Introduction to Time Series Analysis:

Laura Mayoral[1]

[1]Institute for Economic Analysis and Barcelona School of Economics

DG-ECFIN, Nov 2021

WELCOME!

Overview of the course

- This course will provide an introduction to the theory and the application of time series methods in econometrics
- Goal: to develop the skills needed to do empirical analysis that requires the use of time series data.
- Website: http://mayoral.iae-csic.org/timeseries2021/timeseries_2021.htm
- Books: main reference is "Time Series Analysis", by J. Hamilton (1994). See the Syllabus/website for additional references.
- Contact me by email: laura.mayoral@iae.csic.es

I. Introduction

1. Types of data used in econometrics
2. Why is time series analysis useful?
3. Why do we need a course on time series analysis?
4. How do time series data look like?
5. Structure of the course

1. Types of Data

- Three types of data in econometrics:

1. 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

2. Time series data: data collected for a single entity at multiple points in time.
   Example: Monthly consumption in country X.

3. Panel data: data collected observing many individuals who are followed over time.
   Example: Monthly consumption of OECD countries.
   $\longrightarrow$ Macro and micro panels: large/small T (number of periods) in relation to N (size of the cross-section)

2. Why is Time Series Analysis Useful?

- Time series data can answer questions for which cross-sectional data might be inadequate.
  - What is the dynamic effect on a variable of interest, Y, of a change in another variable, X over time?
    - For instance: What is the effect of a change of monetary policy in output and inflation, both initially and subsequently?
  - What is the best forecast of the value of some variable at a future date?
    - For instance: What is the best forecast of inflation and output for the next three terms?

3. Why is a course on Time Series needed?

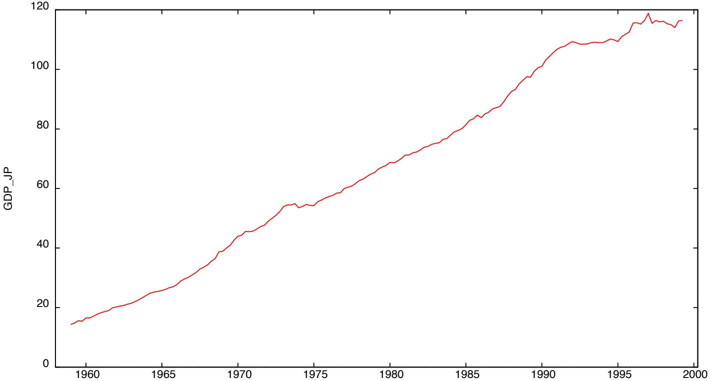
- Key assumption in the analysis of cross-sectional data: Random Sampling
- Random sampling implies that
  - (1) there's a population of interest and (2) an I.I.D. (independent and identically distributed) sample can be drawn from that population.
  - Sample: N independent obs., same distribution. No ordering.

3. Why is a course on Time Series needed?, II

- Time Series data: random sampling doesn't make sense anymore
  - Why? serial correlation (≈ dependence over time). GDP_t cannot be assumed to the independent to GDP_{t+1}
  - Observations can be drawn from different distributions: GDP_t and GDP_{t+1} might have different distributions (there's growth!).
- Thus: serial correlation is a key feature in time series data.
- This has important implications on (among others) (1) the modeling of the data, (2) the appropriate estimation techniques and (3) properties of the estimators.

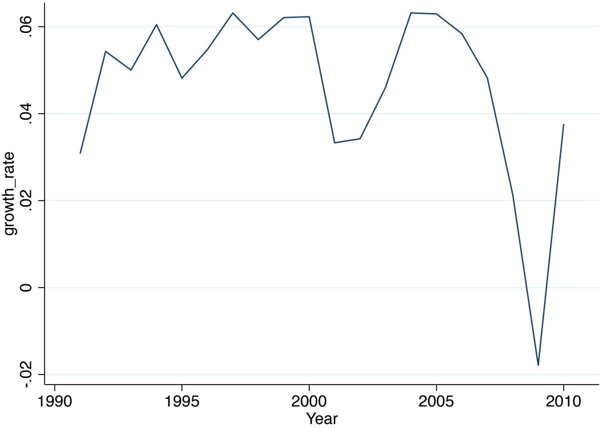
4. How Does Time Series Data look like?

