

Trading-off Bias and Variance When the Size of the Treatment Effect is Bounded

Clément de Chaisemartin

Sciences Po

Toulouse School of Economics

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals
- 6 Applications
- 7 Conclusion

Minimax-linear ATE estimation

- Target: average treatment effect (ATE) τ , weighted average of S CATEs $(\tau_s)_{1 \leq s \leq S}$ with known weights $(p_s)_{1 \leq s \leq S}$ summing to 1: $\tau = \sum_{s=1}^S p_s \tau_s$.
- One has unbiased CATE estimators $(\hat{\tau}_s)_{1 \leq s \leq S}$.
- Examples: stratified RCTs, matching studies.
- Why not just use $\hat{\tau}(\mathbf{p}) := \sum_{s=1}^S p_s \hat{\tau}_s$ to estimate τ ?
- Some CATE estimators may be less precise than others, downweighting imprecise estimators may lead to lower MSE.
- I derive minimax linear estimators of, and confidence intervals (CI) for, τ .
- Effort to derive closed-form expressions, to make estimators transparent.

Assuming bounded CATEs (1/2)

- First, I assume that CATEs' magnitude bounded: $|\tau_s| \leq B$.
- Appeal: applied researchers often have good ex-ante sense of what would be an implausibly large effect.
- Could be based on literature studying related interventions. E.g.: my first application considers effect, on students' test scores, of a school close to "No Excuse" charter schools. In review of four papers estimating 22 CATEs for such schools, none larger than 0.35σ , and for 20 we reject that $> 0.5\sigma$. Then, may be plausible to assume that CATEs $\leq 0.5\sigma$.

Assuming bounded CATEs (2/2)

- Under the assumption that $|\tau_s| \leq B$, I derive a closed-form expression of the minimax-linear estimator of τ .
- Weighted sum of CATE estimators, with ≥ 0 weights summing to ≤ 1 .
- Most precise estimators receive weight equal to p_s .
- Least precise estimators receive weight proportional to one over their variance, and shrunk towards zero.
- Then, using similar ideas as in Donoho (1994), Armstrong and Kolesár (2018a), and Armstrong and Kolesár (2021), I outline simple numerical procedure to approximate CIs for τ with minimax length.

Assuming CATEs heterogeneity is bounded

- Alternative assumption: $|\tau_s - \tau|/|\tau| \leq B. \Rightarrow$ CATEs' coefficient of variation $\leq B$.
 - I am able to derive closed-form of minimax estimator in one special case.
 - Then, shrinks all CATEs towards zero uniformly, multiplying them by the same constant, irrespective of their variances.
- $\Rightarrow$ minimax estimator leads to lower power than unbiased one! Example of well-known fact that minimax estimators do not always lead to maximin tests (Lehmann and Romano, 2005).
- Simple fix: instead of considering unconstrained minimax linear estimator, I propose to consider instead the minimax estimator across all linear estimators that do not downweight the most precise CATE estimator.

Applications: small but non-negligible precision gain

- I revisit Behaghel et al. (2017), SRCT, and Connors et al. (1996), matching study.
- In both cases, assuming that $|\tau_s| \leq B$ for arguably large values of B , minimax estimator close to unbiased one, sd around 10% smaller.
- Gain in precision not offset by bias: estimated RMSE around 10% smaller than that of unbiased estimator, worst-case RMSE around 5-6% smaller.
- Minimax CIs 5 to 7% smaller than CIs based on the unbiased estimator.
- Assuming that $|\tau_s - \tau|/|\tau| \leq B_1$ leads, if anything, to slightly smaller precision gains than assuming that $|\tau_s| \leq B_2 := B_1|\tau|$.
- In SRCT, achieving 5% precision gain by controlling for covariates requires that covariates account for $1 - 0.95^2 \approx 10\%$ of outcome's variance.

