Econometrics with Misaligned Preferences

Jann Spiess

Stanford GSB

May 2023

Challenge and main idea

- Empirical estimates reflect not just data, but also researcher decisions and incentives
- How can we approach statistical decisions when there are conflicts of interest?
- **Approach in my lecture today:** embed econometric tasks in principal-agent framework, implications for pre-analysis plans
- **Broader agenda:** How can we make causal inference and data-driven decisions more efficient and robust?
  - Today: principal–agent model for econometric analysis, PAPs
  - Thursday: principal–agent model for explaining, regulating AI

Lecture outline

1. "Optimal Estimation when Researcher and Social Preferences are Misaligned" (2018; revised 2022)

2. High-level model and integrating machine learning/AI

3. Pre-analysis plans and implementability (with Max Kasy)

4. Summary and conclusion

Lecture outline

1. "Optimal Estimation when Researcher and Social Preferences are Misaligned" (2018; revised 2022)

2. High-level model and integrating machine learning/AI

3. Pre-analysis plans and implementability (with Max Kasy)

4. Summary and conclusion

Challenge

- Empirical estimates reflect not just data,
  but also researcher decisions and incentives
  - *p*-value (Brodeur et al., 2016)
  - Sign (Andrews and Kasy, 2017)
  - Magnitude (Jelveh et al., 2015)
- How can we ensure precise estimation when researchers pursue own goals and engage in specification searches?
- I propose econometric approach rooted in mechanism design that recognizes researchers degrees of freedom and preferences
  1. Constraints we should put on empirical analysis
  2. Estimators that have socially desirable properties
  3. Optimal pre-analysis plans

Treatment effect estimation in experiments ▸ JTPA

- Researcher estimates average treatment effect in experiment

$$y_i = \hat{\alpha} + \underbrace{d_i}_{\text{random treatment}} \hat{\tau} + \overbrace{x_i'}^{\text{additional covariates}} \hat{\gamma} + \hat{\varepsilon}_i$$

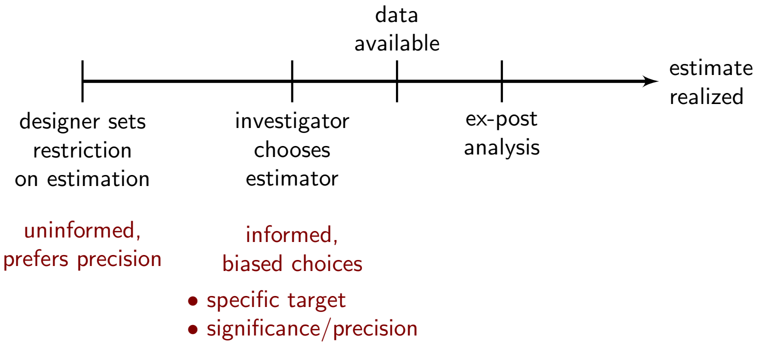
- Simple estimator: treatment–control average difference
- Giving researcher freedom to use control variables
  1. Can improve precision
  2. Can induce bias from specification searches
- One solution: forbid specification searches altogether

→ How to leverage data and researchers expertise,
but not also reflect researchers preferences?

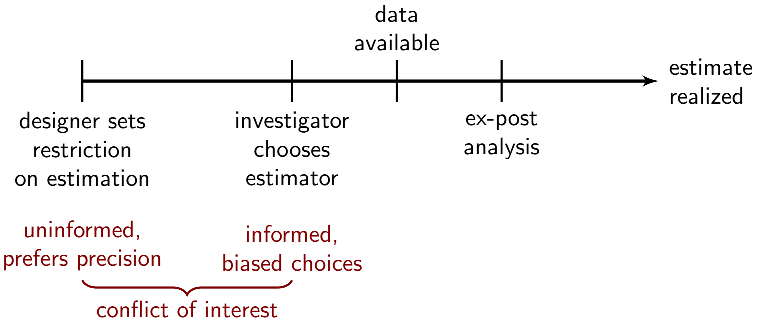
Optimal estimation with specification searches

- Standard econometric approach: statistical problem
  1. Propose an estimator from identification result
  2. Statistical properties, often using large-sample approximations
- My econometric approach: mechanism-design problem
  1. Estimation setup, researcher choice and preferences
  2. Solve for optimal restrictions and estimators in finite samples
- Specific application
  - Precise average treatment effect on experiments
  - Point estimation with explicit preferences beyond p-values
  - Researcher choices, not publication process

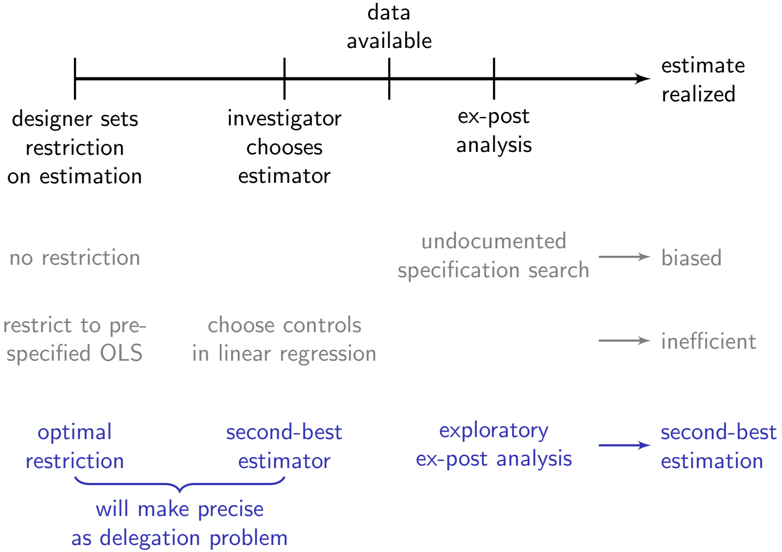
Timeline


data available
designer sets restriction on estimation
investigator chooses estimator
ex-post analysis
estimate realized
uninformed, prefers precision
informed, biased choices
• specific target
• significance/precision

