10. Panel Data and Diff-in-Diffs

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Outline

1. Difference in Differences

2. Empirical Demonstration in R

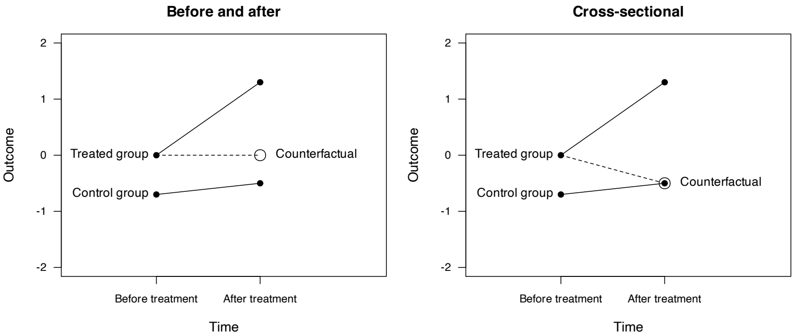
Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates
  - Units with opposite value of some instrument
  - At a discontinuity in treatment assignment (will cover in W13)

- What if we have repeated measurements of the same units?

- Now there are two possible sources of variation to exploit:
  - Exploit **cross-sectional** variation in treatment.
  - Exploit variation in treatment **within a unit over time** (before/after)

Cross-sectional vs Before-and-After



Source: *Chapter 5* of Mostly Harmless Econometrics (Textbook 1) by J. Angrist & J. Pischke

1/ Difference in Differences

Minimum Wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
  - Neighboring PA stays at $4.25
  - We observe employment in both states before and after increase
- Look at eastern PA and NJ fast food restaurants.
  - Similar prices, wages, products, etc.
  - Most likely to be affected by the change.

Minimum Wage Law in PA and NJ

Source: https://www.alamy.com/stock-photo/pennsylvania-map.html

Difference in Differences Design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, the other group remains untreated.

- Groups defined by treatment status in post-period:
  - $G_i = 1$ denote those that are treated at $t = 1$.
  - $G_i = 0$ denote those that are always untreated.

- Treatment status in each period:
  - No treatment in the first period for either group: $D_{i0} = 0$.
  - In treated group: $G_i = 1 \quad \rightsquigarrow \quad D_{i1} = 1$
  - In control group: $G_i = 0 \quad \rightsquigarrow \quad D_{i1} = 0$

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$D_{i0} = 0$	$D_{i1} = 0$
Treated group ($G_i = 1$)	$D_{i0} = 0$	$D_{i1} = 1$

Potential Outcomes Approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d and time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Consistency: $Y_{it} = D_{it} \cdot Y_{it}(1) + (1 - D_{it}) \cdot Y_{it}(0)$
  - We observe control PO for all units in the first period: $Y_{i0}(0) = Y_{i0}$
  - In treated group: $G_i = 1 \quad \rightsquigarrow \quad Y_{i1} = Y_{i1}(1)$
  - In control group: $G_i = 0 \quad \rightsquigarrow \quad Y_{i1} = Y_{i1}(0)$

Identification Problem

- Average treatment effect on the treated (i.e., ATT):

$$
\begin{aligned}
\tau_{\text{ATT}} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1] \\
&= \mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 1] \\
&= \underbrace{\mathbb{E}[Y_{i1} | G_i = 1]}_{\text{(a)}} - \underbrace{\mathbb{E}[Y_{i1}(0) | G_i = 1]}_{\text{(b)}}
\end{aligned}
$$

- Part (a) is just a conditional average of observed data $\rightsquigarrow$ identified.
- Part (b) is a counterfactual: what would the average outcome in the treated group have been if it have been in control?

Three Control Strategies

$$\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)|G_i = 1]$

- **#1: Cross-Sectional Design**
  - Assumption: mean independence of treatment.

$$\mathbb{E}[Y_{i1}(0)|G_i = 1] = \mathbb{E}[Y_{i1}(0)|G_i = 0]$$

  - Use post-treatment control group as control:

$$\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i1}|G_i = 0]$$

Three Control Strategies

$$\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **#2: Before-and-After Design**
  - Assumption: no trends.

$$\mathbb{E}[Y_{i0}(0)|G_i = 1] = \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

  - Use pre-period outcome in treated group:

$$\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]$$

Three Control Strategies

$$\tau_{\text{ATT}} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period ($t = 0$)	Post-period ($t = 1$)
Control group ($G_i = 0$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group ($G_i = 1$)	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **#3: Difference-in-Differences**
  - Assumption: parallel trends.

