

# Master 2 “Economics”

Track Econometrics, Big Data, Statistics  
Magistère option

- Field: Economics
- Duration of studies: 1 year
- Diploma: Master and Magistère
- Language: English

Aix-Marseille School of Economics  
is a department of the Faculty of Economics and Management

# SUMMARY

## TERM 1

Predictive methods .....	2
Machine learning and statistical learning .....	4
Non-parametric methods in econometrics .....	5
Automatic model selection methods .....	6
Methodology of econometrics and statistical studies .....	7
Advanced econometrics .....	8
Programming for big data (Python, SQL, noSQL, etc) .....	9
Softwares for big data .....	10
Big data and economics.....	11
Big data and quantitative marketing.....	12
Big data and finance .....	13
Big data: other applications .....	15
End-of-study project.....	16
IT tools for big data, a deeper view.....	17
Advanced machine learning .....	18

## TERM 2

Transition and duration models .....	16
Models for truncated and censored variables .....	20
Multivariate and nonlinear time series .....	21
Managing big data with SAS.....	22
Hands-on project.....	23
End-of-study internship with report and defence	

## PREDICTIVE METHODS - 24H

### TEACHER

- Pierre MICHEL - [pierre.michel@univ-amu.fr](mailto:pierre.michel@univ-amu.fr)

### CONTENT

This course introduces the main predictive methods based on machine learning. The main objective is to provide students with the knowledge necessary to understand supervised and unsupervised learning methods.

#### Course outline:

1. Introduction
  - a) Introduction to Machine Learning
  - b) Reminders of linear algebra
  - c) Supervised Learning (parametric methods: Linear Regression, Batch Gradient Descent, Stochastic Gradient Descent)
2. Supervised Learning
  - a) Introduction to Kaggle
  - b) Linear algebra and normal equation
  - c) Code vectorization
3. Classification methods
  - a) Parametric methods (logistic and softmax regression)
  - b) Classification and regression trees (CART)
  - c) Advanced tree-based methods (Bagging, Random Forest, Boosting) and variable importance
4. Neural networks
  - a) Introduction to neural networks
  - b) Training a neural network
  - c) Deep learning in Python (keras)
5. Convolutional neural networks (CNN)
  - a) Introduction to CNN
  - b) Training a CNN
  - c) Examples with keras
6. Unsupervised learning
  - a) Complements on unsupervised learning (clustering versus density estimation)
  - b) Clustering methods (k-mans, HCA, DBSCAN)
  - c) Recent approaches in clustering (CUBT, LDA)

*Please note that other methods will be taught in Ewen Gallic's course.*

### PROFESSIONAL SKILLS

- Understand how supervised and unsupervised methods work
- Understand the theory (linear algebra and optimization) underlying those methods
- Being able to code (in Python or R – choice up to students) some algorithms from scratch

### ORGANIZATION

- Sessions alternating theoretical presentations and applications.
- Hands-on mini-projects on toys datasets (Housing, MNIST, Fashion-MNIST, Kaggle datasets, etc).
- The applications will be carried out on computers.

### BIBLIOGRAPHY AND TEXTBOOKS

- Hastie, Trevor, Robert Tibshirani, et Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 5e éd. New York, NY: Springer-Verlag New York Inc., 2009.
- Geron, Aurelien. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd édition. Sebastopol, CA: O'Reilly Media, Inc, USA, 2019.
- Online Course: CS229 - Machine Learning, Stanford University.

### COURSE LANGUAGE

French  English

### MANDATORY PREREQUISITES

Econometrics, Statistics, Basic Python programming skills (numpy, matplotlib).

**RECOMMENDED PREREQUISITES**

Advanced Python programming skills (numpy, matplotlib, scikit-learn, scipy).

**KEYWORDS**

Statistical Learning; Machine Learning; Supervised and Unsupervised Learning; Python; Big Data; Kaggle.

*Updated on April 2021*

# MACHINE LEARNING AND STATISTICAL LEARNING - 24H

## TEACHER

- Ewen GALLIC - [ewen.gallic@univ-amu.fr](mailto:ewen.gallic@univ-amu.fr)

## CONTENT

This course provides a broad introduction to statistical learning and machine learning. The main objective is to provide students with the knowledge necessary to understand machine learning methods.

