Syllabus for Empirical Methods in Economics and Management Studies[1]

Spring 2025

`Version: 2025-03-20`

Instructor:	Chen, Zhiyuan	**Office:**	919 Mingde Business.
E-mail:	chenzhiyuan@rmbs.ruc.edu.cn	**Time:**	Fri. 8:00-11:30
Location:	2203 No. 2 Public Building	**Office hours:**	By appointment
Course Platform:	WeChat/Github		

Description

This course is intended to train students in frontier research methods for good empirical and quantitative studies in economics and management studies.

It emphasizes a collaboration of various quantitative skills in conducting empirical research, as well as a tight connection between the theory and empirical facts. The course begins by introducing classical causal inference framework based on potential outcomes, and then moves to AI-powered causal inference methods. Then it introduces `Stata` and `Python` programming skills using various applications. Importantly We will further discuss to what extent researchers can rely on Diff-in-Diff approach to learn from data and guide policymaking. Some fundamentals and recent developments in machine learning and artificial intelligence will be introduced. The course concludes by discussing the modern academic research using LLM-based research agents.

[1] I will constantly update the syllabus, please make sure to keep track of the latest version.

Prerequisites

This course will assume a knowledge of graduate-level microeconomics, macroeconomics, and econometrics. Some experience in working with `Stata`, `Python`, and `Jupyterlab` is preferred, but not a must for this course. With the advancement in LLM-assisted programming (such as `Copilot, Cursor, Trae`, etc.), students are encouraged to learn the basic programming skills in `Python`.

Requirements and Grading

1. Class participation: Class discussion and interaction are highly valued in this class! It is therefore vital that you come to every class, prepared and ready to contribute.
2. Problem sets: You may be required to do some in-class quizzes and after-class problem sets. Some problem sets will be discussed in class and students are invited to present their solutions.
3. Research project/Presentation: You are also required to complete a research project/presentation:

   - The research project may be including (1) replicate and\or extend an existing paper or (2) a research proposal. I expect a specific research question, hypotheses, data set, and empirical methods. For the research proposal, you need to present it on June 13 and submit it before June 20.
   - The presentation should be on a recent paper from the reading list. You need to fully understand the paper. If possible, you may provide replication results. For the presentation, you need to present it on June 13.

   Students may work in groups; the maximum number of members in each group should not exceed three.
4. Grading: Your grade for the course will be determined by

Class participation(30%) + **Problem sets:** (20%) + **Research project:** (50%)

The Tentative Class Schedule[2]

PART I Reduced-form Tools

Textbooks & Lecture Notes

1. *J.D. Angrist and J.S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press, 2009.
2. * Victor Chernozhukov, Christian Hansen, Nathan Kallus, Martin Spindler and Vasilis Syrgkanis, *Applied Causal Inference Powered by ML and AI*, 2024, [Link to the book]
3. *Matheus Facure, *Causal Inference for Brave and True*, 2025, [Link to Chinese Version]
4. Diff-In-Diffs Resources: Pedro Sant'Anna's website

Programming Prerequisites

1. * Github Intro: `https://cloud.tencent.com/developer/article/1807687`
2. * Python in VS Code: `https://code.visualstudio.com/docs/python/python-tutorial`
3. * GitHub Copilot in VS Code: `https://code.visualstudio.com/docs/copilot/overview`

Class Schedule

- **Class 1–3 (3/21, 3/28, 4/11)– Causal inference, linear regressions, IV regressions, Diff-in-Diffs**

  Programming/Technical References:

[2]Symbol meaning: * means important.

- * Open-source replication files for *Mostly Harmless Econometrics*: https://github.com/vikjam/mostly-harmless-replication
- * Stata_kernel in Jupyterlab: https://kylebarron.dev/stata_kernel/
- * Call Stata from Python: https://www.stata.com/python/pystata18/#

Readings:

- * Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020): "*Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics*." American Economic Review, 110 (11): 3634-60.
- Christine Cai, 2020, *Literature on Recent Advances in Applied Micro Methods.* [PDF]
- Jeff Woodridge 2007 NBER Lecture "Difference in Differences Estimation"
- Abadie, A., and M. D. Cattaneo. (2018): "*Econometric Methods for Program Evaluation*," Annual Review of Economics, 10, 465–503.

- **Class 4 (4/18) More on Diff-in-Diffs: Clustering and Various Extensions of Diff-in-Diffs**

References:

- *Marianne Bertrand, Esther Duflo, Sendhil Mullainathan, *How Much Should We Trust Differences-In-Differences Estimates?*, The Quarterly Journal of Economics, Volume 119, Issue 1, February 2004, Pages 249–275
- *Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A. and Sant'Anna, P.H., 2025. Difference-in-Differences Designs: A Practitioner's Guide. arXiv preprint arXiv:2503.13323.
- *Roth, Jonathan, et al. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." Journal of Econometrics (2023).

