Testing and Sensitivity Analysis for Violations of Parallel Trends

Jonathan Roth

Microsoft

Introduction

- Difference-in-differences and related methods rely on a **parallel trends** assumption.
- In practice, we're often not sure if parallel trends holds!
- Common practice: **test for pre-trends** to assess plausibility.
- Pre-trends testing is *very intuitive*. But it has *limitations.*
- This talk is about those limitations and what we can do about them.

Overview

Limitations of Pre-testing?

1. Low Power – Pre-test may fail to detect violations of PT
2. Distortions from Pre-testing – Selection bias from only analyzing cases with insignificant pre-trend
3. If reject pre-test, what comes next?

Overview

Limitations of Pre-testing?

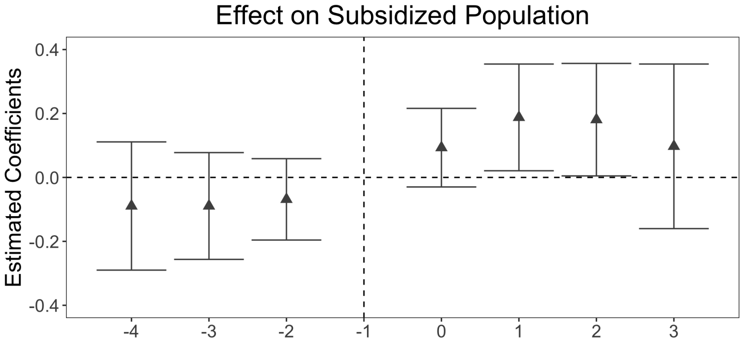
1. Low Power – Pre-test may fail to detect violations of PT
2. Distortions from Pre-testing – Selection bias from only analyzing cases with insignificant pre-trend
3. If reject pre-test, what comes next?

What Can We Do About It?

1. Diagnostics of power and distortions from pre-testing (Roth, 2021, "Pre-Test with Caution..."). See `pretrends` package.
2. Formal sensitivity analysis (Rambachan and Roth, 2020, "An Honest Approach..."). See `HonestDiD` package.

- Focus today will be on problems and implementation of solutions. (Skipping econometric sausage).

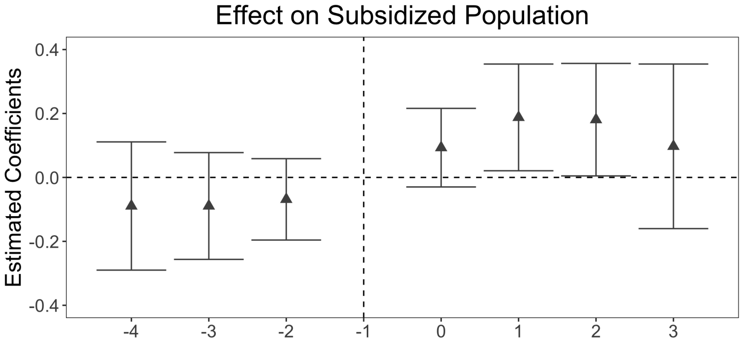
Issue 1 - Low Power



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

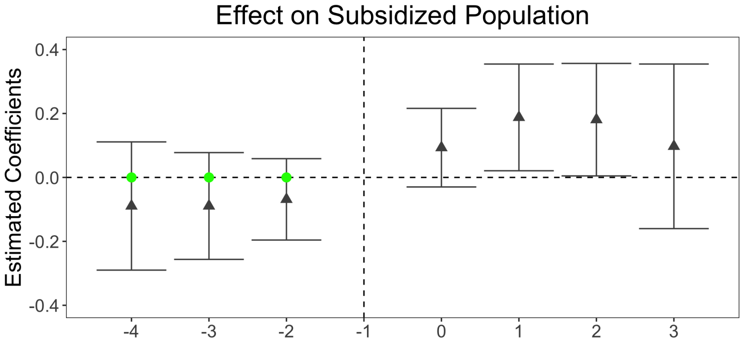
$$Y_{it} = \sum_{k \neq -1} D_{it}^{k} \beta_k + \alpha_i + \phi_t + \epsilon_{it}$$

Issue 1 - Low Power



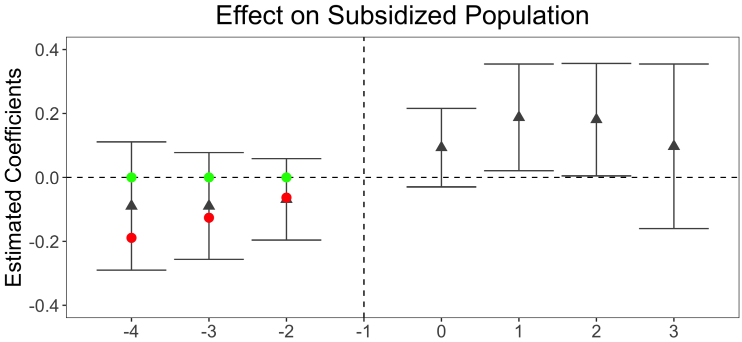
"The estimated coefficients on the leads of treatment ... are statistically indifferent from 0. ... We conclude that the pretreatment trends in the outcomes in both groups of villages are similar, and villages without CGVOs can serve as a suitable control group for villages with CGVOs in the treatment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

Issue 1 - Low Power



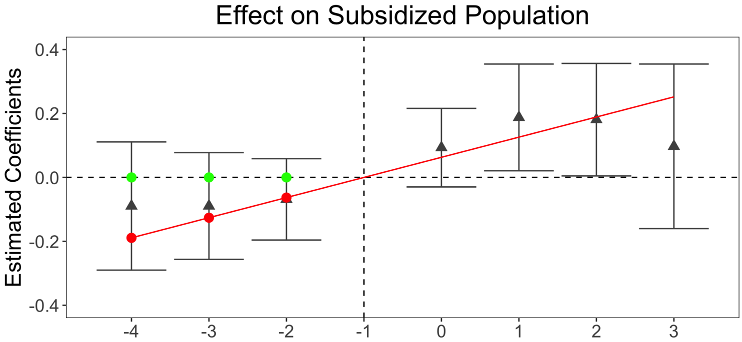
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81