### Course outline:

Part 1. Introduction: Statistical learning and machine learning: what and why?

- 1.1. Definitions
- 1.2. Initial concepts
  - 1.2.1. The estimation of  $f$
  - 1.2.2. Accuracy vs interpretability
  - 1.2.3. Accuracy
  - 1.2.4. Curse of dimensionality

Part 2. Supervised Methods: Regression models

- 2.1. Linear Regression
- 2.2. Quantile Regression
- 2.3. Exercises

Part 3. Supervised learning: Classification

- 3.1. Logistic regression + Exercises
- 3.2. KNN + Exercises
- 3.3. Linear Discriminant Analysis
- 3.4. SVM
- 3.5. Dealing with images
- 3.6. Unboxing the black box (Shap Values)

Part 4. Conclusion

Machine Learning and ethics

*Please note that unsupervised methods will be taught in Pierre Michel's course.*

## PROFESSIONAL SKILLS

- Understand how machine learning algorithms work.
- Understand the basic theory underlying machine learning.
- Being able to code (in Python or R - choice up to students) simple automatic learning algorithms.

## BIBLIOGRAPHY AND TEXTBOOKS

- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). i. New York: springer. doi:10.1007/978-1-4614-7138-7.
- Berk, R. A. (2016). Statistical Learning from a Regression Perspective. Springer Texts in Statistics. doi:10.1007/978-3-319-44048-4.
- Murphy, K. (2012). Machine Learning: A Probabilistic Perspective. The MIT Press. ISBN: ISBN: 9780262018029.

## ORGANIZATION

- Sessions alternating theoretical presentations and applications.
- The applications will be carried out on computers.

## COURSE LANGUAGE

French  English

## MANDATORY PREREQUISITES

Econometrics, Basic Python programming.

## KEYWORDS

Statistical Learning; Machine Learning; Python; R; Big Data.

*Updated on April 2021*

## NON-PARAMETRIC METHODS IN ECONOMETRICS - 24H

### TEACHER

- Costin PROTOPOPESCU - [costin.protopopescu@univ-amu.fr](mailto:costin.protopopescu@univ-amu.fr)

### CONTENT

Non-Parametric methods are statistical techniques that do not require to specify functional forms for objects being estimated. Instead, they let the data itself play and informs the resulting model in a particular manner. Such methods are becoming increasingly popular for applied data analysis, they are best suited to situations involving large data sets for which the number of variables involved is manageable. These methods are often deployed after common parametric specifications are found to be unsuitable for the problem at hand, particularly when formal rejection of a parametric model based on specification tests yields no clues as to the direction in which to search for an improved parametric model.

The job market understood the importance of the non/semi-parametric methods and almost any serious software contains the principal techniques in this area. We illustrate the different models and techniques with R and Matlab. First, R because of the huge number of packages from CRAN, and secondly Matlab because is the easiest environment for programming arrays in econometrics (and typically all objects are arrays in applied econometrics). Both are very representative for the job market. Each lecture will be accompanied by numerical examples and small programming tutorials.

### Course outline:

- Brief Review of Statistics and Probability Concepts used in the course (pre-requisites)
- Empirical Non-Parametric estimators: histogram, empirical distribution, regressogram
- Kernel-based methods in Non-Parametric Econometrics: Parzen-Rosenblatt, Nadaraya-Watson
- Semi-parametric methods: Single index and Additive models
- Splines-based methods in Non-Parametric Econometrics: Smoothing splines and Least-Squares splines
- Series-based methods in Non-Parametric Econometrics: Wavelets, Laguerre and Hermite Polynomials

### PROFESSIONAL SKILLS

- To master the concepts specific to nonparametric methods for modelling the conditional expectation and the conditional variance.
- Estimation and forecasting in i.i.d and non i.i.d. simulated or real data sets.
- Be able to use the nonparametric estimation for driving more structural models.
- To use the basic packages from R and the toolboxes from Matlab for nonparametric and semiparametric modelling.

