Advanced DiD Mixtape Workshop
Violations of Parallel Trends

Jonathan Roth

June 22, 2022

Staggered Timing

- Remember that in the canonical DiD model we had:
  - Two periods and a common treatment date
  - Identification from parallel trends and no anticipation
  - A large number of clusters for inference
- A second literature has focused on relaxing the second assumption: **what if parallel trends may be violated?**
- The ideas from this literature apply even if there is non-staggered timing, although as we'll see, many of the tools can be applied with staggered timing as well. Large clusters is maintained throughout.

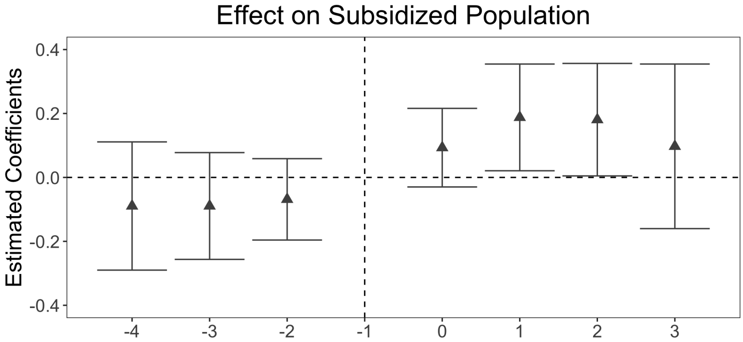
Violations of parallel trends

- Three substrands of this literature:
  - **Parallel trends only conditional on covariates**
  - **Testing for violations of (conditional) parallel trends**
  - **Sensitivity analysis and bounding exercises**
- I will focus on the latter two, (somewhat selfishly) focusing mainly on my own work
  - Roth (Forthcoming, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends")
  - Rambachan and Roth (2022, "A More Credible Approach to Parallel Trends")

Testing for pre-trends

- In practice, we're often unsure *ex ante* about the validity of the parallel trends assumption
- A nice feature of the DiD design is it has a built-in plausibility test:
  Were the groups moving in parallel prior to the treatment?
- Testing for pre-existing trends is a very natural way to assess the plausibility of the PT assumption
- But it also has several *limitations*, highlighted in recent work (Freyaldenhoven et al., 2019; Kahn-Lang and Lang, 2020; Bilinski and Hatfield, 2018b; Roth, Forthcoming)

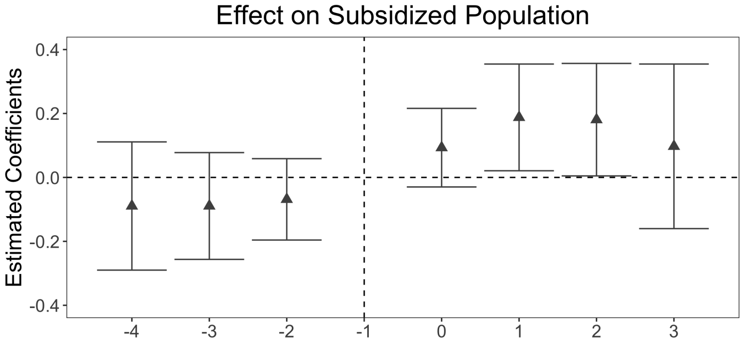
Issue 1 - Low Power



- He & Wang (2017) study impacts of placing college grads as village officials in China
- Use an "event-study" approach comparing treated and untreated villages

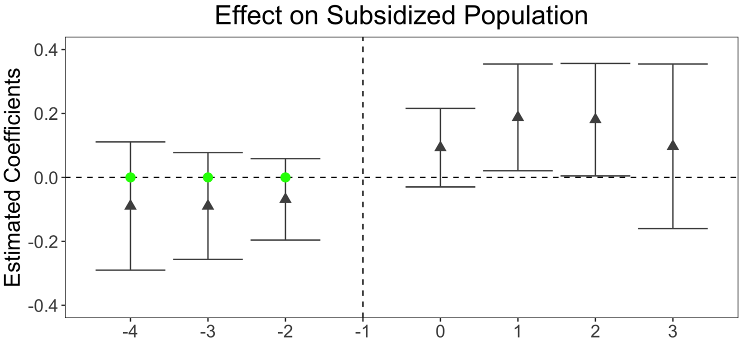
$$Y_{it} = \sum_{k \neq -1} D_{it}^k \beta_k + \alpha_i + \phi_t + \epsilon_{it}$$

