Advanced Quantitative Methods in Political Science: Causal Inference, Panel Data, and Multilevel Modeling

Fall Semester 2024, Aarhus University

(Preliminary syllabus; please make sure to download the final version later)

Seminar Time:	Monday, 3.00–6.00 PM
Seminar Location:	1330-038 Undervisningslokale
Instructor:	Jan P. Vogler
Contact:	jan.vogler@ps.au.dk

Course Abstract:

This course provides students with a comprehensive overview of advanced quantitative applications in political science, including methods of causal inference, panel data analysis, and multilevel modeling. The key goal is to equip the participants with both a strong theoretical background in these methods and the knowledge to apply them in future research projects. The course builds upon skills acquired in foundational methods courses (Methods I and Methods II at Aarhus University) and does not presuppose knowledge of mathematics or probability theory beyond what is introduced there. It consists of four different parts: (1) mathematical foundations, (2) theory and application of causal inference, (3) theory and application of panel data analysis, and (4) theory and application of multilevel modeling. In the first part of the course, because a thorough understanding of advanced quantitative methods requires a strong mathematical foundation, we begin with a review of the concepts in probability theory and regression analysis that are most directly relevant to us. In the second part of the course we discuss the mathematical theory and application of key causal inference tools, including matching, regression discontinuity designs, instrumental variables, and differences-in-differences. In the third part of the class, we consider different types of panel data analysis, including fixed effects models, random effects models, panel models with instrumental variables, and dynamic panel models. In the fourth and final part of the course, we introduce multilevel modeling, including multilevel linear regression, multilevel logistic regression, and multilevel generalized models. During class, students will participate in a series of replication exercises through which they will learn how to apply these tools themselves. By the end of the course, students will formulate their own research question and present a research design that makes use of one of one of the advanced methods they have learned about.

Course Objectives:

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

- Understand how probability theory relates to advanced quantitative methods in political science and apply concepts from the former to the latter.
- Describe the logic behind key tools of causal inference; and answer research questions by using these tools in applied research.
- Explain different panel data structures and employ a variety of tools to analyze these data.
- Recognize when data have a multilevel component and make use of appropriate multilevel models to examine them.
- Formulate their own research question, select an appropriate statistical tool, and describe how the tool can be applied to answer the research question.

Course Requirements:

Prerequisites

Before enrolling in this course, students are expected to have completed the courses Methods I and Methods II at Aarhus University (or their equivalents at other universities). In general, the course is targeted at graduate students who have previously gone through basic training in statistics. No further knowledge is formally required. However, a very high degree of motivation on the part of the students to learn and understand mathematical concepts is expected. Additionally, basic knowledge of mathematical functions and calculus is extremely helpful to succeed in the course. If students are uncertain about the adequacy of their current knowledge for succeeding in the course or if students wish to prepare more comprehensively for the course in advance, especially in terms of basic mathematical concepts, they are welcome to contact me to receive advice in this respect.

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 strongly encouraged to (at least briefly) study the relevant chapters prior to class to gain familiarity 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)

An important practical component of the course are in-class research article presentations and replication exercises. These exercises serve the purpose of making the course more interactive and helping the students bridge the gap between methodological theory and empirical application. There are 8 slots for article presentations and replications available (one in each of the last eight sessions).

We will assign at least 1 person to each of these topics (each topic is a different method). The specific procedure of assigning people to topics will be detailed in the first and/or second sessions of the course. 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 in all 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 also expected to practice the presentation in advance at least once to ensure that they can stay within the suggested time limit. Finally, participants are expected to email me their preliminary PowerPoint slides on Thursday (by 12.00 PM) before the relevant class for final feedback.

Although the replication exercises are not graded, they represent an extremely important component of the learning goals of the course and students are expected to invest time and energy to make them a success for all participants.

Synopsis and Oral Examination

The exam form of this course is a synopsis (no more than 5 pages/2000 words) (33% of the grade) combined with an oral examination (67% of the grade). The synopsis will require the students to specify how they would use one or two of the methods introduced in the course to answer a research question and analyze real-world data. In the oral examination that follows, students' knowledge regarding the application of the chosen method will be tested. The questions in the oral examination may go beyond the content presented in the synopsis and could include additional questions on the applicability/utility of other methods introduced in the course as well as closely related content.