### BIBLIOGRAPHY AND TEXTBOOKS

Given the wide content of the course, there is no single textbook for this course, therefore the lecture handouts are self-contained. However, the content is covered in some classic textbooks, such as:

- Jeff RACINE: "Nonparametric Econometrics, A Primer", Foundations and Trends in Econometrics, Vol. 3, No 1 (2008).
- Adrian PAGAN and Aman ULLAH, "Nonparametric Econometrics", Cambridge University Press, (1999).
- B. P. SILVERMAN: "Density Estimation for Statistics and Data Analysis", Chapman Hall, (1986).
- Wolfgang HÄRDLE: "Applied Nonparametric Regression", Cambridge University Press; Revised edition (1992).
- Denis BOSQ: "Nonparametric Statistics for Stochastic Processes: Estimation And Prediction", Springer-Verlag New York Inc, 2nd ed. (2013).
- Ziyue LIU and Wensheng GUO: "Data Driven Adaptive Spline Smoothing", Statistica Sinica 20 (2010).

### ORGANIZATION

- Each session is accompanied by numerical and empirical examples in R and Matlab

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Foundations in Statistics and Probability - Basic differential and integral calculus.

### RECOMMENDED PREREQUISITES

Basic skills for doing econometrics in any language (R/Matlab/STAT/SAS/Ox/Eviews) - Basic knowledge of statistical inference in econometrics.

*Updated on April 2021*

## AUTOMATIC MODEL SELECTION METHODS - 24H

### TEACHER

- Sullivan HUÉ - [sullivan.hue@univ-amu.fr](mailto:sullivan.hue@univ-amu.fr)

### CONTENT

The objective of this course is to introduce quantitative methods allowing to reduce information. These methods cover different fields of statistics and are based on classical econometric methods (OLS, MLE) or classificatory or principal component methods. The goal is to study methods to do automatic variable selection in large-scale problems and to apply them to real data.

### Course outline:

- Classification methods
- Economic factor models
- Statistical factor models
- Lasso methods
- The so-called « General to Specific » method (Hendry, Gets or Autometrics Methodology)

### PROFESSIONAL SKILLS

- Understanding new methods
- Application on real data
- Learning new tools or econometric softwares dedicated to the reduction of information

### BIBLIOGRAPHY AND TEXTBOOKS

- Doornik, J.A. and Hendry, D.F. (2015). Statistical model selection with “Big Data”, Cogent Economics & Finance, vol 3, n°1, 1-15.
- Hendry, D.F. and Doornik, J.A. (2014). Empirical Model Discovery and Theory Evaluation. Automatic Selection Methods in Econometrics. The MIT Press.
- Richard A. Johnson and Dean W. Wichern, Applied Multivariate Statistical Analysis, Pearson.

### ORGANIZATION

Lecture.

### COURSE LANGUAGE

French  English

### MANDATORY PREREQUISITES

- Advanced statistics
- Introduction to econometrics

### RECOMMENDED PREREQUISITES

Programming (R or Ox)

### KEYWORD

Reduction of information

*Updated on April 2021*

## METHODOLOGY OF ECONOMETRICS AND STATISTICAL STUDIES - 24H

### TEACHER

- PhD Supagro/EDEG Montpellier & INRAE (professional)

### CONTENT

Provide students with a set of rules to be followed in the course of realization of a statistical or econometric study for an organization (be it a private company, a public or private administration, a NGO, etc.).

### Course outline:

- Starting from the beginning: calls for tender, mutual agreement contracts and within organization projects.
- Identifying the question(s) to be answered. Modelling issues.
- Going from the economic model to the econometric model.
- Collecting, inspecting and managing the data.
- Choosing the right statistical or econometric estimation method(s).
- Analysing the results and their consequences in terms of decision/policy. Professional ethics.
- Communicating about the results in oral and written forms.

### PROFESSIONAL SKILLS

Ability to conduct a statistical or econometric study in a professional way, from the start (answers to calls for tender) to the end (communication about the results).

