Learning What to Learn: Experimental Design when Combining Experimental with Observational Evidence

Aristotle Epanomeritakis   Davide Viviano

Department of Economics, Harvard University

Frontiers in Econometric Methods, NBER SI 2026

Problem description

- Growing availability of experiments in social sciences and industry
  - Internal validity: mitigate (potential) bias of obs/ evidence
  - External validity: often limited due to cost/feasibility constraints (Muralidharan and Niehaus, 2017)

- Nevertheless, often experiments are meant to draw broader conclusions
  - GE effect of a cash transfer? (e.g. Attanasio et al., 2012)
  - TE effect on a wider population?(e.g. Gechter et al. 2024)
  - Equilibrium effects in supply/demand system? (e.g. Alcott et al., 2025)

Problem description

- Growing availability of experiments in social sciences and industry
  - Internal validity: mitigate (potential) bias of obs/ evidence
  - External validity: often limited due to cost/feasibility constraints (Muralidharan and Niehaus, 2017)
- Nevertheless, often experiments are meant to draw broader conclusions
  - GE effect of a cash transfer? (e.g. Attanasio et al., 2012)
  - TE effect on a wider population?(e.g. Gechter et al. 2024)
  - Equilibrium effects in supply/demand system? (e.g. Alcott et al., 2025)

$\Rightarrow$ Augment experiments with obs evidence (model-based or reduced-form)

  - Goal is extrapolation of 'localized' effects identified by experiment
  - $\sim 30\%$ of published AEA experimental articles in past decade

Problem description

- Growing availability of experiments in social sciences and industry
  - Internal validity: mitigate (potential) bias of obs/ evidence
  - External validity: often limited due to cost/feasibility constraints (Muralidharan and Niehaus, 2017)

- Nevertheless, often experiments are meant to draw broader conclusions
  - GE effect of a cash transfer? (e.g. Attanasio et al., 2012)
  - TE effect on a wider population?(e.g. Gechter et al. 2024)
  - Equilibrium effects in supply/demand system? (e.g. Alcott et al., 2025)

$\Rightarrow$ Augment experiments with obs evidence (model-based or reduced-form)
  - Goal is extrapolation of 'localized' effects identified by experiment
  - $\sim 30\%$ of published AEA experimental articles in past decade

  $\Rightarrow$ Potential biases may arise (selection, confounding, misspecification)

Problem description

- Growing availability of experiments in social sciences and industry
  - Internal validity: mitigate (potential) bias of obs/ evidence
  - External validity: often limited due to cost/feasibility constraints (Muralidharan and Niehaus, 2017)

- Nevertheless, often experiments are meant to draw broader conclusions
  - GE effect of a cash transfer? (e.g. Attanasio et al., 2012)
  - TE effect on a wider population?(e.g. Gechter et al. 2024)
  - Equilibrium effects in supply/demand system? (e.g. Alcott et al., 2025)

$\Rightarrow$ Augment experiments with obs evidence (model-based or reduced-form)
  - Goal is extrapolation of 'localized' effects identified by experiment
  - $\sim 30\%$ of published AEA experimental articles in past decade

  $\Rightarrow$ Potential biases may arise (selection, confounding, misspecification)

- Which experiments to run (and how) for these extrapolations?

Some stats from NBER SI 2026, Development

Some stats from NBER SI 2026, Development

Look at the list of papers carefully here. How many of these paper use a model or observational evidence in combination with an experiment? https://www.nber.org/conferences/si-2026-development-economics Be very precise

Worked for 11m 47s

The defensible headline is **11 of 17 papers (64.7%)**. Among the **14 papers containing a randomized experiment**, **11 of 14 (78.6%)** combine it with an explicit economic model or a distinct nonexperimental empirical component.

A narrower definition—requiring the model or observational component to provide a separate quantitative input, rather than mainly organizing hypotheses—gives **8 of 17 papers (47.1%)**, or **8 of 14 experimental papers (57.1%)**.

Summary of the paper

- Main contribution: framework for optimal choice of experiments to be combined with external (e.g. observational) evidence for extrapolation
- External evidence may be biased in ways unknown ex-ante
- Costs or feasibility constraints are first order

Summary of the paper

- Main contribution: framework for optimal choice of experiments to be combined with external (e.g. observational) evidence for extrapolation
- External evidence may be biased in ways unknown ex-ante
- Costs or feasibility constraints are first order

- Given existing external evidence, we proceed by
  1. laying out framework for experimental design in this context
  2. selecting which experiments to run from some feasible set
  3. setting experimental precision, e.g. via sample allocation

Summary of the paper

- Main contribution: framework for optimal choice of experiments to be combined with external (e.g. observational) evidence for extrapolation
- External evidence may be biased in ways unknown ex-ante
- Costs or feasibility constraints are first order

- Given existing external evidence, we proceed by
  1. laying out framework for experimental design in this context
  2. selecting which experiments to run from some feasible set
  3. setting experimental precision, e.g. via sample allocation
- We study two research environments where researchers
  - **report an estimate** for the target effect using available data

Summary of the paper

- Main contribution: framework for optimal choice of experiments to be combined with external (e.g. observational) evidence for extrapolation
- External evidence may be biased in ways unknown ex-ante
- Costs or feasibility constraints are first order

- Given existing external evidence, we proceed by
  1. laying out framework for experimental design in this context
  2. selecting which experiments to run from some feasible set
  3. setting experimental precision, e.g. via sample allocation
- We study two research environments where researchers
  - **report an estimate** for the target effect using available data $\Rightarrow$ focus on smooth method of moments estimators

Summary of the paper

- Main contribution: framework for optimal choice of experiments to be combined with external (e.g. observational) evidence for extrapolation
- External evidence may be biased in ways unknown ex-ante
- Costs or feasibility constraints are first order

- Given existing external evidence, we proceed by
  1. laying out framework for experimental design in this context
  2. selecting which experiments to run from some feasible set
  3. setting experimental precision, e.g. via sample allocation
- We study two research environments where researchers
  - **report an estimate** for the target effect using available data $\Rightarrow$ focus on smooth method of moments estimators
  - **report all collected results** to audience that updates beliefs

Related literature

- Experimental design:
  - Variance-optimal designs, with no bias from final experimental estimate (e.g. Bai, 2019; Chaloner and Verdinelli, 1995; Cytrynbaum, 2021; Dette, 1997; Higbee, 2024; Kiefer and Wolfowitz, 1959; Rosenman and Owen, 2021)
  - Test two competing models or assume a priori worst-case biases (e.g Atkinson and Fedorov, 1975; Box and Draper, 1959; Tsirpitzi et al., 2023)
  - Recent work on decision theory focus on researchers have only access to experimental variation (e.g. Abadie and Zhao, 2021; Banerjee et al., 2020; Breza et al., 2025; Dominitz and F. Manski, 2017; Gechter et al., 2024; Hu et al., 2024; Manski and Tetenov, 2016; Olea et al., 2024)

