ECON 730: Causal Inference with Panel Data

Lecture 4: Efficient Estimation for Staggered Rollout Designs

Pedro H. C. Sant'Anna

EMORY UNIVERSITY

Spring 2026

Motivation: Application

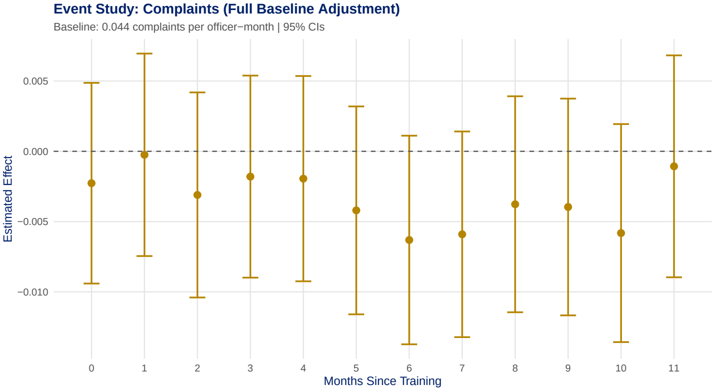
Application: Reducing Police Misconduct

- Reducing police misconduct and the use of force is an important policy objective
- Wood, Tyler and Papachristos (2020a) (PNAS): Randomized rollout of procedural justice training for Chicago police officers
  - Emphasized respect, neutrality, and transparency in the exercise of authority
- Original study found **large and significant** reductions in complaints/use of force
- Wood, Tyler, Papachristos, Roth and Sant'Anna (2020b): a re-analysis using Callaway and Sant'Anna (2021)'s full-adjustment approach ($\beta = 1$):
  - No significant impacts on complaints; borderline effects on force; **wide CIs**

> Two natural questions: Is the sampling-based approach to inference adequate? How should we *optimally* use pre-treatment data when treatment timing is random?

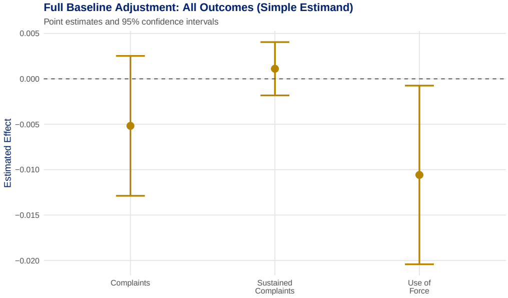
The Precision Problem: Event Study

Re-analysis by Wood et al. (2020b) using full baseline adjustment ($\beta = 1$):



- Officers averaged **0.044 complaints/month** before training
- Estimates suggest ≈11% reduction — but confidence intervals are **wide**

The Precision Problem: All Outcomes



- Wide confidence intervals across **all three outcomes**
- Only use of force is (barely) significant

> Is this imprecision *inherent* to the data, or is there a **more efficient** way to use pre-treatment information?

How should we optimally use pre-treatment data when treatment timing is random?

By finding the efficient adjustment weight.

In Roth and Sant'Anna (2023), we introduce a design-based framework and derive the **efficient estimator** for staggered rollout designs.

Preview of Results

- Roth and Sant'Anna (2023) introduce a **design-based framework** formalizing random treatment timing
- Consider a **large class of causal parameters** aggregating effects across periods and cohorts
- Solve for the **efficient estimator** in a class of pre-treatment adjustment estimators

Key results:

- SE reductions of **2**$\times$ **or more** in Monte Carlo simulations and applications
- Both t-based and **Fisher Randomization Test** inference
- Implemented in the **R** package `staggered`

Roadmap

1. **Framework:** Staggered rollout with random treatment timing
2. **Special Case:** Two-period model (building intuition)
3. **Causal Parameters:** Aggregations of $ATE(g, t)$
4. **Class of Estimators:** $\hat{\theta}_\beta = \hat{\theta}_0 - \hat{X}'\beta$
5. **The Efficient Estimator:** β^* and plug-in version
6. **Inference:** t-based CIs and Fisher Randomization Tests
7. **R Implementation:** The `staggered` package
8. **Application:** Police training revisited

From Experiments to Staggered Designs

- **Lecture 3:** We studied *randomized experiments* with design-based inference
  - Known treatment probabilities $\rightarrow$ unbiased Horvitz-Thompson estimation
  - No use of pre-treatment data for efficiency
- **This lecture:** What if *treatment timing* is the random variable?
  - Staggered rollout: units are treated at different times
  - Pre-treatment outcomes are available — should we use them?

