MACHINE LEARNING AND CAUSAL INFERENCE

Period: a.y. 2025/2026

Class times: II

Instructor:

Prof. Marton Varga
Dept. of Marketing. - Room 4-D1-11
marton.varga@unibocconi.it

Course description

This PhD course provides an in-depth exploration of causal inference and machine learning, with a particular emphasis on their intersection in empirical research. These methods are foundational in contemporary studies **across management and marketing**, playing a critical role in high-impact academic publications.

Throughout the course, you will:

- Develop a strong theoretical understanding of **causal inference**, including the principles underlying causal identification and the estimation of treatment effects.
- Learn how to apply **machine learning techniques** to improve causal estimation, refine predictions, and enhance the robustness of empirical findings.
- Examine **recent high-impact research papers** to understand how these methodologies are implemented in real-world management and marketing studies.
- Gain practical experience with **R programming**, working with key statistical packages and tools for empirical research.

The course combines theoretical foundations with hands-on applications, ensuring that participants are well-equipped to incorporate these advanced methods into their own research. By the end, you will have the knowledge and technical skills to rigorously design, implement, and evaluate empirical studies using the latest advancements in causal inference and machine learning. This course is particularly valuable for PhD students aiming to apply cutting-edge quantitative methods in management, marketing, and related disciplines.
Please note that not all the papers above will be covered in class due to time constraints. The reading list is subject to changes.

Course Material

Session 1: Treatment Estimation

- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics, 10*(1), 465–503.
- Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *The American Economic Review, 80*(3), 313–336.
- Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal, 114*(494), C52–C83.
- Angrist, J. D., & Rokkanen, M. (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association, 110*(512), 1331–1344.
- Card, D., & Krueger, A. (1994). Minimum wages and employment: Case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84*(4), 772–793.
- Anderson, M., & Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal, 122*(563), 957–989.

Session 2: Prediction

- Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning: Methods and applications* (pp. 157–175). Springer.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 58*(1), 267–288.
- Jawadekar, N., Kezios, K., Odden, M. C., Stingone, J. A., Calonico, S., Rudolph, K., & Zeki Al Hazzouri, A. (2023). Practical guide to honest causal forests for identifying heterogeneous treatment effects. *American Journal of Epidemiology, 192*(7), 1155–1165.
- Jakobsen, K. D. (2024). Re: "Practical guide to honest causal forests for identifying heterogeneous treatment effects." *American Journal of Epidemiology, 193*(5), 811–812.

Session 3: State of Machine Learning and Casual Inference in Research

- Athey, S., Imbens, G., Pham, T., & Wager, S. (2017). Estimating average treatment effects: Supplementary analyses and remaining challenges. *American Economic Review, 107*(5), 278–281.
- Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science, 355*(6324), 483–485.
- Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University

of Chicago Press.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives, 28*(2), 29–50.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28.
- Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics, 235*(2), 2218–2244.

Session 4: Advanced Methods for Treatment Estimation

- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review, 93*(1), 113–132.
- 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.
- Ben-Michael, E., Feller, A., & Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 84*(2), 351–381.
- Wing, C., Freedman, S. M., & Hollingsworth, A. (2024). Stacked difference-in-differences (No. w32054). National Bureau of Economic Research.
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242.
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. The Annals of Statistics, 47(5), 1148–1178.
- Athey, S., Blei, D., Donnelly, R., Ruiz, F., & Schmidt, T. (2018). Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. In AEA Papers and Proceedings, 108, 64–67. American Economic Association.
- Athey, S., Imbens, G. W., & Wager, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. Journal of the Royal Statistical Society Series B: Statistical Methodology, 80(4), 597–623.

Session 5: Applications in Management

- Li, Z., & Wang, G. (2024). On-demand delivery platforms and restaurant sales. *Management Science.*
- Li, X., Deng, Y., Manchanda, P., & De Reyck, B. (2022). Can lower (ed) expert opinions lead to better consumer ratings? The case of Michelin

stars. *Forthcoming in Management Science.*

- Yılmaz, Ö., Son, Y., Shang, G., & Arslan, H. A. (2024). Causal inference under selection on observables in operations management research: Matching methods and synthetic controls. *Journal of Operations Management, 70*(5), 831–859.

Session 6: Applications in Marketing

- He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science, 41*(5), 896–921.
- Adalja, A., Liaukonytė, J., Wang, E., & Zhu, X. (2023). GMO and non-GMO labeling effects: Evidence from a quasi-natural experiment. *Marketing Science, 42*(2), 233–250.
- Lambrecht, A., Tucker, C., & Zhang, X. (2024). TV advertising and online sales: A case study of intertemporal substitution effects for an online travel platform. *Journal of Marketing Research, 61*(2), 248–270.
- Donnelly, R., Ruiz, F. J., Blei, D., & Athey, S. (2021). Counterfactual inference for consumer choice across many product categories. *Quantitative Marketing and Economics, 19*(1), 1–39.

Assessment Methods

Effective class participation includes attendance to sessions and making an active and constructive contribution to the discussion having read in advance the suggested papers and making constructive comments. Finally, participants will be asked to submit a proposal for a conceptual paper.

Project: Concept Paper

For this assignment, you are required to create a concept paper on a topic related to management or marketing. The paper should have an empirical focus using secondary (i.e., observational) dataset(s). Importantly, you are not required to download or analyze data: you are only required to conceptualize how you would analyze that data if it were available to you.

Your task is to:

1. Define the literature contribution: Briefly state the existing literature and identify the gap your research will address.

2. Describe the data source(s): Explain the data source(s) you would use for this research and the structure of the data. This could include panel data, cross-sectional data, scraped data, administrative data, or other data collection methods. The suggested data must be observational data (i.e. surveys, lab experiments are not allowed). Cleary define what is the unit of observation.

3. Outline the empirical approach: Discuss the empirical approach you would apply to this data (if it were available to you), including the type of analysis. You should carefully design this with equations and visual illustrations of the identification and/or algorithmic approach. You must apply an empirical approach we covered during class or one that is very closely related.

4. Expected results: Briefly state the expected results or outcomes from your analysis.

The grading will place significant **emphasis on part 2 and 3**, which focuses on how you would analyze the data you suggest. You must apply an empirical approach that we have covered in class or one that is closely related.

The paper's length is capped at 10 pages, using Times New Roman, 12 pt. font, and 2 line spacing. There's no requirement for an extensive literature review, keeping it concise and focused on your idea.

The final grade for this course will be determined based on the following components:

Concept Paper	70%
Class Participation	30%

Faculty Bio.

Marton Varga is an Assistant Professor of Marketing at Bocconi University. His research focuses on online product search and purchasing behavior, particularly in the context of user-generated content. He is also interested in the effects of product recalls, including how they impact dealer prices, as well as how quality cues on online platforms influence pricing. Additionally, he explores how fMRI studies can improve forecasts for product sales. His work has been published in *Journal of Marketing Research*, *Health Economics*, and *Portuguese Economic Journal*. He earned his Ph.D. from INSEAD and Sorbonne University and previously worked as an economist, gaining experience in public policy. He also holds Master's degrees from universities in Budapest and Lisbon.