

ECON 251 Office Hours

Review + Extensions

Elird Haxhiu

Fall 2022

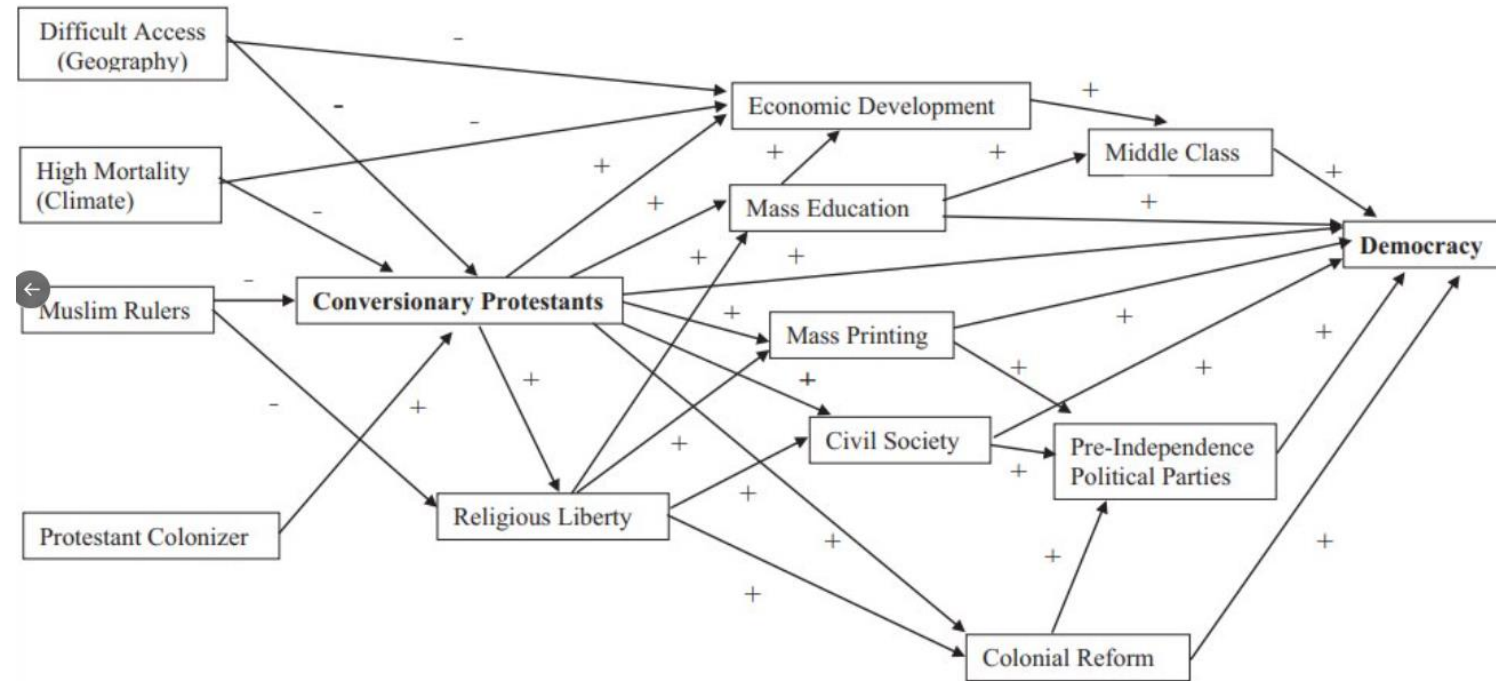
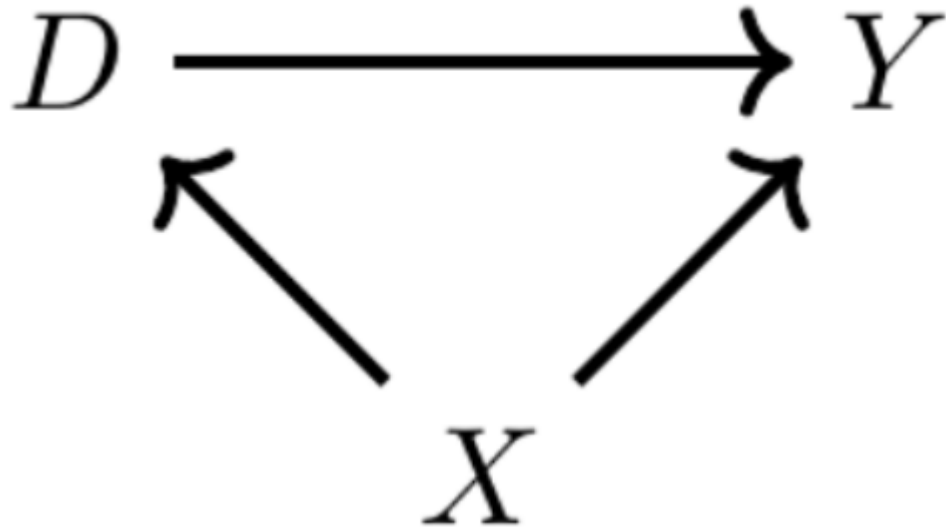