Issue 1 - Low Power



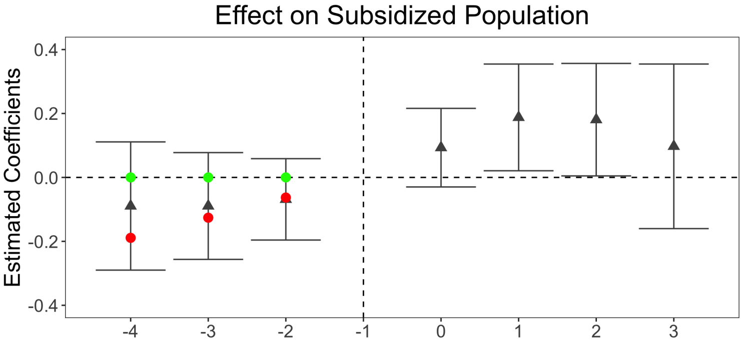
"The estimated coefficients on the leads of treatment ... are statistically indifferent from 0. ... We conclude that the pretreatment trends in the outcomes in both groups of villages are similar, and villages without CGVOs can serve as a suitable control group for villages with CGVOs in the treatment period." (He and Wang, 2017)

Issue 1 - Low Power



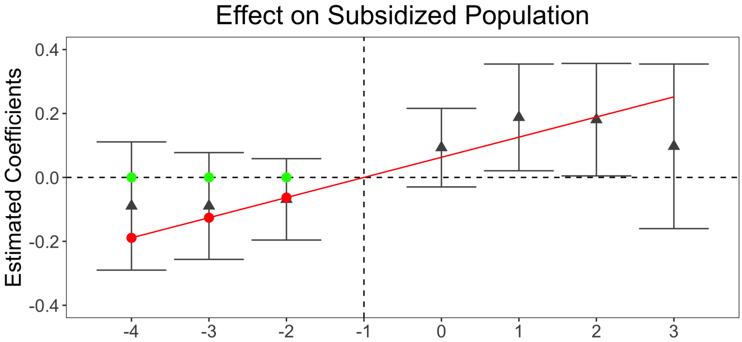
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81

Issue 1 - Low Power



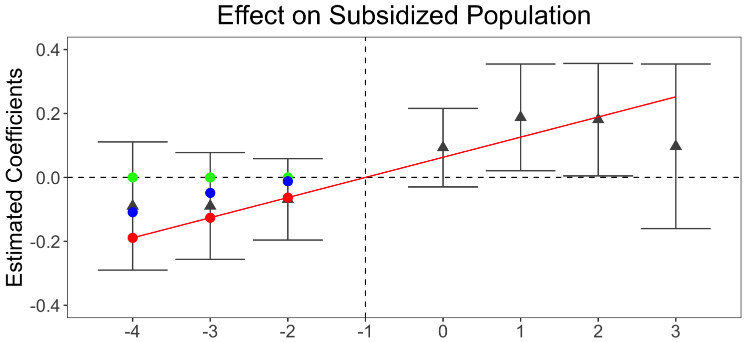
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81

Issue 1 - Low Power



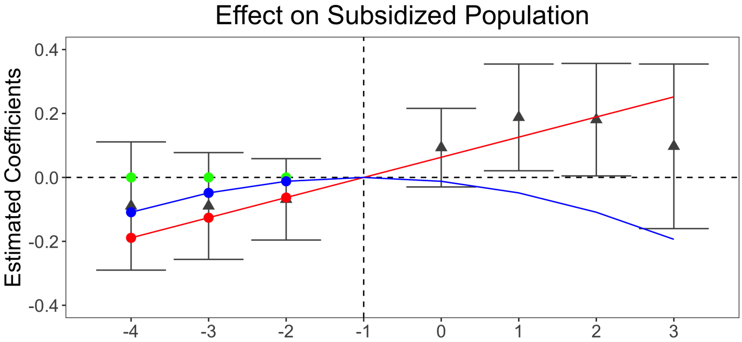
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81

