Quantitative Methods for Causal Inference QMCI

MBR Program (A/I course) WS 2024/2025

Ambre Nicolle ISTO, Munich School of Management, LMU

This version: 11.07.2024. Subject to changes.

Course description

The course will provide PhD students with a comprehensive understanding of contemporary causal inference techniques. Focusing on quasi-experimental methods like Difference-in-Differences, Regression Discontinuity Design, and Synthetic Control Methods, the course will emphasize both theoretical foundations and practical applications. Participants will engage in hands-on empirical exercises relying on datasets from published papers from Economics and Management. The course aims to enhance students' ability to conduct robust causal analysis in their research. Stata will be the software used for examples and solutions. However, participants can use Python or R at their convenience during the hands-on session and during the exam. Prior completion of the course “Quantitative Methods” is required to participate to the course.

Course material

Lecture slides and data files will be made available on Moodle prior to each session. Suggested solutions to empirical exercises will be provided after each session (only Stata code). All material will be in English.

Prerequisites

The course is open to PhD students currently enrolled in the MBR program and is credited as an **A/I course**. Students **must** have participated to the course Quantitative Methods *before* participating to QMCI. Participants to the course should feel comfortable working with Stata – or be proficient in Python or R if they opt for working with alternative softwares.

Organization of the sessions

The course will be organized in four sessions of five hours and a half (22 hours in total). Sessions are held in person at:

- *Iuk-Pool*: Ludwigstr. 28 VG, 2.Stock, Raum 207
- or *Raum III*: Ludwigstr. 28 RG, EG, Raum 023.

Attendance to all teaching sessions is mandatory. If you have to leave earlier or arrive later one of the days, please write in advance to let me know.

Examination

The exam will take the form of an **exam** of 3 hours which will combine:

- a set of theoretical questions
    - 50% of the grade.
    - No books, no notes, or computer will be allowed.
- an empirical “open book” exercise similar to one of the exercises solved during class
    - 50% of the grade.
    - Students will use the software of their choice to perform this exercise.
    - Code, figures and tables produced by the students will be evaluated.

Two dates for the exam are offered (students have to choose *one*): **14.02.2025** or **21.02.2025**, 08:00-12:00. No alternative dates or examination type (e.g. homework) will be offered. If you cannot take the exam, please do not register for the course.

Course structure

The course is organized around six topics.

Each topic will combine:

- a **theoretical section** where we will present the setting, the assumptions, the properties of estimators, as well as the pros and cons of approaches (120 minutes)
- an **applied section** where examples from one or several published papers will be presented and discussed (30 minutes).
- an **exercise** relying on a dataset, solved by the course participants (90 min).
    - Solutions (in Stata) will be provided at the end of the session.

Outline of the course

1. Course presentation: outline, organization, examination
2. Introduction and overview of the methods
3. Matching methods
4. Difference in Differences – advanced tools
5. Synthetic Control
6. Regression Discontinuity Design

A detailed version of the outline is presented at the end of this document, with references to all papers and datasets (subject to changes).

Main references

- Cameron, A. C., & Trivedi, P. K. (2010). Microeconometrics using Stata. Stata press.
- Cameron, A. C., & Trivedi, P. K. (2022). Microeconometrics using Stata: Volume 2 Non Linear Models and Causal Inference Methods. Stata press.
- Cunningham, S. (2020). Causal Inference. The Mixtape, 1.
- Huntington-Klein, N. (2021). The effect : An introduction to research design and causality.

Additional references and datasets (subject to minor changes)

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493-505. [*SCG*]
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495-510. [*SCG, dataset available*]
- Abadie, Alberto. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature* 59.2 (2021): 391-425. [*SCG, dataset available*]
- Almond, D., Chay, K. Y., Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3), 1031-1083. *[PSM]*
- Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- Angrist, J. D., & Pischke, J. S. (2014). Mastering'metrics: The path from cause to effect. Princeton university press.
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. Journal of Economic perspectives, 31(2), 3-32.
- Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates?. *Journal of Financial Economics*, 144(2), 370-395. *[DiD]*
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *The Quarterly journal of economics*, 119(1), 249-275. *[DiD]*
- Boehmer, E., Jones, C. M., & Zhang, X. (2020). Potential pilot problems: Treatment spillovers in financial regulatory experiments. *Journal of Financial Economics*, 135(1), 68-87. *[DiD]*
- Bradley, D., Kim, I., & Tian, X. (2017). Do unions affect innovation?. *Management Science*, 63(7), 2251-2271. *[RDD]*
- Butts, K. (2021). Difference-in-differences estimation with spatial spillovers. arXiv preprint arXiv:2105.03737. *[DiD]*
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230. *[DiD, dataset:see Cunningham blog below, Stata implementation]*
- Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155(2), 138-154. *[PSM, dataset available]*
- Cheng, C., & Hoekstra, M. (2013). Does strengthening self-defense law deter crime or escalate violence?: Evidence from expansions to castle doctrine. *Journal of Human Resources*, 48(3), 821-854. *[DiD, data available]*
- Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. Journal of Political Economy, 122(2), 282-318. *[RDD, data available].*
- Favaron, S. D., Di Stefano, G., & Durand, R. (2022). Michelin is coming to town: Organizational responses to status shocks. *Management Science*, 68(9), 6925-6949. *[DiD, SG, dataset available]*
- Flammer, C., & Bansal, P. (2017). Does a long-term orientation create value? Evidence from a regression discontinuity. *Strategic Management Journal*, 38(9), 1827-1847. *[RDD]*