The core question: In a randomized experiment, should you adjust for baseline covariates? How much? This is the classical **regression adjustment** debate (Freedman, 2008; Lin, 2013). Here, the "covariate" is the *pre-treatment outcome.*

Framework

Framework

- **Finite population:** N units, T periods ($i = 1, \ldots, N$; $t = 1, \ldots, T$)
- **Treatment timing:** Unit i first treated at $G_i \in \mathcal{G} \subset \{1, \ldots, T\} \cup \{\infty\}$
  - $G_i = \infty$: never treated. Treatment is **absorbing** (no switching on/off)
- **Potential outcomes:** $Y_{i,t}(g)$ = outcome for unit i in period t if first treated at g
- **Observed outcomes:** $Y_{i,t} = \sum_g \mathbf{1}[G_i = g]\, Y_{i,t}(g)$

Design-based framework: Potential outcomes $Y_{i,t}(\cdot)$ and cohort sizes $N_g = \sum_i \mathbf{1}[G_i = g]$ are *fixed*; only the treatment assignment G is stochastic.

Connection to Lecture 2: *Same $Y_{i,t}(g)$ notation.* ***Connection to Lecture 3:*** *Same design-based perspective, but now with absorbing staggered treatment.*

Design-Based Uncertainty

Randomization Inference: potential outcomes are **fixed**; only treatment timing G is **permuted**

	Actual Sample				Alternative Sample I				Alternative Sample II				...
Unit	$Y_{i,t}(2)$	$Y_{i,t}(3)$	$Y_{i,t}(4)$	G_i	$Y_{i,t}(2)$	$Y_{i,t}(3)$	$Y_{i,t}(4)$	G_i	$Y_{i,t}(2)$	$Y_{i,t}(3)$	$Y_{i,t}(4)$	G_i	...
1	?	?	✓	4	?	✓	?	3	?	?	✓	4	...
2	✓	?	?	2	?	✓	?	3	?	✓	?	3	...
3	?	?	✓	4	?	?	✓	4	✓	?	?	2	...
4	?	✓	?	3	✓	?	?	2	?	?	✓	4	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
N	✓	?	?	2	✓	?	?	2	?	?	✓	3	...

Potential outcomes are the same across samples. Only G_i (and hence which POs we observe) changes. Contrast with sampling-based inference, where different units are drawn.

Assumption 1: Random Treatment Timing

Random Treatment Timing

Let $G = (G_1, \ldots, G_N)$. Then:

$$\Pr(G = \tilde{g}) = \frac{\prod_{g \in \mathcal{G}} N_g!}{N!} \quad \text{if} \sum_i \mathbf{1}[\tilde{g}_i = g] = N_g \text{ for all } g, \text{ and zero otherwise.}$$

- **Interpretation:** Any permutation of treatment timing preserving cohort sizes is equally likely
- **Examples of plausible random timing:**
  - By design: randomized rollout of police training (Wood et al., 2020a)
  - Quasi-random: timing of parental deaths (Druedahl and Martinello, 2022), health shocks (Fadlon and Nielsen, 2021), office closings (Deshpande and Li, 2019), stimulus payments (Parker, Souleles, Johnson and McClelland, 2013)

Assumption 2: No Anticipation

No Anticipation

For all units i, periods t, and treatment dates $g, g' > t$:

$$Y_{i,t}(g) = Y_{i,t}(g')$$

- **Interpretation:** Outcomes before treatment do not depend on *when* treatment will start
- **Connection to Lecture 3:** Same no-anticipation assumption. Under this assumption, we write $Y_{i,t}(\infty)$ for the "untreated" potential outcome
- **Caveat:** May fail if treatment timing is announced in advance (Malani and Reif, 2015)

Assessing Random Treatment Timing

- Random treatment timing is a **strong assumption** — how can we assess its plausibility?

- **Key idea:** Under random treatment timing + no anticipation, pre-treatment outcomes should be **balanced** across cohorts (just like covariate balance in RCTs!)

- Roth and Sant'Anna (2023) propose pre-tests to assess plausibility of random treatment timing
  - Balance checks implemented in **R** package `staggered`

- If random treatment timing fails, weaker assumptions (covered in later lectures) can still justify causal inference — but without the efficiency gains derived here

> **Bottom line:** Random treatment timing provides *additional structure* beyond basic unconfoundedness. When plausible, it enables more efficient estimation. Balance checks help assess its credibility.

Special Case: Two-Period Model

Two-Period Model: Building Intuition

- **Setup:** $T = 2$, $\mathcal{G} = \{2, \infty\}$. Some units treated in period 2, rest never treated.

- Under randomization + no anticipation, this is analogous to a **cross-sectional randomized experiment**:

Outcome:	$Y_i = Y_{i,t=2}$	(post-treatment outcome)
Covariate:	$X_i = Y_{i,t=1} \equiv Y_{i,t=1}(\infty)$	(pre-treatment outcome)
Treatment:	$D_i = \mathbf{1}[G_i = 2]$	(treatment indicator)

We use the two-period case as a *running example* to build intuition before moving to the general staggered case.