Issue 1 - Low Power



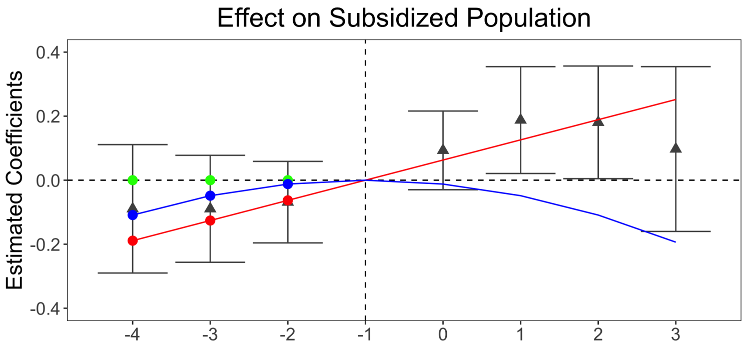
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81
- P-value for $H_0 : \beta_{pre} =$ blue dots: 0.81

Issue 1 - Low Power



- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81
- P-value for $H_0 : \beta_{pre} =$ blue dots: 0.81

Issue 1 - Low Power



- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81
- P-value for $H_0 : \beta_{pre} =$ blue dots: 0.81
- We can't reject zero pre-trend, but we also can't reject pre-trends that under smooth extrapolations to the post-treatment period would produce substantial bias

More systematic evidence

- Roth (forthcoming): simulations calibrated to papers published in *AER*, *AEJ: Applied*, and *AEJ: Policy* between 2014 and mid-2018
  - 70 total papers contain an event-study plot; focus on 12 w/available data
- Evaluate properties of standard estimates/CIs under linear violations of parallel trends against which conventional tests have limited power (50 or 80%):
  1. Bias often of magnitude similar to estimated treatment effect
  2. Confidence intervals substantially undercover in many cases
  3. Distortions from pre-testing can further exacerbate these issues

Issue 2 - Distortions from Pre-testing

- When parallel trends is violated, we will sometimes fail to find a significant pre-trend
- But the draws of data where this happens are a **selected sample**. This is known as *pre-test bias*.
- Analyzing this selected sample introduces additional statistical issues, and can make things worse!

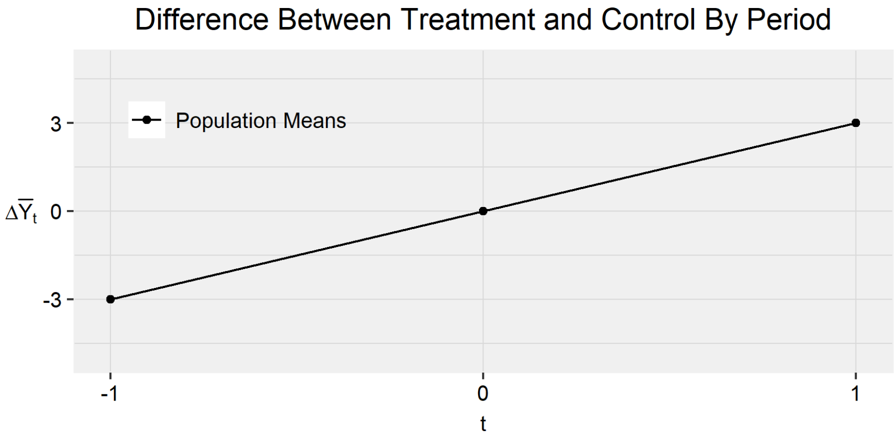
Stylized Three-Period DiD Example

- Consider a 3-period model ($t = -1, 0, 1$) where treatment occurs in last period