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

  - Use pre-period treated outcome plus trend in control group:

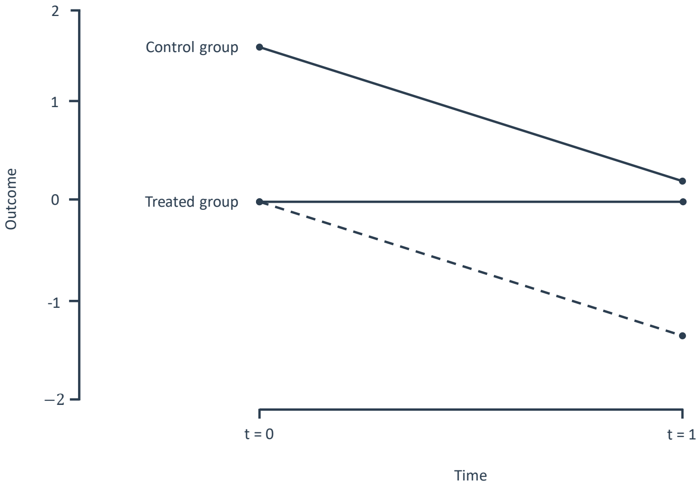
$$\begin{aligned}\tau_{ATT} = &(\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ &- (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])\end{aligned}$$

Parallel Trends

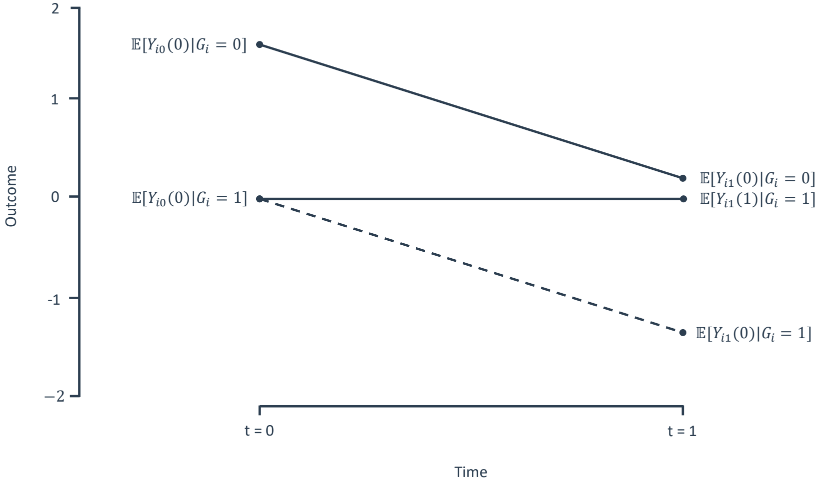
$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of difference-in-differences: **parallel trends**
- Interpretation:
  - Secular trends in the control group is a good proxy for how the treated group would have changed over time without treatment.
    - p.s., a secular trend = a general trend built up over the years.
- Why is this weaker than other assumptions?
  - It allows for time-constant unmeasured confounding b/w Y_{it} and G_i.
  - It allows for (common) secular trends in the outcome over time.
- Not invariant to non-linear transformations:
  - Parallel trends for Y_{it} implies non-parallel trends for $\log(Y_{it})$ vice versa.

Parallel Trends in a Graph



Parallel Trends in a Graph



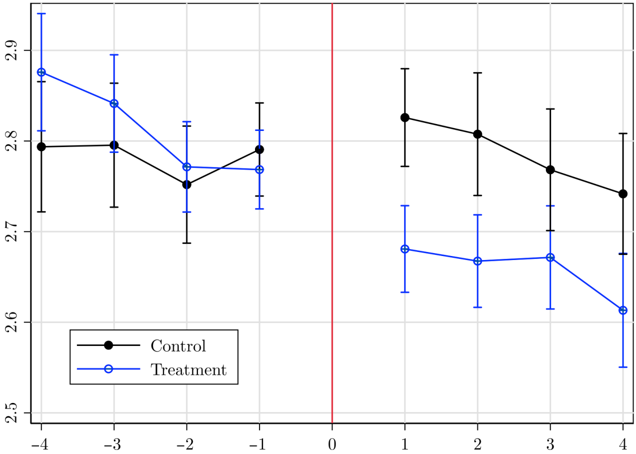
Identification

- Identification result:

