

Semiparametric Continuous Difference-in-Differences with Causal Forests: The Case of Charter School Effect Heterogeneity

Feng Chen¹ Douglas N. Harris² Nicholas Lacoste³

October 13, 2026

¹Liaoning University

²Tulane University, National Center for Research on Education Access and Choice (REACH)

³University of Hawai'i at Mānoa, UHERO

- In an RCT, the ATE is just a regression coefficient:

$$Y_i = \alpha + \tau W_i + \beta X_i + \varepsilon_i$$

- Want heterogeneity? Interact treatment with a covariate:

$$Y_i = \alpha + \tau_1 W_i + \tau_2 W_i X_i + \beta X_i + \varepsilon_i$$

- τ_2 is the *additional* effect of treatment for units with X

- Example: $X_i \in \{0 = \text{male}, 1 = \text{female}\}$

- $\tau_1 = \text{ATE for males}$, $\tau_1 + \tau_2 = \text{ATE for females}$

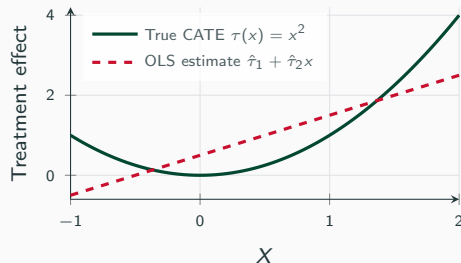
Conditional average treatment effect (CATE)

$$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x] = \tau_1 + \tau_2 x$$

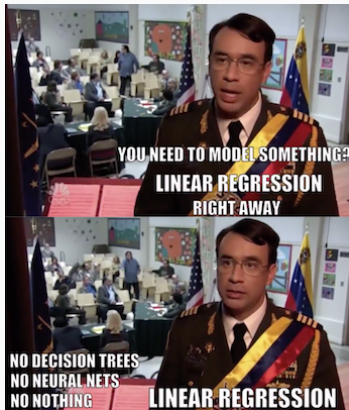
One option: keep adding interactions

$$Y_i = \alpha + \tau_1 W_i + \tau_2 W_i X_{1i} + \tau_3 W_i X_{2i} + \cdots + \beta' \mathbf{X}_i + \varepsilon_i$$

- ▶ Very restrictive: only **linear** heterogeneity
- ▶ Misses nonlinearities:
 $X_1^2, X_1^3, X_1 X_2, X_1^2 X_2, \dots$
- ▶ Wrong functional form $\Rightarrow$ OLS is **biased** for the CATE
- ▶ OLS gives the best *linear* summary of $\tau(x)$, not $\tau(x)$ itself



DGP: $X \sim U[-1, 2]$, $\tau(x) = x^2$. OLS line is the large-sample limit.



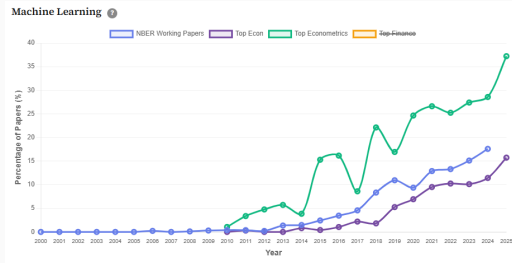
The usual microeconomist approach may not get us what we want if we want **good** estimates of CATE parameters for individual units

Semi-/nonparametric, e.g. partially linear model (Robinson, 1988):

$$Y_i = \alpha(X_i) + \tau(X_i) W_i + \varepsilon_i$$

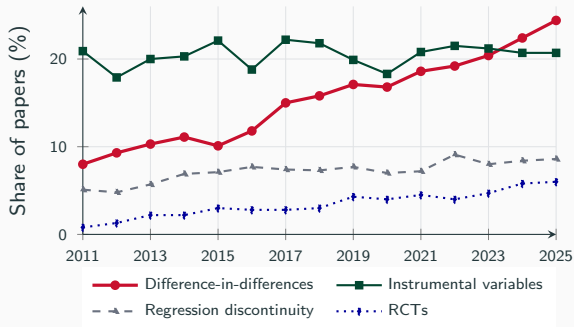
Common estimators

- ▶ Kernel / local polynomial regression
- ▶ Series & sieve estimators (splines)
- ▶ Double/debiased ML (Chernozhukov et al., 2018)
- ▶ Causal forests / GRF (Wager & Athey, 2018)
- ▶ Meta-learners (S-, T-, X-, R-learner)



Source: Paul Goldsmith-Pinkham, methods-use tracker.

ML is fast becoming standard in econometrics, much faster than in other fields of economics



Source: Paul Goldsmith-Pinkham, methods-use tracker.

- ML for heterogeneous effects: almost exclusively **RCT-like randomization**
- But DiD is ~25% of papers in top journals & NBER; RCTs are ~6%

RCT / selection on observables

Conditional independence of assignment (CIA):

$$(Y_i(1), Y_i(0)) \perp W_i \mid X_i$$

Difference-in-differences

Parallel trends in untreated outcomes:

$$\mathbb{E}[\Delta Y_{it}(0) \mid W_i = 1] = \mathbb{E}[\Delta Y_{it}(0) \mid W_i = 0]$$

- ▶ Last 5–7 years: big literature on how DiD estimators behave in different policy settings
 - **Staggered adoption:** Callaway & Sant'Anna (2021); Sun & Abraham (2021); de Chaisemartin & D'Haultfoeulle (2020); Borusyak, Jaravel & Spiess (2024)
 - **Continuous treatment:** Callaway, Goodman-Bacon & Sant'Anna (2026)
 - **Decomposing TWFE:** Goodman-Bacon (2021); Wooldridge (2021)
- ▶ Common theme: the **TWFE estimator does not generally recover the ATT** in many common quasi-experimental designs.
- ▶ A ML estimator that relies on the CIA or behaves like a TWFE estimator is likely problematic in quasi-experimental settings

Takeaway: Flexible estimators of heterogeneous treatment effects need to be adapted for QEDs.

- ▶ **Instrumental variables:** instrumental forests (Athey, Tibshirani & Wager, 2019)
- ▶ **Regression discontinuity:**
 - Honest RD trees (Reguly, 2021)
 - Bayesian causal forests for RD (Alcantara et al., 2024)
- ▶ **Difference-in-differences:**
 - DiDCF (Gavrilova, Langørgen & Zoutman, 2025): simple event-study setting, **binary** treatment, all units treated in the same period
 - Causal forests with fixed effects (Kattenberg et al., 2023)
 - ML-DiD for staggered adoption, binary treatment (Hatamyar et al., 2023)

Takeaway: There is no **causal forest** approach incorporating modern-DiD logic for **staggered adoption** or a **continuous** treatment dose.

1. We extend the DiDCF to staggered adoption settings

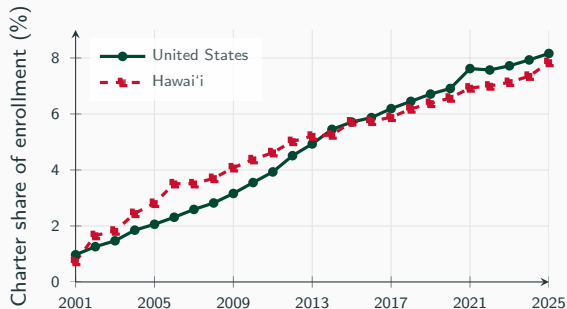
We show that, under binary treatment, the DiDCF can be applied to a staggered adoption setting and be consistent under Callaway and Sant'Anna (2021) conditions.

2. The existing DiDCF breaks under continuous treatment

With a continuous dose, the Gavrilova et al. estimator does not recover the *CATT*, in the same way that TWFE fails to recover the *ATT* (Callaway et al., 2026)

3. We generalize the DiDCF for continuous treatment

We provide a (staggered and unstaggered) DiDCF that recovers the *CATT* under continuous treatment.



Source: NCES Common Core of Data (via Urban Institute Education Data Portal).
Spring of school year.

