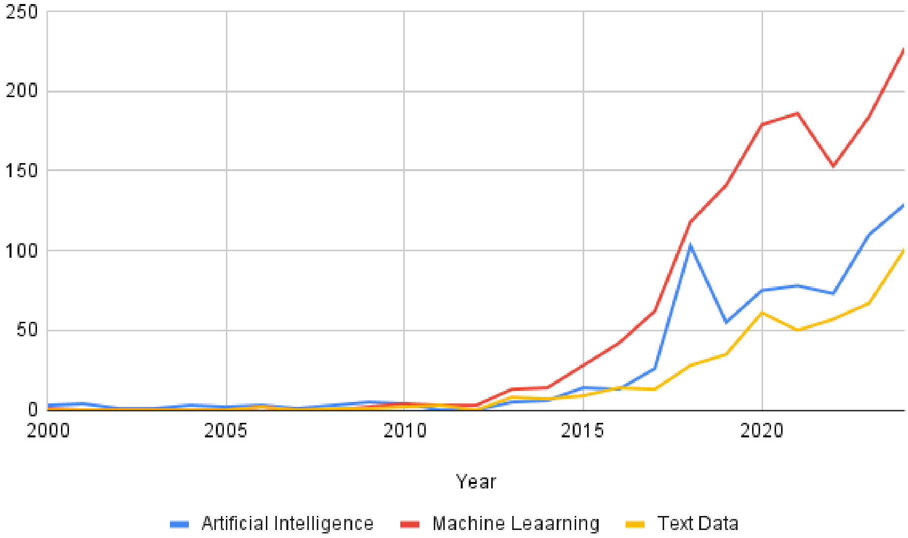
Data Science for Public Policy
From Econometrics to AI

Introduction

ETH Zurich

20/02/2024

Data Science in Economics



Why Data Science for Public Policy?

Why Data Science for Public Policy?

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

Why Data Science for Public Policy?

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

Why Data Science for Public Policy?

- 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

Why Data Science for Public Policy?

- 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 is where tools from **data science** and *econometrics* come in handy
- **Causal inference methods** from applied economics and **machine learning techniques** are perfect to evaluate and support better policies

Welcome to *Data Science for Public Policy*

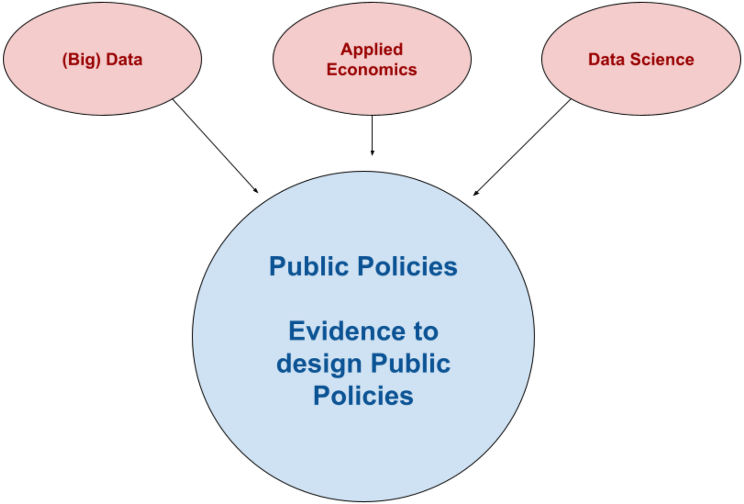
This course provides an introduction to **data science** and **applied economics** methods for **public policy** applications

Welcome to *Data Science for Public Policy*

This course provides an introduction to **data science** and **applied economics** methods for **public policy** applications

Welcome to *Data Science for Public Policy*

This course provides an introduction to **data science** and **applied economics** methods for **public policy** applications



Some Concepts We Will Explore

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

Some Concepts We Will Explore

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

Some Concepts We Will Explore

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

Some Concepts We Will Explore

- (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

Some Concepts We Will Explore

- (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:

Some Concepts We Will Explore

- (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
  - Allow the analysis of different data formats, e.g., text, images, audio
  - 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 did not

Some Examples

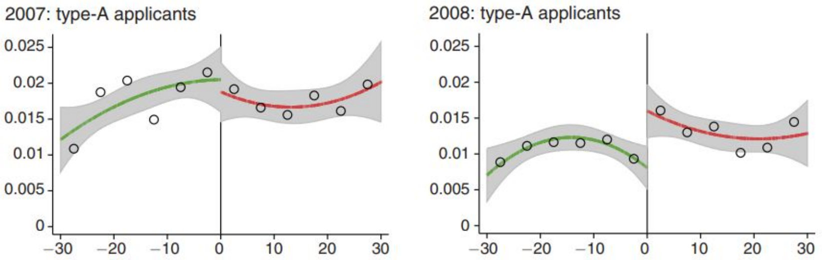
Pinotti 2017 AER - The Effect of Immigrant Legalization on Crime

- Does immigrant legalization reduce crime?
- 180K immigrants to Italy, some sent the application on time, others did 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 did 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

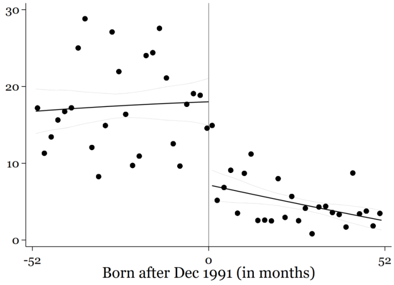
Altindağ et al. 2022 RESTAT - Government-Sponsored Information Campaign on Infant Mortality

