Quantitative Methods II

Professors:
Brenda van Coppenolle brenda.vancoppenolle@sciencespo.fr
Jan Rovny jan.rovny@sciencespo.fr

TA:
Luis Sattelmayer luis.sattelmayer@sciencespo.fr

Course Description

This is an intermediate level course of quantitative methods for social scientists. It builds on Quantitative methods I, in which students learn descriptive statistics, the logic of inference, and OLS regression analysis, as well as basic programming in R. This course, aims to extend this knowledge by introducing the logic of maximum likelihood estimation and causal inference. The course applies MLE to a broad spectrum of practical questions, particularly to the analysis of categorical or ordinal dependent variables, time-series analysis, multi-level modeling, and data reduction techniques. The course also focuses on basics of causal inference. The course is based on a mixture of lectures and practical lab sessions. Students are thus expected to attend the lectures and lab sessions where they will work on practical problem sets. These problem sets will be competed at home as activities. Besides these problem sets and a final exam, the students will write a final project applying the methods learned in class to a research topic of their interest.

Validation

- **Lab acitvities 30%** - continuous assignments throughout the semester
- **Final exam 20%** - short online exam to check progress
- **Final paper 50%** - quantitative analysis of a social science research question. It must use quantitative data to address a substantive issue, using one of the methods learned this semester in Methods 2 - maximum likelihood estimation, multi-level modeling, methods of causal identification (i.e. RDD, IV or Diff-in-Diff), or time-series analysis

Readings

- Long, Scot (1997). *Regression Models for Categorical and Limited Dependent Variables.* Sage
- Gujarati, Damodar 2003. *Basic Econometrics.* McGraw Hill
- Angrist, Joshua D. and Pischke, Jorn-Steffen. (2015) *Mastering Metrics*, Princeton and Oxford: Princeton University Press
- Angrist, Joshua D. and Pischke, Jorn-Steffen. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton and Oxford: Princeton University Press

Course Outline

1: Research Design

Reading:

(*) Angrist and Pischke (2015) Mastering Metrics: Introduction xi - xv, Chapter 1
Angrist and Pischke (2009) Mostly Harmless: Chapter 1 and 2
In this introductory class, we introduce the logic of potential outcomes, the concept of counterfactuals, and we review the logic of regression (see Quantitative Methods I). The goal of this class is to understand the fundamental problem of causal inference as a research challenge. Good research design, in combination with the methods introduced in this class, is crucial to help address the fundamental research challenge.

2: Logic of Maximum Likelihood Estimation (MLE)
Reading: Long pp.25-33
This class introduces the logic of maximum likelihood estimate. It begins with the concept of likelihood, derives the likelihood function, and finally demonstrates how to estimate linear relationships using maximum likelihood.

3: Application of MLE: Logit / Probit
Reading: Gujarati, Chapter 15 "Qualitative Response Regression Models" (Linear Probability Model, Logit and Probit models)
This class turns to the practical application of MLE and introduces the logic of binary logit and probit.

4: Application of MLE: Ordinal Logit / Multinomial Logit
To further expand our MLE tools, in this class we consider how to model ordinal and nominal dependent variables. We focus on the modeling of vote choice.

5: Common Problems, Common Solutions, and Multi-Level Models
Reading: (*) Angrist and Pischke (2015) Mastering Metrics: Chapter 2
Angrist and Pischke (2009) Mostly Harmless: Chapter 3
In this class, we discuss how most common problems with regression can be understood as problems with the model's assumption about the distribution of the error term. We discuss some of the most common solutions, including interactions, polynomials and transformations of the dependent variable. We also introduce the concept of panel data, and of multi-level models that use the time-dimension, or repeated observations for the same unit, to gain traction on the identification of causality, within nested or hierarchical models. Beware of omitted variable bias.

6: Differences-in-Differences
Reading: (*) Angrist and Pischke (2015) Mastering Metrics: Chapter 5
Angrist and Pischke (2009) Mostly Harmless: Chapters 5, and 8
Sometimes we are lucky enough to be able to observe treatment and control groups over time. Comparing the differences before and after treatment allows controlling for unobserved omitted variables that are fixed over time. We introduce the assumptions that need to be fulfilled for estimates to be credibly causally identified. We will also discuss an extensive example of such a research design.

7: Instrumental Variables
Reading: (*) Angrist and Pischke (2015) Mastering Metrics: Chapter 3
Angrist and Pischke (2009) Mostly Harmless: Chapter 4
Sometimes there are unobserved confounders that introduce omitted variable bias. Yet sometimes we are lucky enough to be able to use an instrument to deal with this problem. We introduce the assumptions that need to be fulfilled for estimates to be credibly

causally identified. We will also discuss an extensive example of such a research design.

8: Regression Discontinuity Design 1
Reading: (*) Angrist and Pischke (2015) Mastering Metrics: Chapter 4
Angrist and Pischke (2009) Mostly Harmless: Chapter 6
Sometimes there are unobserved confounders that introduce omitted variable bias. Yet sometimes we are lucky enough that the treatment is delivered along thresholds, which allows us to deal with this problem. In this class we discuss such 'sharp' RDD estimation techniques. We introduce the assumptions that need to be fulfilled for estimates to be credibly causally identified. We will also discuss an extensive example of such a research design.

9: Regression Discontinuity Design 2
Reading: (*) Angrist and Pischke (2015) Mastering Metrics: Chapters 4, and 6
Angrist and Pischke (2009) Mostly Harmless: Chapter 6
Sometimes there are unobserved confounders that introduce omitted variable bias. Yet sometimes we are lucky enough that the treatment is delivered -imperfectly- along thresholds, which allows us to deal with this problem. In this class we discuss such 'fuzzy' RDD estimation techniques. We introduce the assumptions that need to be fulfilled for estimates to be credibly causally identified. We will also discuss an extensive example of such a research design.

10: Time Series: Univariate Time-Series data
Reading: Monogan, pp.157-167
This is an introductory lecture to time-series analysis. It introduces the concept of stationarity and temporal dependence and teaches us to use ARIMA models to assess these.

11: Time Series: Bivariate Time-Series data
Reading: TBA
This lecture considers how to assess relationships between two time-series, and introduces the concepts of pre-whitening, Granger causality, and cointegration.

12: Time Series: Time-Series Cross-Section data
Reading: TBA
This lecture addresses how to deal with time-series across cases. It introduces the methods of fixed and random effects and considers other alternatives.