Angira Shukla
@ShuklaAngira

The class:

The final:

Just kidding of course! :)



Outline

1. Review: HW4 solutions
2. Review: Final exam questions
3. Review: OLS, IV, DID, LPM
4. Recommendations for further reading
 - Directed Acyclic Graphs (DAGs) for intuition and identification
 - Bayesian methods in econometrics
 - James-Stein Paradox and shrinkage estimators

Review HW4 solutions

Review for Final Exam

- HW3 solutions
 - Midterm solutions
 - HW2 solutions
 - HW1 solutions
-
- Discussion notes solutions
 - Lecture notes solutions

Review

1. OLS assumptions and theorems
2. IV assumptions and results
3. DID assumptions and results
4. LPM intuition and results

$$\text{OLS} \quad \min_{\{\beta_0, \beta_1, \dots, \beta_k\}} \frac{1}{N} \sum_{i=1}^N (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_k X_{ik})^2 \quad \Rightarrow \quad \hat{\beta}_j^{OLS}$$

- MLR1 (linear outcome model)
- MLR2 (random sampling)
- MLR3 (no collinearity)
- MLR4 (independence)
- MLR5 (homoskedasticity)
- MLR6 (normality)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + U_i$$

$\{Y_i, X_{i1}, \dots, X_{ik}\}_{i=1}^N$ is random draw

no X_{ij} linear function of any other X_{il}

$$E[U_i | X_{i1}, \dots, X_{ik}] = 0$$

$$\text{Var}(U_i | X_{i1}, \dots, X_{ik}) = \sigma^2$$

$$U_i \sim N(0, \sigma^2)$$

- T1 (unbiased) MLR1+2+3+4 $\Rightarrow \quad E[\hat{\beta}_j^{OLS}] = \beta_j \quad \forall j = \{0, 1, \dots, k\}$
- T2 (efficient-GM) MLR1+2+3+4+5 $\Rightarrow \quad E[\hat{\beta}_j^{OLS}] = \beta_j \quad \forall j$

$$\text{Var}[\hat{\beta}_j^{OLS}] \leq \text{Var}[\hat{\beta}_j^{\text{other linear}}]$$
- T3 (efficient-CL) MLR1+2+3+4+5+6 $\Rightarrow \quad \hat{\beta}_j^{OLS} \sim N(\beta_j, \text{Var}[\beta_j]) \quad \forall j$

Linear Probability Model (LPM)

Binary Outcome

$$Y = \beta_0 + \beta_1 X + U$$
$$Y \in \{0,1\}$$

- Independence + binary Y gives “change in prob($Y=1$)” interpretation (pp!)

$$E[Y|X] = P[Y = 1|X] \cdot 1 + P[Y = 0|X] \cdot 0 = P[Y = 1|X]$$
$$\Rightarrow \beta_1 = \frac{\partial}{\partial X} P[Y = 1|X]$$

- LPM is nice because...

