Causal Inference Workshop

PhD in Sustainable Development

Spring 2026

(Last Updated January 20, 2026)

Course Title: SDEV9280 Causal Inference Workshop
Instructor: Suresh Naidu, sn2430@columbia.edu
Teaching Assistant: Fanyu Wang, fw2397@columbia.edu
Meeting Dates / Times: Fridays, 9:00am - 9:55am, before colloquium
(optional office hours / troubleshooting Fridays, 8:00am - 9:00am)
Location: IAB 823
Credits: 1.5 (Pass/Fail; fulfills Sustainable Development PhD Research Tools requirement)

Course Description

Note: This workshop is the third iteration of the Causal Inference Workshop for students in the PhD program in Sustainable Development. The original workshop was designed by Claire Palandri in 2022 (see syllabus here) and offered again by Anna Papp in 2024 (see materials here). The updated syllabus and slides are extremely heavily based on the amazing materials by Claire and Anna.

This workshop is designed for students in the PhD program in Sustainable Development and covers the fundamental theory and techniques of causal inference. The course is specifically tailored to students trained in econometrics and positioned to conduct interdisciplinary research. It ties back the econometrics approaches covered to the underlying statistical framework, and provides the students with the tools to conduct rigorous empirical analyses and to share and defend their approach in front of both economics and non-economics audiences. Lower-year students are presented the fundamental methods for observational studies; upper-year students can discuss how they employ them in their own current research. Participants are presented with the core methods in the field, their limitations, and best practices, and less common statistical methods relevant for causal inference and Sustainable Development. In addition, the course includes practical coding exercises to help students implement the techniques covered in the workshop.

Course Overview

The course will consist of a weekly class led by the teaching assistant. The 12 weeks are organized into four sections:

A. **Causal inference fundamentals [the aim of causal inference]:** Fundamentals of inferential statistics are reviewed, followed by the theoretical framework of potential outcomes for causal inference, and how it's implemented with observational data through regression modeling.

B. **Design stage, identification strategies [how causal inference is done]:** The most common identification strategies — special cases of regression adapted to particular forms of natural experiments — are reviewed. For each method, the canonical setup is presented in the first part of the session, in particular: the data generating process assumed; the identifying assumptions; the estimand of interest; the estimator used; best practices; strengths and weaknesses. The second part of the session then puts the theory in practice by discussing the analyses of working or published papers: work in progress by a current PhD student and/or a published paper on Sustainable Development. When appropriate, the workshop will also include short coding exercises (in R and Stata) to help students implement the methods discussed.

C. **Analysis stage, stronger causal inferences [how to improve upon causal inference]:** How to obtain stronger causal inferences through steps at the analysis stage. Pre- and post-estimation best practices are presented, including how to support the assumptions on which the inferences rest.

D. **Other topics in causal inference and Sustainable Development:** Less common topics in causal inference are presented. Additionally, topics especially relevant for students in Sustainable Development will be covered, such as causal inference with remotely sensed data.

The workshop does not follow a specific textbook, but the two references in which the participants will find most of the material covered — and that are highly recommended as complements of each other — are *Mostly Harmless Econometrics* (Angrist and Pischke, 2008) and *Regression and Other Stories* (Gelman et al., 2020). References to other useful papers will be added to the syllabus.

Grading

This course is graded on a Pass/Fail basis. The course grade will be mainly based on weekly attendance, and completion of the final assignment due the end of the course (May 1, 2026). The final assignment will involve implementation and modification of one of the short coding exercises.

Accommodations

I am committed to making the Causal Inference workshop an inclusive experience in which every student can participate. Please let me know if you have a disability or a circumstance that may affect your learning in the workshop so that we can determine the best way to make the class accessible for your needs. You should also feel free to contact Disability Services to discuss options for removing barriers in this course.