$\rightarrow$ Our framework studies which experiment to run and how when experimental parameters are combined with observational ones

Related literature

- Experimental design:
  - Variance-optimal designs, with no bias from final experimental estimate (e.g. Bai, 2019; Chaloner and Verdinelli, 1995; Cytrynbaum, 2021; Dette, 1997; Higbee, 2024; Kiefer and Wolfowitz, 1959; Rosenman and Owen, 2021)
  - Test two competing models or assume a priori worst-case biases (e.g Atkinson and Fedorov, 1975; Box and Draper, 1959; Tsirpitzi et al., 2023)
  - Recent work on decision theory focus on researchers have only access to experimental variation (e.g. Abadie and Zhao, 2021; Banerjee et al., 2020; Breza et al., 2025; Dominitz and F. Manski, 2017; Gechter et al., 2024; Hu et al., 2024; Manski and Tetenov, 2016; Olea et al., 2024)

  → Our framework studies which experiment to run and how when experimental parameters are combined with observational ones

- Data fusion and misspecification assume a fixed design (e.g. Andrews et al., 2025, 2020; Andrews and Shapiro, 2021; Armstrong et al., 2024; Armstrong and Kolesár, 2021; Athey et al., 2020, 2025; Bickel, 1984; Bonhomme and Weidner, 2022; de Chaisemartin and D'Haultfœuille, 2020; Kempthorne, 1988; Tsybakov, 1998)

  → Here experimental design: different objective and optimization problem

Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Basic setup

- Estimand $\tau(\theta) \in \mathbb{R}$, with (unknown) parameters $\theta \in \mathbb{R}^p$, and known τ mapping
- Feasibility constraints may prevent learning all parameters experimentally

Basic setup

- Estimand $\tau(\theta) \in \mathbb{R}$, with (unknown) parameters $\theta \in \mathbb{R}^p$, and known τ mapping
- Feasibility constraints may prevent learning all parameters experimentally

0. Existing "observational" estimators (with unknown bias b)

$$\tilde{\theta}^{\text{obs}} \in \mathbb{R}^p, \qquad \mathbb{E}[\tilde{\theta}^{\text{obs}}] - \theta = b$$

Basic setup

- Estimand $\tau(\theta) \in \mathbb{R}$, with (unknown) parameters $\theta \in \mathbb{R}^p$, and known τ mapping
- Feasibility constraints may prevent learning all parameters experimentally

0. Existing "observational" estimators (with unknown bias b)

$$\tilde{\theta}^{\text{obs}} \in \mathbb{R}^p, \qquad \mathbb{E}[\tilde{\theta}^{\text{obs}}] - \theta = b$$

1. Select entries of parameter vector to identify experimentally, $\mathcal{E} \subset \{1, \ldots, p\}$, resulting in additional experimental estimators (under feasibility constraints):

$$\tilde{\theta}^{\text{exp}}_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}, \qquad \mathbb{E}[\tilde{\theta}^{\text{exp}}_{\mathcal{E}}] - \theta_{\mathcal{E}} = 0$$

Basic setup

- Estimand $\tau(\theta) \in \mathbb{R}$, with (unknown) parameters $\theta \in \mathbb{R}^p$, and known τ mapping
- Feasibility constraints may prevent learning all parameters experimentally

0. Existing "observational" estimators (with unknown bias b)

$$\tilde{\theta}^{\text{obs}} \in \mathbb{R}^p, \qquad \mathbb{E}[\tilde{\theta}^{\text{obs}}] - \theta = b$$

1. Select entries of parameter vector to identify experimentally, $\mathcal{E} \subset \{1, \ldots, p\}$, resulting in additional experimental estimators (under feasibility constraints):

$$\tilde{\theta}^{\text{exp}}_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}, \qquad \mathbb{E}[\tilde{\theta}^{\text{exp}}_{\mathcal{E}}] - \theta_{\mathcal{E}} = 0$$

2. Set precision, subject to constraints, where $\Sigma(\mathcal{E}) := \mathbb{V}[\tilde{\theta}^{\text{obs}}, \tilde{\theta}^{\text{exp}}_{\mathcal{E}}]$ (assumed positive definite and bounded)

Basic setup

- Estimand $\tau(\theta) \in \mathbb{R}$, with (unknown) parameters $\theta \in \mathbb{R}^p$, and known τ mapping
- Feasibility constraints may prevent learning all parameters experimentally

0. Existing "observational" estimators (with unknown bias b)

$$\tilde{\theta}^{\text{obs}} \in \mathbb{R}^p, \qquad \mathbb{E}[\tilde{\theta}^{\text{obs}}] - \theta = b$$

1. Select entries of parameter vector to identify experimentally, $\mathcal{E} \subset \{1, \ldots, p\}$, resulting in additional experimental estimators (under feasibility constraints):

$$\tilde{\theta}^{\text{exp}}_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}, \qquad \mathbb{E}[\tilde{\theta}^{\text{exp}}_{\mathcal{E}}] - \theta_{\mathcal{E}} = 0$$

2. Set precision, subject to constraints, where $\Sigma(\mathcal{E}) := \mathbb{V}[\tilde{\theta}^{\text{obs}}, \tilde{\theta}^{\text{exp}}_{\mathcal{E}}]$ (assumed positive definite and bounded)

In the paper:

- Observational and experiment identify moments, instead of parameters
- Σ intended as expected variance; if estimated, error typically of second order

Two illustrative examples

- GE effects of cash transfers [Todd and Wolpin, 2006; Attanasio et al., 2012]
  - $\tau(\theta)$: effect of a cash-transfer policy on school attendance in Kenya
  - θ: responses to the stipend, household income, and eq/ wages
  - $\tilde{\theta}^{\text{obs}}$: estimates from PROGRESA in Mexico, may not transport

Two illustrative examples

- GE effects of cash transfers [Todd and Wolpin, 2006; Attanasio et al., 2012]
  - $\tau(\theta)$: effect of a cash-transfer policy on school attendance in Kenya
  - θ: responses to the stipend, household income, and eq/ wages
  - $\tilde{\theta}^{\text{obs}}$: estimates from PROGRESA in Mexico, may not transport
  - Q: which responses to estimate experimentally, and how precise?
  - Constraint: cost constraints on experiment size; estimate model to extrapolate [Average JPAL experiment $n \approx 3,700$, Viviano et al. (2026)]

Two illustrative examples

- GE effects of cash transfers [Todd and Wolpin, 2006; Attanasio et al., 2012]
  - $\tau(\theta)$: effect of a cash-transfer policy on school attendance in Kenya
  - θ: responses to the stipend, household income, and eq/ wages
  - $\tilde{\theta}^{\text{obs}}$: estimates from PROGRESA in Mexico, may not transport
  - Q: which responses to estimate experimentally, and how precise?
  - Constraint: cost constraints on experiment size; estimate model to extrapolate [Average JPAL experiment $n \approx 3,700$, Viviano et al. (2026)]

- Site selection [Gechter et al., 2024; Olea et al., 2025]
  - $\tau(\theta)$: average effect of scaling the intervention to the full population
  - θ: treatment effects in each region
  - $\tilde{\theta}^{\text{obs}}$: observational/model-based estimates of the regional effects
  - Q: in which region to experiment and how to allocate the sample?