1. Easy to estimate
2. Easy to interpret

- LPM is problematic because...

1. Predicted values of outcome can be outside of $[0,1]$ interval
2. Does not make sense for X to change $P[Y = 1|X]$ linearly
3. Homoskedasticity is always violated: $\text{Var}(Y|X) = P(Y = 1|X)[1 - P(Y = 1|X)]$

Omitted variable bias (OVB)

“True” model $\log Y_i = \alpha + \beta \cdot S_i + \delta^{X \rightarrow Y} \cdot X_i + U_i$ $\text{Cov}(S_i, U_i) = 0$

Our model $\log Y_i = a + b \cdot S_i + E_i$

Auxiliary model $X_i = c + \gamma^{S \rightarrow X} \cdot S_i + \eta_i$

Naively assuming $\text{Cov}(S_i, E_i) = 0$ in our model implies

$$b = \frac{\text{Cov}(S_i, \log Y_i)}{\text{Var}(S_i)}$$

$$= \beta + \gamma^{S \rightarrow X} \cdot \delta^{X \rightarrow Y}$$

$$= \text{causal effect} + (\text{var in } S \text{ related to } X) \cdot (\text{var in } X \text{ related to } Y)_9$$

Mincer (1974) model of earnings

$$\log Y_i = \alpha + \beta \cdot S_i + U_i$$

Instrumental variable (IV) Z_i decomposes

S_i into S_i^X and S_i^N

- First stage generates predicted values for treatment
- We estimate returns β from model

$$\hat{S}_i := \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

$$\log Y_i = \alpha + \beta \cdot \hat{S}_i + U_i$$

- A valid instrument satisfies

1. Relevance
2. Exogeneity
3. Exclusion

$$\text{Cov}(Z_i, S_i) \neq 0$$

$$\text{Cov}(Z_i, U_i) = 0$$

no direct effect of Z_i on Y_i

Doing IV can be worse than OLS

- The OVB formula for OLS implies that it converges to

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{OLS} = \beta + \frac{\text{Cov}(S_i, U_i)}{\text{Var}(S_i)}$$

- We have just shown that the MM estimator converges to

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{MM} = \beta + \frac{\text{Cov}(Z_i, U_i)}{\text{Cov}(Z_i, S_i)}$$

- What if $\text{Cov}(S_i, U_i) = 0$ or $\text{Cov}(Z_i, U_i) = 0$ are not exactly $= 0$?
- Not clear which is more likely to hold without more context, but...
- $\text{Cov}(Z_i, S_i) \approx 0$ (weak IV) \Rightarrow minor violations of IV exogeneity lead to large bias!