Related papers improving precision of ATE estimators

- Crump et al. (2009):
 - Redefine target as ATE in subpopulation whose ATE can be estimated most precisely: observations with propensity score in $[0.1, 0.9]$ interval.
 - Instead, minimax estimators do not change goalpost.
 - In my second empirical application minimax estimator almost as precise as trimming estimator of Crump et al. (2009).
- Kwon and Sun (2025):
 - Derive minimax length CIs for the ATE under assumption that variance of CATEs is bounded: $\sum_{s=1}^S p_s (\tau_s - \tau)^2 \leq B$.
 - Applied researchers may find it harder to choose an ex-ante plausible bound for variance of CATEs than for their magnitude, because until now empirical papers rarely estimate that variance.

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals
- 6 Applications
- 7 Conclusion

Set up

Definition 1

(ATE and CATEs) Target is τ , a weighted average of S conditional average treatment effects $(\tau_s)_{1 \leq s \leq S}$, with weights $(p_s)_{1 \leq s \leq S}$ that are known, positive, and sum to 1:

$$\tau = \sum_{s=1}^S p_s \tau_s. \quad (1)$$

Assumption 1

One observes random variables $(\widehat{\tau}_s)_{1 \leq s \leq S}$ such that:

- 1** $E(\widehat{\tau}_s) = \tau_s$ for all s .
- 2** $\text{cov}(\widehat{\tau}_s, \widehat{\tau}_{s'}) = 0$ for all $s' \neq s$.
- 3** $V_s := V(\widehat{\tau}_s) < +\infty$.

Stratified Randomized Controlled Trials (SRCTs)

- SRCT with S strata indexed by s .
- τ_s : CATE in stratum s , p_s : share of population s accounts for, τ : ATE.
- $\hat{\tau}_s$: DIM in stratum s .
- n_s : # of units in s , $n_{1,s}$: # treated, $n_{0,s} = n_s - n_{1,s}$: # control, $\mathbf{D}_s$: treatments of all units in s , $\mathbf{Y}_s$: their potential outcomes.
- Assume that:

$$\text{1 } \forall (d_1, \dots, d_{n_s}) : \sum_{i=1}^{n_s} d_i = n_{1,s}, P(\mathbf{D}_s = (d_1, \dots, d_{n_s}) | \mathbf{Y}_s) = \frac{1}{\binom{n_s}{n_{1,s}}}.$$

$$\text{2 } (\mathbf{D}_s, \mathbf{Y}_s) \text{ mutually independent across } s.$$

Holds by design in SRCTs. Implies Points 1 and 2 of Assumption 1 hold.

- If units in s randomly drawn from superpopulation,

$$V_s = \frac{\sigma_{0,s}^2}{n_{0,s}} + \frac{\sigma_{1,s}^2}{n_{1,s}}, \quad (2)$$

so Point 3 of Assumption 1 holds if $\sigma_{0,s}^2 < +\infty$ and $\sigma_{1,s}^2 < +\infty$.

Estimation under CIA, using AIPW estimator (1/2)

- s indexes iid units drawn from superpopulation.
- $Y_s(0)$ and $Y_s(1)$: potential outcomes, D_s and Y_s : treatment and observed outcome, X_s : covariates.
- CIA:

$$(Y_s(0), Y_s(1)) \perp\!\!\!\perp D_s | X_s. \quad (3)$$

- Let $\tau_s = E(Y_s(1) - Y_s(0) | X_s)$, $p_s = 1/S$.
- Then, $\tau = 1/S \sum_{s=1}^S E(Y_s(1) - Y_s(0) | X_s)$: SATE considered by Crump et al. (2009).

Estimation under CIA, using AIPW estimator (2/2)

■ $e(X_s) := P(D_s = 1|X_s), \forall d \in \{0, 1\} \mu_d(X_s) := E(Y_s|D_s = d, X_s).$

■ Let

$$\hat{\tau}_s = \mu_1(X_s) - \mu_0(X_s) + D_s \frac{Y_s - \mu_1(X_s)}{e(X_s)} - (1 - D_s) \frac{Y_s - \mu_1(X_s)}{1 - e(X_s)}. \quad (4)$$

$1/S \sum_{s=1}^S \hat{\tau}_s$: oracle version of AIPW estimator of Robins et al. (1994).