Approximate linearity and sensitivity

- Assumption (First-order approximation) At a preliminary value θ_0^{obs},

$$\tau(\theta) = \underbrace{\tau(\theta_0^{\text{obs}}) + \omega^\top(\theta - \theta_0^{\text{obs}})}_{\text{Linearized } \tau(\theta)}$$

Approximate linearity and sensitivity

- Assumption (First-order approximation) At a preliminary value θ_0^{obs},

$$\tau(\theta) = \underbrace{\tau(\theta_0^{\text{obs}}) + \omega^\top(\theta - \theta_0^{\text{obs}})}_{\text{Linearized } \tau(\theta)} + \underbrace{O\left(\|\theta - \theta_0^{\text{obs}}\|^2\right)}_{\text{second-order } o(n^{-1/2})}, \qquad \omega = \frac{\partial \tau(\theta)}{\partial \theta}\bigg|_{\theta = \theta_0^{\text{obs}}}.$$

Approximate linearity and sensitivity

- Assumption (First-order approximation) At a preliminary value θ_0^{obs},

$$\tau(\theta) = \underbrace{\tau(\theta_0^{\text{obs}}) + \omega^\top(\theta - \theta_0^{\text{obs}})}_{\text{Linearized } \tau(\theta)} + \underbrace{O\left(\|\theta - \theta_0^{\text{obs}}\|^2\right)}_{\text{second-order } o(n^{-1/2})}, \qquad \omega = \left.\frac{\partial \tau(\theta)}{\partial \theta}\right|_{\theta=\theta_0^{\text{obs}}}.$$

  - Under local asymptotics, remainder negligible: b of the same or different order than standard errors, while local to zero $\Rightarrow \theta_0^{\text{obs}}$ is informative
  - ω analogous to *sensitivity* in Andrews et al. (2020): first-order effect of b

Approximate linearity and sensitivity

- Assumption (First-order approximation) At a preliminary value θ_0^{obs},

$$\tau(\theta) = \underbrace{\tau(\theta_0^{\text{obs}}) + \omega^\top(\theta - \theta_0^{\text{obs}})}_{\text{Linearized } \tau(\theta)} + \underbrace{O\left(\|\theta - \theta_0^{\text{obs}}\|^2\right)}_{\text{second-order } o(n^{-1/2})}, \qquad \omega = \frac{\partial \tau(\theta)}{\partial \theta}\Big|_{\theta=\theta_0^{\text{obs}}}.$$

  - Under local asymptotics, remainder negligible: b of the same or different order than standard errors, while local to zero $\Rightarrow \theta_0^{\text{obs}}$ is informative
  - ω analogous to *sensitivity* in Andrews et al. (2020): first-order effect of b

- Key feature: $\tau(\cdot)$ is a *known* parametric function
  - Makes the researcher's structural assumptions explicit
  - Shows which parameters matter most for experimental design

Outline

1 Setup: baseline case

2 Optimal design for common estimators

3 General reporting rules

4 Extensions

5 Illustration

6 Conclusion

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \big) \cdot \tilde{\theta}_j^{\text{obs}}$$

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}}\big) \cdot \tilde{\theta}_j^{\text{obs}} + \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \tilde{\theta}_j^{\text{exp}}$$

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}}\big) \cdot \tilde{\theta}_j^{\text{obs}} + \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \tilde{\theta}_j^{\text{exp}}$$

  - γ can depend on $\mathcal{E}, \Sigma$, or replaced by estimate $\hat{\gamma} \to_p \gamma$ for some vector γ.
  - Important restriction is γ is not random

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}}\big) \cdot \tilde{\theta}_j^{\text{obs}} + \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \tilde{\theta}_j^{\text{exp}}$$

  - γ can depend on $\mathcal{E}, \Sigma$, or replaced by estimate $\hat{\gamma} \to_p \gamma$ for some vector γ.
  - Important restriction is γ is not random

- We focus on mean squared error, for given $(\mathcal{E}, \Sigma)$

$$\text{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \mathbb{E}_{\mathcal{E},\Sigma,b}\left[\Big(\tau\big(\hat{\theta}(\gamma)\big) - \tau(\theta)\Big)^2\right]$$

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \Big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}}\Big) \cdot \tilde{\theta}_j^{\text{obs}} + \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \tilde{\theta}_j^{\text{exp}}$$

  - γ can depend on $\mathcal{E}, \Sigma$, or replaced by estimate $\hat{\gamma} \to_p \gamma$ for some vector γ.
  - Important restriction is γ is not random

- We focus on mean squared error, for given $(\mathcal{E}, \Sigma)$

$$\text{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \mathbb{E}_{\mathcal{E},\Sigma,b}\left[\Big(\tau\big(\hat{\theta}(\gamma)\big) - \tau(\theta)\Big)^2\right]$$

- Because b unknown, simple uncertainty set is ℓ_∞ (arbitrary norms possible)

$$\mathcal{B}(B) \equiv \{b : \ \|b\|_\infty \leq B\}, \qquad B : \ \text{radius of the bias}$$

Description of the researcher's problem

- Compute estimator $\tau(\hat{\theta}(\gamma))$ for weights $\gamma \in \mathbb{R}^{|\mathcal{E}|}$,

$$\hat{\theta}_j(\gamma) = \big(1 - \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}}\big) \cdot \tilde{\theta}_j^{\text{obs}} + \gamma_j \cdot \underbrace{1\{j \in \mathcal{E}\}}_{\text{experiment is available}} \tilde{\theta}_j^{\text{exp}}$$

  - γ can depend on $\mathcal{E}, \Sigma$, or replaced by estimate $\hat{\gamma} \to_p \gamma$ for some vector γ.
  - Important restriction is γ is not random

- We focus on mean squared error, for given $(\mathcal{E}, \Sigma)$

$$\text{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \mathbb{E}_{\mathcal{E},\Sigma,b}\left[\Big(\tau\big(\hat{\theta}(\gamma)\big) - \tau(\theta)\Big)^2\right]$$

- Because b unknown, simple uncertainty set is ℓ_∞ (arbitrary norms possible)

$$\mathcal{B}(B) \equiv \{b: \ \|b\|_\infty \leq B\}, \qquad B: \text{ radius of the bias}$$

- An oracle that knows B would pick the minimax design

$$\text{MSE}^*(B) \ \equiv \inf_{\text{feasible } (\mathcal{E},\Sigma,\gamma)} \ \sup_{b \in \mathcal{B}(B)} \text{MSE}_b(\mathcal{E}, \Sigma, \gamma)$$

Oracle and regret

- Problem: we do not know B and want to be robust (or we may only have a very conservative bound)

Oracle and regret

- Problem: we do not know B and want to be robust (or we may only have a very conservative bound)

- The researcher minimizes the worst-case proportional regret

$$\mathcal{R}(\mathcal{E}, \Sigma, \gamma) \equiv \sup_{B \geq 0} \frac{\sup_{b \in \mathcal{B}(B)} \mathrm{MSE}_b(\mathcal{E}, \Sigma, \gamma)}{\mathrm{MSE}^*(B)}$$

$\Rightarrow$ Common performance measure under fixed design.

