ECON 730: Causal Inference with Panel Data

Lecture 3: Panel Experiments and Dynamic Causal Effects

Pedro H. C. Sant'Anna

Spring 2026

Motivation: Why Panel Experiments?

Why Study Panel Experiments?

- Last lecture: Potential outcomes depend on **treatment histories**, not just current treatment
- Today: What can we learn when treatments are **randomly assigned** over time?
- Panel experiments provide the **experimental foundation** for understanding:
  - How to define causal effects when past treatments affect current outcomes
  - Why standard estimators can fail
  - What identification looks like under randomization
- This builds intuition for the **observational methods** (DiD, etc.) we'll study later

Main reference: Bojinov, Rambachan and Shephard (2021) — "Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective"

The Carryover Problem

- In panel experiments, we randomly assign treatments over multiple periods
- **Key challenge:** Past treatments may affect current outcomes (*carryover effects*)
- **Examples:**
  - A/B testing at tech companies: User behavior today depends on past experiences
  - Clinical trials with repeated dosing: Drug effects accumulate over time
  - Pricing experiments: Past prices affect current demand through learning
- Standard cross-sectional methods assume $Y_{it}(d)$ — i.e., the outcome depends only on *current* treatment
- But the truth may be $Y_{it}(d_{i,1}, d_{i,2}, \ldots, d_{i,T})$ — i.e., the outcome depends on the *entire history*

Roadmap for Today

1. **Framework:** Potential outcomes indexed by treatment paths
2. **Estimands:** Lag-p dynamic causal effects
3. **Identification & Estimation:** Sequential randomization + Horvitz-Thompson
4. **Special Cases:** RCT, Bernoulli, staggered adoption as illustrations
5. **Application:** Prisoners' dilemma experiment
6. **Why not Fixed Effects?** Bias under carryover + serial correlation
7. **Bridge:** From experiments to observational data

Framework: Potential Outcomes with Treatment Paths

Setup: The Potential Outcome Panel

- **Units:** $i \in \{1, \ldots, N\}$ observed over **periods:** $t \in \{1, \ldots, T\}$
- **Treatment:** $D_{it} \in \mathcal{D}$ assigned to unit i at time t
  - For binary treatment: $\mathcal{D} = \{0, 1\}$
- **Treatment path** for unit i: The sequence of all treatments up to time T

$$\mathbf{d}_{i,1:T} = (d_{i,1}, d_{i,2}, \ldots, d_{i,T}) \in \mathcal{D}^T$$

- **Cross-sectional assignment** at time t: All treatments at period t

$$\mathbf{d}_{1:N,t} = (d_{1,t}, d_{2,t}, \ldots, d_{N,t}) \in \mathcal{D}^N$$

Potential Outcomes Depend on Treatment Paths

Definition (Potential Outcome)

The **potential outcome** for unit i at time t along treatment path $\mathbf{d}_{i,1:T} \in \mathcal{D}^T$ is:

$$Y_{it}(\mathbf{d}_{i,1:T})$$

- In principle, Y_{it} can depend on the **entire** treatment path
- This allows arbitrary spillovers **across time** within a unit
- We assume **no interference across units** (SUTVA): Y_{it} doesn't depend on $\mathbf{d}_{j,1:T}$ for $j \neq i$

Key Assumption: Non-Anticipation

Assumption (Non-Anticipating Potential Outcomes)

For all units i, periods t, and treatment paths $\mathbf{d}_{i,1:T}, \tilde{\mathbf{d}}_{i,1:T} \in \mathcal{D}^T$:

$$Y_{it}(\mathbf{d}_{i,1:T}) = Y_{it}(\tilde{\mathbf{d}}_{i,1:T}) \quad \text{whenever} \quad \mathbf{d}_{i,1:t} = \tilde{\mathbf{d}}_{i,1:t}$$

Interpretation:

- Potential outcomes at time t only depend on treatments **up to time** t
- **Future** treatments don't affect **current** outcomes
- But **past and current** treatments can have **arbitrary effects**

Under non-anticipation: $Y_{it}(\mathbf{d}_{i,1:t})$ instead of $Y_{it}(\mathbf{d}_{i,1:T})$

Q: What if treatment is announced in advance?

Key Assumption: Non-Anticipation

Assumption (Non-Anticipating Potential Outcomes)

For all units i, periods t, and treatment paths $\mathbf{d}_{i,1:T}, \tilde{\mathbf{d}}_{i,1:T} \in \mathcal{D}^T$:

$$Y_{it}(\mathbf{d}_{i,1:T}) = Y_{it}(\tilde{\mathbf{d}}_{i,1:T}) \quad \text{whenever} \quad \mathbf{d}_{i,1:t} = \tilde{\mathbf{d}}_{i,1:t}$$

Interpretation:

- Potential outcomes at time t only depend on treatments **up to time** t
- **Future** treatments don't affect **current** outcomes
- But **past and current** treatments can have **arbitrary effects**