■ Under (3), $E(\hat{\tau}_s|X_s) = \tau_s$: Point 1 of A1 holds conditional on X_s .

■ Point 2 mechanically holds as units are assumed to be iid.

■ If for all $d \in \{0, 1\}, V(Y_s(d)|X_s) = \sigma_d^2(X_s) < +\infty$, then

$$V(\hat{\tau}_s|X_s) = \frac{\sigma_0^2(X_s)}{1 - e(X_s)} + \frac{\sigma_1^2(X_s)}{e(X_s)}, \quad (5)$$

so Point 3 of A1 also holds conditional on X_s .

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs**
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals
- 6 Applications
- 7 Conclusion

Bounded CATEs

Assumption 2

There is a known $B \in \mathbb{R}_+ \setminus \{0\}$ such that $|\tau_s| \leq B$ for all s .

Worst-case MSE

For any $1 \times S$ deterministic vector $\mathbf{w} = (w_1, \dots, w_S)$, let

$$\hat{\tau}(\mathbf{w}) = \sum_{s=1}^S w_s \hat{\tau}_s. \quad (6)$$

$\hat{\tau}(\mathbf{w})$: linear combination of estimators $\hat{\tau}_s$.

Lemma 1

(Worst-case MSE of $\hat{\tau}(\mathbf{w})$)

If Assumptions 1 and 2 hold,

$$E \left((\hat{\tau}(\mathbf{w}) - \tau)^2 \right) \leq \overline{\text{MSE}}(\mathbf{w}) := \sum_{s=1}^S w_s^2 V_s + B^2 \left(\sum_{s=1}^S |w_s - p_s| \right)^2.$$

The upper bound in the previous display is sharp.

Minimax-linear estimator

Wlog: $p_1 V_1 \leq p_2 V_2 \leq \dots \leq p_S V_S$. Let

$$\underline{s} = \min \left\{ s \in \{1, \dots, S\} : \frac{1}{\frac{1}{B^2} + \sum_{s'=s}^S \frac{1}{V_{s'}}} \sum_{s'=s}^S p_{s'} < p_s V_s \right\}.$$

Let

$$w_{B,s}^M = p_s \text{ for all } s < \underline{s}$$

$$w_{B,s}^M = \frac{1}{V_s} \frac{1}{\frac{1}{B^2} + \sum_{s'=\underline{s}}^S \frac{1}{V_{s'}}} \sum_{s'=\underline{s}}^S p_{s'} \text{ for all } s \geq \underline{s}. \quad (7)$$

Theorem 1

If Assumption 1 holds, $\mathbf{w}_B^M = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^S} \overline{\text{MSE}}(\mathbf{w})$.

Minimax estimator: weighted sum of $\hat{\tau}_s$ s, with ≥ 0 weights summing to ≤ 1 .
 For precise $\hat{\tau}_s$ (with low value of $p_s V_s$), optimal weight: p_s . For imprecise $\hat{\tau}_s$, optimal weight proportional to one over its variance, and shrunk towards zero.

Feasible estimator

- In an SRCT, under homoscedasticity assumption minimax estimator feasible: only depends on known quantities.
- In general, infeasible: depends on unknown quantities, that have to be estimated.
- In SRCT, to estimate $V_s = \frac{\sigma_{0,s}^2}{n_{0,s}} + \frac{\sigma_{1,s}^2}{n_{1,s}}$ one can replace $\sigma_{0,s}^2$ and $\sigma_{1,s}^2$ by variances of outcome in the control and treatment groups, respectively.
- In matching study, to estimate V_s one can replace $\sigma_0^2(X_s)$, $\sigma_1^2(X_s)$, and $e(X_s)$ in (5) by some estimators.

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity**
- 5 Minimax-length confidence intervals
- 6 Applications
- 7 Conclusion

Bounded heterogeneity

Assumption 3

S is even, $p_s = 1/S$, and $V_s < V_{s+1}$ for all $s \leq S - 1$.