$\Rightarrow$ Here oracle and researcher choose the design.

Solution to the researcher's problem

- Standard decomposition:

$$\sup_{b \in \mathcal{B}(B)} \mathrm{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \underbrace{\alpha(\mathcal{E}, \Sigma, \gamma)}_{\text{Variance}} + \underbrace{B^2\, \beta(\mathcal{E}, \gamma)}_{\text{Bias}^2},$$

Solution to the researcher's problem

- Standard decomposition:

$$\sup_{b \in \mathcal{B}(B)} \text{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \underbrace{\alpha(\mathcal{E}, \Sigma, \gamma)}_{\text{Variance}} + \underbrace{B^2 \, \beta(\mathcal{E}, \gamma)}_{\text{Bias}^2},$$

$$\alpha(\mathcal{E}, \Sigma, \gamma) \equiv \underbrace{\mathbb{V}_{\mathcal{E},\Sigma}\big[\, \tau(\hat{\theta}(\gamma)) \,\big]}_{\text{Variance}}$$

Solution to the researcher's problem

- Standard decomposition:

$$\sup_{b\in\mathcal{B}(B)} \text{MSE}_b(\mathcal{E},\Sigma,\gamma) = \underbrace{\alpha(\mathcal{E},\Sigma,\gamma)}_{\text{Variance}} + \underbrace{B^2\,\beta(\mathcal{E},\gamma)}_{\text{Bias}^2},$$

$$\alpha(\mathcal{E},\Sigma,\gamma) \equiv \underbrace{\mathbb{V}_{\mathcal{E},\Sigma}\big[\,\tau(\hat{\theta}(\gamma))\,\big]}_{\text{Variance}} \qquad \beta(\mathcal{E},\gamma) \equiv \Bigg(\underbrace{\|\omega\|_1 - \|\omega_{\mathcal{E}}\|}_{\substack{\text{observational}\\ \text{bias contribution}}} + \underbrace{|1-\gamma|^\top|\omega_{\mathcal{E}}|_1}_{\substack{\text{Bias from}\\ \text{shrinkage}}}\Bigg)^2$$

Solution to the researcher's problem

- Standard decomposition:

$$\sup_{b \in \mathcal{B}(B)} \mathrm{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \underbrace{\alpha(\mathcal{E}, \Sigma, \gamma)}_{\text{Variance}} + \underbrace{B^2\, \beta(\mathcal{E}, \gamma)}_{\text{Bias}^2},$$

$$\alpha(\mathcal{E}, \Sigma, \gamma) \equiv \underbrace{\mathbb{V}_{\mathcal{E},\Sigma}\left[\, \tau(\hat{\theta}(\gamma))\, \right]}_{\text{Variance}} \qquad \beta(\mathcal{E}, \gamma) \equiv \Bigg(\underbrace{\|\omega\|_1 - \|\omega_{\mathcal{E}}\|}_{\substack{\text{observational}\\ \text{bias contribution}}} + \underbrace{|1-\gamma|^\top |\omega_{\mathcal{E}}|_1}_{\substack{\text{Bias from}\\ \text{shrinkage}}} \Bigg)^2$$

- Define $\alpha^\star$ and $\beta^\star$ their minima within the feasible set of $\mathcal{E}, \Sigma, \gamma$

Solution to the researcher's problem

- Standard decomposition:

$$\sup_{b \in \mathcal{B}(B)} \mathrm{MSE}_b(\mathcal{E}, \Sigma, \gamma) = \underbrace{\alpha(\mathcal{E}, \Sigma, \gamma)}_{\text{Variance}} + \underbrace{B^2 \, \beta(\mathcal{E}, \gamma)}_{\text{Bias}^2},$$

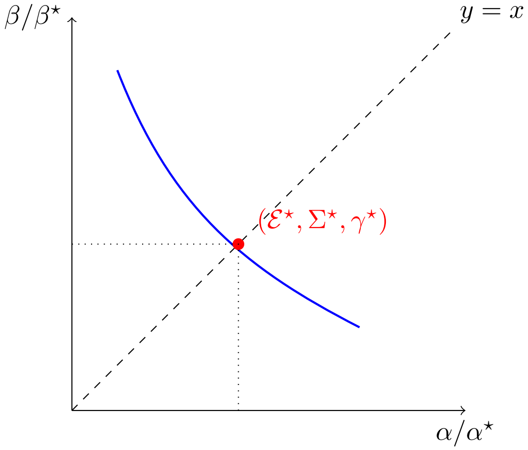
$$\alpha(\mathcal{E}, \Sigma, \gamma) \equiv \underbrace{\mathbb{V}_{\mathcal{E},\Sigma}\left[\tau(\hat{\theta}(\gamma))\right]}_{\text{Variance}} \qquad \beta(\mathcal{E}, \gamma) \equiv \Big(\underbrace{\|\omega\|_1 - \|\omega_{\mathcal{E}}\|}_{\substack{\text{observational}\\ \text{bias contribution}}} + \underbrace{|1-\gamma|^\top |\omega_{\mathcal{E}}|_1}_{\substack{\text{Bias from}\\ \text{shrinkage}}} \Big)^2$$

- Define $\alpha^\star$ and $\beta^\star$ their minima within the feasible set of $\mathcal{E}, \Sigma, \gamma$

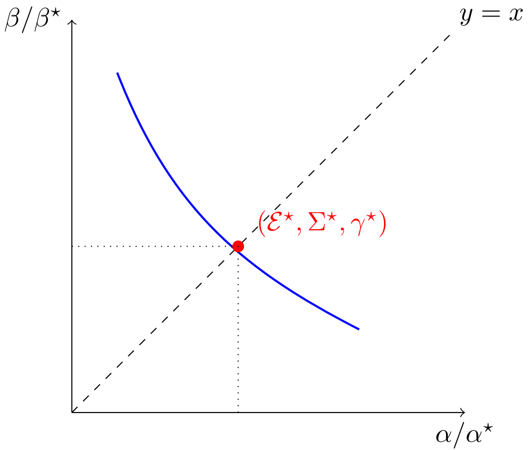
- **Theorem.** For any $(\mathcal{E}, \Sigma, \gamma) \in \mathcal{D}$,

$$\mathcal{R}(\mathcal{E}, \Sigma, \gamma) \ = \ \max\Bigg\{ \underbrace{\frac{\alpha(\mathcal{E}, \Sigma, \gamma)}{\alpha^\star}}_{\text{"Variance regret"}} , \underbrace{\frac{\beta(\mathcal{E}, \gamma)}{\beta^\star}}_{\text{"Bias regret"}} \Bigg\}.$$

Graphical intuition for interior solutions Example



Graphical intuition for interior solutions Example



- Key insight is to establish quasi-convexity: simple and tractable solution!
- It leverages and applies to class of (asymptotic) linear estimators

Receipt

- Pre-specify model(s) τ and compute ω through its derivative (Jacobian)
- For candidate experiment $\mathcal{E}$ input the variance as in power analyses.