Detailed requirements for the synopsis (33% of the grade): For the synopsis, students are expected to (1) identify a research question that interests them and that can be answered by applying one of the empirical tools that we have learned about and (2) write a plan for how to pursue this specific research question. The synopsis is meant to include: (1) The specific question, its connection to the course, and relevance (~400-500 words), (2) an outline of your own argument or theory that addresses the research question (~400-500 words), (3) a description of the actual data that allow you to evaluate your theory/argument (~200-300 words), and (4) a clear explanation of why and how a specific empirical tool introduced in the course can be applied to these data (~600-700 words). The expected total length of the synopsis is 4–5 pages or 1,600 to 2,000 words (excluding references). The deadline of the research plan is Monday, January 6, 2025 (please send it to me by email in Word *and* PDF format).

Detailed requirements for the oral examination (67% of the grade): During the oral examination, students will be asked questions about the synopsis they have written. Questions may cover the entirety of the content of the synopsis, including the research question, the theory, the available data, and the chosen method. As touched upon above, questions during the oral examination may not be limited to the content of the synopsis, but could cover other methods introduced in the course and their applicability/utility as well as closely related topics.

Late Assignment Policy

If you cannot finish an assignment (especially the synopsis) on time for a legitimate reason, including, for example, emergencies and illness, please contact me as soon as possible. In coordination with the department administration and in line with the department's examination rule, we will deal with such situations on a case-by-case basis.

Attendance Policy

Students are expected to come to as many sessions of the course as possible. In line with the official department policy, they are required to participate in at least 80% of all course sessions. Furthermore, students must notify the lecturer if they are unable to attend class. Students are also expected to actively participate in the course, for example by answering questions and participating in in-class exercises. Participation in the sessions is highly relevant to students' learning success. Moreover, students are expected to give a presentation and conduct a replication exercise in one of the sessions. These will be individually assigned. In the first course session, the activity requirements

and the expectations for the students' participation in general will be presented in more detail. All other policies regarding attendance are in line with departmental standards.

Grading:

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

- 33%: Written synopsis (maximum 5 pages)
- 67%: Oral examination

Inclusion:

An essential goal of the course 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 course. 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 course. 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. If AI is used in any part of the writing (or coding) process, it should be indicated clearly. A succinct and precise statement regarding the extent and purpose of AI use is necessary in any such case. 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.

In addition, all general rules established at Aarhus University apply. For details, see the following website: https://studerende.au.dk/en/examinations/cheating-at-exams

Textbooks:

The course is primarily based on the following books. All students are required to acquire either a physical or an electronic copy of them:

- Moore, W. H., & Siegel, D. A. (2013). *A mathematics course for political and social research*. Princeton University Press. **[Moore & Siegel]**
    - This book is available to Aarhus students as e-book here: link
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. **[Cunningham]**
    - This book is available to Aarhus students as e-book here: link
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. **[Gelman & Hill]**
    - This book is available to Aarhus students as e-book here: link

Important Dates and Deadlines:

- First class: Monday, August 26, 2024
- Holiday break day (no class): Monday, October 14, 2024
- In-class presentation and replication: Individually assigned/information in 1st or 2nd session
- Last class: Monday, December 2, 2024
- Synopsis due: Monday, January 6, 2025
- Oral examinations: Monday and Tuesday, January 13 and 14, 2025

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 (ideally 2-3 hours in length) 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 MATHEMATICAL FOUNDATIONS

The first part of the course introduces the students to the class schedule and to the mathematical foundations of advanced quantitative methods, especially in terms of probability theory and the mathematical basics of regression analysis. We will also review the R statistical programming language as this will be an essential tool for the replication exercises that participants will be engaged in throughout the course.

1. **Introduction and Course Overview: "Advanced Quantitative Methods in Political Research"**

   The Relevance of Quantitative Methods for Research in Political Science (August 26, 2024)

Required Reading:

- Required readings for the first session will be made available electronically.

2. **Mathematical Foundations/Probability Theory, I** (September 2, 2024)

Required Readings:

- **Moore & Siegel,** chap. 9 ("An Introduction to Probability") (pp. 175-197)
- **Moore & Siegel,** chap. 10 ("An Introduction to (Discrete) Distributions") (pp. 198-241)

3. **Mathematical Foundations/Probability Theory, II** (September 9, 2024)

(*We will have to reschedule this class due to conference travel.)

