Introduction to Econometrics (Part. I)

Giovanni Angelini

Rimini, 9 September 2013 - 19 September 2013

Information

Giovanni Angelini

PhD students in 'Statistical Methodology for Scientific Research'
Department of Statistics "P.Fortunati" - University of Bologna

giovanni.angelini3@unibo.it

Time Table

- 09/09/2013 10-13, 14-17 Alberti 10
- 11/09/2013 10-13 Alberti 10, 14-17 Green Lab
- 13/09/2013 10-13 Alberti 10, 14-17 Green Lab
- 17/09/2013 10-13 Alberti 10, 14-17 Green Lab
- 19/09/2013 10-13 Alberti 10, 14-17 Green Lab

Outline

- Econometrics definition:
  - Econometrics analysis
  - Econometrics model
  - Different type of data
- Uni-EquationI Model
- Recall of probability: conditional distribution and expectations
- Examples

What is Econometrics and what is econometric analysis?

- **Econometrics is a branch of social sciences that uses quantitative methods to investigate economic phenomena with respect to which observations (data) are available.**
- **Economic phenomenon of interest:** consumption, production, good markets, financial markets.
- **The observations (data)** are treated as realizations of an underlying **stochastic process** which is unknown to the researcher, known as the **data generating process (DGP).**

- All techniques and methods of **probability** and **statistical** analysis can be used to infer the relationships of interest.
- Other than the observations (data), the researcher must have a **theoretical view** about the model that might have generated the data.
- **Theory & Data**: both contribute to the formulation of an econometric model.
- The econometric analysis of an economic phenomenon represents a summary of the theory used to explain the phenomenon and of the available observations (data).
- **Perspective**: the econometric analysis of an economic phenomenon can be carried out at different levels: micro level or macro level, with time-series data, with cross-section data or with panel data.

- **Micro:** the analysis of the phenomenon of interest is carried out with reference to individual behaviour (e.g. consumption of individuals or firms) and the observations refer to these individuals:
  - **Theory referes to the behaviour of individuals**
  - **Micro Data**
  - **Microeconometrics.**
- **Macro**: the analysis of the phenomenon of interest is carried out at an aggregate level (e.g. production of the industural sector or of the whole country) and the observations are the result of an aggregation process or refer to variables that are common to all individuals:
  - **Theory refers to agrgegate behaviour**
  - **Macro Data**
  - **Macroeconometrics.**

- **Time-series:** the phenomenon of interest (micro or macro) is investigated over time:
  - **Time series data.**
- **Cross-section:** the phenomenon of interest (micro or macro) is investigated by observing many units (individuals) at the same point in time:
  - **Cross-section data.**
- **Panel:** the phenomenon of interest (micro or macro) is investigated by observing many units (individuals) at different points in time:
  - **Panel data.**

- **Quantitative methods**: the econometric anlysis makes use of equations and data that formalize the behavior of economic agents (at the micro or macro level) and/or explain the relationships between the variables that are important to explain the phenomenon of
  interest. These equations generally involve a set of parameters of interest that admit economic interpretation. These parameters are typically unknown to the researcher.
- **One of the main objectives of the econometric analysis is to make inference about them and testing whether some implications and or restrictions implied by the theory are satisfied by the observed data.**
- This requires the use of **estimation techniques** and **hypothesis testing procedures** (inductive process).

What is an econometric model?

- Our objective is the investigation of an economic phenomenon of interest.
- Examples:
  - the impact of government spending and taxes on GDP growth;
  - the impact of a monetary policy shock on output and unemployment;
  - the impact of an earthquake on the economic activity;
  - the effect of oil price on petrol price and quantity;
  - the effect on consumption of a 1% increase of the price of cigarettes;
  - the effect on a firm's sales of an advertising campaign;
  - the effect of university enrolment of an increase of the student tax rate;
  - the effect of stock index fluctuations on portfolio allocation.
- To address all these questions we have to **combine theory & data.**

- Let y be the $g \times 1$ vector containing the variables we want to explain: henceforth y will denote the **endogenous variable**.
- We wish to model y :
  - we need to understand which are the determinats of y in order to **explain**, **forecast** or **put forth policy actions**. In other words, we want to answer the list of questions seen before.
- The endogenous variables y can be either **quantitative** (generally continuos), e.g. a short term interest rate, consumption, production, etc. or **qualitative**, e.g. y measures the participations of househols in financial market (namely y might be equal to one if the household purcheses financial assets or zero if the household does not purchases financial assets, etc).