Receipt

- Pre-specify model(s) τ and compute ω through its derivative (Jacobian)
- For candidate experiment $\mathcal{E}$ input the variance as in power analyses. Example

$$\Sigma = \begin{pmatrix} \Sigma_{\text{obs}} & 0_{p\times|\mathcal{E}|} \\ 0_{|\mathcal{E}|\times p} & \Sigma^{\text{exp}} \end{pmatrix}, \qquad \Sigma^{\text{exp}} = \text{diag}\left(\left\{ \frac{S_j^2}{n_j} \right\}_{j\in\mathcal{E}} \right)$$

where S_j^2 is the (expected) variance for treatment arm j.

Receipt

- Pre-specify model(s) τ and compute ω through its derivative (Jacobian)
- For candidate experiment $\mathcal{E}$ input the variance as in power analyses. Example

$$\Sigma = \begin{pmatrix} \Sigma_{\text{obs}} & 0_{p\times|\mathcal{E}|} \\ 0_{|\mathcal{E}|\times p} & \Sigma^{\text{exp}} \end{pmatrix}, \qquad \Sigma^{\text{exp}} = \text{diag}\left(\left\{\frac{S_j^2}{n_j}\right\}_{j\in\mathcal{E}}\right)$$

  where S_j^2 is the (expected) variance for treatment arm j.
- Trade-off sensitivity and variance via Theorem 1 to choose $\mathcal{E}$ and n_j

Receipt

- Pre-specify model(s) τ and compute ω through its derivative (Jacobian)
- For candidate experiment $\mathcal{E}$ input the variance as in power analyses.
  Example

$$\Sigma = \begin{pmatrix} \Sigma_{\text{obs}} & 0_{p\times|\mathcal{E}|} \\ 0_{|\mathcal{E}|\times p} & \Sigma^{\text{exp}} \end{pmatrix}, \qquad \Sigma^{\text{exp}} = \text{diag}\left(\left\{\frac{S_j^2}{n_j}\right\}_{j\in\mathcal{E}}\right)$$

  where S_j^2 is the (expected) variance for treatment arm j.
- Trade-off sensitivity and variance via Theorem 1 to choose $\mathcal{E}$ and n_j

$\Rightarrow$ Use mixed-integer quadratic program for optimization [R package soon]:
  - $n_j^\star$ minimizes variance conditional on experiment/estimator choice
  - convex optimization for estimator
  - mixed-integer variables for experiment choice

Good... when is this useful and when it is not?

- Our leading case is where some extrapolation occurs: $\beta^\star > 0$
  - No feasible experiment identifies $\tau(\theta)$ without external evidence
  - Proportional regret implies: bias-variance trade-off + tractable
  - By contrast, additive regret prioritizes bias lexicographically (and minimax MSE uninformative without restrictions on B)

Good... when is this useful and when it is not?

- Our leading case is where some extrapolation occurs: $\beta^\star > 0$
  - No feasible experiment identifies $\tau(\theta)$ without external evidence
  - Proportional regret implies: bias-variance trade-off + tractable
  - By contrast, additive regret prioritizes bias lexicographically (and minimax MSE uninformative without restrictions on B)

- Boundary case outside our leading applications: $\beta^\star = 0$
  - A feasible design can identify the full target $\tau(\theta)$ without bias
  - Solution over asymptotic linear estimators becomes lexicographic
  - Non-linear estimators useful, but studied for fixed design (Armstrong et al., 2024; de Chaisemartin and D'Haultfœuille, 2020)

Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer
Choose exp/
and precision

$\longrightarrow$

Report
$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}^{\text{exp}}_{\mathcal{E}})$
together with Σ

$\longrightarrow$

Audience
Use own beliefs
to estimate $\tau(\theta)$

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer		**Report**		**Audience**
Choose exp/ and precision	$\longrightarrow$	$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}^{\text{exp}}_{\mathcal{E}})$ together with Σ	$\longrightarrow$	Use own beliefs to estimate $\tau(\theta)$

- Audience prior $(\theta, b) \sim \pi$,

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer		**Report**		**Audience**
Choose exp/ and precision	$\longrightarrow$	$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}_{\mathcal{E}}^{\text{exp}})$ together with Σ	$\longrightarrow$	Use own beliefs to estimate $\tau(\theta)$

- Audience prior $(\theta, b) \sim \pi$, researcher reports $\tilde{\theta}_{\mathcal{E},\Sigma} | \theta, b \sim \mathcal{N}\left(\begin{bmatrix} b + \theta \\ \theta_{\mathcal{E}} \end{bmatrix}, \Sigma \right)$

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer	Report	Audience
Choose exp/ and precision	$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}^{\text{exp}}_{\mathcal{E}})$ together with Σ	Use own beliefs to estimate $\tau(\theta)$

- Audience prior $(\theta, b) \sim \pi$, researcher reports $\tilde{\theta}_{\mathcal{E},\Sigma} | \theta, b \sim \mathcal{N}\left(\begin{bmatrix} b+\theta \\ \theta_{\mathcal{E}} \end{bmatrix}, \Sigma\right)$

$\Rightarrow$ Expected audience loss depends on the chosen design:

$$\underbrace{\mathbb{E}_\pi[\tau(\theta)|\tilde{\theta}]}_{\text{Audience posterior}},$$

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer
Choose exp/
and precision

→

Report
$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}_{\mathcal{E}}^{\text{exp}})$
together with Σ

→

Audience
Use own beliefs
to estimate $\tau(\theta)$

- Audience prior $(\theta, b) \sim \pi$, researcher reports $\tilde{\theta}_{\mathcal{E},\Sigma} | \theta, b \sim \mathcal{N}\left(\begin{bmatrix} b+\theta \\ \theta_{\mathcal{E}} \end{bmatrix}, \Sigma\right)$

$\Rightarrow$ Expected audience loss depends on the chosen design:

$$\underbrace{\mathbb{E}_\pi[\tau(\theta)|\tilde{\theta}]}_{\text{Audience posterior}} \quad , \qquad \underbrace{L_\pi(\mathcal{E}, \Sigma) = \mathbb{E}_\pi\left[\left(\tau(\theta) - \delta_\pi(\tilde{\theta}_{\mathcal{E},\Sigma})\right)^2\right]}_{\text{Audience loss}},$$

Researchers may report the full evidence

- So far: choose $(\mathcal{E}, \Sigma)$ and report estimator
- Next: choose $(\mathcal{E}, \Sigma)$, report the full evidence and let the audience update