$$\tau_{\text{ATT}} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

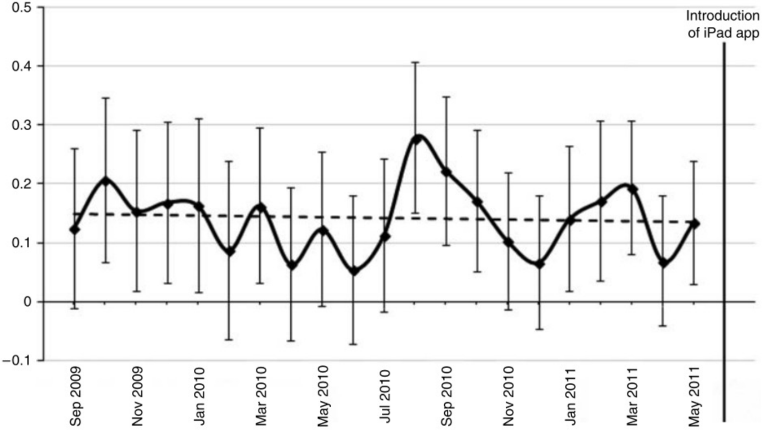
- Threat to identification: non-parallel trends
    - **Unmeasured time-variant confounders**
    - **Ashenfelter's dip**: empirical finding that people who enroll in job training programs see their earnings decline prior to that training (link).
- Falsification test: check pre-treatment parallel trends.
    - Doesn't imply parallel trends hold for the post-period, however!
    - *For a formal treatment of pretesting pitfalls (low power, conditioning distortion), see Roth (2022, AER Insights).*

Checking Parallel Trends (Kretschmer & Peukert, 2020)



Source: Kretschmer, Tobias, and Christian Peukert. "Video killed the radio star? Online music videos and recorded music sales." Information Systems Research 31, no. 3 (2020): 776-800.

Checking Parallel Trends (Xu et al., 2017)



Source: Xu, Kaiquan, Jason Chan, Anindya Ghose, and Sang Pil Han. "Battle of the channels: The impact of tablets on digital commerce." Management Science 63, no. 5 (2017): 1469-1492.

Event-Study Plot: Pre + Post Coefficients

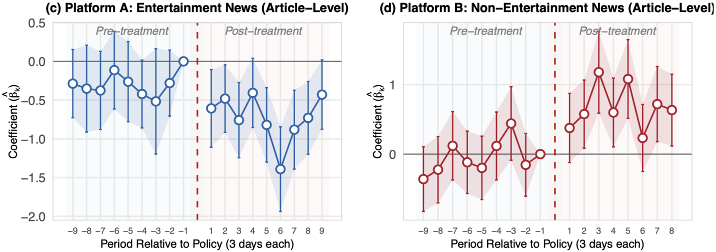


Figure 1. Event Study Estimates of Policy Effects (Reference: k = −1)

Ongoing research on platform governance and self-regulation; data from two largest digital news platforms in Asia.

- Convention: reference $k = -1$, full pre + post $\widehat{\beta}_k$, pointwise CI + shaded band; two panels reveal sign-flip across platforms (heterogeneous TE).
- **Model-free visual evidence is primary** in design-based causal inference. Statistical tests follow as confirmation, not the main argument.

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\text{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i\{Y_{i1} - Y_{i0}\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\{Y_{i1} - Y_{i0}\}}_{\text{average trend in control group}}$$

  - Standard errors from standard difference in means.

- Regression implementation:
  - Regress $\Delta Y_i = Y_{i1} - Y_{i0}$ on G_i.
  - Use (cluster) robust SEs.

- Also possible to use DID on repeated cross sections.

DID and Linear Two-Way Fixed Effects

- Linear two-way (group and time) fixed effects model:

$$Y_{it} = \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - D_{it} is treatment status for unit i at time t (i.e., $G_i \times t$).
  - Be sure to cluster by unit (or level of treatment assignment).
- Coefficient on D_{it} equivalent to DID estimation.
- Only holds for the 2 group, 2 period case!
  - Large new literature on interpretation of TWFE in more general cases.
  - Basically, TWFE is an odd weighted average of DID effects with sometimes negative weights.