In this course we focus on the case in which y is a vector containing quantitative variables.

- Henceforth the vector y will be denoted as the vector of **endogenous variable**.

$$y := \begin{pmatrix} y_1 \\ \vdots \\ y_g \end{pmatrix}$$

- According to our theoy and/or experience, we have in mind a situation in which y can be explained in terms of a $f \times 1$ vector of **quantitative (stochastic)** variables

$$\tilde{x} := \begin{pmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_f \end{pmatrix}$$

- *For each unit (time) i of our sample, we treat y_i and x_i as random (stochastic) vector variables!*

- The vectors

$$y_i := \begin{pmatrix} y_{1,i} \\ \vdots \\ y_{g,i} \end{pmatrix} \quad , \quad \tilde{x}_i := \begin{pmatrix} \tilde{x}_{1,i} \\ \vdots \\ \tilde{x}_{f,i} \end{pmatrix}$$

  denote the observation on y and x relative to the i-th individual or at time i.
- Sometimes we need to include among the explanatory variables also some deterministic variables, e.g. constant, linear trend, dummy variables, etc. We define

$$d_i := \begin{pmatrix} d_{1,i} \\ \vdots \\ d_{h,i} \end{pmatrix}$$

  the $h \times 1$ vector of **deterministic components**.

- Data (dataset), our observations are:

$$y_1 := \begin{pmatrix} y_{1,1} \\ \vdots \\ y_{g,1} \end{pmatrix}, \; y_2 := \begin{pmatrix} y_{1,2} \\ \vdots \\ y_{g,2} \end{pmatrix}, \ldots, \; y_n := \begin{pmatrix} y_{1,n} \\ \vdots \\ y_{g,n} \end{pmatrix}$$

$$\tilde{x}_1 := \begin{pmatrix} \tilde{x}_{1,1} \\ \vdots \\ \tilde{x}_{f,1} \end{pmatrix}, \; \tilde{x}_2 := \begin{pmatrix} \tilde{x}_{1,2} \\ \vdots \\ \tilde{x}_{f,2} \end{pmatrix}, \ldots, \; \tilde{x}_n := \begin{pmatrix} \tilde{x}_{1,n} \\ \vdots \\ \tilde{x}_{f,n} \end{pmatrix}$$

$$d_1 := \begin{pmatrix} d_{1,1} \\ \vdots \\ d_{h,1} \end{pmatrix}, \; d_2 := \begin{pmatrix} d_{1,2} \\ \vdots \\ d_{h,2} \end{pmatrix}, \; \ldots, \; d_n := \begin{pmatrix} d_{1,n} \\ \vdots \\ d_{h,n} \end{pmatrix}.$$

- In our course we focus on the case in which y_i is a scalar (i.e. $g = 1$, we have a single endogenous variable) and the relationhisp between y_i and $x_i{:=}(\tilde{x}_i, 1)'$ ($d_i = 1$ we have a constant) is given by

$$y_i = \beta' x_i + u_i \quad , \quad i = 1, ..., n$$

  where the vector of unknown parameters is $\theta{:=}(\beta', \sigma_u^2)$ with $\sigma_u^2{:=}E(u_i^2)$.

This is the uni-equational linear regression model.

- y_i is a scalar, $\tilde{x}_i$ and d_i vectors.

$$x_i := \begin{pmatrix} \tilde{x}_i \\ d_i \end{pmatrix} \text{ is } k \times 1 \quad (k := f + h) \textbf{ explanatory variables}$$

- We have in mind a relationship of the form

$$y_i = E(y_i \mid x_i) + u_i \quad , \quad i = 1, 2, ..., n \tag{1}$$

where

$$E(y_i \mid x_i) = x_i' \beta \ , \quad i = 1, 2, ..., n$$

$$u_i := y_i - E(y_i \mid x_i) \ , \quad i = 1, 2, ..., n.$$

- In the model

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n.$$

u_i **is not observable** ! **It is not part of the data** !
u_i captures our uncertainly about the relationship between y and x.