Assumption 4

$\tau \neq 0$, and there is a known $B \in \mathbb{R}_+$ such that $\left| \frac{\tau_s - \tau}{\tau} \right| \leq B \forall s$.

A2: distance between τ_s and τ no larger than $B|\tau|$. Implies

$$\frac{\sqrt{1/S \sum_{s=1}^S (\tau_s - \tau)^2}}{|\tau|} \leq B. \quad (8)$$

With $B = 0$, A2 requires that all CATEs are equal. With $B \leq 1$, requires all CATEs have same sign.

Worst-case MSE

$\forall \mathbf{w} \in \mathbb{R}_+^S$, let (s) be permutation of $\{1, \dots, S\}$ such that $s \mapsto w_{(s)}$ decreasing.

Lemma 2

(Worst-case MSE of $\hat{\tau}(\mathbf{w})$ under Assumption 4) If A1, A3, and A4 hold,

$$E \left(\left(\hat{\tau}(\mathbf{w}) - \tau \right)^2 \right) \leq \sum_{s=1}^S w_s^2 V_s + \tau^2 \max \left[\left((B+1) \sum_{s=1}^{S/2} (w_{(s)} - 1/S) - (B-1) \sum_{s=S/2+1}^S (w_{(s)} - 1/S) \right)^2, \left((B+1) \sum_{s=S/2+1}^S (w_{(s)} - 1/S) - (B-1) \sum_{s=1}^{S/2} (w_{(s)} - 1/S) \right)^2 \right].$$

Proof: worst-case + (resp. -) bias reached letting

$$\tau_{(s)}^+ = (B+1)\tau 1\{s \leq S/2\} - (B-1)\tau 1\{s > S/2\} \text{ (resp.}$$

$$\tau_{(s)}^- = (B+1)\tau 1\{s > S/2\} - (B-1)\tau 1\{s \leq S/2\}).$$

Minimax-linear positive estimator (1/3)

Let

$$V(\mathbf{w}) = \sum_{s=1}^S w_s^2 V_s,$$

$$\overline{SQB}^H(\mathbf{w}) = \tau^2 \max \left[\left((B+1) \sum_{s=1}^{S/2} (w_{(s)} - 1/S) - (B-1) \sum_{s=S/2+1}^S (w_{(s)} - 1/S) \right)^2, \right. \\ \left. \left((B+1) \sum_{s=S/2+1}^S (w_{(s)} - 1/S) - (B-1) \sum_{s=1}^{S/2} (w_{(s)} - 1/S) \right)^2 \right],$$

$$\overline{MSE}^H(\mathbf{w}) = V(\mathbf{w}) + \overline{SQB}^H(\mathbf{w}).$$

Let

$$\mathbf{w}^H = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}_+^S} \overline{MSE}^H(\mathbf{w}).$$

Minimax across linear combinations with ≥ 0 weights, to avoid sign reversal.

Minimax-linear positive estimator (2/3)

Recall: $V_1 < V_2 < \dots < V_S$, and (s) permutation such that $s \mapsto w_{(s)}$ decreasing, dependence on $s \mapsto w_s$ left implicit.

Let $[s]$ denote permutation such that $s \mapsto w_{[s]}^H$ decreasing.

Lemma 3

$\sum_{s=1}^S w_{[s]}^H \leq 1$, and

$$\begin{aligned} & \left| (B+1) \sum_{s=1}^{S/2} (w_{[s]}^H - 1/S) - (B-1) \sum_{s=S/2+1}^S (w_{[s]}^H - 1/S) \right| \\ & \leq \left| (B+1) \sum_{s=S/2+1}^S (w_{[s]}^H - 1/S) - (B-1) \sum_{s=1}^{S/2} (w_{[s]}^H - 1/S) \right|. \end{aligned}$$

Lemma 4

$w_s^H \geq w_{s+1}^H$ for all $s \leq S-1$.