Issue 1 - Low Power



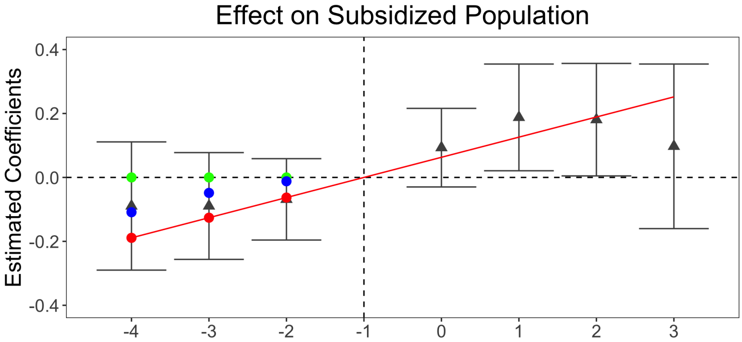
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81

Issue 1 - Low Power



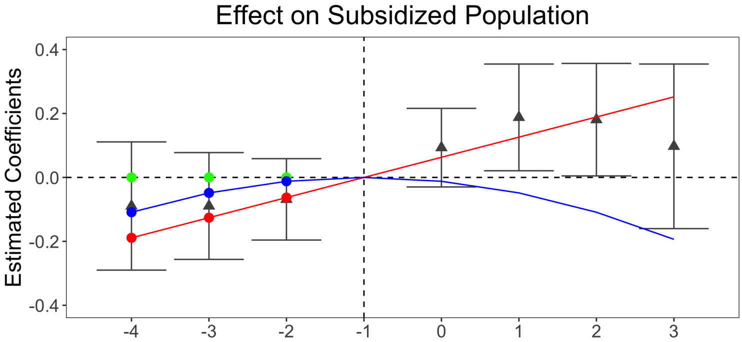
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81

Issue 1 - Low Power



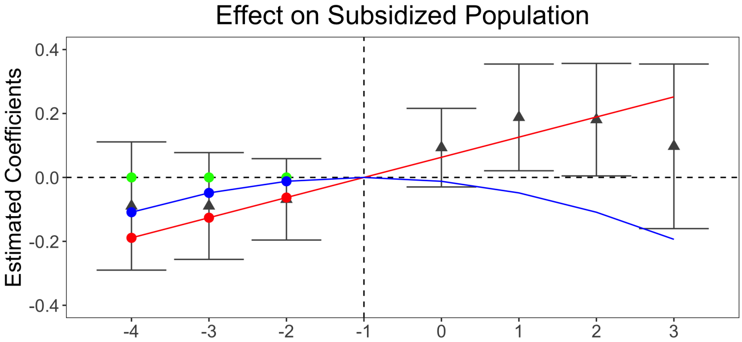
- P-value for $H_0 : \beta_{pre} =$ green dots (no pre-trend): 0.81
- P-value for $H_0 : \beta_{pre} =$ red dots: 0.81
- P-value for $H_0 : \beta_{pre} =$ blue dots: 0.81

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 (2019): 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

- Suppose parallel trends is violated.
- Because pre-test doesn't have perfect power, sometimes we won't find a significant pre-trend.
- But the draws of data where this happens are a **selected sample**.
- This introduces additional statistical issues, and can make things worse!
- I'll illustrate this with a sample example; See Roth (2021) for more details.

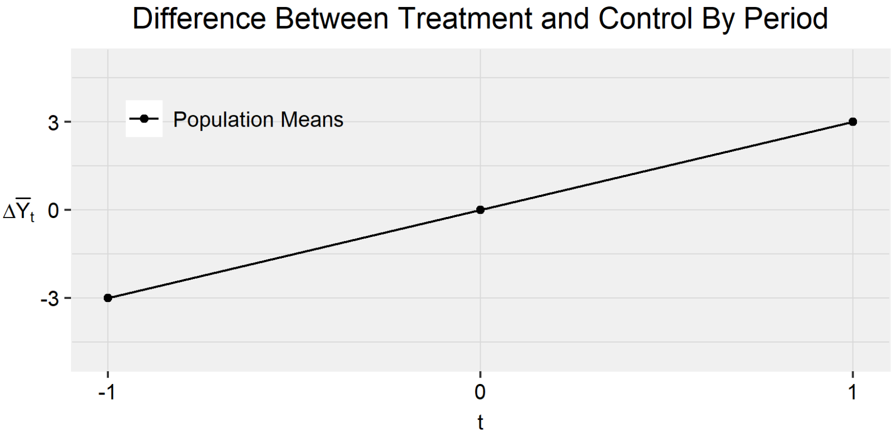
Stylized Three-Period DiD Example

- Three periods $t \in \{-1, 0, 1\}$, with the treatment group receiving treatment between $t = 0$ and $t = 1$.

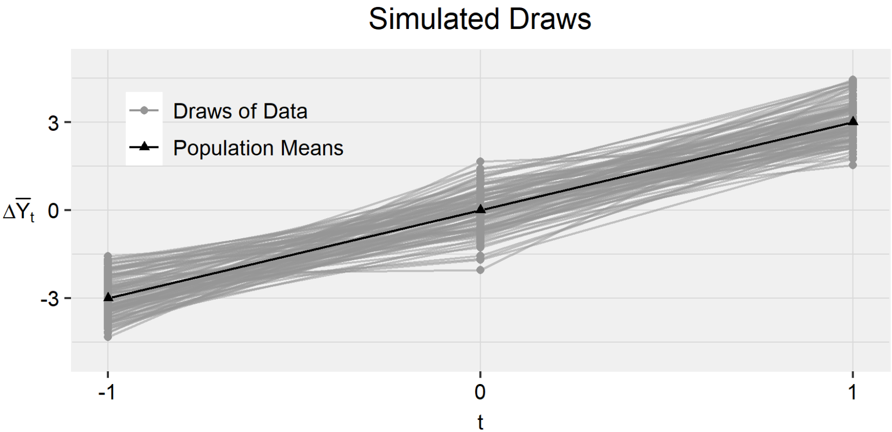
- No causal effect of treatment: $\tau = 0$

- In population, treatment group is on a linear trend relative to the control group with slope δ
  - Control group mean: $\mu_t^C = E[Y_{it} | Control] = 0$
  - Treatment group mean: $\mu_t^T = E[Y_{it} | Control, Treated] = \delta t$