Obviously, since y_i is a scalar, u_i will be a scalar as well.

Probability Distribution for a Discrete variable

- The probability distribution of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur. **This probability sum to one**.
- The cumulative probability distribution is the probability that the random variable is less than or equal to a particoular value.
- *Example of a roll of a dice.*

Probability Distribution for a Continuous variable

- Because a continuous random variable can take a continuum of possible values, the probability distribution used for discrete variables, which list the probability of each possible value of the random variable, is not suitable for continuous variables. The probability is summarized by the probability density function. The area under the probability density function between two points is the probability that the random variable falls between this two points.
- The cumulative probability distribution for a continuous variable is defined just as it is for discrete random variable.

Expected Value

- Let X a random variable with pdf $f(x)$, the expected value of X is:

$$E(X) = \mu_x = \sum_{x \subset X} x f(x), \text{ in discrete case}$$
$$E(X) = \mu_x = \int_{x \subset X} x f(x), \text{ in continuos case}$$

- Properties:
  - $E(aX + b) = aE(X) + b$
  - $E(aX_1 + ... + nX_n) = aE(X_1) + ... + nE(X_n)$
  - if X and Y are indipendent $->$ $E(XY) = E(X)E(Y)$
- Exercize: find $E(X)$ where X is the outcome of rolling a die

Variance

- Let X a random variable, thevariance of X, denoted by Var(X) or σ_x^2 is defined as:

$$
\begin{aligned}
Var(X) &= \sigma_x^2 = E[(X - E(X))^2] = E(X^2) - E(X)^2 \\
&= E\left[(X - \mu_x)^2\right]
\end{aligned}
$$

- Properties:
  - $Var(aX + b) = a^2 Var(X)$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
  - $Var(aX_1 + ... + nX_n + d) = a^2 Var(X_1) + ... + n^2 Var(X_n)$
- Exercize: find $Var(X)$ where X is the outcome of rolling a die

Covariance and Correlation

- Covariance and Correlation measure the dependence between two random variable.

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= E\left[(X-\mu_x)(Y-\mu_Y)\right] \\ Corr(X,Y) &= \frac{Cov(X,Y)}{\sqrt{Var(X)+Var(Y)}} \end{aligned}$$

Conditional distribution and expectations

- Let $w_i = \begin{pmatrix} y_i \\ z_i \end{pmatrix}$ be bi-variate continuous random vector with density $f_{y,w}(w_i;\theta)$ (joint distribution). Marginal distribution of y_i :

$$f_y(y_i;\theta_y) = \int_z f(w_i;\theta)dz_i$$

from which

$$E(y_i) = \int_y y_i \; f_y(y_i;\theta_y)dy_i$$

Marginal distribution of z_i :

$$f_z(z_i;\theta_z) = \int_y f(w_i;\theta)dy_i$$

- from which

$$E(z_i) = \int_z z_i \; f_z(z_i; \theta_y) dz_i$$

Conditional distribution of y_i given z_i :

$$f_{y|z}(y_i \mid z_i; \theta_{y|z}) = \frac{f_{y,w}(w_i; \theta)}{f_z(z_i; \theta_z)}$$

from which we derive the well known factorization:

$$\underset{\text{joint}}{f_{y,w}(w_i; \theta)} \;=\; \underset{\text{conditional}}{f_{y|z}(y_i \mid z_i; \theta_{y|z})} \times \; \underset{\text{marginal}}{f_z(z_i; \theta_z)}.$$

and

$$E(y_i \mid z_i) = \int_y y_i \; f_{y|z}(y_i \mid z_i; \theta_{y|z}) dy_i.$$

Intuitively, $E(y_i \mid z_i)$ is as a function of z_i.