### ORGANIZATION

8 sessions of 3 hours each including a personalized supervision of students about one econometric project they have to produce in (an)other course(s).

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Basics of linear and non linear econometrics.

Mastering of at least one econometric software (SAS, Stata, R, etc.)

*Updated on April 2021*



## ADVANCED ECONOMETRICS - 24H

### TEACHER

- Emmanuel FLACHAIRE - [emmanuel.flachaire@univ-amu.fr](mailto:emmanuel.flachaire@univ-amu.fr)

### CONTENT

The goal of this course is to present advanced methods in econometrics for distributional analysis, regression and classification models. The course will present theoretical foundations and underlying intuition of each method, as well as several empirical examples.

#### Course outline:

1. Resampling Methods
  - Pseudo-random generator
  - Monte Carlo experiments
  - Bootstrap and permutation tests
2. Nonparametric Econometrics
  - Density estimation
  - Regression splines
  - Finite mixture models
3. Econometrics and Machine Learning
  - Philosophy and general principle
  - Resampling-based methods and algorithms
  - Misspecification detection

#### BIBLIOGRAPHY AND TEXTBOOKS

- Ahmada et Flachaire (2011) Non-Parametric Econometrics, Oxford University Press.
- Efron et Hastie (2016) Computer Age Statistical Inference, Cambridge University Press.

#### COURSE LANGUAGE

French       English

*Updated on April 2021*

## PROGRAMMING FOR BIG DATA (PYTHON, SQL, NOSQL, ETC) - 24H

### TEACHERS

- Airbus Helicopters (professional)
- CCI Aix Marseille Provence (professional)

### CONTENT

This course is aimed at teaching the basics of computer programming, with emphasis on its use in Big Data. Students will first become familiar with database management. They will then learn the basics of programming with the computer language Python.

### Course outline:

1. Database Management (relational model, relational algebra, SQL language...)
2. Introduction to Python
3. Creating Functions
4. Introduction to Numpy
5. Data manipulation with Pandas
6. Visualization
7. Parallel programming

### PROFESSIONAL SKILLS

- Being able to create a database and query it to extract information.
- Knowing how to load data into Python, manipulate it and format it.
- Being able to understand and use functions from Python libraries.
- Acquire knowledge on distributed computing.

### BIBLIOGRAPHY AND TEXTBOOKS

- Rockoff, L. (2016). The language of SQL. Addison-Wesley Professional.
- VanderPlas, J. (2016). A Whirlwind Tour of Python. O'Reilly Media, Inc.
- Lutz, M. (2013). Learning Python: Powerful Object-Oriented Programming. O'Reilly Media, Inc.
- <https://www.python.org/>

### ORGANIZATION

Sessions alternating theoretical presentations and applications.  
The applications will be carried out on computers.

### COURSE LANGUAGE

French  English

*Updated on April 2021*

## SOFTWARES FOR BIG DATA - 24H

### TEACHER

- CCI Aix Marseille Provence (professional)

### CONTENT

The objective of this course is to teach Python and its most used libraries for data analysis and data visualization. It aims to make students masters in data handling, manipulation, and visualization.

#### Course outline:

- Basics of web: HTML, CSS, Bootstrap, deployment
- Web scraping using Python and its limits
- Data handling using Pandas
- Data visualization using Python, best practices
- Building interactive and dynamic dashboards using Dash from Plotly
- Git, GitHub, GitLab for code management
- Data Analytics Softwares

Final project: build your own interactive dashboard.

### PROFESSIONAL SKILLS

- Master data handling, data manipulation and data visualization using Python and scientific libraries
- Be able to design dashboards and to extract key indicators from data
- Understand the importance of Git and git based platforms

### ORGANIZATION

Lecture. Evaluation based on a group project and multiple short individual projects. Build an interactive dashboard, oral presentation.

### COURSE LANGUAGE

French  English

### MANDATORY PREREQUISITES

- Programming skills (Python / R)
- Database skills

### KEYWORDS

Programming language, data handling, data visualization, data manipulation, Python.

