Introduction to Time Series Analysis

Lecture 1
Univariate Time Series Models

Mattia Guerini

Scuola Superiore Sant'Anna

30 March 2016

Planned schedule

In this lecture we will try to answer to the following questions:

- how to analyse macroeconomic data?
- which are the statistical tools more apt for analysing macroeconomic data?
- which are the basic properties of a time series?
- which are the most important classes of univariate processes and which properties they have?

Macroeconomics

Macroeconomics *is the study of the economy taken as a whole; whereas* **microeconomics** *is the study of a part of the economy (particular people, households, firms, markets, and so forth), taking the remainder as given.*

Hoover (2012)

The big picture: analysing economy-wide phenomena [...] Although economists generally separate themselves into distinct macro and micro camps, macroeconomic phenomena are the product of all the microeconomic activity in an economy. The precise relationship between macro and micro is not particularly well understood [...]

The Economist

Which variables?

The three main indicators for evaluating a country performance have historically been:

- **Gross Domestic Product**: the market value of the final goods and services produced by labour and property located within the borders of a country within a definite period.
- **Unemployment rate**: the percentage of the total labour force that is unemployed but actively seeking employment and willing to work.
- **Inflation rate**: the rate at which the *general level of prices* for goods and services is changing.
  - **GDP Deflator**: the ratio between nominal and real GDP.
  - **Consumer Price Index (CPI)**: a weighted average price where the weights reflect the relative importance of different goods and services in the consumption bundle of typical consumers.

Which dataset?

- **Cross Section**: data for which the time dimension is held fixed and each point refers to a separate unit
  - **Assumption**: the N observations are considered to be a realization of N i.i.d. draws (g.e. weight and height)
  - **Statistical tool**: realization of a random variable as a metaphor of the i^{th} observation.
- **Time Series**: data for which the unit dimension is held fixed and each point refers to a separate time period
  - **Property**: the assumption above is not valid here because two subsequent observation are typically not independent.
  - **Statistical tool**: stochastic processes as a metaphor for the generation of the observed time series.

Why not Standard Regression? (I)

Suppose you want to estimate the following model

$$x_t = \alpha + \beta x_{t-1} + \varepsilon_t, \quad \epsilon_t \sim^{iid} \mathcal{N}(0, \sigma_\varepsilon^2) \tag{1}$$

where x_t is a time series variable $x_t = (x_1, \ldots x_T)$ and x_{t-1} its lagged values.

For the OLS to provide a "good" (i.e. unbiased) estimation of α, β we need the Gauss-Markov assumptions to be satisfied

- **A1**: $E[\varepsilon_t] = 0$
- **A2**: $(x_1, \ldots x_T)$ and $(\varepsilon_1, \ldots \varepsilon_T)$ are independent
- **A3**: $Var[\varepsilon_t] = \sigma_\epsilon^2$
- **A4**: $Cov[\varepsilon_t, \varepsilon_{t-j}] = 0$

but when the data are time-series, the assumption **A2** (*exogeneity of the regressor*) does not hold.

Why not Standard Regression? (II)

Indeed we can rewrite equation 1 as (by iterating forward):

$$
\begin{aligned}
x_1 &= \alpha + \beta x_0 + \varepsilon_1 \\
x_2 &= \alpha(1+\beta) + \beta^2 x_0 + (\varepsilon_2 + \beta\varepsilon_1) \\
&\vdots \\
x_t &= \alpha(1+\beta+\cdots+\beta^{t-1}) + \beta^t x_0 + (\varepsilon_t + \beta\varepsilon_{t-1} + \cdots + \beta^{t-1}\varepsilon_1)
\end{aligned}
$$

Then by multiplying successively by $\varepsilon_t, \varepsilon_{t-1}, \ldots$ it's easy to check that:

$$
\begin{aligned}
E\left[x_{t-1}\varepsilon_t\right] &= 0 \\
E\left[x_{t-1}\varepsilon_{t-1}\right] &= \sigma_\varepsilon^2 \\
E\left[x_{t-1}\varepsilon_{t-2}\right] &= \beta\sigma_\varepsilon^2
\end{aligned}
$$

Thus the assumption of *exogeneity* does not hold when the regressor is a lagged value of the dependent variable.
OLS estimations will be biased.