Designer
Choose exp/
and precision

$\longrightarrow$

Report
$\tilde{\theta}_{\mathcal{E},\Sigma} = (\tilde{\theta}^{\text{obs}}, \tilde{\theta}_{\mathcal{E}}^{\text{exp}})$
together with Σ

$\longrightarrow$

Audience
Use own beliefs
to estimate $\tau(\theta)$

- Audience prior $(\theta, b) \sim \pi$, researcher reports $\tilde{\theta}_{\mathcal{E},\Sigma} | \theta, b \sim \mathcal{N}\left(\begin{bmatrix} b + \theta \\ \theta_{\mathcal{E}} \end{bmatrix}, \Sigma \right)$

$\Rightarrow$ Expected audience loss depends on the chosen design:

$$\underbrace{\mathbb{E}_\pi[\tau(\theta)|\tilde{\theta}]}_{\text{Audience posterior}} \quad , \qquad \underbrace{L_\pi(\mathcal{E}, \Sigma) = \mathbb{E}_\pi\left[\Big(\tau(\theta) - \delta_\pi(\tilde{\theta}_{\mathcal{E},\Sigma}) \Big)^2 \right]}_{\text{Audience loss}},$$

$\Rightarrow$ Problem is... π unknown ex-ante

Regret retains the same variance/sensitivity components

- For each $B \geq 0$ consider priors $\pi \in \Pi_B$ satisfying for every j

$$\mathbb{E}_\pi[b_j^2|\theta] \leq B^2 \qquad (\Rightarrow \text{similar interpretation of } B)$$

Regret retains the same variance/sensitivity components

- For each $B \geq 0$ consider priors $\pi \in \Pi_B$ satisfying for every j

$$\mathbb{E}_\pi[b_j^2|\theta] \leq B^2 \qquad (\Rightarrow \text{similar interpretation of } B)$$

- **Prop.** Under the audience model, the loss for given B is

$$\bar{L}_B(\mathcal{E},\Sigma) \equiv \sup_{\pi \in \Pi_B} L_\pi(\mathcal{E},\Sigma) = \min_\gamma \left\{ \underbrace{\alpha(\mathcal{E},\Sigma,\gamma)}_{\text{Same variance!}} + B^2 \underbrace{\beta(\mathcal{E},\gamma)}_{\text{Same sensitivity!}} \right\}.$$

Regret retains the same variance/sensitivity components

- For each $B \geq 0$ consider priors $\pi \in \Pi_B$ satisfying for every j

$$\mathbb{E}_\pi[b_j^2|\theta] \leq B^2 \qquad (\Rightarrow \text{similar interpretation of } B)$$

- **Prop.** Under the audience model, the loss for given B is

$$\bar{L}_B(\mathcal{E}, \Sigma) \equiv \sup_{\pi \in \Pi_B} L_\pi(\mathcal{E}, \Sigma) = \min_\gamma \left\{ \underbrace{\alpha(\mathcal{E}, \Sigma, \gamma)}_{\text{Same variance!}} + B^2 \underbrace{\beta(\mathcal{E}, \gamma)}_{\text{Same sensitivity!}} \right\}.$$

$\Rightarrow$ **Proportional regret**: convex given $\mathcal{E}$, running over grid for B

$$\mathcal{R}^{\text{aud}}(\mathcal{E}, \Sigma) = \sup_{B \geq 0} \left\{ \bar{L}_B(\mathcal{E}, \Sigma) \Big/ \bar{L}_B(\mathcal{E}_B^\star, \Sigma_B^\star) \right\}.$$

$\Rightarrow$ Intuitive and feasible to optimize!

Regret retains the same variance/sensitivity components

- For each $B \geq 0$ consider priors $\pi \in \Pi_B$ satisfying for every j

$$\mathbb{E}_\pi[b_j^2|\theta] \leq B^2 \qquad (\Rightarrow \text{similar interpretation of } B)$$

- **Prop.** Under the audience model, the loss for given B is

$$\bar{L}_B(\mathcal{E},\Sigma) \equiv \sup_{\pi\in\Pi_B} L_\pi(\mathcal{E},\Sigma) = \min_\gamma \left\{ \underbrace{\alpha(\mathcal{E},\Sigma,\gamma)}_{\text{Same variance!}} + B^2 \underbrace{\beta(\mathcal{E},\gamma)}_{\text{Same sensitivity!}} \right\}.$$

$\Rightarrow$ **Proportional regret**: convex given $\mathcal{E}$, running over grid for B

$$\mathcal{R}^{\text{aud}}(\mathcal{E},\Sigma) = \sup_{B\geq 0} \left\{ \bar{L}_B(\mathcal{E},\Sigma) \Big/ \bar{L}_B(\mathcal{E}_B^\star,\Sigma_B^\star) \right\}.$$

$\Rightarrow$ Intuitive and feasible to optimize!

$\Rightarrow$ **Theorem** (informal): For a single but arbitrary estimator, audience regret is *lower bound* and linear regret is a *upper bound* for its regret

Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Extensions summary

- Observational and experimental moment implied conditions

Extensions summary

- Observational and experimental moment implied conditions
  - $\Rightarrow$ Results directly extend to this setting: Proportional regret is

$$\max\{\alpha/\alpha^\star, \beta/\beta^\star\}$$

where α and β depend on the weighting matrix and moment Jacobian

Extensions summary

- Observational and experimental moment implied conditions

  $\Rightarrow$ Results directly extend to this setting: Proportional regret is

$$\max\{\alpha/\alpha^\star, \beta/\beta^\star\}$$

  where α and β depend on the weighting matrix and moment Jacobian

- Vector-valued τ
- Non linear $\tau(\theta)$ and non linear moments: local asymptotics
- Arbitrary norms/weighted prior set on biases
- Minimizing confidence interval length
- Known upper bound $B \leq \bar{B}$

Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Application: school stipends in Kenya

- Q: GE effect on school attendance of school stipend in Kenya?
- Only small-scale experiments feasible; the target includes GE feedback through income and wages (Attanasio et al., 2012; Todd and Wolpin, 2006)

Application: school stipends in Kenya

- Q: GE effect on school attendance of school stipend in Kenya?
- Only small-scale experiments feasible; the target includes GE feedback through income and wages (Attanasio et al., 2012; Todd and Wolpin, 2006)
- A school-choice model and labor-market clearing decompose the target:

$$\tau(\theta) \approx \underbrace{\theta_1}_{\substack{\text{Income}\\ \text{effect}}} \cdot \underbrace{\theta_M}_{\substack{\text{Income}\\ \text{multiplier}}} + \underbrace{\theta_2}_{\substack{\text{Direct}\\ \text{stipend effect}}} + \underbrace{\theta_3}_{\substack{\text{Wage}\\ \text{effect}}} \cdot \underbrace{w_0(\theta)}_{\substack{\text{Wage}\\ \text{multiplier}}}$$

  where $w_0(\theta)$ is the wage response implied by market clearing

Application: school stipends in Kenya

- Q: GE effect on school attendance of school stipend in Kenya?
- Only small-scale experiments feasible; the target includes GE feedback through income and wages (Attanasio et al., 2012; Todd and Wolpin, 2006)
- A school-choice model and labor-market clearing decompose the target:

$$\tau(\theta) \approx \underbrace{\theta_1}_{\text{Income effect}} \cdot \underbrace{\theta_M}_{\text{Income multiplier}} + \underbrace{\theta_2}_{\text{Direct stipend effect}} + \underbrace{\theta_3}_{\text{Wage effect}} \cdot \underbrace{w_0(\theta)}_{\text{Wage multiplier}}$$

  where $w_0(\theta)$ is the wage response implied by market clearing

- Existing evidence:
  - Mexico PROGRESA estimates $\theta_{1:3}$; possible external validity bias
  - Kenyan estimate θ_M and the labor-demand elasticity in $w_0(\theta)$
  - Important feature: researchers ex-ante describe potential biases for extrapolation. Here consider bias only for estimates from Mexico.

