Policy Learning with Unobserved Heterogeneity

Giacomo Opocher

University of Bologna

September 13, 2025

General Problem

- Map resources into welfare through policies (e.g. transfers to foster development).

$\rightarrow$ **Pr1:** Evaluate a given policy's effects (e.g. effects on profits).

$\rightarrow$ **Pr2:** Decide who to treat in the full population (e.g. Indian entrepreneurs).

- **Decide who to treat**:

$\rightarrow$ Easy if the policy's effects are constant.

$\rightarrow$ Hard if heterogeneous.

Problem Description

- Policies' effects vary between individuals.

$\rightarrow$ The same policy can help some and harm others.

e.g. Alt, Lassen, and Marshall (2016), Hussam, Rigol, and Roth (2022), Biroli et al. (2025), Brynjolfsson, Li, and Raymond (2025).

- Policymakers test policies before implementing them at scale.

$\rightarrow$ Treat only those expected to benefit.

- Econometric approach: Policy Learning.

$\rightarrow$ Use RCTs to learn assignment rules that *perform well* in the population.

e.g. Kitagawa and Tetenov (2018), Mbakop and Tabord-Meehan (2021), Athey and Wager (2021).

- State of the art: all relevant dimensions are observed.

$\rightarrow$ **When** and **how** to account for **unobserved heterogeneity** in policy learning?

Stylized Example - Setting Hussam, Rigol, and Roth, 2022 (AER)

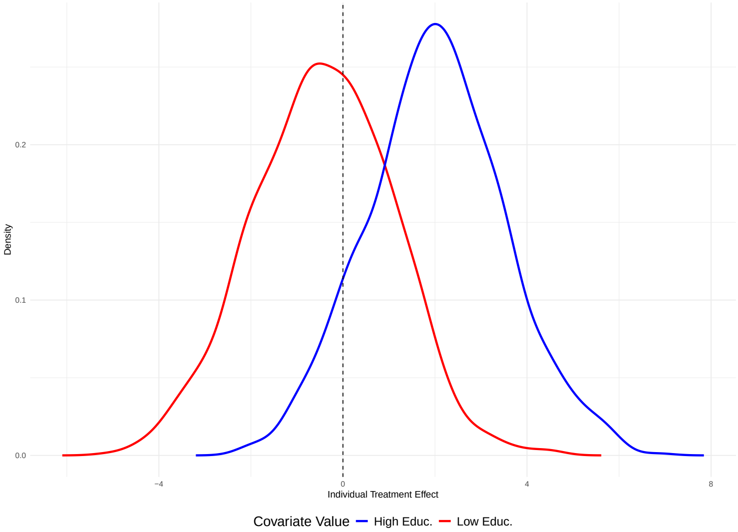
- Binary Policy: $D_i \in \{0, 1\}$ (e.g. cash transfer to micro-entrepreneurs).
- Random Sample $\mathcal{S}$ from a population of interest $\mathcal{P}$ (e.g. Indian entrepreneurs).
- RCT to evaluate the effects on Y_i (e.g. profits).
- For simplicity, $X_i \in \{h, l\}$: $\tau(l) < 0 < \tau(h)$ (e.g. high/low education).

$\rightarrow$ Who to treat in $\mathcal{P}$? Covariate-Based Policy Rule: $G_x = \mathbb{1}(X_i = h)$.

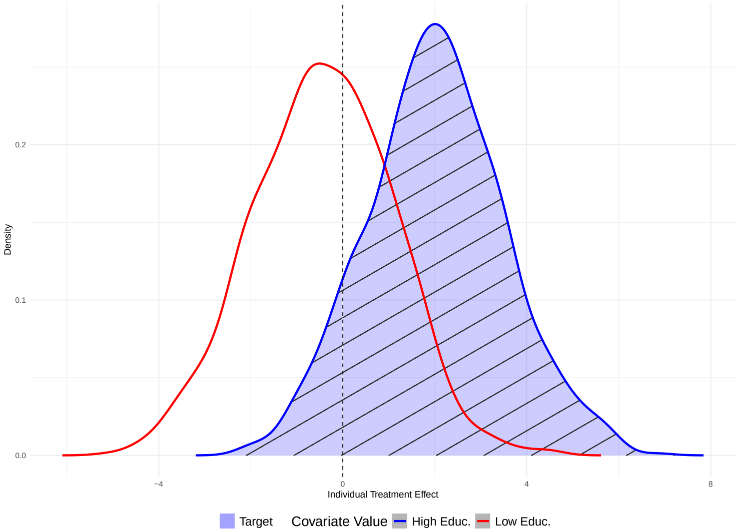
- Assume now observed, and **unobserved** heterogeneity: $\tau(X_i, A_i)$.
- For simplicity, $A_i \in \mathbb{R}$ (e.g. business skills).
- Denote with $\tau(h), \tau(l)$ the avg. effect for $X = h, l$: $\tau(l) < 0 < \tau(h)$.

$\rightarrow$ Is G_x still *optimal*?

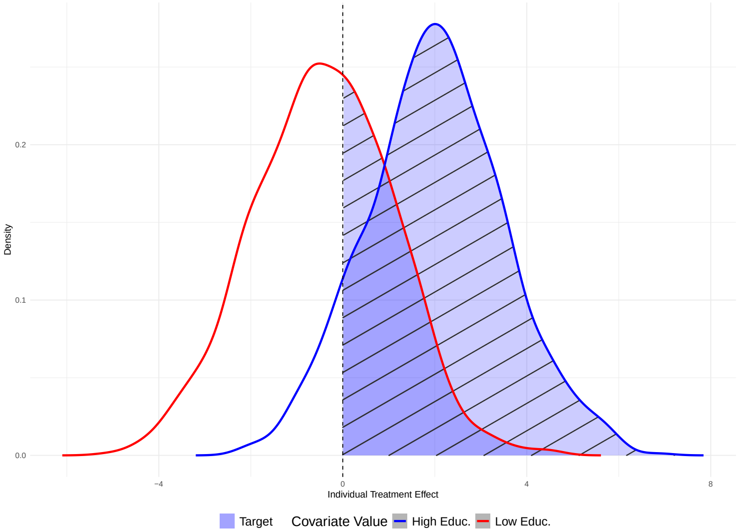
Stylized Example - Individual Treatment Effects



Stylized Example - Covariate-Based Policy Rule



Stylized Example - Oracle Policy Rule



Research Question

- **Problem 1**: we only observe $\mathcal{S}$, and we do not know counterfactuals.

$\rightarrow$ **Solution**: Empirical Welfare Maximization (Kitagawa and Tetenov, 2018).

- **Problem 2**: we don't observe the realizations of A_i: α_i (e.g. business skills).