*Updated on April 2020*

## BIG DATA AND ECONOMICS - 24H

### TEACHERS

- Assistance publique Hôpitaux de Paris (professional)
- ECJA law firm (professional)

### CONTENT

The objective is to introduce several questions associated with the use of big data in order to understand the context, impacts and limits of this new technology. Particular issues will be presented to students: e.g. access, security, innovation and opportunities in order to specifically identify how big data can boost local development, and allow to design new public policies.

### Course outline:

Part I: Data and local development

1. Open data, Smart Region and innovation
2. Contracts and legal aspect of the platform FlexGrid, a tool for energy transition

Part II: Big data and security

1. Public security and cities security
2. Cybersecurity

Part II: Big data, a decision-making support

### PROFESSIONAL SKILLS

- Analyze a specific issue,
- Identify the means of data,
- Identify how to use a particular tool to a specific question in order to improve a decision,
- General knowledge on legal and economic aspect associated with the use of big data.

### ORGANIZATION

The course spans 24h, and different professional speakers will introduce the specific topics, either by slots of 3h, or an all-day lecture.

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Public economics, economic analysis.

### RECOMMENDED PREREQUISITES

Big data initiation.

*Updated on April 2021*

## BIG DATA AND QUANTITATIVE MARKETING - 24H

### TEACHERS

- IPSOS (professional)

### CONTENT

Understand how data analytics, machine and deep learning methods on large volumes of structured or unstructured data allow to better model, predict or describe consumer behaviour. Understand how data analytics-based modelling impacts marketing decisions and how recent advances are changing market research and marketing science.

#### Course outline:

- The 3 main uses of data in marketing analytics:
  - Model and describe behaviours / segment consumers
  - Predict behaviours based on a set of variables
  - Receive real-time information on consumer behaviours
- The types of data used in quantitative marketing:
  - Aggregate statistical/econometric data and time series
  - Individual observations in data sets (survey data, databases, etc.)
  - User generated content (text and images)
  - Datafied objects
    - How data is changing and why: from solicited to unsolicited, from structured to unstructured, from manual datafication to algorithm-led datafication, from text to data, etc.
    - The main statistical methods used in data analytics
    - The code languages used in advanced data analytics
    - Algorithms, machine learning, deep learning and artificial intelligence

### PROFESSIONAL SKILLS

- Statistics
- Econometrics
- Computer science / coding languages

### BIBLIOGRAPHY AND TEXTBOOKS

<https://www.ipsos.com/en-us/ipsos-science-center>

- Davy Cielen and Arno D.B. Meysman, 2016: Introducing Data Science: Big Data, Machine Learning, and more, using Python Tools.
- Andrew McAfee and Erik Brynjolfsson in Harvard Business Review, October 2012: Big Data: The Management Revolution.

### ORGANIZATION

Case studies.

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Curiosity.

### RECOMMENDED PREREQUISITES

Statistics, Econometrics, R, Python, SPSS.

*Updated on April 2021*

## BIG DATA AND FINANCE - 24H

### TEACHERS

- Crédit Agricole CACIB (professional)
- WeeFin (professional)

### CONTENT

This lecture exposes some recent developments of “big data analytics” in market finance. It is formalized and proposes some practical examples. There will be five sections. The first reminds some stochastic calculus definitions. The second aims at to present the link between integrated variance, quadratic variation and realized variance. Then, we present several estimators for high frequency realized volatility/correlation and we explain why it is crucial for a option market maker to use this information to adjust his option prices with respect to a schedule of future events (example FX market). The next section presents an engine to select optimal government bonds to hedge a portfolio. It is based on a relative value analysis with respect to a cubic spline theoretical yield curve. For corporate bonds, the dimension increases severely. We then have to use some clustering methods (section four) to reduce the dimension. A method for optimal bond index replication in high dimension is proposed in section 5.