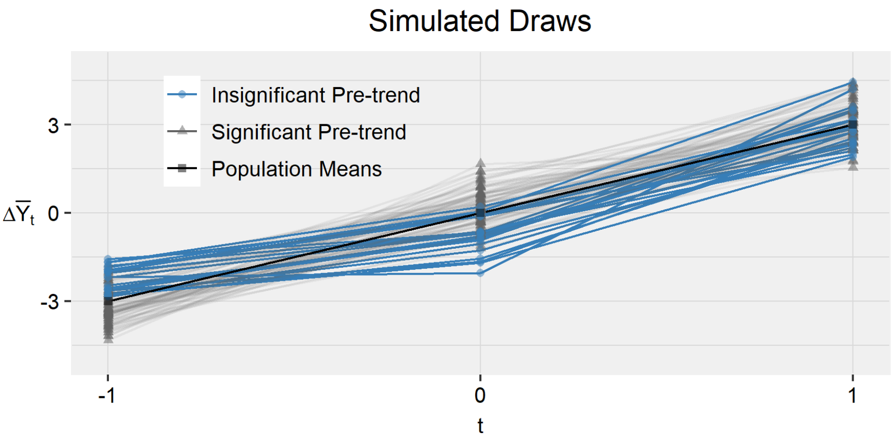
- Realized outcomes:
  - $\bar{Y}_t^C = \mu_t^C + \epsilon_t^C$
  - $\bar{Y}_t^T = \mu_t^T + \epsilon_t^T$
  - Independent normal errors: $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$



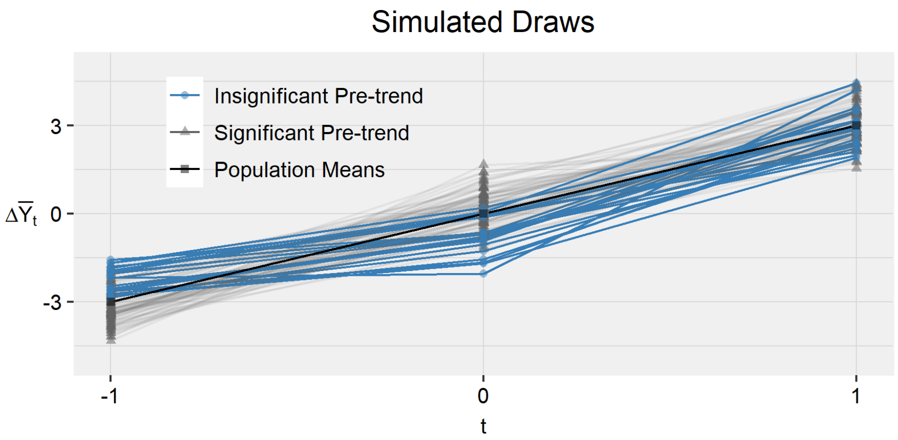
- Example: In population, there is a linear difference in trend with slope 3



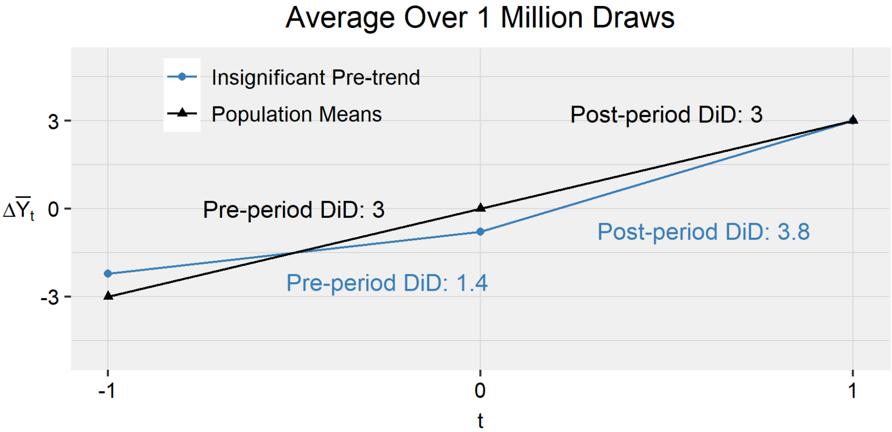
- Example: In population, there is a linear difference in trend with slope 3
- In actual draws of data, there will be noise around this line



- 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

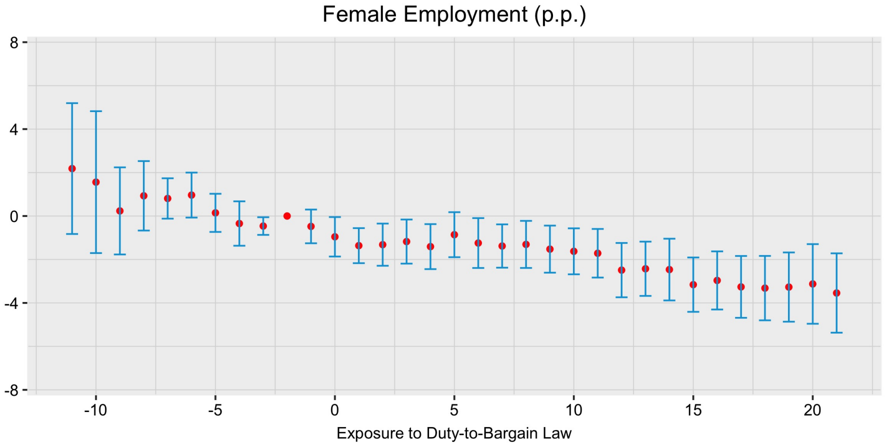


- In some of the draws of the data, highlighted in blue, the difference between period -1 and 0 will be insignificant
- In the insignificant draws, we tend to underestimate the difference between treatment and control at $t = 0$



- 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

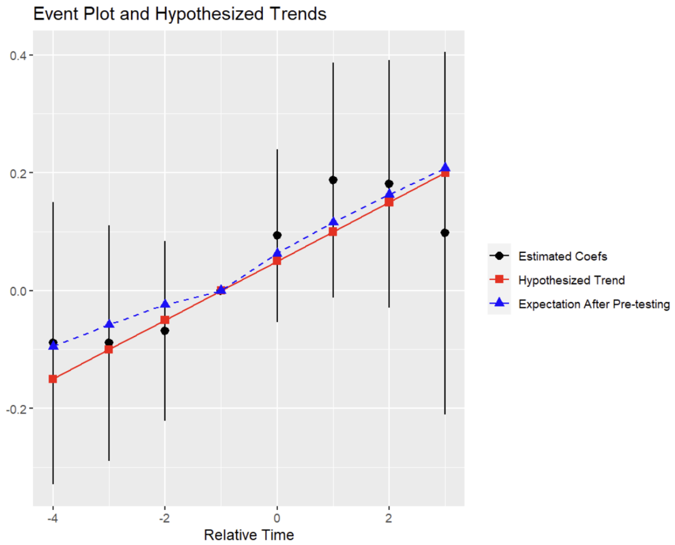
Issue 3 - Fail Pre-test, What Next?