DID vs Lagged Dependent Variable

- Parallel trends is **scale-dependent** (recall: PT in Y implies non-PT in $\log Y$). When PT is questionable, a common alternative is to adjust for the **lagged dependent variable** (LDV).

- Alternative identification assumption (ignorability given past outcome):

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends (distribution-level vs. mean-level restriction).
  - Benefit over PT: **scale-free**.

- Different ideas about why there is imbalance on the LDV:
  - **DID**: time-constant unmeasured confounder creates imbalance.
  - **LDV**: previous outcome directly affects treatment assignment.

DID/LDV Bracketing (Angrist & Pischke, 2009; Ding & Li, 2019)

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{\text{LDV}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i1}}_{\text{difference in post period}} - \widehat{\rho}_{\text{LDV}} \underbrace{\left\{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i0}\right\}}_{\text{difference in pre period}}$$

- If $\widehat{\rho}_{\text{LDV}} = 1$ (no mean-reversion) then $\widehat{\tau}_{\text{DID}} = \widehat{\tau}_{\text{LDV}}$; if $0 \leq \widehat{\rho}_{\text{LDV}} < 1$ (partial mean-reversion):
    - Treated has higher pre-period baseline ($\overline{Y}_0^T > \overline{Y}_0^C$) $\rightsquigarrow \widehat{\tau}_{\text{LDV}} > \widehat{\tau}_{\text{DID}}$.
    - Treated has lower pre-period baseline ($\overline{Y}_0^T < \overline{Y}_0^C$) $\rightsquigarrow \widehat{\tau}_{\text{DID}} > \widehat{\tau}_{\text{LDV}}$.
- Bracketing: under PT *or* LDV, τ_{att} lies between $\widehat{\tau}_{\text{LDV}}$ and $\widehat{\tau}_{\text{DID}}$:

$$\mathbb{E}[\widehat{\tau}_{\text{LDV}}] \geq \tau_{\text{att}} \geq \mathbb{E}[\widehat{\tau}_{\text{DID}}]$$

    - Use as upper/lower bounds; holds nonparametrically (Ding & Li, 2019).

Nonparametric Identification Strategy

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | \mathbf{X}_i, G_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | \mathbf{X}_i, G_i = 0]$$

- What does this assumption say?
    - Potential trend under control is the same for the control & treated groups, conditional on covariates.
        - $\rightsquigarrow$ Units that are similar at baseline will follow similar paths under no treatment.
- **Matching**: conduct DID analysis on units with similar values of $\mathbf{X}_i$
    - *We will discuss more when we cover matching methods (e.g., PSM, CEM)!*
- Beyond covariate matching, **SC/SDiD** families also handle fragile PT (see handout 10(b)).

Inference Caveats (Bertrand-Duflo-Mullainathan, 2004)

- DID with many time periods and serially correlated outcomes: **standard OLS SEs are dangerously anti-conservative**.
- Three reinforcing factors specific to DID:
  - Long panels (DD literature averages ≈ 16.5 periods).
  - Outcomes typically have strong positive serial correlation.
  - Treatment indicator I_{st} changes very little within a unit over time.
- Placebo evidence: random "laws" on CPS female wages, 21-year panel.
  - Nominal 5% test $\rightsquigarrow$ empirical rejection rate of true null up to **45%**.
- Solutions tested, by sample size of clusters:
  - Parametric AR(1) correction: *poor* (mis-specifies the DGP).
  - **Block bootstrap**: good when number of states is large.
  - **Cluster-robust SE at the level of treatment assignment**: good for moderate cluster counts.
  - **Collapse panel to 2-period** (pre vs. post averages): good for small cluster counts.

TWFE Decomposition (Goodman-Bacon, 2021)

- TWFE specification with $T > 2$ periods and staggered timing:

$$Y_{it} = \alpha_i + \lambda_t + \tau D_{it} + \varepsilon_{it}$$

- Goodman-Bacon (2021): $\widehat{\tau}_{\text{TWFE}}$ is a **weighted average of all** 2×2 **DID comparisons**, weights $\propto$ group sizes $\times$ within-group treatment variance.
- Three flavors of 2×2 (same cohort can be "not-yet-treated" or "already-treated" depending on the comparison window):
  - Treated cohort vs. never-treated: **clean**.
  - Early-treated vs. later cohort, *before later's adoption* (later = not-yet-treated): **clean**.
  - Late-treated vs. earlier cohort, *after earlier's adoption* (earlier = already-treated, still under its own TE): **forbidden**.