$\rightarrow$ **Potential solution**: use *estimates* or *proxies*: $\hat{\alpha}_i$ (e.g. fixed effects, factors, principal components; satellite data, survey questions, ...).

Can ***estimated*** *latent characteristics improve policy recommendations when treatment effects vary along their* ***true*** *values?*

This Paper in One Slide

*Can **estimated** latent characteristics improve policy recommendations when treatment effects vary along their **true** values?*

- **It depends:** trade-off importance of α vs. estimation error in $\hat{\alpha}$.

→ **Regret bounds** on welfare for policy rules that include $\hat{\alpha}$ or not.

→ **Connection** with shrinkage: new **regret bounds** for shrinkage-based rules.

- **New class of rules**: set $\hat{\alpha}$'s **importance** depending on the welfare gain produced by including it in the decision rule.

→ Data-driven method to set $\hat{\alpha}$'s **importance** via cross-validation.

→ Adaptively achieves near-optimal welfare.

- Empirical application in development economics Hussam et al., 2022 (AER).

→ Including proxies **halves** the probability of producing welfare losses.

Related Literature

- **Policy Learning**.

  e.g. Manski (2004), Bhattacharya and Dupas (2012), Kitagawa and Tetenov (2018), Kitagawa and Tetenov (2021), Mbakop and Tabord-Meehan (2021), Athey and Wager (2021), Viviano and Bradic (2024), Viviano (2024).

$\rightarrow$ Unobserved heterogeneity introduces a new approximation-estimation error trade-off.

$\rightarrow$ Connection with shrinkage (Yamin, 2025; Moon, 2025).

$\rightarrow$ Data-driven procedure can solve this trade-off.

- **Applied Microeconomics** (development, education, political economy, labor).

  e.g. Leuven, Oosterbeek, and Klaauw (2010), Alt, Lassen, and Marshall (2016), Hussam, Rigol, and Roth (2022), Bryan, Karlan, and Osman (2024), Biroli et al. (2025), Brynjolfsson, Li, and Raymond (2025).

$\rightarrow$ How to scale up interventions when treatment effects vary between individuals.

$\rightarrow$ We can evaluate policy recommendations before recommending them!

Roadmap

Potential Outcomes

Consider:

$$(Y_i(0), X_i, A_i) \sim P_{y(0),x,\alpha}, \quad D_i \sim \mathcal{B}(1, e(Z_i))$$

that takes values $(y_i(0), x_i, \alpha_i) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{A}$ and $e(Z_i) = p$.

Potential Outcomes:

$$Y_i(0), \quad Y_i(1) = Y_i(0) + \tau(X_i, A_i)$$

Observable Data:

$$(Y_i, X_i, D_i), \quad Y_i = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0)$$

Policy Rules and Classes

Policy Rule: A mapping from variables ($\mathcal{Z} = \mathcal{X}$ or $\mathcal{Z} = \mathcal{X} \times \mathcal{A}$) to target set $\{0, 1\}$:

$$G_z : \mathcal{Z} \to \{0, 1\}$$

Classes of rules:

$$\underbrace{\mathcal{G}_x = \{G_x : \mathcal{X} \to \{0,1\}\}}_{\text{Covariate-Based (CB)}}, \quad \underbrace{\mathcal{G}_{x,\alpha} = \{G_{x,\alpha} : \mathcal{X} \times \mathcal{A} \to \{0,1\}\}}_{\alpha\text{-Augmented } (\alpha\text{-CB})}$$

Example: Grant Allocation.

- G_x: Assign by age and education.
- $G_{x,\alpha}$: Also include business skills α.

Feasible α-Augmented Rules

The realized value of A_i, α_i is not observed, but it can be estimated with $\hat{\alpha}_i$.

$$\hat{\alpha}_i = \alpha_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

Feasible α-Augmented rules:

$$\underbrace{\mathcal{G}_{x,\hat{\alpha}} = \{G_{x,\hat{\alpha}} : \mathcal{X} \times \mathcal{A} \to \{0,1\}\}}_{\hat{\alpha}\text{-Augmented } (\hat{\alpha}\text{-CB})}$$

Example: (Linear) Threshold Rules

Threshold rules:

- CB rules:

$$\mathcal{G}_x = \{G_x = \mathbb{1}(x > t)\}$$

- α-CB rules:

$$\mathcal{G}_{x,\alpha} = \{G_{x,\alpha} = \mathbb{1}(x + \alpha > t)\}$$

- $\hat{\alpha}$-CB rules:

$$\mathcal{G}_{x,\hat{\alpha}} = \{G_{x,\hat{\alpha}} = \mathbb{1}(x + \hat{\alpha} > t)\}$$

Welfare and Policy Learning

Welfare generated by a (general) Policy G_z:

$$W(G_z) := \frac{1}{n}\sum_{i=1}^{n} [Y_i(1) \cdot 1(i \in G_z) + Y_i(0) \cdot 1(i \notin G_z)]$$

Oracle Rule: Find G_z that maximizes expected welfare:

$$G_z^* := \arg\max_{G_z \in \mathcal{G}_z} E_{P^n}[W(G_z)]$$

Challenge: G_z^* depends on counterfactuals and solves a population-wide problem. Need for a feasible empirical analogue.

Empirical Welfare Maximization (EWM) - Kitagawa and Tetenov (2018)

EWM Rule:

$$\hat{G}_z := \arg\max_{G_z \in \mathcal{G}_z} \{W_n(G_z)\}$$

with

$$W_n(G_z) := \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Y_i D_i}{e(Z_i)} \cdot 1(i \in G_z) + \frac{Y_i(1-D_i)}{1-e(Z_i)} \cdot 1(i \notin G_z)\right]$$

Regret:

$$R(\hat{G}_z) := E_{P^n}[W(G_z^*) - W(\hat{G}_z)]$$

Measures average welfare loss from using $\hat{G}_z$ *instead of* G_z^*.

Roadmap

Assumption 1 - Kitagawa and Tetenov (2018) (1/2)

i. Bounded Outcomes

$$|Y_i| \leq M/2$$

Potential outcomes are uniformly bounded by a constant.

ii. Clean Design (Unconfoundedness + SUTVA)

$$D_i \perp (Y_i(0), Y_i(1)) | (X_i, A_i) \quad ; \quad Y_i(D_i, \mathbf{D}_{-i}) = Y_i(D_i)$$

Treatment assignment is as good as random; no spillovers.

Assumption 1 - Kitagawa and Tetenov (2018) (2/2)

iii. Strict Overlap

$$\Pr(D_i = 1 | X_i, A_i) \in [k, 1-k] \text{ for some } k > 0$$

All units have a positive chance of receiving treatment.

iv. Finite VC-Dimension

$$\text{VC}(\mathcal{G}_z) = v_z < \infty$$

The policy class has finite complexity.