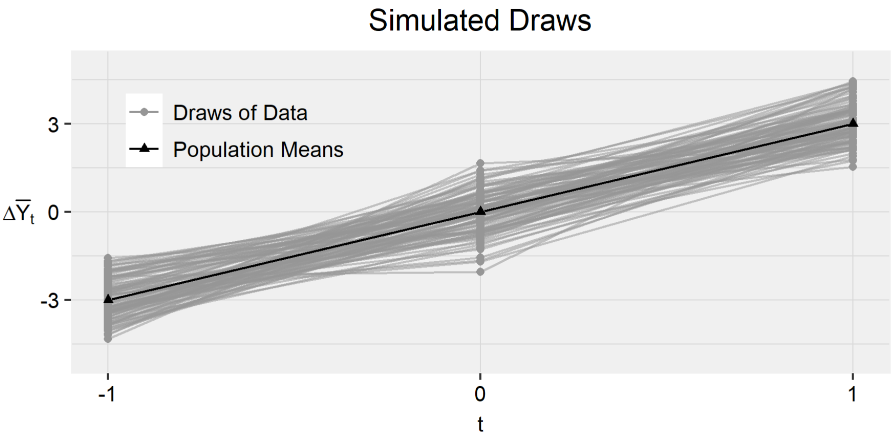
- No causal effect of treatment: $\tau = 0$

- In population, treatment group is on a linear trend relative to the control group with slope δ
  - Control group mean in period t: $\mu_t^C = 0$
  - Treatment group mean in period t: $\mu_t^T = \delta t$

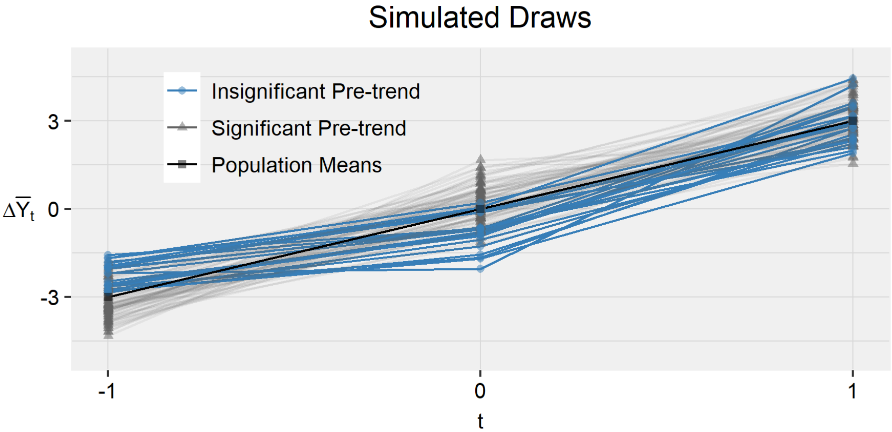
- Realized outcomes:
  - $\bar{Y}_t^C = \mu_t^C + \epsilon_t^C$
  - $\bar{Y}_t^T = \mu_t^T + \epsilon_t^T$
  - Independent normal errors: $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$



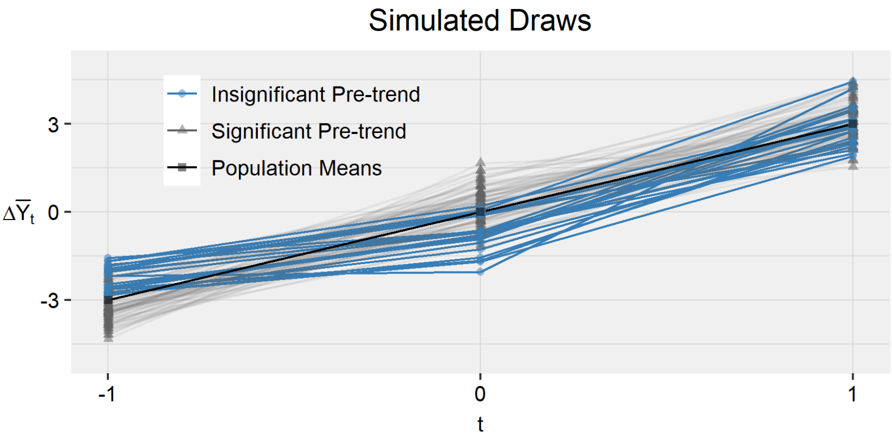
- Example: In population, there is a linear difference in trend with slope 3



