Econometrics III: Applied Methods
NovaSBE, 2026

Problem Set 3: Difference-in-Differences

Exercise A: DiD Foundations

A brief warm-up on the mechanics of DiD.

Card and Krueger (1994) studied the effect of New Jersey (NJ)'s 1992 minimum wage increase on fast-food employment, using Pennsylvania (PA) as a control. The outcome in the table shows average full-time equivalent (FTE) employment per restaurant:

	Before (Feb 1992)	**After** (Nov 1992)
New Jersey (treated)	20.44	21.03
Pennsylvania (control)	23.33	21.17

1.1 Compute the DiD estimate. Explain the parallel trends assumption for this setting and give one concrete reason why it might fail.

1.2 Suppose employment in both states follows the model:

$$Y_{st} = \alpha_s + \delta_t + \beta \cdot D_{st} + \gamma_s \cdot t + \varepsilon_{st}$$

where γ_s is a state-specific linear trend. Show that if $\gamma_{NJ} \neq \gamma_{PA}$, the DiD estimator is biased for β. Derive the bias term, and explain it.

1.3 A skeptic argues: "New Jersey's economy was stronger than Pennsylvania's in 1992, so employment would have grown faster in NJ even without the minimum wage increase." Translate this concern into the model from 1.2. Would DiD overestimate or underestimate the employment effect of the minimum wage?

1.4 Suppose workers can easily commute across the NJ–PA border for work. Explain what might happen when NJ raises its minimum wage. How might this cross-border mobility affect the DiD estimate? Would it lead to overestimation or underestimation of the employment effect?

Exercise B: Formal Analysis of Threats to Identification

This exercise develops formal models for common threats to DiD. For each threat, you will derive the bias that arises.

B.1 Ashenfelter's Dip and Mean Reversion

Consider a job training program. Let $Y_{it}(0)$ be potential earnings without training. Suppose:

$$Y_{it}(0) = \mu_i + \delta_t + \eta_{it}$$

where μ_i is a permanent individual component, δ_t is a common time effect, and η_{it} is a transitory shock with $\mathbb{E}[\eta_{it}] = 0$, $\text{Cov}(\eta_{it}, \eta_{it'}) = 0$ for $t \neq t'$, and $\eta_{it} \perp (\mu_i, \delta_t)$.

Workers decide whether to enroll in training ($D_i = 1$) based on their period-1 outcome. Specifically, they enroll if $Y_{i1}(0) < c$ for some threshold.

1.1 Explain intuitively why participants might enroll if $Y_{i1}(0) < c$ in this kind of context. Explain intuitively why $\mathbb{E}[\eta_{i1}|D_i = 1] < 0$.

1.2 Assume the true treatment effect is τ (so $Y_{i2} = Y_{i2}(0) + \tau \cdot D_i$). Derive an expression for the DiD estimator, defined as:

$$\hat{\tau}^{DiD} = \mathbb{E}[Y_{i2} - Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i2} - Y_{i1}|D_i = 0]$$

Show that DiD is biased and determine the sign of the bias.

1.3 This bias is sometimes called "mean reversion bias." Explain why: what happens to the transitory component η_{it} between periods 1 and 2, and why does this look like a treatment effect?

1.4 Would this bias show up in a pre-trends test if we had data from period 0?

B.2 Anticipation Effects

Consider a policy announced at $t = 0$ but implemented at $t = 1$. Let treatment be $D_i = 1$ for affected units. Suppose the true potential outcome model is:

$$Y_{it}(d) = \alpha_i + \delta_t + \tau^{post} \cdot d \cdot \mathbb{1}(t \geq 1) + \tau^{antic} \cdot d \cdot \mathbb{1}(t = 0) + \varepsilon_{it}$$

where τ^{post} is the post-implementation effect and τ^{antic} is the anticipation effect (behavioral response to the announcement).

The researcher, unaware of anticipation, defines $t = 0$ as "pre-treatment" and estimates:

$$\hat{\tau}^{DiD} = (\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$$

2.1 Derive the expression for the researcher's DiD estimate and explain why it is wrong.

2.2 The researcher runs an event study with periods $t \in \{-1, 0, 1, 2\}$, using $t = -1$ as the reference period. Derive the population values of β_0, β_1, and β_2.

2.3 Looking at the event study, how would the researcher detect anticipation effects? What pattern would signal this problem?