Under non-anticipation: $Y_{it}(\mathbf{d}_{i,1:t})$ instead of $Y_{it}(\mathbf{d}_{i,1:T})$

Q: What if treatment is announced in advance? *A: Define the "treatment" as the announcement, not implementation. Non-anticipation then holds relative to the announcement date.*

Connection to Lecture 2: Treatment Sequences

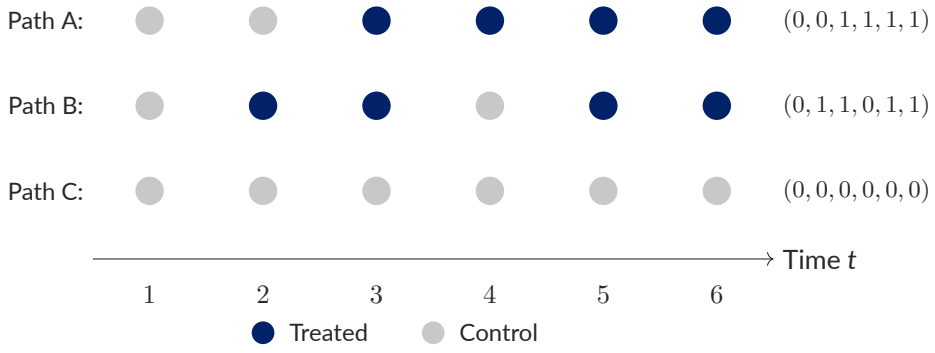
Lecture 2 introduced: $Y_{it}(\mathbf{d}_{i,1:T})$ where $\mathbf{d}_{i,1:T} = (d_{i,1}, \ldots, d_{i,T})$ (Robins)

What's new in this lecture?

- **Non-anticipation:** Restricts dependence to treatments *up to time t*
- **Tractable estimands:** Focusing on recent treatment history instead of the full path
- **How experiments help:** Known assignment probabilities $\rightarrow$ design-based inference
- **When standard methods fail:** Why fixed effects can mislead under carryover

This lecture provides the *experimental foundation* for treatment path dependence. Later: DiD uses parallel trends instead of randomization.

Treatment Path Visualization



Key insight: Different paths lead to different potential outcomes at $t = 6$:

$$Y_{i6}(\text{Path A}) \neq Y_{i6}(\text{Path B}) \neq Y_{i6}(\text{Path C})$$

Estimands: Lag-p Dynamic Causal Effects

Defining Dynamic Causal Effects

A **dynamic causal effect** compares potential outcomes along different treatment paths:

$$\tau_{it}(\mathbf{d}_{i,1:t}, \tilde{\mathbf{d}}_{i,1:t}) := Y_{it}(\mathbf{d}_{i,1:t}) - Y_{it}(\tilde{\mathbf{d}}_{i,1:t})$$

Problem: The number of comparisons grows exponentially with t. **Solution:** Lag-p effects.

Definition (Lag-p Dynamic Causal Effect)

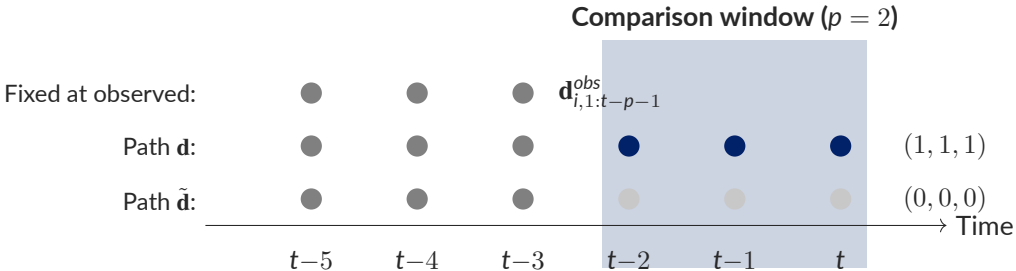
For $0 \leq p < t$ and treatment sequences $\mathbf{d}, \tilde{\mathbf{d}} \in \mathcal{D}^{p+1}$:

$$\tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p) := Y_{it}(\mathbf{d}^{obs}_{i,1:t-p-1}, \mathbf{d}) - Y_{it}(\mathbf{d}^{obs}_{i,1:t-p-1}, \tilde{\mathbf{d}})$$

where $\mathbf{d}^{obs}_{i,1:t-p-1}$ denotes unit i's **realized** path up to period $t-p-1$.

Note: $p < t$ ensures there are enough past periods. Choosing p: bias-variance tradeoff (larger p captures more carryover but needs more data).

Interpreting Lag-p Effects



Lag-p effect with $p = 2$, $\mathbf{d} = (1,1,1)$, $\tilde{\mathbf{d}} = (0,0,0)$:

$$\tau_{it}((1,1,1),(0,0,0);2) = Y_{it}(\mathbf{d}^{obs}_{i,1:t-3},1,1,1) - Y_{it}(\mathbf{d}^{obs}_{i,1:t-3},0,0,0)$$

Effect of treatment in periods $t-2, t-1, t$ vs. control, conditional on the observed earlier path.