"[There is] clear evidence of differential pretreatment trends... [O]ur empirical approach does not appear valid for women; we cannot draw strong conclusions" (Lovenheim and Willen, 2019)

Solution 1 - 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.

Solution 2 - An Honest Approach to Parallel Trends

- Rather than pre-test, place restrictions on way that parallel trends can be violated
- Restrictions can formalize the logic motivating pre-trends testing!
  - Logic of pre-trends testing: Pre-period difference in trends is **informative** about counterfactual post-period difference in trends
  - Formalize this by placing **restrictions** on the relationship between pre-period and counterfactual post-period differences in trends

Solution 2 - An Honest Approach to Parallel Trends

- Rather than pre-test, place restrictions on way that parallel trends can be violated
- Restrictions can formalize the logic motivating pre-trends testing!
  - Logic of pre-trends testing: Pre-period difference in trends is **informative** about counterfactual post-period difference in trends
  - Formalize this by placing **restrictions** on the relationship between pre-period and counterfactual post-period differences in trends
- Derive confidence sets that are **uniformly valid** so long as the difference in trends satisfies these restrictions
- Recommend **sensitivity analysis:** How informative must the pre-period differential trend be about the counter-factual post-period differential trend to obtain informative inference?

Motivating Model: 3-period Diff-in-Diff

- Three periods: $t = -1, 0, 1$.
- There is a treated population ($D = 1$) who receives treatment between period 0 and 1; the control population ($D = 0$) does not receive treatment
- We observe an outcome Y_{it} for a panel of N_1 treated and N_0 control units.
- Potential outcomes: $Y_{it}(1)$, $Y_{it}(0)$. Observe $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$. Assume treatment has no effect before implementation: $Y_{it}(1) = Y_{it}(0)$ for $t < 1$.
- Target parameter is average treatment-on-treated (ATT) in period $t = 1$,

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

Motivating Model (continued)

- Suppose we estimate the event-study specification:

$$Y_{it} = \lambda_i + \phi_t + \sum_{s \neq 0} \beta_s \times 1[t = s] \times D_i + \epsilon_{it}.$$

- In this simple setting, $\hat{\beta}_1$ and $\hat{\beta}_{-1}$ are difference-in-differences of sample means:

$$\hat{\beta}_1 = \underbrace{(\bar{Y}_{D=1,t=1} - \bar{Y}_{D=1,t=0})}_{\text{Dif for treated group } (t=0 \text{ to } 1)} - \underbrace{(\bar{Y}_{D=0,t=1} - \bar{Y}_{D=0,t=0})}_{\text{Dif for control group } (t=0 \text{ to } 1)}$$

$$\hat{\beta}_{-1} = \underbrace{(\bar{Y}_{D=1,t=-1} - \bar{Y}_{D=1,t=0})}_{\text{Dif for treated group } (t=0 \text{ to -}1)} - \underbrace{(\bar{Y}_{D=0,t=-1} - \bar{Y}_{D=0,t=0})}_{\text{Dif for control group } (t=0 \text{ to -}1)}$$

Motivating Model (continued)

- Taking expectations and re-arranging, we obtain

$$\mathbb{E}\left[\hat{\beta}_1\right] = \tau_{ATT} + \underbrace{\mathbb{E}\left[Y_{i,t=1}(0) - Y_{i,t=0}(0) \mid D_i = 1\right] - \mathbb{E}\left[Y_{i,t=1}(0) - Y_{i,t=0}(0) \mid D_i = 0\right]}_{\text{Post-period differential trend } := \delta_1},$$

- Thus, $\hat{\beta}_1$ is biased for τ_{ATT} if there is a post-period differential trend ($\delta_1 \neq 0$)
- We don't observe δ_1 directly, but identify its pre-period analog

$$\mathbb{E}\left[\hat{\beta}_{-1}\right] = \underbrace{\mathbb{E}\left[Y_{i,t=-1}(0) - Y_{i,t=0}(0) \mid D_i = 1\right] - \mathbb{E}\left[Y_{i,t=-1}(0) - Y_{i,t=0}(0) \mid D_i = 0\right]}_{\text{Pre-period differential trend } := \delta_{-1}}.$$

- Typical parallel trends assumption: $\delta_{-1} = \delta_1 = 0$.

- **Goal**: Conduct inference on τ_{ATT} while relaxing the assumption that δ_1 is exactly 0.

- **Goal**: Conduct inference on τ_{ATT} while relaxing the assumption that δ_1 is exactly 0.
- If δ_1 left unrestricted, then τ_{ATT} is not identified.
  But the intuition behind pre-trends testing is that δ_{-1} is informative about δ_1.

- **Goal**: Conduct inference on τ_{ATT} while relaxing the assumption that δ_1 is exactly 0.

- If δ_1 left unrestricted, then τ_{ATT} is not identified.
  But the intuition behind pre-trends testing is that δ_{-1} is informative about δ_1.

- **Our approach:** Formalize intuition of pre-trends testing via restrictions on δ_1 given δ_{-1}: require that $\delta = (\delta_{-1}, \delta_1)' \in \Delta$ for some class of possible underlying trends Δ.

- **Goal**: Conduct inference on τ_{ATT} while relaxing the assumption that δ_1 is exactly 0.

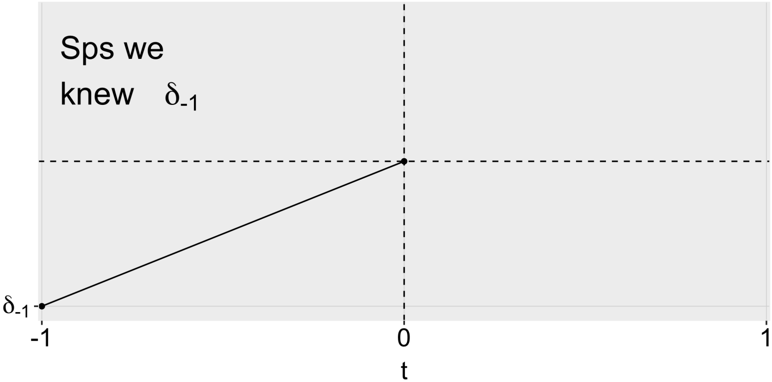
- If δ_1 left unrestricted, then τ_{ATT} is not identified.
But the intuition behind pre-trends testing is that δ_{-1} is informative about δ_1.