### Course outline:

1. Update of Stochastic Calculus
  - a. Brownian motion and random walk
  - b. Quadratic variation and integrated variance for a geometric Brownian motion
  - c. Link between integrated variance and realized volatility
2. Realized volatility/correlation
  - a. Estimators for intraday and non-synchronous data
  - b. Realized correlation and quadratic covariation, estimator
  - c. Epps effect
  - d. Application to FX market to evaluate the impact of some periodic events
3. Bonds portfolio, automatic engine
  - a. Definitions, pricing (yield, bond, duration)
  - b. Use of a reference yield curve to detect the best bond to select for hedging purpose
  - c. Cubic spline for the reference curve
  - d. Comparison metrics
  - e. Application to government bonds
4. Unsupervised clustering methods
  - a. K-means
  - b. K-means++
  - c. Model based (optimal K, penalization)
  - d. EM algorithm
5. Problem of dimension for corporate bond index replication
  - a. Bloomberg-Barclays indexes
  - b. Ill conditioned correlation matrix
  - c. Deep Embedded Clustering
  - d. Optimal replication in high dimension

### PROFESSIONAL SKILLS

- Financial markets big data analytics.
- Quantitative analysis, Strat.

### BIBLIOGRAPHY AND TEXTBOOKS

- Brigo, D. and Mercurio, F., 'Interest Rate Models - Theory and Practice', 1998, 981p, Springer.
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, 'An Introduction to Statistical Learning with Applications in R', 2013, Springer.

### ORGANIZATION

The whole lecture last for 12h. Slides available.

### COURSE LANGUAGE

- French  English

### RECOMMENDED PREREQUISITES

Financial econometrics, options, numerical optimization.

**KEYWORDS**

Realized variance, intraday implied volatility, factorial analysis, clustering, bonds, bias/variance decomposition.

*Updated on October 2020*

## BIG DATA: OTHER APPLICATIONS - 24H

### TEACHERS

- Airbus Helicopters (professional)
- EURA NOVA - GlaxoSmithKline (GSK) Belgique (professional)

### CONTENT

### PROFESSIONAL SKILL(S)

### BIBLIOGRAPHY AND TEXTBOOKS

### ORGANIZATION

### COURSE LANGUAGE

- French       English

### RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*



## END-OF-STUDY PROJECT

TEACHER

-

CONTENT

PROFESSIONAL SKILL(S)

BIBLIOGRAPHY AND TEXTBOOKS

ORGANIZATION

COURSE LANGUAGE

French  English

RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*

## IT TOOLS FOR BIG DATA, A DEEPER VIEW - 24H

### TEACHER

- EQUANCY (professional)

### CONTENT

#### Course outline:

- Hadoop. HDFS. MapReduce. Stockage et calculs distribués. Déploiement d'un cluster.
- Préparation, stockage et traitement des big data : Pandas, Hive and Pig
- Data visualisation avec matplotlib & seaborn
- Alternatives : solutions propriétaires, bases NoSQL, ElasticSearch

### PROFESSIONAL SKILL(S)

### BIBLIOGRAPHY AND TEXTBOOKS

### ORGANIZATION

### COURSE LANGUAGE

French  English

### RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*

## ADVANCED MACHINE LEARNING - 24H

### TEACHER

- Badih GHATTAS - [badih.ghattas@univ-amu.fr](mailto:badih.ghattas@univ-amu.fr)

### CONTENT

#### Course outline:

- Advanced feature extraction (TF\_IDF, Word2Vec, etc.)
- Deep learning (TensorFlow, Theano)
- Text Mining

### PROFESSIONAL SKILL(S)

### BIBLIOGRAPHY AND TEXTBOOKS

### ORGANIZATION

### COURSE LANGUAGE

French  English

### RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*

## TRANSITION AND DURATION MODELS - 24H

### TEACHER

- Christian SCHLUTER - [christian.schluter@univ-amu.fr](mailto:christian.schluter@univ-amu.fr)

### CONTENT

Students will study models of transitions and durations, and learn how to estimate these using real-world data.