Special Cases: Lag-0 and Contemporaneous Effects

Lag-0 dynamic causal effect ($p = 0$):

$$\tau_{it}(d, \tilde{d}; 0) = Y_{it}(\mathbf{d}^{obs}_{i,1:t-1}, d) - Y_{it}(\mathbf{d}^{obs}_{i,1:t-1}, \tilde{d})$$

- Compares treatment d vs. $\tilde{d}$ at time t only, holding the **entire past** fixed
- If **no carryover**: reduces to $Y_{it}(1) - Y_{it}(0)$ (standard effect)

> **Key:** The lag-0 effect is *conditional on history*. For the same unit, $\tau_{it}(1, 0; 0)$ can differ depending on prior treatment status!

Example: Suppose the treatment is a pain medication.

- If unit i was treated yesterday: tolerance may develop $\rightarrow$ *smaller* $\tau_{it}(1, 0; 0)$
- If unit i was *not* treated recently: full drug effect $\rightarrow$ *larger* $\tau_{it}(1, 0; 0)$

Average Lag-p Dynamic Causal Effects

Definition (Average Lag-p Dynamic Causal Effects)

For $p < T$ and $\mathbf{d}, \tilde{\mathbf{d}} \in \mathcal{D}^{p+1}$:

Time-t average (across units): $\bar{\tau}_{\cdot t}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{N} \sum_{i=1}^{N} \tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

Unit-i average (across time): $\bar{\tau}_{i \cdot}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{T-p} \sum_{t=p+1}^{T} \tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

Total average (across all i, t): $\bar{\tau}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{N(T-p)} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

Note: Sums start at $t = p + 1$ because we need p prior periods. Compare to Lecture 2: ATE(t), unit-specific, overall ATE.

Identification and Estimation

The Fundamental Problem of Causal Inference (Revisited)

Recall from Lecture 2: We cannot learn causal effects directly from data without structure.

The fundamental problem:

- For each unit *i* at time *t*, we observe **one** outcome along the realized path
- But potential outcomes exist for **every** possible treatment path $\mathbf{d}_{i,1:t} \in \mathcal{D}^t$
- With binary treatment and *T* periods: 2^T potential outcomes per unit, observe only 1

What structure can help?

- In observational settings: parallel trends, selection on observables, etc.
- In **experiments**: we *know* the assignment mechanism

This lecture: Exploit known assignment probabilities from randomization to construct unbiased estimators via inverse probability weighting.

Assignment Mechanism: Sequential Randomization

What makes this a panel *experiment*? The assignment mechanism is known.

Definition (Sequentially Randomized Assignments)

Assignments are **sequentially randomized** if for all $t \in \{1, \ldots, T\}$:

$$\Pr(\mathbf{D}_{1:N,t} | \mathbf{D}_{1:N,1:t-1}, \mathbf{Y}_{1:N,1:T}) = \Pr(\mathbf{D}_{1:N,t} | \mathbf{D}_{1:N,1:t-1}, \mathbf{Y}_{1:N,1:t-1}(\mathbf{D}_{1:N,1:t-1}))$$

Interpretation:

- Assignment at t can depend on **past assignments** and **past observed outcomes**
- But not on **future** potential outcomes or **counterfactual** past outcomes
- This is the panel analogue of "unconfounded" assignment
- In words: knowing which outcomes *would have been* realized under alternative paths provides no additional information about current assignment

Important: Since we're in an *experiment*, assignment probabilities are **known** to the researcher.

Individualistic Assignments

Definition (Individualistic Assignment)

Assignments are **individualistic** for unit i if:

$$\Pr(D_{it}|D_{-i,t}, \mathcal{F}_{1:N,t-1,T}) = \Pr(D_{it}|\mathbf{D}_{i,1:t-1}, \mathbf{Y}_{i,1:t-1})$$

where $\mathcal{F}_{1:N,t,T}$ is the filtration generated by treatments and potential outcomes.

Interpretation:

- Unit i's assignment depends only on its **own** past, not on other units
- Conditional on own history, assignments are **independent across units**

Example: Bernoulli assignment where $\Pr(D_{it} = 1) = q$ for all i, t independently.

The Adapted Propensity Score: Definition

Definition (Adapted Propensity Score)

For unit i at time t and treatment sequence $\mathbf{d} = (d_{t-p}, \ldots, d_t) \in \mathcal{D}^{p+1}$:

$$\pi_{i,t-p}(\mathbf{d}) := \Pr(\mathbf{D}_{i,t-p:t} = \mathbf{d} | \mathbf{D}_{i,1:t-p-1}, \mathbf{Y}_{i,1:t-1})$$

What this measures: Probability of observing path $\mathbf{d}$ over periods $t-p$ to t, conditional on the unit's **past** assignment and outcome history. (The subscript $t-p$ denotes the *start* of the treatment window.)

Why "adapted"?