Timeline


data
available
estimate
realized
designer sets
restriction
on estimation
investigator
chooses
estimator
ex-post
analysis
uninformed,
prefers precision
informed,
biased choices
conflict of interest

Timeline


data
available
estimate
realized
designer sets
restriction
on estimation
investigator
chooses
estimator
ex-post
analysis
no restriction
undocumented
specification search
biased
restrict to pre-
specified OLS
choose controls
in linear regression
inefficient
optimal
restriction
second-best
estimator
exploratory
ex-post analysis
second-best
estimation
will make precise
as delegation problem

Results

1. Designer's solution: bias restriction
2. Investigator's solution: flexible unbiased estimators
   - Sample-splitting ensures unbiasedness
   - Prediction yields efficiency
3. Implementation: optimal pre-analysis plans
   - Specification searches without bias
   - Data distribution instead of pre-specification

Context

- Specification searches, researcher incentives, pre-analysis plans
  Leamer (1974); Glaeser (2006); Olken (2015); Coffman and Niederle (2015); Young (2017); Andrews and Kasy (2017)
- Delegation as mechanism-design problem Holmström (1978, 1984); Alonso and Matouschek (2008); Frankel (2014)
- Decision-theoretic approaches to experimental design Kasy (2016); Banerjee et al. (2016, 2017)
- Covariate adjustments and bias Freedman (2008); Lin (2013); Bloniarz et al. (2016); Wager et al. (2016); Wu and Gagnon-Bartsch (2017)
- Machine learning in causal inference Farrell (2015); Athey and Imbens (2016); Chernozhukov et al. (2017a)
- Sample-splitting as orthogonalization Hájek (1962); Angrist et al. (1999); Hansen and Racine (2012); Schorfheide and Wolpin (2012, 2016); Chernozhukov et al. (2017b); Wager and Athey (2017)
- Hold-out in multiple testing Dahl et al. (2008); Dwork et al. (2015); Fafchamps and Labonne (2016); Anderson and Magruder (2017)

Econometric exercise

- Target: sample-average treatment effect (Neyman, 1923)

$$\underset{\substack{|\\\text{potential}\\\text{outcomes}}}{\tau_\theta} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{(y_i(1) - \overbrace{y_i(0)}^{\text{potential outcome}})}_{\text{causal effect on } i}$$

- Data:

$$z = (y_i, \overset{\substack{\text{treatment}\\|}}{d_i}, \underbrace{x_i}_{\text{take sample as given}})_{i=1}^{n} \in \mathcal{Z} \qquad \overbrace{y_i}^{\text{finite support}} = y_i(\underbrace{d_i}_{\text{random (prob } p)})$$

- Goal: estimator $\hat{\tau} : \mathcal{Z} \to \mathbb{R}$

Example (Average-difference estimator)

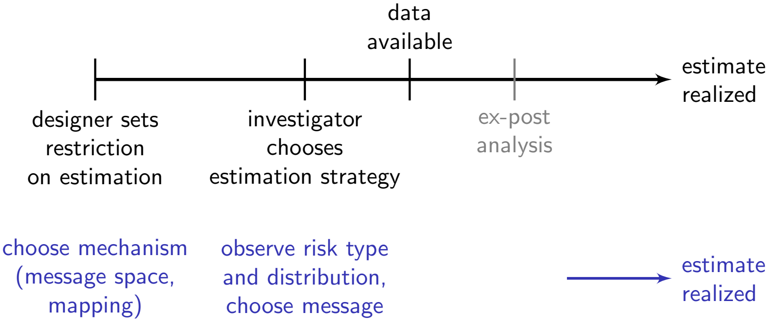
$$\hat{\tau}(z) = \frac{1}{n_1} \sum_{d_i=1} y_i - \frac{1}{n_0} \sum_{d_i=0} y_i$$

Preferences

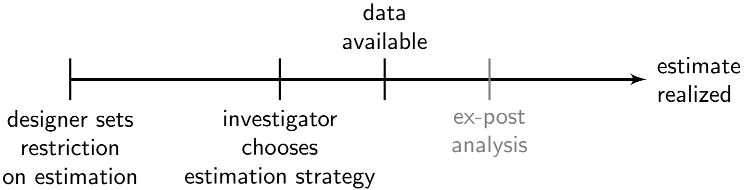
- Designer: $\mathrm{MSE}_\theta(\hat{\tau}) = \mathrm{E}_\theta[(\hat{\tau}(z) - \tau_\theta)^2] \to \min$
- Investigator: $\mathrm{RISK}_\theta(\hat{\tau}) = \mathrm{E}_\theta[(\hat{\tau}(z) - \tilde{\tau}_\theta)^2] \to \min$
  for some target $\tilde{\tau} : \Theta \to \mathbb{R}$
- No best estimator for all $\theta \to$ weigh by $\theta \sim \pi$ (Wald, 1950)
  - Investigator minimizes $\mathrm{E}_\pi \mathrm{RISK}_\theta(\hat{\tau})$
  - Designer *would want to* minimize $\mathrm{E}_\pi \mathrm{MSE}_\theta(\hat{\tau})$
- Distribution π private information of investigator

  $\to$ Designer faces a delegation problem

Timeline


data available
designer sets restriction on estimation
investigator chooses estimation strategy
ex-post analysis
estimate realized
choose mechanism (message space, mapping)
observe risk type and distribution, choose message
estimate realized