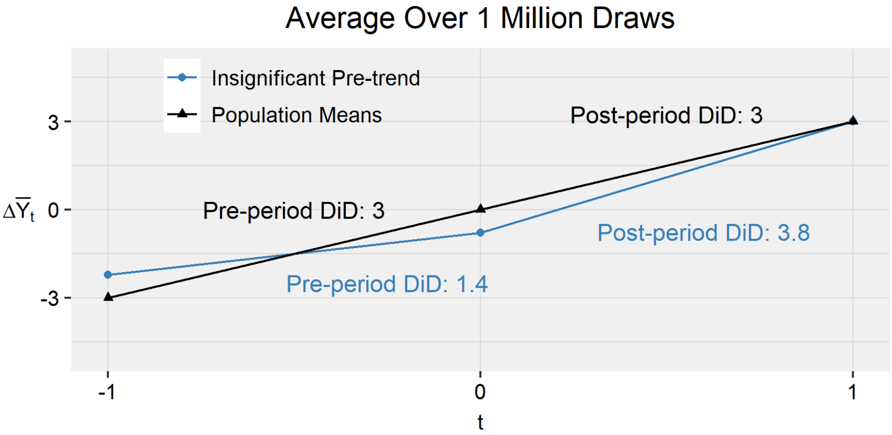
- Example: In population, there is a linear difference in trend with slope 3
- In actual draws of data, there will be noise around this line



- Example: In population, there is a linear difference in trend with slope 3
- In some of the draws of the data, highlighted in blue, the difference between period -1 and 0 will be insignificant



- In some of the draws of the data, highlighted in blue, the difference between period -1 and 0 will be insignificant
- In the insignificant draws, we tend to underestimate the difference between treatment and control at $t = 0$



- In the insignificant draws, we tend to underestimate the difference between treatment and control at $t = 0$
- As a result, the DiD between period 0 and 1 tends to be particularly large when we get an insignificant pre-trend

To Summarize

What are the Limitations of Pre-trends Testing?

1. Low Power – May not find significant pre-trend even if PT is violated
2. Pre-testing Issues – Selection bias from only analyzing cases with insignificant pre-trend
3. If reject pre-trends test, what comes next?

To Summarize

What are the Limitations of Pre-trends Testing?

1. Low Power – May not find significant pre-trend even if PT is violated
2. Pre-testing Issues – Selection bias from only analyzing cases with insignificant pre-trend
3. If reject pre-trends test, what comes next?

What Can We Do About It?

1. Diagnostics of power and distortions from pre-testing (Roth, Forthcoming, "Pre-Test with Caution..."). See `pretrends` package. Details
2. Formal sensitivity analysis that avoids pre-testing (Rambachan and Roth, "A More Credible Approach..."). See `HonestDiD` package.

"A More Credible Approach to Parallel Trends"

- The intuition motivating pre-trends testing is that the pre-trends are informative about counterfactual post-treatment trends
- Formalize this by imposing the restriction that the counterfactual difference in trends can't be "too different" than the pre-trend
- This allows us to bound the treatment effect and obtain uniformly valid ("honest") confidence sets under the imposed restrictions
- Enables **sensitivity analysis:** How different would the counterfactual trend have to be from the pre-trends to negate a conclusion (e.g. a positive effect)?

Restrictions on Violations of PT

- Consider the 3-period model ($t = -1, 0, 1$) where treatment occurs in last period

- Let δ_1 be the violation of PT:

$$\delta_1 = \mathbb{E}\left[Y_{i,t=1}(0) - Y_{i,t=0}(0) \,|\, D_i = 1\right] - \mathbb{E}\left[Y_{i,t=1}(0) - Y_{i,t=0}(0) \,|\, D_i = 0\right]$$

- We don't directly identify δ_1, but we do identify its pre-treatment analog, δ_{-1}:

$$\delta_{-1} = \mathbb{E}\left[Y_{i,t=-1}(0) - Y_{i,t=0}(0) \,|\, D_i = 1\right] - \mathbb{E}\left[Y_{i,t=-1}(0) - Y_{i,t=0}(0) \,|\, D_i = 0\right]$$

- Key idea: restrict possible values of δ_1 given δ_{-1}
  Intuitively, counterfactual trend can't be too different from pre-trend