- The propensity score can **change over time** as information accumulates
- But since the experiment is **designed**, we *know* these probabilities

Assumption (Probabilistic Assignment (Overlap))

There exist constants $0 < c_L < c_U < 1$ such that $c_L < \pi_{i,t-p}(\mathbf{d}) < c_U$ for all $i, t, \mathbf{d}$

Computing the Adapted Propensity Score

Step 1: Identify the experimental design (Bernoulli, block, adaptive, etc.)

Step 2: Compute period-by-period probabilities. For path $\mathbf{d} = (d_{t-p}, \ldots, d_t)$:

$$\pi_{i,t-p}(\mathbf{d}) = \prod_{s=t-p}^{t} \Pr(D_{is} = d_s | \text{history up to } s-1)$$

Step 3: Apply to specific designs

- **iid Bernoulli** with $\Pr(D_{it} = 1) = q$: $\pi_{i,t-p}(\mathbf{d}) = q^{\#\{s:d_s=1\}} \cdot (1-q)^{\#\{s:d_s=0\}}$
- **Example:** $p = 2$, $q = 0.5$, $\mathbf{d} = (1, 0, 1)$: $\pi_{i,t-2}(1, 0, 1) = (0.5)^2 \cdot (0.5)^1 = 0.125$

Key insight: We only need $\pi_{i,t-p}(\mathbf{d})$ for the **observed** path — and we know it by design!

Horvitz-Thompson Estimator: Building Block

Definition ((i,t)-th Contribution to Lag-p Effect Estimator)

For $\mathbf{d}, \tilde{\mathbf{d}} \in \mathcal{D}^{p+1}$:

$$\hat{\tau}_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p) = \frac{Y_{it} \cdot \mathbf{1}\{\mathbf{D}_{i,t-p:t} = \mathbf{d}\}}{\pi_{i,t-p}(\mathbf{d})} - \frac{Y_{it} \cdot \mathbf{1}\{\mathbf{D}_{i,t-p:t} = \tilde{\mathbf{d}}\}}{\pi_{i,t-p}(\tilde{\mathbf{d}})}$$

Intuition:

- When unit i follows path $\mathbf{d}$: contribute $Y_{it}/\pi_{i,t-p}(\mathbf{d})$
- When unit i follows path $\tilde{\mathbf{d}}$: contribute $-Y_{it}/\pi_{i,t-p}(\tilde{\mathbf{d}})$
- Otherwise: contribute zero. IPW corrects for different path probabilities

This is an **estimation building block**. We aggregate these (i,t) contributions into **plug-in averages** to estimate population-level effects.

Plug-in Average Estimators

Definition (Plug-in Average Estimators)

Time-t average: $\hat{\bar{\tau}}_{\cdot t}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{N} \sum_{i=1}^{N} \hat{\tau}_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

Unit-i average: $\hat{\bar{\tau}}_{i\cdot}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\tau}_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

Total average: $\hat{\bar{\tau}}(\mathbf{d}, \tilde{\mathbf{d}}; p) := \frac{1}{N(T-p)} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \hat{\tau}_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$

These estimate $\bar{\tau}_{\cdot t}$, $\bar{\tau}_{i\cdot}$, and $\bar{\tau}$ from earlier — **unbiased under randomization.**

Key Result: Unbiasedness

Theorem (Bojinov et al. (2021), Theorem 3.1)

Under individualistic and probabilistic assignment:

$$\mathbb{E}[\hat{\tau}_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p) | \mathcal{F}_{i,t-p-1}] = \tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p)$$

The estimation error is a **martingale difference sequence** through time and **conditionally independent** across units.

Implications:

- The estimator is **unbiased** for the true lag-p dynamic causal effect
- Unbiasedness is over the **randomization distribution** (design-based)
- No assumptions on the outcome model needed!

Variance: Can be **conservatively estimated** from the data. Plug-in averages are also unbiased.

What Does Design-Based Unbiasedness Mean?

Setup: $N = 4$, binary treatment, Bernoulli(0.5), **at period** t ($\mathcal{F}_{i,t-p-1}$ realized). $2^4 = 16$ possible assignments.

D_{1t}	D_{2t}	D_{3t}	D_{4t}	$\hat{\bar{\tau}}_{\cdot t}$
0	0	0	1	some number
		⋮		⋮
1	1	1	0	some number

"Unbiased" = average of $\hat{\bar{\tau}}_{\cdot t}$ over all 16 assignments equals $\bar{\tau}_{\cdot t}$.

Design-based $\neq$ model-based. Potential outcomes are *fixed*; only assignments are random. We average over the *randomization distribution*, not over hypothetical repeated samples.

Full panel: Total randomization space is 2^{NT}, but unbiasedness works **period by period** (Theorem 3.1 conditions on $\mathcal{F}_{i,t-p-1}$), then aggregates via iterated expectations.