- ▶ Serve ~8% of U.S. public school students (3.9M in 2024–25)
- ▶ 1,000+ districts in 45 states have had at least one charter
- ▶ Market shares from 0 to ~100% (New Orleans)
- ▶ Charter schools may (or may not) improve student outcomes by offering more tailored programs to serve the specific needs of local students

Average effects are positive but small

- ▶ Urban lotteries: large gains, e.g. ≈ 0.4 SD/yr in math in Boston (Abdulkadiroğlu et al., 2011)
- ▶ National lottery study: no average impact (Gleason et al., 2010)
- ▶ National matched comparison: +0.01 SD math, +0.03 SD reading per year (CREDO, 2023)
- ▶ Market level: +10 pp share $\Rightarrow$ +0.01 SD test scores, +1–2 pp graduation (Chen & Harris, 2023)

... but performance is widely dispersed

- ▶ *Market Setting*: large gains in urban areas and for disadvantaged students; ≈ 0 or negative elsewhere (Gleason et al., 2010; Angrist et al., 2013)
- ▶ *States*: charters beat local TPS in ELA in 18 of 31 states, in math just 12 (CREDO, 2023)
- ▶ *Model*: “No Excuses” practices explain $\sim 45\%$ of NYC charter effectiveness (Dobbie & Fryer, 2013); virtual charters strongly negative
- ▶ The distribution of charter school performance is wider than the TPS at both tails

- ▶ **Optimal policy design:** which districts should expand charters?
 - Are certain school districts better/worse candidates for charter schools? E.g. urban vs. rural, low- vs. high-performing, low- vs. high-income
- ▶ **Policy levers:** do funding, accountability, authorizers, or regulation change charter effectiveness?
⇒ is there an “optimal” design of charter law?
- ▶ **How much is too much?** Is there an optimal charter market share?

Takeaway: Can we predict where charters will work, and why?

1. **Small positive average effects:** +10 pp charter share $\Rightarrow$ +0.01–0.04 SD test scores, +0.5–1 pp graduation
2. **Large heterogeneity:** middle 90% of district *CATT*s spans -0.3 to $+0.3$ SD; about half are negative
3. **Baseline achievement dominates:** top variable importance for math and ELA ($2\times$ the next); effects largest where baseline scores are low. Demographics matter less, and nonlinearly
4. **Resources matter:** higher per-pupil and charter spending, lower student–teacher ratios $\Rightarrow$ larger effects
5. **State charter laws don't:** caps, accountability, exemptions carry almost no variable importance
6. **Diminishing returns:** marginal returns to additional charters are negative beyond 50-60% charter market share

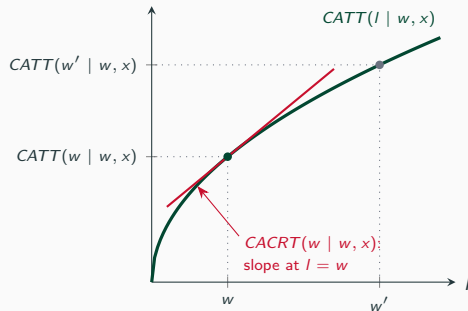
- ▶ School districts (or district-grades) $i \in \{1, \dots, N\}$, observed over time periods $t \in \{1, \dots, T\}$
- ▶ Outcomes $Y_{i,t} \in \mathbb{R}$: graduation rate (AFGR), Math, and ELA scores
- ▶ District-level covariates: $X_i \in \mathcal{X}$
- ▶ Continuous charter intensity $W_{i,t} \in [0, 1]$
- ▶ **Staggered adoption:** timing cohort $G_i \in \mathcal{G}$, with $G = \infty$ for never-treated districts
 - Always-treated excluded: $\mathcal{G} \subseteq \{2, \dots, T, \infty\}$; ever-treated cohorts $\bar{\mathcal{G}} = \mathcal{G} \setminus \{\infty\}$
- ▶ **Potential outcomes:** $Y_{i,t}(g, w)$ if first treated in cohort g under dose w
 - Untreated: $Y_{i,t}(0) = Y_{i,t}(\infty, 0)$; observed: $Y_{i,t} = Y_{i,t}(G_i, W_i)$
- ▶ **Base period** for cohort g (the period before first treatment): $b_g = g - 1$

Conditional average treatment effect on the treated

$$\begin{aligned} CATT(g, t, w, x \mid g, w, x) \\ \equiv \mathbb{E}[Y_{i,t}(g, w) - Y_{i,t}(0) \mid G_i = g, W_i = w, X_i = x] \end{aligned}$$

Conditional average causal response to treatment

$$\begin{aligned} CACRT(g, t, w, x \mid g, w, x) &\equiv \left. \frac{\partial CATT(g, t, l, x \mid g, w, x)}{\partial l} \right|_{l=w} \\ &= \left. \frac{\partial \mathbb{E}[Y_{i,t}(g, l) \mid G_i = g, W_i = w, X_i = x]}{\partial l} \right|_{l=w} \end{aligned}$$



Districts with covariates x that received dose w ; (g, t) suppressed. $CATT(w' \mid w, x)$: effect of a dose they did not receive. Adapted from Callaway, Goodman-Bacon & Sant'Anna (2026), Fig. 2

Step 1: aggregate over event times $e = t - g$ within dose groups

$$\widetilde{CATT}^{dose, es}(w, x \mid w, e, x) = \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^T \phi^{dose, es}(g, t, w, x \mid e) CATT(g, t, w, x)$$

$$\widetilde{CACRT}^{dose, es}(w, x \mid w, e, x) = \sum_{g \in \bar{\mathcal{G}}} \sum_{t=2}^T \phi^{dose, es}(g, t, w, x \mid e) CACRT(g, t, w, x)$$

$\phi^{dose, es}(g, t, w, x \mid e)$ = the share of districts with dose w that belong to cohort g

Step 2: aggregate over dose: average effect for districts with covariates x treated for e periods

$$CATT^{es}(e, x) = \mathbb{E} \left[\widetilde{CATT}^{dose, es}(W_i, x \mid W_i, e, x) \mid G_i + e \in [2, T], G_i \leq T \right]$$

$$CACRT^{es}(e, x) = \mathbb{E} \left[\widetilde{CACRT}^{dose, es}(W_i, x \mid W_i, e, x) \mid G_i + e \in [2, T], G_i \leq T \right]$$

1. Random sampling
2. Unbounded propensity scores
3. No anticipation
4. Dose support & smoothness

Assumption 5 (Conditional Parallel Trends)

For all $g \in \bar{\mathcal{G}}$, $t = 2, \dots, T$, $w \in \mathcal{W}_+$, and $x \in \mathcal{X}$:

$$\begin{aligned}\mathbb{E}[Y_{i,t}(0) - Y_{i,b_g}(0) \mid G_i = g, W_i = w, X_i = x] \\ = \mathbb{E}[Y_{i,t}(0) - Y_{i,b_g}(0) \mid G_i = \infty, W_i = 0, X_i = x]\end{aligned}$$

The expected counterfactual change in outcomes between period t and b_g for districts in cohort g who experienced dose w with covariates x , had they not been treated, is equal to the observed change in outcomes among never-treated districts, conditional on X

We can identify the *CATT* with a diff-in-diff estimator which compares the change in outcomes from b_g to t for districts in group g with dose w and covariates x to the same change in outcomes from never treated units with covariates x :

Lemma 1 (Identification of the *CATT*)

Under Assumptions 1–5, for all $g \in \bar{\mathcal{G}}$, $t \geq g$, $w \in \mathcal{W}_+$, and $x \in \mathcal{X}$:

$$\begin{aligned} CATT(g, t, w, x \mid g, w, x) = & \mathbb{E}[Y_{i,t} - Y_{i,b_g} \mid G_i = g, W_i = w, X_i = x] \\ & - \mathbb{E}[Y_{i,t} - Y_{i,b_g} \mid G_i = \infty, W_i = 0, X_i = x] \end{aligned}$$

Assumption 6 (Conditional Strong Parallel Trends)

Both of the following are true. (a) For all $g \in \bar{\mathcal{G}}$, $t = 2, \dots, T$, $w, l \in \mathcal{W}_+$, and $x \in \mathcal{X}$:

$$\begin{aligned} \mathbb{E}[Y_{i,t}(g, w) - Y_{i,b_g}(g, w) \mid G_i = g, W_i = l, X_i = x] \\ = \mathbb{E}[Y_{i,t}(g, w) - Y_{i,b_g}(g, w) \mid G_i = g, W_i = w, X_i = x] \end{aligned}$$

(b) Assumption 5 holds.

