Recent Advances in Difference-in-Differences

Pedro H. C. Sant'Anna
Microsoft & Vanderbilt University

One World YoungStats Webinar, December 15, 2021

Summary

- Clément gave a nice overview about the state-of-the-art DiD literature with variation in treatment timing (based on his paper with Xavier d'Haultfoeuille, "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey").

- Jonathan talked about DiD with potential violations of parallel trends (based on his AER:I paper, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends", and his paper with Ashesh Rambachan, "An Honest Approach to Parallel Trends").

Summary (Cont.)

- Brant talked about DiD with a continuous treatment (based on his paper with me and Andrew Goodman-Bacon, "Difference-in-Differences with a Continuous Treatment").

- Lihua talked about how to estimate weighted-averages of unit-and-time-specific treatment effects in setups with flexible treatment allocations when one is willing to explicitly model the treatment assignment mechanism (based on his paper with Dmitry Arkhangelsky, Guido Imbens, and Xiaoman Luo, "Double-Robust Two-Way-Fixed-Effects Regression For Panel Data").

What does unify all these papers?

Focus on providing more reliable tools for empirical researchers

Heterogeneous-robust DiD

- As Clément and Xavier's survey paper make it very clear, empirical researchers can now use several "heterogeneity-robust" DiD estimators.
- **How does one choose which to use?**

Heterogeneous-robust DiD

- As Clément and Xavier's survey paper make it very clear, empirical researchers can now use several "heterogeneity-robust" DiD estimators.
- **How does one choose which to use?**
- That is question and design dependent:
  - Is your treatment binary and staggered?
  - What is the type of parallel trends you are willing to use?
  - Are you comfortable of ruling out carryover effects?
  - What is your target causal parameter of interest?

Heterogeneous-robust DiD

- As Clément and Xavier's survey paper make it very clear, empirical researchers can now use several "heterogeneity-robust" DiD estimators.
- **How does one choose which to use?**
- That is question and design dependent:
  - Is your treatment binary and staggered?
  - What is the type of parallel trends you are willing to use?
  - Are you comfortable of ruling out carryover effects?
  - What is your target causal parameter of interest?
- **No estimator dominates all the others: they should be viewed as complements!**

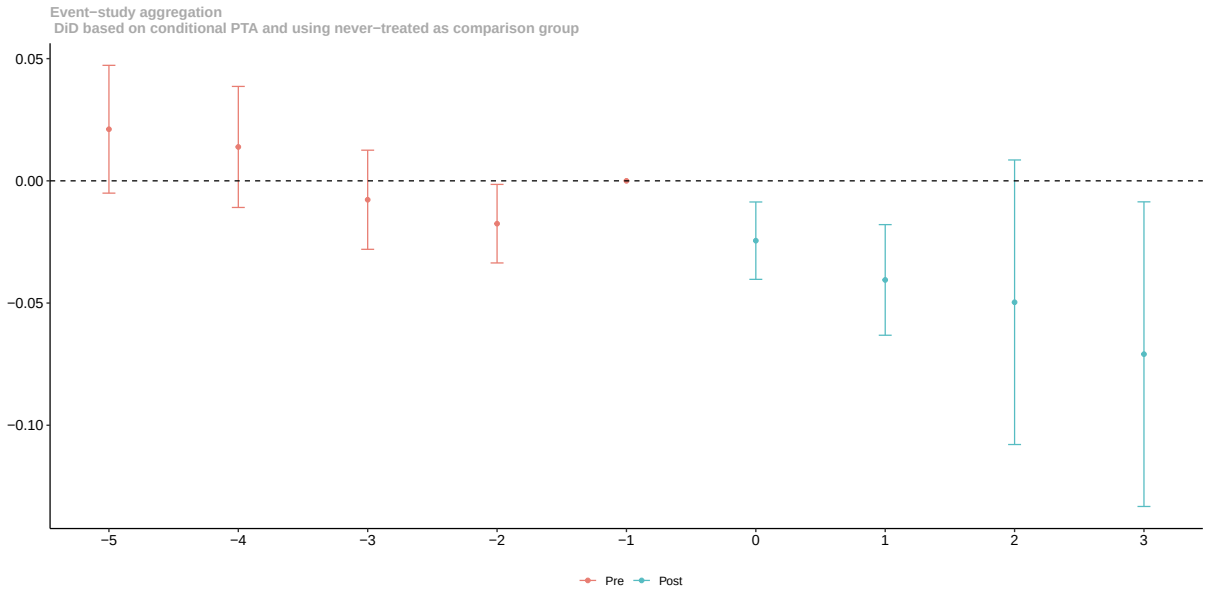
In my view, this message should be stressed more. Kudos to dCdH!