Timeline



restrict estimators to set $\mathcal{C}^D$

choose estimator $\hat{\tau}^I$ from $\mathcal{C}^D$

estimate $\hat{\tau}^I(z)$ realized

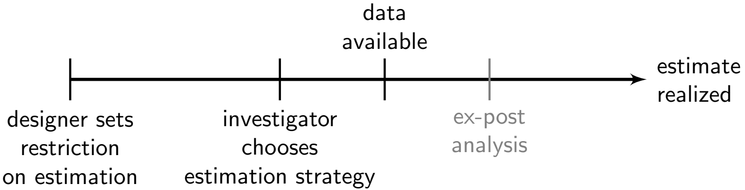
$\arg\min_{\mathcal{C}} \sup_{\text{RISK} \in \mathcal{R}} \mathrm{E}_\eta \text{MSE}_\theta(\hat{\tau}^I)$

$\arg\min_{\hat{\tau} \in \mathcal{C}^D} \mathrm{E}_\pi \text{RISK}_\theta(\hat{\tau})$

Example (First-best)

$\mathcal{R} = \{\text{MSE}\} \quad \rightarrow \quad \mathcal{C}^D$ unrestricted

Timeline



restrict estimators to set $\mathcal{C}^D$

choose estimator $\hat{\tau}^I$ from $\mathcal{C}^D$

→ estimate $\hat{\tau}^I(z)$ realized

$\arg\min_{\mathcal{C}} \sup_{\text{RISK} \in \mathcal{R}} \text{E}_\eta \text{MSE}_\theta(\hat{\tau}^I)$

$\arg\min_{\hat{\tau} \in \mathcal{C}^D} \text{E}_\pi \text{RISK}_\theta(\hat{\tau})$

Example (Left with no choice)

$\mathcal{R}$ unrestricted $\quad \rightarrow \quad \mathcal{C}^D = \{\hat{\tau}^D\}$

Results

1. Designer's solution: bias restriction
2. Investigator's solution: flexible unbiased estimators
   - Sample-splitting ensures unbiasedness
   - Prediction yields efficiency
3. Implementation: optimal pre-analysis plans
   - Specification searches without bias
   - Data distribution instead of pre-specification

Bias–variance trade-off

$$
\begin{aligned}
\mathsf{MSE}_\theta(\hat{\tau}) &= \mathsf{E}_\theta[(\hat{\tau}(z) - \tau_\theta)^2] \\
&= (\underbrace{\mathsf{E}_\theta[\hat{\tau}(z)] - \tau_\theta}_{\text{bias}})^2 + \underbrace{\mathsf{Var}_\theta(\hat{\tau})}_{\text{variance}}
\end{aligned}
$$

- Generally improve precision by allowing for bias
- Researcher may have different preference over trade-off

Bias–variance trade-off

$$\begin{aligned}\mathsf{RISK}_\theta(\hat{\tau}) &= \mathsf{E}_\theta[(\hat{\tau}(z) - 42)^2] \\ &= (\underbrace{\mathsf{E}_\theta[\hat{\tau}(z)] - \tau_\theta}_{\text{bias}} - (\tau_\theta - 42))^2 + \underbrace{\mathsf{Var}_\theta(\hat{\tau})}_{\text{variance}}\end{aligned}$$

- Generally improve precision by allowing for bias
- Researcher may have different preference over trade-off

Bias–variance trade-off

$$\begin{aligned}\mathsf{RISK}_\theta(\hat\tau) &= \mathsf{E}_\theta[(\hat\tau(z) - \tau_\theta - K)^2] \\ &= (\underbrace{\mathsf{const}_\theta}_{\text{bias}} - K)^2 + \underbrace{\mathsf{Var}_\theta(\hat\tau)}_{\text{variance}}\end{aligned}$$

- Generally improve precision by allowing for bias
- Researcher may have different preference over trade-off
- Among fixed-bias estimators, choices are aligned
- But is it worth the cost?

Designer's solution

Assumptions (Risk functions, random treatment, support)

- $\mathcal{R} = \{\mathrm{RISK}; \mathrm{RISK}_\theta(\hat{\tau}) = \mathrm{E}_\theta[(\hat{\tau}(z) - \tilde{\tau}_\theta)^2]$ for some $\tilde{\tau} : \Theta \to \mathbb{R}\}$
- Treatment random, outcomes have finite support
- π has full support η-a.s.

Theorem (Fixed bias is minimax optimal) ‣ Proof sketch

There exists $\beta : \Theta \to \mathbb{R}$ such that

$$\mathcal{C}^*_\beta \in \arg\min_{\mathcal{C}} \sup_{\mathrm{RISK} \in \mathcal{R}} \mathrm{E}_\eta \mathrm{MSE}(\hat{\tau}^I)$$

where $\mathcal{C}^*_\beta$ fixes biases $\beta_\theta = \mathrm{E}_\theta[\hat{\tau}] - \tau_\theta$ for all $\theta \in \Theta$

Aligned delegation analogy

Treatment-effect estimation	Grading (Frankel, 2014)
Designer	School principal
Researcher	Teacher
Estimation	Grading
Prior distribution	Student performance
Fix the bias	Fix the grade average

Estimation with fixed bias

$$E_\theta[\hat{\tau}(z)] = \tau_\theta \cdot \lambda \qquad \longleftrightarrow \qquad \hat{\tau}^I(z) = \Big(\overbrace{\hat{\tau}_0^D(z)}^{\substack{\text{chosen by designer,} \\ \text{unbiased}}} + \underbrace{\hat{\delta}^I(z)}_{\substack{\text{chosen by investigator,} \\ \text{mean-zero}}} \Big) \cdot \overbrace{\lambda}^{\text{chosen by designer}}$$

- Uninformed about preference → fix the bias
- Uninformed about treatment effect → to zero
  (invariant hyperprior/extend minimax)
- Some knowledge about distribution → e.g. shrinkage