The Practical Problem

We need therefore to tackle the problem from a different perspective and with time series data typically we try the following one:

- We have a macroeconomic time series ...
- assuming that it can be considered as a single realization of a stochastic process ...
- which kind of stochastic process provides realizations that are the most similar as possible to our time series?

In practice, we look for the *stochastic process* that most likely had generated the data we can observe in our time series.

Stochastic Process

Definition

A **stochastic process** is a collection of random variables representing the evolution of some system over time.

Intuitively we can think at a stochastic process as an infinite sequence of random variables. A time series of length T then, is not considered to be a sequence of T distinct draws from T random variables but as a unique draw of a stochastic process.

Properties of a stochastic process

- it is possible to define a density function for the stochastic process $f(\ldots, x_{t-1}, x_t, x_{t+1}, \ldots)$
- it is possible to find marginal densities for each subset of the process
- it is possible (if the marginal densities have moments) to compute mean, variance and covariance of the process

We will here always assume (unless explicitly told) that the moments exist.

Ergodicity, Stationarity, Memory

Three additional properties of a stochastic process are necessary for making inference

- **Ergodicity**: the fact that observing only a unique (but sufficiently long) draw of the process, is sufficient for knowing its whole characteristics.
- **Stationarity**: the fact that the process is stable in time, so that the statistical properties of the process are constant for the whole observation interval.
- **Memory**: the extent to which the future path of a process can be predicted, knowing its past history.

In what follows we will always assume ergodicity is verified while we will study what stationarity means and how it affects macro data analysis.

Moments of a Stochastic Process

The low-order moments of a stochastic process are defined as:

- **Mean**: $\mu_t = E[x_t]$
- **Variance**: $\sigma_t^2 = E[(x_t - \mu)^2]$
- **Autocovariances**: $\gamma_{j,t} = E[(x_t - \mu)(x_{t-j} - \mu)]$

Note that all the moments have a subscript for t meaning that they can possibly vary over time.

Stationarity

There are two different versions of this concept.

Definition
A finite process $(x_1, \ldots x_T)$ is said to be **strongly stationary** if, for every $k > 0$, the joint distributions of the collections $(x_t, \ldots, x_{t+k})$ do not depend in any way on t.

Definition
A finite process $(x_1, \ldots x_T)$ is said to be **weakly stationary** if the mean $E(\mathbf{x_t})$, the variance $Var(\mathbf{x_t})$ and the autocovariances $Cov(\mathbf{x_t}, \mathbf{x_{t-j}})$ for $j > 0$ are all independent of t.

Note:

- strong stationary $\Rightarrow$ weak stationary
- weak stationary $\nRightarrow$ strong stationary

From now on when we will mention stationary, we will refer to the weak stationary definition.

Memory

Starting from the variance ($\gamma_0 = \sigma^2$) and the autocovariances (γ_j) of a process we can define its autocorrelation function as

$$\rho_j = \frac{\gamma_j}{\gamma_0}$$

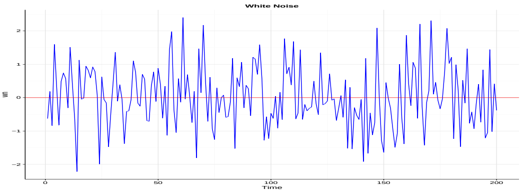
Then we can divide the processes in two categories according to their autocorrelation function:

- **Short memory**: if the autocorrelation function is decaying in j and it decays faster than at a certain rate;
- **Long memory**: if the autocorrelation funciton is not decaying in j or if it does not decay to zero after a certain number of lags $\bar{J}$.

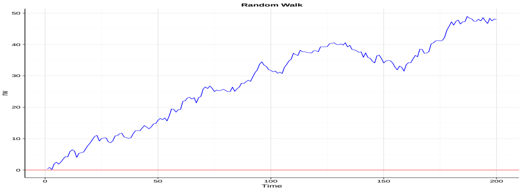
Note: the “true” definition of short/long memory is rather technical, but for our purpose here, it is sufficient to catch the broad meaning.