Some Examples:
1) GDP Japan (yearly data)



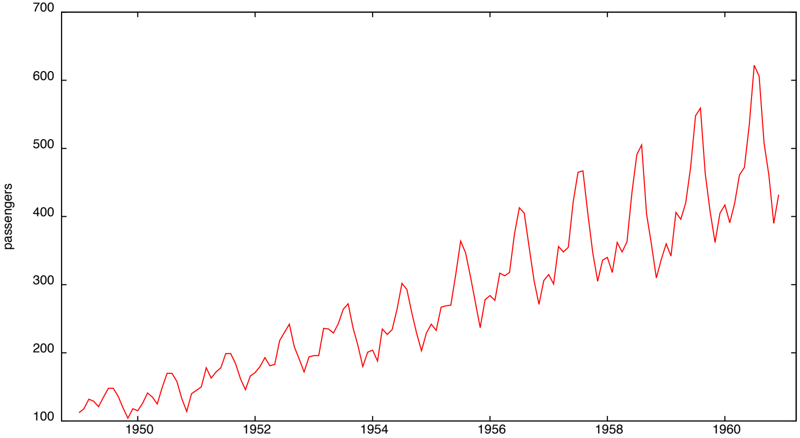
Examples, II

1) GDP growth, US (yearly data)



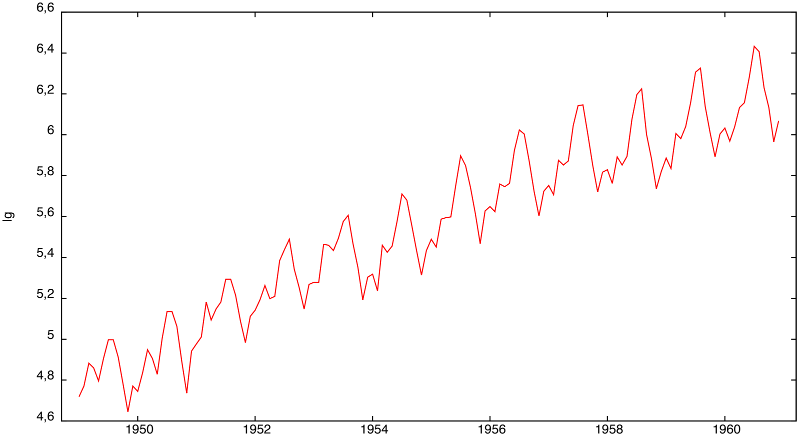
Examples, III

Airline passengers (monthly data)



Examples, IV

Log airline passengers (monthly data)



How Does Real Time Series Data look like?

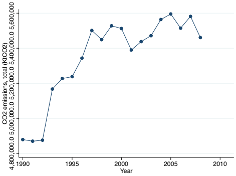
Time series data often display

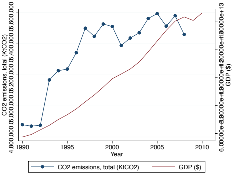
- trends
- seasonality
- Random fluctuations

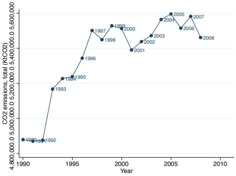
In this course we'll focus on modeling random fluctuations and trends

Plotting time series data with STATA

STATA provides good tools to draw nice time series plots. See the code in Lecture1.do







5. Structure of the course

Two important dimensions:

- Models for univariate vs. models for multivariate processes
- Models for stationary versus models for non-stationary models

	Stationary	Non-stationary
Univariate	Block 1	Block 3
Multivariate	Block 2	Block 4

Block 1: Univariate Stationary Processes

Block 1: Three main sections

- Preliminary Concepts
- Models for univariate stationary processes: ARMA models
- Estimation: (next handout).

II. Key preliminary concepts

1. Stochastic process
2. Univariate versus Multivariate time series
3. Autocovariance function
4. Stationarity
5. Examples of stationary processes
6. The Lag operator
7. Filters

1. Stochastic Process

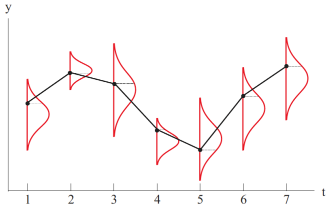
- A stochastic process is a collection of random variables $\{X_t(\omega),\ t \in T,\ \omega \in \Omega\}$ defined on a probability space $\{\Omega, \mathcal{F}, P\}$.

- if T=time, then it's a time series process.
- A stochastic process can be discrete or continuous according to whether T is continuous, *e.g.*, $T = \mathbb{R}$, or discrete, e.g., $T = \mathbb{Z}$.

- We'll mostly work with discrete time series, but also with continuous stochastic processes (Brownian Motion).

Stochastic Process, II

- A realization of a time series process is the observed time series data points.
- Inference is tough: from each random variable we only observe 1 value!



