CAUSAL INFERENCE IN POLITICAL SCIENCE RESEARCH

(Ph.D.-Level Class)

Summer Semester 2023, University of Konstanz

Seminar Time:	Monday, 11.45 AM–1.15 PM (11.45–13.15 Uhr)
Seminar Location:	C358 (in-person seminar)
Instructor:	Jan P. Vogler
Contact:	jan.vogler@uni-konstanz.de

Please note: The first class will be virtual and will take place on April 6 at 4.30 PM. Further details are below.

Course Abstract:

A large share of advanced research in political science theorizes about and tests *causal* relationships between social, economic, and political phenomena. Given the importance of identifying causal relationships to state-of-the-art research in the discipline, this class aims to introduce graduate students to the mathematical foundations, theory, and methods of causal inference. After an initial review of probability theory and regression analysis, students learn about directed acyclic graphs and the potential outcomes model, which will be the theoretical basis for the remainder of the class. In subsequent sessions, the class will cover the key methods of causal inference with observational data. Specifically, students will learn about the following topics: (1) matching, (2) regression discontinuity, (3) instrumental variables, (4) panel data, (5) differences-in-differences, and (6) synthetic controls. An essential component of the class will be student-led presentations and replication exercises of outstanding recent research articles in political science. These exercises are meant to help students bridge the gap between the theory and practice of causal inference.

Course Objectives:

By the end of the class, students will be able to:

- Describe the relevance of causal inference and related research strategies to political science.
- Understand the mathematical foundations of advanced quantitative methods and how they relate to the goal of causal identification.

- Elaborate on the utility of directed acyclic graphs and the properties of the potential outcomes causal model.
- Apply a wide variety of key causal inference tools to real-world data, including matching, regression discontinuity, differences-in-differences, and instrumental variables approaches.
- Combine knowledge in all of the aforementioned areas to write their own research plans and research papers using the theory and tools of causal inference.

Course Requirements:

Useful Background Knowledge

Good knowledge of mathematics, statistics, and the programming language R is a prerequisite of the class. While we will spend some time at the beginning of the course with reviewing content in these areas, the class will likely be very challenging to any students who do not already have related foundational knowledge. In general, the course is targeted at graduate students who have previously already gone through some basic training in all of the aforementioned areas.

Preparation for Class, Weekly Readings, and Participation

Students are asked to carefully read and follow the relevant textbook chapters while they are enrolled in the class. While the students are encouraged to study the relevant chapters prior to class to familiarize themselves with terms, notation, and content, they may also choose to first come to class and consult the textbook afterwards, if this approach is more conducive to their learning.

In-Class Research Article Presentations and Replications (Individual or Group-Based)

Students will sign up for in-class research article presentations and replication exercises (50% of the course grade). Currently, there are 6 slots for article presentations and replications available (one in each of the last six sessions of the class).

We will assign at least 1 person to each of these topics (each topic is a different causal inference approach). The specific procedure of assigning people to topics will be detailed in the first and/or second sessions of the class. For each topic, the assigned student(s) will need to identify an article published in a leading journal of political science in 2015 or later that is of interest to them, applies the respective method to real-world data, and for which replication data is available online.

If only one person is assigned to a topic, this person will be expected to provide a 30-minute (PowerPoint/Beamer) presentation that consists of:

- A 10-minute summary of the article's topic and theory
- A 10-minute summary of the article's empirical test (including a clear connection to the content of this week's class and reading)
- A 10-minute replication of the article's main results (which means ***only*** the results from the main body of the study, not from the appendix/supplementary material)

If two people are assigned to a topic, they will be expected to provide a 40-minute (PowerPoint/ Beamer) presentation that consists of:

- All of the above (points 1–3)
- A 10-minute replication of the article's robustness checks in the appendix/supplementary material (in addition to a replication of the main results)

If three or four people are assigned to a topic, they will be expected to provide a 50-minute (PowerPoint/ Beamer) presentation, that consists of:

- All of the above
- A 10-minute extension of the article's findings with additional robustness checks and/or further empirical analyses (the expectations for this part will be higher depending on the number of people)

In addition to preparing the presentation itself, the assigned students are expected to provide a detailed and annotated R script that replicates the study's results and runs without further modification from beginning to end. Annotation of the R script can be copied from the original replication files but should provide additional clarification on coding steps in cases where insufficient information is provided by the study's original author.