- **Our approach:** Formalize intuition of pre-trends testing via restrictions on δ_1 given δ_{-1}: require that $\delta = (\delta_{-1}, \delta_1)' \in \Delta$ for some class of possible underlying trends Δ.
  - Under such restrictions, τ_{ATT} is typically set-identified

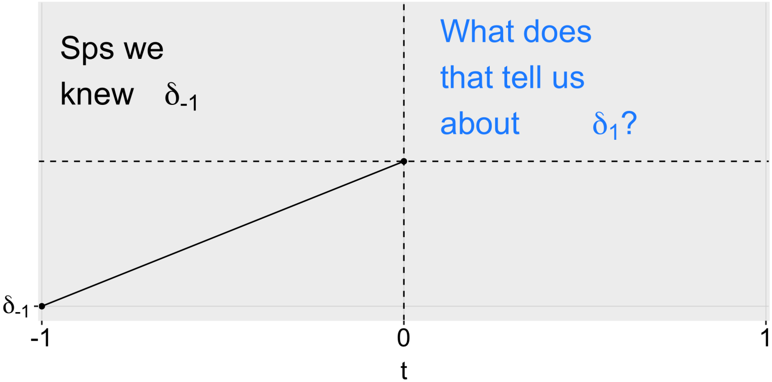
- Derive confidence sets that are uniformly valid for τ_{ATT} so long as $\delta \in \Delta$
  - Such confidence sets are "honest" with respect to Δ (Li, 1989) Details

- **Goal**: Conduct inference on τ_{ATT} while relaxing the assumption that δ_1 is exactly 0.
- If δ_1 left unrestricted, then τ_{ATT} is not identified.
  But the intuition behind pre-trends testing is that δ_{-1} is informative about δ_1.
- **Our approach:** Formalize intuition of pre-trends testing via restrictions on δ_1 given δ_{-1}: require that $\delta = (\delta_{-1}, \delta_1)' \in \Delta$ for some class of possible underlying trends Δ.
  - Under such restrictions, τ_{ATT} is typically set-identified
- Derive confidence sets that are uniformly valid for τ_{ATT} so long as $\delta \in \Delta$
  - Such confidence sets are "honest" with respect to Δ (Li, 1989) Details
- Recommend **sensitivity analysis** with respect to Δ: how do our conclusions change under different assumptions about what violations of parallel trends we're willing to allow for ex ante?

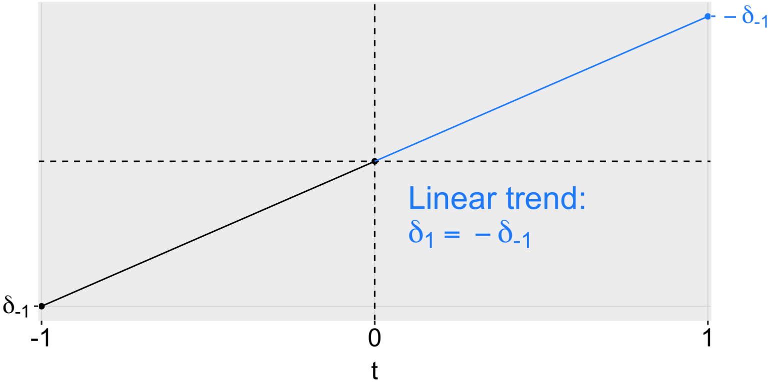
Intuition for Δ



Intuition for Δ

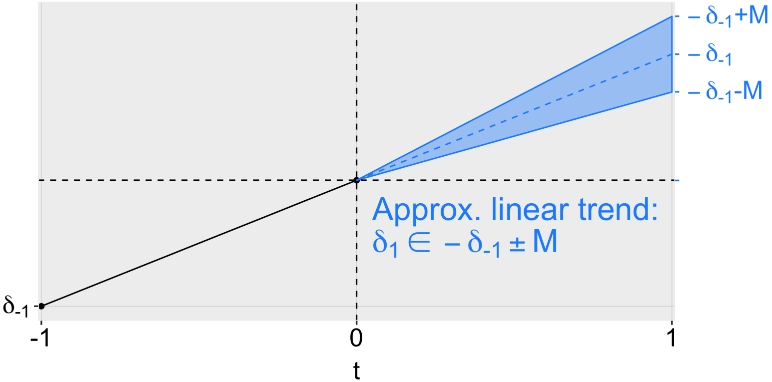


Intuition for Δ



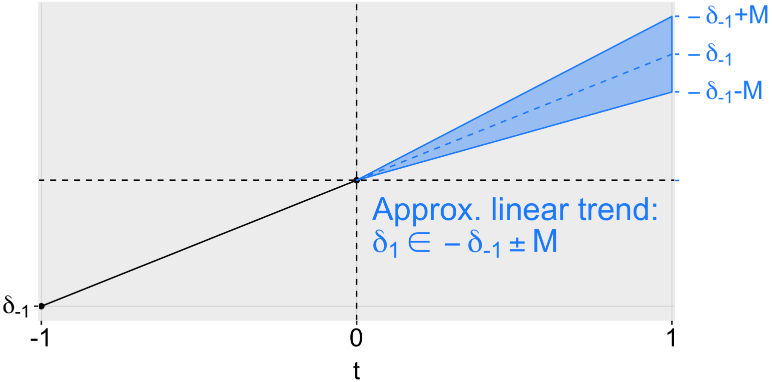
- Assuming linear trends corresponds with $\Delta = \{(\delta_{-1}, \delta_1)' : \delta_1 = -\delta_{-1}\}$.

Intuition for Δ



- Bound approximation error by M: $\Delta^{SD}(M) = \{(\delta_{-1}, \delta_1)' \,:\, \delta_1 \in -\delta_{-1} \pm M\}$.