Inference: Finite Population CLTs

Theorem (Bojinov et al. (2021), Theorem 3.2)

Under individualistic, probabilistic assignment with bounded potential outcomes:

Time-t specific CLT (as $N \to \infty$):

$$\frac{\sqrt{N}\{\hat{\bar{\tau}}_{\cdot t}(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}_{\cdot t}(\mathbf{d}, \tilde{\mathbf{d}}; p)\}}{\sigma_{\cdot t}} \xrightarrow{d} N(0, 1)$$

Unit-i specific CLT (as $T \to \infty$):

$$\frac{\sqrt{T - p}\{\hat{\bar{\tau}}_{i \cdot}(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}_{i \cdot}(\mathbf{d}, \tilde{\mathbf{d}}; p)\}}{\sigma_{i \cdot}} \xrightarrow{d} N(0, 1)$$

Total average CLT (as $NT \to \infty$):

$$\frac{\sqrt{N(T - p)}\{\hat{\bar{\tau}}(\mathbf{d}, \tilde{\mathbf{d}}; p) - \bar{\tau}(\mathbf{d}, \tilde{\mathbf{d}}; p)\}}{\sigma} \xrightarrow{d} N(0, 1)$$

Inference: Key Points

Key implications:

- **Time-specific inference** ($N \to \infty$): Test effects at a given period across units
- **Unit-specific inference** ($T \to \infty$): Test effects for a given unit across time
- **Total average inference** ($NT \to \infty$): Pool across both dimensions

Practical considerations:

- Variance can be **conservatively estimated** from the data
- Enables valid confidence intervals and hypothesis tests

Two testing approaches:

1. *Conservative tests*: CLT + variance upper bound $\to$ tests **weak null** $H_0 : \bar{\tau} = 0$ (average effect is zero)
2. *Randomization tests*: Exact tests under **sharp null** $\tau_{it} = 0$ for *all* (i, t) (no unit has any effect)

Randomization Tests: The Idea

Under the **sharp null** $H_0 : \tau_{it}(\mathbf{d}, \tilde{\mathbf{d}}; p) = 0$ for *all* (i, t):

1. We can **impute all missing potential outcomes**: if no unit has any effect, then $Y_{it}(\mathbf{d}) = Y_{it}(\tilde{\mathbf{d}})$ for all paths $\rightarrow$ observed outcome = counterfactual outcome
2. Compute the HT test statistic for **every possible assignment**, not just the one that occurred
3. ***p*-value** $=$ fraction of possible assignments producing a test statistic as extreme as the one observed

Advantages:

- **Exact**: No asymptotic approximation needed — valid in finite samples
- Follows directly from Fisher's classical randomization inference

Limitation: The sharp null ($\tau_{it} = 0$ for *every* unit at *every* time) is stronger than the weak null ($\bar{\tau} = 0$, the *average* is zero). The CLT-based conservative test handles the weak null.

Identification and Estimation

Illustrative Special Cases

Special Case 1: Single Treatment Date (RCT)

Setup: N units, $T = 3$ periods. Period $t{=}1$ is pre-treatment. Treatment randomized at $t{=}2$ only: $\Pr(D_{i2} = 1) = q$. No treatment at $t{=}1$ or $t{=}3$.

All units share the same pre-treatment path, so $\mathbf{d}_{i,1}^{obs} = 0$ for everyone.

- **Lag-0 effect at $t = 2$:** $\tau_{i2}(1, 0; 0) = Y_{i2}(0, 1) - Y_{i2}(0, 0)$ $\Rightarrow$ **Standard ATE**
- **Lag-1 effect at $t = 3$:** $\tau_{i3}(1, 0; 1) = Y_{i3}(0, 1, 0) - Y_{i3}(0, 0, 0)$ $\Rightarrow$ **Carryover effect**

Adapted propensity score: $\pi_{i,2}(1) = q$, $\pi_{i,2}(0) = 1 - q$ — known by design.

Absorbing variant: If once treated, always treated: paths become $(0, 0, 0)$ and $(0, 1, 1)$. Then lag-1 at $t{=}3$ is $Y_{i3}(0, 1, 1) - Y_{i3}(0, 0, 0)$ — conflating contemporaneous and carryover. This is the staggered adoption challenge.

Key insight: Even in a simple RCT with follow-up, a naïve "treatment effect" conflates contemporaneous and lagged effects. The lag-p framework separates them cleanly.

Special Case 2: iid Bernoulli Assignment

Setup: Each unit independently assigned $D_{it} = 1$ with probability q at *every* period.

- **Treatment paths:** All 2^T paths are possible, each with known probability
- **Adapted propensity score** for path $\mathbf{d} = (d_{t-p}, \ldots, d_t) \in \mathcal{D}^{p+1}$:

$$\pi_{i,t-p}(\mathbf{d}) = q^k \cdot (1-q)^{p+1-k}, \quad k = \#\{s : d_s = 1\}$$

- Same for all units and all time periods — no dependence on history
- **HT estimator:** IPW corrects for different path probabilities; naïve mean comparison would not

Notable feature: **No serial correlation** in treatment — successive treatments are independent: $\text{Cov}(D_{it}, D_{is}) = 0$ for $s \neq t$. We will see why this independence property matters when we discuss fixed effects estimators.