Assumption 2 - Novel

i. Proxy Representation

$\hat{\alpha}_i$ can be written as $\hat{\alpha}_i = \alpha_i + \varepsilon_i$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

ii. Lipschitz Treatment Effects

$$|\tau(x, \alpha + \gamma) - \tau(x, \alpha)| \leq L \cdot |\gamma|, \quad L \in \mathbb{R}^+$$

Small changes in α lead to smooth changes in τ.

Regret Bound for Covariate-Based Rules

Theorem 1: Under Assumptions 1–2, regret of Covariate-Based policy rules satisfies:

$$R(\hat{G}_x) := E_{P^n}[W(G^*_{x,\alpha}) - W(\hat{G}_x)] \leq C_1 \frac{M}{k}\sqrt{\frac{v_x}{n}} + \sigma_{\tau|x}$$

where C_1 is a universal constant and $\sigma_{\tau|x} = E_X\left[\sqrt{\mathrm{Var}(\tau(X_i, A_i)|X_i)}\right]$

- **Two terms:**
  1. $C_1 \frac{M}{k}\sqrt{\frac{v_x}{n}}$: regret due to empirical analogue.
  2. $\sigma_{\tau|x}$: regret due to ignoring α.

Regret Bound for $\hat{\alpha}$-Augmented Rules

Theorem 2: Under Assumptions 1–2, regret of $\hat{\alpha}$-Augmented policy rules satisfies:

$$R(\hat{G}_{x,\hat{\alpha}}) := E_{P^n}[W(G^*_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}})] \leq 2C_1 \frac{M}{k}\sqrt{\frac{v_{x,\hat{\alpha}}}{n}} + c_1\sigma_\varepsilon$$

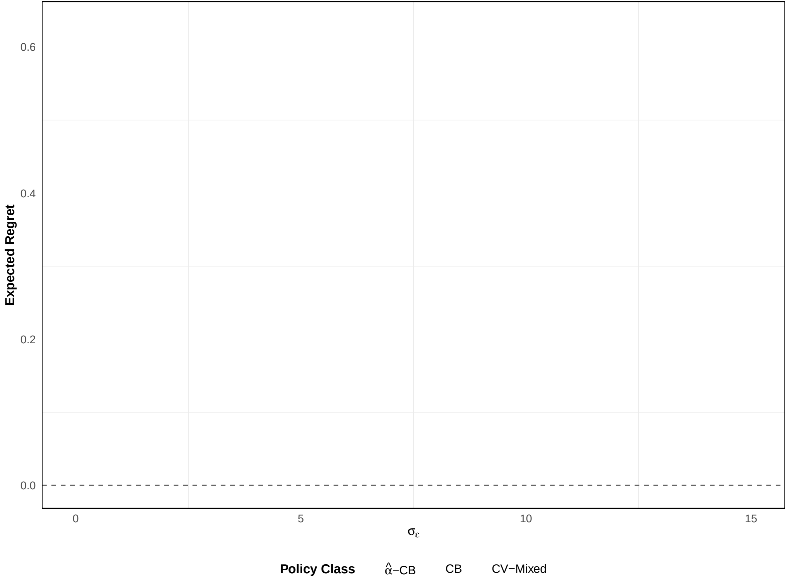
where C_1 is a universal constant and $c_1 := 4L\sqrt{2/\pi}$.

- **Two terms:**
  1. $2C_1 \frac{M}{k}\sqrt{\frac{v_{x,\hat{\alpha}}}{n}}$: regret due to empirical analogue.
  2. $c_1\sigma_\varepsilon$: regret due to noisy estimates of α.

- **Key insight:** trade-off between estimation (due to noise in $\hat{\alpha}$) and approximation (due to importance of α) error.

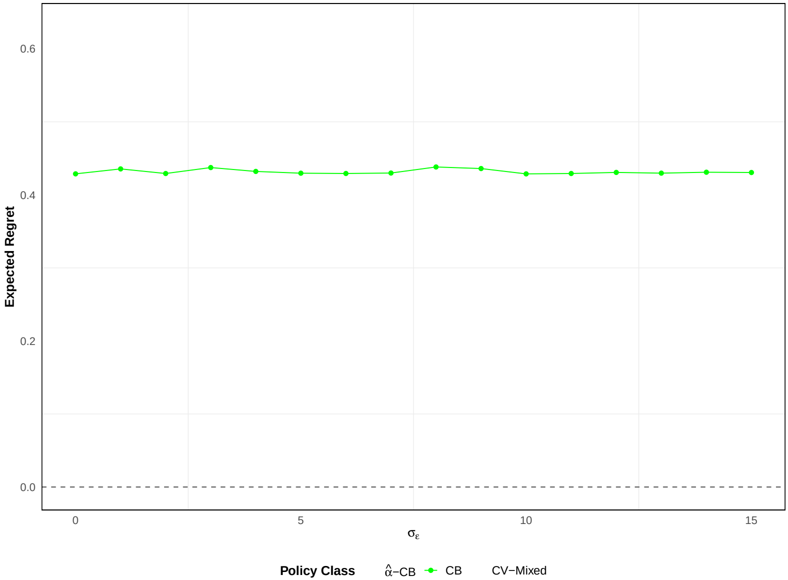
Simulations Results - Regret

Simulations DGP



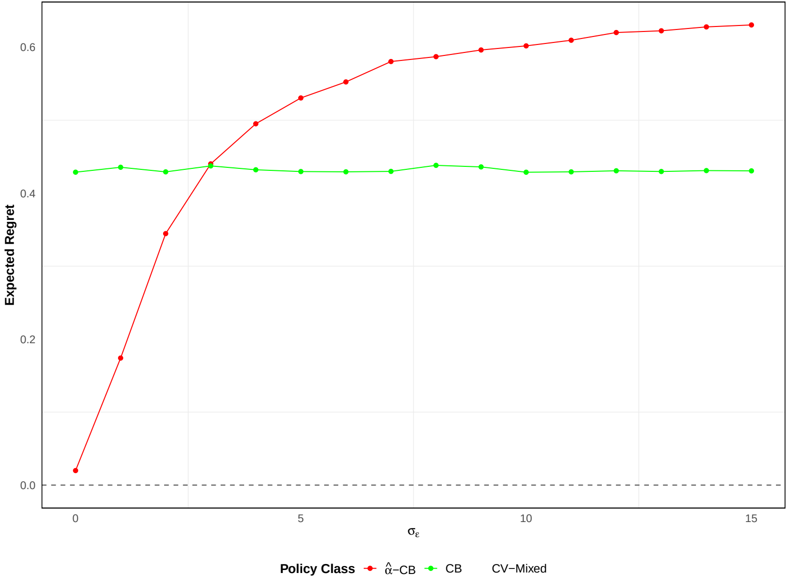
Simulations Results - Regret CB Rules

Simulations DGP



Simulations Results - Regret $\hat{\alpha}$-CB Rules

Simulations DGP　Back to Empirics



Motivation for Shrinkage Steps

Formal motivation:

$$
\begin{aligned}
R(\hat{G}_{x,\hat{\alpha}}) &= E_{P^n}[W(G_z^*) - W(\hat{G}_{x,\hat{\alpha}})] = \\
&= E_{P^n}[W(G_z^*) - W(\hat{G}_{x,\alpha})] + E_{P^n}[W(\hat{G}_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}})] = \\
&= \underbrace{E_{P^n}[W(G_{x,\alpha}^*) - W(\hat{G}_{x,\alpha})]}_{A} + \underbrace{E_{P^n}[W(\hat{G}_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}})]}_{B}
\end{aligned}
$$