2. Univariate vs. Multivariate time series processes

- Univariate Analysis: models serial correlation among the random variables in a scalar stochastic process.
  - Example: $\{GPD\}_{t=1}^{T}$

$$GDP_1, GDP_2, \ldots, GDP_T$$

- Multivariate analysis: models serial correlation among the random variables in a vector stochastic process. vector stochastic processes.
  - Example:$\{X\}_{t=1}^{T}$ where,

$$X_t' = (GDP_t, CONS_t, INV_t, INFL_t)'$$

$$X_1, X_2, \ldots, X_T$$

- In this handout we will focus on UNIVARIATE time series

3. Autocovariance function

- It measures linear dependence among the random variables of a stochastic process $\{X_t,\ t \in \mathbb{Z}\}$
- It 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.

Autocovariance function, II

Definition

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

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

4. Stationarity

- Loosely speaking: processes whose statistical properties remain constant over time.
- Depending on what remains “constant” over time, several ways of defining this concept. (Strict vs Weak stationarity)
- We’ll focus on the most employed one: weak Stationarity

4. Stationarity: Weak, Second order or Covariance Stationarity

Definition

Weak, second order or covariance 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) $\mathrm{E}(X_t) = m$ for all t
iii) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ for all r, s, $t \in \mathbb{Z}$.

Weak Stationarity: comments, I

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

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

- Weak 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.**

Weak Stationarity: comments, II

- 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 γ_h will denote the h-th autocovariance of X_t.

Weak Stationarity: comments, III

The autocorrelation function

Definition

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

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

Some examples of stationary processes

- These simple processes are used as building blocks of more complex processes.

  Example 1:

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

  - Provided the first and second order moments exist (i.e., the mean and the variance), an *iid* sequence is covariance stationary.

Some examples of stationary processes, II

Example 2:

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

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

- A white noise sequence is always stationary
- An $i.i.d$ sequence with zero mean and variance σ^2 is also white noise.
- A white noise sequence, however, doesn't need to be i.i.d.

Some examples of stationary processes, III

Example 3:

- 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

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

  where $\{\varepsilon_t\}$ is a white noise process
- A MA is stationary for any value of θ.

Some examples of stationary processes, IV

To see this notice that:

$$E(X_t) = \mu$$

$$Var(X_t) = \sigma_\epsilon^2(1+\theta^2)$$

$$\gamma(t, t+1) = \sigma^2\theta$$

and for $h > 1$

$$\gamma(t, t+h) = 0.$$

Some examples of stationary processes, V

Example 4:

- 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 $|\phi| < 1$.

Take a look at some simulated data

- The graphs below correspond to simulated data.
- Plot of 1) the time series, 2) autocorrelation function and 3) partial correlation function
- Partial autocorrelation function: plots the correlation between y_t and y_{t-k} once the effect of $y_{t-1}, \ldots, y_{t-k-1}$ has been taken into account.

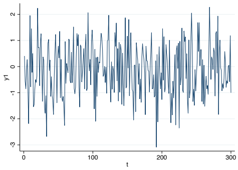
  i.e., the partial correlation of order k is given by the coefficient α_k in the linear projection:

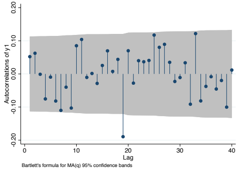
$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} \cdots + \alpha_k y_{t-k} + u_t$$

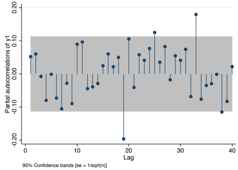
Take a look at some simulated data, II

You can find the code to generate these graphs in STATA: see Lecture1.do

- IID process: plot, autocorrelation function and partial autocorrelation function

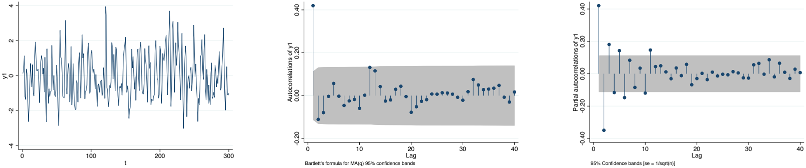






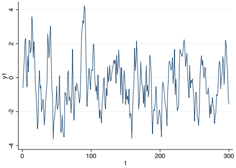
Simulated MA(1) process

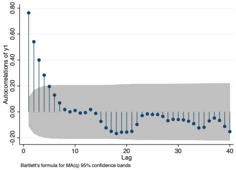
MA(1) process, $\theta = 0{,}8$: plot, autocorrelation function and partial autocorrelation function