- We can further generalize the concept of conditional expectations. Let v a $p \times 1$ stochastic vector and $\mathcal{F}$ a set containing random variables and all of their linear combinations (sigma-algebra or sigma-field). Then

$$E(v \mid \mathcal{F}) = \begin{pmatrix} E(v_1 \mid \mathcal{F}) \\ \vdots \\ E(v_p \mid \mathcal{F}) \end{pmatrix}$$

is a stochastic vector.

Properties

- *Linearity*:

$$E(a'v_1 + b'v_2 \mid \mathcal{F}) = a'E(v_1 \mid \mathcal{F}) + b'E(v_2 \mid \mathcal{F})$$

  where v_1 and v_2 stochastic vectors and a and b deterministic vectors of conformable dimensions;
- *Inclusion rule*: if $v \in F$, then

$$E(v \mid \mathcal{F}) = vE(1 \mid \mathcal{F}) = v$$

- *Law of iterated expectations 1*: Suppose that $G \subseteq F$; then

$$E\left[(v \mid \mathcal{F}) \mid \mathcal{G}\right] = E(v \mid \mathcal{G}).$$

- *Law of iterated expectations 2*:

$$E(v) = E_{\mathcal{F}}\left[E(v \mid \mathcal{F})\right]$$

  (This rule established a link between unconditional expectations and

The model

- $$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

  implies that for each i we have implicitly in mind the $(f+1) \times 1$ vector

  $$w_i := \begin{pmatrix} y_i \\ \tilde{x}_i \end{pmatrix} \quad \begin{matrix} 1 \times 1 \\ f \times 1 \end{matrix}$$

  which is the **whole vector of stochastic observable** variables.

- Thus the dataset is

  $$w_1, w_2, ..., w_n$$

  $$d_1, d_2, ..., d_n.$$

  Since w_i is a stochastic vector, it has distribution function, $f_w(w_i, \theta_w)$, where θ_w denotes the vector of parameters.

- As is known, one can always write

$$\underbrace{f_w(w_i, \theta_w)}_{\text{joint}} = \underbrace{f_{y \cdot \tilde{x}}(y_i \mid \tilde{x}_i \ , \theta_{y \cdot \tilde{x}})}_{\text{conditional}} \times \underbrace{f_{\tilde{x}}(\tilde{x}_i \ , \theta_{\tilde{x}})}_{\text{marginal}}$$

where $f_{y \cdot \tilde{x}}(y_i \mid \tilde{x}_i \ , \theta_{y \cdot \tilde{x}})$ is the **conditonal** distribution of y_i given x_i and $f_{\tilde{x}}(\tilde{x}_i \ , \theta_{\tilde{x}})$ is the **marginal** distribution of $x_i . \theta_{y \cdot x}$ is the vector of parameters of the conditional density: it should be clear that β is somehow related to $\theta_{y \cdot \tilde{x}}$! $\theta_{\tilde{x}}$ is the vector of parameters of the marginal density of $\tilde{x}_i$: it seems that with model (1) we are not interested in these parameters.

- Note that in general

$$g(x_i,\beta) := E(y_i \mid x_i) \; := \int y_i f_{y\cdot\tilde{x}}(y_i \mid \tilde{x}_i, \theta_{y\cdot\tilde{x}}) dy_i$$

To simply the general setup, we usually **assume** that

$$E(y_i \mid x_i) := g(x_i,\beta) := \text{ is a linear function of } x_i$$

in particular

$$E(y_i \mid x_i) := g(x_i,\beta) := \beta' x_i = x_i'\beta.$$

For the moment we focus only on the case in whic the function g is linear in x!

- In model $yi = g(x_i, \beta) + u_i$, $i = 1, 2, ..., n$ the function $g(xi, \beta)$ denotes the conditional expectation of y_i given $_{xi}$ and u_i is the disturbance (or error) term that captures the portion of y_i that is not explained by $E(y_i|x_i)$.
- u_i is not observable ! It is not part of the data !
- u_i captures our uncertainly about the relationship between y and x.