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

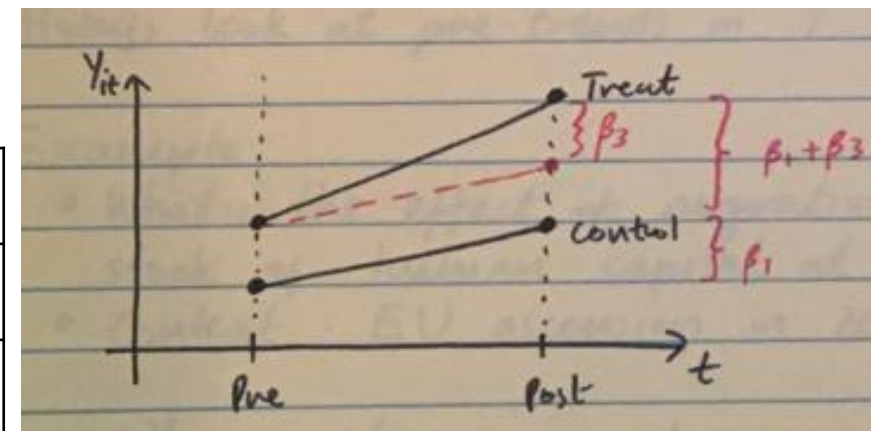
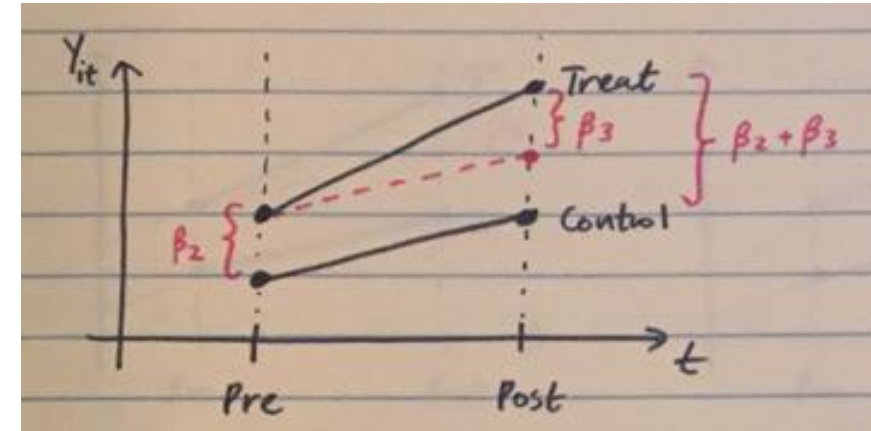
2 periods (before/after) and 2 groups (treated/control)

Y_{it} := outcome of interest

$P_t := 1\{t \text{ is after treatment occurs}\}$

$T_i := 1\{i \text{ is treated/exposed}\}$

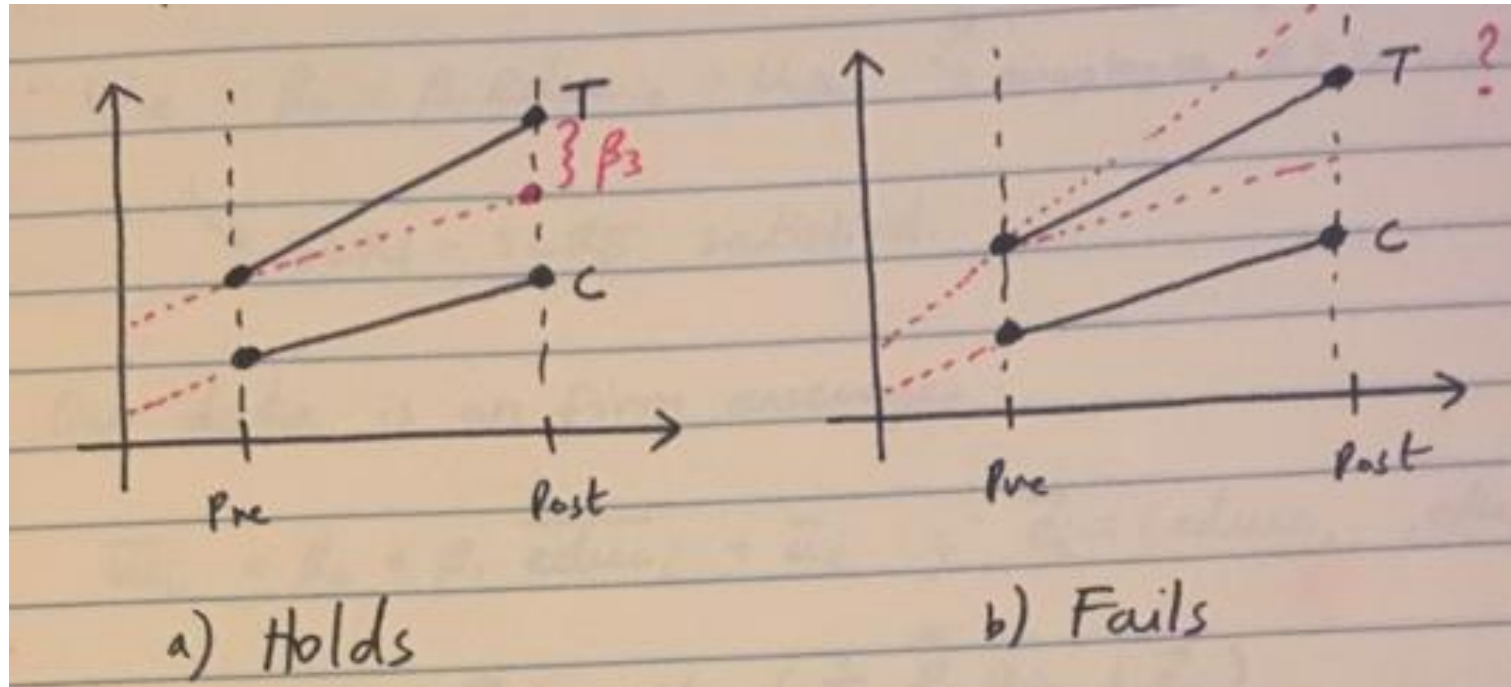
$$Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 [P_t \cdot T_i] + U_i$$



