Causal Machine Learning References

- Xiaoyang Ye (xiaoyang.ye26@gmail.com)

- updated: 12.31.2024

Part 1. Causal Inference

Section 1.1 Textbooks

- Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press. https://www.mostlyharmlesseconometrics.com/
- Pearl, J. (2009). Causality. Cambridge University Press.
- Wooldridge, Jeffrey. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed. MIT Press.
- Hayashi, F. (2011). Econometrics. Princeton University Press.
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
- Cunningham, S. (2021). Causal inference: The mixtape. Yale University Press. https://mixtape.scunning.com/
- Huntington-Klein, N. (2021). The effect: An introduction to research design and causality. CRC Press. https://theeffectbook.net/
- Hansen, B. (2022). Econometrics. Princeton University Press.
- Facure, M. (2023). Causal Inference in Python: Applying Causal Inference in the Tech Industry. O'Reilly Media, Inc. https://matheusfacure.github.io/python-causality-handbook/landing-page.html
- Ding, P. (2024). A first course in causal inference. CRC Press. https://arxiv.org/pdf/2305.18793

Section 1.2 Reviews

- LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. American Economic Review, 604-620.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. Review of Economics and Statistics, 86(1), 4-29.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322-331.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47(1), 5-86.
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. Journal of Economic perspectives, 31(2), 3-32.
- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. Annual Review of Economics, 10, 465-503.
- Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. Journal of Economic Literature, 58(4), 1129-1179.

- Angrist, J. D. (2022). Empirical strategies in economics: Illuminating the path from cause to effect. Econometrica, 90(6), 2509-2539.
- Imbens, G. W. (2022). Causality in Econometrics: Choice vs Chance. Econometrica, 90(6), 2541-2566.
- Imbens, G. W. (2024). Causal inference in the social sciences. Annual Review of Statistics and Its Application, 11: 123-152.

Section 1.3 RCT

- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. Handbook of Development Economics, 4, 3895-3962.
- Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In Handbook of Economic Field Experiments (Vol. 1, pp. 73-140). North-Holland.
- DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. Econometrica, 90(1), 81-116.
- List, J. (2025). Experimental Economics: Theory and Practice. University of Chicago Press.

- Baird, S., Bohren, J. A., McIntosh, C., & Özler, B. (2018). Optimal design of experiments in the presence of interference. Review of Economics and Statistics, 100(5), 844-860.
- Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Quarterly Journal of Economics, 134(2), 557-598.
- Azevedo, E. M., Deng, A., Montiel Olea, J. L., Rao, J., & Weyl, E. G. (2020). A/b testing with fat tails. Journal of Political Economy, 128(12), 4614-000.
- Banerjee, A. V., Chassang, S., Montero, S., & Snowberg, E. (2020). A theory of experimenters: Robustness, randomization, and balance. American Economic Review, 110(4), 1206-1230.
- Bai, Y. (2022). Optimality of matched-pair designs in randomized controlled trials. American Economic Review, 112(12), 3911-3940.
- Bai, Y., Romano, J. P., & Shaikh, A. M. (2022). Inference in experiments with matched pairs. Journal of the American Statistical Association, 117(540), 1726-1737.
- Bai, Y., Jiang, L., Romano, J. P., Shaikh, A. M., & Zhang, Y. (2024). Covariate adjustment in experiments with matched pairs. Journal of Econometrics, 241(1), 105740.
- Lu, X., Liu, T., Liu, H., & Ding, P. (2023). Design-based theory for cluster rerandomization. Biometrika, 110(2), 467-483.
- Ghanem, D., Hirshleifer, S., & Ortiz-Becerra, K. (2023). Testing attrition bias in field experiments. Journal of Human Resources.
- Fuster, A., & Zafar, B. (2023). Survey experiments on economic expectations. In Handbook of Economic Expectations (pp. 107-130). Academic Press.
- Vazquez-Bare, G. (2023). Identification and estimation of spillover effects in randomized experiments. Journal of Econometrics, 237(1), 105237.
- Chen, J., & Andrews, I. (2023). Optimal conditional inference in adaptive experiments. arXiv preprint arXiv:2309.12162.
- Muralidharan, K., Romero, M., & Wüthrich, K. (2024). Factorial designs, model selection, and (incorrect) inference in randomized experiments. Review of Economics and Statistics, 1-44.