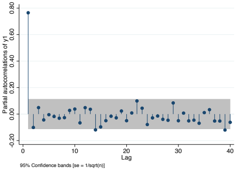


Simulated stationary AR(1) process

AR(1) process, $\phi = 0{,}8$: plot, autocorrelation function and partial autocorrelation function

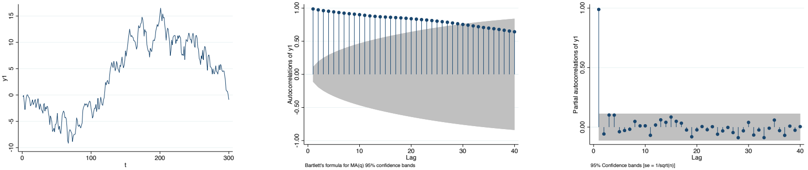






Simulated non-stationary AR(1) process (Random walk)

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



6. The Lag operator

- The models we're going to build often contain lags of other variables
- The lag operator L will provide a convenient way of writing these models
- Definition: The lag operator L maps a sequence $\{x_t\}$ into a sequence $\{y_t\}$ such that

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

The Lag operator, II

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

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

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

Filters

- We can also form polynomials in L:
  $A_p(L) = 1 + a_1 L + a_2 L^2 + ... + a_p L^p$, such that

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

- $A_p(L)$ is a filter
- Filters can contain a finite or an infinite number of terms.
- Examples: (1-L) (first differences), $\sum_{t=0}^{\infty} \psi_j L^j$ (Wold decomposition)

Inverting filters

- Filters can be inverted.
- Problem: for a given polynomial $\Phi_p(L)$, find $\Phi_p^{-1}(L)$ such that $\Phi_p(L)\,\Phi(L)^{-1} = 1$.
- (Why is this useful? we will use it when building time series models)
- This amounts to looking for the values of the coefficients α_i of $\alpha(L)$, that verify
    - $\Phi(L)^{-1} = \alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + ...$, that verify that
    - $\Phi_p(L)\,\Phi(L)^{-1} = 1$.

Inversion of polynomials in L: An example

Example I

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,

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

and

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

Inversion of polynomials in L: An example, II

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:

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

Note: Inversion can only be performed provided $|\phi| < 1$. Why?

Inversion of polynomials in L, Example II

Example II

Let p=2. Find the inverse of $\Phi_2(L) = (1 - \phi L - \phi_2 L^2)$.

- Two ways of doing this:

  1. Employ the same technique as before (i.e., match the coefficients associated to the same power of L). Drawback: it can get complicated for $p > 1$

  2. First factor the polynomial and then apply the same technique as before for each of the factors.

Inversion of polynomials in L: Example II

Problem: Invert $\Phi_2(L) = 1 - \phi_1 L - \phi_2 L^2$

- Let $1/\lambda_1$ and $1/\lambda_2$ be the roots of $\Phi_2(L)$,

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

- Provided $|\lambda_1|, |\lambda_2| < 1$, (or, equivalently, provided the roots are larger than 1 in absolute value)

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

III. Models for Stationary Processes

1. Overview
2. The Wold Theorem
3. MA processes
4. AR processes
5. ARMA processes: stationarity and invertibility

1. Overview

- Do economic time series typically show constant statistical properties over time? (NO!)
- Why do we always start by studying stationary processes?
- Three reasons
    1. Wold decomposition: allows us to write very simple (LINEAR) models for any weak stationary processes.
    2. Standard estimation, standard asymptotics, standard inference.
    3. Simple linear stationary models are used as “building blocks”for more complex models

Wold Theorem

- Let Z_t be a covariance-stationary process with mean μ. Then,

$$Z_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad (1)$$

where

- $\psi_0 = 1$
- (square summability): $\sum \psi_j^2 < \infty$
- $\{\epsilon_t\}$ is a white noise process with variance σ^2

The Wold Theorem, II

- Very IMPORTANT theorem: any stationary process can ALWAYS be written as a linear process. This process is a MA(∞): a linear combination of a white noise process.

- Very easy to manipulate analytically:

$$E(Z_t) = \mu; Var(Z_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

Particular case: MA(p)

If the order of the MA is finite, then the process is a MA(q) process, for some finite value of q:

$$Z_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

An alternative representation: From MA(∞) to AR(∞)

- By inverting the polynomial $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$, one can obtain an alternative representation: Autoregressive representation

- AR(∞): Let $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 \cdots = \Psi(L)^{-1}$. Multiply both sides of (1) by $\Psi(L)^{-1}$:

$$\begin{aligned}
\Psi(L)^{-1} Z_t &= \Psi(L)^{-1}\Psi(L)\epsilon_t \\
\Phi(L) Z_t &= \epsilon_t \\
Z_t &= \phi Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \epsilon_t
\end{aligned}$$

- $Z_t \sim AR(\infty)$ process
- Does this representation always exist?
- Answer: NO. Why?
- All roots of $\Psi(L)$ need to be larger than 1 in absolute value.