- The **linear uni-equational model** becomes

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

where

$$E(u_i) := E\left(E(u_i \mid x_i)\right) = E\left(0\right) = 0.$$

$$Var(u_i^2) := E(u_i^2) := ?$$

- Why $E(u_i \mid x_i) = 0$?

$$E(u_i \mid x_i) := E(y_i - E(y_i \mid x_i) \mid x_i) = E(y_i \mid x_i) - E(y_i \mid x_i) = 0.$$

(Law of iterated expectations).
The disturbance (error) has no sistematic effect.

- What about $Var(u_i) := E(u_i^2)$?
- In general, we have no information about $Var(u_i) := E(u_i^2)$.
- We have to formulate hypotheses about $Var(u_i)$, depending on the phenomenon under investigation.
- Assume, for instance that

$$E(u_i^2 \mid x_i) := \sigma_u^2 \quad , \quad i = 1, ..., n.$$

Then $E(u_i^2) := E\left(E(u_i^2 \mid x_i)\right) := E\left(\sigma_u^2\right) = \sigma_u^2$.
This is the hypothesis of homoschedasticity.

- Homoschedasticity is a strong assumption: all units in the sample have the same variance, or the variability of the endogenous variables do not change over time !

- Note that

$$Var(y_i \mid x_i) := E(y_i^2 \mid x_i) - [E(y_i \mid x_i)]^2$$
$$= E(y_i^2 \mid x_i) - (\beta' x_i)^2.$$

Now

$$E(y_i^2 \mid x_i) := E([\beta' x_i + u_i)]^2 \mid x_i)$$
$$= E([u_i^2 + (\beta' x_i)^2 + 2u_i \beta' x_i] \mid x_i)$$
$$E(u_i^2 \mid x_i) + E((\beta' x_i)^2 \mid x_i) + 2E(u_i \beta' x_i \mid x_i)$$
$$= E(u_i^2 \mid x_i) + (\beta' x_i)^2.$$

Thus

$$\begin{aligned} Var(y_i \quad \mid \quad & x_i) := E(y_i^2 \mid x_i) - [E(y_i \mid x_i)]^2 \\ = \quad & E(u_i^2 \mid x_i) + (\beta' x_i)^2 - (\beta' x_i)^2 := \sigma_u^2. \end{aligned}$$

- The linear regression model with homoschedasticity and with the additional condition **(uncorrelation)**

$$Cov(u_i,\, u_j){:=}\sigma_{i,j}{:=}0 \quad \forall i \neq j$$

will be denoted as **the classical linear regression model.**

- This model will be written as

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

with

$$E(u_i){=}\ E_{x_i}\left(E(u_i \mid x_i)\right) = E_{x_i}\left(0\right){=}0.$$
$$Var(u_i^2){:=}E(u_i^2){:=}\sigma_u^2 \quad \forall i$$
$$Cov(u_i,\, u_j){:=}\sigma_{i,j}{:=}0 \quad \forall i \neq j$$

The unknown parameters are $\theta{:=}(\beta',\, \sigma_u^2)'$, where $\dim(\theta){:=}k+1$. *This means that if* $n > k+1$ *we can estimate* θ *using* n *observations.*

- Take the model

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n \tag{2}$$

with

$$E(u_i) = 0$$

$$Var(u_i^2) := E(u_i^2) := \sigma_u^2 \quad \forall i$$

$$Cov(u_i, u_j) := \sigma_{i,j} := 0 \quad \forall i \neq j$$

The unknown parameters are $\theta := (\beta', \sigma_u^2)'$.

- x_i is said to be **weaky exogenous with respect to** θ: this means that the knowledge of the marginal distribution $f_{\tilde{x}}(\tilde{x}_i\ ,\ \theta_{\tilde{x}})$ does not provide additional information about the parameteres of interest (θ).