Unbiased estimation

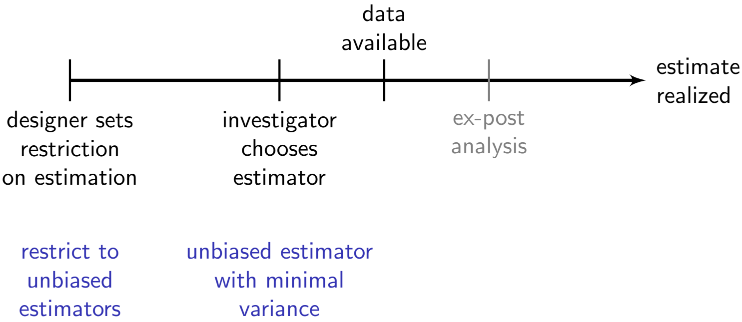
$$\mathsf{E}_\theta[\hat{\tau}(z)] = \tau_\theta \qquad \longleftrightarrow \qquad \hat{\tau}^I(z) = \overbrace{\hat{\tau}_0^D(z)}^{\text{chosen by designer, unbiased}} + \underbrace{\hat{\delta}^I(z)}_{\text{chosen by investigator, mean-zero}}$$

- In finite samples, aligns precision relative to *some* goal
- In large samples, once asymptotic Normality established, also:
  - Low p-value
  - Small standard error
  - Tight confidence interval

$\mathcal{N}(\tau, \sigma^2)$

- Does *not* align investigator who does *not* want to reject null

Timeline


data available
estimate realized
designer sets restriction on estimation
investigator chooses estimator
ex-post analysis
restrict to unbiased estimators
unbiased estimator with minimal variance

Results

1. Designer's solution: bias restriction
2. Investigator's solution: flexible unbiased estimators
   - Sample-splitting ensures unbiasedness
   - Prediction yields efficiency
3. Implementation: optimal pre-analysis plans
   - Specification searches without bias
   - Data distribution instead of pre-specification

Linear regression

$$y_i = \hat{\alpha} + d_i \underbrace{\hat{\tau}}_{\text{biased}} + \overbrace{x_i' \hat{\gamma}}^{\text{inefficient}} + \hat{\varepsilon}_i$$

(e.g. Freedman, 2008)

Causal effect on i (Rubin, 1974)

$$\tau_i = \underbrace{y_i(1) - \overbrace{y_i(0)}^{\text{potential outcome}}}_{\text{causal effect on } i}$$

Causal effect on i (Rubin, 1974)

$$\tau_i = \underbrace{\$19,320 - \overbrace{\$18,478}^{\text{earnings without training}}}_{\text{causal effect on } i} = \$842$$

Estimating τ_i

- For

$$y_i = y_i(\underbrace{d_i}_{\text{prob } p = .5}) = \begin{cases} \$19,320, & d_i = 1 \\ \$18,478, & d_i = 0 \end{cases}$$

estimate

$$\hat{\tau}_i = 2(2d_i - 1)y_i = \begin{cases} +\$39,640, & d_i = 1 \\ -\$36,956, & d_i = 0 \end{cases}$$

is unbiased for $\tau_i = \$842$ (e.g. Athey and Imbens, 2016)

- Extremely high variance

Estimating τ_i

- For

$$y_i = y_i(\underbrace{d_i}_{\text{prob } p=.5}) = \begin{cases} \$19,320, & d_i = 1 \\ \$18,478, & d_i = 0 \end{cases}$$

estimate

$$\hat{\tau}_i = 2(2d_i - 1)(y_i - \overbrace{\phi_i}^{=\$19{,}000}) = \begin{cases} +\$640, & d_i = 1 \\ +\$1,044, & d_i = 0 \end{cases}$$

is unbiased for $\tau_i = \$842$ (e.g. Athey and Imbens, 2016)

- Less variance through regression adjustment
- Unbiased provided ϕ_i uses only data from *other* units

Unbiasedness is sample-splitting (I)

Assumptions (Randomization I, finite support)

- Treatment is randomized independently with probability p
- Outcomes have finite support

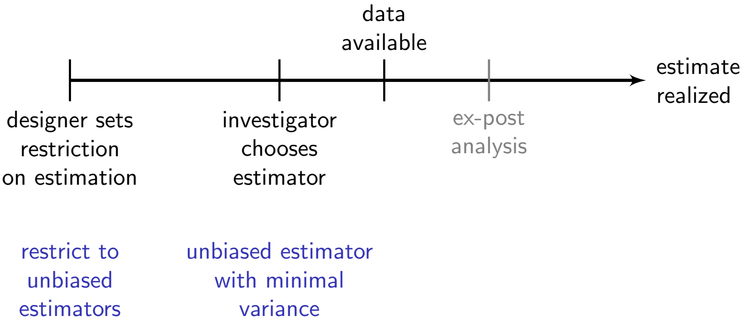
Lemma (Characterization of unbiased estimators, I) ‣ Proof sketch

For known p, $\hat{\tau}$ is unbiased *if and only if*

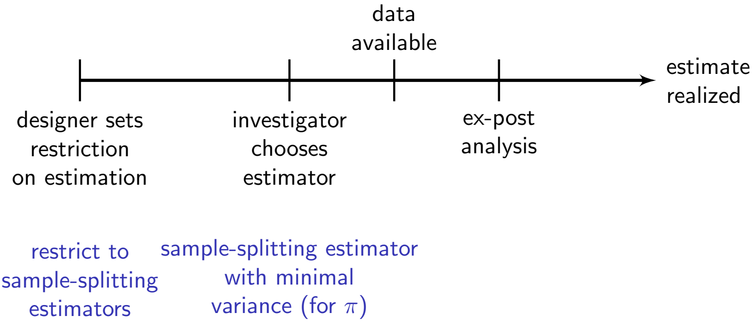
$$\hat{\tau}(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{d_i - p}{p(1-p)}(y_i - \overbrace{\phi_i(z_{-i})}^{\text{leave-one-out adjustment}})$$