Intuition for Δ



- Can extend to multiple periods by restricting change in slope (second differences) by M,

$$\Delta^{SD}(M) := \{\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M\}$$

Other Possibilities for Δ

- Economic knowledge often implies sign or shape restrictions
  - Secular trends $\rightarrow$ monotonicity
  - Known simultaneous policy changes $\rightarrow$ sign restrictions
- Can allow extrapolation error to depend on magnitude of pre-trend
  - Incorporates intuition that pre-trends closer to 0 more informative
- Framework flexibly accommodates many other restrictions
  - Allow for any constraints expressed by linear inequalities, $\Delta = \{\delta : A\delta \leq d\}$
  - Convexity/concavity, bounds on magnitudes, Ashenfelter's dip, etc.

General Model

- Consider finite-sample normal model

$$\hat{\beta}_n \sim \mathcal{N}(\beta, \Sigma_n)$$

where $\hat{\beta}_n \in \mathbb{R}^{\bar{T}+\underline{T}}$ has $\underline{T}$ pre-treatment coefficients and $\bar{T}$ post-treatment coefficients,

$$\beta = \underbrace{\begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix}}_{\text{Treatment effect}} + \underbrace{\begin{pmatrix} \delta_{pre} \\ \delta_{post} \end{pmatrix}}_{\text{Bias from trend}}, \qquad \Sigma_n = \frac{1}{n}\Sigma^*.$$

- Viewed as asymptotic approximation: $\sqrt{n}(\hat{\beta}_n - \beta) \to_d \mathcal{N}(0, \Sigma^*)$ for wide range of DGPs.
  - Compatible w/approaches that correct weighting issues with staggered timing / heterogeneity
- In paper, show our finite-sample results translate to uniform asymptotic statements.

Target Parameter and Identification

- **Target parameter**: $\theta := l'\tau_{post}$.
  - E.g. effect for a single post-period, average across all post-periods
- Researcher specifies a set of possible differential trends Δ, motivated by economic intuition or context-specific knowledge.
- Under the restriction that $\delta \in \Delta$, one can obtain bounds on θ given the observed distribution ("set-identification").
- The identified set is the set of values θ consistent with β and $\delta \in \Delta$,

$$\mathcal{S}(\Delta, \beta) := \left\{ \theta \,:\, \exists \delta \in \Delta, \tau_{post} \in \mathbb{R}^{\bar{T}} \text{ s.t. } l'\tau_{post} = \theta, \beta = \delta + \begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix} \right\}, \tag{1}$$

.

Coverage

- We propose confidence sets $\mathcal{C}_n$ with correct coverage,

$$\inf_{\delta \in \Delta, \tau} \inf_{\theta \in \mathcal{S}(\Delta, \delta+\tau)} \mathbb{P}_{(\delta,\tau,\Sigma_n)} \left(\theta \in \mathcal{C}_n\right) \geq 1 - \alpha,$$

  in the sense that each point in the identified set is covered with probability $1 - \alpha$.

- Intuitively, this means that if the actual difference in trends is in the imposed Δ, we're guaranteed to include the true parameter at least $1 - \alpha$ fraction of the time!

Fixed length confidence intervals

- Consider CIs of the form $(a + v'\hat{\beta}_n) \pm \chi_n$ — linear combinations of event-study coeffs.
- For given (a, v), calculate maximum possible bias over set of possible δ's,

$$\bar{b}(a, v) = \max_{\delta \in \Delta, \tau} \left| \mathbb{E}_{(\delta, \tau, \Sigma_n)} \left[a + v'\hat{\beta}_n - l'\tau \right] \right|.$$

- Use a CI length that ensures coverage under this maximal bias.
  - This is the $1 - \alpha$ quantile of $\left| \mathcal{N}\left(\bar{b}(a, v),\ v'\Sigma_n v \right) \right|$.
- Find the estimator (a^*, v^*) that minimizes the length, $2\chi_n(a^*, v^*)$.
  - Optimal v^* trades off bias and variance, since χ_n increasing in $\bar{b}$ and $v'\Sigma v$

Fixed length confidence intervals

- Consider CIs of the form $(a + v'\hat{\beta}_n) \pm \chi_n$ — linear combinations of event-study coeffs.
- For given (a, v), calculate maximum possible bias over set of possible δ's,

$$\bar{b}(a, v) = \max_{\delta \in \Delta, \tau} \left| \mathbb{E}_{(\delta, \tau, \Sigma_n)} \left[a + v'\hat{\beta}_n - l'\tau \right] \right| .$$

- Use a CI length that ensures coverage under this maximal bias.
  - This is the $1 - \alpha$ quantile of $\left| \mathcal{N}\left(\bar{b}(a, v),\ v'\Sigma_n v \right) \right|$.
- Find the estimator (a^*, v^*) that minimizes the length, $2\chi_n(a^*, v^*)$.
  - Optimal v^* trades off bias and variance, since χ_n increasing in $\bar{b}$ and $v'\Sigma v$
- FLCIs have near-optimal expected length for certain types of restrictions (Armstrong and Kolesar, 2018, 2020)
  - Results apply for $\Delta = \Delta^{SD}$ when δ is linear, but not for other leading examples
  - Expected length of 95% CI within 28% of optimum when conditions hold Details

A Recipe for Practice

1. **Estimate an "event-study"** specification with a causal interpretation under a parallel trends assumption and asymptotic normal distribution

2. **Conduct sensitivity analysis** — report confidence sets under different assumptions about the set of possible differences in trends Δ

   Example: Recall $\Delta^{SD}(M)$ bounds change in slope of trend by M

   - How do confidence sets grow as we increase M?
   - At what value of M can we no longer reject null hypotheses of interest?

3. **Provide economic benchmarks** for evaluating the different choices of Δ

   - Magnitudes of potential confounds
   - Calibration to placebo groups or periods

- Provide an R package `HonestDiD` for easy and fast implementation

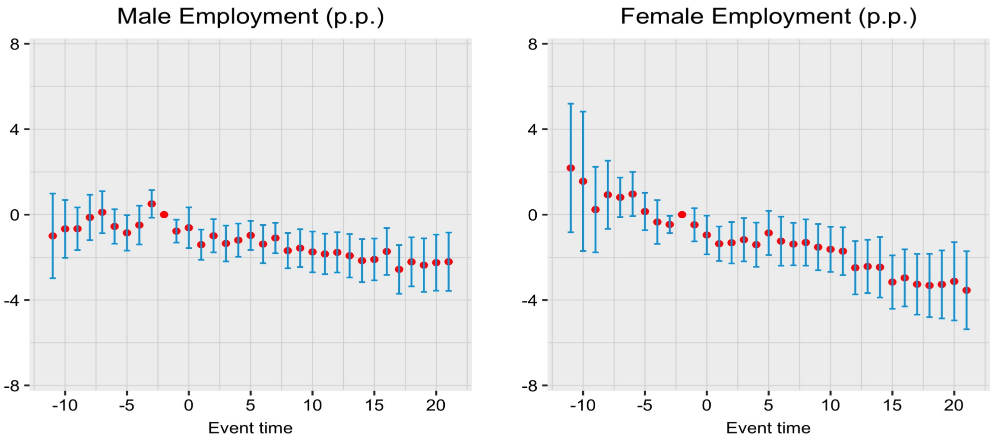
Lovenheim and Willen (2019)