And:

$$
B \leq KT18 + 4L \cdot \sqrt{2/\pi} \cdot \sqrt{E_{P^n}\left[\frac{1}{n}\sum_{i \in \mathcal{S}^n}(A_i - \hat{\alpha}_i)^2\right]}
$$

Empirical Bayes - $\hat{\alpha}^{EB}$-Augmented Rules

Assumption 3
The unobserved factor A_i is normally distributed: $A_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$.

Define:

$$\hat{\alpha}_i^{EB} = \omega \hat{\alpha}_i + (1 - \omega)\bar{\hat{\alpha}}$$

where $\omega = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + \sigma_\alpha^2}$.

The class of $\hat{\alpha}^{EB}$-Augmented Rules is then defined as:

$$\mathcal{G}_{x,\hat{\alpha}^{EB}} = \{G_{x,\hat{\alpha}^{EB}} : \mathcal{X} \times \mathcal{A}^B \to \{0, 1\}\}, \quad \mathcal{A}^B = \{\hat{\alpha}^{EB} | \mathcal{S}_n\}$$

Regret Bound on $\hat{\alpha}^{EB}$-Augmented Rules

Theorem 2: Under Assumptions 1–2, regret of $\hat{\alpha}$-Augmented policy rules satisfies:

$$R(\hat{G}_{x,\hat{\alpha}}) := E_{P^n}[W(G^*_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}})] \leq 2C_1 \frac{M}{k}\sqrt{\frac{v_{x,\hat{\alpha}}}{n}} + c_1\sigma_\varepsilon$$

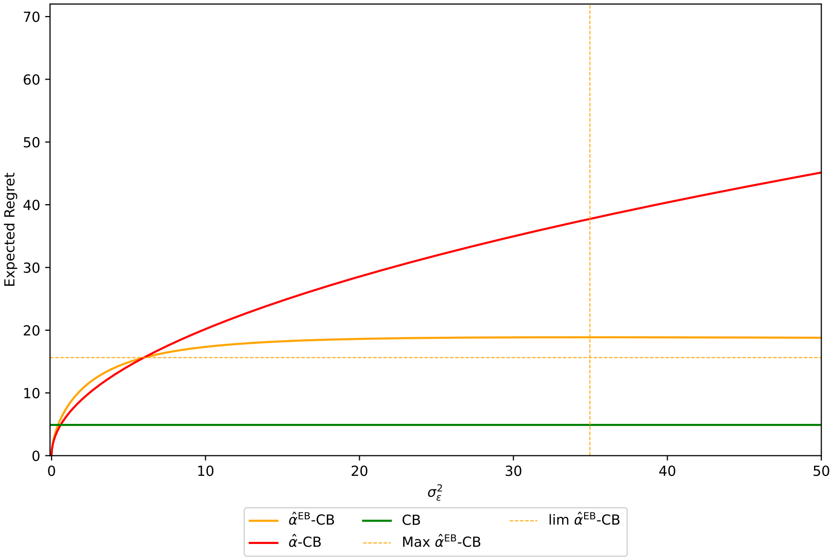
Proposition 1: Under Assumptions 1-3, regret of $\hat{\alpha}^{EB}$-Augmented policy rules satisfies:

$$R(\hat{G}_{x,\hat{\alpha}^{EB}}) := E_{P^n}[W(G^*_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}^{EB}})] \leq$$

$$\leq 2C_1 \frac{M}{k}\sqrt{\frac{v_{x,\hat{\alpha}^{EB}}}{n}} + c_1\left[\frac{1+\sqrt{n}}{\sqrt{n}}(\omega\sigma_\varepsilon + (1-\omega)\sigma_\alpha) + \frac{\sigma_\varepsilon}{\sqrt{n}}\right]$$

where C_1 is a universal constant and $c_1 := 4L\sqrt{2/\pi}$.

Asymptotic Comparison - Analytical Bounds



Data-Driven Shrinkage based on Welfare

Mixed Rules:

$$\mathcal{G}(\lambda) = \{G(\lambda) : \mathcal{X} \times \tilde{\mathcal{A}} \to \{0,1\}\}, \quad \tilde{\mathcal{A}} := \Lambda \times \mathcal{A} := \{\lambda\hat{\alpha} \in \mathbb{R} : \lambda \in \Lambda,\ \hat{\alpha} \in \mathcal{A}\}$$

For any class $\mathcal{G}$, $\mathcal{G}(0) = \mathcal{G}_x$, and $\mathcal{G}(1) = \mathcal{G}_{x,\hat{\alpha}}$.

How to select λ?

- Split $\mathcal{S}$ of size N into a training set $\mathcal{S}_{\text{train}}$, and a validation set $\mathcal{S}_{\text{val}}$.
- Estimate $\hat{G}(\lambda)$ on $\mathcal{S}_{\text{train}}$ and select $\hat{\lambda}$ on $\mathcal{S}_{\text{val}}$:

$$\hat{G}(\hat{\lambda}) := \underset{\lambda \in \Lambda, i \in \mathcal{S}_{\text{val}}}{\text{argmax}} \{W_m(\hat{G}(\lambda))\}$$

Cross-Validated Mixed Rules - Example Algorithm

Mixed (threshold) **Rules**:

$$G(\lambda) = \mathbb{1}(x + \lambda \cdot \hat{\alpha} > t(\lambda)), \quad \lambda \in \Lambda \subset [0, 1]$$

$\rightarrow$ $\lambda = 0$: Covariate-Based rule G_x

$\rightarrow$ $\lambda = 1$: $\hat{\alpha}$-Augmented rule $G_{x,\hat{\alpha}}$

Idea: Learn the optimal weight to attach to $\hat{\alpha}$ via cross-validation using the welfare generated out-of-sample.