- "Leave-one-out potential outcomes" (LOOP) estimator (Wu and Gagnon-Bartsch, 2017), going back to Aronow and Middleton (2013); Horvitz and Thompson (1952)

Timeline


data
available
estimate
realized
designer sets
restriction
on estimation
investigator
chooses
estimator
ex-post
analysis
restrict to
unbiased
estimators
unbiased estimator
with minimal
variance

Timeline


data available
estimate realized
designer sets restriction on estimation
investigator chooses estimator
ex-post analysis
restrict to sample-splitting estimators
sample-splitting estimator with minimal variance (for π)

Investigator's choice of adjustment

- Estimate of τ_i:

$$\hat{\tau}_i = 2(2d_i - 1)(y_i - \phi_i)$$

- Mistake at τ_i:

$$\hat{\tau}_i - \tau_i = 2(2d_i - 1)\left(\frac{y_i(1) + y_i(0)}{2} - \phi_i\right)$$

$\rightarrow$ Optimal infeasible choice:

$$\phi_i = \bar{y}_i = \frac{y_i(1) + y_i(0)}{2}$$

$\rightarrow$ Optimal feasible choice: best prediction of $\bar{y}_i$

Investigator's second-best analysis plan

Theorem (Solution of the investigator)

For known treatment probability p and prior π with

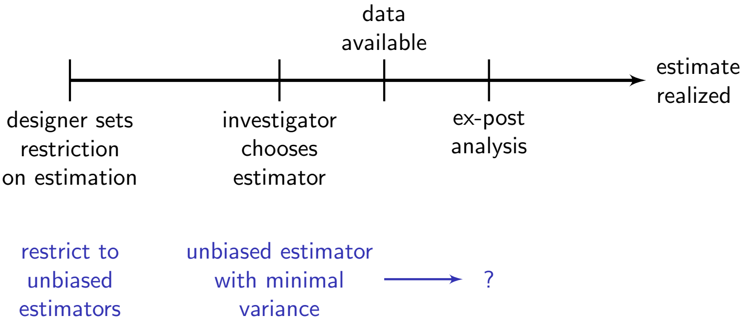
$$\mathrm{E}_\pi[\mathrm{E}_\pi[\bar{y}_j|y_i(1), z_{-ij}]|z_{-i}] = \mathrm{E}_\pi[\mathrm{E}_\pi[\bar{y}_j|y_i(0), z_{-ij}]|z_{-i}]$$

for $\bar{y}_i = (1-p)y_i(1) + py_i(0)$ the investigator chooses

$$\hat{\tau}(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{d_i - p}{p(1-p)}(y_i - \mathrm{E}_\pi[\bar{y}_i|z_{-i}])$$

- Adjustment $\mathrm{E}_\pi[\bar{y}_i|z_{-i}]$ minimize prediction risk
$$\mathrm{E}[w(d_i)(\hat{y}_i - y_i)^2]$$
with larger weight $w(d_i) = \left(\frac{(d_i-p)}{p(1-p)}\right)^2$ on smaller group
- Duality also holds in asymptotic approximation for K-fold

Timeline


data available
estimate realized
designer sets restriction on estimation
investigator chooses estimator
ex-post analysis
restrict to unbiased estimators
unbiased estimator with minimal variance
?

Results

1. Designer's solution: bias restriction
2. Investigator's solution: flexible unbiased estimators
   - Sample-splitting ensures unbiasedness
   - Prediction yields efficiency
3. Implementation: optimal pre-analysis plans
   - Specification searches without bias
   - Data distribution instead of pre-specification

“Cross-estimation” (Wager et al., 2016) implementation

$$\hat{\tau}(z) = \frac{2}{n}\sum_{i=1}^{n}(2d_i - 1)(y_i - \hat{y}_i)$$

“Cross-estimation” (Wager et al., 2016) implementation

$$\hat{\tau}(z) = \frac{2}{n}\sum_{i=1}^{n}(2d_i - 1)(y_i - \hat{f}_i(x_i))$$

Sample	1	2	3	4	5	6	
			Build $\hat{f}_i$ from				*Adjust at x_i*
Split 1		2	3	4	5	6	1
Split 2	1		3	4	5	6	2
Split 3	1	2		4	5	6	3
Split 4	1	2	3		5	6	4
Split 5	1	2	3	4		6	5
Split 6	1	2	3	4	5		6

"Cross-estimation" (Wager et al., 2016) implementation

$$\hat{\tau}(z) = \frac{2}{n}\sum_{i=1}^{n}(2d_i - 1)(y_i - \hat{f}_i(x_i))$$

Sample	1	2	3	4	5	6	
	Build $\hat{f}_i$ from						*Adjust at x_i*
Split 1		2	3	4	5	6	1
Split 2	1		3	4	5	6	2
Split 3	1	2		4	5	6	3
Split 4	1	2	3		5	6	4
Split 5	1	2	3	4		6	5
Split 6	1	2	3	4	5		6

First solution

$$\hat{\tau} \qquad \longleftrightarrow \qquad \hat{y}_i$$

- Pre-specify an algorithm that engages in specification searches
  - Divide the sample into K folds
  - Go through every fold k
    1. Train prediction function $\hat{f}$ on (y_j, d_j, x_j), j not in fold k with
    $$\mathrm{E}[w(d)(y - \hat{f}(x))^2] \to \min$$
    2. Adjust y_i by $\hat{f}(x_i)$, i in fold k
  - Estimate ATE from adjusted outcome

$$\mathrm{Var}(\hat{\tau}) \approx \frac{1}{np(1-p)} \left(\mathrm{E}[w(d)(y - \hat{f}(x))^2] - p(1-p)\tau \right)$$