- **Crucial point**:

$$Cov(x_i, u_i) := E(x_i u_i) = ? \text{ it is a } k \times 1 \text{ vector !}$$

$$E(x_i u_i) := E_{x_i}\left(E(x_i u_i \mid x_i)\right) = E_{x_i}\left(x_i E(u_i \mid x_i)\right)$$
$$= E_{x_i}\left(0_{k\times 1}\right) = 0_{k\times 1}.$$

- In the linear regression model

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

the regressors and the disturbance term must be uncorrelated for each i:

$$Cov(x_i, u_i) := E(x_i u_i) = 0_{k\times 1}.$$

This certainly happens when x_i is not stochastic (there are no regressors $\tilde{x}_i$s in x_i) but must be true also when x_i do is stochastic.

- Imagine to pre-multiply both sides of

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

by x_i:

$$x_i y_i = x_i x_i'\beta + x_i u_i \quad , \quad i = 1, 2, ..., n$$

then take expectations of both sides:

$$E(x_i y_i) = E(x_i x_i')\beta + E(x_i u_i) \quad , \quad i = 1, 2, ..., n.$$

If $E(x_i u_i) = 0_{k\times 1}$, one can interpret the **population parameter** β as

$$\beta = \Big[\underset{k\times k}{E(x_i x_i')}\Big]^{-1} \underset{k\times 1}{E(x_i y_i)}$$

on conditon that the $k \times k$ matrix $E(x_i x_i')$ be invertible. Note that $E(x_i x_i')$ is the matrix of second order - non centred - moments of regressors.

- $E(x_i u_i) = 0_{k\times 1}$ implies

$$E(x_i y_i) = E(x_i x_i')\beta \quad , \quad i = 1, 2, ..., n$$

which implies

$$\beta = \Big[\underset{k\times k}{E(x_i x_i')}\Big]^{-1} \underset{k\times 1}{E(x_i y_i)}.$$

- Simple estimation rule (dates back to the method of moments): replace population moments with sample moments:

$$\hat{\beta} := \underset{k\times k}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right]}^{-1} \underset{k\times 1}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right]} := (X'X)^{-1}(X'y)$$

- **Generalization**
  - Assume, for instance, that

$$E(u_i^2 \mid x_i) := h(x_i) := \alpha x_{1,i}^2$$

where $\alpha > 0$. This is an example of **hypothesis of conditional heteroshedasticity.**

  - Is in this case heterosckedasticity also unconditional?

$$E(u_i^2) := E_{x_i}\left(E(u_i^2 \mid x_i)\right)$$

$$= E_{x_i}\left(\alpha x_{1,i}^2\right) = \alpha^2 E_{x_i}\left(x_{1,i}^2\right).$$

If $E_{x_i}\left(x_{1,i}^2\right)$ depends on i we **also have unconditional heteroschedasticity, otherwise unconditional homoskedasticity.**

  - This serves to show that we can have situations characterized by conditional heterosckedasticity but unconditional homoschedasticity.

- The linear regression model with unconditional heteroskedasticity will be denoted as **the generalized linear regression model:**

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

$$E(u_i) = 0 \;\; , \;\; E(u_i^2) = \sigma_i^2.$$

$$Cov(u_i,\, u_j) := \sigma_{i,j} := 0 \quad \forall i \neq j$$

$$Cov(x_i, u_i) := E(x_i u_i) = 0_{k\times 1} \;\; , \quad i = 1, 2, ..., n.$$

The unknown parameters are $\theta := (\beta',\, \sigma_1^2, \sigma_2^2, ..., \sigma_n^2)'$, where

$$\dim(\theta) := k + n.$$

This means that with n observations it is impossible to estimate θ (unless with have some prior information about $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$).

- If
$$Cov(u_i,\ u_j) := \sigma_{i,j} \neq 0 \quad \forall i \neq j$$
then unknown parameters also include all covariances, hence
$$\dim(\theta) := k + n + \frac{1}{2}n(n-1)$$
where note that
$$\begin{pmatrix} n \\ 2 \end{pmatrix} := \frac{1}{2}n(n-1).$$
- This means that with n observations **it is impossible to estimate** θ (unless with have some prior information about $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ and the covariances $\sigma_{i,j}$).
- With the generalized linear regression model **we have a problem**. In general, it is not possible to estimate all parameters, unless some prior information is used.