Pre-tests and parallel trends

- As Jon emphasized, currently common practice on pre-test has limitations with important practical consequences.
- However, as a good econometrician, instead of sitting in our Ivory Tower, Jon provides several practical, easy-to-use tools that can alleviate some of these problems.
- In my view, the sensitivity analysis procedures that he and Ashesh provide are fundamental to improve the reliability and transparency of DiD procedures.
- Although Jon didn't emphasize in his talk, the sensitivity analyse tools he (and Ashesh) propose can be **combined** with modern *Heterogeneous Robust* DiD procedures.
- Let's show this using the `did` and `HonestDiD` `R` packages, which implements Callaway and Sant'Anna (2021) and Rambachan and Roth (2021), respectively.

Combining Callaway and Sant'Anna (2021) and Rambachan and Roth (2021)

```
# Install the packages (I used the Github versions)
devtools::install_github("bcallaway11/did"); devtools::install_github("asheshrambachan/HonestDiD")
#Load the packages
library(did); library(HonestDiD); library(dplyr); library(here)
# Load data used by Callaway and Sant'Anna (2021)
min_wage <- readRDS((here("data",'min_wage_CS.rds')))
#-------------------------------------------------------------------------------------------
# Formula for covariates
xformla <- ~ region + (medinc + pop ) + I(pop^2) + I(medinc^2)  + white + hs  + pov
#-------------------------------------------------------------------------------------------
# Estimate ATT(g,t)'s using DR DiD with never-treated as comparison group
CS_never_cond <- did::att_gt(yname="lemp", tname="year", idname="countyreal", gname="first.treat",
                             xformla = xformla, control_group="nevertreated", data = min_wage,
                             panel = TRUE, base_period="universal", bstrap = TRUE, cband = TRUE)
# compute event-study aggregation
CS_es_never_cond <- aggte(CS_never_cond, type = "dynamic", min_e = -5, max_e = 5)
ggdid(CS_es_never_cond,
      title = "Event-study aggregation \n DiD based on conditional PTA and using never-treated as
      comparison group ")
```

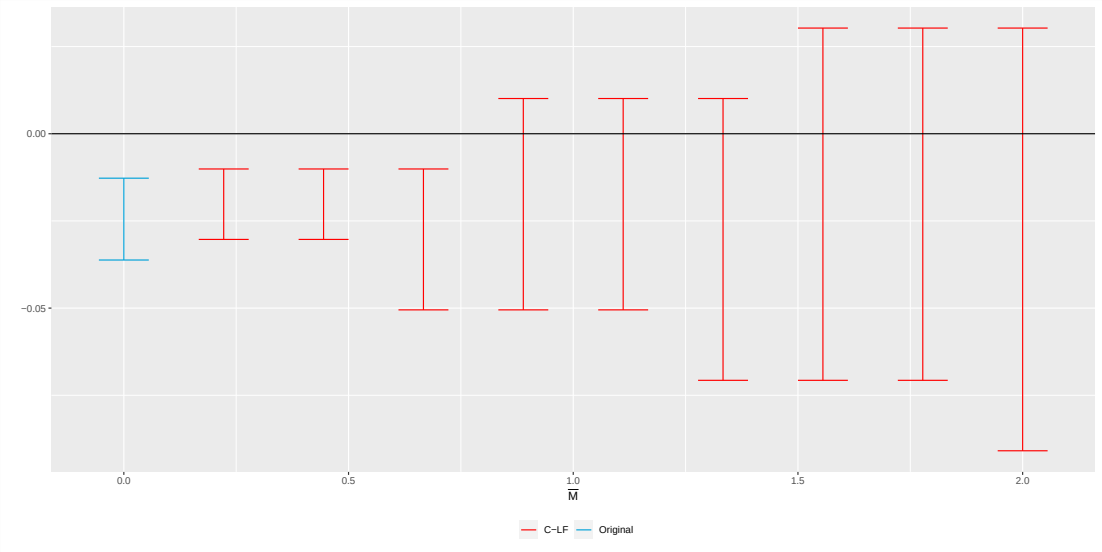

Event−study aggregation
DiD based on conditional PTA and using never−treated as comparison group
0.05
0.00
−0.05
−0.10
−5
−4
−3
−2
−1
0
1
2
3
Pre
Post