Group-Time ATTs (Callaway & Sant'Anna, 2021)

- Identification target: ATT(g, t), the ATT for cohort g (units first treated at time g) at calendar time t.

$$\text{ATT}(g, t) = \mathbb{E}[Y_t(1) - Y_t(0) \mid G_i = g]$$

- Each ATT(g, t) is a 2×2 DID: cohort g vs. **never-treated** (or **not-yet-treated**) as control; estimated via OR / IPW / **doubly-robust** (Sant'Anna & Zhao, 2020). No already-treated unit ever used as control.
- Aggregation menu: overall ATT, event-study by exposure $e = t - g$, dynamic / cohort / calendar-time effects.
- Pre-trend testing: ATT(g, t) for $t < g$ should be 0 under PT. R: `did::att_gt()` + `did::aggte()`.
- Other alternatives in the same spirit: **Sun & Abraham (2021, JoE)**; **Borusyak, Jaravel & Spiess (2024, ReStud)**. CSA's ATT(g, t) is the most widely adopted in applied work.

Recent Discussions on DiD

- **Baker, Larcker & Wang (2022, JFE)** "How Much Should We Trust Staggered DID Estimates?"
  - Of 744 DD papers in top-5 finance/accounting journals (2000–2019), **55% use staggered DID**.
  - Replications of Beck et al. (2010, JF) and Fauver et al. (2017, JFE): **TWFE estimates change substantively** under CSA / SA / stacked DID ⇝ re-examine published findings.

- **Roth, Sant'Anna, Bilinski & Poe (2023, JoE)** "What's Trending in DID?" practitioner checklist:
  1. *Is everyone treated at the same time?*
     - Yes ⇝ classic 2×2 or event-study; No ⇝ CSA / SA / BJS.
  2. *Is parallel trends defensible?*
     - Conditional PT (covariates) and/or sensitivity analysis (Rambachan-Roth, R: `HonestDiD`).
  3. *Many treated and untreated clusters?*
     - Yes ⇝ cluster-robust SE; No ⇝ block bootstrap or randomization inference (Bertrand et al., 2004).

2/ Empirical Demonstration in R

Empirical Demonstration: Roadmap

- Two parts, each run in R with full console input + output:
- **Part A** (Card & Krueger, 1994): canonical 2×2 DID on NJ-PA fast-food data.
    - Illustrate the LDV bracketing inequality $\mathbb{E}[\widehat{\tau}_{\text{LDV}}] \geq \tau_{\text{att}} \geq \mathbb{E}[\widehat{\tau}_{\text{DID}}]$ (Ding & Li, 2019) in actual data.
- **Part B** (Cheng & Hoekstra, 2013, JHR): Castle Doctrine (criminal- / self-defense policy) staggered adoption $\rightsquigarrow$ staggered DiD.
    - TWFE $\rightsquigarrow$ Goodman-Bacon decomposition $\rightsquigarrow$ Callaway-Sant'Anna.
    - Reveal **sign-flip across estimators** and the underlying cohort heterogeneity.
- Synthetic Control and SDiD empirical demo (Cal Prop 99) are in companion file `ECI 10(b).` Handout.

Card-Krueger Data Setup

```
> library(tidyverse); library(fixest); library(estimatr)
> minwage <- read_csv(url("https://bit.ly/3QZmodH"), show_col_types = FALSE) |>
+   mutate(state   = if_else(location == "PA", "PA", "NJ"),
+          treated = if_else(location == "PA", 0, 1),
+          part_prop_before = partBefore / (fullBefore + partBefore),
+          part_prop_after  = partAfter  / (fullAfter  + partAfter),
+          trend            = part_prop_after - part_prop_before)

> head(minwage |> select(chain, location, state, treated,
+                        part_prop_before, part_prop_after, trend), 4)
# A tibble: 4 x 7
  chain      location state treated part_prop_before part_prop_after  trend
  <chr>      <chr>    <chr>   <dbl>            <dbl>           <dbl>  <dbl>
1 wendys     PA       PA          0           0.5             1      0.500
2 wendys     PA       PA          0           0.8125          0.0968 -0.716
3 burgerking PA       PA          0           0.4118          0.5455  0.134
4 burgerking PA       PA          0           0.6296          0.2571 -0.372

> minwage |> count(location)
# A tibble: 5 x 2
  location      n
1 PA           67
2 centralNJ    45
3 northNJ     146
4 shoreNJ      33
5 southNJ      67
```