This is a major task and students are expected to spend at least two weeks (ideally three) with preparing their presentations and conducting the replication exercises. Students should plan ahead for the possibility that they have questions on/issues with the replication code and may need to contact the article's author to provide further clarification. Students are expected to practice the presentation in advance at least once to ensure that they can stay within the suggested time limit. Participants are expected to email me their preliminary PowerPoint slides on Thursday (by 12.00 PM or 12.00 Uhr mittags) before the relevant class for feedback.

Written Assignment

There will be one written assignment in this class. This written assignment is a research plan that is due at the end of the class.

Requirements for the research plan (50% of the course grade): As the final assignment of the class, students are expected to (1) identify a research question that interests them and that can be answered by applying one of the tools of causal inference that we have learned about and (2) write a research plan on how to pursue this specific research question. The plan is meant to include: (1) The specific question, its connection to the class, and relevance (~500–650 words), (2) an outline of your own argument or theory that addresses the research question, with an emphasis on the causal process you suggest (~500–650 words), (3) a description of the actual data that allow you to evaluate your theory/argument (~500–650 words), and (4) a clear explanation of why and how a specific tool of causal inference can be applied to these data (~500–650 words). The expected total length of the research plan is 8–10 pages or 2,000 to 2,500 words (excluding references). The deadline of the research plan is Monday, August 15, 2022 (please send it to me by email in Word *and* PDF format).

Late Assignment Policy

If you cannot finish an assignment (especially the research plan) on time for a legitimate reason, including, for example, emergencies and illness, please contact me. Please do so ideally before the assignment is due, but at most seven days after the deadline. In case of illness, a statement by your doctor is needed. Depending on the situation, I will provide you with an alternative assignment and/or deadline. If no legitimate reason is provided for late assignments, 10% of the point total for the assignment are subtracted for every day that the assignment is late, and a point total of 0 is awarded if the assignment is more than seven days late.

Attendance Policy

You are allowed to miss a maximum of three classes. Out of those three, you can miss one class without informing me. However, it is encouraged that you inform me in advance. If you miss more than one class, I expect you to inform me about this in advance. If you miss more than three classes (for any reason), you will automatically fail the class. Please note that this course moves at a very high speed and that missing classes will negatively affect your ability to understand the material of future sessions. Therefore, it is strongly encouraged to avoid missing any classes.

Grading:

Based on the above requirements, the course grade will consist of the following elements:

- 50%: In-class presentation and replication exercise (individual or group-based)
- 50%: Research plan (~8–10 pages or ~2000–2500 words)

Inclusion:

An essential goal of the class is to create an open and welcoming discussion atmosphere. Diversity of opinions, constructive discussion, and mutual respect are at the core of academic discourse and will be key elements of this class. A heterogeneity in backgrounds, experiences, and identities will greatly benefit us by allowing us to learn from each other and expand our thinking. All students are encouraged to voice their opinions and to do so in a way that displays respect for the opinions of other students in the class. Students who believe that these goals are inhibited in any way should contact me so that we can discuss their concerns.

Academic Integrity:

A second essential goal of mine is to uphold the standards of academic integrity in this class. With respect to individual assignments, it is expected that all work submitted is entirely done by the person who submits it. Similarly, group-based work is expected to be done only by those who are officially assigned to the specific task. As detailed above, students may contact the author of the study if they

have clarification questions on any aspect of it. Moreover, all literature used in writing a paper must be referenced. Students are expected to always use quotation marks when they directly quote the words or statements of others. If you have any questions about academic integrity, please contact me so that we can discuss them.

Textbook:

The class is primarily based on the following book. All students are required to acquire a copy of it:

- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. **[Cunningham]**

Please note that an electronic campus license of the text book has been made available by the library. University of Konstanz students can access this version at the following URL (on campus or via VPN connection):
https://www.degruyter.com/document/doi/10.12987/9780300255881/html

Important Dates and Deadlines:

- First class (virtual): **Monday, April 6, 2023 (4.30 PM)**
- Holiday break day I (no class): Monday, May 1, 2023
- Holiday break day II (no class): Monday, May 29, 2023
- Holiday break day III (no class): Monday, June 5, 2022
- In-class presentation and replication: Individually assigned
- Last class: Monday, July 17, 2023
- Research plan due: Monday, August 14, 2023

Office Hours:

If you would like to speak with me, please contact me by email to set up an appointment. In the email, please also include the specific reason why you would like to speak with me and provide me with at least three different dates at and time frames during which you are available. I will then schedule a meeting with you.

