tbs EDUCATION
Toulouse Business School

MSc AIBA

UE 1 - Analytical Theory, Methods and Models

Econometrics and Statistical Models

Time series and business data

General Presentation

Anne VANHEMS
a.vanhems@tbs-education.fr

EFMD EQUIS ACCREDITED · AACSB ACCREDITED · Accredited by Association of MBAs

www.tbs-education.fr

General organisation

- **Econometrics and statistical models**
  - ➔ From October 24 to October 28

- **Time series and business data**
  - ➔ From January 11 to January 13

Both courses use R software

Joint evaluation

- **A single project by groups of 4 students**
  - ➔ Topic/dataset related to one of the 2 courses.

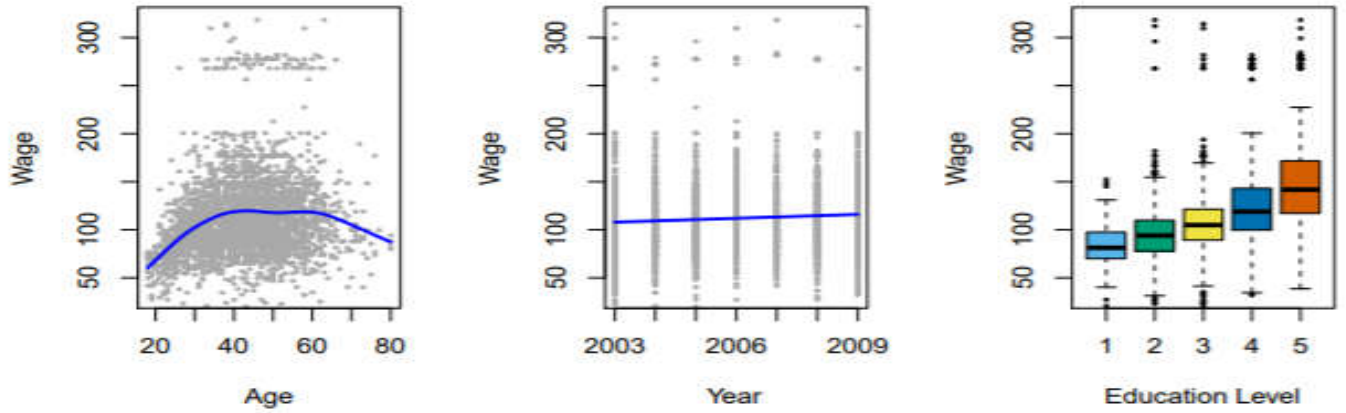
- **Two individual evaluations (MCQ)**
  - ➔ One for each course
  - ➔ Date to be fixed

Course #1

Econometrics and Statistical Models

Course #1: Econometrics and Statistical Models

- Statistical learning: inference and prediction of an *output variable* Y given a set of *input variables* $X_1, \ldots, X_k$



Income survey data for males from the central Atlantic region of the USA in 2009.

Source: JAMES, Gareth, WITTEN, Daniela, HASTIE, Trevor, *et al. An introduction to statistical learning*. New York : springer, 2013.

The supervised learning problem

Starting point:

- Outcome measurement Y (also called dependent variable, response, target).
- Vector of p predictor measurements X (also called inputs, regressors, covariates, features, independent variables).
- In the *regression problem*, Y is quantitative (e.g price, blood pressure).
- In the *classification problem*, Y takes values in a finite, unordered set (survived/died, digit 0-9, cancer class of tissue sample).
- We have training data $(x_1, y_1), \ldots, (x_N, y_N)$. These are observations (examples, instances) of these measurements.

Objectives

On the basis of the training data we would like to:

- Accurately predict unseen test cases.
- Understand which inputs affect the outcome, and how.
- Assess the quality of our predictions and inferences.

Philosophy

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.
- One has to understand the simpler methods first, in order to grasp the more sophisticated ones.
- It is important to accurately assess the performance of a method, to know how well or how badly it is working [simpler methods often perform as well as fancier ones!]
- This is an exciting research area, having important applications in science, industry and finance.
- Statistical learning is a fundamental ingredient in the training of a modern *data scientist.*