- Fetter, D. K. (2013). How do mortgage subsidies affect home ownership? Evidence from the mid-century GI Bills. American Economic Journal: Economic Policy, 5(2), 111-147. *[RDD, data available].*
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277. *[DiD]*
- Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Science*, 38(2), 193-225.
- Hitt, L. M., & Frei, F. X. (2002). Do better customers utilize electronic distribution channels? The case of PC banking. *Management Science*, 48(6), 732-748. *[Matching]*
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
- Kessler, J. B., & Roth, A. E. (2012). Organ allocation policy and the decision to donate. *American Economic Review*, 102(5), 2018-2047. *[DiD, dataset available]*
- Kreitmeier, D. & Raschky,, P. (2023): The Unintended Consequences of Censoring Digital Technology – Evidence from Italy's ChatGPT Ban. *Working paper*. *[DiD]*
- Kretschmer, T., & Peukert, C. (2020). Video killed the radio star? Online music videos and recorded music sales. *Information Systems Research*, 31(3), 776-800. *[DiD]*
- Lo, D., Brahm, F., Dessein, W., & Minami, C. (2022). Managing with Style? Microevidence on the Allocation of Managerial Attention. *Management Science*, 68(11), 8261-8285. *[DiD, dataset available]*
- Manacorda, M., Miguel, E., Vigorito, A. (2011). Government transfers and political support. *American Economic Journal: Applied Economics*, 3(3), 1-28. *[RDD, dataset available]*
- Olden, A., & Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*, 25(3), 531-553. *[Triple DiD]*
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- Zohrehvand, A., Doshi, A. R., & Vanneste, B. S. (2023). Generalizing event studies using synthetic controls: An application to the dollar tree–family dollar acquisition. *Long Range Planning*, 102392.
- Zuo, G. W. (2021). Wired and hired: Employment effects of subsidized broadband Internet for low-income Americans. *American Economic Journal: Economic Policy*, 13(3), 447-482. *[Triple DiD, dataset available]*

Additional resources used during the course

- Stata simulation Cunningham: Heterogeneous treatment effects: https://causalinf.substack.com/p/att-estimation-using-regression-and?utm_source=post-email-title&publication_id=306886&post_id=113916004&isFreemail=true
- Cunningham: Group-Time Heterogeneous ATT – Callaway and Sant'Anna (2020) estimator:
- Synthetic Control Groups Toolbox Stata:: https://yiqingxu.org/packages/fect/stata/fect_md.html, paper: https://polmeth.mit.edu/sites/default/files/documents/Yiqing_Xu.pdf
- Dataset, partially simulated, from Courthoud (2022). Synthetic Control with Python: https://matteocourthoud.github.io/post/synthetic_control. Python code to be adapted in Stata.

Detailed outline (subject to changes)

Session 1

[L0] Course presentation: outline, organization, and examination

[L1] Introduction and overview of the methods

- Why do we need "causal inference" tools?
- Exogenous vs. endogenous treatments
- The Average Treatment Effect
- The Counterfactual

Main references: Cameron and Trivedi (2022), Cunningham (2020), Gordon et al (2019).

Session 2

[L2] Matching methods

- Inverse Probability Weights
- Propensity Score Matching (PSM) & Nearest-Neighbors
- Coarsened Exact Matching (CEM)

Main references: Almond et al (2005), Cattaneo (2010), Cunningham (2020), Hitt and Frei (2002), Huntington-Klein (2021).

Datasets: `birth_weight` from Cattaneo (2010), `training_example` from Cunningham (2020)

Session 3

[L3a] Difference in Differences – Advanced tools

- Brief reminders on the DiD approach
- Unobserved individual heterogeneity
  - Repeated cross sections vs. panel data
  - Bias of the TWFE and the Bacon decomposition
- Heterogeneous treatment effects
  - Over time and across cohorts
  - Challenges associated with staggered DiD

Main references: Baker et al (2022), Callaway and Sant'Anna (2021), Cunningham (2020), Goodman-Bacon (2021), Kretschmer and Peukert (2020).

Datasets: `castle_doctrine_law` from Cheng and Hoekstra (2013), `callaway` from Cunningham (subset from Callaway and Sant'Anna (2021)).

Session 4

[L3b] Difference in Differences – Advanced tools

- Triple difference models
- Contamination, (spatial) spillovers: challenges and possible solutions
- Serial correlation and mismeasurement of standard errors

Main references : Bertrand et al (2004), Butt (2021), Boehmer et al (2020), Cameron and Trivedi (2022), Cunningham (2020), Olden and Moen (2022), Zuo (2021).

Datasets: `management_style` from Lo et al (2022), `organs` from Kessler and Roth (2012), `michelin` from Favaron et al (2022).

Session 5

[L4] Synthetic Control

- Comparative case studies
- Data requirements
- Estimation procedure
- Advantages and limitations of the method

Main references: Abadie et al (2010), Abadie et al (2015), Abadie (2021), Cameron and Trivedi (2022), Cunnigham (2020), Zohrehvand et al (2023).

Datasets: `texas` from Cornwell and Cunningham (2016), `repgermany` from Abadie et al (2015), `selfdriving_cars` from Courthoud (2022).

Session 6

[L5] Regression Discontinuity Design

- The RD design: running variable, cutoff, bandwidth
- Data requirement
- Estimation procedure
- Sharp and Fuzzy RDD
- Regression Discontinuity in Time (RDiT)
- Regression Kink Design

Main references: Cameron and Trivedi (2022), Cunningham (2020), Huntington-Klein (2021), Bradley et al (2017), Flammer and Bansal (2017).

Datasets: `gov_transfs` from Manacorda et al (2011), `mortgage` from Feller (2013), `diplomas` from Clark and Martorell (2014).