- **To sum up:**
  - Uni-equational linear regression model

$$y_i = x_i'\beta + u_i \quad , \quad i = 1, 2, ..., n$$

$$\begin{cases} \text{classical:} & \begin{matrix} Cov(u_i \ , \ u_j) := 0 \quad , \forall i \neq j \\ E(u_i^2) := \sigma_u^2 \quad , \forall i \end{matrix} \\ \\ \text{generalized:} & \begin{matrix} Cov(u_i \ , \ u_j) := \sigma_{i,j} \text{ zero or not} \\ E(u_i^2) := \sigma_i^2 \ , \ i = 1, ..., n \end{matrix} \end{cases}$$

EXAMPLE

- How do central banks fix short term interest rates ? (How do central banks run monetary policy ?). Currently used model (Taylor rule):

$$R_i = \varphi_1 b_i + \varphi_2 \pi_i + u_i \quad , \quad i = 1, 2, ..., n$$

i = denotes **time**;
n= the total number of available observations (n is usually called T when i denotes time);
R_i = short term interest rate fixed by the central bank at period i or for country i (policy rate);
b_i= measure of the business cycle status (capacity utilization, output gap) at period i or for country i ;

π_i=inflation rate at time i or for the country i;
u_i=is an **unobservable disturbance term** that can be interpreted as the part of R_i that we are not able to explain on the basis of our theory or view of the phenomenon;
φ_1 =parameter ($\varphi_1 \geq 0$) that captures the impact of economic activity on interest rate setting (thus if the economic activity increases of 1%, the long-run response of the policy rate is φ_1);
φ_2 =parameter ($\varphi_2 \geq 0$) that captures the impact of inflation on interest rate setting (thus if inflation rises of 1%, the long-run response of the policy rate is φ_2).

- The **estimation** of the parameters φ_1 and φ_2 from the data, i.e. from the observations

$$R_1, R_2, ..., R_n$$
$$b_1, b_2, ..., b_n$$
$$\pi_1, \pi_2, ..., \pi_n$$
$$\text{no direct obs. on } u_i \text{ !!}$$

is of crucial importance for central banks to understand whether they are stabilizing the economy or not.

- For instance, it can be shown that if $\varphi_2 \leq 1$, then the central bank does not control the economy sufficiently well (it is not aggressively enough) and this would imply that monetary policy does not contrast inflation !

- This means once we have estimated the parameters φ_1 and φ_2 from the data, **and we are relatively sure that the estimated model addresses the data sufficiently well**, we are also interested in testing (linear) restrictions of the type

$$H_0 : \varphi_2 = 1$$

against

$$H_1 : \varphi_2 \neq 1$$

Note that the null H_0 here means that the response of the central bank to inflation is not aggresively enough and that there is no inertia in the policy rule.

- Before drawing any meaningful conclusion from an estimated model we have to be sure that the model describes the observed data sufficiently well.
  Reasonable statistical approximation of the Data Generating Process (DGP).
- How do we evaluate **the data adequacy** of the estimated model ?

$$\text{R-squared} = \frac{\sum_{i=1}^{n}(\hat{R}_i - \bar{R})^2}{\sum_{i=1}^{n}(R_i - \bar{R})^2} = \frac{\text{DEV-REG}}{\text{DEV-TOT}}$$

may be a misleading indicator !

- To see this imagine that we estimate

$$R_i = \rho R_{i-1} + \beta_1 b_i + \beta_2 \pi_i + u_i$$

$$\begin{aligned} \beta_1 &= (1-\rho)\varphi_1 \quad , (\hat{\varphi}_1 = \frac{\hat{\beta}_1}{1-\hat{\rho}}) \\ \beta_2 &= (1-\rho)\varphi_2 \quad , (\hat{\varphi}_2 = \frac{\hat{\beta}_2}{1-\hat{\rho}}) \end{aligned}$$

by OLS. We use (actual) U.S. data.

- Which are **the properties** of the OLS estimator i**n this context** ?