Two-Period: Target Parameter and Estimator Class

Target: $\theta = \frac{1}{N}\sum_{i=1}^{N}[Y_{i,2}(2) - Y_{i,2}(\infty)]$

Class of estimators:

$$\hat{\theta}_\beta = \underbrace{(\bar{Y}_{22} - \bar{Y}_{2\infty})}_{\text{Post-treatment diff}} - \beta \underbrace{(\bar{Y}_{12} - \bar{Y}_{1\infty})}_{\text{Pre-treatment diff}}$$

where $\bar{Y}_{sg} = N_g^{-1}\sum_{i:\, G_i = g} Y_{i,s}$ is the period-s sample mean for cohort g, and $\beta \in \mathbb{R}$.

$\beta = 0$:	**No adjustment** (difference-in-means)	*ignores pre-treatment data*
$\beta = 1$:	**Full baseline adjustment**	*subtracts entire pre-treatment diff*
$\beta \in (0,1)$:	**Partial adjustment**	*intermediate correction*
$\beta < 0$ or $\beta > 1$:	**Extrapolation**	*also allowed; optimal β^* may lie here*

Isomorphic to **regression adjustment in experiments** (Freedman, 2008; Lin, 2013). The pre-treatment outcome $Y_{i,1}$ serves as the covariate.

Causal Parameters

Building Block: Average Treatment Effects

The **average effect** of switching treatment timing from g' to g at period t:

$$\tau_{t,gg'} = \frac{1}{N} \sum_{i=1}^{N} \left[Y_{i,t}(g) - Y_{i,t}(g') \right]$$

General Scalar Estimand

$$\theta = \sum_{t,g,g'} a_{t,gg'} \, \tau_{t,gg'}$$

where $a_{t,gg'} \in \mathbb{R}$ are known weights with $a_{t,gg'} = 0$ if $t < \min(g, g')$.

Connection to Lecture 2: $\tau_{t,g\infty} = ATE(g,t)$ is the group-time treatment effect. All aggregation parameters from Lecture 2 are special cases of θ.

Aggregation Parameters

Simple weighted average: $\theta^{simple} = \frac{1}{\sum_t \sum_{g \le t} N_g} \sum_t \sum_{g \le t} N_g \cdot ATE(g,t)$

Calendar-time aggregate: $\theta^{calendar} = \frac{1}{T} \sum_t \theta_t, \quad \text{where } \theta_t = \frac{1}{\sum_{g \le t} N_g} \sum_{g \le t} N_g \cdot ATE(g,t)$

Cohort aggregate: $\theta^{cohort} = \frac{1}{\sum_{g \ne \infty} N_g} \sum_{g \ne \infty} N_g \, \theta_g, \quad \text{where } \theta_g = \frac{1}{T-g+1} \sum_{t \ge g} ATE(g,t)$

Event-study (lag l): $\theta_l^{ES} = \frac{1}{\sum_{g:g+l \le T} N_g} \sum_{g:g+l \le T} N_g \cdot ATE(g, g+l)$

*These are the same aggregation parameters from **Lecture 2** (Callaway and Sant'Anna, 2021). Here we apply them to the design-based finite-population setting.*

Class of Estimators

The Class of Estimators

Estimator Class

Let $\bar{Y}_{tg} = N_g^{-1} \sum_i \mathbf{1}[G_i = g]\, Y_{i,t}$ be the sample mean for cohort g at period t, and $\hat{\tau}_{t,gg'} = \bar{Y}_{tg} - \bar{Y}_{tg'}$.

Plug-in estimator: $\hat{\theta}_0 = \sum_{t,g,g'} a_{t,gg'}\, \hat{\tau}_{t,gg'}$

General class:

$$\hat{\theta}_\beta = \hat{\theta}_0 - \hat{X}'\beta$$

where $\hat{X}$ is an M-vector of **pre-treatment comparisons:** $\hat{X}_j = \sum_{(t,g,g'):g,g'>t} b^j_{t,gg'}\, \hat{\tau}_{t,gg'}$

Under no anticipation, $\mathbb{E}[\hat{X}] = 0$. The vector $\hat{X}$ compares cohorts *before either was treated* — any pre-treatment difference is "noise" from randomization.

Existing Estimators as Special Cases

Callaway and Sant'Anna (2021): For estimating $ATE(g, t)$ using never-treated:

$$\hat{\tau}^{CS}_{t,g} = \underbrace{(\bar{Y}_{tg} - \bar{Y}_{t\infty})}_{\hat{\theta}_0:\ \text{post-treatment diff}} - \underbrace{(\bar{Y}_{g-1,g} - \bar{Y}_{g-1,\infty})}_{\hat{X}:\ \text{pre-treatment diff}}$$

This is $\hat{\theta}_\beta$ with $\beta = 1$: **full baseline adjustment** using period $g-1$ as the pre-treatment reference.