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

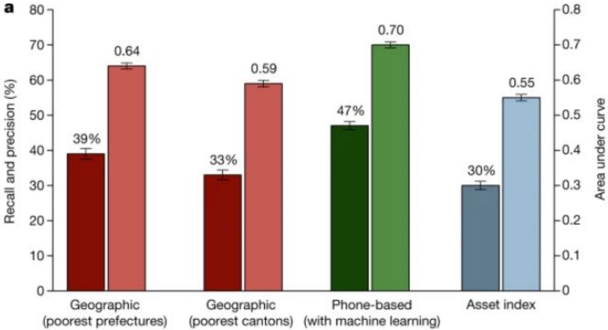
- Can machine learning improve humanitarian aid targeting?
- Cash transfer program in Togo for COVID-19 relief aid
- Targeting based on non-traditional "big" data (satellites and phone)

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
$\hookrightarrow$ Ratio over Random Audits			[1.788]	[1.927]	[1.783]
Audit Rate, if Corrupt	0.036	0.036	0.076	0.119	0.074
$\hookrightarrow$ Ratio over Random Audits			[2.714]	[4.246]	[2.644]

Some Examples

Hansen et al. 2023 QJE - Transparency and Deliberation Within the FOMC

- How does transparency affect monetary policy makers' deliberations?

Some Examples

Hansen et al. 2023 QJE - Transparency and Deliberation Within the FOMC

- How does transparency affect monetary policy makers' deliberations?
- Combine text analysis and a change in policy about publication of transcripts

Some Examples

Hansen et al. 2023 QJE - Transparency and Deliberation Within the FOMC

- How does transparency affect monetary policy makers' deliberations?
- Combine text analysis and a change in policy about publication of transcripts
- Isolate topics that capture policy preferences

Some Examples

Hansen et al. 2023 QJE - Transparency and Deliberation Within the FOMC

- How does transparency affect monetary policy makers' deliberations?
- Combine text analysis and a change in policy about publication of transcripts
- Isolate topics that capture policy preferences

Difference-in-Differences Results for Policy Strategy Discussion (FOMC2): Topic Measures

Main regressors	Concentration (1)	Quant (2)	Avg Sim (B) (3)	Avg Sim (D) (4)	Avg Sim (KL) (5)	Pr (No Dissent) (6)
D(Trans) × Fed experience	−0.00077**	−0.00011	−0.00019	−0.00041***	−0.00040	−0.0015**
	[.014]	[.323]	[.222]	[.006]	[.377]	[.025]
Fed experience	−0.21***	−0.0035	−0.057	−0.11***	−0.22**	−0.41**
	[.000]	[.911]	[.140]	[.006]	[.045]	[.031]
Rookie effect	8.9**	5.6	0.4	5.5***	1.1	3.5**

Notes. This table reports the results of estimating (DinD) on FOMC member statements from the monetary policy strategy discussion. Dependent variable definitions are in Table IV. Coefficients are labeled according to significance (***$p < .01$, **$p <. 05$, *$p < .1$) while brackets below coefficients report p-values calculated using Driscoll-Kraay standard errors. The rookie effect reports the estimated coefficient on $D(Trans)_t \times FedExp_{it}$ multiplied by 20 (approximate difference in experience between the two modes in Figure VI) as a percentage of the average value of the dependent variable before November 1993. These effects carry the same star labels as the corresponding estimated coefficient on $D(Trans)_t \times FedExp_{it}$.

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

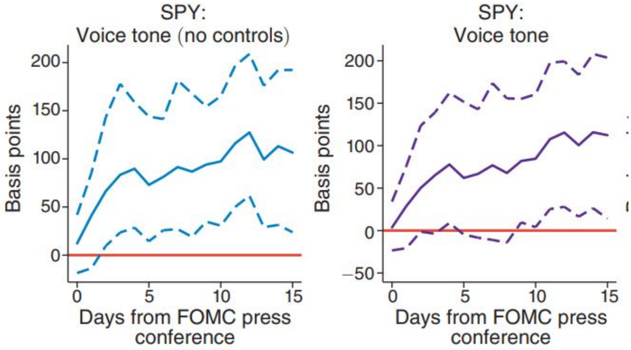
Gorodnichenko et al. 2023 AER - The Voice of Monetary Policy

- 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



Logistics

- Instructors: Sergio Galletta (sergio.galletta@gess.ethz.ch) & Elliott Ash (ashe@ethz.ch) & Claudia Marangon (claudia.marangon@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

Assessment

- Paper presentations (April 17th to May 15th) – **20% of the final grade**
- Final exam (May 22nd) – **30% of the final grade**
- Individual or group project (due on July 22nd) – **50% of the final grade**

Assessment: Paper Presentations

- Group presentations – 2-4 members
- From April 17th to May 15th – 4 presentations per day
  - 15 minutes presentations (only clarifying questions)
  - 5-10 minutes questions and discussion
- Choose a paper from our reading list *OR* propose one in coordination with us
- Sign-up here **by March 13th**

Assessment: Final Project

- Research paper: pick a topic you are interested in!
- Individual or in groups of 2 to 4
- Timeline (more details on syllabus):
  - Choose a topic and sign-up here by **April 9th**
  - Submit 1/2 pages outline by **April 17th**
  - Submit 10 minutes video presentations by **May 30th**
  - Submit final paper with replication package by **July 20th**
- You can reach out to us at any time to discuss your project

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)