Estimation on actual U.S. data
R_i=‘FFC’
b_i=‘YGAPC’
π_i = ‘INFC’

Modello 1: Stime OLS usando le 94 osservazioni 1985:2–2008:3
Variabile dipendente: FFC

Variabile	Coefficiente	Errore Std.	Statistica t	p-value
const	-0.0161560	0.0114880	-1.4063	0.1631
YGAPC	0.0301904	0.00792405	3.8100	0.0003
INFC	0.165453	0.0519592	3.1843	0.0020
FFC_1	0.931884	0.0228120	40.8507	0.0000

Media della variabile dipendente	-0.00921613
D.S. della variabile dipendente	0.545118
Somma dei quadrati dei residui	1.11546
Errore standard dei residui ($\hat{\sigma}$)	0.111328
R^2	0.959637
$\bar{R}^2$ corretto	0.958291
$F(3, 90)$	713.247
Statistica Durbin-Watson	0.745132
Coefficiente di autocorrelazione del prim'ordine	0.627353
Statistica h di Durbin	6.20191
Log-verosimiglianza	75.0192
Criterio di informazione di Akaike	-142.03
Criterio bayesiano di Schwarz	-131.86
Criterio di Hannan-Quinn	-137.92

- Imagine now that we estimate by OLS (on the same period, same data) **the same model**, but written (**re-parameterized**) as

$$R_i - R_{i-1} = \rho R_{i-1} - R_{i-1} + \beta_1 b_i + \beta_2 \pi_i + u_i$$

$$\Delta R_i = \underbrace{(\rho - 1)}_{\alpha} R_{i-1} + \beta_1 b_i + \beta_2 \pi_i + u_i$$

Modello 3: Stime OLS usando le 94 osservazioni 1985:2–2008:3
Variabile dipendente: d_FFC

Variabile	Coefficiente	Errore Std.	Statistica t	p-value
const	-0.0161560	0.0114880	-1.4063	0.1631
YGAPC	0.0301904	0.00792405	3.8100	0.0003
INFC	0.165453	0.0519592	3.1843	0.0020
FFC_1	-0.0681159	0.0228120	-2.9860	0.0036

Media della variabile dipendente	-0.0173404
D.S. della variabile dipendente	0.122765
Somma dei quadrati dei residui	1.11546
Errore standard dei residui ($\hat{\sigma}$)	0.111328
R^2	0.204175
$\bar{R}^2$ corretto	0.177647
$F(3, 90)$	7.69672
p-value per $F()$	0.000123542
Statistica Durbin-Watson	0.745132
Coefficiente di autocorrelazione del prim'ordine	0.627353
Log-verosimiglianza	75.0192
Criterio di informazione di Akaike	-142.03
Criterio bayesiano di Schwarz	-131.86
Criterio di Hannan-Quinn	-137.92

- How do we evaluate **the data adequacy** of the estimated model ?
- The level of R-squared changes with the chosen parameterization !
- Evaluating the data adequacy of a model is a critical process more involving than simply looking at the level of R-squared.
- Econometricians call this process: "**analysis of the correct specification**" or "**diagnostic analysis**".

- **To sum up:**
  - The objective of this course is to provide methods for the **quantitative treatment of economic phenomena investigated over time**.
  - The variable of interest (endogenous) can be a scalar or a vector. In the first case we deal with an **uni-equational model**. In the second case we deal with a **system of equations**.
  - An econometric model typically summarizes two aspects: **1. the theoretical view** about the phenomenon under study; 2. **the available data**.
  - The parameters of the econometric models usually capture the behaviour of economic agents or policy institutions, or can be expressed a function of these parameters. **Structural parameters**. **They are the object of inference.**

- **To sum up:**
  - Assuming (at least temporarily) that the model is correctly specified, **does the applied estimator is 'the right one'** I can select amongst a range of alternatives ? (Does it exist 'the best' estimator for the problem at hand ?);
  - Before we can use an estimated econometric model for interpreting, forecasting, testing hypotheses of interest and taking policy actions, we have to make sure that the model is able to account for important aspects of the data. **The analysis of the correct specification (diagnostic analysis) plays a crucial role.**
  - The evaluation of the data adequacy (it is not only 'the goodness of fit') of the econometric model is a non-trivial process which requires addressing many questions.