Examples (I)

A stationary process: the White Noise ($x_t = \alpha + \varepsilon_t$)

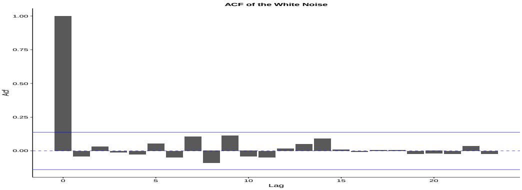


A non-stationary process: the Random Walk ($x_t = \alpha + x_{t-1} + \varepsilon_t$)

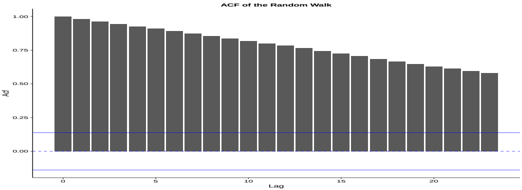


Examples (II)

A short memory process: the White Noise ($x_t = \alpha + \varepsilon_t$)



A longer memory process: the Random Walk ($x_t = \alpha + x_{t-1} + \varepsilon_t$)



The White Noise

We here check that the WN process is stationary.

$$x_t = \varepsilon_t, \quad \text{with} \quad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2)$$

To check for (weak) stationarity is sufficient to compute the mean, the variance and the autocovariance and show that they do not depend on t (i.e. they are constant over time).

- $\mu = E[x_t] = 0$
- $\sigma^2 = Var[x_t] = \sigma_\varepsilon^2$
- $\gamma_j = Cov[x_t, x_{t-j}] = 0$

Then the White Noise is stationary.

The Random Walk

We here check that the RW process is non-stationary.

$$x_t = x_{t-1} + \varepsilon_t, \quad \text{with} \quad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2), \quad x_0 = 0$$

To check for (weak) stationarity is sufficient to compute the mean, the variance and the autocovariance and show that they do not depend on t (i.e. they are constant over time).

- $\mu = E[x_t] = 0$
- $\sigma^2 = Var[x_t] = t\sigma_\varepsilon^2$
- $\gamma_j = Cov[x_t, x_{t-j}] = (t-j)\sigma_\varepsilon^2$

Then the Random Walk is non-stationary.

How to detect non-stationarity?

Apart from the graphical methods, we can test for non-stationarity using the Dickey-Fuller statistic. If a series is non-stationary we will also say that the series has a **unit-root**.

Dickey Fuller Test
Starting for the model of interest, by means of simple differentiation we can obtain and estimate

$$\begin{aligned} x_t &= \beta x_{t-1} + \varepsilon_t \\ \Delta x_t &= (\beta - 1) x_{t-1} + \varepsilon_t \\ \Delta x_t &= \eta x_{t-1} + \varepsilon_t \end{aligned}$$

and test the null hypothesis $H0 : \eta = 0$ against the alternative $H1 : \eta < 0$.

- If $\hat{\eta} = 0 \Rightarrow \beta = 1$ (i.e. non-stationary series)
- If $\hat{\eta} < 0 \Rightarrow \beta < 1$ (i.e. stationary series)

How to make a series stationary?

Assume the series we have tested with the DF-test has a unit root, we then know that OLS is biased (see slides above). Then what can we do to make the series stationary?

The simplest method is that of differentiating.

Indeed if x_t is non-stationary, Δx_t might be so.
For example, the difference of a RW is a WN.

$$x_t = x_{t-1} + \varepsilon_t$$

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t$$

Note:

- if after differentiation the variable is still non-stationary, you can proceed in taking differences many times $(\Delta x_t - \Delta x_{t-1})$
- a series that is stationary after 1 differentiation is said to be integrated of order 1: $I(1)$
- a series that is stationary after d differentiations is said to be integrated of order d: $I(d)$

The Lag Operator

Before moving to the most important family of stochastic processes (at least for time series econometrics analysis) we need to introduce the Lag operator (L).

$$Lx_t = x_{t-1}$$

Properties

- $L^j x_t = x_{t-j}$
- $L(\varphi x_t) = \varphi L x_t = \varphi x_{t-1}$
- $L(x_t + y_t) = Lx_t + Ly_t = x_{t-1} + y_{t-1}$
- $\lambda(L) = (1 - \lambda_1 L - \lambda_2 L^2 - \cdots - \lambda_p L^p)$

Autoregressive Processes AR(p)

Many macroeconomic time series are represented by means of an autoregressive process; in fact these processes allow to take into account autocorrelation and to represent the fact that a variation in the variable today might persist up to p periods.