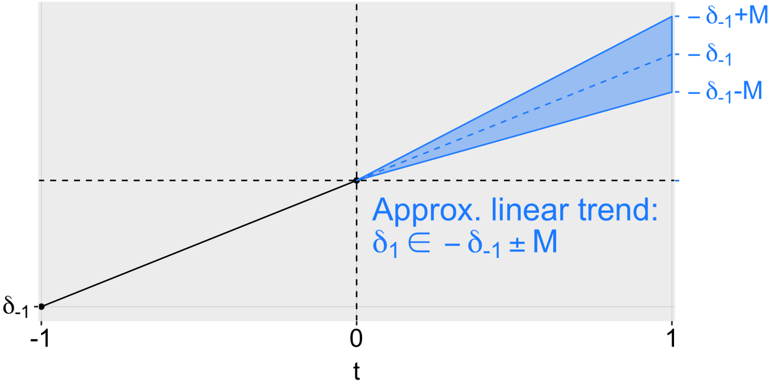
Examples of Restrictions on δ

- **Bounds on relative magnitudes:** Require that $|\delta_1| \leq \bar{M}|\delta_{-1}|$

Examples of Restrictions on δ

- **Bounds on relative magnitudes:** Require that $|\delta_1| \leq \bar{M}|\delta_{-1}|$

- **Smoothness restriction:** Bound how far δ_1 can deviate from a linear extrapolation of the pre-trend: $\delta_1 \in [-\delta_{-1} - M, -\delta_1 + M]$



Benzarti & Carloni (2019)

- BC study the incidence of a cut in the value-added tax on sit-down restaurants in France. France reduced the VAT on restaurants from 19.6 to 5.5 percent in July of 2009.

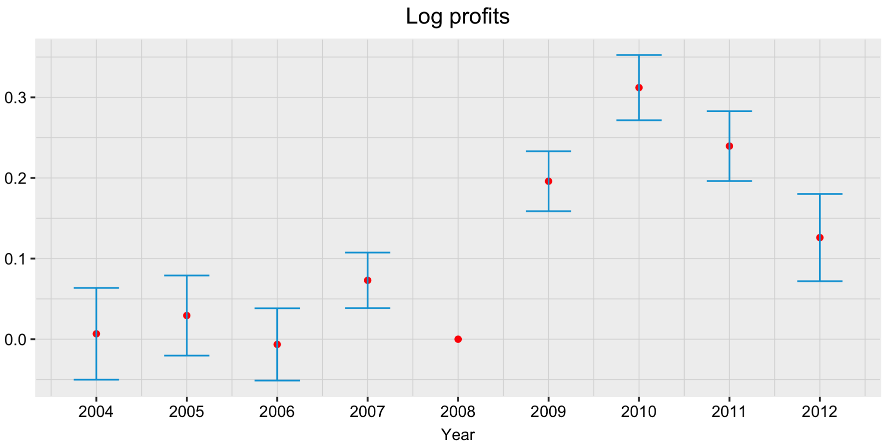
- BC analyze the impact of this change using a difference-in-differences design comparing restaurants to a control group of other market services firms

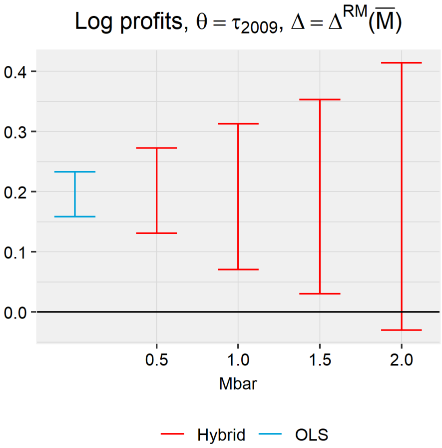
$$Y_{irt} = \sum_{s=2004}^{2012} \beta_s \times 1[t = s] \times D_{ir} + \phi_i + \lambda_t + \epsilon_{irt}, \tag{1}$$

  - Y_{irt} = outcome of interest for firm i in region r
  - D_{ir} = indicator if firm i in region r is a restaurant
  - Φ_i, λ_t = firm and year FEs

- Outcomes of interest include firm profits, prices, wage bill & employment. We focus on impact on profits in first year after reform.

Event-study coefficients for log profits





- "Breakdown" $\bar{M}$ for null effect is ~ 2
- Can rule out a null effect unless allow for violations of PT 2x larger than the max in pre-period

So to summarize!

- Tests of pre-trends are intuitive but not a panacea!
- In particular, they may suffer from low power and introduce pre-test bias
- Roth (Forthcoming) and Rambachan and Roth (2022) provide tools for diagnostics and sensitivity analysis
- And these tools play nicely with recent estimators developed for heterogeneous treatment effects; see Pedro Sant'Anna's website for examples!