Minimax-linear positive estimator (3/3)

$$\overline{\text{MSE}}^{H,d}(\mathbf{w}) := V(\mathbf{w}) + \tau^2 \left((B-1) \sum_{s=1}^{S/2} w_s - (B+1) \sum_{s=S/2+1}^S w_s + 1 \right)^2,$$

$$\mathcal{R} := \left\{ \mathbf{w} \in \mathbb{R}^S : \sum_{s=1}^S w_s - 1 \leq 0, \forall s \in \{1, \dots, S-1\} : w_{s+1} - w_s \leq 0, -w_S \leq 0 \right\}.$$

Theorem 2

If A1, A3, and A4 hold:

- 1 $\mathbf{w}^H = \operatorname{argmin}_{\mathbf{w} \in \mathcal{R}} \overline{\text{MSE}}^{H,d}(\mathbf{w})$.
- 2 If $B \geq 1$ and $\frac{V_S}{1/S \sum_{s'=1}^S V_{s'}} \leq B+1$,

$$w_1^H = w_S^H = \frac{\tau^2}{\frac{1}{S^2} \sum_{s=1}^S V_s + \tau^2} 1/S.$$

Minimax but powerless?

- Point 2 of Thm 2: if $B \geq 1$ and CATES' largest variance no larger than $(B + 1)$ times average variance, $\hat{\tau}(\mathbf{w}^H)$ shrinks uniformly all estimators $\hat{\tau}_s$ by the same constant.
 - Shrinkage may reduce power: CI for τ based on $\hat{\tau}(\mathbf{w}^H)$ contains zero more often than the CI based on $\hat{\tau}(\mathbf{p})$.
 - Intuition: even without accounting for the bias of $\hat{\tau}(\mathbf{w}^H)$, t-stats of $\hat{\tau}(\mathbf{w}^H)$ and $\hat{\tau}(\mathbf{p})$ are equal.
- ⇒ $\hat{\tau}(\mathbf{w}^H)$ unappealing alternative to $\hat{\tau}(\mathbf{p})$.

Power-aware minimax estimator

- Whenever $w_1^H < 1/S$ (even the most precisely estimated CATE is shrunk towards zero), one can construct an alternative estimator whose bias-aware CI will contain zero less often: $\tilde{w}_s := 1/S/w_1^H \times w_s^H$.
- Proposal: consider instead

$$\mathbf{w}^P = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}_+^S: w_{(1)} \geq 1/S} \overline{\text{MSE}}^H(\mathbf{w}),$$

minimax-linear estimator constrained to not shrink at least one estimator.

- Alternative idea:

$$\mathbf{w}^\Sigma = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}_+^S: \sum_s w_s = 1} \overline{\text{MSE}}^H(\mathbf{w}),$$

minimax-linear estimator constrained to have weights summing to one.

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals**
- 6 Applications
- 7 Conclusion

Normality assumption

In this section, I derive Cls for τ under following assumption:

Assumption 5

$\hat{\tau}(\mathbf{w})$ follows a normal distribution.

Then,

$$\hat{\tau}(\mathbf{w}) - \tau \sim \mathcal{N}(B(\mathbf{w}), \sigma^2(\mathbf{w})).$$

Conjecture:¹ Cls asymptotically valid without that assumption.

¹Based on Armstrong and Kolesár (2018b) for similar minimax Cls in different contexts.

Minimax-length CI with bounded CATEs (1/2)

Under Assumption 2,

$$|B(\mathbf{w})| \leq \overline{|B|}(\mathbf{w}) := B \sum_{s=1}^S |w_s - p_s|. \quad (9)$$

Then, to form CI for τ , I follow Donoho (1994); Armstrong and Kolesár (2018a, 2021), and use CI centered around $\hat{\tau}(\mathbf{w})$, with minimax length. As shown by those papers, equal to

$$\hat{\tau}(\mathbf{w}^{\text{CI}}) \pm Q_{1-\alpha}(\overline{|B|}(\mathbf{w}^{\text{CI}}), \sigma(\mathbf{w}^{\text{CI}})),$$

where $Q_{1-\alpha}(\mu, \sigma) :=$ quantile of order $1 - \alpha$ of $|X|$ where $X \sim \mathcal{N}(\mu, \sigma^2)$, and $\mathbf{w}^{\text{CI}}$ minimizer of $Q_{1-\alpha}(\overline{|B|}(\mathbf{w}), \sigma(\mathbf{w}))$ across all $\mathbf{w} \in \mathbb{R}^S$.