#### Course outline:

This course is an introduction to modelling transitions into a state of interest (such as the transition into employment from unemployment) and durations (such as unemployment, survival of patients after medical treatment or firms after a financial crash). We start with the basic building blocks (Poisson processes, Markovian transitions, hazard models), and then develop methods for estimation using Maximum Likelihood. Duration data might be incomplete (hence are censored) in that we might not observe exits (individuals might still be in the state of interest at the end of the observation window), and unobserved heterogeneity introduces fundamental identification challenges. Throughout all methods will be illustrated using examples in the software R, and we will consider several articles that have applied these methods. Several exercise sets will help students deepen their understanding of the theory.

#### (I) Introduction to Poisson and counting processes

The Poisson process is the classic counting process that models the arrival of new events, and thus transitions (increments) and durations (inter-arrival times). We will study this model, which has several interesting features, such as the independence of increments (transitions), so the Poisson process is a special Markov process (which exhibits a lack of memory property). Application & illustrations in R: The number of doctor visits following a major health care reform (Poisson regressions). Search models of unemployment.

#### (II) Introduction to Markov processes

The Poisson process has increments that are independent, hence satisfy the Markov (lack of memory) property. We generalise this idea, consider transition probabilities between states, and look at the evolution of the Markov process in time. Applications & illustrations: We will study numerical examples using R, and several applications such as Nakajima (2007, ReStud), "Measuring Peer Effects on Youth Smoking Behaviour", and Topa (2001, ReStud), "Social Interactions, Local Spillovers and Unemployment."

#### (III) Duration and survival analysis: Hazard models

The Poisson model gives a simple duration model, but in many empirical situations this is too limiting as exit rates (hazards) form the state of interest are constant. In many situations the exit rate depends on the duration of stay, for instance the longer the unemployment spell the less likely the exit (duration dependence). We will study the main modelling objects (hazard rate and survival function), examine several parametric models (e.g. the Weibull model), and incorporate observed heterogeneity across individuals (Cox's proportional hazard (PH) model). In order to accommodate unmeasured heterogeneity, we extend the PH model to the mixed proportional hazard (MPH) model. This unobserved heterogeneity will introduce a fundamental identification problem since duration dependence might be confounded by dynamic sorting (individuals of a latent "high" type tending to exit the state of interest more quickly). Since the models are fully parametrically specified, it is natural to estimate these using Maximum Likelihood. One particular characteristic of duration data is that we might not observe an individual's exit from the state of interest; hence these duration data might be incomplete (hence are censored), and this needs to be taken into account in the estimation. Applications in R: Survival probabilities for smokers, criminal recidivism.

### PROFESSIONAL SKILLS

Master the theory of transition and duration models, learn how to estimate such models in practicing real-world data, understand and address the empirical challenges that arise in empirical work.

### BIBLIOGRAPHY AND TEXTBOOKS

The Lecture handouts are fairly self-contained. Supplementary reading is:

- J. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*.
- van den Berg, G. J. (2001) "Duration models: Specification, identification, and multiple duration". In: Heckman J. J., Leamer E. (eds.), *Handbook of Econometrics*, Vol V, Chapter 55. Amsterdam: North-Holland, 3381–3460.

### ORGANIZATION

Lectures and exercises.

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Basic econometrics.

*Updated on April 2021*

## MODELS FOR TRUNCATED AND CENSORED VARIABLES - 24H

### TEACHER

- Christophe MULLER - [christophe.muller@univ-amu.fr](mailto:christophe.muller@univ-amu.fr)

### CONTENT

The main objective of this course is to provide the students with a synthetic framework so that they can thoroughly understand and efficiently apply to concrete cases the main estimation and test techniques for limited-dependent variable models with data censoring or truncation.

Course outline:

1. Censored and Truncated Data, Limited Dependent Variable Distributions and Selection Bias.
2. Inference with M-estimators
3. Point and Partial Identification
4. The Tobit Model
5. Other Models with Censored or Truncated Dependent Variables
6. Empirics

### PROFESSIONAL SKILLS

The targeted skills are primarily an understanding of the reasoning for justifying and using econometric techniques, and their practical application.