- Always unbiased, quality estimable $\longrightarrow$ nonparametrics (Wager et al., 2016; Wu and Gagnon-Bartsch, 2017)
  - Model selection, model averaging, shrinkage

Second solution

$$\hat{\tau}(z) = \frac{2}{n}\sum_{i=1}^{n}(2d_i - 1)(y_i - \hat{f}_i(x_i))$$

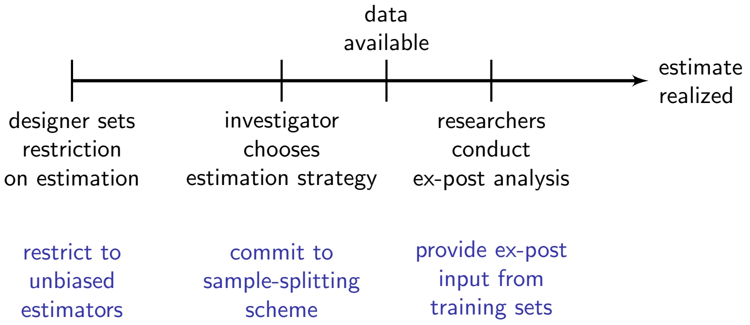
Sample	1	2	3	4	5	6	
			Build $\hat{f}_i$ from				*Adjust at x_i*
Researcher 1		2	3	4	5	6	1
Researcher 2	1		3	4	5	6	2
Researcher 3	1	2		4	5	6	3
Researcher 4	1	2	3		5	6	4
Researcher 5	1	2	3	4		6	5
Researcher 6	1	2	3	4	5		6

Second solution

$$\hat{\tau}(z) = \frac{2}{n}\sum_{k=1}^{K}\sum_{i \in S^k}(2d_i - 1)(y_i - \hat{f}^k(x_i))$$

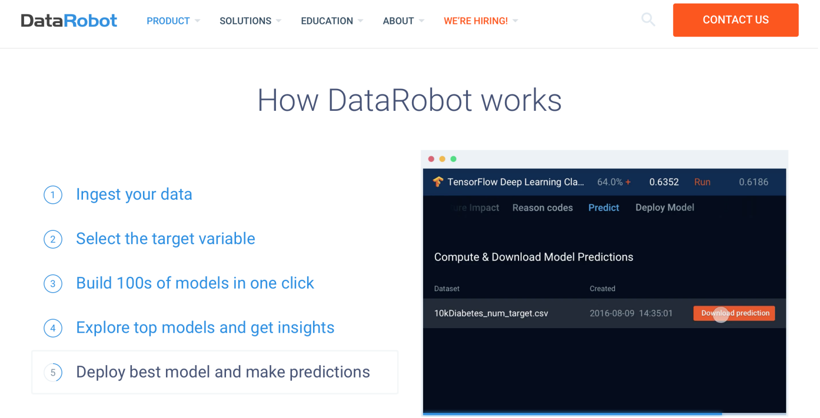
Sample	1	2	3	4	5	6	
		Build $\hat{f}^k$ from					*Adjust at x_i*
Researcher 1				4	5	6	1 2 3
Researcher 2	1	2	3				4 5 6

Timeline

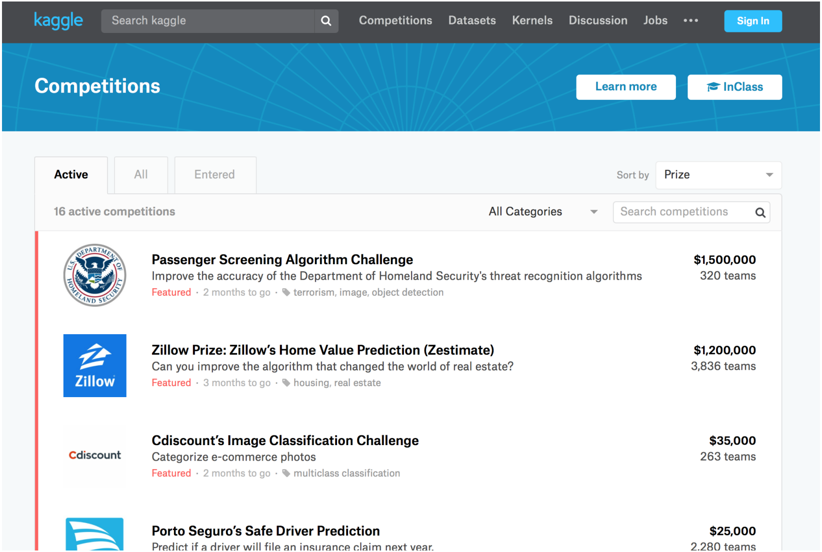

data available
estimate realized
designer sets restriction on estimation
investigator chooses estimation strategy
researchers conduct ex-post analysis
restrict to unbiased estimators
commit to sample-splitting scheme
provide ex-post input from training sets

Prediction on demand

▶▶ Summary



Causal prediction competitions



Results

1. Designer's solution: bias restriction
2. Investigator's solution: flexible unbiased estimators
   - Sample-splitting ensures unbiasedness
   - Prediction yields efficiency
3. Implementation: optimal pre-analysis plans
   - Specification searches without bias
   - Data distribution instead of pre-specification

Conclusion

- Econometric approach that acknowledges researcher degrees of freedom and preferences $\rightarrow$ research protocols
  - Experimental analysis
  - + Endogenous treatment

- Connection between causal estimation and nonparametric prediction $\rightarrow$ beneficial specification searches
  - Control variables
  - + Other implicit prediction tasks, e.g. instrumental variables