	Before	After	After – Before
Control	β_0	$\beta_0 + \beta_1$	β_1
Treated	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Treat – Control	β_2	$\beta_2 + \beta_3$	β_3

Parallel Trends Assumption = exposed units Y without policy T would have changed like unexposed units Y

- PTA is an untestable assumption, just like OLS exogeneity or IV exogeneity
- However, if we have access to more data before policy, we can assess how likely it is to hold in practice... commonly known as “checking for pre-trends”
- One reason why people seem to like DD... visual check of identifying assumption!




Outline

1. Review: HW4 solutions
2. Review: Final exam questions
3. Review: OLS, IV, DID, LPM
4. Recommendations for further reading
 - Directed Acyclic Graphs (DAGs) for intuition and identification
 - Bayesian methods in econometrics
 - James-Stein Paradox and shrinkage estimators

Causal Inference
The Mixtape

- Welcome
- 1 Introduction
 - 2 Probability and Regression Review
 - 3 **Directed Acyclic Graphs**
 - 4 Potential Outcomes Causal Model
 - 5 Matching and Subclassification
 - 6 Regression Discontinuity
 - 7 Instrumental Variables
 - 8 Panel Data
 - 9 Difference-in-Differences
 - 10 Synthetic Control
 - 11 Conclusion
- Mixtape Sessions
- Teaching Resources
- Acknowledgments

3 Directed Acyclic Graphs



Causal Inference:

The Mixtape.

Buy the print version today:

[Buy from Amazon](#) [Buy from Yale Press](#)

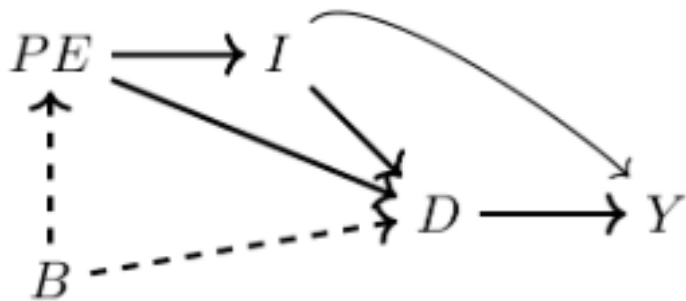
The history of graphical causal modeling goes back to the early twentieth century and Sewall Wright, one of the fathers of modern genetics and son of the economist Philip Wright. Sewall developed path diagrams for genetics, and Philip, it is believed, adapted them for econometric identification ([Matsueda 2012](#)).¹

But despite that promising start, the use of graphical modeling for causal inference has been largely ignored by the economics profession, with a few exceptions ([J. Heckman and Pinto 2015](#); [Imbens 2019](#)). It was revitalized for the