- ▶ Districts in the same $(g \times x)$ cell share a common path under any dose w
- ▶ ▶ Dose-response curve I.e. every district in cohort g with covariates x has the same dose-response path
- ▶ Assumptions 5 and 6 relax PT-MP and SPT-MP of Callaway, Goodman-Bacon & Sant'Anna (2026): they need only hold within groups of districts with the same covariate values

Lemma 2 (Identification of the *CACRT*)

Under Assumptions 1–6, for all $g \in \bar{\mathcal{G}}$, $t \geq g$, $w \in \mathcal{W}_+^c$, and $x \in \mathcal{X}$:

$$CACRT(g, t, w, x \mid g, w, x) = \left. \frac{\partial \mathbb{E}[Y_{i,t} - Y_{i,b_g} \mid G_i = g, W_i = l, X_i = x]}{\partial l} \right|_{l=w}$$

- ▶ Under Assumptions 1–6, the *CACRT*($\cdot$) is identified as the derivative of the *CATT*($\cdot$)
- ▶ Conditional analogue of Theorem C.1 of Callaway, Goodman-Bacon & Sant'Anna (2026)

Consider an unconditional TWFE model:

$$Y_{i,t} = \theta_t + \eta_i + \beta^{twfe} W_i \cdot Post_t + v_{i,t}$$

Callaway, Goodman-Bacon & Sant'Anna (2026) show that:

- ▶ **Under parallel trends:** β^{twfe} = weighted average of *ACRTs* + selection bias in dose levels
- ▶ **Under strong parallel trends:** no bias. β^{twfe} is *some* weighted average of $ACRT(\cdot)$'s, but OLS weights do not match the true dose distribution so $\beta^{twfe} \neq \mathbb{E}[ACRT(W_i) \mid W_i > 0]$
- ▶ You can never recover the *ATT* with a TWFE estimator

Callaway et al. (2026): recenter by the untreated, regress among the treated

$$\tilde{\Delta} Y_i = \Delta Y_i - \mathbb{E}_n[\Delta Y \mid W = 0], \quad \tilde{\Delta} Y_i = ATT(W_i) + \varepsilon_i \quad (W_i > 0)$$

Our extension: recenter **conditionally on X** , separately in each cohort g and period t

$$\tilde{\Delta} Y_{i,g,t} = \Delta Y_{i,g,t} - \hat{m}_{g,t}(X_i), \quad m_{g,t}(x) = \mathbb{E}[\Delta Y_{i,g,t} \mid G_i = \infty, W_i = 0, X_i = x]$$

$$\tilde{\Delta} Y_{i,g,t} = CATT(G, T, W, X) + \varepsilon_{i,g,t}$$

- ▶ Lemma 1 $\Rightarrow \mathbb{E}[\tilde{\Delta} Y_{i,g,t} \mid G_i = g, W_i = w, X_i = x] = CATT(g, t, w, x \mid g, w, x)$
- ▶ $CATT(\cdot)$ can be **linear** (CS-DiD) or **semiparametric** (DiDCF)

$$\tilde{\Delta} Y_{i,g,t} = \alpha_0 + \alpha_1 X_i + \tau_0 W_i + \tau_1 (W_i X_i) + \varepsilon_{i,g,t}$$

- Estimated on treated districts, separately in each cohort $\times$ time (g, t) cell
- $\tilde{\Delta} Y_{i,g,t} = \Delta Y_{i,g,t} - \hat{m}_{g,t}(X_i)$, with baseline conditional mean from OLS on never-treated districts:

$$\hat{m}_{g,t}(x) = \hat{\gamma}_0 + \hat{\gamma}_1 x = \hat{\mathbb{E}}[\Delta Y_{i,g,t} \mid G_i = \infty, W_i = 0, X_i = x]$$

With $\hat{\alpha}(x) = \hat{\alpha}_0 + \hat{\alpha}_1 x$ and $\hat{\tau}(x) = \hat{\tau}_0 + \hat{\tau}_1 x$:

$$\widehat{CATT}(g, t, w, x) = \hat{\alpha}(x) + \hat{\tau}(x) w$$

$$\widehat{CACRT}(g, t, w, x) = \hat{\tau}(x) \quad (\text{under Assumption 6})$$

Two-period regression between t and b_g on $\mathcal{S}_g = \{i : G_i = g\} \cup \{i : G_i = \infty\}$:

$$Y_{i,g,t} = \mu_i + \lambda_t + Post_t \left[\left(\gamma_0 + \gamma_1 X_i \right) + \mathbb{1}\{G_i = g\}(\alpha_0 + \alpha_1 X_i) + W_i(\tau_0 + \tau_1 X_i) \right] + \varepsilon_{i,g,t}$$

$Post_t = 0$ in period b_g , so the first difference between t and b_g is:

$$\Delta Y_{i,g,t} = \underbrace{(\lambda_t - \lambda_{b_g}) + \gamma_0 + \gamma_1 X_i}_{\hat{m}_{g,t}(x)} + \mathbb{1}\{G_i = g\}(\alpha_0 + \alpha_1 X_i) + W_i(\tau_0 + \tau_1 X_i) + \Delta \varepsilon_{i,g,t}$$

For a treated district, $\mathbb{1}\{G_i = g\} = 1$; subtract $\hat{m}_{g,t}(x)$ from both sides:

$$\tilde{\Delta} Y_{i,g,t} = \alpha_0 + \alpha_1 X_i + \tau_0 W_i + \tau_1 (W_i X_i) + \varepsilon_{i,g,t}$$

Takeaway: The CS-DiD is a generalized DiD with covariate-specific trends and heterogeneous level and dose effects.

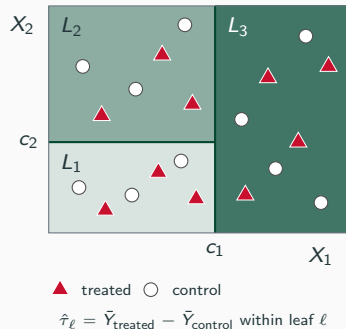
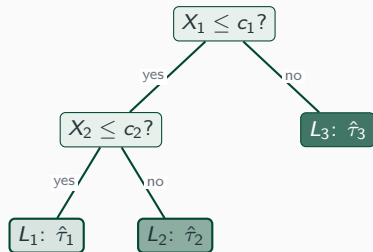
Recall the general form of the *CATT* model:

$$\tilde{\Delta} Y_{i,g,t} = CATT(G, T, W, X) + \varepsilon_{i,g,t}$$

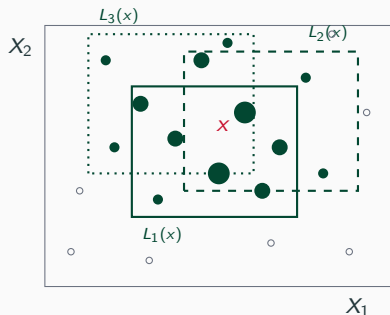
Partially linear model (Robinson 1988): coefficients are unknown functions of X

$$Y_{i,g,t} = \mu_i + \lambda_t + Post_t \mathbb{1}\{W_{i,t} > 0\} \underbrace{(\alpha(X_i) + \tau(X_i)W_{i,t})}_{CATT} + \varepsilon_{i,g,t}$$

- $CATT(g, t, w, x) = \alpha(x) + \tau(x)w$ and, under Assumption 6, $CACRT(g, t, w, x) = \tau(x)$
- Task: estimate $\{\alpha(x), \tau(x)\}$ nonparametrically



- Each split picks the X_k and cutoff c that make the effects in the two children as different as possible: $\max_{X_k, c} N_{C_1} N_{C_2} (\hat{\tau}_{C_1} - \hat{\tau}_{C_2})^2$
- **Honesty:** one half of the subsample chooses the splits, the other half estimates the effect in each leaf



Leaves containing x in three trees; dot size $\propto \omega_i(x)$

Wager & Athey (2018); Athey, Tibshirani & Wager (2019).

