i)$ and $ii)$ imply that $\bar{X}_T$ converges in mean square to $\mu$.

- thus, the sample mean is **consistent** provided $\gamma(T) \to 0$ (remember that this is the ergodicity condition).

- Then, this theorem presents **a weak Law of Large Numbers for stationary and ergodic processes.**

- Stationarity alone is not enough. Consider this example: V$\sim N\left(0, \sigma^2\right)$. Suppose X$_t = V$ for all $t$. Then, X$_t$ is stationary and cov(X$_t, X_s) = \sigma^2$ for all t, s. Notice that covariances are not summable and that a LLN does not hold. Why? This example is equivalent of having a single observation of a random variable.

**Theorem 5** *(Central limit theorem for dependent processes) Let* $\{X_t\}$ *be a stationary sequence given by* $X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *where* $\{\varepsilon_t\}$ *is an iid(0,*$\sigma^2$*) sequence of random variables and* $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty$ *and* $\sum_{j=0}^{\infty} \psi_j \neq 0$ *then*

$$\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right).$$

Proof. See Brockell and Davis, (1991), Section 7.3.

The limit of $TE(\bar{X}_T - \mu)^2$, $\sum_{h=-\infty}^{\infty} \gamma(n)$, is called the **long run variance of** $\bar{X}_T$. Later on in the course we will discuss how to estimate this quantity.

## Sample autocovariance and sample autocorrelation

**Definition 13** *The sample autocovariance of* $\{x_1, x_2, ..., x_T\}$ *is defined as*

$$\hat{\gamma}(h) = T^{-1} \sum_{j=1}^{T-h} \left(x_{j+h} - \bar{X}_T\right)\left(x_j - \bar{X}_T\right), \ 0 \leq h \leq T, \qquad (3)$$

*where* $\bar{X}_T = T^{-1} \sum_{j=1}^{T} x_j$.

**Definition 14** *The sample autocorrelation of* $\{x_1, x_2, ..., x_T\}$ *is defined as*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \ 0 \leq h \leq T. \qquad (4)$$

- The estimators in $(3)$ and in $(4)$ are biased but under certain conditions they are asymptotically unbiased.

- A remark:

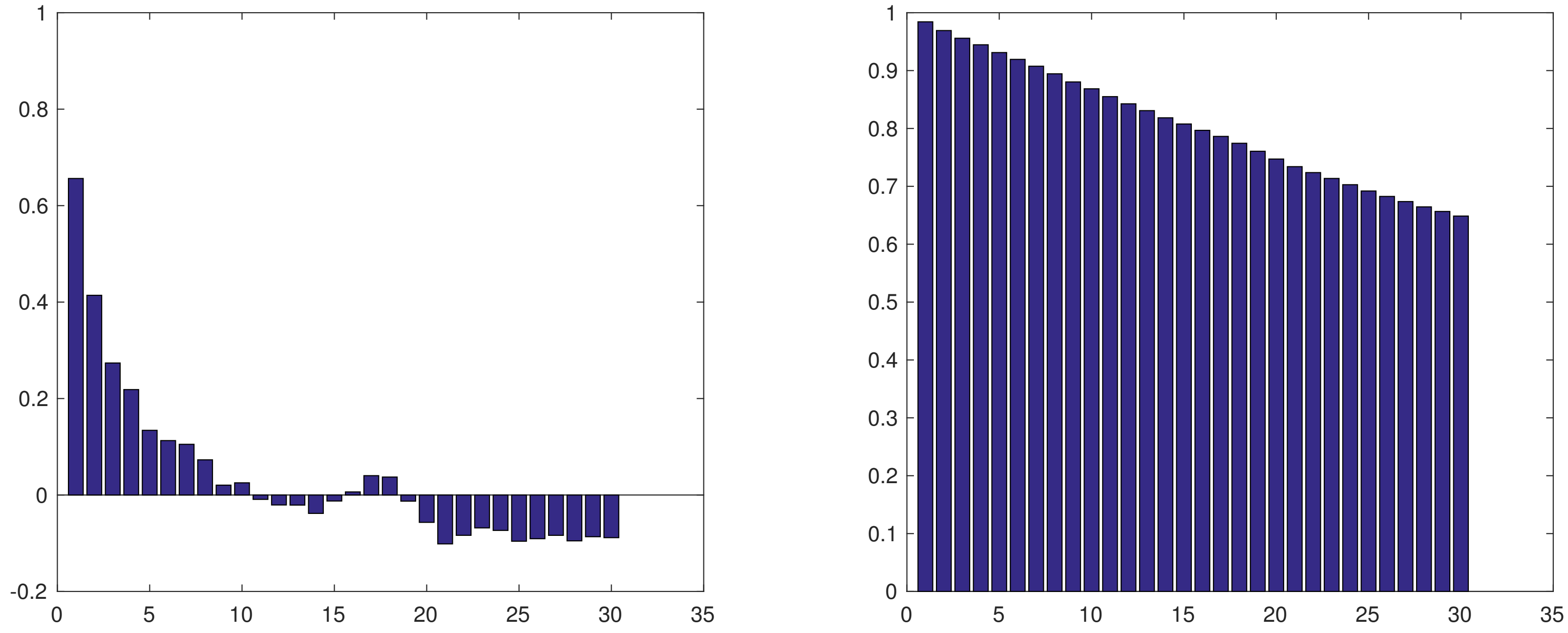
- The sample autocovariance and sample autocorrelation can be computed for any data set, and are not restricted to realizations of a stationary process.