Course Structure and Week-by-Week Schedule

A. Causal inference fundamentals

1. Week 1, 01/23/26: Overview; inferential statistics fundamentals

[theory]

- Regression models as conditional distributions; assumptions of the classical linear regression model, and the estimator properties depending on them; dealing with some departures from the usual assumptions: sandwich estimators; limited y models.

2. Week 2, 01/30/26: The potential outcomes framework and identification

[theory]

- The Neyman-Rubin causal model or potential outcomes framework; identification from independence assumptions; regression on the treatment recovers an average treatment effect (relation between observed and potential outcomes); simplest case & extensions (limited Y, covariates X, continuous D).

B. Design stage: Identification strategies

3. Week 3, 02/06/26: Instrumental Variables (IV) and Regression Discontinuity (RD)

[theory]

- IV, theory: instruments and compliance behavior; two-stage least squares; local average treatment effect (LATE); computing average complier characteristics and getting more out of a LATE.
- RDD, theory: deterministic but discontinuous assignment; estimation with flexible functional forms; Sharp RD, Fuzzy RD (imperfect compliance).

4. Week 4, 02/13/26: Instrumental Variables (IV) and Regression Discontinuity (RD)

[application and coding]

- Working paper by current PhD student and/or published paper related to Sustainable Development.
- Short coding exercise related to the identification strategy.

5. Week 5, 02/20/26: Difference-in-Differences (DiD), Triple Differences (DiDiD), Event Studies

[theory]

- DiD, theory: different models of the counterfactual; pre-trends; justifying a third difference.
- New TWFE literature, theory: Pitfalls of weighted sums of the average treatment effects with two-way fixed effect estimators; estimators robust to heterogeneous treatment effects; continuous TWFE.

6. Week 6, 02/27/26: Difference-in-Differences (DiD), Triple Differences (DiDiD), Event Studies

[application and coding]

- Working paper by current PhD student and/or published paper related to Sustainable Development.
- Short coding exercise related to the identification strategy.

7. Week 7, 03/06/26: Synthetic Control, Synthetic Difference-in-Differences (SynthDiD)

[theory, application and coding]

- Synthetic control/DiD, theory: creating an optimized counterfactual; synthetic difference-in-differences.
- Working paper by current PhD student and/or published paper related to Sustainable Development.
- Short coding exercise related to the identification strategy.

C. Analysis stage: Steps for stronger causal inferences

8. Week 8, 03/13/26: Limitations of identification strategies; pre-estimation steps (restructuring data), estimation (controls) and post-estimation steps (supporting assumptions)

[theory]

- Limitations of identification strategies; steps can be taken pre-/during-/post-estimation.
- Exploratory data analyses; restructuring of data; propensity score matching.
- Required/forbidden controls; optional good/bad controls; bias amplification.
- Modeling assumptions; back to inference fundamentals: post-estimation model diagnostics. Fake data simulations: fit the model to simulated data where you know the true parameter values.
- Identifying assumptions; show a balance test table and do falsification tests. Examples of falsification tests for each identifying assumption of common identification strategies (IV, RDD, DiD).

Spring Break, 03/20/26

IPWSD, 03/27/26

D. Other topics in causal inference and Sustainable Development

9. Week 9, 04/03/26: Inference

[theory]

- Design-based vs sampling-based inference; randomization inference, bootstrapping standard errors.

SusDeveR, 04/10/26

10. Week 10, 04/17/26: Field experiment in Sustainable Development

[theory and application]

- Basics of designing and implementing a field experiment in Sustainable Development and analysis of experimental data.
- Working paper by current PhD student and/or other relevant working/published paper.

11. Week 11, 04/24/26: Weather/climate data and regressions, remote sensing data for Sustainable Development

[theory, application, and coding]

- How to use weather/climate data in causal inference; other common topics of interest related to Sustainable Development.
- Potential challenges with using remotely sensed data in causal inference; possible solutions.
- Working paper by current PhD student and/or other relevant working/published paper.
- Short coding exercise with example.

12. Week 12, 05/01/26: Other topics and Wrap-up

Topics TBD based on student interests and needs.