Data Science for Public Policy
From Econometrics to AI

Introduction

Dr. Sergio Galletta

ETH Zurich

22/02/2024

Welcome

- Good **public policies** are essential for a well-functioning economy and society

Welcome

- Good **public policies** are essential for a well-functioning economy and society
  - We want policies to be correctly designed

Welcome

- Good **public policies** are essential for a well-functioning economy and society
  - We want policies to be correctly designed
  - We want policies to be correctly evaluated

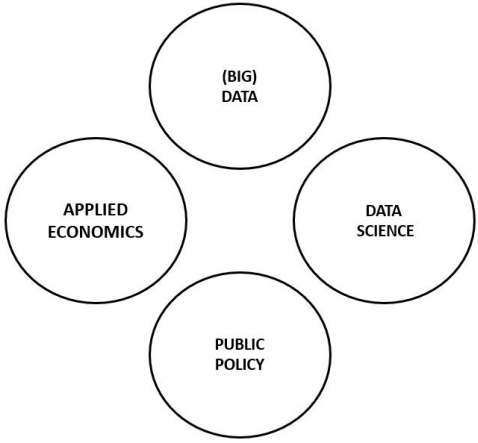
Welcome

- Good **public policies** are essential for a well-functioning economy and society
  - We want policies to be correctly designed
  - We want policies to be correctly evaluated
- This course provides an introduction to **data science** and **applied economics** methods for **public policy** applications

Welcome

- Good **public policies** are essential for a well-functioning economy and society
  - We want policies to be correctly designed
  - We want policies to be correctly evaluated
- This course provides an introduction to **data science** and **applied economics** methods for **public policy** applications
- **Causal inference methods** from applied economics and **machine learning techniques** are perfect to evaluate and support better policies

Welcome



Welcome

- (Big) Data are relevant to improve and assess public policies

Welcome

- (Big) Data are relevant to improve and assess public policies
  - What is Big Data?
  - Does Big mean better?

Welcome

- (Big) Data are relevant to improve and assess public policies
  - What is Big Data?
  - Does Big mean better?
- Applied (micro)Econometrics

Welcome

- (Big) Data are relevant to improve and assess public policies
  - What is Big Data?
  - Does Big mean better?
- Applied (micro)Econometrics
  - Focus on causality
  - Tool kit based on quasi-experimental setting

Welcome

- (Big) Data are relevant to improve and assess public policies
  - What is Big Data?
  - Does Big mean better?
- Applied (micro)Econometrics
  - Focus on causality
  - Tool kit based on quasi-experimental setting
- Data Science:

Welcome

- (Big) Data are relevant to improve and assess public policies
  - What is Big Data?
  - Does Big mean better?
- Applied (micro)Econometrics
  - Focus on causality
  - Tool kit based on quasi-experimental setting
- Data Science:
  - ML/AI optimal methods to make predictions
  - Perfect for generating risk assessment and tailoring actions

Some Examples

Pinotti 2017 AER - The Effect of Immigrant Legalization on Crime

- Does immigrant legalization reduce crime?

Some Examples

Pinotti 2017 AER - The Effect of Immigrant Legalization on Crime

- Does immigrant legalization reduce crime?
- 180K immigrants to Italy, some sent the application on time, others not

Some Examples

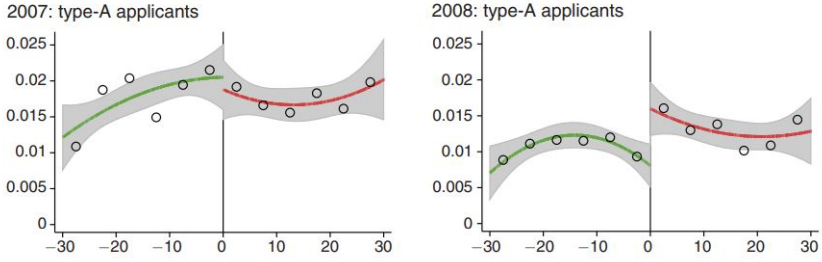
Pinotti 2017 AER - The Effect of Immigrant Legalization on Crime

- Does immigrant legalization reduce crime?
- 180K immigrants to Italy, some sent the application on time, others not
- Perfect setting for a causal estimation (randomness)

Some Examples

Pinotti 2017 AER - The Effect of Immigrant Legalization on Crime

- Does immigrant legalization reduce crime?
- 180K immigrants to Italy, some sent the application on time, others not
- Perfect setting for a causal estimation (randomness)



Some Examples

Altindağ et al 2022 RESTAT - Government-Sponsored Information Campaign on Infant Mortality

- Does government-sponsored campaigns affect health outcomes?