Special Case 3: Staggered Adoption

Setup: Treatment is **absorbing** — once treated, always treated. Units adopt at different times $g \in \{1, \ldots, T\}$.

- **Treatment paths collapse to:** $(0, \ldots, 0, 1, \ldots, 1)$ with switch point g
- **Lag-p effect** for unit treated at g, evaluated at $t \geq g + p$ (unit must have been treated for at least $p+1$ consecutive periods):
  Compares "treated for $p+1$ consecutive periods" vs. "not yet treated"
- **Connection to Lecture 2:** Closely related to $ATT(g, t)$ — the group-time treatment effect
- **Adapted propensity score:** $\Pr(\text{adopt at } g \mid \text{not yet adopted by } g-1)$ — the *hazard* of adoption

Notable feature: Treatment is **perfectly serially correlated** once adopted — if $D_{it} = 1$, then $D_{is} = 1$ for all $s > t$. Contrast with iid Bernoulli above!

We'll see why this serial correlation structure matters for FE estimators.

HT Estimator: A Worked Example

Setup: $N = 4$ units, period $t = 2$, Bernoulli(0.5), lag-0 effect $\hat{\tilde{\tau}}_{\cdot 2}(1, 0; 0)$.

Unit i	D_{i2}	Y_{i2}	HT contribution: $\frac{Y_{i2} \cdot \mathbf{1}\{D_{i2}=1\}}{\pi_{i2}(1)} - \frac{Y_{i2} \cdot \mathbf{1}\{D_{i2}=0\}}{\pi_{i2}(0)}$
1	1	8	$8/0.5 = 16$
2	0	3	$-3/0.5 = -6$
3	1	6	$6/0.5 = 12$
4	0	2	$-2/0.5 = -4$

$$\hat{\tilde{\tau}}_{\cdot 2} = \tfrac{1}{4}(16 - 6 + 12 - 4) = 4.5$$

Check: With equal propensity scores ($\pi_{i2}(1) = 0.5$ for all i), HT = difference in means: $\bar{Y}_{\text{treated}} - \bar{Y}_{\text{control}} = 7 - 2.5 = 4.5$ ✓

What if propensity scores are **unequal**? Does simple mean comparison still work?

Why IPW Matters: Unequal Propensity Scores

Same units, but now an **adaptive design**: propensity scores vary across units.

Unit i	D_{i2}	Y_{i2}	$\pi_{i2}(D_{i2})$	HT contribution
1	1	8	0.8	$8/0.8 = 10$
2	0	3	0.7	$-3/0.7 = -4.29$
3	1	6	0.3	$6/0.3 = 20$
4	0	2	0.4	$-2/0.4 = -5$

$$\hat{\bar{\tau}}^{HT}_{\cdot 2} = \tfrac{1}{4}(10 - 4.29 + 20 - 5) = 5.18$$

Naïve difference in means: $\bar{Y}_{\text{treated}} - \bar{Y}_{\text{control}} = 7 - 2.5 = 4.5 \neq 5.18$

Why the difference? Unit 3 ($D_{i2} = 1$, $\pi = 0.3$) was *unlikely* to be treated — its outcome is more "informative" about treatment effects and gets **upweighted**. Naïve means ignore this, producing bias.

Application: Prisoners' Dilemma Experiment

Application: Rational Cooperation in Games

Setting: Andreoni and Samuelson (2006) study cooperative behavior in a twice-repeated prisoners' dilemma.

Design: $N = 110$ participants, $T = 20$ rounds, randomly matched into pairs each round. **Outcome:** cooperation in period 1 of the game.

Treatment ($D_{it} \in \{0, 1\}$): The payoff structure parameter λ is **randomly varied** each round.

- $D_{it} = 1$ (high λ): Payoffs reward *patience* — cooperation is rational
- $D_{it} = 0$ (low λ): Payoffs reward *immediacy* — defection is tempting

> Treatment is randomly assigned each round $\Rightarrow$ **Bernoulli-like** design. But past game structures may affect current behavior through **learning** — exactly the carryover concern.

Results: Dynamic Causal Effects in the Experiment

	Lag-p			
	$p = 0$	$p = 1$	$p = 2$	$p = 3$
Point estimate $\hat{\bar{\tau}}^{\dagger}(1, 0; p)$	0.285	0.058	0.134	0.089
Conservative p-value	0.000	0.226	0.013	0.126
Randomization p-value	0.000	0.263	0.012	0.114

†: Overall average HT estimator $\hat{\bar{\tau}}$ as defined earlier, pooling across all units and rounds. Conservative tests use CLT; randomization tests simulate under sharp null.

Findings:

- **Strong contemporaneous effect** ($p = 0$): Treatment increases cooperation by 28.5 pp
- **Suggestive lag-2 effects** ($p = 2$): Past structures affect behavior (p-value $= 0.012$)

Interpretation: The lag-2 effect suggests *learning dynamics* — players update beliefs based on past experiences. This carryover would bias naïve FE regressions.

Why Not Fixed Effects?

A Natural Question

We've established that Horvitz-Thompson estimators are unbiased.