Other estimators in the class:

- Sun and Abraham (2021): Last-treated cohort as comparison ($\beta = 1$)
- de Chaisemartin and D'Haultfœuille (2020): Equivalent to CS for instantaneous event-study ($\beta = 1$)
- **TWFE**: Also in the class (Athey and Imbens, 2021), but with potentially **negative weights**

> **Key observation:** All existing approaches use $\beta = 1$ (or fixed). None optimize over β!

The Efficient Estimator

Unbiasedness: All β Work

Proposition (Unbiasedness — Lemma 2.1)

Under Assumptions 1 (Random Treatment Timing) and 2 (No Anticipation):

$$\mathbb{E}[\hat{\theta}_\beta] = \theta \quad \text{for all } \beta \in \mathbb{R}^M$$

Proof sketch:

1. Under randomization: $\mathbb{E}[\bar{Y}_{tg}] = \mathbb{E}_{fin}[Y_{i,t}(g)]$, so $\mathbb{E}[\hat{\theta}_0] = \theta$
2. Under no anticipation: $\mathbb{E}[\hat{X}] = 0$ (comparing cohorts pre-treatment)
3. Therefore: $\mathbb{E}[\hat{\theta}_\beta] = \mathbb{E}[\hat{\theta}_0] - \mathbb{E}[\hat{X}]'\beta = \theta - 0 = \theta \quad \square$

Since all β give **unbiased** estimators, we are free to choose β to **minimize variance**.

The Efficient β^*

Proposition 2.1: The variance of $\hat{\theta}_\beta$ is uniquely minimized at:

$$\beta^* = \text{Var}[\hat{\text{X}}]^{-1}\,\text{Cov}[\hat{\text{X}}, \hat{\theta}_0]$$

Proof: This is just **OLS**! We are minimizing

$$\text{Var}[\hat{\theta}_0 - \hat{\text{X}}'\beta] = \text{Var}\big[(\hat{\theta}_0 - \theta) - (\hat{\text{X}} - 0)'\beta\big]$$

over β. The solution is the best linear predictor of $\hat{\theta}_0$ given $\hat{\text{X}}$.

Intuition: Adjust more for pre-treatment differences when they are *more predictive* of post-treatment differences. The optimal β^* balances full adjustment ($\beta = 1$) and no adjustment ($\beta = 0$).

Key Property: β^* Depends Only on Estimable Quantities

Joint variance structure:

$$\text{Var}\begin{pmatrix}\hat{\theta}_0 \\ \hat{X}\end{pmatrix} = \begin{pmatrix} V_{\theta_0} & V_{\theta_0,X} \\ V_{X,\theta_0} & V_X \end{pmatrix}$$

where the components involve:

- $S_g = \text{Var}_{fin}[\mathbf{Y}_i(g)]$: finite-population variance for cohort g — **estimable**
- $S_\theta = \text{Var}_{fin}\left[\sum_g A_{\theta,g}\mathbf{Y}_i(g)\right]$: cross-PO variance — **not estimable**

> **Critical insight:** $\beta^* = V_X^{-1}V_{X,\theta_0}$ depends only on S_g (estimable!), not on S_θ. This is because V_X and V_{X,θ_0} involve only *within-cohort* covariances.

We estimate S_g with the within-cohort sample covariance $\hat{S}_g = (N_g - 1)^{-1}\sum_i \mathbb{1}[G_i = g](\mathbf{Y}_i - \bar{\mathbf{Y}}_g)(\mathbf{Y}_i - \bar{\mathbf{Y}}_g)'$.

Two-Period Efficient β^*: Intuition

In the two-period case:

$$\beta^* = \frac{N_\infty}{N}\beta_\infty + \frac{N_2}{N}\beta_2$$

where β_g is the regression coefficient from regressing $Y_{i,2}(g)$ on $Y_{i,1}$ (plus constant).

Scenario	β^*	**Optimal estimator**
$Y_{i,2}(g)$ uncorrelated with $Y_{i,1}$	≈ 0	No adjustment
High autocorrelation ($\beta_g \approx 1$)	≈ 1	Full adjustment
Intermediate autocorrelation	$\in (0,1)$	Partial adjustment
Mean reversion ($\beta_g < 0$)	< 0	*Reverse adjustment*
Low autocorrelation + heterogeneity	> 1	*Over-adjustment*

Connection to experiments: This is exactly the Lin (2013) result for covariate adjustment in randomized experiments, with $Y_{i,1}$ as the "covariate."

Connection: From Lecture 3 to Lecture 4

- **Lecture 3:** Design-based Horvitz-Thompson estimation
  - Known propensity scores $\rightarrow$ unbiased estimation via IPW
  - *No covariate adjustment*

- **Lecture 4:** Design-based estimation *with covariate adjustment*
  - Pre-treatment outcomes serve as "covariates"
  - $\hat{\theta}_0$ is a Horvitz-Thompson-type estimator
  - $\hat{X}'\beta$ is the covariate adjustment
  - Choosing β^* optimally $\rightarrow$ efficient estimation

Bridge: Lecture 3 established that HT-type estimators are unbiased under randomization. This lecture asks: among all unbiased estimators, which is *most precise*?