- *Cameron, A.Colin, and Douglas L. Miller. "*A Practitioner's Guide to Cluster-Robust Inference*." Journal of Human Resources, vol. 50, no. 2, 2015, pp. 317–372.
- Abadie, A., A. Diamond, and A. J. Hainmueller. (2010): "*Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program*," Journal of the American Statistical Association, 105, 493–505.
- Abadie, Alberto. "*Semiparametric difference-in-differences estimators.*" The Review of Economic Studies 72.1 (2005): 1-19.
- Arkhangelsky, Dmitry, et al. *Synthetic difference in differences.* No. w25532. National Bureau of Economic Research, 2019.
- Athey, S., and Imbens, G. W. (2022). *Design-based analysis in difference-in-differences settings with staggered adoption.* Journal of Econometrics, 226(1), 62-79.
- Callaway, Brantly, and Pedro HC Sant'Anna. "*Difference-in-differences with multiple time periods.*" Journal of Econometrics 225.2 (2021): 200-230.
- De Chaisemartin, Clement, and Xavier d'Haultfoeuille. "*Two-way fixed effects estimators with heterogeneous treatment effects.*" American Economic Review 110.9 (2020): 2964-96.
- Goodman-Bacon, Andrew. "*Difference-in-differences with variation in treatment timing.*" Journal of Econometrics 225.2 (2021): 254-277.
- Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. "*Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme.*" The Review of Economic Studies 64.4 (1997): 605-654.
- Sun, Liyang, and Sarah Abraham. "*Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.*" Journal of Econometrics 225.2 (2021): 175-199.

PART II Causal Inference Powered by ML and AI

- **Class 5–6 (4/25, 5/9)** – Machine Learning with Applications in Python

  References

  - *Gentzkow, M., B. Kelly, and M. Taddy. (2019): "*Text as Data*," Journal of Economic Literature, 57, 535–74. [PDF]
  - *Athey, S., and G. W. Imbens. (2019): "*Machine Learning Methods That Economists Should Know About*," Annual Review of Economics, 11, 685–725. [PDF]
  - *Cameron, A. Colin, 2019, *Machine Learning Methods in Economics*, slides

  Readings:

  A. Machine Learning in Economics

  - *Hoberg, G., and G. Phillips. (2016), *Text-based network industries and endogenous product differentiation*, Journal of Political Economy, 124, 1423–65.
  - *Susan Athey (2018), *The Impact of Machine Learning on Economics*. [PDF]
  - Sendhil Mullainathan and J. Spies, 2017, *Machine Learning: An Applied Econometric Approach*, Journal of Economic Perspectives, 87-106.
  - *Gareth James, Daniela Witten, Trevor Hastie and Robert Tibsharani (2013), *An Introduction to Statistical Learning: with Applications in R*, Springer. [Python Code]
  - Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

  B. Text-analysis

  - *Scott R. Baker, Nicholas Bloom, Steven J. Davis, *Measuring Economic Policy Uncertainty*, The Quarterly Journal of Economics, Volume 131, Issue 4, November 2016, Pages 1593–1636.

- *Kelly, B. T., D. Papanikolaou, A. Seru, and M. Taddy. (2020): "*Measuring Technological Innovation over the Long Run*," NBER working paper w25266.
- Raffo, J., and S. Lhuillery. (2009): *"How to play the ' Names Game ': Patent retrieval comparing different heuristics,"* Research Policy, 38, 1617–27.

- **Class 7 (5/16)** – AI-powered Causal Inference

  References

  - Causal Machine Learning's GitHub Repository

  Readings

  - Vashishtha, A., Reddy, A. G., Kumar, A., Bachu, S., Balasubramanian, V. N., and Sharma, A. (2023). *Causal inference using LLM-guided discovery.* arXiv preprint arXiv:2310.15117.

- **Class 8 (5/23)**– Policy Learning through Experimentation and Observational Studies

  - * Banerjee, A.V., Chassang, S., Montero, S., Snowberg, E., 2020. *A Theory of Experimenters: Robustness, Randomization, and Balance.* American Economic Review 110, 1206–1230.
  - Higbee, S.D., *Experimental Design for Policy Choice*, 2025, Job Market Paper, University of Chicago.
  - Wang, Shaoda, and David Y. Yang. *Policy Experimentation in China: the Political Economy of Policy Learning.* NBER No. w29402, 2021.

- **Class 9 (5/30)** – Academic Research with LLM-based Research Agents

  References:

  - *Manning, B.S., Zhu, K., Horton, J.J., 2024, Automated Social Science: Language Models as Scientist and Subjects.https://arxiv.org/abs/2404.11794

  - *Novy-Marx, R. and Velikov, M. Z. (2025). AI-powered (finance) scholarship (No. w33363). National Bureau of Economic Research.
  - Gao, Chen, et al. "Large language models empowered agent-based modeling and simulation: A survey and perspectives." Humanities and Social Sciences Communications 11.1 (2024): 1-24.

- **Class 10 (6/6)** – Student presentations: *Presenting a paper from the reading list OR Presenting the proposal of the research project*