Some Examples

Altindağ et al 2022 RESTAT - Government-Sponsored Information Campaign on Infant Mortality

- Does government-sponsored campaigns affect health outcomes?
- Sudden infant death syndrome (SIDS) campaign in Denmark in 1991

Some Examples

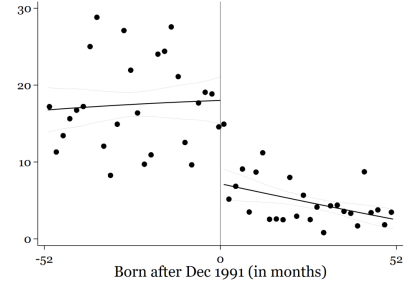
Altindağ et al 2022 RESTAT - Government-Sponsored Information Campaign on Infant Mortality

- Does government-sponsored campaigns affect health outcomes?
- Sudden infant death syndrome (SIDS) campaign in Denmark in 1991
- Some was exposed to the campaign and some were not

Some Examples

Altindağ et al 2022 RESTAT - Government-Sponsored Information Campaign on Infant Mortality

- Does government-sponsored campaigns affect health outcomes?
- Sudden infant death syndrome (SIDS) campaign in Denmark in 1991
- Some was exposed to the campaign and some were not



(A) SIDS mortality

Some Examples

Aiken et al 2022 NATURE - Machine learning to improve targeting of humanitarian aid

- Can machine learning improve humanitarian aid targeting?

Some Examples

Aiken et al 2022 NATURE - Machine learning to improve targeting of humanitarian aid

- Can machine learning improve humanitarian aid targeting?
- Cash transfer program in Togo for COVID-19 relief aid

Some Examples

Aiken et al 2022 NATURE - Machine learning to improve targeting of humanitarian aid

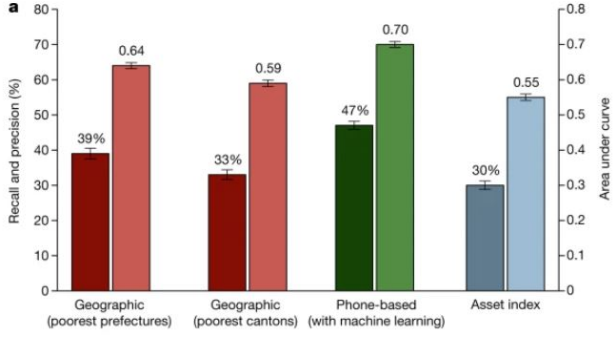
- Can machine learning improve humanitarian aid targeting?
- Cash transfer program in Togo for COVID-19 relief aid
- Targeting based on non-traditional "big" data (satellites and phone)

Some Examples

Aiken et al 2022 NATURE - Machine learning to improve targeting of humanitarian aid

- Can machine learning improve humanitarian aid targeting?
- Cash transfer program in Togo for COVID-19 relief aid
- Targeting based on non-traditional "big" data (satellites and phone)

Fig. 1: Comparing Novissi targeting to alternatives.



Some Examples

Ash et al 2024 AEJPOL - Machine learning to detect corruption

- Can machine learning support anti-corruption policy?

Some Examples

Ash et al 2024 AEJPOL - Machine learning to detect corruption

- Can machine learning support anti-corruption policy?
- Brazil anti-corruption policy vs ML detection

Some Examples

Ash et al 2024 AEJPOL - Machine learning to detect corruption

- Can machine learning support anti-corruption policy?
- Brazil anti-corruption policy vs ML detection
- Audit municipality based on ML risk score based on local budget

Some Examples

Ash et al 2024 AEJPOL - Machine learning to detect corruption

- Can machine learning support anti-corruption policy?
- Brazil anti-corruption policy vs ML detection
- Audit municipality based on ML risk score based on local budget

Table 4: Performance Metrics for Targeted Auditing Policies

	Status Quo (Lottery)		Targeted Audits		Fair Targeting
	(1a)	(1b)	(2a)	(2b)	(3)
Evaluation Sample	All (Sim)	Audited	All (Sim)	Audited	All (Sim)
Corruption Rate, if Audited	0.486	0.458	0.871	0.883	0.868
↪ Ratio over Random Audits			[1.788]	[1.927]	[1.783]
Audit Rate, if Corrupt	0.036	0.036	0.076	0.119	0.074
↪ Ratio over Random Audits			[2.714]	[4.246]	[2.644]

Some Examples

Gentzkow et al 2019 ECTM - Political polarization and Text as Data

- Is the US politics become more polarized?