But practitioners often use simpler **fixed effects** estimators.

Do they work under carryover effects?

Spoiler: In general, **no**. FE estimators can be substantially biased when there are carryover effects *and* serial correlation in treatment assignment.

The Appeal of Fixed Effects

Common practice: Run OLS with unit fixed effects:

$$Y_{it} = \alpha_i + \beta D_{it} + \epsilon_{it}$$

Why it seems reasonable:

- Controls for time-invariant unit heterogeneity (α_i)
- $\hat{\beta}$ should capture the "treatment effect"
- Simple, widely available in standard software

Question: Does the unit FE estimator $\hat{\beta}_{UFE}$ recover a meaningful causal effect when there are **carryover effects**?

Answer: In general, **no** — and the bias can be substantial.

Linear Potential Outcome Panel Model

Definition (Bojinov et al. (2021), Definition 7)

A **linear potential outcome panel** satisfies:

$$Y_{it}(d_{i,1:t}) = \beta_{it,0} d_{it} + \beta_{it,1} d_{i,t-1} + \cdots + \beta_{it,t-1} d_{i,1} + \epsilon_{it}$$

Interpretation:

- $\beta_{it,0}$: **Contemporaneous effect** — effect of current treatment on current outcome
- $\beta_{it,s}$ for $s > 0$: **Carryover effect** — effect of treatment s periods ago
- $\epsilon_{it} = Y_{it}(\mathbf{0})$: Potential outcome under never-treated path

This is a **structural model** for potential outcomes that separates contemporaneous from carryover effects. It nests the no-carryover case ($\beta_{it,s} = 0$ for $s > 0$).

The Unit Fixed Effects (UFE) Estimator

Definition: The **unit fixed effects (UFE) estimator** is:

$$\hat{\beta}_{UFE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Y}_{it}\tilde{D}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{D}_{it}^2}$$

where $\tilde{A}_{it} = A_{it} - \bar{A}_{i\cdot}$ denotes the **within-unit deviation** from unit i's time average.

Equivalently: OLS coefficient from regressing Y_{it} on D_{it} with unit fixed effects:

$$Y_{it} = \alpha_i + \beta D_{it} + \text{error}_{it}$$

UFE uses **within-unit variation** in treatment over time to identify effects. It removes time-invariant unit heterogeneity.

Bias of the Unit Fixed Effects Estimator

Proposition (Bojinov et al. (2021), Proposition 4.1)

Under a linear potential outcome panel, as $N \to \infty$:

$$\hat{\beta}_{UFE} \xrightarrow{p} \underbrace{\frac{\sum_{t=1}^{T} \tilde{\kappa}_{D,\beta,t,t}}{\sum_{t=1}^{T} \tilde{\sigma}^2_{D,t}}}_{\text{Target}} + \underbrace{\frac{\sum_{t=1}^{T} \sum_{s=1}^{t-1} \tilde{\kappa}_{D,\beta,t,s}}{\sum_{t=1}^{T} \tilde{\sigma}^2_{D,t}}}_{\text{Carryover Bias}} + \underbrace{\frac{\sum_{t=1}^{T} \tilde{\delta}_t}{\sum_{t=1}^{T} \tilde{\sigma}^2_{D,t}}}_{\text{Specification Error}}$$

(Each quantity defined on the next slide.)

Three components:

1. **Target:** A variance-weighted average of contemporaneous effects $\beta_{it,0}$
2. **Carryover bias:** Arises when past treatment affects outcomes *and* treatment is serially correlated
3. **Specification error:** Arises if untreated potential outcomes $Y_{it}(\mathbf{0})$ vary over time

UFE is unbiased only when *both* carryover bias and specification error are zero.

Unpacking Proposition 4.1: The Key Quantities

Within-unit deviation: $\tilde{D}_{it} = D_{it} - \bar{D}_{i\cdot}$, where $\bar{D}_{i\cdot} = T^{-1} \sum_{t=1}^{T} D_{it}$

Treatment variation: $\tilde{\sigma}^2_{D,t} = \lim_{N\to\infty} N^{-1} \sum_i \text{Var}(\tilde{D}_{it})$

How much within-unit treatment variation exists at time t?

Effect–correlation interaction: $\tilde{\kappa}_{D,\beta,t,s} = \lim_{N\to\infty} N^{-1} \sum_i \beta_{it,s} \cdot \text{Cov}(\tilde{D}_{it}, \tilde{D}_{is})$

How much does carryover from period s "leak" into the period-t estimate?

Specification error: $\tilde{\delta}_t$ captures time-varying untreated outcomes

Vanishes if $Y_{it}(\mathbf{0})$ is time-invariant (conditional on unit FE)

The critical insight: Carryover bias $= \sum_{s<t} \beta_{it,s} \cdot \text{Cov}(\tilde{D}_{it}, \tilde{D}_{is})$. This is the product of **carryover effects** and **serial correlation in treatment**. If *either* is zero, the bias vanishes.

Discussion: What Target Parameter?

