DEPARTMENT OF POLITICAL AND SOCIAL SCIENCES

SPS 3rd term workshop 2024 – 2025

Causal Inference with Longitudinal Data using G-Methods

Professor: Sascha Riaz

Guest lecturer: Alejandra Rodríguez-Sánchez (University of Potsdam)

Contact: Sofia.Altesini@eui.eu

Dates and schedule:

Monday	14 April @ Badia Fiesolana, Seminar Room 2 (BF156) from 09:00 until 18:00
Tuesday	15 April @ Badia Fiesolana, Seminar Room 2 (BF156) from 09:00 until 18:00
Wednesday	16 April @ Badia Fiesolana, Seminar Room 2 (BF156) from 12:00 until 18:00

Credits: 10

Workshop Outline / Syllabus

G-methods, originally developed in epidemiology, are gaining traction in the social sciences as powerful tools for addressing causal inference challenges in longitudinal settings. These methods are specifically designed to handle repeated exposures over time, time-varying confounders, and treatment-confounder feedback loops. By accounting for dynamic confounding, G-methods such as inverse probability of treatment weighting (IPTW) and the G-formula can, under certain assumptions, estimate causal effects of multiple exposures.

In this seminar, we will explore scenarios where G-methods are essential and focus on two key approaches: IPTW and the G-formula. To deepen our understanding, we will simulate data-generating processes that include various forms of dynamic confounding and demonstrate how results differ when using other longitudinal analysis methods (e.g., fixed effects models) compared to G-methods. Finally, participants will have the opportunity to apply these techniques to a dataset of their choice or simulate data that mirrors empirical settings, allowing for hands-on practice with the methods discussed.

Outline

1. Foundations of Causal Inference (14th morning session)
Why Causal Inference?
Observational studies vs. randomized experiments
The problem of confounding and bias
Potential outcomes and counterfactuals
Directed Acyclic Graphs as tools to identify confounding and selection bias

Identifying causal paths and backdoor adjustment sets

2. Time-Varying Treatments and Confounders (14^{th} morning session)
Why is confounding in time-varying settings harder to solve?
Handling longitudinal data with time-dependent variables

3. Emulating a "target trial" (14^{th} afternoon session)
The value of thinking through hypothetical target trials

4. Estimation approaches (15^{th} morning and afternoon session)

Inverse Probability of Treatment Weighting (IPTW)

Conceptual foundation: weighing by the inverse probability of receiving treatment.
Estimating marginal effects and balancing covariates.
Strengths and weaknesses of IPTW.
Using logistic regression to calculate propensity scores in R.
Generating stabilized weights to reduce variance.
Using Machine Learning to calculate weights in R
Some diagnostics for IPTW

The G-Formula: Theory and Practice

What is the G-Formula? Theoretical foundation: conditional expectations and the law of iterated expectations
When to Use the G-Formula
Implementing the G-Formula in R
Estimating causal effects with static and dynamic interventions.
R packages: glm for modeling, gformula for simulation-based methods.

4. Simulation-Based Approach to Causal Inference (15^{th} afternoon session and 16^{th} afternoon session)

The Role of Simulations in Statistics and Causal Inference:
Simulating time-varying data for treatment and confounders.
Building Simulations in R
Simulating datasets with known causal structures.
Incorporating time-varying confounders and dynamic treatments.
Hands-On Exercise:
Comparing naïve regression with G-Formula results
Create a simulated dataset with longitudinal data and apply both the G-Formula and IPTW.

5. Empirical data project (16^{th} afternoon session – optional? Only for those interested)
Hands-On Project:
Estimating causal effects of a treatment on an outcome using simulated longitudinal data
(Student led project)

Course Materials

Software: R (with packages: dagitty, ipw, gformula, survey, lme4, etc.).
Readings: Selected chapters from *Causal Inference: What If* (Hernán & Robins) and supplementary articles.
Datasets: Simulated and real-world datasets (e.g., health, education, or economics).