Lecture outline

1. "Optimal Estimation when Researcher and Social Preferences are Misaligned" (2018; revised 2022)
2. High-level model and integrating machine learning/AI
3. Pre-analysis plans and implementability (with Max Kasy)
4. Summary and conclusion

Delegation approach to econometric decisions

1. **Designer** observes η and chooses $\mathcal{C} \subseteq \mathbb{R}^{\mathcal{Z}}$ to minimize
$$\mathsf{E}_\eta \, \mathsf{E}_\pi L^D(\hat{\tau}(\mathcal{C}); \theta)$$

2. **Researcher** observes $\pi \sim \mathsf{P}_\eta$ and chooses $\hat{\tau} \in \mathcal{C}$ to minimize
$$\mathsf{E}_\pi L^R(\hat{\tau}; \theta)$$

Delegation approach to econometric decisions

1. **Designer** observes η and chooses $\mathcal{C} \subseteq \mathbb{R}^{\mathcal{Z}}$ to minimize
$$\mathsf{E}_\eta \, \mathsf{E}_\pi L^D(\hat{\tau}(\mathcal{C}); \theta)$$

2. **Researcher** observes $\pi \sim \mathsf{P}_\eta$ and chooses $\hat{\tau} \in \mathcal{C}$ to minimize
$$\mathsf{E}_\pi L^R(\hat{\tau}; \theta)$$
by specifiying a function class $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$, loss function $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and mapping $T : \mathcal{F} \to \mathcal{C}, \hat{f} \mapsto \hat{\tau}$

3. **Machine-learning algorithm** observes data $z \sim \mathsf{P}_\theta$, chooses $\hat{f}$ to minimize (optimistically)
$$\mathsf{E}_\pi[\mathsf{E}_\theta[\ell(\hat{f}(x), y)]|z]$$
or (practically)
$$\mathbb{E}_z[\ell(\hat{f}(x), y)]$$

Delegation view on integrating ML into econometrics

- **Goal:** Assume we want to choose $\hat{\tau} \in \mathcal{C} \subseteq \mathbb{R}^{\mathcal{Z}}$ to minimize

$$\mathsf{E}_\pi L(\hat{\tau}; \theta)$$

- **Delegation view:** Design a function class $\mathcal{F}$, loss function ℓ, optimization routine (empirical risk minimization)

$$\arg\min_f \mathbb{E}_z[\ell(f(x), y)],$$

  and mapping $T : \mathcal{F} \to \mathcal{C}$
  - **Robustness:** $T(\hat{f}) \in \mathcal{C}$ for all $\hat{f} \in \mathcal{F}$
  - **Efficiency:** $\hat{\tau} = T(\hat{f})$ good solution to original goal

Data-driven decisions with multiple objectives across domains

- Strategic classification (Hardt et al., 2016)
- Manipulation-proof machine learning (Björkegren et al., 2020)
- Performative prediction (Perdomo et al., 2020)
- Regulation of AI (Rambachan et al., 2020)
- AI alignment (Hadfield-Menell and Hadfield, 2019)
- Prediction-powered inference (Angelopoulos et al., 2023)

Lecture outline

1. "Optimal Estimation when Researcher and Social Preferences are Misaligned" (2018; revised 2022)

2. High-level model and integrating machine learning/AI

3. Pre-analysis plans and implementability (with Max Kasy)

4. Summary and conclusion

Pre-analysis plans and implementability (with Max Kasy)

- Delegation with misaligned preferences, private info

Pre-analysis plans and implementability (with Max Kasy)

- Delegation with misaligned preferences, private info
  - Designer wants to implement a mapping

    $$a : (\text{private info}, \text{data}) \mapsto \text{decision} \in \{\text{accept}, \text{deny}\},$$

    but lacks private info
  - Researcher has private info, but always prefers accept

First idea: role of pre-analysis plans

- Common view: pre-analysis plan (PAP) ensures valid inference
- <u>First idea:</u> PAPs increase implementable decision rules
  - Baseline (no PAP): mechanism of form

    $$(\text{post-data message}, \text{data}) \mapsto \text{decision}$$

    limits which decision rules a can be implemented
  - Pre-commitment (PAP): mechanisms

    $$(\text{pre-data message}, \text{data}) \mapsto \text{decision}$$

    increase space of implementable decision rules a

$\rightarrow$ Characterization of implementable decision rules and optimal PAPs (allows for simplicity constraints on message space)

Second idea: PAPs with partial verifiability

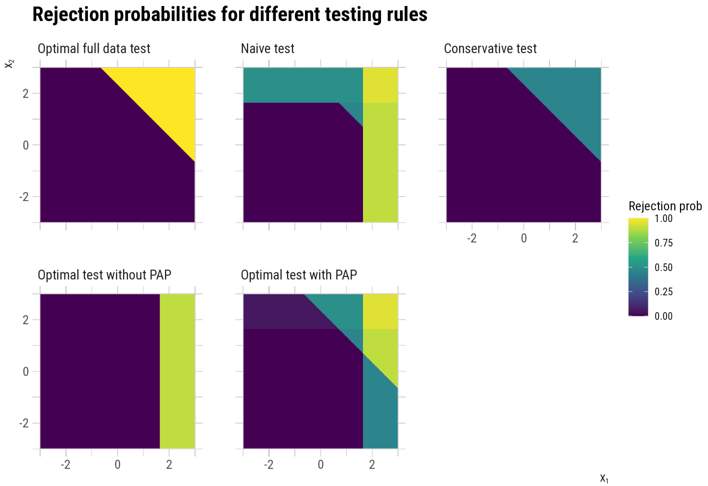
- Data availability ex-ante uncertain, may be selectively reported
- Second idea: Value of PAPs with partial verifiability
  - Designer wants to implement a mapping

    $$(\text{private info}, \text{available data}) \mapsto \text{decision}$$

    but does not know which data is available
  - Researcher learns availability, decides what to report
  - Mechanisms with PAP of form

    $$(\text{pre-data message}, \text{reported data}) \mapsto \text{decision}$$

Illustration: joint testing of $\theta \leq 0$, $X_1, X_2 \sim \mathcal{N}(\theta, 1)$



$X_1, X_2 \sim N(\theta, 1)$, independently. H0: $\theta < 0$. Probabilities of observing X_1 and X_2 are 0.9 and 0.5.

