Conducting Inference in Difference-in-Differences with Different Numbers of Clusters: A Flow Chart with Advice

By Patrick Button
(pbutton@tulane.edu)

Last Updated: April 29, 2020

START HERE: How many clusters (groups) do you have in your difference-in-differences?

2

This is a very difficult case that could lead to major problems. For examples of how a two group DiD is problematic, see:

Donald, Stephen G, and Kevin Lang. 2007. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics* 89 (2): 221–33. https://doi.org/10.1162/rest.89.2.221.

TL;DR: If you only have two time periods, the treatment effect is probably not identified. If you have more than two time periods, you can identify the treatment effect, but, by using their more realistic approach, you could have low statistical power. See, also, the box to the right for other issues.

The approach in Donald and Lang (2007) may also not perform well, see Conley and Taber (2010) below. However, it's not possible to construct Conley-Taber confidence intervals with just two groups. Synthetic control is also not possible. Bootstrapping is also not possible.

3 to 19

This is a tricky case. Two assymptotic assumptions are likely violated:
1) The number of groups approaches infinitity
2) The number of *treated* groups approaches infinity.

The violation of 1) suggests you should do a cluster bootstrap. However, the violation of 2), suggests you should use Conley-Taber confidence intervals, but you do not have a sufficient number of groups to follow this approach. Cluster bootstrapping is your best bet, but it imperfect. Use a wild cluster bootstrap if your clusters vary "wildly" in size, otherwise just do a regular cluster bootstrap.

20+

How many of these groups are ever treated?

< 11

Please see this paper, which details how, despite having lots of groups, an assymptotic assumption regarding the number of treated groups is likely violated:

Conley, Timothy G., and Christopher R. Taber. 2011. "Inference with Difference in Differences with a Small Number of Policy Changes." *Review of Economics and Statistics* 93 (1): 113–25. https://doi.org/10.1162/REST_a_00049.

TL;DR: Your inference will incorrect, with too much Type 1 error and standard errors/confidence intervals that are unrealistically small. Use Conley-Taber confidence intervals, as detailed in their paper. However, it's only really possible to construct these if you have enough control groups (at least 20)

20+

Do your clusters have "wildly" different sizes (e.g., some have 10 observations and some have 100)?

Yes

You should probably use a wild cluster bootstrap, which performs better when clusters have different sizes.

For more on this, see this helpful article for practitioners:

Mackinnon, James G. 2019. "How Cluster-Robust Inference Is Changing Applied Econometrics." *Canadian Journal of Economics* 52 (3): 851–81. https://doi.org/10.1111/caje.12388

No

How many clusters do you have again?

50+

It seems like you are probably fine doing regular clustering on group, but please read the box below.

< 50

11 to 19

In this case you are in a gray area. Conley and Taber (2010) may provide benefits over cluster bootstrapping, so I would usually recommend that unless you're closer to 19 and/or you have "wildly" different cluster sizes.

You may also want to consider a synthetic control case study. See:

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.http.

For an example of a paper that does this by pooling more treated groups together, see:

Kreif, Noemi, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton. 2016. "Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units." *Health Economics* 25: 1514–28. https://doi.org/10.1002/hec.3258.

You should probably use a cluster bootstrap. It is unlikely that the asymptotic assumption that the number of groups approaches infinity holds. See, e.g.,

Cameron, A.Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–72. https://doi.org/10.3368/jhr.50.2.317.

TL;DR: If you don't bootstrap, your inference will incorrect, with too much Type 1 error and standard errors/ confidence intervals that are unrealistically small.

"But I thought I could just do regular clustering if I have at least 42 clusters!?"

While the "rule of thumb", from the fantastic, but sometimes outdated, Mostly Harmless Economics text, is that with at least 42 groups, regular clustering on group is ok, this can be misleading (see Mackinnon, 2019, citation in the top right box).

The more supported "rule of thumb" seems to be 50, but the whole idea of a rule of thumb is highly inflexible.

But with around 40 to 49, and even with the low 50s, we definitely get into a gray area. There are likely benefits to bootstrapping but they may not be worth the cost if bootstrapping is time-intensive or adds complications. So, the best approach depends on circumstances.

Disclaimer:

This flow chart, made using LucidChart, is a work-in-progress and may contain errors, especially since econometric methods are consistently evolving. It is a summary that is geared towards pointing researchers towards papers they should read, and making them aware of econometric issues that are often overlooked.

If something seems incorrect or if you have comments, please email me at pbutton@tulane.edu. Thanks so much!

You are free to use this as long as you keep this note in here. If you'd like to modify this, I can provide the original file to you, and I would love to know how you're changing it as it may give me ideas for how to improve this.