Rambachan and Roth (2021) after Callaway and Sant'Anna (2021)

```
# Brant has written a wrapper for HonestDiD that allows one to use aggte did outputs as inputs
# Here we apply the wrapper, and use the ``relative magnitude'' type of sensitivity analysis
# Doing it for instantaneous treatment effect, e = 0
hd_cs_rm_never <- honest_did(CS_es_never_cond,
                             e = 0,
                             type="relative_magnitude")

# Plot results
cs_HDiD_relmag <- createSensitivityPlot_relativeMagnitudes(hd_cs_rm_never$robust_ci,
                                        hd_cs_rm_never$orig_ci)
cs_HDiD_relmag
```

Sensitivity Analysis based on “relative magnitude” restrictions



Question:

What are reasonable values for M (smoothness-type) or $\overline{M}$ (relative magnitude-type) to use in the sensitivity analysis?

Another central theme in modern DiD:

Heterogeneity, and parameters of interest

Importance of being careful about parameter of interest

- With binary treatments and staggered adoption, the literature has somehow stressed the pitfalls of using variants of the TWFE regression

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

- Issue is that, under some assumptions,

$$\beta = \sum_{t,g} w_{t,g} \cdot ATT(g,t),$$

but the weights $w_{t,g}$ are not guaranteed to be convex, i.e., they can be negative; see, e.g., Athey and Imbens (2021), Borusyak, Jaravel and Spiess (2021), de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021).

Importance of being careful about parameter of interest

- With binary treatments and staggered adoption, the literature has somehow stressed the pitfalls of using variants of the TWFE regression

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

- Issue is that, under some assumptions,

$$\beta = \sum_{t,g} w_{t,g} \cdot ATT(g,t),$$

but the weights $w_{t,g}$ are not guaranteed to be convex, i.e., they can be negative; see, e.g., Athey and Imbens (2021), Borusyak et al. (2021), de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021).

- What if the weights were convex? Would this be "fine"?

Importance of being careful about parameter of interest

- With binary treatments and staggered adoption, the literature has somehow stressed the pitfalls of using variants of the TWFE regression

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

- Issue is that, under some assumptions,

$$\beta = \sum_{t,g} w_{t,g} \cdot ATT(g,t),$$

but the weights $w_{t,g}$ are not guaranteed to be convex, i.e., they can be negative; see, e.g., Athey and Imbens (2021), Borusyak et al. (2021), de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021).

- What if the weights were convex? Would this be "fine"?

- LATE and MTE IV literature have being debating this issue for the last 20 years: **What is the causal question of interest? That should help us picking "good"**

What if treatment is continuous?

- With continuous treatments, this becomes even more important, as discussed by Brant and in our paper, Callaway, Goodman-Bacon and Sant'Anna (2021)

- Even with two periods, with no units being treated in period $t = 1$, some units remaining untreated at period $t = 2$, and the others receiving different dosages d, the β from the TWFE regression

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

  can have **very different** causal interpretations!

What if treatment is continuous?

- Under a "strong parallel trends" assumption, we have:
  - If we were to use "slope effects" as "building blocks":

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l) ACR(l)\, dl + w_0 \frac{ATE(d_L)}{d_L},$$

    where $ACR(d) = \frac{\partial \mathbb{E}[Y_t(d)]}{\partial d}$, and all weights are non-negative and integrate to one.

  - If we were to use "level effects" as "building blocks":

$$\beta^{twfe} = \int_{\mathcal{D}_+} w_1^{\text{alt}}(l) \frac{ATE(l)}{l}\, dl$$

    where $ATE(d) = \mathbb{E}[Y_t(d) - Y_t(0)]$, and the weights integrate to one but are non-convex (i.e., can be negative).