### BIBLIOGRAPHY AND TEXTBOOKS

- Amemiya, T., Advanced Econometrics, Basil Blackwell, 1989.
- Cameron, C. and P. K. Trivedi, Microeconometrics Using Stata, Revised Edition, Stata Press, 2010.
- Davidson, R. and McKinnon, J.G., Estimation and Inference in Econometrics, 1993.
- Davidson, R. and McKinnon, J.G., Econometric Theory and Methods, 2004.
- Gouriéroux, C., Économétrie des Variables Qualitatives, Economica, Paris, 1989.
- Greene, W.H., Econometric Analysis, Prentice Hall, 2018.
- Lollivier, S., Économétrie avancée des variables qualitatives, Economica, Paris, 2016.
- Long, J.S. and J. Freese, Regression Models for Categorical Dependent Variables Using Stata, Stata Press, 2006.
- Maddala, G.S., Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, 1988.
- Manski, C., Identification Problems in the Social Sciences, Harvard University Press, 1995.
- Stata Manual
- Wooldridge, J.W., Introduction Econometrics: A Modern Approach, South-Western College Publishing, 2012.

### ORGANIZATION

- Professorial theoretical lecture and applications
- Written final exam
- A few empirical application sessions on computer

### COURSE LANGUAGE

French  English

### FUNDAMENTAL PREREQUISITES

Linear model econometrics, basic mathematics of the optimisation, linear algebra, Stata software, Probit model.

### KEYWORDS

Econometrics, Truncated dependent variable, Censored Dependent variable, Truncated and Censored Data, Limited Dependent Variable.

*Updated on January 2023*

## MULTIVARIATE AND NONLINEAR TIME SERIES - 24H

### TEACHER

-

### CONTENT

Make the students able to cope with multivariate, nonlinear time series analysis.

#### Course outline:

- Introduction to time series analysis
- Nonlinear models (TAR, STAR, MS, tests, forecasting)
- Multivariate models (VAR, SVAR, cointegration)
- Factor analysis (Macro, fundamental, PCA)
- Multivariate volatility models (VEC, BEKK, DCC, OGARCH)

#### PROFESSIONAL SKILLS

- Command of descriptive analysis tools
- Knowledge of key nonlinear and multivariate models
- Command of diagnosis tools
- Understanding of models' performance in forecasting

#### BIBLIOGRAPHY AND TEXTBOOKS

- Tsay, R. S., Analysis of Financial Time Series, Wiley, ed. 3, 2010.

Autres références :

- Hamilton, J., Time Series Analysis, Princeton University Press, 1994.
- Diebold, F. X., Time Series econometrics, online on author's webpage, 2017.

#### ORGANIZATION

- 8\*3hours with both theory and practice
- Applications on R and OxMetrics
- Mandatory preliminary readings before each session

#### COURSE LANGUAGE

French  English

#### FUNDAMENTAL PREREQUISITES

Statistics. Econometrics.

#### RECOMMENDED PREREQUISITES

Time series analysis.

*Updated on April 2021*

## MANAGING BIG DATA WITH SAS - 24H

### TEACHER

- Freelance in data science (professional)

### CONTENT

#### Course outline:

- SAS and Big Data, an introduction
- The SAS procedures for Big Data (1/2)
- The SAS procedures for Big Data (2/2)

### PROFESSIONAL SKILL(S)

### BIBLIOGRAPHY AND TEXTBOOKS

### ORGANIZATION

### COURSE LANGUAGE

French  English

### RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*

## HANDS-ON PROJECT - 24H

### TEACHER

- EQUANCY (professional)

### CONTENT

### PROFESSIONAL SKILL(S)

### BIBLIOGRAPHY AND TEXTBOOKS

### ORGANIZATION

### COURSE LANGUAGE

French  English

### RECOMMENDED/FUNDAMENTAL PREREQUISITES

*Update in progress*



**Aix-Marseille School of Economics**

AMU-AMSE

5-9 boulevard Maurice Bourdet - CS 50498  
13205 Marseille cedex 1

Tél.: +33.(0)4.13.55.25.23  
infoecole@amse-aixmarseille.fr

[amse-feg.univ-amu.fr](http://amse-feg.univ-amu.fr)  
[www.amse-aixmarseille.fr](http://www.amse-aixmarseille.fr)