Optimality of CV-Mixed Rules

Proposition 3: Under Assumption 1, with probability at least $1 - \delta$,

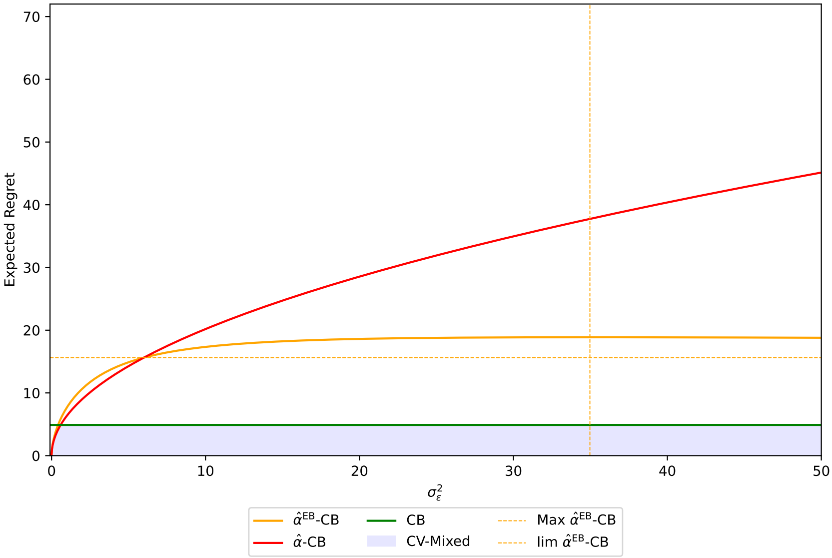
$$E_{P^n}[W(\hat{G}(\hat{\lambda}))] \geq \max\{E_{P^n}[W(\hat{G}(0))],\ E_{P^n}[W(\hat{G}(1))]\} - 2\gamma_m,$$

where:

$$\gamma_m := M \cdot \sqrt{\frac{1}{2m} \log\left(\frac{2r}{\delta}\right)}$$

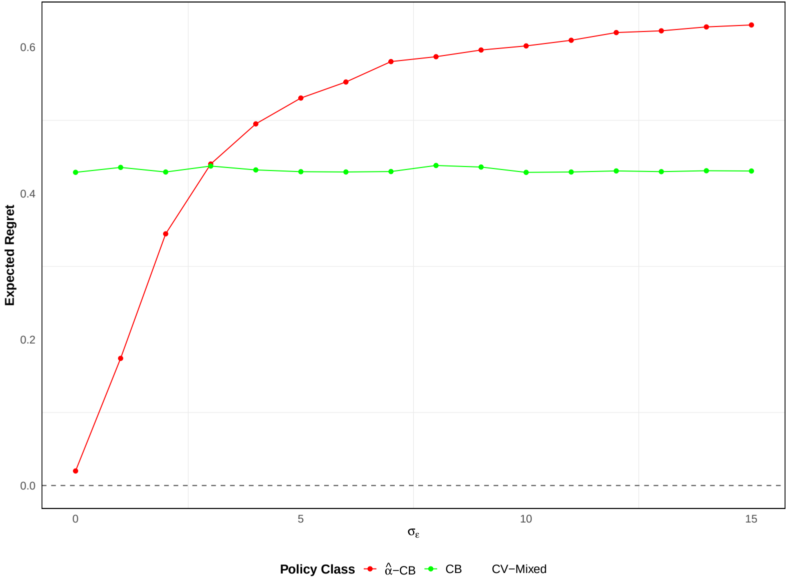
- **Key Insight:** with high probability, CV-Mixed rules outperform the best between Covariate-Based and $\hat{\alpha}$-Augmented rules.

Asymptotic Comparison - Analytical Bounds



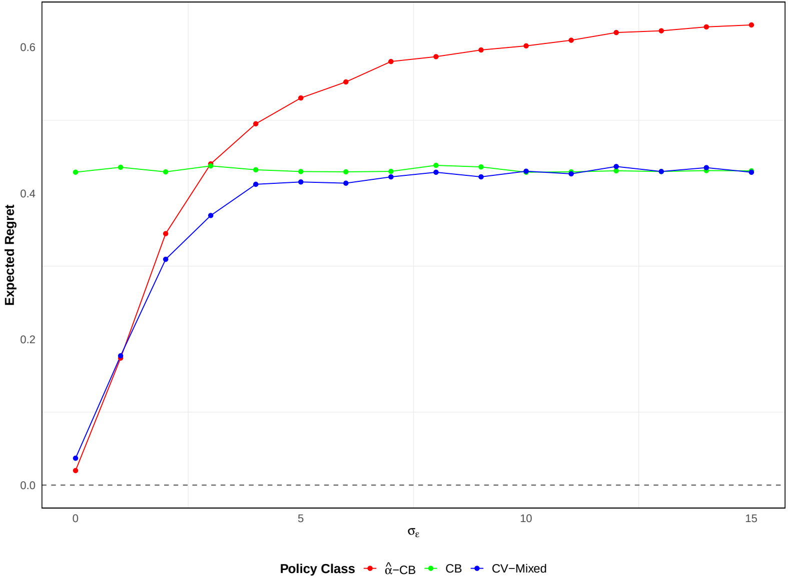
Simulation Results - Regret CV-Mixed Rules

Simulations DGP

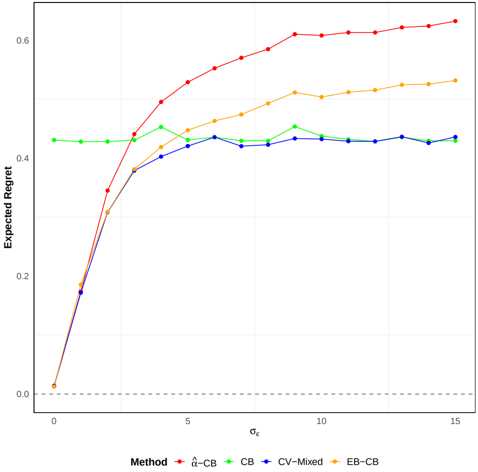


Simulation Results - Regret CV-Mixed Rules

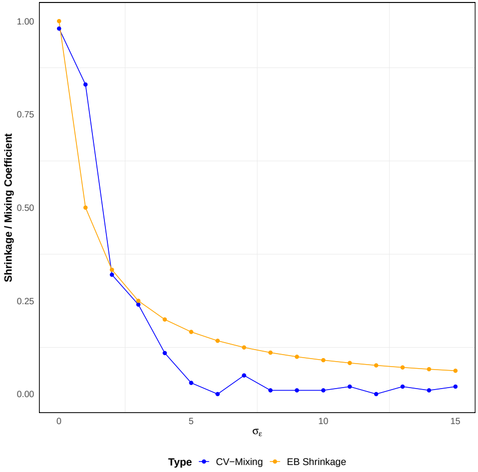
Is $\hat{\lambda}$ Interpretable? Different $\sigma_{\tau|x}$ Back to Empirics



Simulation Results - $\hat{\lambda}$ and ω Back



Panel A. Regret



Panel B. Shrinking Parameters

This Presentation

Empirical Application - Setting

Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field (Hussam, Rigol, and Roth, 2022 (AER)):

- RCT with 1500 Indian microentrepreneurs.
- Treatment: cash for business development.
- Outcome: profits.
- Heterogeneity dimension: community ratings as a proxy for business skills.