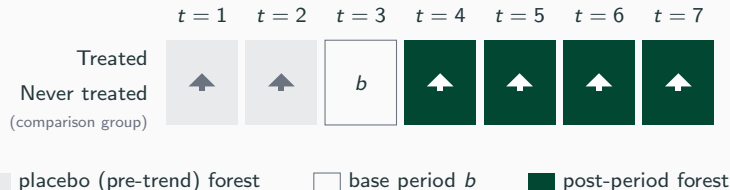
- Grow B honest trees, each on a random subsample
- Weight on district i = how often it shares a leaf with x :

$$\omega_i(x) = \frac{1}{B} \sum_{b=1}^B \frac{\mathbb{1}\{X_i \in L_b(x)\}}{|L_b(x)|}$$

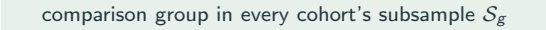
- Local, ω -weighted least squares:

$$(\hat{\alpha}(x), \hat{\tau}(x)) = \arg \min_{\alpha, \tau} \sum_i \omega_i(x) (Y_i - \alpha - \tau W_i)^2$$

- Neighbors are chosen by **where effects differ**, not by distance in X

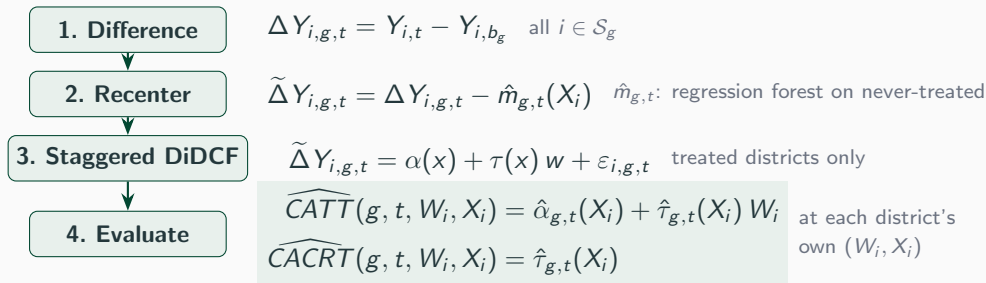


- One treated group, all treated in the same period; treatment is binary ($W_i \in \{0, 1\}$) and absorbing
- For each period $t \neq b$: difference to the base period, $\Delta Y_{i,t} = Y_{i,t} - Y_{i,b}$, then fit a causal forest of $\Delta Y_{i,t}$ on W_i and X_i (treated + never treated)
- Under conditional parallel trends, $\hat{\tau}_t(x)$ estimates $CATT(t, x) = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0) \mid W_i = 1, X_i = x]$, predicted for treated districts with trees not trained on them
- One forest per $t \approx$ forcing the first split on t ; forests for $t < b$ are placebo tests of conditional parallel trends

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Cohort $g = 3$							
Cohort $g = 4$							
Cohort $g = 5$							
Never treated							

 placebo (pre-trend) forest  base period b_g  post-period forest

- Each cell: outcome differenced to the cohort's own base period, $\Delta Y_{i,g,t} = Y_{i,t} - Y_{i,b_g}$, on $\mathcal{S}_g = \{i : G_i = g\} \cup \{i : G_i = \infty\}$
- Extends DiDCF (Gavrilova, Langørgen & Zoutman, 2025) to staggered timing
- Separate forests $\approx$ forcing the first splits on g and t : no comparisons across cohorts or periods (cf. Sun & Abraham, 2021)



- Consistent and asymptotically normal for the $CATT$ under Assumptions 1–5 (+ regularity); $\hat{\tau}$ for the $CACRT$ under Assumption 6
- Average within cohort, then across cohorts with weights $\hat{\phi}_g = n_g / \sum_{g'} n_{g'} \Rightarrow \text{event-study } \widehat{CATT}^{es}(e)$

Implementation: `grf::lm_forest`, 1,500 trees, honest half-samples, enrollment weights, clustered by district.

Test scores: SEDA

- ▶ Math & ELA, grades 3–8 (standardized)
- ▶ Unit: district $\times$ grade (66,765 units)
- ▶ Panel: 2009–2018
- ▶ 11,399 districts: 1,597 ever / 9,802 never treated

Graduation: AFGR

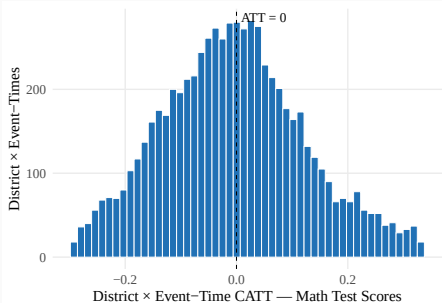
- ▶ Averaged freshman graduation rate
- ▶ Unit: district
- ▶ Panel: 1995–2009
- ▶ 10,629 districts: 1,073 ever / 9,556 never treated

Both samples

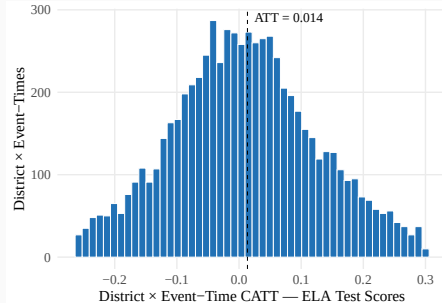
- ▶ Source: National Longitudinal School Database (CCD, SEDA, SAIPE)
- ▶ Market = geographic district; outcomes average TPS + charters $\Rightarrow$ market-level effects
- ▶ Dose $W_{i,t}$: charter enrollment share (grade-specific for SEDA), lagged one year

Category	Variables
Student demographics	% White, % Black, % Hispanic, % free/reduced-price lunch, % special education
Market setting	Baseline outcome (Y_{g-2}), urbanicity (urban, suburban, town, rural), log enrollment, magnet schools, city population, university in city
School finance	Per-pupil spending, per-pupil revenue, charter per-pupil spending, TPS–charter spending gap, teacher salary, student–teacher ratio
Charter sector	Shares of charters authorized by the district (LEA), run by a CMO, run by an EMO
State charter policy	NAPCS (2014) ratings: equitable funding, no caps on growth, performance-based contracts, transparent startup, clear renewal/nonrenewal, exemption from regulations

Math: distribution of $\widehat{CATT}$



ELA: distribution of $\widehat{CATT}$



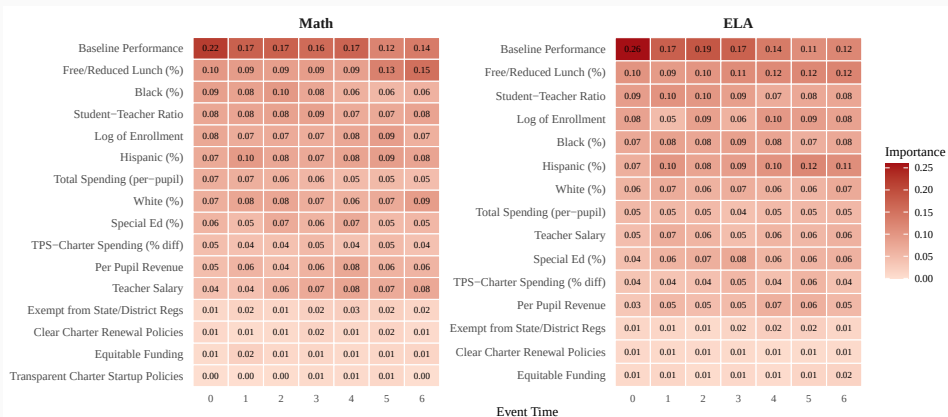
► Math: $\widehat{ATT} = 0.00002$ s.d.

► ELA: $\widehat{ATT} = 0.014$ s.d.

► Graduation (AFGR)

Which variables drive heterogeneity?