Plug-In Estimator

The Plug-In Efficient Estimator

- **Problem:** β^* depends on unknown $S_g = \text{Var}_{fin}[\mathbf{Y}_i(g)]$
- **Solution:** Replace S_g with sample analogue $\hat{S}_g$, compute $\hat{\beta}^*$
- **Plug-in estimator:** $\hat{\theta}_{\hat{\beta}^*} = \hat{\theta}_0 - \hat{X}'\hat{\beta}^*$

Regularity Conditions

As $N \to \infty$: (i) Cohort shares converge: $N_g/N \to p_g \in (0,1)$; (ii) Variances S_g have positive definite limits; (iii) Lindeberg condition holds.

Proposition 2.2: $\sqrt{N}(\hat{\theta}_{\hat{\beta}^*} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_*^2)$. **No efficiency loss** from estimating β^* — the plug-in achieves the same asymptotic variance as the oracle.

Inference

Variance Estimation and t-Based Inference

- **Challenge:** The variance of $\hat{\theta}_{\beta*}$ contains $S_\theta = \text{Var}_{fin}\left[\sum_g A_{\theta,g}\mathbf{Y}_i(g)\right]$, which depends on covariances of potential outcomes **never observed together**

- **Conservative approach:** Neyman-style variance estimator
  - Ignores S_θ (treats it as zero), replaces S_g with $\hat{S}_g$
  - Conservative: overestimates variance $\rightarrow$ valid but wide CIs

- **Less conservative:** Roth and Sant'Anna (2023) show how to estimate the part of S_θ explained by $\hat{X}$
  - Tighter CIs while remaining conservative

> ***t*-based CI:** $\hat{\theta}_{\hat{\beta}*} \pm z_{1-\alpha/2} \cdot \hat{\sigma}_{**}/\sqrt{N}$, where $\hat{\sigma}^2_{**}$ converges to an upper bound on the true variance.

What about Fisher Randomization Tests?

Fisher Randomization Tests

Under random treatment timing, we can construct **permutation tests**:

1. Compute the studentized test statistic $t = \hat{\theta}_{\hat{\beta}^*} / \widehat{se}$
2. For each permutation π of G, compute t^π
3. p-value $= \Pr_\pi(|t^\pi| \geq |t_{obs}|)$

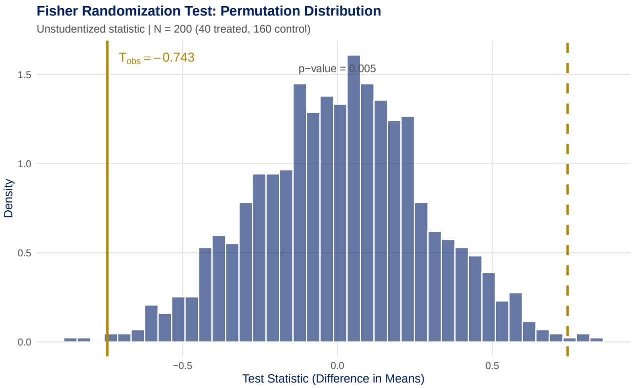
Proposition 2.3: The studentized FRT is:

1. **Exact under sharp null** (H_0^S: $Y_i(g) = Y_i(g')$ for all i, g, g')
2. **Asymptotically valid under weak null** (H_0^W: $\theta = 0$)

Studentization is essential! Without it, FRTs may not have correct size for the weak null (Wu and Ding, 2020; Zhao and Ding, 2020).

Visualizing Fisher Randomization Tests

- Illustration with simulated synthetic data
- Under the null, the observed assignment is "just another permutation"
- **Procedure:** (1) Compute T_{obs} from actual data (2) Re-shuffle treatment $\rightarrow$ recompute T^{π} (3) p-value $= \frac{\#\{|T^{\pi}| \geq |T_{\text{obs}}|\}}{B}$

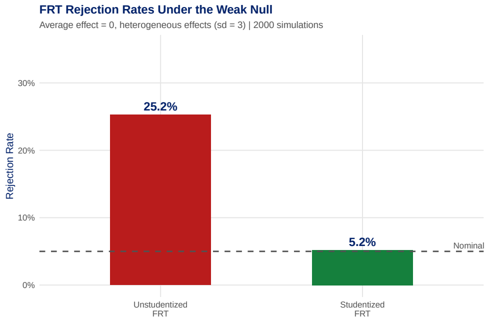


Why Studentization Matters

Simulation with synthetic data.

Sharp null (H_0^S): $Y_i(g) = Y_i(g')$ for all i, g, g' $\rightarrow$ FRT is exact

Weak null (H_0^W): $\theta = 0$ $\rightarrow$ FRT may over-reject!



