Applied Econometrics: ML
Module 2, 2025

Course Information

Instructor: Jizhou Liu
Office: PHBS Building, Room 659
Phone: 86-755-2603-9236
Email: jizhou.liu@phbs.pku.edu.cn
Office Hour: TBD

Teaching Assistant:
Phone: TBD
Email: TBD

Classes:
Lectures: Mon & Thur 13:30-15:20
Venue: PHBS Building, Room TBD

Course Website:
If any.

1. Course Description

1.1 Context

Course overview:

This course is designed for PhD students, focusing on Machine Learning and Causal Inference methods, such as Augmented Inverse Probability Weighting estimators, Double/Debiased Machine Learning, Causal Forest, Causal Matrix Completion and Randomized trials (A/B testing).

Prerequisites:

Students are expected to have a solid graduate-level understanding of econometrics or statistics.

1.2 Textbooks and Reading Materials

Textbooks:

Stefan Wager (2024). Causal Inference: A Statistical Learning Approach

Victor Chernozhukov, Christian Hansen, Nathan Kallus, Martin Spindler, Vasilis Syrgkanis (2025). Applied Causal Inference Powered by ML and AI

Reading Materials:

1. Athey, S., Bayati, M., Doudchenko, N., Imbens, G. W., and Khosravi, K. (2018a). Matrix completion methods for causal panel data. arXiv preprint arXiv:1710.10251.

2. Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523):1228–1242.

3. Abadie, A., Diamond, A., and 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.

4. Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12):4088–4118.

5. Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. Journal of the American Statistical Association, 116(536):1789–1803.

6. Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. Annals of Statistics, 40(2):1263–1282.

7. Young, A. (2018). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. The Quarterly Journal of Economics, 134 557–598.

8. Canay, I. A. and Kamat, V. (2017). Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. The Review of Economic Studies, 85 1577–1608.

9. Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018), "Inference Under Covariate-Adaptive Randomization,"Journal of the American Statistical Association,113,1784–1796.

10. Susan Athey, Raj Chetty, Guido W Imbens, Hyunseung Kang, The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely, The Review of Economic Studies, 2025 ;, rdaf087

11. Bai, Yuehao, Joseph P. Romano, and Azeem M. Shaikh. 2021. "Inference in Experiments with Matched Pairs." Journal of the American Statistical Association 1–12.

2. Learning Outcomes

2.1 Intended Learning Outcomes

Learning Goals	**Objectives**	**Assessment (YES with details or NO)**
1. Our graduates will be effective communicators.	1.1. Our students will produce quality business and research-oriented documents.	Yes
	1.2. Students are able to professionally present their ideas and also logically explain and defend their argument.	Yes
2. Our graduates will be skilled in team work and leadership.	2.1. Students will be able to lead and participate in group for projects, discussion, and presentation.	Yes
	2.2. Students will be able to apply leadership theories and related skills.	Yes
3. Our graduates will be trained in ethics.	3.1. In a case setting, students will use appropriate techniques to analyze business problems and identify the ethical aspects, provide a solution and defend it.	Yes
	3.2. Our students will practice ethics in the duration of the program.	Yes

4. Our graduates will have a global perspective.	4.1. Students will have an international exposure.	Yes
5. Our graduates will be skilled in problem-solving and critical thinking.	5.1. Our students will have a good understanding of fundamental theories in their fields.	Yes
	5.2. Our students will be prepared to face problems in various business settings and find solutions.	Yes
	5.3. Our students will demonstrate competency in critical thinking.	Yes

2.2 Course specific objectives

This course aims to help students develop the ability to read, interpret, and critically evaluate econometrics and statistics research papers on causal inference and machine learning. Students will learn the foundational econometric and statistical frameworks that underpin modern causal machine learning methods and gain the skills to independently learn and adapt new techniques as the field evolves. The course also equips students with a practical toolbox for addressing industry-motivated causal inference problems, bridging rigorous methodology with real-world applications.

2.3 Assessment/Grading Details

Assessment Methods	Brief Description (Optional)	Weight
A1. In-class & forum participation	Attendance & discussions	10%
A2. Assignments	Effort and accuracy	20%
A3. Paper presentation and Project	Effort and accuracy	70%
	Total	100%

2.4 Academic Honesty and Plagiarism

It is important for a student's effort and credit to be recognized through class assessment. Credits earned for a student work due to efforts done by others are clearly unfair. Deliberate dishonesty is considered academic misconducts, which include plagiarism; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.

All assessments are subject to academic misconduct check. Misconduct check may include reproducing the assessment, providing a copy to another member of faculty, and/or communicate a copy of this assignment to the PHBS Discipline Committee. A suspected plagiarized document/assignment submitted to a plagiarism checking service may be kept in its database for future reference purpose.

Where violation is suspected, penalties will be implemented. The penalties for academic misconduct may include: deduction of honour points, a mark of zero on the assessment, a fail grade for the whole course, and reference of the matter to the Peking University Registrar.

AI tools requirements：

Using AI tools to complete assignments or assessments without the approval of the course instructor will be regarded as an act of academic dishonesty. Depending on the severity of the situation, penalties will be implemented in accordance with the provisions of the Peking University Graduate Student Handbook.

For more information of plagiarism, please refer to *PHBS Student Handbook*.

3. Topics, Teaching and Assessment Schedule

Topic 1 — Potential Outcomes & Identification
Potential outcome (PO) framework; ATE/CATE/ITE; SUTVA; ignorability/unconfoundedness; RCT as identification device; overlap/positivity.

Topic 2 — IPW, Oracle IPW, AIPW/DR; DML1 vs. DML2
Propensity score as covariate balancing; Oracle IPW vs. estimated IPW; augmentation; influence functions; Neyman orthogonality; cross-fitting; DML1 vs. DML2 constructions and inference.

Topic 3 — Trees→Random Forests→Causal Forests
CART bias–variance; honesty; splitting for heterogeneity; asymptotics & variance for GRF; inference on CATE

Topic 4 — Causal Matrix Completion I (Foundations)
Netflix competition as motivation; low-rank structure & nuclear-norm penalization; missing not at random vs. panel missingness; bridging prediction and counterfactuals.

Topic 5 — Causal Matrix Completion II (Panel Causal Effects)
Matrix completion for panel causal data; relation to interactive fixed effects; comparisons

Topic 6 — Synthetic Control (SC) & Synthetic DiD (SDID) + Augmented SC
SC identification & convex weights; placebo and inference; SDID construction and properties; Augmented SC (balance + outcome modeling).

Topic 7 — Randomized Trials & A/B Testing; Role of Covariates
Revisiting unconfoundedness under randomization; blocking/stratification; matched pairs; covariate-adaptive randomization; rerandomization; variance impacts; covariate adjustment post-randomization.

Topic 8 — Randomization Inference (RI) & Conditional RI; ML-Assisted Tests
Fisher exact test; sharp vs. weak nulls; studentized statistics; conditional randomization tests; using ML predictions as test statistics while preserving validity.

4. Miscellaneous

TBD