Main point of the paper: **demonstrate that community knowledge can help target high-growth microentrepreneurs**.

Policy Learning Exercise

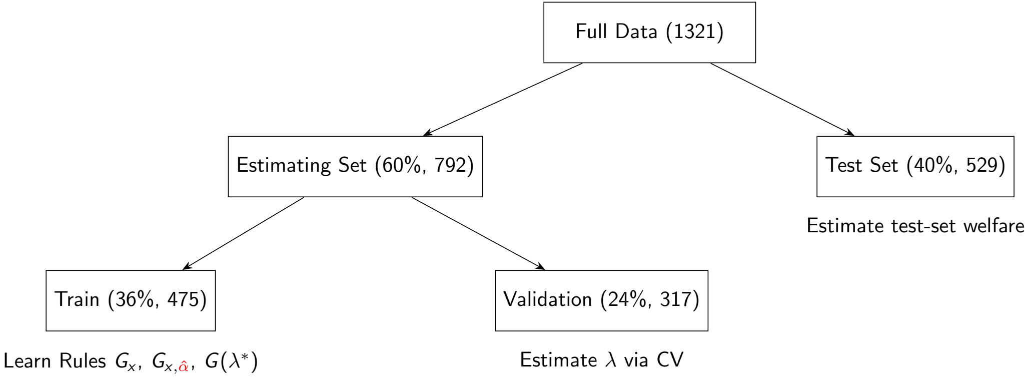
Does including community ratings as a targeting variable *actually* increase welfare?

- Status Quo: don't scale up.
- Random rule G_{rand}: scale up randomly.
- CB threshold rules G_x: age, education.
- $\hat{\alpha}$-CB threshold rules $G_{x,\hat{\alpha}}$: covariates + community rating.
- CV-Mixed threshold rules $G(\lambda)$: selects the weight λ of community ratings via cross-validation.

Ranking Rules

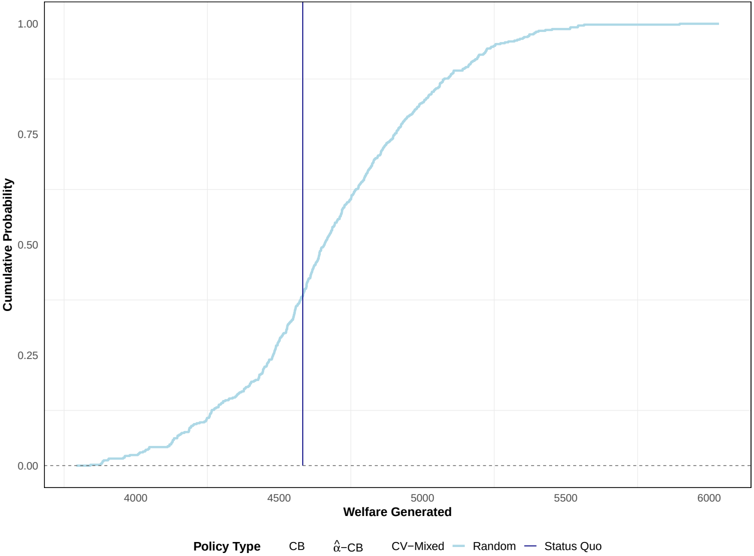
Formal Algorithm | Welfare

Randomly split the data into an estimating and testing sample:

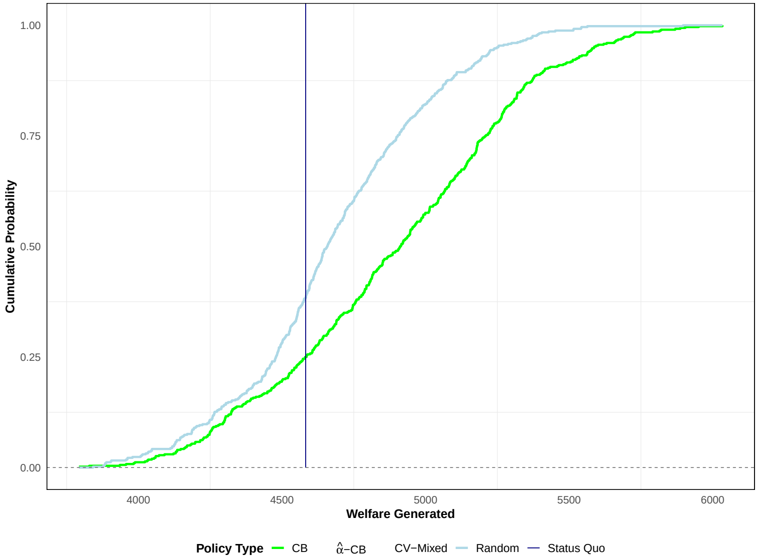


Repeat the random split $B = 500$ times.