2.4 Suppose we are studying a corporate tax increase. If firms accelerate investment to $t = 0$ to avoid higher taxes in $t \geq 1$, what are the signs of τ^{antic} and τ^{post}? How does this affect the DiD estimate?

B.3 Heterogeneous Trends and the Bias Formula

Suppose the true model is:

$$Y_{it}(0) = \alpha_i + \delta_t + \gamma_i \cdot t + \varepsilon_{it}$$

where γ_i is a unit-specific linear trend. Treatment occurs at $t = T^*$ for treated units.

3.1 Show that parallel trends holds if and only if $\mathbb{E}[\gamma_i|D_i = 1] = \mathbb{E}[\gamma_i|D_i = 0]$.

3.2 Define $\bar{\gamma}_1 = \mathbb{E}[\gamma_i|D_i = 1]$ and $\bar{\gamma}_0 = \mathbb{E}[\gamma_i|D_i = 0]$. Show that for a 2×2 DiD comparing periods t and t' (with $t' > t \geq T^*$):

$$\text{Bias} = (\bar{\gamma}_1 - \bar{\gamma}_0) \cdot (t' - t)$$

3.3 Using the result from 3.2, explain why DiD bias grows with the length of the post-treatment window. What does this imply for research design?

3.4 Suppose treatment is more likely in regions with stronger economic growth (higher γ_i). What is the sign of the bias? Relate this to omitted variable bias: what is the "omitted variable" and why is it correlated with treatment?

Exercise C: Pre-testing Pitfalls (Roth, 2022)[1]

This exercise explores limitations of pre-trends testing, following Roth (2022).

C.1 The Power Problem

Use the dataset `pretrends_power.csv`, containing 1,000 simulated DiD studies. Variables: `study_id`, `beta_pre` (estimated pre-trend), `se_pre` (standard error), `beta_post` (estimated treatment effect). The true pre-trend is 0.5 (parallel trends violated) and the true treatment effect is 1.0.

1.1 For each study, test $H_0 : \beta_{pre} = 0$ at the 5% level. What fraction of studies "pass" (fail to reject)?

1.2 What fraction of studies can reject a pre-trend of magnitude 0.5 or larger? What does this say about power?

1.3 Compute the "breakdown pre-trend" for each study: $|\hat{\beta}_{pre}| + 1.96 \times SE_{pre}$. Explain what this quantity represents. What is the average breakdown pre-trend? How large is it compared to the true treatment effect? Comment.

1.4 Based on these results, explain why a non-significant pre-trend is weak evidence that parallel trends holds.

C.2 Selection Bias from Pre-testing

Continue with `pretrends_power.csv`.

2.1 Among studies that pass the pre-trends test, what is the average values of $\hat{\beta}_{pre}$ and $\hat{\beta}_{post}$? Compare to the true values from the data generating process.

2.2 Explain intuitively why conditioning on passing induces bias. What type of noise realizations are selected?

2.3 What are the implications for interpreting published DiD studies that report clean pre-trends?

C.3 Controlling for Linear Pre-trends

Use the dataset `linear_pretrends.csv`, containing panel data with a differential pre-trend (treated group grows faster). The true treatment effect is $\tau = 0.5$ (a level shift at treatment).

3.1 Estimate standard TWFE: $Y_{it} = \alpha_i + \delta_t + \beta \cdot D_{it} + \varepsilon_{it}$. Report $\hat{\beta}$.

3.2 Estimate TWFE with group-specific linear trends: $Y_{it} = \alpha_i + \delta_t + \gamma(D_i \times t) + \beta \cdot D_{it} + \varepsilon_{it}$. Report $\hat{\beta}$.

3.3 Plot mean outcomes over time for treated and control groups. Overlay the fitted pre-period linear trends (estimated from $t < 6$ only) extrapolated to all periods. What do you observe about the pre-trends?

3.4 Create a "residual variation" plot showing deviations from the fitted pre-period trends:

- For each group, fit a linear trend using pre-treatment data only ($t < 6$)

[1] Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights* 4(3): 305–22.

- Compute residuals as: $\text{resid}_{gt} = \bar{Y}_{gt} - \widehat{\text{trend}}_g(t)$
- Plot mean residuals over time, separately for treated and control

What identifying variation remains after removing linear trends? What does the "kink" at the treatment date represent?

