



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Big Data in Economics

2021-2-F9101Q018-F9101Q018M

Learning objectives

At the end of the course, you will learn how to:

- Use big data and machine learning for causal inference.
- Understand the advantages and value added of using big data for applied research in social sciences.

At the end of the course, you will be able to:

- Be familiar with some of the most important approaches to program evaluation to address a variety of policy-relevant research questions.
- Effectively use big data to address important applied research questions.

Contents

This course introduces the emerging field that merges Economics and Data Science to answer policy relevant research questions. The main goal of the course is to discuss how to use big data to answer relevant research questions in several applications.

We will discuss three main topics:

- 1) Causal Inference and Big Data.
- 2) Machine Learning and Causal Inference.
- 3) Empirical Applications Using Big Data.

Detailed program

Topic 1: Causal Inference and Big Data.

- Ø Causality, internal and external validity.
- Ø Big data: new frontiers for economic analysis.

References:

- * Angrist and Pischke, *Mostly Harmless Econometrics*, Princeton and Oxford University Press, 2009, Chapter 1 pages 3-8.
- Athey, Susan, 2017. "Beyond prediction: Using big data for policy problems", *Science*, 355, 483-485.
- *Athey, Susan. 2017. "The Impact of Machine Learning on Economics", mimeo.
- Athey S. and G. Imbens. 2016. "The State of Applied Econometrics – Causality and Policy Evaluation", <https://arxiv.org/pdf/1607.00699.pdf>
- Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and Selection of Human Capital with Machine Learning." *American Economic Review*, 106(5): 124–27.
- Einav L., and J. Levin. 2013. "The Data Revolution and Economic Analysis", NBER Working Paper 19035
- *Einav L., and J. Levin. 2014. "Economics in the Age of Big Data", *Science*, Vol 346, Issue 6210: 1243089.
- *Kleinberg, John, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems" *American Economic Review*, 105(5): 491–95.
- *Kleinberg, John, Jens Ludwig, and Sendhil Mullainathan. 2016. "A Guide to Solving Social Problems with Machine Learning", *Harvard Business Review*.
- Kleinberg, John, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions", *Quarterly Journal of Economics*, <https://doi.org/10.1093/qje/qjx032>
- Mullainathan S., and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach", *Journal of Economic Perspectives*, Volume 31, Number 2, Pages 87–106.

- Pearl, J. 2018. "Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution", Technical Report R-475.
- *Shmueli, G. 2010. "To Explain or to Predict?", *Statistical Science*, Vol. 25, No. 3, 289-310.
- Varian, H. 2013. "Beyond Big Data", Paper Presented at the NABE Annual Meeting, September 10, 2013, San Francisco, CA.
- *Varian, H. 2014. "Big Data: New Tricks for Econometrics", *Journal of Economic Perspectives* 28, 3-28
- Wager, S. and Susan Athey. 2017. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests", *Journal of the American Statistical Association*.

Topic 2: Machine Learning and Causal Inference.

References:

- *Athey S. and G. Imbens. 2015. "Machine Learning Methods in Economics and Econometrics", *American Economic Review: Papers and Proceedings*, 105(5): 476-480.
- Athey S. and G. Imbens. 2016. "The Econometrics of Randomized Experiments", <https://arxiv.org/abs/1607.00698>
- Blake Thomas, Chris Nosko, and Steven Tadelis. 2015. "Consumer Heterogeneity and Paid Search Effectiveness: a Large-Scale Field Experiment", *Econometrica*, Vol. 83, No. 1, pp. 155-174.
- Brodersen Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy and Steven L. Scott. 2015. "Inferring Causal Impact Using Bayesian Structural Time-Series Models", *The Annals of Applied Statistics*, Vol. 9 No. 1, pages 247-274.
- Rubin, D. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66: 688-701.
- *Varian, H. (2016) "Causal Inference in Economics and Marketing", *PNAS*, Vol. 113, No. 27, pages 7310-7315.