An illustration with three potential experiments

Feasible small scale experiments	θ_j	Interpretation	ω_j	Σ_{jj}^{obs}
Unconditional transfer (UCT)	θ_1	Direct income effect	0.25	$4.9/10^8$
Conditional transfer (CCT)	θ_2	Direct stipend effect	0.11	$1.11/10^5$
Randomized job earnings (Wage)	θ_3	Direct wage effect	0.11	$1.18/10^5$

- Feasible set: one cash-transfer arm, both transfer arms, or wage exp

An illustration with three potential experiments

Feasible small scale experiments	θ_j	Interpretation	ω_j	Σ_{jj}^{obs}
Unconditional transfer (UCT)	θ_1	Direct income effect	0.25	$4.9/10^8$
Conditional transfer (CCT)	θ_2	Direct stipend effect	0.11	$1.11/10^5$
Randomized job earnings (Wage)	θ_3	Direct wage effect	0.11	$1.18/10^5$

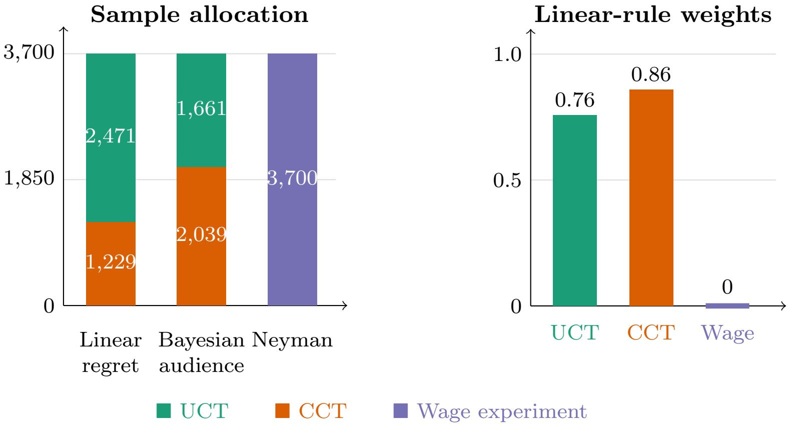
- Feasible set: one cash-transfer arm, both transfer arms, or wage exp
- For each experiment j, researchers construct
  - $\tilde{\theta}_j^{\text{exp}}$: mean difference betw treated/controls in arm j per \$ spent,
  - Outcome: school attendance.

An illustration with three potential experiments

Feasible small scale experiments	θ_j	Interpretation	ω_j	Σ_{jj}^{obs}
Unconditional transfer (UCT)	θ_1	Direct income effect	0.25	$4.9/10^8$
Conditional transfer (CCT)	θ_2	Direct stipend effect	0.11	$1.11/10^5$
Randomized job earnings (Wage)	θ_3	Direct wage effect	0.11	$1.18/10^5$

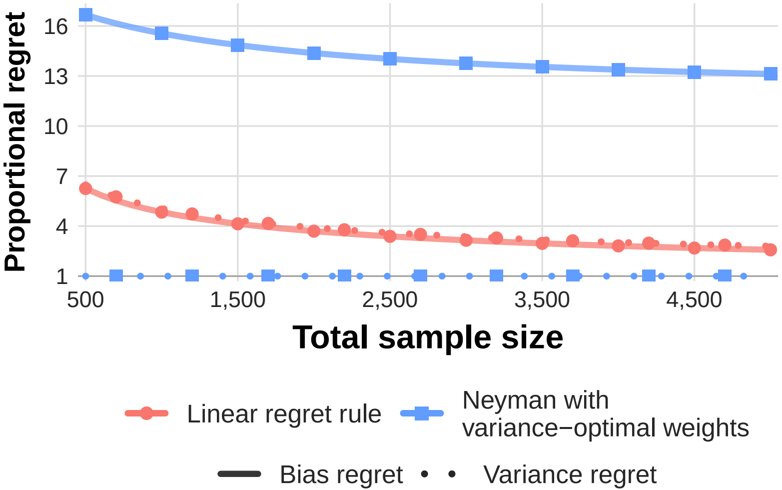
- Feasible set: one cash-transfer arm, both transfer arms, or wage exp
- For each experiment j, researchers construct
  - $\tilde{\theta}_j^{\text{exp}}$: mean difference betw treated/controls in arm j per $ spent,
  - Outcome: school attendance.
- Variance calibration: use Kenyan school attendance with homoskedastic treated and controls

Robust criteria choose both cash-transfer arms



- $n = 3700$, average JPAL sample size
- Results are robust for other sample sizes

Variance minimization vs robust design: regret



Outline

1. Setup: baseline case
2. Optimal design for common estimators
3. General reporting rules
4. Extensions
5. Illustration
6. Conclusion

Conclusion: receipt for practice

- This paper studies experimental design with observational data
- Experiments "complement" previous estimates for extrapolation

Framework for the design of the experiment that proceeds as follows:

- Think ahead about extrapolation exercise and sources of (local) biases
- Compute sensitivities and specify Σ (as for power analyses)
- Balance bias/variance regret via simple optimization

Conclusion: receipt for practice

- This paper studies experimental design with observational data
- Experiments "complement" previous estimates for extrapolation

Framework for the design of the experiment that proceeds as follows:

- Think ahead about extrapolation exercise and sources of (local) biases
- Compute sensitivities and specify Σ (as for power analyses)
- Balance bias/variance regret via simple optimization

Applicability include

- Field experiments in development
- Industrial organization, labor economics, public economics, ...

$\Rightarrow$ Part of agenda at intersection of development and metrics

Thank you!

References I

Abadie, A. and J. Zhao (2021). Synthetic controls for experimental design. *arXiv preprint arXiv:2108.02196*.

Andrews, I., N. Barahona, M. Gentzkow, A. Rambachan, and J. M. Shapiro (2025). Structural estimation under misspecification: Theory and implications for practice. *The Quarterly Journal of Economics*, qjaf018.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2020). Transparency in structural research. *Journal of Business & Economic Statistics 38*(4), 711–722.