Table of contents

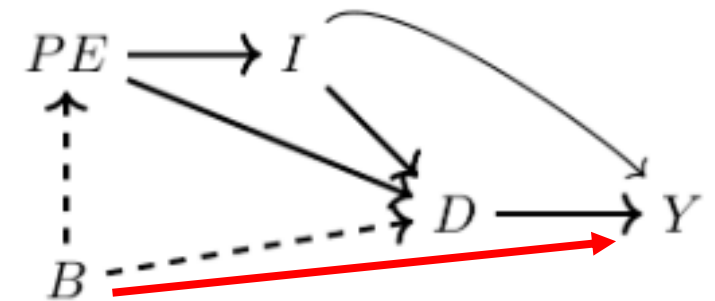
- Introduction to DAG Notation**
- A simple DAG
- Colliding
- Backdoor criterion
- More examples of collider bias
- Discrimination and collider bias
- Sample selection and collider bias
- Collider bias and police use of force
- Conclusion

Example: Mincer (1974) model



Now that we have a DAG, what do we do? I like to list out all direct and indirect paths (i.e., backdoor paths) between D and Y . Once I have all those, I have a better sense of where my problems are. So:

1. $D \rightarrow Y$ (the causal effect of education on earnings)
2. $D \leftarrow I \rightarrow Y$ (backdoor path 1)
3. $D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)
4. $D \leftarrow B \rightarrow PE \rightarrow I \rightarrow Y$ (backdoor path 3)



Bayesian (vs frequentist) statistics



Psychon Bull Rev (2018) 25:155–177
DOI 10.3758/s13423-017-1272-1

BRIEF REPORT

Bayesian data analysis for newcomers

John K. Kruschke¹ · Torrin M. Liddell¹

Published online: 12 April 2017
© Psychonomic Society, Inc. 2017

Abstract This article explains the foundational concepts of Bayesian data analysis using virtually no mathematical notation. Bayesian ideas already match your intuitions from everyday reasoning and from traditional data analysis. Simple examples of Bayesian data analysis are presented that illustrate how the information delivered by a Bayesian

This article explains the... The article uses virtual... emphasizes are on establishing... disabusing misconception... the many reasons to be... intervals (see, for exam...

Psychon Bull Rev (2018) 25:178–206
DOI 10.3758/s13423-016-1221-4

BRIEF REPORT

The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective

John K. Kruschke¹ · Torrin M. Liddell¹

Published online: 7 February 2017
© Psychonomic Society, Inc. 2017

Abstract In the practice of data analysis, there is a conceptual distinction between hypothesis testing, on the one hand, and estimation with quantified uncertainty on the other. Among frequentists in psychology, a shift of emphasis from hypothesis testing to estimation has been dubbed “the New Statistics” (Cumming, 2014). A second conceptual