What if treatment is continuous?

- Under a "strong parallel trends" assumption, we have:
  - If we were to use "slope effects" as "building blocks":

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l) ACR(l)\, dl + w_0 \frac{ATE(d_L)}{d_L},$$

    where $ACR(d) = \frac{\partial \mathbb{E}[Y_t(d)]}{\partial d}$, and all weights are non-negative and integrate to one.

  - If we were to use "level effects" as "building blocks":

$$\beta^{twfe} = \int_{\mathcal{D}_+} w_1^{\text{alt}}(l) \frac{ATE(l)}{l}\, dl$$

    where $ATE(d) = \mathbb{E}[Y_t(d) - Y_t(0)]$, and the weights integrate to one but are non-convex (i.e., can be negative).

- Same estimator and same assumptions, but sharply different interpretations!

In my view, whenever it is possible, we should be clear about the causal parameter of interest from the very beginning!

But what if we don't want to drop TWFE regressions?

What if don't want to dump TWFE regressions?

- Lihua presented a nice paper where the starting point, at least implicitly, it the fact that empirical researchers may not want to stop using TWFE regressions, but still want to allow for heterogeneous treatment effects.

- Furthermore, they want to allow for general binary treatment selection patterns where units can turn treatment on and off, see, also, de Chaisemartin and D'Haultfœuille (2020, 2021).

- Here, though, Arkhangelsky, Imbens, Lei and Luo (2021) don't want to rely on parallel trends assumptions!

- How that is possible, you may ask?

What if don't want to dump TWFE regressions?

- Arkhangelsky et al. (2021) propose to augmented the TWFE with unit-specific weights, where these weights are constructed base on the researcher' ability to model the assignment mechanism.

- The assignment mechanism can depend on covariates, past treatment assignments, time-varying unobservables, but not outcomes.

- They consider two scenarios/setups: one under which the assignment probability is know (design-based inference) and one which it is not (sampling-based inference).

What if don't want to dump TWFE regressions?

- They show that as long as the model for the assignment mechanism is correctly specified *or* the TWFE regression is correctly specified (homogeneous treatment effects), their proposed estimator is consistent for a double-average of the unit-and-time specific TE.
- What is unique about this paper is that it does not rely on a parallel trends condition but on a latent ignorability condition.
  - This allows the authors to identify causal parameters other than a ATT.
  - Plus, they can use different comparison groups (e.g.,always treated units) to construct their estimator.

What if don't want to dump TWFE regressions?

- They show that as long as the model for the assignment mechanism is correctly specified *or* the TWFE regression is correctly specified (homogeneous treatment effects), their proposed estimator is consistent for a double-average of the unit-and-time specific TE.
- What is unique about this paper is that it does not rely on a parallel trends condition but on a latent ignorability condition.
  - This allows the authors to identify causal parameters other than a ATT.
  - Plus, they can use different comparison groups (e.g.,always treated units) to construct their estimator.
- I wonder if this last point should be stressed more. You can consider several designs other than DiD!
- I would love to see a deeper discussion around the empirical setups you foresee people actually using your proposed tools.

Conclusion

Conclusion

- We are living in an exciting time!
- Somehow we are now embracing heterogeneity and accepting that heterogeneous treatment effects are the norm and not the exception.
- This somehow stress that is better to avoid talking about "the" causal effect, but rather to talk about "a" causal effect.
- I personally love the idea of econometricians and empirical researchers working side-by-side, helping each other to improve research practice.
- Thanks!

References

Arkhangelsky, Dmitry, Guido W. Imbens, Lihua Lei, and Xiaoman Luo, "Double-Robust Two-Way-Fixed-Effects Regression For Panel Data," *arXiv:2107.13737 [econ.EM]]*, July 2021.

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

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," Technical Report, Working Paper 2021.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, (2), 200–230.

_ , Andrew Goodman-Bacon, and Pedro H.C. Sant'Anna, "Difference-in-Differences with a Continuous Treatment," *arXiv:2107.02637*, 2021.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

_ **and** _ , "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *arXiv:2007.04267 [econ]*, May 2021. arXiv: 2007.04267.

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

Rambachan, Ashesh and Jonathan Roth, "An Honest Approach to Parallel Trends," *Working Paper*, 2021.

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