An autoregressive process of order p, AR(p), is a stochastic process in the form

$$x_t = \alpha + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \cdots + \lambda_p x_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2)$$
$$\lambda(L) x_t = \alpha + \varepsilon_t$$

In particular we here focus only on AR(1) process to understand their basic properties, therefore

$$x_t = \alpha + \lambda_1 x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2)$$

Note:

- if $\lambda_1 = 0 \rightarrow$ White Noise + constant
- if $\lambda_1 = 1 \rightarrow$ Random Walk + constant

Moments of the AR(1)

The AR(1) process might be both stationary and non-stationary according to the value of the parameter λ_1 (properties of geometric series).

If it is stationary, then its low-order moments are:

- $E[x_t] = \frac{\alpha}{1-\lambda_1}$
- $Var[x_t] = \frac{\sigma_\varepsilon^2}{1-\lambda_1^2}$
- $Cov[x_t, x_{t-j}] = \lambda_1^j \frac{\sigma_\varepsilon^2}{1-\lambda_1^2}$

Therefore in the AR(1) case the autocorrelation is:

- $\rho_j = \lambda^j$

and the process has therefore a decaying (short) memory.

Moving Average Processes MA(q)

MA(q) processes are another important class of stochastic processes for macroeconomics, because they can represent the effects of exogenous shocks to a macroeconomic variable, whose effect last for q periods.

A Moving Average process of order q, MA(q), is a stochastic process in the form

$$x_t = \alpha + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \cdots + \vartheta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2) \quad \forall t$$
$$x_t = \alpha + \vartheta(L)\varepsilon_t$$

In particular we here focus only MA(1) process to understand their basic properties, therefore

$$x_t = \alpha + \varepsilon_t + \vartheta_1 \epsilon_{t-1}, \qquad \varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2) \quad \forall t$$

Moments of the MA(1)

The low-order moments of the MA(1) (for $\alpha = 0$) are:

- $E\left[x_t\right] = 0$
- $Var\left[x_t\right] = (1 + \vartheta_1^2)\sigma_\varepsilon^2$
- $Cov\left[x_t, x_{t-1}\right] = \vartheta_1 \sigma_\varepsilon^2$

Therefore the MA(1) process is always stationary, for any value of ϑ.

In general any MA(q) process is stationary and has a memory equal to q, meaning that the covariance is different to zero up to q lags.

ARMA (p,q) processes

This is the most important family of univariate time series models for economists because combine AR(p) and MA(q) in the form:

$$x_t = \alpha + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \cdots + \lambda_p x_{t-p} + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \cdots + \vartheta_q \varepsilon_{t-q}$$

with: $\varepsilon_t \sim^{iid} N(0, \sigma_\varepsilon^2) \quad \forall t$

This class is important because:

- The AR(p) component represents persistence in a variable dynamics
- The MA(q) component represents the effects of current and past shock to that dynamics.
- Any stationary time series can be represented as an ARMA(p,q) model (*Wald Theorem*).

Box and Jenkins Procedure

Remember that our aim was to estimate a dynamic model: one that contains as independent variables the lagged values of the dependent one.

In order to do that, we have to follow the Box-Jenkins procedure.

1. **Preliminary Analysis**: graphical inspection and test for stationarity. If the series is not stationary, transform it to be so.
2. **Model Identification**: by means of the Autocorrelation Function identify the number of lags (p,q).
3. **Parameters Estimation**: by means of OLS or Maximum-Likelihood.
4. **Model Checking**: if the model is well specified and well estimated, the residuals should be White Noise.

But why do we need that?

The aim of using this kind of analysis might be explained in the following way:

- We know that stochastic processes have certain characteristics
- and if we get to know them, we can explain and predict the behaviour of the process over time.
- Therefore, if a macroeconomic variable can be represented as a well known stochastic process,
- then we can explain and predict the dynamic of the macro variables.

But, up to now only worked with 1-variable at the time studying its behaviour only as a function of its own past (univariate case).

We know by the way that macro variables are closely linked one with the other, so we need to switch to a multivariate case.