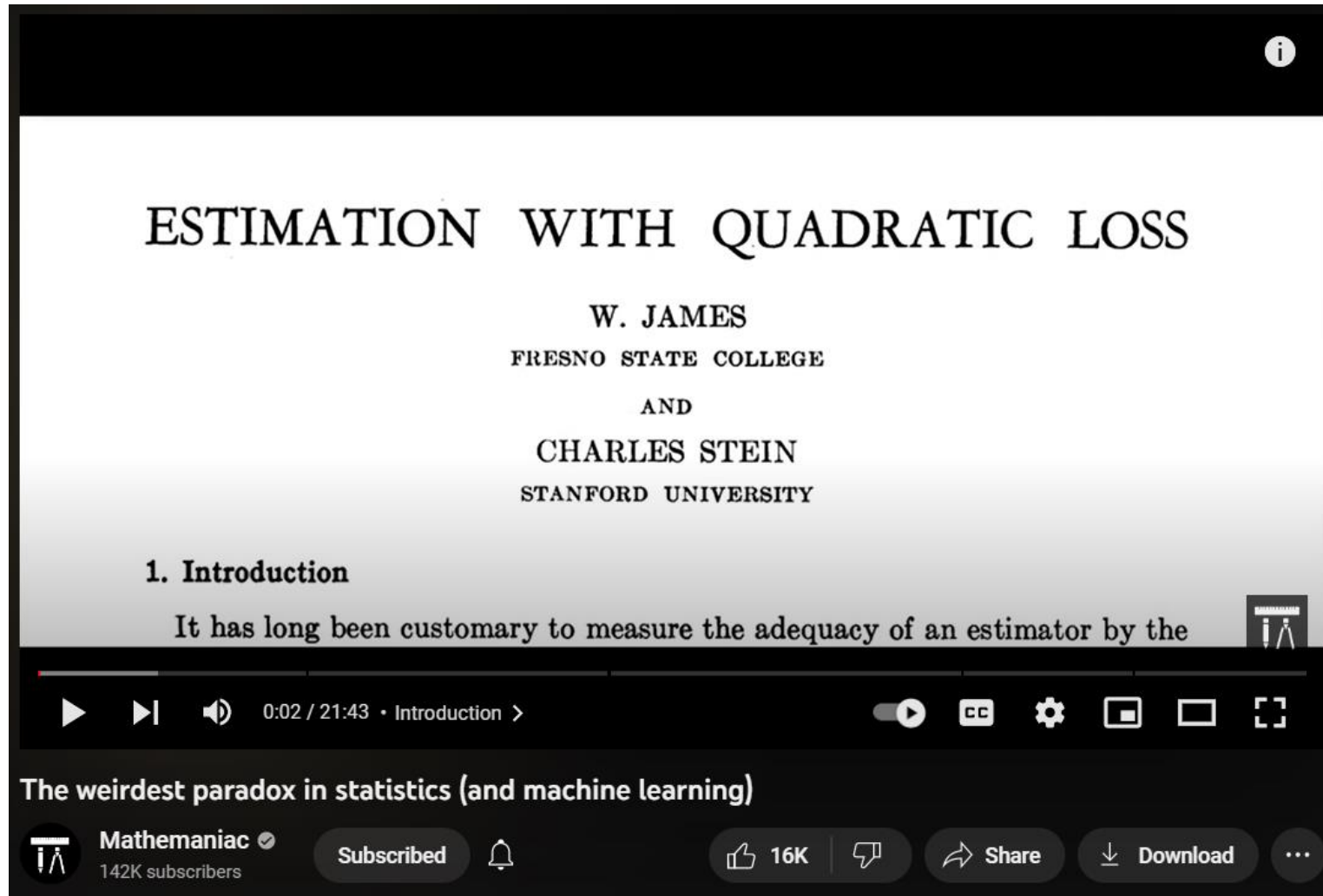
to eschew NHST, with its seductive lapse to black-and-white thinking about the presence or absence of effects. There are also many reasons to promote instead a cumulative science that incrementally improves estimates of magnitudes and uncertainty. These reasons were recently highlighted in a prominent statement from the American Statistical Asso-

James-Stein (1961) estimator (and paradox)

Warm up = prove that sample mean is OLS when model has no covariates

James-Stein (1961) estimator (and paradox)

<https://www.youtube.com/watch?v=cUqoHQQDinCM>



The image shows a YouTube video player interface. The video title is "ESTIMATION WITH QUADRATIC LOSS". The authors listed are "W. JAMES" from "FRESNO STATE COLLEGE" and "CHARLES STEIN" from "STANFORD UNIVERSITY". The video is at the 0:02 mark of a 21:43 duration, titled "Introduction". The video content shows the text "It has long been customary to measure the adequacy of an estimator by the". The video player controls include play, next, volume, and a progress bar. Below the video player, the channel name "Mathemaniac" is displayed with a verified badge and 142K subscribers. The video has 16K likes and a "Share" button. The video title "The weirdest paradox in statistics (and machine learning)" is also visible.

ESTIMATION WITH QUADRATIC LOSS

W. JAMES
FRESNO STATE COLLEGE

AND

CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

It has long been customary to measure the adequacy of an estimator by the

0:02 / 21:43 • Introduction >

The weirdest paradox in statistics (and machine learning)

Mathemaniac ✓
142K subscribers

Subscribed

16K

Share

Download

Thank you for a fun semester! I learned a lot :)