The previous condition is called the invertibility condition for MA processes and it allows us to write the AR representation.

From the Wold Theorem to ARMA processes

- The Wold theorem is extremely useful but also has limitations:
- The MA(∞) –and the AR(∞)– models contain an infinite number of parameters!
- A simplifying assumption: (ARMA assumption):

$$\sum_{j=0}^{\infty} \psi_j L^j = \Psi(L) = \frac{\Theta_q(L)}{\Phi_p(L)}$$

where $\Theta_q(L)$ and $\Phi_p(L)$ contain q and p lags, respectively. It follows that:

$$Z_t = \frac{\Theta_q(L)}{\Phi_p(L)}\epsilon_t \Rightarrow \Phi_p(L) Z_t = \Theta_q(L)\epsilon_t$$

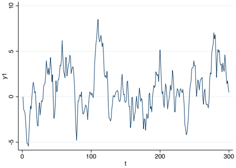
ARMA process

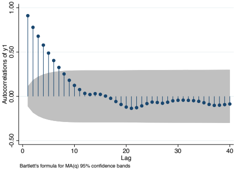
$$Z_t = \frac{\Theta_q(L)}{\Phi_p(L)}\epsilon_t \Rightarrow \Phi_p(L)Z_t = \Theta_q(L)\epsilon_t$$

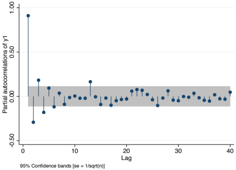
- AR component: $\Phi_p(L)$, p coefficients
- MA component: $\Theta_q(L)$, q coefficients
- Main advantage of ARMA processes: approximates the Wold representation with a finite number of parameters ($\approx$ parsimonious)

ARMA processes: simulated data

ARMA(2,2), with $\phi_1 = 0{,}7$, $\phi_2 = 0{,}2$, $\theta_1 = -0{,}6$ and $\theta_2 = 0{,}08$







Behavior of AC and PAC in ARMA processes: summary

Table 7.1. Indicators of p, d, and q

Process	Autocorrelation function	Partial autocorrelations function
Nonstationary	Autocorrelations do not die out They remain large or diminish approximately linearly	
Stationary	After the first few lags, autocorrelations die out (collapse toward 0 in some combination of exponential decay or damped oscillation)	
AR(p)	Autocorrelations die out	Partial autocorrelations cut off after the first p lags
MA(q)	Autocorrelations cut off after the first q lags	Partial autocorrelations die out
ARMA(p,q)	Autocorrelations die out after first $q-p$ lags	partial autocorrelations die out after first $p-q$ lags

ARMA processes: Invertibility and Stationarity

- Stationarity: a process is stationary if it has constant mean and variance and autocovariance function only depending on distance between observations.
- Invertibility: a process is invertible if it admits an autoregressive representation.

ARMA processes: Invertibility

- The AR polynomial of the ARMA process is already in autoregressive form
- We need to look at the MA(q) polynomial.
- The MA(q) can be inverted provided all its roots are larger than 1 in absolute value

ARMA processes: Stationarity

Stationarity means that we can write the process in MA form (Wold Theorem)

- The MA(q) is obviously not problematic
- We need to look at the AR(p) polynomial.
- The AR(q) is stationary (and therefore, can be written in MA form) provided all its roots are larger than 1 in absolute value

ARMA processes: Summary

An ARMA(p,q) process is stationary and invertible provided:

- The roots of the AR polynomial are larger than 1 in absolute value
- The roots of the MA polynomial are larger than 1 in absolute value

Why are these properties important?

- Stationarity: as we will see in Block III, without stationarity estimation and inference becomes complicated (no LLN or CLT!)

- Invertibility: We can't write the AR representation without this property. Important if we want to estimate a VAR!

Wrapping up

- This handout describes the theory behind ARMA models.
- These models provide a linear representation with a finite number of parameters for a stationary univariate process x_t
- ARMA processes are derived from the Wold theorem plus the "ARMA" assumption
- Models contain lags of x_t and a linear combination of a white noise process
- Next thing in the agenda: model specification, estimation and inference.