Minimax-length CI with bounded CATEs (2/2)

$\mathbf{w}^{\text{CI}}$ can be found following similar steps as Armstrong and Kolesár (2018a):

- 1 One can show that $0 \leq w_s^{\text{CI}} \leq p_s$. Therefore,
 $|B|(\mathbf{w}) := B \sum_{s=1}^S (p_s - w_s) \in [0, B]$.
- 2 Then, for any $M : 0 < M < B$, let $\mathbf{w}_M^{\text{CI}}$ be minimizer of $\sigma^2(\mathbf{w})$ subject to $B \sum_{s=1}^S (p_s - w_s) \leq M$ and $w_s \leq p_s$. For any $k \in \{1, \dots, S\}$, let

$$\lambda(k) = \frac{\sum_{s=k}^S p_s - M/B}{B \sum_{s=k}^S 1/V_s},$$

and let

$$\mathcal{S}_{\text{CI}} = \{s : \lambda(s) > 0, \lambda(s)B/V_s < p_s, \forall s' < s : \lambda(s)B/V_{s'} \geq p_{s'}\}.$$

Proposition 1

$\mathcal{S}_{\text{CI}} \neq \emptyset$, and $\exists s_{\text{CI}} \in \mathcal{S}_{\text{CI}} : w_{M,s}^{\text{CI}} = p_s 1\{s < s_{\text{CI}}\} + \lambda(s_{\text{CI}})B/V_s 1\{s \geq s_{\text{CI}}\}$.

- 3 $\mathbf{w}^{\text{CI}}$: argmin of $Q_{1-\alpha}(M, \sigma(\mathbf{w}_M^{\text{CI}}))$ across $M \in (0, B)$ (univariate grid search).

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals
- 6 Applications**
- 7 Conclusion

Behaghel et al. (2017): Background

- SRCT to estimate effect of boarding school for low-SES students in France.
- School's pedagogy is similar to US “No Excuse” charter schools.
- Capacity constraints at gender $\times$ grade level.
- In 2009 and 2010, admission lotteries in 14 gender $\times$ grade strata.
- Offer probability varies from 0.17 to 0.93 across strata.
- Outcome: math test score 2 years after lottery.
- τ_s : ITT of receiving offer to join school in stratum s .
- FS effect: 1.34, so τ_s effect of spending 1.34 years in the school.

Behaghel et al. (2017): Choosing B

- Closest intervention: Curto and Fryer Jr (2014), who study “No Excuse” charter boarding school in DC. One year spent in the school increases students’ math test scores by $0.23\sigma_0$.
 - Also estimate 8 CATEs (males/females, above/below median etc.)
 - All between 0.04 and $0.36\sigma_0$, and for 7 one rejects at 90% level that effect $\geq 0.5\sigma_0$.
- Angrist et al. (2010), Dobbie and Fryer Jr (2011), and Abdulkadiroğlu et al. (2011) study successful non-boarding “No Excuse” charter schools in New-York and Boston. Together, estimate 14 CATEs, all between 0.18 and $0.36\sigma_0$, and for 13 one rejects at 90% level that effect $\geq 0.5\sigma_0$.
- Therefore, I assume $|\tau_s| \leq B = 1.34 \times 0.5$.

