Universitat Pompeu Fabra Barcelona — Department of Economics and Business

MASTER OF RESEARCH IN ECONOMICS, FINANCE AND MANAGEMENT

RESEARCH SEMINAR: DATA SCIENCE	31587

COURSE

TERM 1	COMPULSORY	WORKLOAD: 75H

This information is for the 2026/27 session.

Professor

Prof. David Rossell

Teaching guide

Overview

This research seminar forms the first half of a two-course series that complements the follow-up "**Topics in Data Science**". Together, they explore advanced, research-oriented methods for contemporary data analysis in Economics.

Part 1. Research seminar in Data Science – Foundations

Regression settings with high-dimensional parameter spaces, Bayesian inference, and causal inference applications. You'll learn to navigate datasets where the number of predictors can vastly exceed the number of observations, and gain a solid grounding in probabilistic machine learning.

Part 2. Topics in Data Science – State-of-the-art methods

Flexible, cutting-edge techniques: non-linear additive models, neural networks (deep learning), tree-based methods (random forests, boosting), latent variable models for text, and non-parametric approaches to causal inference. Real-world applications will illustrate how these methods help solve economic questions.

Why this matters.

Modern Economics increasingly demands the tools of Data Science and Machine Learning. For example, Nobel laureate Christopher A. Sims has long argued for formulating economic theories as probabilistic models, using Bayesian inference to rigorously evaluate policy uncertainty and causal mechanisms—and these core ideas are intertwined with the methods taught in this sequence[1].

Recent **NBER Summer Institute** events reinforce this shift: the 2025 Methods Lecture by Susan Athey on **bandit designs and policy learning**, and the Frontier Econometric Methods track co-organized by Susan Athey, Isaiah Andrews, and Christopher Walters showcase how causal machine learning is shaping empirical economics[2].

Similarly, the **Chicago Booth Machine Learning in Economics Summer Institute (MLESI 2025)** offers advanced workshops on supervised/unsupervised ML, text analysis, experimental design, fairness, and their intersection with causal inference[3]. Additional examples include the Harvard Economics PhD's "Big Data & ML" syllabus, the Chicago Booth courses on Bayesian inference and AI for Economics, and Bocconi's PhD modules on text analysis, networks, and causal inference.

Data analysis methods are evolving to cope with increasingly challenging problems, a case in point being high-dimensional situations where one considers a large number of parameters or models. For instance, one may have a regression where the number of covariates far exceeds the sample size, or may want to simplify the interpretation of complex data via clustering, text or latent variable analysis, or may want to make predictions in challenging settings using Deep Learning. Engaging effectively in such research, either from a methodological or applied perspective, requires one to understand, and when needed modify or extend, such methodology. Just as importantly, they need to communicate such ideas effectively to a potentially non-expert audience.

Learning outcomes: improved familiarity with selected research topics in Data Science, at a level sufficient to critically appraise, modify and apply novel methods, the ability to carry them out using modern Statistical software, and to communicate their analyses and findings. The course also intends to provide students with applied data analysis skills useful for their MRes thesis and subsequent research work.

Pre-requisites: the course is designed for MRes students who are familiar with basic Statistical inference, specifically linear regression and maximum likelihood estimation. Basic R programming skills are beneficial, though examples and links to learning resources will be provided for students who are not familiar with R.

1https://www.nobelprize.org/prizes/economic-sciences/2011/sims/lecture/

2https://www.nber.org/conferences/si-2025-frontier-econometric-methods

3https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/opportunities/event---aiesi

Content

Introduction and Bayesian inference

We review classical results from maximum likelihood estimation and computational methods such as the bootstrap, and we introduce the basic Bayesian paradigm for statistical inference, its use for model selection, parameter estimation and prediction, and standard computational tools such as Gibbs or Metropolis-Hastings. We learn programming skills in R.

Variable selection in high-dimensional regression

We review the fundamental penalized likelihood and Bayesian frameworks for linear regression with a large number of variables. We shall discuss the relative merits of popular strategies such as LASSO, adaptive LASSO and related penalties to help decouple variable selection from prediction and various Bayesian strategies to achieve good performance in high dimensions. We will discuss some theoretical, practical and computational considerations.

Application to causal inference

We will discuss how to apply the earlier strategies to settings where one has an outcome, some treatments of interest, and a potentially large set of control variables.

References

The books below provide a good introduction to a substantial part of the topics covered in this course (and many others), however we shall complement them with a number of additional selected research manuscripts.

- Andrew Gelman, John B. Carlin, Hals S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin. Bayesian Data Analysis (3rd edition). CRC Press, 2013.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- Trevor Hastie, Robert Tishirani, Martin Wainwright. Statistical learning with sparsity. The LASSO and its generalizations. CRC press.
- Mahlet Tadesse, Marina Vannucci, M. (Eds.). (2021). Handbook of Bayesian variable selection.

Assessment and Grading System

The course will be delivered in a combination of regular lectures, computer-based seminars where students get hands-on experience with the taught data analysis methods, and presentations by students (on published manuscripts chosen by the students and on the final project). The idea is to give students an entry point to understand Data Science, the flexibility to deepen their understanding on the topics they find more interesting / relevant for their own research, and the ability to continue learning about this topic in the future if they so wish.

Assessment and Grading System

Students will be asked to orally present 1 research paper of their choice (40% of the final mark), some selected exercises from the seminar sessions (10% of final mark) and a written report (50% of the final mark). This project will be decided by the students but must be pre-approved by the lecturer, and should involve the application, critical assessment or extension of the research methods seen in class. The content can be theoretical, empirical, a practical application or a combination of the former. Students who also take the “Topics in Data Science” course can submit a single joint project for both courses, specific instructions on the format will be given in class.

Attendance is a compulsory requirement for this course. In the event of an absence, students must communicate it to the professor(s) and provide an appropriate justification. Unjustified absences may ultimately imply that a student cannot be evaluated for the course.