Course Schedule Begins on the Following Page.

COURSE SCHEDULE:

PART I: INTRODUCTION TO THE CLASS AND REVIEW OF STATISTICAL FOUNDATIONS

The first part of the class introduces the students to the class schedule and to the mathematical foundations of causal inference. We will also review the R statistical programming language as this will be an essential tool for the replication exercises.

1. **Introduction and Course Overview: "Causal Inference in Political Research"** (April 6, 2023)

   The Relevance of Causal Inference for Modern Research in Political Science (Online Class)

   This class will take place on April 6 at 4.30 PM (16.30 Uhr) Central European Time.

Required Reading:

- Required readings for the first session are posted on ZEUS.

Zoom Link (*Zoom registration is required, please register before the class): https://uni-konstanz-de.zoom.us/j/91853989482?pwd=Tzh0Q2dRemNsdGx6TDYvOGtVSVdvdz09

2. **R Statistical Programming Language Review** (April 17, 2023)

Required Readings:

Venables, W. N., Smith, D. M., & R Development Core Team. (2009). *An introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics.* Available at: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

3. **Probability and Regression Review, I** (April 24, 2023)

Required Readings:

- **Cunningham,** chap. 1 ("Introduction") (Background Reading)
- **Cunningham,** chap. 2 ("Probability and Regression Review"), pp. 16–52

— NO CLASS ON MAY 1, 2023 (HOLIDAY BREAK)! —

4. Probability and Regression Review, II (May 8, 2023)

Required Readings:

- **Cunningham,** chap. 2 ("Probability and Regression Review"), pp. 52–95

PART II: THEORETICAL FOUNDATIONS OF CAUSAL INFERENCE

The second part of the class deals with the theoretical foundations of causal inference and covers directed acyclic graphs and the potential outcomes causal model. These two courses represent essential building blocks of a comprehensive understanding of causal inference techniques.

5. Directed Acyclic Graphs (May 15, 2023)

Required Readings:

- **Cunningham,** chap. 3 ("Directed Acyclic Graphs")

6. Potential Outcomes Causal Model (May 22, 2023)

Required Reading:

- **Cunningham,** chap. 4 ("Potential Outcomes Causal Model")

— NO CLASS ON MAY 29, 2023 (HOLIDAY BREAK)! —

— NO CLASS ON JUNE 5, 2023 (HOLIDAY BREAK)! —

PART III: METHODS OF CAUSAL INFERENCE AND THEIR APPLICATION

The third part of the class covers the key tools of causal inference research, including techniques like matching, regression discontinuity, and instrumental variables, among others. For this part of the class, the students will also be expected to prepare research article presentations and replications.

7. Matching and Subclassification (June 12, 2023)

Required Reading:

- **Cunningham,** chap. 5 ("Matching and Subclassification")

Practical Exercise:

- Research article presentation and replication exercise I (focus on matching)

8. Regression Discontinuity (June 19, 2023)

Required Reading:

- **Cunningham,** chap. 6 ("Regression Discontinuity")

Practical Exercise:

- Research article presentation and replication exercise II (focus on regression discontinuity)

9. Instrumental Variables (June 26, 2023)

Required Reading:

- **Cunningham,** chap. 7 ("Instrumental Variables")

Practical Exercise:

- Research article presentation and replication exercise III (focus on instrumental variables)

10. Panel Data **(July 3, 2023)**

Required Reading:

- **Cunningham,** chap. 8 ("Panel Data")

Practical Exercise:

- Research article presentation and replication exercise IV (focus on panel data)

11. Differences-in-Differences **(July 10, 2023)**

Required Reading:

- **Cunningham,** chap. 9 ("Differences-in-Differences")

Practical Exercise:

- Research article presentation and replication exercise V (focus on differences-in-differences)

12. Synthetic Controls **(July 17, 2023)**

Required Readings:

- **Cunningham,** chap. 10 ("Synthetic Controls")

Practical Exercise:

- Research article presentation and replication exercise VI (focus on synthetic controls)

Dates and Deadlines at the End of the Semester:

- Research Plan Due: Monday, August 14, 2023