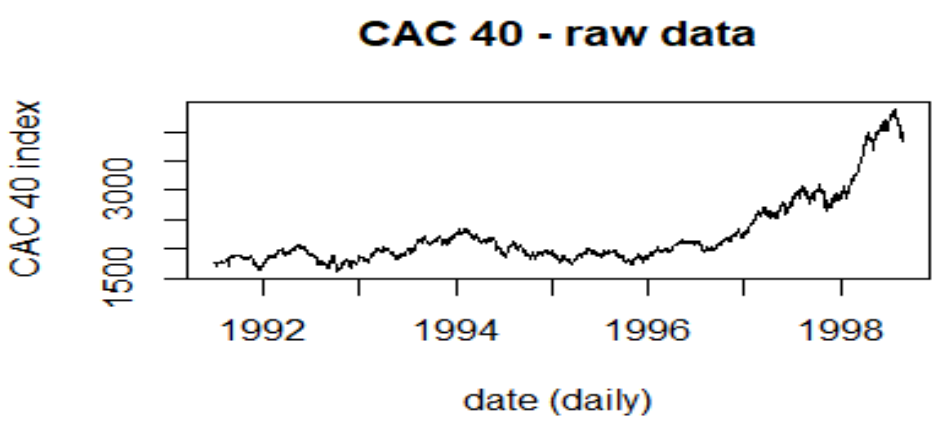
Course #1 content

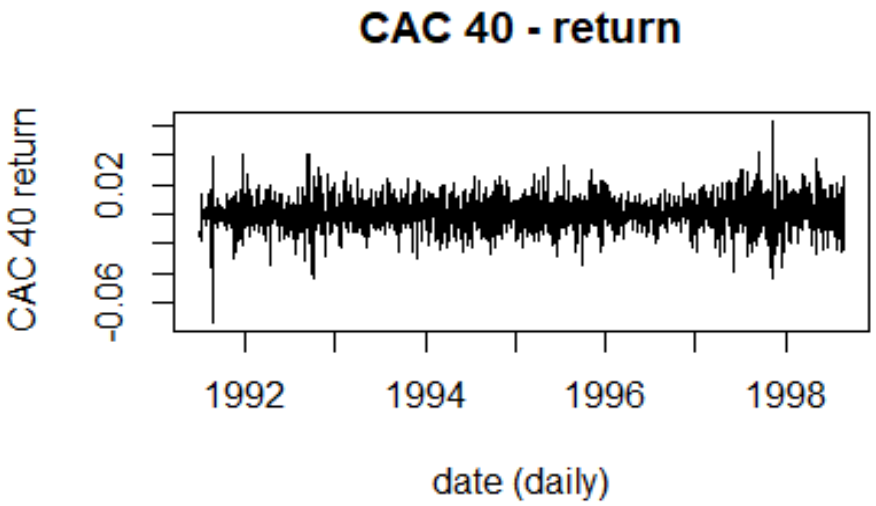
- Class #1: data pre-processing
- Class #2: Linear regression and standard non linear extensions
- Class #3: Polynomial and spline smoothing
- Class #4: Generalized additive models (GAM)
- Class #5: GAMs for classification methods

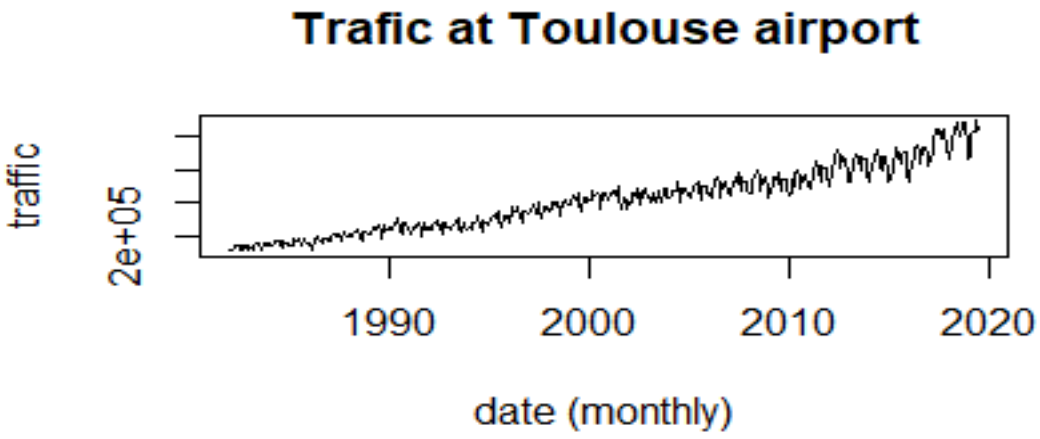
Course #2

Time series and business data

Course #2: Time series and business data







Course #2 objectives:

- Y_t: random variable indexed by time t

Examples: CAC 40, Danone stock, Toulouse airport traffic, sales, turnover

- Frequency can vary: daily, monthly, quaterly, annually, …
- We observe values of Y_t, for t=1, … T
- Objective: build a statistical model in order to forecast values of Y at dates T+1, T+2, T+3, ….

Course #2 content

- **Description** of a time series: trend, seasonality, random component
- **Stationarity**: *crucial assumption* to check before modelling and forecasting
- ***ARMA models***: *standard models* used to estimate and forecast stationary processes
- ***Box-Jenkins methodology***: *practical procedure* to forecast given raw data.

Course #1
vs
Course #2

Course #1 vs Course #2

Course #1:

- building an econometric model, describing relationship between *variable of interest Y* with *other economic quantities X* (***Econometrics and Statistical Models course***)
- (+): gives economic content to predictions
- (-): unless we can predict X, we cannot forecast Y

Course #2:

- different route: a pure *Time Series* approach
- making use of information in past values of a variable Y for forecasting
- (+): one can fit an approximation to almost any time series you like (requires data on time series)
- (-): not embedded within any underlying theoretical model ➔ economic relevance of the chosen model ?

References

- *An introduction to statistical learning*, G. James, D. Witten, T. Hastie, R. Tibshirani. New York : springer, 2013.
- *Data Science and Big Data Analytics, Discovering, Analyzing, Visualizing and Presenting Data*, EMC Education Services, Wiley 2015
- *Applied Econometrics*, D.Asteriou & S.Hall, Palgrave Macmillan 2015
- *Introduction to Econometrics*, 3rd Edition, J. H. Stock & M.W. Watson, Pearson 2011