Under heterogeneous effects, unstudentized FRTs have **severe size distortions**. *Studentization restores correct size* (Wu and Ding, 2020; Zhao and Ding, 2020).

Simulation: `frt_simulation.R`

FRT for Callaway and Sant'Anna Estimators

- The randomization-based approach can also be applied to Callaway and Sant'Anna (2021) estimators
- Recall: CS is a special case with $\beta = 1$
- Thus, one can use FRTs to conduct "design-based" inference with CS estimators
- Rambachan and Roth (2025): FRT results are likely conservative even without full random timing

Practical implication: Even if you use CS estimators (under weaker assumptions discussed in later lectures), you can supplement with FRT-based *p*-values when random timing is plausible. The `staggered` package implements this.

R Implementation

R Package: staggered

Installation: `install.packages("staggered")`

Main functions:

- `staggered()`: Efficient estimator (Roth and Sant'Anna, 2023)
- `staggered_cs()`: Callaway and Sant'Anna (2021) estimator
- `staggered_sa()`: Sun and Abraham (2021) estimator

Built-in dataset: `pj_officer_level_balanced` (police training application)

```
library(staggered)
data(pj_officer_level_balanced)
# 5,537 officers, 72 periods, 47 cohorts
# Variables: uid, period, first_trained,
#            complaints, sustained, force
```

Basic Usage: Efficient Estimator

```r
# Simple average treatment effect
staggered(df = pj_officer_level_balanced,
          i = "uid", t = "period",
          g = "first_trained", y = "complaints",
          estimand = "simple")
```

Output: Returns `estimate`, `se`, `se_neyman`

```r
# With Fisher Randomization Test
staggered(df = pj_officer_level_balanced,
          i = "uid", t = "period",
          g = "first_trained", y = "complaints",
          estimand = "simple",
          compute_fisher = TRUE,
          num_fisher_permutations = 500)
```

Additional output: `fisher_pval` (permutation p-value)

Multiple Estimands

```r
# Calendar-time weighted average
staggered(df = pj_officer_level_balanced,
          i = "uid", t = "period",
          g = "first_trained", y = "complaints",
          estimand = "calendar")

# Cohort-weighted average
staggered(..., estimand = "cohort")

# Event-study: effects at lags 0 through 11
staggered(df = pj_officer_level_balanced,
          i = "uid", t = "period",
          g = "first_trained", y = "complaints",
          estimand = "eventstudy",
          eventTime = 0:11)
```

Event-study returns one row per event time, each with its own estimate and SE.

Comparing Estimators

```r
# Efficient estimator (Roth & Sant'Anna)
res_eff <- staggered(df = pj_officer_level_balanced,
    i = "uid", t = "period", g = "first_trained",
    y = "complaints", estimand = "simple")

# Callaway & Sant'Anna (2021)
res_cs <- staggered_cs(df = pj_officer_level_balanced,
    i = "uid", t = "period", g = "first_trained",
    y = "complaints", estimand = "simple")

# Sun & Abraham (2021)
res_sa <- staggered_sa(df = pj_officer_level_balanced,
    i = "uid", t = "period", g = "first_trained",
    y = "complaints", estimand = "simple")

# Unadjusted (no pre-treatment adjustment)
res_unadj <- staggered(df = pj_officer_level_balanced,
    i = "uid", t = "period", g = "first_trained",
    y = "complaints", estimand = "simple",
    beta = 1e-16, use_DiD_A0 = TRUE)
```

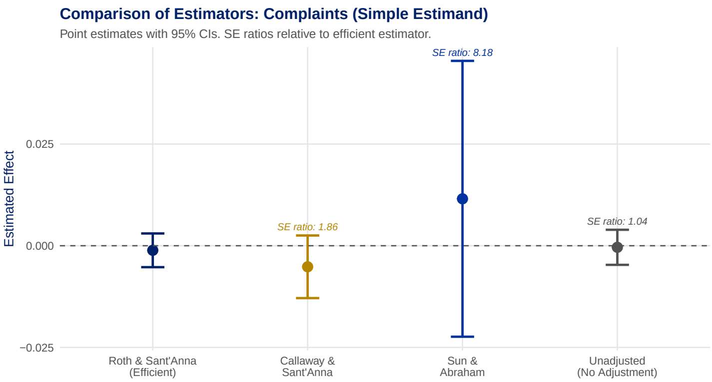
Application: Police Training Revisited

Application: Procedural Justice Training

- **Setting:** Wood et al. (2020a) — randomized rollout of procedural justice training to Chicago police officers
- **Data:** 5,537 officers, 72 monthly periods, 47 treatment cohorts (cohort sizes: 3–575)
- **Outcomes:**
  - Complaints against officers
  - Sustained complaints
  - Officer use of force
- **Treatment timing:** Randomized by design $\rightarrow$ random treatment timing holds

