



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Big Data in Economics

2021-2-F9101Q018-F9101Q018M

Obiettivi formativi

Al termine del corso, gli studenti avranno imparato come:

- Usare big data e machine learning per la stima di un impatto causale.
- Capire vantaggi e valore aggiunto dell'utilizzo di big data per ricerca applicata nelle scienze sociali.

Al termine del corso, gli studenti sapranno:

- Utilizzare alcune tra le più importanti tecniche di program evaluation per rispondere a domande di ricerca che pongono problemi di policy-relevance.
- Utilizzare in modo efficace i big data per rispondere a importanti domande di ricerca applicata.

Contenuti sintetici

Questo corso introduce il field emergente che nasce dalla fusione di Economia e Data Science per rispondere a domande di policy-relevance. L'obiettivo principale del corso è discutere come utilizzare i big data per rispondere a importanti domande di ricerca in diverse applicazioni.

Discuteremo i seguenti tre argomenti principali:

- 1) Inferenza Causale e Big Data.

2) Machine Learning e Inferenza Causale.

3) Applicazioni Empiriche che utilizzano Big Data.

Programma esteso

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.

Prerequisiti

Principi di econometria applicata e metodi quantitativi di statistica applicata.

Metodi didattici

Nell'anno accademico 2020-2021 il corso sarà erogato online con una struttura che prevede una forte interazione studenti-docente. Il corso utilizzerà diversi strumenti didattici tra cui brevi lezioni registrate, attività di autovalutazione dei contenuti, webconferences e incontri in streaming, progetti e attività settimanali (individuali o di gruppo).

Modalità di verifica dell'apprendimento

L'esame consiste in due componenti che contribuiscono al voto finale secondo il seguente dettaglio:

- 60%: progetto di applicazione dei modelli e metodi ai dati da svolgere in gruppi di 2-3 studenti.
- 40%: esame orale.

Testi di riferimento

Libri di testo: per questo corso non c'è un libro di testo di riferimento. Segue un elenco di alcuni testi di riferimento per gli argomenti di econometria trattati nel corso. Tutti i libri di testo di seguito riportati sono disponibili in formato e-book tranne Wooldridge (2020) che è disponibile presso la Biblioteca di Ateneo sede Centrale e Sede di Scienze.

Avanzato:

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

Semplice/meno tecnico:

- 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)

Articoli e capitoli di libro: la discussione di ciascuno dei tre argomenti di cui discuteremo nel corso farà riferimento agli articoli elencati nel programma dettagliato.

Periodo di erogazione dell'insegnamento

Secondo semestre.

Lingua di insegnamento

Inglese.