Canonical 2×2 DID + First-Difference Regression

```
# (A) Canonical 2x2 DID via difference-in-means on trend = Y_post - Y_pre
> n1 <- sum(minwage$treated); n0 <- sum(1 - minwage$treated); c(n1, n0)
[1] 291  67
> tau_DID <- mean(minwage$trend[minwage$treated == 1]) -
+            mean(minwage$trend[minwage$treated == 0])
> tau_DID
[1] -0.06155831

# (B) Regression implementation: lm_robust on trend ~ treated, HC2 SE
> fd_fit <- lm_robust(trend ~ treated, data = minwage, se_type = "HC2")
> summary(fd_fit)$coefficients[, c("Estimate", "Std. Error", "t value", "Pr(>|t|)")]
               Estimate Std. Error    t value  Pr(>|t|)
(Intercept)  0.03768357 0.04195211  0.8982522 0.3696582
treated     -0.06155831 0.04551265 -1.3525540 0.1770566
```

Same point estimate via both routes; HC2 SE on the FD regression coincides with the Neyman diff-in-means SE (Samii & Aronow, 2012).

TWFE Regression with Chain-Clustered SE

```
# Reshape to long format, run TWFE with state + year FE, cluster by chain
> minwage_long <- minwage |>
+   pivot_longer(cols = c(part_prop_after, part_prop_before),
+                names_to = "post", values_to = "part_prop") |>
+   mutate(post = if_else(post == "part_prop_before", 0, 1),
+          year = if_else(post == 0, 1991, 1992))
> twfe_fit <- feols(part_prop ~ treated:post | state + year,
+                   data = minwage_long, cluster = ~chain)
> summary(twfe_fit)
OLS estimation, Dep. Var.: part_prop
Observations: 716
Fixed-effects: state: 2,  year: 2
Standard-errors: Clustered (chain)
              Estimate Std. Error  t value Pr(>|t|)
treated:post -0.061558   0.021945 -2.80517 0.067562 .
---
RMSE: 0.240757     Adj. R2: -2.787e-4     Within R2: 0.00248
```

Same point estimate as the FD regression; clustering on `chain` (Burger King, KFC, Wendy's, . . .) tightens the SE substantively because within-chain shocks are correlated.

LDV Regression and Bracketing Check

```
# LDV regression: Y_post = alpha + rho * Y_pre + tau * G
> ldv_fit <- lm_robust(part_prop_after ~ part_prop_before + treated,
+                      data = minwage, se_type = "HC2")
> coef(ldv_fit)
     (Intercept) part_prop_before          treated
       0.5940267        0.1937457       -0.0507227

# Bracketing check (0 <= rho < 1; treated has higher baseline Y_0)
> c(tau_LDV = coef(ldv_fit)["treated"], tau_DID = tau_DID,
+   baseline_gap = mean(minwage$part_prop_before[minwage$treated == 1]) -
+                  mean(minwage$part_prop_before[minwage$treated == 0]))
     tau_LDV.treated         tau_DID    baseline_gap
        -0.05072270    -0.06155831      0.01342929

# Theory: tau_LDV >= tau_ATT >= tau_DID -> observed: -0.0507 >= tau_ATT >= -0.0616  (verified)
```

Castle Doctrine (Cheng & Hoekstra, 2013, JHR): TWFE Baseline

21 US states adopted "Castle Doctrine" / "Stand Your Ground" laws in 2005–2009, expanding lethal-force self-defense rights beyond the home. Original finding: +8% homicide, no deterrent effect on burglary/robbery ⇝ violence *escalation*, not deterrence.

```
> library(bacondecomp); library(fixest); library(did)
> data(castle, package = "bacondecomp")  # 50 states x 11 years (2000-2010)
> table(if_else(is.na(castle$effyear), 0L, as.integer(castle$effyear)))
   0 2005 2006 2007 2008 2009
 319   11  143   44   22   11

# TWFE estimate (clustered by state)
> twfe_fit <- feols(l_homicide ~ post | state + year, data = castle, cluster = ~state)
> summary(twfe_fit)
OLS estimation, Dep. Var.: l_homicide
Observations: 550
Fixed-effects: state: 50,  year: 11
Standard-errors: Clustered (state)
     Estimate Std. Error t value Pr(>|t|)
post 0.081812   0.058874  1.3896  0.17093
---
RMSE: 0.176318     Adj. R2: 0.899604     Within R2: 0.013406
```

- Naive TWFE post coefficient: $\widehat{\tau} = +0.082$ (n.s.); suspect under staggered TE heterogeneity (recall Goodman-Bacon).