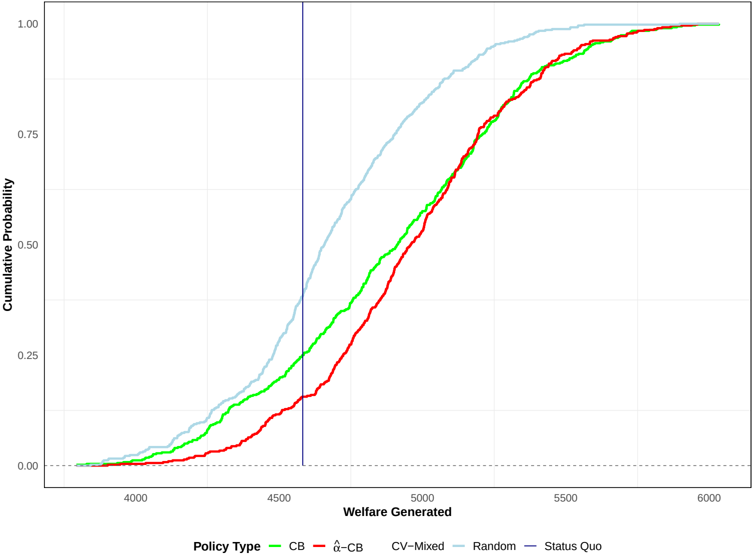
Distribution of Welfare - Random Rules



Distribution of Welfare - CB Rules

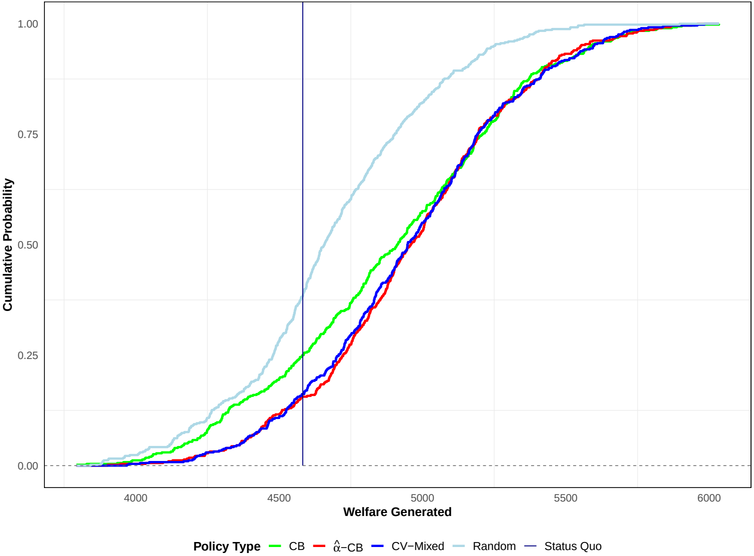


Distribution of Welfare - $\hat{\alpha}$-CB Rules



Distribution of Welfare - CV-Mixed Rules

Noise Increase Table Summary



Ethical Policy Learning

- Allocation of a scholarship.
- The policymaker can access genetic data as a proxy for individuals' innate ability.
- High-ability (tyoe H) students benefit more from the policy.

Would that be ethical to assign the scholarship based on the genetic information?

I define as ethical the set of allocations that do not enlarge pre-existing differences in outcomes between groups defined through the proxy of the unobserved characteristic:

$$\max_{G \in \mathcal{G}} E_{P^n}[W(G)] \quad \text{s.to:} \quad \mathcal{G} = \{G : E_{P^n}[W(G)|i \in H] - E_{P^n}[W(G)|i \in L] \leq Q_a\}$$

Conclusions
Future Directions

- **Main insight**: Including estimated latent variables introduces an approximation-estimation error trade-off.

$\rightarrow$ Improves policy recommendations if α's importance $>$ $\hat{\alpha}$'s estimation error.

- **CV-Mixed rules**: adaptively set the importance of $\hat{\alpha}$ via cross-validation.

$\rightarrow$ Theoretically and empirically shown to achieve **near-optimal welfare**.

- **Empirical application** in development economics.

$\rightarrow$ Intuitive procedure to rank policy recommendations.

Thanks for your attention!

For any comment: giacomo.opocher2@unibo.it

Simulation Study - Design

Back to Regret Back to CB Back to $\hat{\alpha}$-CB Back to CV-Mixed Back to Different γ

Data-Generating Process:

- Covariates: $X_i \in \mathbb{R}^1$.
- Unobserved characteristic: $\alpha_i \sim N(0, \sigma_\alpha^2)$.
- Unobserved characteristic's estimate: $\hat{\alpha}_i = \alpha_i + N(0, \sigma_\varepsilon^2)$.
- Potential outcomes:

$$Y_i(0) = g(X_i) + A_i + \varepsilon_i, \quad Y_i(1) = Y_i(0) + \tau(X_i, A_i)$$

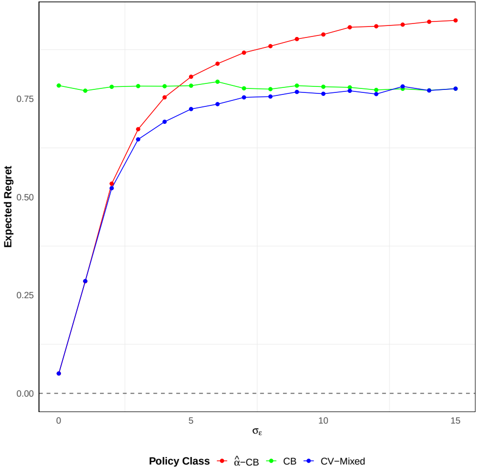
- Treatment assignment: $D_i \sim \mathit{Bernoulli}(0.5)$.

Treatment Effect:

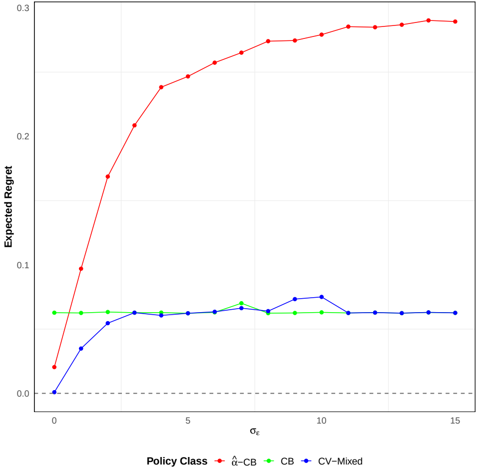
$$\tau(x, \alpha) = x + \gamma \cdot \alpha$$

Linear in (x, α) with varying γ to control unobserved heterogeneity.

Simulation Results - Different $\sigma_{\tau|x}$ Back to CV-Mixed Simulations DGP



Panel A. High $\sigma_{\tau|x}$



Panel B. Low $\sigma_{\tau|x}$

$R(G_{x,\hat{\alpha}})$ - Formal Steps Back

$$E_{P^n}[W(\hat{G}_{x,\alpha}) - W(\hat{G}_{x,\hat{\alpha}})] \leq KT18 + 4 \cdot E_{P^n}\left[\frac{1}{n}\sum_{i=1}^{n}(\tau(X_i, A_i) - \tau(X_i, \hat{\alpha}))\right] \leq$$

$$\leq KT18 + 4 \cdot E_{P^n}\left[\frac{1}{n}\sum_{i \in \mathcal{S}^n}|\tau(X_i, A_i) - \tau(X_i, \hat{\alpha})|\right]$$

$$\leq KT18 + 4L \cdot E_{P^n}\left[\frac{1}{n}\sum_{i \in \mathcal{S}^n}|A_i - \hat{\alpha}_i|\right] =$$

$$= KT18 + 4L \cdot \sqrt{2/\pi} \cdot \sigma_\varepsilon = KT18 + 4L \cdot \sqrt{2/\pi} \cdot \sqrt{E_{P^n}\left[\frac{1}{n}\sum_{i \in \mathcal{S}^n}(A_i - \hat{\alpha}_i)^2\right]}$$

Cross-Validated Mixed Policy Rules - Algorithm

Back

Algorithm Cross-Validated Mixed Rules

Require: Data $(X_i, \hat{\alpha}_i, Y_i, D_i)$ for $i = 1, \ldots, n+m$, grid $\Lambda = \{\lambda_1, \ldots, \lambda_r\} \subset [0, 1]$.

```
Randomly split data into training set S_train of size n and validation set S_val of size m.
for each λ ∈ Λ do:
    Estimate thresholds t(λ) on S_train.
    Define G(t̂(λ)).
    Estimate empirical welfare W_m(G(t̂(λ))) on S_val.
end for.
Estimate λ̂ = arg max_{λ∈Λ} W_m(G(t̂(λ))).
return Final policy rule Ĝ(λ̂).
```

Randomly split data into training set $\mathcal{S}_{\text{train}}$ of size n and validation set $\mathcal{S}_{\text{val}}$ of size m.

for each $\lambda \in \Lambda$ **do**:

  Estimate thresholds $t(\lambda)$ on $\mathcal{S}_{\text{train}}$.