- For stationary processes both functions will show a rapid decay towards zero as $h$ increases.

- However, for non-stationary data, these functions will exhibit quite different behavior. For data containing a trend, $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ will exhibit very slow decay as $h$ increases and for data with a substantial cyclical component, $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ will exhibit a similar cyclical behavior with the same periodicity. Thus, $\hat{\gamma}(.)$, $\hat{\rho}(.)$ are often employed as identification tools.

# Some examples

- The following graphs present the sample autocorrelation function of a stationary AR(1) process, $X_t = 0.7X_{t-1} + \varepsilon_t$ (left side) and that of a non-stationary random walk process (right side), $Y_t = Y_{t-1} + \varepsilon_t$.



# Asymptotic properties of the sample autocorrelation function

**Theorem 6** *Let* $\{X_t\}$ *be a stationary process given by* $X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *where* $\{\varepsilon_t\}$ *is an iid(0,*$\sigma^2$) *sequence of random variables and* $\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty$ *and E*$\left(\varepsilon_t^4\right) < \infty$, *then*

$$\sqrt{T}(\hat{\rho}(h) - \rho(h)) \xrightarrow{d} N(0, W),$$

*where* $\rho(h)' = \begin{pmatrix} \rho(1) & \rho(2) & \ldots & \rho(h) \end{pmatrix}$, $\hat{\rho}(h)' = \begin{pmatrix} \hat{\rho}(1) & \hat{\rho}(2) & \ldots & \hat{\rho}(h) \end{pmatrix}$ *and W is the variance-covariance matrix whose (i,j) element is given by Bartlett's formula:*

$$\begin{aligned} w_{ij} &= \sum_{k=-\infty}^{\infty} \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) \\ &\quad -2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i). \end{aligned} \tag{5}$$

*Proof. See Brockell and Davis, 1991, Section 7.3.*

■ An interesting particular case is when the process $\{X_t\}$ is an iid sequence. Then $\rho(h) = 0$ for all $h \neq 0$. In this case, the variance-covariance matrix W simplifies notably and from $(5)$ it is obtained that

$$w_{ij} = \begin{cases} 1 & if \;\; i = j \\ 0 & otherwise \end{cases}$$

Thus if we plot the sample autocorrelation function as a function of $h$, approximately 95% of the correlations should lie between the bounds $\pm 1.96\sqrt{T}$.

# Partial autocorrelation function, PACF (conditional correlation)

- This function gives the correlation between two random variables, $X_t$ and $X_{t+k}$, after their mutual linear dependency on the intervening variables $X_{t+1}$, ..., $X_{t+k-1}$ has been removed.

It is the following partial correlation:

$$\alpha_k^{(k)} = corr(X_t, X_{t+k} | X_{t+1}, ..., X_{t+k-1})$$

**Definition 15** *The $k$-th population partial autocorrelation (denoted as $\alpha_k^{(k)}$) of a stationary process is defined by*

$$\alpha_k^{(1)} = corr(X_{t+1}, X_t),$$

*and*

$$\alpha_k^{(k)} = corr(X_{t+k} - \breve{X}_{t+k}, X_t - \breve{X}_t), \text{ for } k > 1$$

*where $\breve{X}_i$, i={t+k,t} denote the linear projections of $X_{t+k}$ and $X_t$ on $\{1, X_{t+1}, ..., X_{t+k-1}\}$, respectively.*

# PACF: computation

- To compute the $m-th$ partial correlation, one simply has to run an OLS regression including the most recent $m$ values of the variable.

- The last coefficient would be the $mth-$ autocorrelation, that is,

$$X_t = c + \hat{\alpha}_1^{(m)} X_{t-1} + ... + \hat{\alpha}_m^{(m)} X_{t-m} + \hat{e}_t,$$

where $\hat{e}_t$ denotes OLS residuals. The estimated coefficient $\hat{\alpha}_m^{(m)}$ is the $m-th$ partial autocorrelation.

# References

Hamilton: Chapter 1, 7.

Brockwell, P. J., and A. Davis (1991): Chapter 1, 6.

Stock and Watson, Chapter 14.