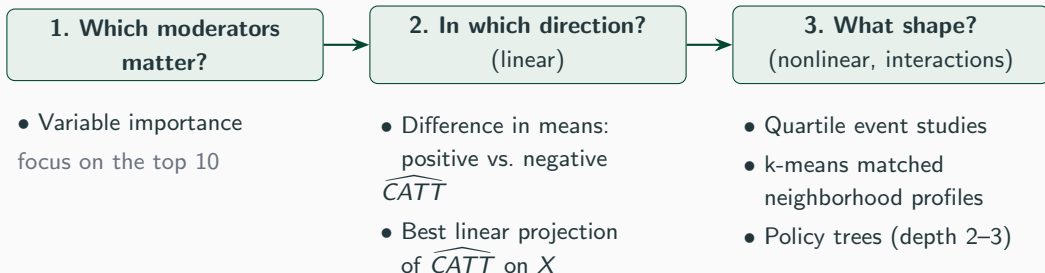
Results



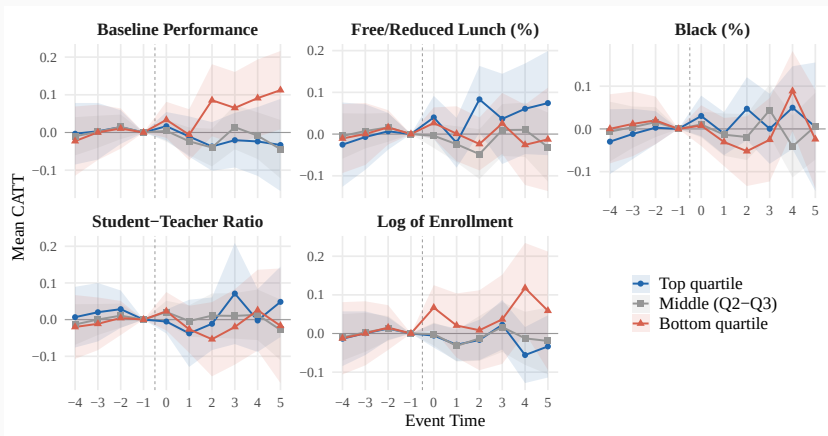
Variable importance: depth-weighted share of splits on each moderator, averaged over cohorts. Moderators with importance < 0.01 in every period not shown.

► Graduation (AFGR)

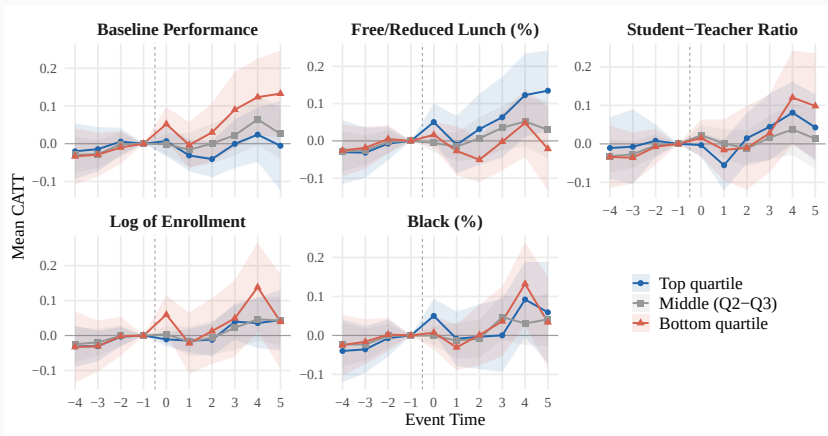
The DiDCF estimates $\widehat{CATT}(g, t, w, x)$ and $\widehat{CACRT}(g, t, w, x)$ as flexible, high-dimensional, nonlinear functions of x : easy to estimate, hard to read.



No single summary is decisive. More concrete conclusions come from patterns that are consistent across different aggregations.



Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Shaded bands: pointwise 95% confidence intervals.



Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Shaded bands: pointwise 95% confidence intervals.

► Graduation (AFGR)

Math

Moderator	$\widehat{CATT} > 0$	$\widehat{CATT} < 0$	Diff.	p
Baseline Performance	-0.31	-0.05	-0.26	<0.01
Free/Reduced Lunch (%)	0.53	0.46	0.06	<0.01
Black (%)	0.20	0.16	0.04	<0.01
Student-Teacher Ratio	15.97	16.36	-0.40	<0.01
Log of Enrollment	8.43	8.65	-0.22	<0.01
Hispanic (%)	0.21	0.18	0.03	0.02
Total Spending (\$1,000s/pupil)	12.18	11.78	0.40	0.02
White (%)	0.55	0.61	-0.06	<0.01
Special Ed (%)	0.13	0.13	0.00	0.39
TPS-Charter Spending (% diff)	0.16	0.22	-0.06	<0.01

ELA

Moderator	$\widehat{CATT} > 0$	$\widehat{CATT} < 0$	Diff.	p
Baseline Performance	-0.33	-0.14	-0.19	<0.01
Free/Reduced Lunch (%)	0.51	0.47	0.04	<0.01
Student-Teacher Ratio	16.13	16.24	-0.11	0.46
Log of Enrollment	8.46	8.65	-0.19	<0.01
Black (%)	0.18	0.17	0.01	0.40
Hispanic (%)	0.21	0.18	0.03	<0.01
White (%)	0.56	0.61	-0.05	<0.01
Total Spending (\$1,000s/pupil)	12.31	11.59	0.72	<0.01
Teacher Salary (\$1,000s)	94.97	93.29	1.68	0.07
Special Ed (%)	0.13	0.13	0.00	0.21

Mean moderator value among treated district-grades whose average post-treatment $\widehat{CATT}$ is positive vs. negative; p from Welch t -test. Top 10 moderators by variable importance at $e = 0$.

► Graduation (AFGR)

Math

Moderator	Coef.	SE	p
Baseline Performance	-0.007	0.007	0.31
Free/Reduced Lunch (%)	0.006	0.035	0.87
Black (%)	-0.013	0.050	0.80
Student-Teacher Ratio	0.000	0.002	0.90
Log of Enrollment	-0.032	0.005	<0.01
Hispanic (%)	-0.086	0.048	0.07
Total Spending (\$1,000s/pupil)	0.004	0.003	0.15
White (%)	-0.108	0.050	0.03
Special Ed (%)	-0.163	0.143	0.25
TPS-Charter Spending (% diff)	-0.007	0.024	0.77
Units	1,718		

ELA

Moderator	Coef.	SE	p
Baseline Performance	-0.005	0.006	0.43
Free/Reduced Lunch (%)	0.038	0.030	0.21
Student-Teacher Ratio	-0.004	0.002	0.03
Log of Enrollment	-0.033	0.005	<0.01
Black (%)	-0.031	0.041	0.45
Hispanic (%)	0.006	0.040	0.88
White (%)	-0.074	0.042	0.07
Total Spending (\$1,000s/pupil)	0.001	0.002	0.70
Teacher Salary (\$1,000s)	0.000	0.000	0.79
Special Ed (%)	0.061	0.118	0.60
Units	1,718		

$$\overline{CATT}_i = \beta_0 + X_i' \beta + \varepsilon_i$$

OLS, one observation per treated unit; $\overline{CATT}_i$ = unit i 's mean post-treatment $\widehat{CATT}$; X_i = all moderators plus missing-value indicators. Coefficients shown for the top 10 by variable importance at $e = 0$. ► Graduation (AFGR)

	Neighborhood (highest to lowest ATT)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline Performance	-0.68	-1.86	-0.83	-1.19	-0.72	1.41	-0.36	0.38	0.19	0.61
Free/Reduced Lunch (%)	0.64	0.81	0.49	0.70	0.65	0.19	0.65	0.42	0.40	0.31
Black (%)	0.32	0.74	0.03	0.07	0.47	0.05	0.06	0.17	0.03	0.04
Student-Teacher Ratio	13.17	14.46	18.42	21.92	14.45	15.24	15.64	14.92	14.19	19.93
Log of Enrollment	8.65	8.21	7.50	9.24	9.34	8.63	9.26	9.47	7.48	8.79
Hispanic (%)	0.34	0.05	0.16	0.67	0.10	0.06	0.71	0.16	0.05	0.12
Total Spending (\$1,000s/pupil)	23.2	12.8	10.3	11.0	11.5	13.7	10.3	10.4	12.8	10.5
White (%)	0.27	0.17	0.75	0.19	0.39	0.82	0.19	0.63	0.89	0.76
Special Ed (%)	0.16	0.14	0.12	0.10	0.14	0.13	0.10	0.11	0.15	0.12
TPS-Charter Spending (% diff)	0.30	0.11	0.59	0.10	0.14	0.20	0.04	0.11	0.16	0.40
ATT	0.08	0.06	0.04	0.02	0.01	0.00	-0.02	-0.02	-0.03	-0.05
ACRT	0.19	0.32	0.38	0.39	-0.03	0.37	0.23	0.03	0.21	0.31
Treated units	62	136	204	109	211	170	132	191	299	204