3.5 What is the identifying assumption when controlling for linear trends? When is this appropriate vs. problematic?

C.4 Sensitivity Analysis with `HonestDiD` (Rambachan & Roth, 2023)

Rambachan & Roth (2023) propose a more principled approach to handling potential violations of parallel trends. Rather than assuming parallel trends holds exactly, they allow for violations but impose *smoothness restrictions*—the idea that violations of parallel trends cannot change too abruptly over time.

Use the dataset `event_study.csv`*, which contains panel data with variables* `unit`*,* `time` *($t = 1, \ldots, 10$),* `treated`*, and* `outcome`*. Treatment occurs at $t = 6$ for treated units.*

4.1 Estimate an event study specification with event-time dummies from $k = -4$ to $k = 4$, omitting $k = -1$. Create an event study plot with 95% confidence intervals.

In R: Use `fixest::feols()` with the `i()` syntax for event-time interactions.

In Stata: Create event-time dummies interacted with treatment, then use `reghdfe` with unit and time fixed effects. Omit the $k = -1$ dummy as the reference period.

4.2 Define the "relative magnitudes" restriction: the post-treatment violation of parallel trends is no more than $\bar{M}$ times the maximum pre-treatment violation. Formally:

$$|\delta_{post}| \leq \bar{M} \cdot \max_{k<0} |\delta_k|$$

where δ_k is the differential trend at event-time k. Explain intuitively why this is a reasonable assumption. When might it fail?

4.3 Using your event study estimates, compute robust confidence intervals for the treatment effect under different values of $\bar{M}$ (e.g., $\bar{M} \in \{0, 0.5, 1, 1.5, 2\}$).

In R: Install `HonestDiD` and use `createSensitivityResults_relativeMagnitudes()`. You will need to extract the coefficient vector and variance-covariance matrix from your event study regression.

In Stata: Install `honestdid` (via `ssc install honestdid`) and use `honestdid` after your event study regression.

4.4 Create a sensitivity plot showing how the confidence interval for the treatment effect changes as $\bar{M}$ increases. At what value of $\bar{M}$ does the confidence interval first include zero (the "breakdown value")?

4.5 Interpret your results. If the breakdown value is large, what does this say about the robustness of your findings?

4.6 Compare the `HonestDiD` approach to simply controlling for linear pre-trends (C.3.2). What are the advantages and disadvantages of each approach?

Exercise D: Staggered Treatment and TWFE

This exercise examines why two-way fixed effects (TWFE) can fail with staggered treatment timing and dynamic treatment effects.

Consider a setting with $T = 8$ time periods and four groups:

- **Early** (E): treated starting at $t = 2$
- **Middle** (M): treated starting at $t = 4$
- **Late** (L): treated starting at $t = 6$
- **Never** (N): never treated

Treatment effects are *dynamic*, growing linearly with time since treatment:

$$\tau_{g,t} = \tau_0 + \delta \cdot (t - g) \quad \text{for } t \geq g$$

where $\tau_0 = 1$ is the immediate effect and $\delta = 0.5$ captures growth, and g denotes the time that the group started treatment (e.g., $g = 2$ for group E, $g = 4$ for group M, etc.).

1 Compute the treatment effect for each cohort at $t = 8$: find $\tau_{E,8}$, $\tau_{M,8}$, and $\tau_{L,8}$.

2 Goodman-Bacon (2021) shows that TWFE is a weighted average of binary comparisons, including "forbidden" comparisons that use *already-treated* units as controls. Explain why using already-treated units as controls is problematic when effects are dynamic. What bias does this introduce?

3 *(R or Stata)* Use the dataset `staggered_did.csv` with variables `unit`, `time`, `cohort` (2, 4, 6, or 99 for never), `treat_post`, and `outcome`. Estimate the TWFE regression:

$$Y_{it} = \alpha_i + \delta_t + \beta \cdot D_{it} + \varepsilon_{it}$$

Report $\hat{\beta}$.

4 *(R or Stata)* Estimate three "clean" DiDs using only never-treated as controls, comparing $t = 1$ (pre) to $t = 8$ (post):

(a) Early vs. Never
(b) Middle vs. Never
(c) Late vs. Never

Report all three estimates. Compare to your answer in question 1.

5 Compare the TWFE estimate (question 3) to the simple average of your clean DiD estimates (question 4). Is TWFE higher or lower? Explain the difference using your answer to question 2.

6 Under what conditions is the standard TWFE estimator valid despite staggered timing? What must be true about δ?