- LW study how exposure to laws that strengthen teachers unions as a child affects labor market outcomes in adulthood
- Duty-to-bargain (DTB) laws passed mainly in 1960s-1980s; LW examine outcomes for adults in ACS from 2005-2012 who were students around the time these laws passed
- Use an "event-study specification" comparing students across states and birth cohorts:

$$Y_{sct} = \sum_{r=-11}^{21} D_{scr}\beta_r + X'_{sct}\gamma + \lambda_{ct} + \phi_s + \epsilon_{sct}.$$

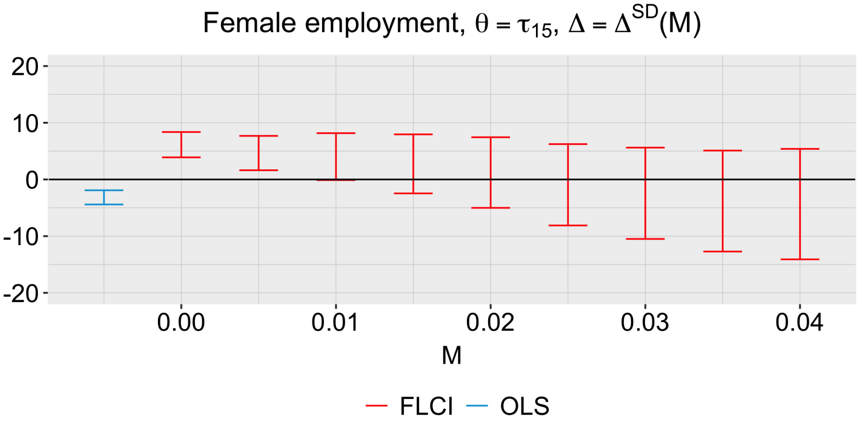
  - Y_{sct} = avg. outcome for the cohort born in state s in cohort c in ACS calendar year t.
  - D_{scr} = indicator for whether state s passed a DTB law r years before cohort c turned age 18.
  - $X_{sct}, \lambda_{ct}, \phi_s$ = time-varying controls, cohort-survey year FEs, state-of-birth FEs

Event-study coefficients for employment



Sensitivity plot - male employment

Sensitivity plot - female employment

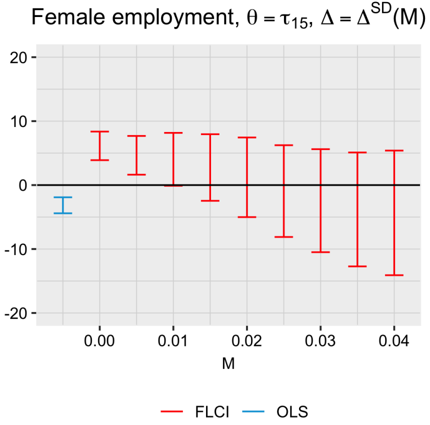


Calibrating M Using Knowledge of Possible Confounds

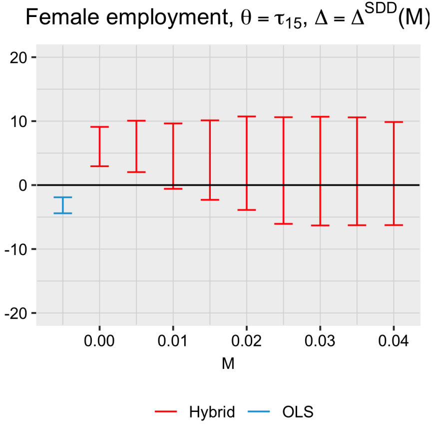
- A concern: differential trends in education quality, even absent DTB laws, may lead to non-parallel trends
- Thought exercise: if this were the mechanism, what does M imply about the possible differential evolution of teacher quality?
- Calibrate using Chetty, Friedman and Rockoff (2014) estimates:
  1SD change in teacher value-added $\rightarrow$ 0.4 pp increase in employment.
- Breakdown value of $M = 0.01$ therefore corresponds with changes in slope comparable to 0.025 SDs of TVA.

Incorporating shape restrictions

- LW argue: Observed pre-trends for women likely arise from "secular changes in women's educational and labor market outcomes"
- Formal version: Additionally impose **monotonicity** if think such trends would have continued absent treatment.
  - Imposing this additional shape restriction enables us to obtain tighter lower bounds



(a) W/o monotonicity



(b) W/monotonicity

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.

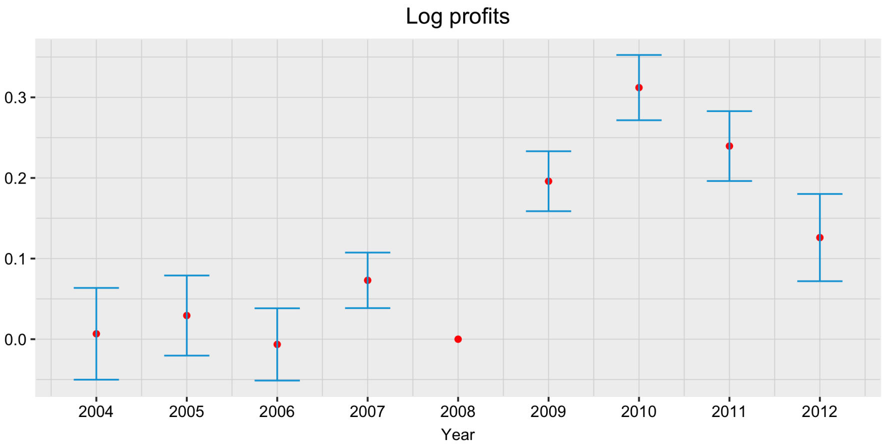
- BC analyze the impact of this change using a difference-in-differences design comparing restaurants to a control group of other market services firms