  Define $G(\hat{t}(\lambda))$.

  Estimate empirical welfare $W_m(G(\hat{t}(\lambda)))$ on $\mathcal{S}_{\text{val}}$.

end for.

Estimate $\hat{\lambda} = \arg\max_{\lambda \in \Lambda} W_m(G(\hat{t}(\lambda)))$.

return Final policy rule $\hat{G}(\hat{\lambda})$.

Can we compute the distribution of the performance? Back

Consider different realizations of the sample splitting:

Algorithm Welfare Evaluation

for $b = 1$ to $B = 500$ **do**
    Set random seed to b.
    Random split: $\mathcal{S} = \mathcal{S}^b_{\text{est}} \cup \mathcal{S}^b_{\text{est}}$ and $\mathcal{S}^b_{\text{est}} = \mathcal{S}^b_{\text{train}} \cup \mathcal{S}^b_{\text{val}}$, $\mathcal{S}^b_{\text{train}} \cap \mathcal{S}^b_{\text{val}} \cap \mathcal{S}^b_{\text{test}} = \emptyset$
    Estimate Rules:
    Estimate G_x and $G_{x,\hat{\alpha}}$ using $\mathcal{S}^b_{\text{train}}$.
    Estimate $G(\lambda^*)$ using $\mathcal{S}^b_{\text{train}}$ and $\mathcal{S}^b_{\text{val}}$.
    Evaluate Rules:
    Estimate $\hat{W}^b_{\text{test}}(\hat{G}_{\text{rand}})$, $\hat{W}^b_{\text{test}}(\hat{G}_x)$, $\hat{W}^b_{\text{test}}(\hat{G}_{x,\hat{\alpha}})$, and $\hat{W}^b_{\text{test}}(\hat{G}(\lambda^*))$.
end for

Estimated Rules and Welfare

Back

Rules are defined as: $\hat{G}_z := \arg\max_{G_z \in \mathcal{G}_z} \{W_n(G_z)\}$, where:

$$W_n(G_z) := \frac{1}{475} \sum_{i \in \mathcal{S}} \left[\frac{Y_i D_i}{0.3} \cdot 1(i \in G_z) + \frac{Y_i(1 - D_i)}{0.7} \cdot 1(i \notin G_z) \right]$$

and:

- Y_i is profits 30 days after the intervention.
- D_i takes value one if i received the grant.

What if Community Rankings Were More Noisy? Back to Welfare

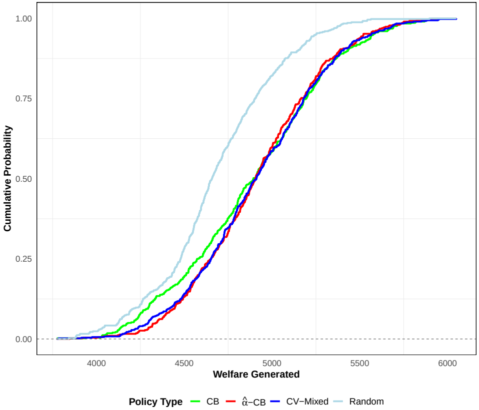
Are these findings robust to an increase in σ^2_ε? Add random noise $\zeta_i \sim \mathcal{N}(0, \sigma^2_\zeta)$ to the original variable:

$$\tilde{\alpha}_i = \hat{\alpha} + \zeta_i$$

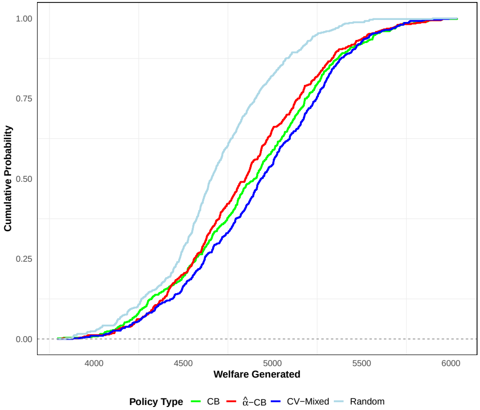
And apply the same algorithm to compute welfare gains.

Noise Increase - Welfare Gains

Simulations 1 Simulations 2 Back to Graph



Panel A. $\sigma_\zeta^2 = 1$



Panel B. $\sigma_\zeta^2 = 5$

Welfare Gains - Summary Table

Back to Graph

How can we summarize welfare gains?

Policy Rule	Harm Rate	Rand.	CB	$\hat{\alpha}$-CB	CV-Mixed
Status Quo	-	+98\$ (+2%)	+310\$ (+7%)	+380\$ (+8%)	+375\$ (+8%)
Rand.	0.38	-	+212\$ (+5%)	+282\$ (+6%)	+277\$ (+6%)
CB	0.25	-	-	+69\$ (+1%)	+64\$ (+1%)
$\hat{\alpha}$-CB	0.16	-	-	-	-5\$ (-0%)
CV-Mixed	0.16	-	-	-	-
Status Quo	4,582\$	-	-	-	-

Notes: Each cell reports the difference in mean welfare between the policy class in the column and the one in the row. Positive values indicate that the column policy performs better. Harm Rate denotes the probability that the policy yields lower welfare than the status quo.

Future Directions (1/2)

Three different policy learning problems:

- Find a treatment rule that generalizes well. **Today!**
- Find an optimal subset of a given sample. **Work in Progress...**
  - $\rightarrow$ New matching estimator to estimate ITEs and find the subgroup that maximizes *Synthetic* Welfare.
- Treat/not decision on a single unit. **Work in Progress...**
  - $\rightarrow$ New probabilistic bounds on ITEs.

Future Directions (2/2) Conclusions

Policy recommendations meet policy learning.

- Take most cited papers with an RCT published on a top-5.
- Formalize their policy recommendations.
- Evaluate their performance.
- Compare them with what policy learning would suggest.