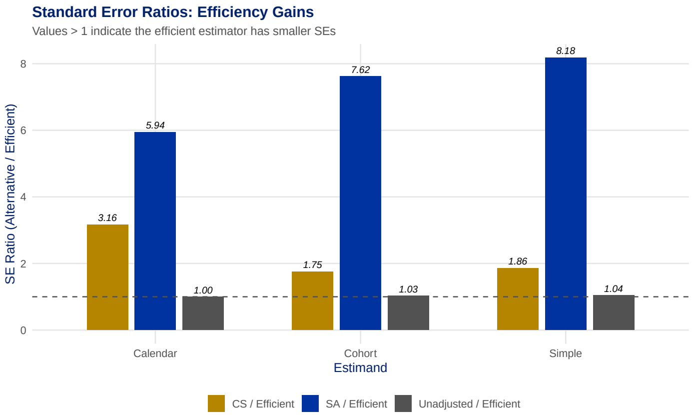
Comparison: Efficient estimator vs. Callaway-Sant'Anna vs. Sun-Abraham

Application Results: Complaints



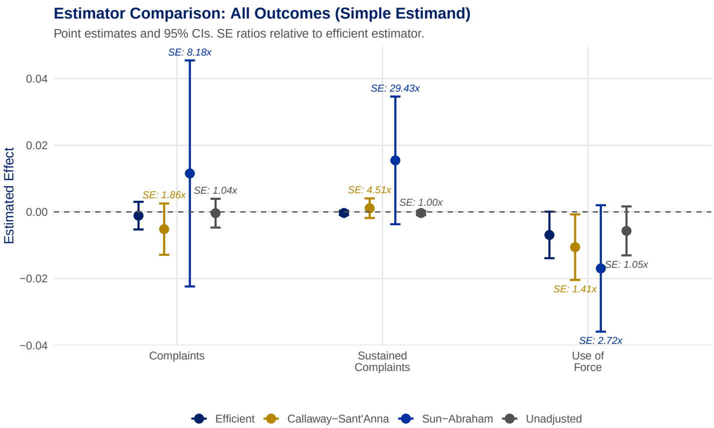
SE ratios annotated above each CI. Large gains over CS and SA; modest gains over unadjusted ($\beta = 0$) – efficiency gains vary by application.

Application Results: Efficiency Gains Across Estimands



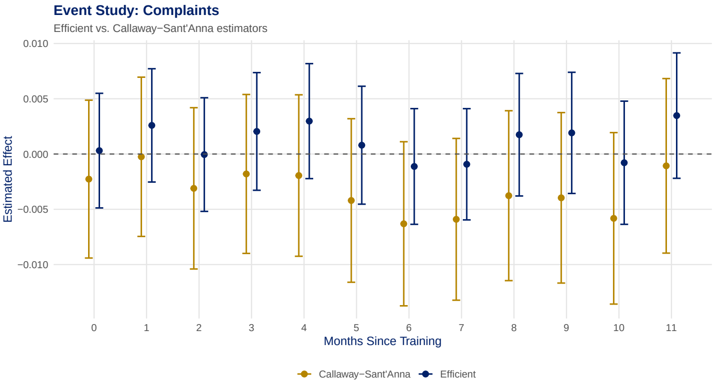
SE ratios > 1 indicate the efficient estimator has smaller standard errors.

Application Results: All Outcomes



Efficiency gains vary across outcomes. Largest gains for sustained complaints; modest for use of force. SE ratios annotated above each CI.

Event Study: Complaints



Efficient estimator (blue) produces tighter CIs at every event time, enabling sharper inference about dynamic treatment effects.

Can we assess the validity of the assumptions?

Balance Checks for Random Treatment Timing

- Under random treatment timing, pre-treatment outcomes should be **balanced** across cohorts
- Test $H_0 : \mathbb{E}[\hat{X}] = 0$ (pre-treatment differences are mean-zero)
- Roth and Sant'Anna (2023): Propose balance checks implemented in `staggered`

Results from police training data:

- **Balanced** on complaints and sustained complaints (main sample)
- **Imbalanced** when including pilot participants/special units (known to violate randomization)
- Pre-treatment event-study: no significant anticipation effects

Balance checks provide evidence *for* or *against* the random timing assumption — analogous to covariate balance checks in RCTs.

Take-Away

Take-Away Message

1. When treatment timing is random, estimators that use *fixed* adjustment weights ($\beta = 0$ or $\beta = 1$) "leave money on the table"
2. Roth and Sant'Anna (2023) show how to use additional information to "collect the money"
3. Estimators and inference easily implemented in **R** via the `staggered` package
4. **Recommendation:** Use when treatment timing is (quasi-)random
   - Other procedures remain valid under weaker assumptions (covered in upcoming lectures)
   - Efficiency gains come from exploiting the additional structure of random timing

Next lecture: Observational panel data — when treatment timing is *not* random.