Ø Randomized and natural experiments

References:

- *Angrist and Pischke, *Mostly Harmless Econometrics*, Princeton and Oxford University Press, 2009, Chapter 2 pages 11-24.
- Deaton and Cartwright. 2017. "Understanding and misunderstanding randomized controlled trials". *Social Science & Medicine*, <https://doi.org/10.1016/j.socscimed.2017.12.005>
- Duflo, E., Glennerster, R. and Kremer, M. (2008) "Using Randomization in Development Economics Research: A Toolkit" In T. Schultz and John Strauss, eds., *Handbook of Development Economics*. Vol. 4. Amsterdam and New York: North Holland.
- *Ludwig, J., S. Mullainathan and J. Spiess. 2019. "Augmenting Pre-Analysis Plans with Machine

Learning", *American Economic Review Papers and Proceedings*, Vol. 109.

- Manski, C. (1996) "Learning about Treatment Effects from Experiments with Random Assignment of Treatment", *Journal of Human Resources*, 31: 709-733.
- Meyer, B.D. (1995) "Natural and Quasi-Experiments in Economics", *Journal of Business and Economic Statistics*, 13(2): 151-161.
- * Stock and Watson, *Introduction to Econometrics*, 3rd edition, Chapter 13 pages 511-529 and 538-540.

Ø Differences-in-differences estimator

References:

- Bertrand, M., Duflo, E., and S. Mullainathan (2004) "How much should we trust differences-in-differences estimates?", *Quarterly Journal of Economics*, 119(1): 249-75.
- Meyer, B.D. (1995) "Natural and Quasi-Experiments in Economics", *Journal of Business and Economic Statistics*, 13 (2): 151-161.
- * Stock and Watson, *Introduction to Econometrics*, 3rd edition, Chapter 10 pages 389-422.
- * Stock and Watson, *Introduction to Econometrics*, 3rd edition, Chapter 13 pages 532-535.

Topic 3: Empirical Applications Using Big Data.

Ø **Students' presentations:** present your own work, one of the papers from a list of suggested papers that will be provided, or a paper of your choice that uses machine learning methods, possibly replicating the results of the paper you choose to present.

References: a list of papers will be provided.

Prerequisites

Principles of applied econometrics and statistical quantitative methods for data analysis.

Teaching methods

In the 2020-2021 academic year, the course will be taught online and will be characterized by a strong interaction between the students and the teacher. The course will use several teaching tools such as short recorded lectures, self-assessment activities, webconferences and streaming sessions, projects and weekly activities (individually or in a small group).

Assessment methods

The exam consists in two parts that contributes to the final mark in the following way:

- 60%: group project (2-3 students) that applies the models and methods to the data.
- 40%: oral exam.

Textbooks and Reading Materials

Textbooks: there is no given recommended textbook for this course. Below you can find a list of some textbooks that can be used as reference for the main Econometrics topics that we will discuss in the course. All listed textbooks are available as e-books with the exception of Wooldridge (2020), which is available at the University Library (both Central Site and Science Site).

Advanced:

- W. H. Greene. *Econometric Analysis*, 5th Edition, Prentice Hall International, 2003.

Simpler/less math:

- J. Wooldridge. *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning, 2020. (for IV and 2 stage least squares)

- Stock and Watson, *Introduction to Econometrics*, 3rd Edition. (Basic statistics and regression analysis; companion website with datasets and files to replicate empirical results: http://wps.aw.com/aw_stock_ie_3/178/45691/11696965.cw/index.html)

- Angrist and Pischke, *Mostly Harmless Econometrics*, Princeton and Oxford University Press, 2009. (Excellent for concept of causality, experiments, diff-in-diff, and RD)

Journal articles and book chapters: each of the three main topics will make reference to the articles listed in the detailed program above.

Semester

Second semester.

Teaching language

English.