- Treated district-grades grouped into 10 neighborhoods by k-means on all standardized moderators, each weighted by its variable importance

	Neighborhood (highest to lowest ATT)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline Performance	-0.80	-1.64	1.20	-1.44	-0.90	-0.69	0.09	0.59	0.55	-0.40
Free/Reduced Lunch (%)	0.64	0.79	0.19	0.74	0.64	0.50	0.42	0.35	0.32	0.58
Student-Teacher Ratio	12.84	14.33	14.67	21.96	15.88	18.57	13.89	21.32	17.53	14.65
Log of Enrollment	8.60	8.38	8.79	9.32	9.20	7.66	7.29	9.28	8.49	9.62
Black (%)	0.31	0.69	0.07	0.06	0.06	0.04	0.03	0.06	0.04	0.38
Hispanic (%)	0.35	0.06	0.07	0.70	0.68	0.16	0.06	0.27	0.07	0.12
White (%)	0.28	0.20	0.79	0.17	0.21	0.75	0.87	0.52	0.86	0.47
Total Spending (\$1,000s/pupil)	23.7	12.5	14.3	10.9	10.3	10.6	12.3	11.5	10.4	11.1
Teacher Salary (\$1,000s)	129.1	92.8	99.5	123.6	83.5	95.5	79.0	117.1	91.4	87.3
Special Ed (%)	0.16	0.14	0.13	0.10	0.10	0.13	0.14	0.11	0.13	0.13
ATT	0.08	0.07	0.05	0.03	0.03	0.03	0.01	-0.02	-0.02	-0.02
ACRT	0.36	0.25	-0.02	0.55	0.36	0.21	0.12	0.06	0.03	-0.18
Treated units	58	188	183	92	147	196	257	93	257	247

- Treated district-grades grouped into 10 neighborhoods by k-means on all standardized moderators, each weighted by its variable importance

► Graduation (AFGR)

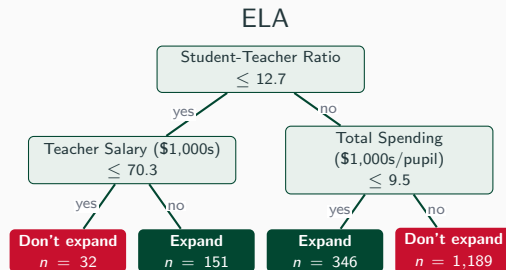
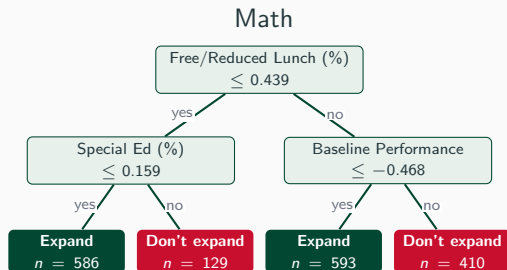
- ▶ Which districts should *marginally* expand charter intensity to raise the outcome? Decision rule $\pi : \mathcal{X} \rightarrow \{0, 1\}$, with $\pi(x) = 1$ meaning expand
- ▶ Choose π to minimize regret relative to the best rule in a class Π (Kitagawa & Tetenov, 2018):

$$R(\pi) = \max_{\pi' \in \Pi} V(\pi') - V(\pi), \quad V(\pi) = \mathbb{E}[Y_{i,t}(\pi(X_i)) - Y_{i,t}(0)]$$

- ▶ Athey & Wager (2021): estimate the value with doubly robust scores $\hat{\Gamma}_i$ for the $CACRT$ and solve

$$\hat{\pi} = \arg \max_{\pi \in \Pi} \frac{1}{n} \sum_{i=1}^n (2\pi(X_i) - 1) \hat{\Gamma}_i$$

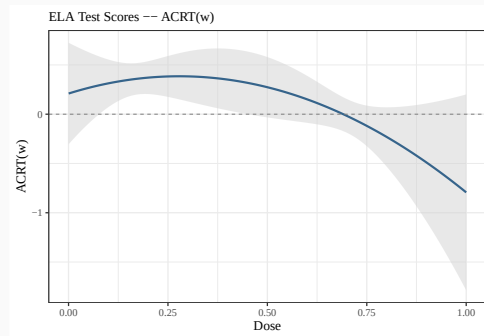
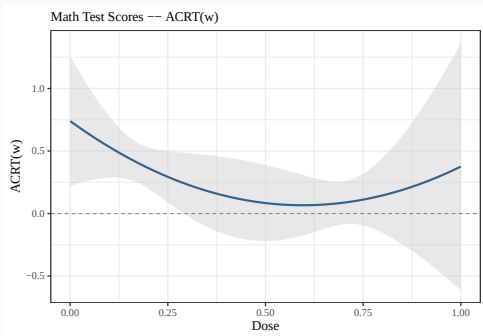
- ▶ Π = decision trees of depth 2 or 3: an exhaustive search finds the regret-minimizing tree in the class



Splits on the 10 moderators with the highest variable importance at $e = 0$; n = treated district-grades in each leaf.

► Depth-3 trees

► Graduation (AFGR)



$$\Delta Y_{i,g,t} = \psi^K(W_i)' \beta_{g,t} + \varepsilon_{i,g,t}, \quad \widehat{ACRT}_{g,t}(w) = \partial \psi^K(w)' \hat{\beta}_{g,t}$$

Treated units in each post-treatment (g, t) cell; ψ^K : cubic B-spline in the dose; cells averaged as in Chen et al. (2025). Pointwise 95% confidence bands.

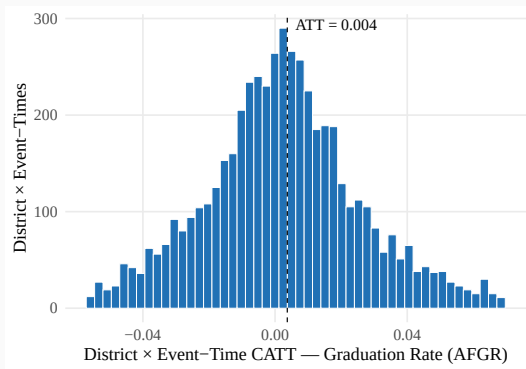
► Graduation (AFGR)

- ▶ **Method:** a DiD causal forest for staggered adoption and a continuous dose, plus a toolkit for reading the estimated $CATT$ / $CACRT$ functions
- ▶ **Heterogeneity:** average market-level effects near zero mask wide variation; district-level $\widehat{CATT}$ s span roughly ± 0.3 s.d.
- ▶ **Baseline achievement** is the most important moderator: charter expansion does more where test scores start low
- ▶ **Demographics and school spending** matter, but nonlinearly and through interactions
- ▶ **State charter-policy ratings** barely register; marginal effects shrink as charter share grows

Thank you!

nlacoste@hawaii.edu

Appendix: distribution of $\widehat{CATT}$, graduation (AFGR)

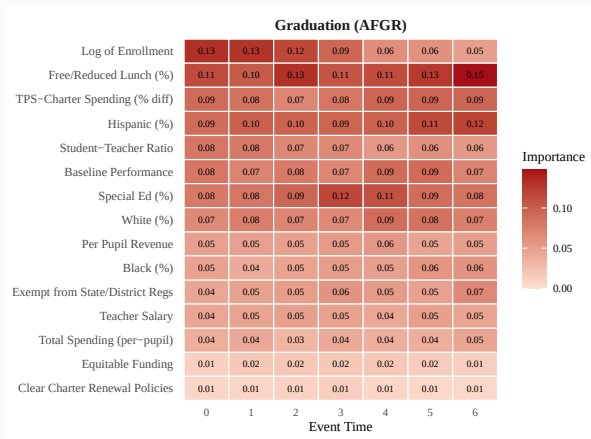


► AFGR: $\widehat{ATT} = 0.004$ (0.4 percentage points)

[◀ Back to test scores](#)

$\widehat{ATT}$ = mean of $\widehat{CATT}$ over treated districts $\times$ post-treatment years; histogram trims the top and bottom 5%.