Illustration: joint testing of $\theta \leq 0$, $X_1, X_2 \sim \mathcal{N}(\theta, 1)$

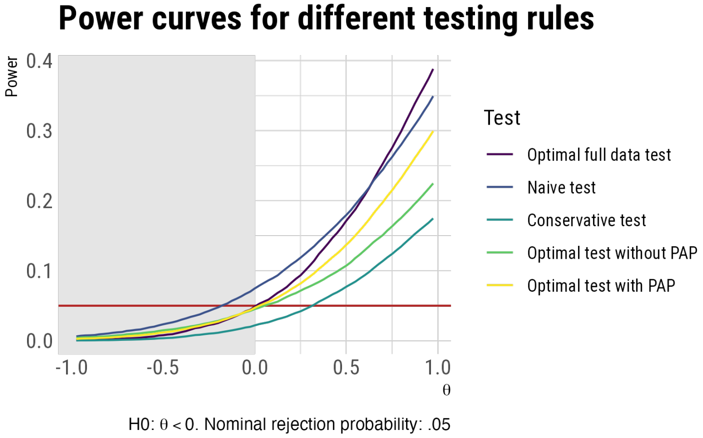
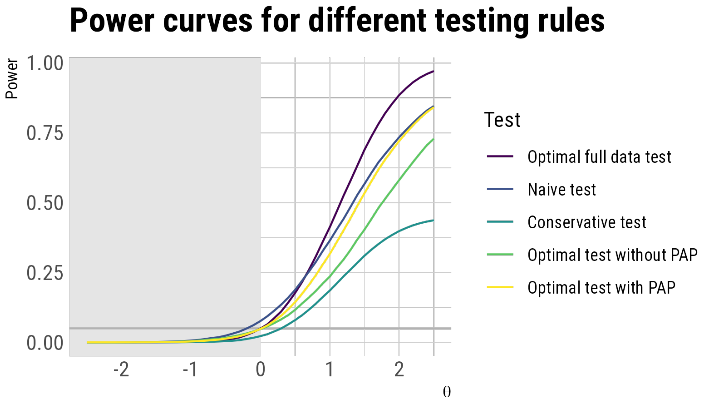
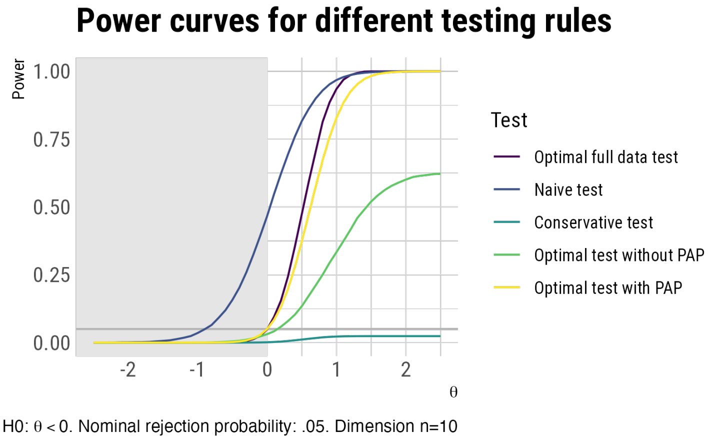


Illustration: joint testing of $\theta \leq 0$, $X_1, X_2 \sim \mathcal{N}(\theta, 1)$



H0: $\theta < 0$. Nominal rejection probability: .05. Dimension n=2

Illustration: joint testing of $\theta \leq 0$, $X_1, X_2 \sim \mathcal{N}(\theta, 1)$



Lecture outline

1. "Optimal Estimation when Researcher and Social Preferences are Misaligned" (2018; revised 2022)

2. High-level model and integrating machine learning/AI

3. Pre-analysis plans and implementability (with Max Kasy)

4. Summary and conclusion

Summary and conclusion

- Empirical estimates reflect not just data, but also researcher decisions and incentives
- How can we approach statistical decisions when there are conflicts of interest?
- **Approach in my lecture today:** embed data-driven decisions in principal-agent framework
  - Can be good frame to diagnose and address misalignment
  - Allows leveraging formal tools from mechanism design
- Has been and can be applied widely across fields
  - Design of pre-analysis plans
  - Integrating ML into causal inference
  - Regulation of AI

jspiess@stanford.edu

References (I)

Angelopoulos, A. N., Bates, S., Fannjiang, C., Jordan, M. I., and Zrnic, T. (2023). Prediction-powered inference. *arXiv preprint arXiv:2301.09633*.

Björkegren, D., Blumenstock, J. E., and Knight, S. (2020). Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*.

Hadfield-Menell, D. and Hadfield, G. K. (2019). Incomplete contracting and ai alignment. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 417–422.

Hardt, M., Megiddo, N., Papadimitriou, C., and Wootters, M. (2016). Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, 77(5):1683–1701.

Perdomo, J., Zrnic, T., Mendler-Dünner, C., and Hardt, M. (2020). Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR.

Rambachan, A., Kleinberg, J., Mullainathan, S., and Ludwig, J. (2020). An economic approach to regulating algorithms. Technical report, National Bureau of Economic Research.