Other Related Papers

- Other bounding exercises (Manski and Pepper, 2018; Ye et al., 2021)
- Non-inferiority approaches to pre-testing (Bilinski and Hatfield, 2018a; Dette and Schumann, 2020)
- Impose structure on the confounds (Freyaldenhoven et al., 2019)

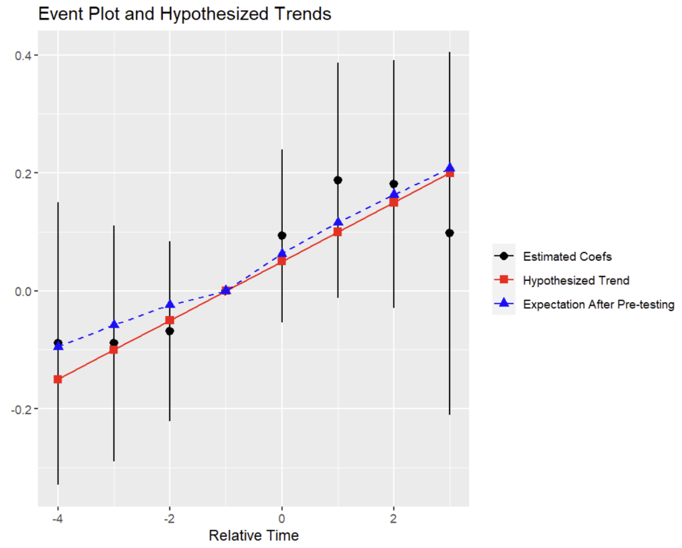
Thank you!

Additional Resources

- Roth (2021), "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends"
  - Paper; `staggered` package ; Shiny app
- Rambachan and Roth (2022), "A More Credible Approach to Parallel Trends"'
  - Paper; `HonestDiD` package ; Vignette

Pre-testing Diagnostics

- A "low-touch" intervention is to evaluate the likely power/distortions from pre-testing under *context-relevant* violations of parallel trends
- Enter the `pretrends` package / Shiny app


Event Plot and Hypothesized Trends
0.4
0.2
0.0
-0.2
-4
-2
0
2
Relative Time
Estimated Coefs
Hypothesized Trend
Expectation After Pre-testing

Power	Bayes.Factor	Likelihood.Ratio
0.33	0.76	1.23

- **Power.** Chance find significant pre-trend under hypothesized trend.
- **Bayes Factor.** Relative chance you pass the pre-test under hypothesized trend versus under parallel trends.
- **Likelihood Ratio.** Likelihood of observed pre-trend coefs under hypothesized trend versus under parallel trends.

Pros and Cons

Pros

- Very intuitive, easy to visualize.
- Helps identify when pre-testing may be least effective
- Requires minimal changes from standard practice

Cons

- Power will always be < 1, so no guarantee of unbiasedness/correct inference
- Need to specify the hypothesized trend. Will sometimes be difficult to summarize over many of these.
- Still not clear what to do when reject the pre-test.

Bilinski, Alyssa and Laura A. Hatfield, "No Free Lunch: A non-inferiority approach to model assumption tests," *arXiv:1805.03273 [stat]*, April 2018.

_ **and** _ , "Seeking evidence of absence: Reconsidering tests of model assumptions," *arXiv:1805.03273 [stat]*, May 2018.

Dette, Holger and Martin Schumann, "Difference-in-Differences Estimation Under Non-Parallel Trends," *Working Paper*, 2020.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro, "Pre-event Trends in the Panel Event-Study Design," *American Economic Review*, 2019, *109* (9), 3307–3338.

Kahn-Lang, Ariella and Kevin Lang, "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business & Economic Statistics*, 2020, *38* (3), 613–620.

Manski, Charles F. and John V. Pepper, "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *The Review of Economics and Statistics*, 2018, *100* (2), 232–244.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, Forthcoming.

Ye, Ting, Luke Keele, Raiden Hasegawa, and Dylan S. Small, "A Negative Correlation Strategy for Bracketing in Difference-in-Differences," *arXiv:2006.02423 [econ, stat]*, 2021.