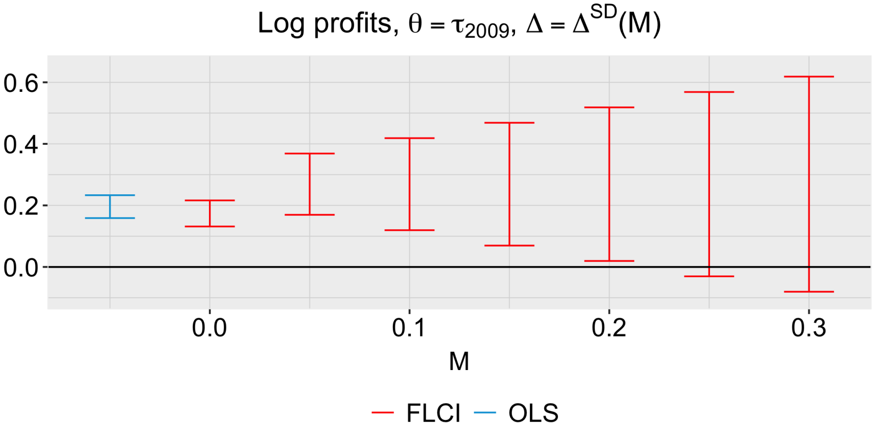
$$Y_{irt} = \sum_{s=2004}^{2012} \beta_s \times 1[t = s] \times D_{ir} + \phi_i + \lambda_t + \epsilon_{irt}, \quad (2)$$

  - 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

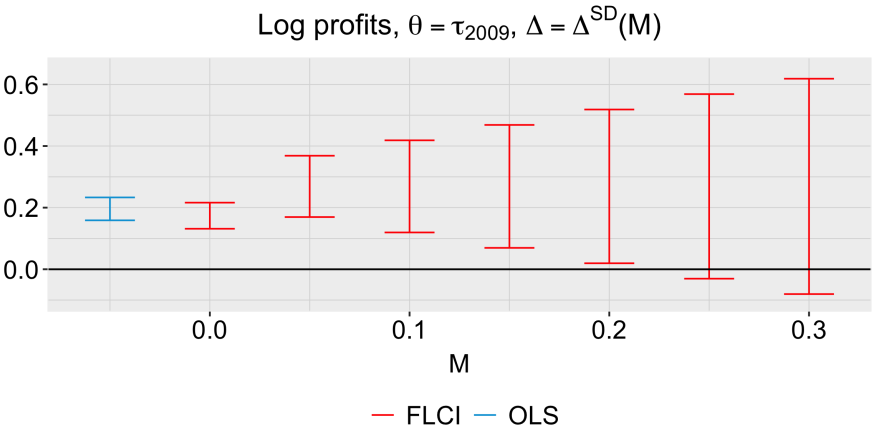
- Outcomes of interest include firm profits, prices, wage bill & employment. We focus on impact on profits in first year after reform.

Event-study coefficients for log profits





- "Breakdown" M for null effect is 0.22 (22 log points)



- "Breakdown" M for null effect is 0.22 (22 log points)
- 95% interval for largest pre-period change in slope: [0.09, 0.21].
- Rules out null effect unless allow for changes in slope that are larger than the upper bound on maximum change in slope in pre-periods

Incorporating Context-Specific Knowledge

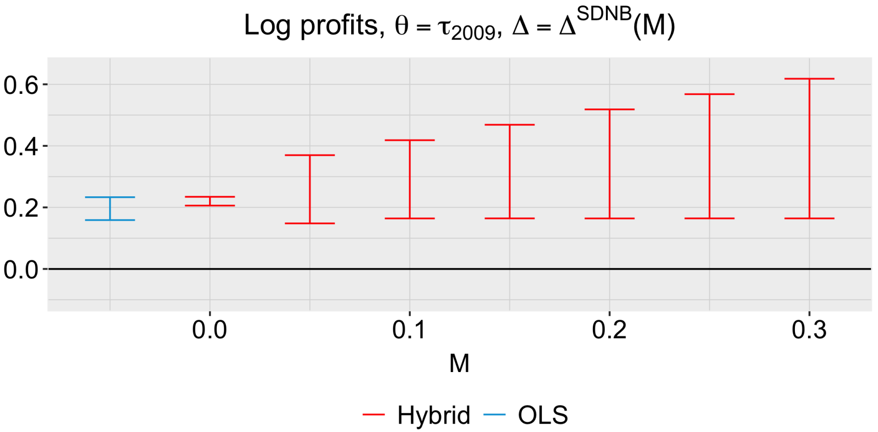
- BC argue that estimates of the treatment effect on profits likely understate true effect because of confounding policy changes
  - VAT cut occurred at the same time that a payroll subsidy for restaurants was terminated

- BC write: "A conservative interpretation of our results is that we are estimating a lower bound on the effect of the VAT cut on profits"

- Formal version: add **sign restrictions** to Δ, e.g.

$$\Delta^{SDNB} := \Delta^{SD} \cap \{\delta \, : \, \delta_{post} \leq 0\}$$

.



- With added sign restrictions, l.b. of confidence set never substantially below OLS l.b.

Wrapping Up

- Tests of pre-trends are intuitive but not a panacea!
- Roth (2021) and Rambachan and Roth (2020) provide tools for diagnostics and sensitivity analysis
- It's important to incorporate *context-specific* knowledge when using these tools.
- **Think about how parallel trends may be violated in your context!**
  This puts the "econ" back into "econometrics"

Additional Resources

- Roth (2021), "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends"
  - Paper; `staggered` package ; Shiny app

- Rambachan and Roth (2020), "An Honest Approach to Parallel Trends"'
  - Paper; `HonestDiD` package ; Vignette

Thank you!

Armstrong, Timothy and Michal Kolesar, "Optimal Inference in a Class of Regression Models," *Econometrica*, 2018, *86*, 655–683.

__ **and** __ , "Sensitivity Analysis using Approximate Moment Condition Models," *Quantitative Economics*, 2020. Forthcoming.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, September 2014, *104* (9), 2593–2632.

Li, Ker-Chau, "Honest Confidence Regions for Nonparametric Regression," *The Annals of Statistics*, 1989, *17* (3), 1001–1008.

Lovenheim, Michael F. and Alexander Willen, "The Long-Run Effects of Teacher Collective Bargaining," *American Economic Journal: Economic Policy*, 2019, *11* (3), 292–324.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *Working paper*, 2019.