Andrews, I. and J. M. Shapiro (2021). A model of scientific communication. *Econometrica 89*(5), 2117–2142.

Armstrong, T. B., P. Kline, and L. Sun (2024). Adapting to misspecification.

Armstrong, T. B. and M. Kolesár (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics 12*(1), 77–108.

Athey, S., R. Chetty, and G. Imbens (2020). Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint arXiv:2006.09676*.

Athey, S., R. Chetty, and G. Imbens (2025). Using experiments to correct for selection in observational studies.

Atkinson, A. C. and V. Fedorov (1975). The design of experiments for discriminating between two rival models. *Biometrika 62*(1), 57–70.

References II

Attanasio, O. P., C. Meghir, and A. Santiago (2012). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. *The Review of Economic Studies 79*(1), 37–66.

Bai, Y. (2019). Optimality of matched-pair designs in randomized controlled trials. *Available at SSRN 3483834*.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020). A theory of experimenters: Robustness, randomization, and balance. *American Economic Review 110*(4), 1206–1230.

Bickel, P. J. (1984). Parametric robustness: small biases can be worthwhile. *The Annals of Statistics 12*(3), 864–879.

Bonhomme, S. and M. Weidner (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics 13*(3), 907–954.

Box, G. E. and N. R. Draper (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association 54*(287), 622–654.

Breza, E., A. G. Chandrasekhar, and D. Viviano (2025). Generalizability with ignorance in mind: learning what we do (not) know for archetypes discovery. *arXiv preprint arXiv:2501.13355*.

Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical science*, 273–304.

References III

Cytrynbaum, M. (2021). Optimal stratification of survey experiments. *arXiv preprint arXiv:2111.08157*.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Empirical mse minimization to estimate a scalar parameter. *arXiv preprint arXiv:2006.14667*.

Dette, H. (1997). Designing experiments with respect to 'standardized'optimality criteria. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*(1), 97–110.

Dominitz, J. and C. F. Manski (2017). More data or better data? a statistical decision problem. *The Review of Economic Studies 84*(4), 1583–1605.

Gechter, M., K. Hirano, J. Lee, M. Mahmud, O. Mondal, J. Morduch, S. Ravindran, and A. S. Shonchoy (2024). Selecting experimental sites for external validity. *arXiv preprint arXiv:2405.13241*.

Higbee, S. D. (2024). Experimental design for policy choice.

Hu, Y., H. Zhu, E. Brunskil, and S. Wager (2024). Minimax-regret sample selection in randomized experiments. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pp. 1209–1235.

Kempthorne, P. J. (1988). Controlling risks under different loss functions: The compromise decision problem. *The Annals of Statistics 16*(4), 1594–1608.

Kiefer, J. and J. Wolfowitz (1959). Optimum designs in regression problems. *The annals of mathematical statistics 30*(2), 271–294.

References IV

Manski, C. F. and A. Tetenov (2016). Sufficient trial size to inform clinical practice. *Proceedings of the National Academy of Sciences 113*(38), 10518–10523.

Olea, J. L. M., B. Prallon, C. Qiu, J. Stoye, and Y. Sun (2024). Externally valid selection of experimental sites via the k-median problem. *arXiv preprint arXiv:2408.09187*.

Rosenman, E. T. and A. B. Owen (2021). Designing experiments informed by observational studies. *Journal of Causal Inference 9*(1), 147–171.

Todd, P. E. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review 96*(5), 1384–1417.

Tsirpitzi, R. E., F. Miller, and C.-F. Burman (2023). Robust optimal designs using a model misspecification term. *Metrika 86*(7), 781–804.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the sobolev classes. *The Annals of Statistics 26*(6), 2420–2469.

Viviano, D., K. Wüthrich, and P. Niehaus (2026). A model of multiple hypothesis testing. *Review of Economic Studies*, rdag026.

Linear design remains useful for nonlinear reporting

- Third case: researchers report a *single, arbitrary* estimator δ, with loss

$$\mathbb{E}_\pi\left[\left(\tau(\theta) - \delta(\tilde{\theta}_{\mathcal{E},\Sigma})\right)^2\right], \qquad (\theta, b) \sim \pi.$$

$\Rightarrow$ Special case: asymptotically linear estimators considered before

$\Rightarrow$ Let $\mathcal{R}^{\text{single}}(\mathcal{E}, \Sigma)$ be the worst-case proportional regret over all δ

Linear design remains useful for nonlinear reporting

- Third case: researchers report a *single, arbitrary* estimator δ, with loss

$$\mathbb{E}_\pi\left[\left(\tau(\theta) - \delta(\tilde{\theta}_{\mathcal{E},\Sigma})\right)^2\right], \qquad (\theta, b) \sim \pi.$$

  $\Rightarrow$ Special case: asymptotically linear estimators considered before

  $\Rightarrow$ Let $\mathcal{R}^{\text{single}}(\mathcal{E}, \Sigma)$ be the worst-case proportional regret over all δ

- **Theorem** Under the same class of priors $\pi \in \Pi_B$ as before

$$\underbrace{\mathcal{R}^{\text{aud}}(\mathcal{E}, \Sigma)}_{\text{Audience regret}} \leq \underbrace{\mathcal{R}^{\text{single}}(\mathcal{E}, \Sigma)}_{\text{arbitrary single estimator}} \leq \underbrace{\min_\gamma \max\left\{\frac{\alpha(\mathcal{E}, \Sigma, \gamma)}{\alpha^\star}, \frac{\beta(\mathcal{E}, \gamma)}{\beta^\star}\right\}}_{\text{linear estimator regret}}.$$

Linear design remains useful for nonlinear reporting

- Third case: researchers report a *single, arbitrary* estimator δ, with loss

$$\mathbb{E}_\pi\left[\left(\tau(\theta)-\delta(\tilde{\theta}_{\mathcal{E},\Sigma})\right)^2\right], \qquad (\theta,b)\sim\pi.$$

  $\Rightarrow$ Special case: asymptotically linear estimators considered before

  $\Rightarrow$ Let $\mathcal{R}^{\text{single}}(\mathcal{E},\Sigma)$ be the worst-case proportional regret over all δ

- **Theorem** Under the same class of priors $\pi \in \Pi_B$ as before

$$\underbrace{\mathcal{R}^{\text{aud}}(\mathcal{E},\Sigma)}_{\text{Audience regret}} \le \underbrace{\mathcal{R}^{\text{single}}(\mathcal{E},\Sigma)}_{\text{arbitrary single estimator}} \le \underbrace{\min_{\gamma}\max\left\{\frac{\alpha(\mathcal{E},\Sigma,\gamma)}{\alpha^\star},\frac{\beta(\mathcal{E},\gamma)}{\beta^\star}\right\}}_{\text{linear estimator regret}}.$$

  - Exact nonlinear design loses quasi-convexity and simple expression

  $\Rightarrow$ linear criterion remains useful for design as a surrogate objective!

Bayesian regret for given bias B ($n = 3700$)