- Xiong, R., Athey, S., Bayati, M., & Imbens, G. (2024). Optimal experimental design for staggered rollouts. Management Science, 70(8), 5317-5336.
- Athey, S., Chetty, R., Imbens, G. W., & Kang, H. (2024). The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely.
- Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., & Pathak, P. A. (2017). Research design meets market design: Using centralized assignment for impact evaluation. Econometrica, 85(5), 1373-1432.
- Andrews, I., Kitagawa, T., & McCloskey, A. (2024). Inference on winners. The Quarterly Journal of Economics, 139(1), 305-358.
- List, J. A., Muir, I., & Sun, G. (2024). Using machine learning for efficient flexible regression adjustment in economic experiments. Econometric Reviews, 1-39.

Section 1.4 RD

- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2024). A practical introduction to regression discontinuity designs: Extensions. Cambridge University Press.

- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69(1), 201-209.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142(2), 615-635.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. Journal of Economic Literature, 48(2), 281-355.
- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. The Review of Economic Studies, 79(3), 933-959.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82(6), 2295-2326.
- Card, D., Lee, D. S., Pei, Z., & Weber, A. (2015). Inference on causal effects in a generalized regression kink design. Econometrica, 83(6), 2453-2483.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. Review of Economics and Statistics, 101(3), 442-451.
- Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. Journal of the American Statistical Association, 115(531), 1449-1455.
- Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. Annual Review of Economics, 14(1), 821-851.
- Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., & Pathak, P. (2022). Breaking ties: Regression discontinuity design meets market design. Econometrica, 90(1), 117-151.

Section 1.5 IV

- Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). Journal of Economic literature, 48(2), 399-423.

- Wooldridge, J. M. (2015). Control function methods in applied econometrics. Journal of Human Resources, 50(2), 420-445.
- Mogstad, M., & Torgovitsky, A. (2024). Instrumental Variables with Unobserved Heterogeneity in Treatment Effects (No. w32927). National Bureau of Economic Research.

- Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. Econometrica, 62(2), 467-475.
- Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American statistical Association, 90(430), 431-442.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American statistical Association, 91(434), 444-455.
- Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. Journal of Human Resources, 441-462.
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. Journal of Economic Perspectives, 15(4), 69-85.
- Carneiro, P., Heckman, J. J., & Vytlacil, E. J. (2011). Estimating marginal returns to education. American Economic Review, 101(6), 2754-2781.

- Wooldridge, J. M. (2015). Control function methods in applied econometrics. Journal of Human Resources, 50(2), 420-445.

- Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics, 11(1), 727-753.
- Blandhol, C., Bonney, J., Mogstad, M., & Torgovitsky, A. (2022). When is TSLS actually late? (No. w29709). Cambridge, MA: National Bureau of Economic Research.
- Adao, R., Kolesár, M., & Morales, E. (2019). Shift-share designs: Theory and inference. The Quarterly Journal of Economics, 134(4), 1949-2010.
- Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586-2624.
- Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. The Review of Economic Studies, 89(1), 181-213.
- Heinesen, E., Hvid, C., Kirkebøen, L. J., Leuven, E., & Mogstad, M. (2022). Instrumental variables with unordered treatments: Theory and evidence from returns to fields of study (No. w30574). National Bureau of Economic Research.
- Mogstad, M., Torgovitsky, A., & Walters, C. R. (2024). Policy evaluation with multiple instrumental variables. Journal of Econometrics, 105718.
- Canay, I. A., Mogstad, M., & Mountjoy, J. (2024). On the use of outcome tests for detecting bias in decision making. Review of Economic Studies, 91(4), 2135-2167.
- Frandsen, B., Lefgren, L., & Leslie, E. (2023). Judging judge fixed effects. American Economic Review, 113(1), 253-277.
- Chyn, E., Frandsen, B., & Leslie, E. C. (2024). Examiner and Judge Designs in Economics: A Practitioner's Guide (No. w32348). National Bureau of Economic Research.
- Borusyak, K., Hull, P., & Jaravel, X. (2004). A Practical Guide to Shift-Share Instruments (No. w33236). National Bureau of Economic Research.

Section 1.6 DID (& SC)

- Roth, J., Sant'Anna, P. H., 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.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. The Econometrics Journal, 26(3), C1-C30.
- Arkhangelsky, D., & Imbens, G. (2024). Causal models for longitudinal and panel data: A survey. The Econometrics Journal, 27(3), C1-C61.

- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. American Economic Review, 84(4), 772-793.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. Quarterly Journal of Economics, 119(1), 249-275.
- Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. Econometrica, 74(2), 431-497.
- 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.
- Doudchenko, N., & Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis (No. w22791). National Bureau of Economic Research.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964-2996.
- Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. Journal of Econometrics, 219(1), 101-122.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175-199.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12), 4088-4118.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. Journal of the American Statistical Association, 116(536), 1716-1730.
- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature, 59(2), 391-425.
- Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. Journal of Econometrics, 226(1), 62-79.
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. Review of Economic Studies, 90(5), 2555-2591.
- Roth, J., & Sant'Anna, P. H. (2023). Efficient estimation for staggered rollout designs. Journal of Political Economy Microeconomics, 1(4), 669-709.
- Roth, J. (2024). Interpreting event-studies from recent difference-in-differences methods. arXiv preprint arXiv:2401.12309.

- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). Difference-in-differences with a continuous treatment (No. w32117). National Bureau of Economic Research.
- Bellégo, C., Benatia, D., & Dortet-Bernadet, V. (2024). The chained difference-in-differences. Journal of Econometrics, 105783.

Section 1.7 Matching

- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1), 31-72.
- Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. Econometrica, 84(2), 781-807.

- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.
- Smith, J. A., & Todd, P. E. (2001). Reconciling conflicting evidence on the performance of propensity-score matching methods. American Economic Review, 91(2), 112-118.
- Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. Review of Economics and Statistics, 84(1), 151-161.
- Heckman, J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. Review of Economics and Statistics, 86(1), 30-57.
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. Journal of Econometrics, 125(1-2), 305-353.
- Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4), 962-973.
- Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. Econometrica, 74(1), 235-267.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.
- King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. Political Analysis, 27(4), 435-454.
- Black, B. S., Lalkiya, P., & Lerner, J. Y. (2020). The trouble with coarsened exact matching. Northwestern Law & Econ Research Paper.
- Ma, X., & Wang, J. (2020). Robust inference using inverse probability weighting. Journal of the American Statistical Association, 115(532), 1851-1860.

Section 1.8 Additional Topics

Linear Regression

- Goldsmith-Pinkham, P., Hull, P., & Kolesár, M. (2024). Contamination Bias in Linear Regressions. American Economic Review.
- Chen, J., & Roth, J. (2024). Logs with zeros? Some problems and solutions. The Quarterly Journal of Economics, 139(2), 891-936.

Standard Errors

- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317-372.
- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. Econometrica, 88(1), 265-296.
- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering?. The Quarterly Journal of Economics, 138(1), 1-35.
- MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. Journal of Econometrics, 232(2), 272-299.
- De Chaisemartin, C., & Ramirez-Cuellar, J. (2024). At what level should one cluster standard errors in paired and small-strata experiments?. American Economic Journal: Applied Economics, 16(1), 193-212.

P-hacking

- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. American Economic Review, 110(11), 3634-3660.
- Kasy, M. (2021). Of forking paths and tied hands: Selective publication of findings, and what economists should do about it. Journal of Economic Perspectives, 35(3), 175-192.
- Elliott, G., Kudrin, N., & Wüthrich, K. (2022). Detecting p-hacking. Econometrica, 90(2), 887-906.
- Lang, K. (2023). How Credible is the Credibility Revolution? (No. w31666). National Bureau of Economic Research.
- Brodeur, A., Carrell, S., Figlio, D., & Lusher, L. (2023). Unpacking p-hacking and publication bias. American Economic Review, 113(11), 2974-3002.
- Brodeur, A., Cook, N. M., Hartley, J. S., & Heyes, A. (2024). Do Preregistration and Preanalysis Plans Reduce p-Hacking and Publication Bias? Evidence from 15,992 Test Statistics and Suggestions for Improvement. Journal of Political Economy Microeconomics, 2(3), 527-561.
- Dreber, A., Johannesson, M., & Yang, Y. (2024). Selective reporting of placebo tests in top economics journals. Economic Inquiry.

Weighting

- Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for?. Journal of Human Resources, 50(2), 301-316.

Binscatter

- Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2024). On binscatter. American Economic Review, 114(5), 1488-1514.

Many people and replicability

- Miguel, E. (2021). Evidence on research transparency in economics. Journal of Economic Perspectives, 35(3), 193-214.
- Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., ... & Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. Economic Inquiry, 59(3), 944-960.

- Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H., Adem, M., Adriaans, J., ... & Van Assche, J. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proceedings of the National Academy of Sciences, 119(44), e2203150119.
- Ankel-Peters, J., Fiala, N., & Neubauer, F. (2023). Do economists replicate?. Journal of Economic Behavior & Organization, 212, 219-232.
- Ankel-Peters, J., Fiala, N., & Neubauer, F. (2023). Is economics self-correcting? Replications in the American Economic Review. Economic Inquiry.
- Dreber, A., & Johannesson, M. (2024). A framework for evaluating reproducibility and replicability in economics. Economic Inquiry.
- Fišar, M., Greiner, B., Huber, C., Katok, E., Ozkes, A. I., & Management Science Reproducibility Collaboration. (2024). Reproducibility in Management Science. Management Science, 70(3), 1343-1356.

Selection bias in surveys

- Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., & Van Dijk, W. (2021). Selection in surveys: Using randomized incentives to detect and account for nonresponse bias (No. w29549). National Bureau of Economic Research.
- Dutz, D., Greenstone, M., Hortaçsu, A., Lacouture, S., Mogstad, M., Shaikh, A. M., ... & van Dijk, W. (2024). Nonrepresentativeness in Population Health Research: Evidence from a COVID-19 Antibody Study. American Economic Review: Insights, 6(3), 313-323.

Part 2. Machine Learning

Section 2.1 Textbooks and Reviews

- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An Introduction to Statistical Learning: With Applications in Python. Springer Nature.
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). Dive into Deep Learning. Cambridge University Press.
- Bishop, C. M., & Bishop, H. (2024). Deep Learning: Foundations and Concepts. Springer Nature.

Section 2.2 Papers

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention Is All You Need. NIPS 2017.

- TO ADD: landmark papers in machine learning - https://github.com/daturkel/learning-papers

Section 3.1 Textbooks and Reviews

- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11(1), 685-725.
- Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024). Applied causal inference powered by ML and AI. https://causalml-book.org/
- Wager, S. (2024). Causal Inference: A Statistical Learning Approach. https://web.stanford.edu/~swager/causal_inf_book.pdf
- Kanus, M. Causal ML. https://github.com/MCKnaus/causalML-teaching

Section 3.2 Double Machine Learning & Other Learners

- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1-C68.
- Colangelo, K., & Lee, Y. Y. (2020). Double debiased machine learning nonparametric inference with continuous treatments. arXiv preprint arXiv:2004.03036.
- Semenova, V., & Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. The Econometrics Journal, 24(2), 264-289.
- Knaus, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. The Econometrics Journal, 25(3), 602-627.
- Foster, D. J., & Syrgkanis, V. (2023). Orthogonal statistical learning. The Annals of Statistics, 51(3), 879-908.

Section 3.3 Heterogeneous Treatment Effects

- Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India (No. w24678). National Bureau of Economic Research.
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests.
- 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., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. Observational studies, 5(2), 37-51.

Section 3.4 Adaptive Experiments

- Kasy, M., & Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. Econometrica, 89(1), 113-132.
- Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., & Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. Proceedings of the National Academy of Sciences, 118(15), e2014602118.
- Wei, W., Ma, X., & Wang, J. (2024). Fair adaptive experiments. Advances in Neural Information Processing Systems, 36.

Section 3.5 Policy Learning

- Athey, S., & Wager, S. (2021). Policy learning with observational data. Econometrica, 89(1), 133-161.
- Zhou, Z., Athey, S., & Wager, S. (2023). Offline multi-action policy learning: Generalization and optimization. Operations Research, 71(1), 148-183.

Section 3.6 Big Data Causal Inference

- Angrist, J. D., & Frandsen, B. (2022). Machine labor. Journal of Labor Economics, 40(S1), S97-S140.
- Farrell, M. H., Liang, T., & Misra, S. (2021). Deep neural networks for estimation and inference. Econometrica, 89(1), 181-213.
- Ash, E., & Hansen, S. (2023). Text algorithms in economics. Annual Review of Economics, 15(1), 659-688.