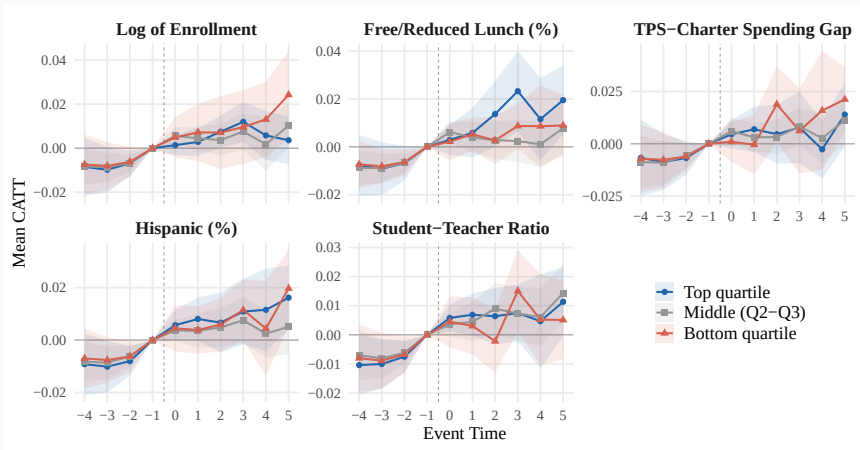
Appendix: variable importance, graduation (AFGR)



Moderators with importance < 0.01 in every period not shown.

[◀ Back to test scores](#)

Appendix: event studies by moderator quartile, graduation (AFGR)



Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Shaded bands: pointwise 95% confidence intervals.

[◀ Back to test scores](#)

Appendix: difference in means, graduation (AFGR)

Moderator	$\widehat{CATT} > 0$	$\widehat{CATT} < 0$	Diff.	p
Log of Enrollment	8.69	8.75	-0.05	0.57
Free/Reduced Lunch (%)	0.24	0.22	0.02	0.18
TPS-Charter Spending (% diff)	0.23	0.19	0.05	0.13
Hispanic (%)	0.13	0.10	0.03	0.01
Student-Teacher Ratio	18.01	17.85	0.16	0.44
Baseline Performance	0.75	0.76	-0.01	0.56
Special Ed (%)	0.10	0.11	-0.01	<0.01
White (%)	0.70	0.73	-0.03	0.13
Per Pupil Revenue (\$1,000s)	6.29	6.31	-0.02	0.90
Black (%)	0.12	0.12	0.00	0.86

Mean moderator value among treated districts whose average post-treatment $\widehat{CATT}$ is positive vs. negative; p from Welch t -test. Top 10 moderators by variable importance at $e = 0$.

[◀ Back to test scores](#)

Appendix: best linear projection, graduation (AFGR)

Moderator	Coef.	SE	<i>p</i>
Log of Enrollment	−0.001	0.001	0.63
Free/Reduced Lunch (%)	0.013	0.008	0.12
TPS-Charter Spending (% diff)	0.005	0.006	0.38
Hispanic (%)	0.048	0.011	<0.01
Student-Teacher Ratio	0.000	0.001	0.81
Baseline Performance	−0.014	0.009	0.13
Special Ed (%)	−0.035	0.020	0.08
White (%)	0.035	0.010	<0.01
Per Pupil Revenue (\$1,000s)	−0.001	0.001	0.59
Black (%)	0.025	0.011	0.02
Units			855

$$\overline{CATT}_i = \beta_0 + X_i' \beta + \varepsilon_i$$

OLS, one observation per treated district; $\overline{CATT}_i$ = district i 's mean post-treatment $\widehat{CATT}$; X_i = all moderators plus missing-value indicators. Coefficients shown for the top 10 by variable importance at $e = 0$.

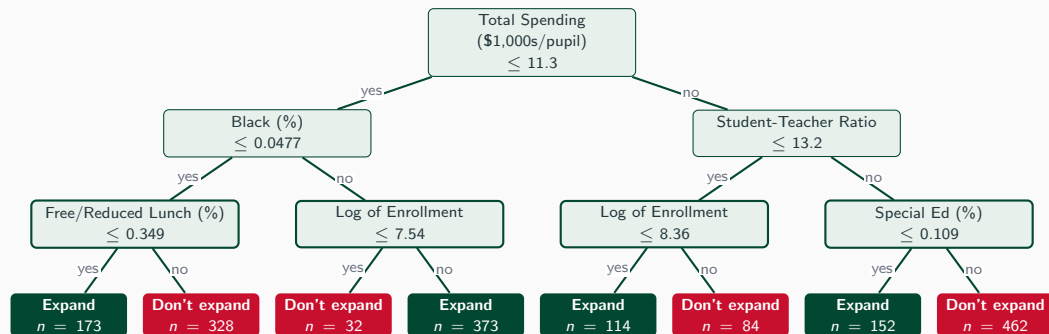
Appendix: matched neighborhoods, graduation (AFGR)

	Neighborhood (highest to lowest ATT)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of Enrollment	9.40	8.92	10.57	9.51	7.57	10.07	8.80	7.05	8.14
Free/Reduced Lunch (%)	0.58	0.46	0.42	0.27	0.14	0.18	0.13	0.38	0.11
TPS-Charter Spending (% diff)	0.06	0.27	0.32	0.19	0.19	0.12	0.40	0.05	0.14
Hispanic (%)	0.78	0.05	0.13	0.31	0.02	0.10	0.04	0.13	0.03
Student-Teacher Ratio	18.50	16.08	17.36	22.18	15.71	17.18	20.11	15.84	18.44
Baseline Performance	0.63	0.59	0.52	0.75	0.87	0.70	0.82	0.74	0.80
Special Ed (%)	0.10	0.10	0.09	0.10	0.12	0.13	0.11	0.15	0.04
White (%)	0.15	0.40	0.22	0.49	0.93	0.69	0.90	0.72	0.91
Per Pupil Revenue (\$1,000s)	5.7	6.0	7.9	5.2	6.6	5.7	6.7	8.1	5.4
Black (%)	0.04	0.46	0.62	0.08	0.02	0.18	0.03	0.04	0.04
ATT	0.015	0.008	0.005	0.005	0.000	-0.001	-0.001	-0.011	-0.014
ACRT	-0.047	-0.040	-0.012	-0.017	-0.098	0.038	0.011	-0.023	-0.028
Treated units	36	58	49	90	163	126	138	77	118

- Treated districts grouped by k-means on all standardized moderators, each weighted by its variable importance; 9 neighborhoods after folding one below 15 districts into its nearest neighbor
- Cells: average moderator value in each neighborhood; columns ordered by the neighborhood's post-treatment *ATT*

[◀ Back to test scores](#)

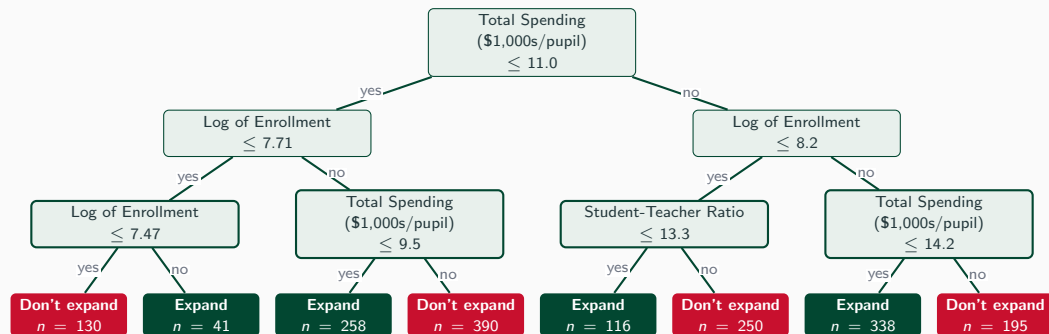
Appendix: policy tree (depth 3), math



n = treated district-grades in each leaf.