Required Readings:

- **Moore & Siegel,** chap. 11 ("Continuous Distributions") (pp. 242-272)

4. **Mathematical Foundations/Probability Theory, III** (September 16, 2024)

Required Readings:

- **Cunningham,** chap. 2 ("Probability and Regression Review") (pp. 16-95)

5. Statistical Programming in R (September 23, 2024)

Required Readings:

- Venables, W. N., Smith, D. M., & R Development Core Team. (2024). *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 (pp. 1-103)

PART II: CAUSAL INFERENCE

The second part of the course deals with the theory and practice of causal inference. We begin by introducing directed acyclic graphs and the potential outcomes causal model. These two courses represent essential building blocks of a comprehensive understanding of causal inference techniques. Then, we move on to discuss various causal inference techniques, including matching, regression discontinuity, instrumental variables, and differences-in-differences designs. Each session will be accompanied by a student-led replication exercise.

6. Directed Acyclic Graphs and the Potential Outcomes Causal Model (September 30, 2024)

Required Readings:

- **Cunningham,** chap. 3 ("Directed Acyclic Graphs") (pp. 96-118)
- **Cunningham,** chap. 4 ("Potential Outcomes Causal Model") (pp. 119-174)

Recommended Readings:

- **Cunningham,** chap. 1 ("Introduction") (Background Reading: No Discussion)

7. Matching and Subclassification (October 7, 2024)

Required Reading:

- **Cunningham,** chap. 5 ("Matching and Subclassification") (pp. 175-240)

Practical Exercise:

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

— NO CLASS ON OCTOBER 14, 2024 (HOLIDAY BREAK)! —

8. Regression Discontinuity (October 21, 2024)

Required Reading:

- **Cunningham,** chap. 6 ("Regression Discontinuity") (pp. 241-314)

Practical Exercise:

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

9. Instrumental Variables (October 28, 2024)

Required Reading:

- **Cunningham,** chap. 7 ("Instrumental Variables") (pp. 315-385)

Practical Exercise:

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

10. Differences-in-Differences (November 4, 2024)

Required Reading:

- **Cunningham,** chap. 9 ("Differences-in-Differences") (pp. 406-510)

Practical Exercise:

- Research article presentation and replication exercise IV (focus on diff-in-diff)

PART III: PANEL DATA

The third part of the course covers the key approaches to dealing with panel data. Specifically, we will begin by looking at the peculiarities of panel data. Then, we will introduce both fixed-effects and random-effects models. Finally, we will look at instrumental variable approaches with panel data and dynamic panel models.

11. Panel Data, I (November 11, 2024)

Required Reading:

- **Cunningham,** chap. 8 ("Panel Data") (pp. 386-405)

Practical Exercise:

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

12. Panel Data, II (November 18, 2024)

Required Reading:

- Wawro, G. (2002). Estimating dynamic panel data models in political science. *Political Analysis*, 10(1), 25-48.
- Garrett, G., & Mitchell, D. (2001). Globalization, government spending and taxation in the OECD. *European Journal of Political Research*, *39*(2), 145-177.
- Plümper, T., Troeger, V. E., & Manow, P. (2005). Panel data analysis in comparative politics: Linking method to theory. *European Journal of Political Research*, *44*(2), 327-354.
- Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, *3*(1), 133-153.

Practical Exercise:

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

PART IV: MULTILEVEL MODELING

The fourth and final part of the course deals with multilevel modeling. We begin by discussing multilevel data structures. Then, we introduce a number of empirical tools that allow us to analyze such data, including multilevel linear models, multilevel logistic models, and multilevel generalized linear models.

13. Multilevel Modeling, I (November 25, 2024)

Required Readings:

- **Gelman & Hill,** chap. 11 ("Multilevel structures") (pp. 237-250)
- **Gelman & Hill,** chap. 12 ("Multilevel linear models: the basics") (pp. 251-278)

Practical Exercise:

- Research article presentation and replication exercise VII (focus on multilevel modeling)

14. Multilevel Modeling, II **(December 2, 2024)**

Required Readings:

- **Gelman & Hill,** chap. 13 ("Multilevel linear models: varying slopes, non-nested models, and other complexities") (pp. 279-300)
- **Gelman & Hill,** chap. 14 ("Multilevel logistic regression") (pp. 301-324)
- **Gelman & Hill,** chap. 15 ("Multilevel generalized linear models") (pp. 325-342)

Practical Exercise:

- Research article presentation and replication exercise VIII (focus on multilevel modeling)

Dates and Deadlines at the End of the Semester:

- Synopsis due: Monday, January 6, 2025
- Oral examinations: Monday and Tuesday, January 13 and 14, 2025