Section 3.7 GenAI

- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.
- Bonnefon, J. F., Rahwan, I., & Shariff, A. (2024). The moral psychology of Artificial Intelligence. Annual Review of Psychology, 75(1), 653-675.
- Ross, J., Kim, Y., & Lo, A. W. (2024). LLM economicus? Mapping the Behavioral Biases of LLMs via Utility Theory. arXiv preprint arXiv:2408.02784.
- Korinek, A. (2024). LLMs Level Up—Better, Faster, Cheaper: June 2024 Update to Section 3 of "Generative AI for Economic Research: Use Cases and Implications for Economists," Published in the. Journal of Economic Literature, 61(4), 1-38.
- Binz, M., Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., ... & Schulz, E. (2024). Centaur: a foundation model of human cognition. arXiv preprint arXiv:2410.20268.
- Manning, B. S., Zhu, K., & Horton, J. J. (2024). Automated social science: Language models as scientist and subjects (No. w32381). National Bureau of Economic Research.
- Ludwig, J., Mullainathan, S., & Rambachan, A. (2024). Large Language Models: An Applied Econometric Framework. arXiv preprint arXiv:2412.07031.

Section 3.8 Algorithms and Human Behaviors

- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), 237-293.
- Rambachan, A. (2024). Identifying prediction mistakes in observational data. The Quarterly Journal of Economics, qjae013.
- Angelova, V., Dobbie, W. S., & Yang, C. (2023). Algorithmic recommendations and human discretion (No. w31747). National Bureau of Economic Research.
- Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). Combining human expertise with artificial intelligence: Experimental evidence from radiology (No. w31422). National Bureau of Economic Research.
- Ludwig, J., & Mullainathan, S. (2024). Machine learning as a tool for hypothesis generation. The Quarterly Journal of Economics, 139(2), 751-827.
- Ludwig, J., & Mullainathan, S. (2021). Fragile algorithms and fallible decision-makers: lessons from the justice system. Journal of Economic Perspectives, 35(4), 71-96.

- Mullainathan, S., & Obermeyer, Z. (2022). Diagnosing physician error: A machine learning approach to low-value health care. The Quarterly Journal of Economics, 137(2), 679-727.
- Ludwig, J., Mullainathan, S., & Rambachan, A. (2024, May). The Unreasonable Effectiveness of Algorithms. In AEA Papers and Proceedings (Vol. 114, pp. 623-627). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.
- Alur, R., Laine, L., Li, D., Raghavan, M., Shah, D., & Shung, D. (2024). Auditing for human expertise. Advances in Neural Information Processing Systems, 36.
- Mullainathan, S., & Rambachan, A. (2024). From predictive algorithms to automatic generation of anomalies (No. w32422). National Bureau of Economic Research.
- Caplin, A., Martin, D. J., & Marx, P. (2023). Rationalizable Learning (No. w30873). National Bureau of Economic Research.
- Stoddard, G., Fitzpatrick, D. J., & Ludwig, J. (2024). Predicting Police Misconduct (No. w32432). National Bureau of Economic Research.
- Ben-Michael, E., Greiner, D. J., Huang, M., Imai, K., Jiang, Z., & Shin, S. (2024). Does AI help humans make better decisions? A methodological framework for experimental evaluation. arXiv preprint arXiv:2403.12108.

Part 4. Coding

- **Python for Econometrics** (by Fabian H. C. Raters)
  - https://pyecon.org/down/pyecon.pdf
- **Applied Machine Learning in Python** (by Michael Pyrcz)
  - https://geostatsguy.github.io/MachineLearningDemos_Book/intro.html
- **Machine Learning-Based Causal Inference** (by Susan Athey)
  - https://d2cml-ai.github.io/mgtecon634_py/md/intro.html
- **Coding for Economists** (by Arthur Turrell)
  - https://aeturrell.github.io/coding-for-economists/intro.html
- **Python for Data Science** (by Arthur Turrell et al.)
  - https://aeturrell.github.io/python4DS/welcome.html
- **Introduction to Python for Econometrics, Statistics and Data Analysis** (by Kevin Sheppard)
  - https://www.kevinsheppard.com/files/teaching/python/notes/python_introduction_2021.pdf
- **Stata-to-Python equivalents** (by Daniel M. Sullivan)
  - https://www.danielmsullivan.com/pages/tutorial_stata_to_python.html
- **Microeconometrics Using Stata** (by Colin Cameron & Pravin Trivedi)
  - https://cameron.econ.ucdavis.edu/musbook/MUS2_Draft_Contents_November_2020.pdf