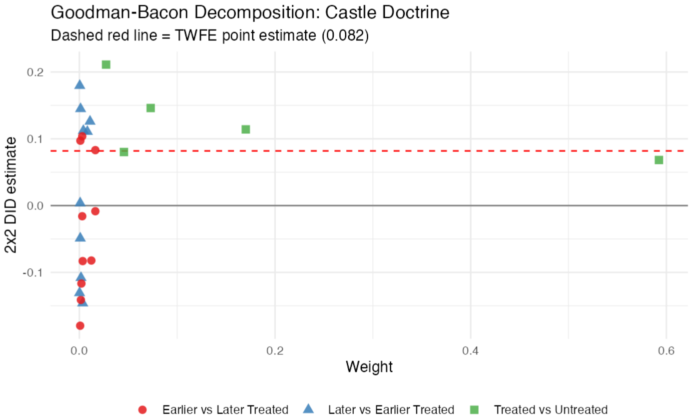
Goodman-Bacon Decomposition on Castle Data

```
> bacon_out <- bacon(l_homicide ~ post, data = castle,
+                    id_var = "state", time_var = "year", quietly = TRUE)
> bacon_out |> group_by(type) |>
+   summarize(total_weight = sum(weight),
+             avg_estimate = sum(estimate * weight) / sum(weight))
# A tibble: 3 x 3
  type                       total_weight avg_estimate
  <chr>                             <dbl>        <dbl>
1 Earlier vs Later Treated         0.0411      -0.0136
2 Later vs Earlier Treated         0.0508       0.0710
3 Treated vs Untreated             0.908        0.0880

> sum(bacon_out$estimate * bacon_out$weight)  # reconstructs TWFE
[1] 0.08181225
```

- TWFE point estimate (+0.082) is a weighted average across 25 different 2×2 DDs.
- “Treated vs Untreated” (clean) carries 91% of weight with avg estimate $+0.088$.
- “Later vs Earlier Treated” (forbidden) contributes $+0.071$ at 5% weight.

Goodman-Bacon Scatter (Castle Data)



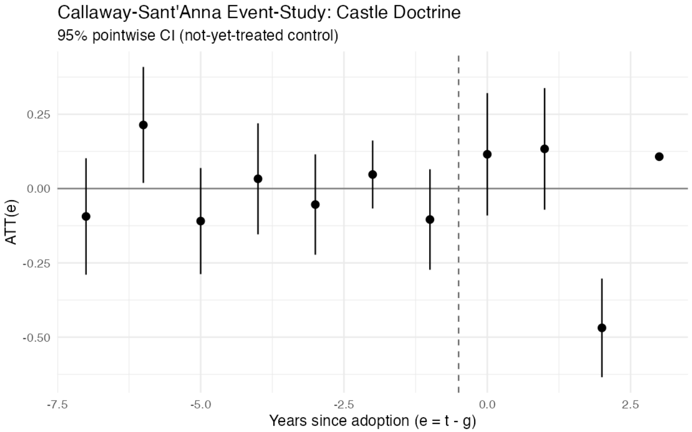
Each point: one 2×2 DID. x = weight, y = estimate; dashed red = TWFE estimate. "Treated vs Untreated" dominates, but forbidden-comparison types still inject variance.