Summary of Key Results

Result	**Statement**
Unbiasedness	$\hat{\theta}_\beta$ is unbiased for θ for *any* β (Lemma 2.1)
Efficiency	$\beta^* = \mathrm{Var}[\hat{X}]^{-1}\mathrm{Cov}[\hat{X}, \hat{\theta}_0]$ minimizes variance (Prop. 2.1)
Feasibility	Plug-in $\hat{\beta}^*$ achieves same asymptotic variance as oracle (Prop. 2.2)
Inference	Studentized FRT: exact under sharp null, valid under weak null (Prop. 2.3)
Nesting	$\beta{=}0$, $\beta{=}1$, CS, SA all special cases with fixed β
Gains	SE reductions of 1.4–8.4$\times$ in application

References

Athey, Susan and Guido Imbens, "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption," *Journal of Econometrics*, 2021, (Forthcoming).

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Deshpande, Manasi and Yue Li, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, November 2019, *11* (4), 213–248.

Druedahl, Jeppe and Alessandro Martinello, "Long-Run Saving Dynamics: Evidence from Unexpected Inheritances," *Review of Economics and Statistics*, 2022, *104* (5), 1079–1095.

Fadlon, Itzik and Torben Heien Nielsen, "Family Labor Supply Responses to Severe Health Shocks: Evidence from Danish Administrative Records," *American Economic Journal: Applied Economics*, July 2021, *13* (3), 1–30.

Freedman, David A., "On regression adjustments to experimental data," *Advances in Applied Mathematics*, 2008, *40* (2), 180–193.

Lin, Winston, "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique," *Annals of Applied Statistics*, March 2013, *7* (1), 295–318.

Malani, Anup and Julian Reif, "Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform," *Journal of Public Economics*, 2015, *124*, 1–17.

Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland, "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, October 2013, *103* (6), 2530–2553.

Rambachan, Ashesh and Jonathan Roth, "Design-Based Uncertainty for Quasi-Experiments," *Journal of the American Statistical Association*, 2025.

Roth, Jonathan and Pedro H. C. Sant'Anna, "Efficient Estimation for Staggered Rollout Designs," *Journal of Political Economy: Microeconomics*, 2023, *1* (4), 669–709.

Sun, Liyan and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2).

Wood, George, Tom R. Tyler, and Andrew V. Papachristos, "Procedural justice training reduces police use of force and complaints against officers," *Proceedings of the National Academy of Sciences*, May 2020, *117* (18), 9815–9821.

___ , ___ , ___ , Jonathan Roth, and Pedro H. C. Sant'Anna, "Revised Findings for "Procedural Justice Training Reduces Police Use of Force and Complaints Against Officers"," *SocArXiv*, 2020. See also PNAS correction, vol. 118(27), 2021.

Wu, Jason and Peng Ding, "Randomization Tests for Weak Null Hypotheses in Randomized Experiments," *Journal of the American Statistical Association*, May 2020, pp. 1–16. arXiv: 1809.07419.

Zhao, Anqi and Peng Ding, "Covariate-adjusted Fisher randomization tests for the average treatment effect," *arXiv:2010.14555 [math, stat]*, November 2020. arXiv: 2010.14555.

Appendix

Monte Carlo: Two-Period Design

DGP: Draw $\mathbf{Y}_i(\infty) \sim \mathcal{N}(0, \Sigma_\rho)$; set $Y_{i,2}(2) = Y_{i,2}(\infty) + \gamma(Y_{i,2}(\infty) - \mathbb{E}_{fin}[Y_{i,2}(\infty)])$

- $\gamma \in \{0, 0.5\}$: treatment effect heterogeneity; $\rho \in \{0, 0.5, 0.99\}$: autocorrelation
- $N_2 = N_\infty = N/2$, $N \in \{25, 1000\}$

Comparison: Plug-in efficient vs. full adjustment ($\beta = 1$) vs. no adjustment ($\beta = 0$)

Key findings:

- All estimators unbiased; coverage $\approx 95\%$
- $\rho = 0.99$: full adjustment optimal ($\beta^* \approx 1$); $\rho = 0$: no adjustment optimal ($\beta^* \approx 0$)
- Intermediate ρ: plug-in can be **1.7$\times$ more precise** than full adjustment

Monte Carlo: Staggered Design (Calibrated to Application)

Setup: Calibrated to police training data (72 periods, 48 cohorts, 7,785 officers). Sharp null: $Y_{it}(g)$ = observed outcome for all g.

Comparison: Plug-in efficient vs. Callaway-Sant'Anna vs. Sun-Abraham

Key findings:

- Plug-in: approximately unbiased, 93–96% coverage; FRT size $\approx 5\%$
- **Efficiency gains vs. CS:** SE ratio 1.39–1.85 (equivalent to 3.4$\times$ the sample size)
- **Efficiency gains vs. SA:** SE ratio 3.0–6.86

> Substantial efficiency gains are achievable in realistic staggered designs when treatment timing is random.