Behaghel et al. (2017): Results

Table 1: Minimax estimators and CI in Behaghel et al. (2017)

	$\hat{\tau}(\mathbf{p})$	$\hat{\tau}(\mathbf{w}_{0.5}^M)$	$\hat{\tau}(\mathbf{w}_{0.6}^M)$	$\hat{\tau}(\mathbf{w}_{0.5}^{\text{hom}})$	$\hat{\tau}(\mathbf{w}^{\text{CI}})$
Point estimate	0.278	0.268	0.270	0.273	0.270
S.e.	0.134	0.122	0.124	0.124	
S.e./S.e. $\hat{\tau}(\mathbf{p})$	1	0.908	0.929	0.925	
RMSE/RMSE $\hat{\tau}(\mathbf{p})$	1	0.908	0.929	0.925	
$\overline{\text{RMSE}}/\overline{\text{RMSE}} \hat{\tau}(\mathbf{p})$	1	0.950	0.959	0.981	
95% level CI	[0.015,0.540]				[0.021,0.519]

Connors et al. (1996)

Matching study of right heart catheterization (RHC) on patient mortality. Propensity score and conditional means estimated using all (71) covariates in the data, following Crump et al. (2009).

Table 2: Minimax estimators and CI in Connors et al. (1996)

	$\hat{\tau}(\mathbf{p})$	$\hat{\tau}(\mathbf{w}_{0.2}^M)$	$\hat{\tau}(\mathbf{w}_{0.3}^M)$	$\hat{\tau}(\mathbf{w}^{CI})$
Point estimate	-0.064	-0.065	-0.066	-0.064
S.e.	0.016	0.014	0.015	
S.e./S.e. $\hat{\tau}(\mathbf{p})$	1	0.893	0.921	
RMSE / RMSE $\hat{\tau}(\mathbf{p})$	1	0.893	0.921	
RMSE/RMSE $\hat{\tau}(\mathbf{p})$	1	0.940	0.958	
95% level CI	[-0.095, -0.033]		[-0.093, -0.036]	

Outline

- 1 Overview
- 2 Setup and examples
- 3 Minimax-linear estimator of τ with bounded CATEs
- 4 Minimax-linear estimator of τ with bounded heterogeneity
- 5 Minimax-length confidence intervals
- 6 Applications
- 7 Conclusion**

Take-away

- Minimax-linear estimators of, and CIs for, an average treatment effect (ATE) that can be decomposed as a weighted average of conditional average treatment effects (CATEs), under various restrictions on the CATEs:
 - Bound on their magnitude.
 - Bound on their heterogeneity.
- In two empirical applications, small but non-negligible precision gains.
- Interesting avenue for future research: power-aware minimax estimators.

Thank you!

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011). Accountability and flexibility in public schools: Evidence from boston's charters and pilots. *The Quarterly Journal of Economics* 126(2), 699–748.

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, and C. R. Walters (2010). Inputs and impacts in charter schools: Kipp lynn. *American Economic Review* 100(2), 239–43.

Armstrong, T. B. and M. Kolesár (2018a). Optimal inference in a class of regression models. *Econometrica* 86(2), 655–683.

Armstrong, T. B. and M. Kolesár (2018b). Optimal inference in a class of regression models. *Econometrica* 86(2), 655–683.

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

Behaghel, L., C. De Chaisemartin, and M. Gurgand (2017). Ready for boarding? the effects of a boarding school for disadvantaged students. *American Economic Journal: Applied Economics* 9(1), 140–164.

- Connors, A. F., T. Speroff, N. V. Dawson, C. Thomas, F. E. Harrell, D. Wagner, N. Desbiens, L. Goldman, A. W. Wu, R. M. Califf, et al. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. *Jama* 276(11), 889–897.
- Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96(1), 187–199.
- Curto, V. E. and R. G. Fryer Jr (2014). The potential of urban boarding schools for the poor: Evidence from seed. *Journal of Labor Economics* 32(1), 65–93.
- Dobbie, W. and R. G. Fryer Jr (2011). Are high-quality schools enough to increase achievement among the poor? evidence from the harlem children's zone. *American Economic Journal: Applied Economics* 3(3), 158–87.
- Donoho, D. L. (1994). Statistical estimation and optimal recovery. *The Annals of Statistics*, 238–270.
- Kwon, S. and L. Sun (2025). Estimating treatment effects under bounded heterogeneity. *arXiv preprint arXiv:2510.05454*.

Lehmann, E. L. and J. P. Romano (2005). *Testing statistical hypotheses*. Springer.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association* 89(427), 846–866.