Callaway-Sant'Anna Group-Time ATT (Castle Data)

```
> castle_cs <- castle |>
+   mutate(id = as.integer(factor(state)),
+          cohort = if_else(is.na(effyear), 0L, as.integer(effyear)))
> cs_fit <- att_gt(yname="l_homicide", tname="year", idname="id", gname="cohort",
+                  data=castle_cs, control_group="notyettreated",
+                  bstrap=TRUE, cband=TRUE, panel=TRUE)
> aggte(cs_fit, type="simple", na.rm=TRUE)
      ATT     Std. Error     [ 95%  Conf. Int.]
 -0.0329         0.057     -0.1446      0.0787
> aggte(cs_fit, type="dynamic", na.rm=TRUE)  # event-study
Dynamic Effects:
 Event time Estimate Std. Error [95% Simult. Conf. Band]
         -2   0.0473     0.0583       -0.0878      0.1823
         -1  -0.1041     0.0862       -0.3038      0.0955
          0   0.1153     0.1050       -0.1279      0.3584
          1   0.1334     0.1044       -0.1082      0.3751
          2  -0.4686     0.0847       -0.6646     -0.2726 *
```

- CSA overall ATT (not-yet-treated control): -0.033 (*opposite sign* from TWFE!).
- Event-study reveals positive on-impact effect but *negative* at $e = 2$ years.

CSA Event-Study Plot (Castle Data)



ATT by years since adoption. Pre-treatment estimates ($e < 0$) ≈ 0 supports parallel trends. Post-treatment small initially, negative and significant at $e = 2$.

Takeaway: TWFE vs CSA on Castle Doctrine

- Same data, two estimators:
  - TWFE post: $\widehat{\tau} = +0.082$ (SE 0.059, n.s.)
  - CSA overall: $\widehat{\tau} = -0.033$ (SE 0.057, n.s.)
  - $\rightsquigarrow$ Point estimates differ in **sign** (both n.s. in $N = 550$, but sign-flip is striking).

- Why? Strong **cohort heterogeneity** in ATT(g, t):
  - 2005 cohort: ATT$(2005, 2007) = +0.297$ (sig) – consistent with Cheng-Hoekstra's escalation story.
  - 2006, 2008 cohorts: ATT$(g, t) \approx -0.3$ to -0.5 at later event-times.
  - TWFE weights ($\propto$ group size $\times$ within-group treatment variance) favor the large 2006 cohort + clean Treated-vs-Untreated comparisons (91% of TWFE weight). CSA's uniform-over-cohorts aggregation treats each cohort equally $\rightsquigarrow$ sign flips.

- **Pedagogy**: A single TWFE number can hide sign-heterogeneity across cohorts. Decompose (Goodman-Bacon) and re-estimate (CSA) before reporting in any modern staggered design.

On to the Presentations & Discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/

Appendix

Basic Idea Behind Fixed Effects

- Given a panel data (repeated measurements of the same units),

$$\begin{aligned} Y_{it} &= \alpha + \tau D_{it} + \mathbf{X}'_{it}\beta + \varepsilon_{it} \\ &= \alpha + \tau D_{it} + \mathbf{X}'_{it}\beta + \underbrace{\gamma_i + \lambda_t + \upsilon_{it}}_{\varepsilon_{it} \text{ decomposed}} \end{aligned}$$

  - Individual fixed effects (i.e., γ_i): unobserved innate differences between individuals that are time-invariant.
    - Unobserved factors captured by γ_i, e.g.,:
      1) ability or preferences (at the individual-level),
      2) atmosphere or management style (at the firm-level), or
      3) brand image or fashion trend (at the product-level).
    - Observed factors? $\rightsquigarrow$ captured with the covariates (i.e., $\mathbf{X}'_{it}$)!
  - Time fixed effects (i.e., λ_t): Impacts common to all groups but vary by year (e.g., seasonality)
  - Idiosyncratic error term (i.e., υ_{it})

Basic Idea Behind Fixed Effects

- ⇝ FE model takes the original econometric error term (i.e., ε_{it}) and decomposes it into three pieces (i.e., γ_i, λ_t, and υ_{it})!
- Alternative to FE specification is random effects (e.g., Wooldridge, 2006) which "ignores" the indiv. & time FE (in cases where it will not create any bias from ignoring it).
  - In this case, the FEs remain as part of the econometric error term.
  - Wooldridge, Jeffrey M. "Introductory econometrics: A modern approach, 3rd." New York: Thomson 53 (2006).

Semiparametric Estimation w/ Repeated Outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}\left[Y_i(1) - Y_i(0) | G_i = 1\right] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

- Reweights control group to have the same distribution of $\mathbf{X}_i$ as treated group.
- Have to estimate the propensity score $\mathbb{P}(G_i = 1 \mid \mathbf{X}_i)$
    - Issue: possible model misspecification!