Even **without** carryover or specification error, UFE estimates:

$$\frac{\sum_{t=1}^{T} \tilde{\kappa}_{D,\beta,t,t}}{\sum_{t=1}^{T} \tilde{\sigma}^2_{D,t}} = \frac{\sum_t \left(N^{-1} \sum_i \beta_{it,0} \cdot \text{Var}(\tilde{D}_{it}) \right)}{\sum_t \left(N^{-1} \sum_i \text{Var}(\tilde{D}_{it}) \right)}$$

$\rightarrow$ **Weights** $\propto \text{Var}(\tilde{D}_{it})$: Units/periods with more treatment variation get more weight.

Key questions:

- Are these weights **policy-relevant**? Do we care more about (i, t) with high treatment variation?
- What if effects are **heterogeneous** ($\beta_{it,0}$ varies)? Weighted average $\neq$ simple avg.
- How does this compare to **equally-weighted** averages from Horvitz-Thompson?

Takeaway: Unbiasedness requires specifying the *target*. UFE is unbiased for a *particular* weighted average — but weights come from *design*, not policy.

When Does UFE Bias Matter?

Carryover bias $\propto$ (carryover effects) $\times$ (serial correlation in treatment)

	Serial Correlation in D_{it}	
	No (iid Bernoulli)	Yes (staggered)
No Carryover ($\beta_{it,s} = 0$ for $s > 0$)	**Unbiased**	**Unbiased**
Carryover Present ($\beta_{it,s} \neq 0$)	**Unbiased**	**BIASED**

Surprising: Under iid Bernoulli, UFE is unbiased *even with carryover*! Why? $\text{Cov}(\tilde{D}_{it}, \tilde{D}_{is}) = 0$ for $s \neq t$, so the carryover bias term vanishes.

Why staggered = guaranteed bias: Under staggered adoption, $D_{it} = 1$ for all $t \geq g_i \Rightarrow$ $\text{Cov}(\tilde{D}_{it}, \tilde{D}_{is}) > 0$ for post-adoption periods. Any $\beta_{it,s} \neq 0$ makes the bias non-zero.

Two-Way Fixed Effects: Same Issue

TWFE: $Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \text{error}_{it}$

Proposition (Bojinov et al. (2021), Proposition 4.2)

TWFE has the **same three-component bias structure** as UFE, with different weights.

Important distinction from DiD literature:

- **Staggered DiD issues** (Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak, Jaravel and Spiess, 2024; Imai and Kim, 2021): arise even *without* carryover, under PT assumptions
- **Here:** bias arises specifically *because of* carryover effects

Athey and Imbens (2021) also take a design-based approach to staggered adoption, but don't focus on carryover effects.

Takeaway: When carryover effects are possible, use Horvitz-Thompson estimators — not unit FE or TWFE.

From Experiments to Observational Data

From Experiments to Observational Data

So far: Propensity scores *known* by design ⇒ Horvitz-Thompson is unbiased.

What if we don't run the experiment? Blackwell and Glynn (2018) address this using the *same* framework:

- Same potential outcomes indexed by treatment histories: $Y_{it}(\mathbf{d}_{i,1:t})$
- Same estimands: Contemporaneous effect ↔ lag-0; Lagged effects ↔ lag-p
- **Key difference:** Identification via *sequential ignorability* (selection on observables) — propensity scores must be **estimated**, not known

Blackwell and Glynn's key insight: With time-varying covariates affected by past treatment, standard regression cannot consistently estimate lagged effects (*post-treatment bias*). Solutions exist (inverse probability weighting, structural models) — see Blackwell and Glynn (2018) for details.

> **Limitation:** Sequential ignorability *cannot* handle time-constant unmeasured confounders ⇒ motivates **DiD** and parallel trends.

Key Takeaways

Key Takeaways

What's unique to this lecture:

1. **Lag-p effects:** Tractable treatment path comparisons nesting familiar settings
2. **Design-based identification:** Known propensity scores $\rightarrow$ unbiased HT estimation
3. **Finite population CLTs:** Valid inference without outcome model assumptions
4. **FE bias:** Bias $\propto$ (carryover) $\times$ (serial correlation) $-$ guaranteed under staggered adoption
5. **Roadmap:** Experiments $\rightarrow$ selection on observables (Blackwell and Glynn, 2018) $\rightarrow$ DiD (parallel trends)

Main message: When carryover effects are possible, don't default to TWFE. DiD provides an alternative that handles unmeasured time-constant confounders.

Next lecture: Efficient Estimation with Staggered Designs.

References

Andreoni, James and Larry Samuelson, "Building Rational Cooperation," *Journal of Economic Theory*, 2006, *127* (1), 117–154.

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

Blackwell, Matthew and Adam N. Glynn, "How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables," *American Political Science Review*, 2018, *112* (4), 1036–1049.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard, "Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective," *Quantitative Economics*, 2021, *12* (4), 1171–1196.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2).

Imai, Kosuke and In Song Kim, "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data," *Political Analysis*, 2021, *29* (3), 405–415.

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