[◀ Back to depth-2 trees](#)

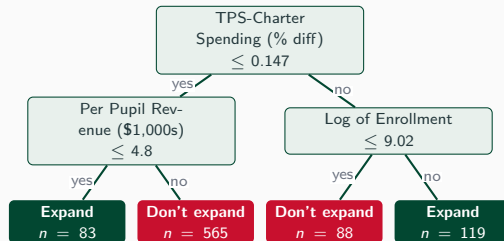
Appendix: policy tree (depth 3), ELA



n = treated district-grades in each leaf.

[◀ Back to depth-2 trees](#)

Appendix: policy tree (depth 2), graduation (AFGR)

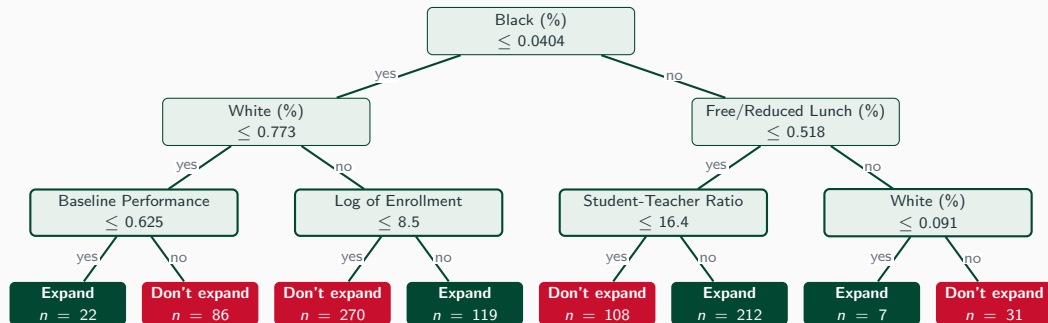


n = treated districts in each leaf.

► Depth 3

◄ Back to test scores

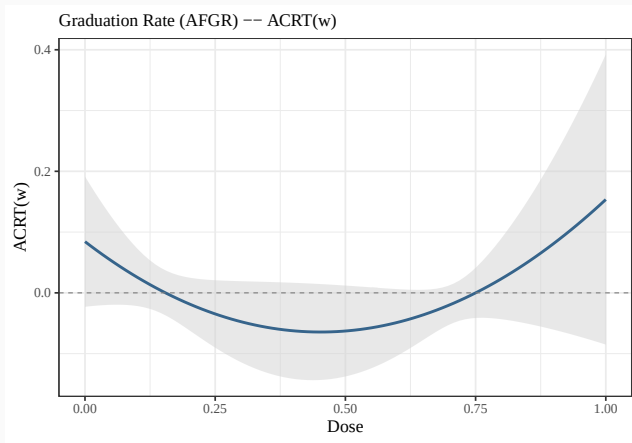
Appendix: policy tree (depth 3), graduation (AFGR)



n = treated districts in each leaf.

[◀ Back to test scores](#)

Appendix: dose response $ACRT(w)$, graduation (AFGR)

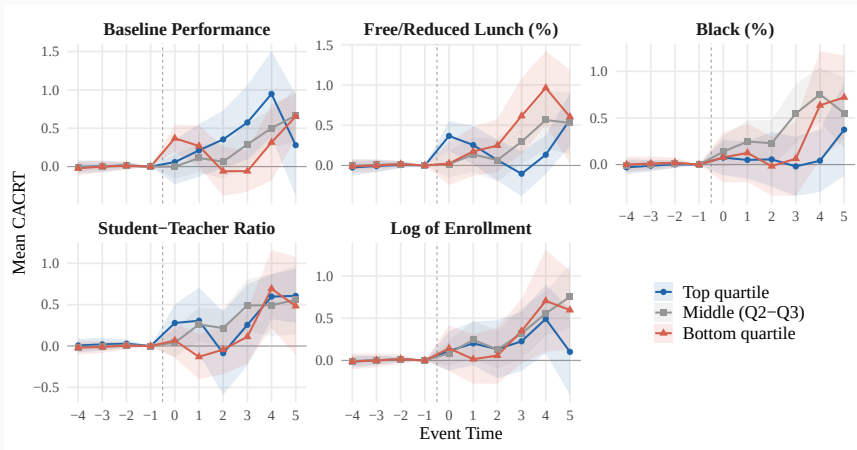


$$\Delta Y_{i,g,t} = \psi^K(W_i)' \beta_{g,t} + \varepsilon_{i,g,t}, \quad \widehat{ACRT}_{g,t}(w) = \partial \psi^K(w)' \hat{\beta}_{g,t}$$

Same construction as the test-score curves. Pointwise 95% confidence bands.

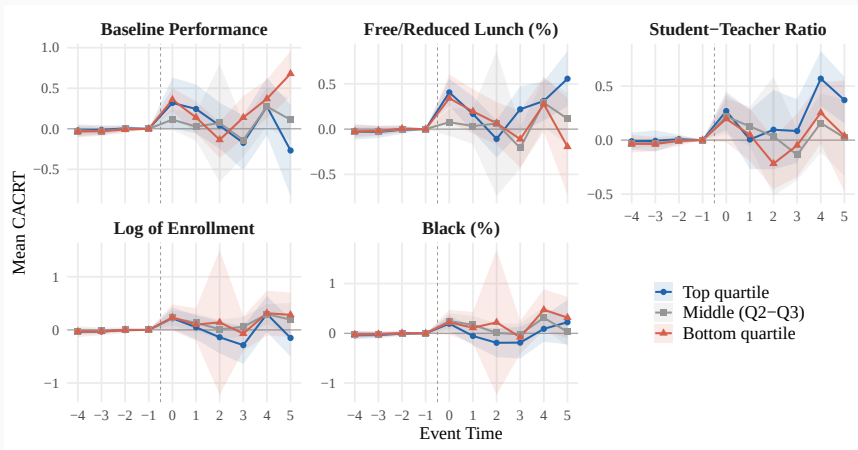
[◀ Back to test scores](#)

Appendix: CACRT event studies by moderator quartile, math



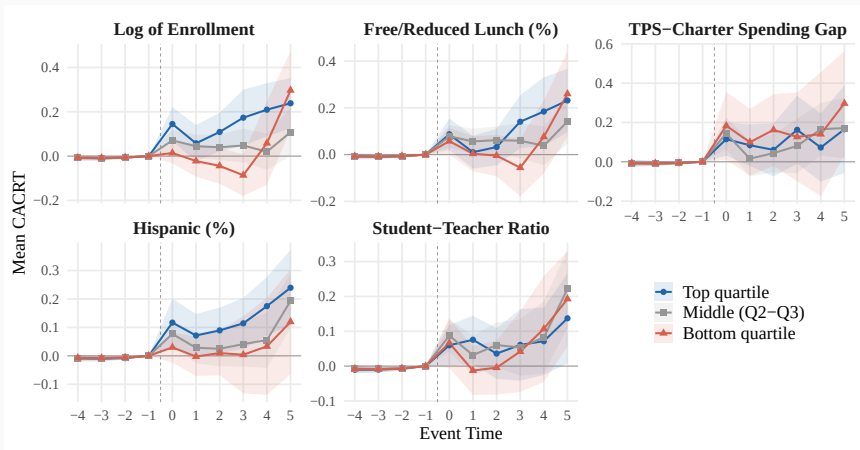
Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Pre-periods are the placebo $\widehat{CATT}$ s (no dose variation before entry). Shaded bands: pointwise 95% confidence intervals.

Appendix: CACRT event studies by moderator quartile, ELA



Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Pre-periods are the placebo $\widehat{CATT}$ s (no dose variation before entry). Shaded bands: pointwise 95% confidence intervals.

Appendix: CACRT event studies by moderator quartile, graduation (AFGR)



Treated districts split at the 25th and 75th percentiles of each moderator; five moderators with the highest variable importance at $e = 0$. Pre-periods are the placebo $\widehat{CATT}$ s (no dose variation before entry). Shaded bands: pointwise 95% confidence intervals.