Some Examples

Gentzkow et al 2019 ECTM - Political polarization and Text as Data

- Is the US politics become more polarized?
- Text analysis and machine learning predictions

Some Examples

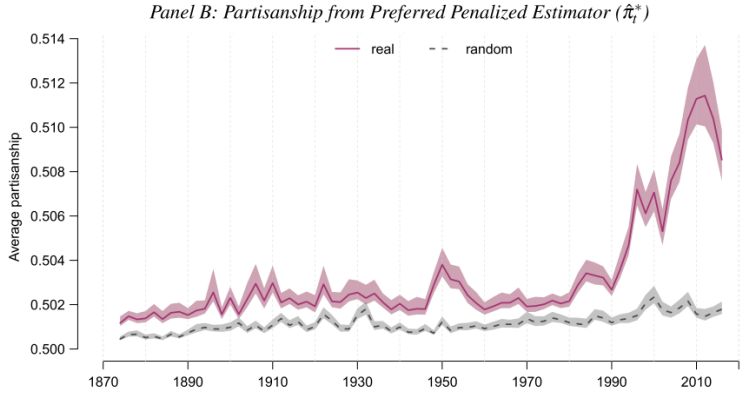
Gentzkow et al 2019 ECTM - Political polarization and Text as Data

- Is the US politics become more polarized?
- Text analysis and machine learning predictions
- Partisanship: ease of deducing congressperson's party from one statement

Some Examples

Gentzkow et al 2019 ECTM - Political polarization and Text as Data

- Is the US politics become more polarized?
- Text analysis and machine learning predictions
- Partisanship: ease of deducing congressperson's party from one statement



Some Examples

Gorodnichenko et al 2023 AER - The Voice of Monetary Policy

- Can the tone of voices of Central Bankers affect the market?

Some Examples

Gorodnichenko et al 2023 AER - The Voice of Monetary Policy

- Can the tone of voices of Central Bankers affect the market?
- Federal Open Market Committee nonverbal communication during

Some Examples

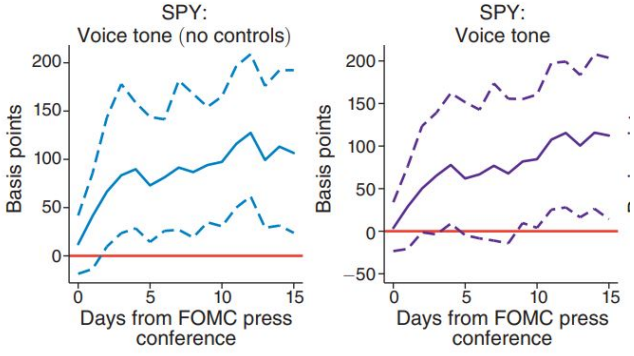
Gorodnichenko et al 2023 AER - The Voice of Monetary Policy

- Can the tone of voices of Central Bankers affect the market?
- Federal Open Market Committee nonverbal communication during
- Emotion Detection Using Neural Networks

Some Examples

Gorodnichenko et al 2023 AER - The Voice of Monetary Policy

- Can the tone of voices of Central Bankers affect the market?
- Federal Open Market Committee nonverbal communication during
- Emotion Detection Using Neural Networks



Practical Information

- Instructors: Sergio Galletta (sergio.galletta@gess.ethz.ch) & Elliott Ash (ashe@ethz.ch) & Christoph Gössmann (christoph.goessmann@gess.ethz.ch)
- Lecture Time: Thursdays 12:15-14:00
- Location: In person LFW B 1
- Office hours: By appointment via email
- Most of the important information are in the syllabus and we update the course material on github

Last year projects

- Does the text sentiment of the voting question influence the outcome of Swiss voting? (Paper)
- The causal impact of rainfall on wealth inequality in Vietnam (Paper)
- Who Pays for the Church? Electoral Institutions and Religious Clientelism in Post-War Italy (Paper)
- Modelling Patient Risk and Extraneous Causal Factors in Physician-Decision Making (Paper)
- Renewable Energy Innovation Across the World (APP)