Frontier Topics in Empirical Economics: Week 11
Standard Error Issues

Zibin Huang [1]

[1]College of Business, Shanghai University of Finance and Economics

December 2, 2024

Introduction: Nonstandard Standard Error Issues

- Inference is important in practice: Data $\Longrightarrow$ Target distribution
- How accurate is our estimate? How confident are we on our results?
- In traditional inference, we have two assumptions:
  - Uncertainty comes from random-sampling, asymptotics when $n \to \infty$
  - i.i.d. sample, no correlations
- What if these two assumptions are violated?

Introduction: Nonstandard Standard Error Issues

- In this lecture, we consider two cases
- First, when n is naturally limited (e.g. number of provinces)
- Another type of uncertainty becomes important: Design-based uncertainty
- Second, when i.i.d. fails and errors are clustered
- We have to incorporate this structure in inference
- Angrist calls them "Nonstandard Standard Error Issues"

Design-based Uncertainty

- In usual case, when we talk about inference, what is that?
- We have a target parameter: *"estimand"* β (Target)
- We want to recover it using an *"estimator"* (Method) $\hat{\beta}$ with a sample from the population, which gives you a result called *"estimate"* $\hat{\beta} = 0.5$ (Result)
- This process is called *estimation*, or statistical inference (Process)

Design-based Uncertainty

- Usually, we consider sampling-based uncertainty
- Each time you draw a new sample, it gives you a new estimate from your estimation process
- When sample changes, your estimation result changes
- Uncertainty comes from sampling process
- Thus, you have a standard error for your estimation
- But is this the only uncertainty in empirical research?
- Today, we are going to introduce the second source of uncertainty

Design-based Uncertainty

- Design-based uncertainty, introduced by Abadie et al. (2020)
- It is the uncertainty coming from the treatment assignment
- Treatment X_i is no longer considered fixed
- In some cases, person 1 is treated; in other cases, person 1 is not treated
- The potential outcome you observed is different when treatment is randomly changed
- We show that this helps you to understand uncertainty of estimation when you have non-negligible sample size

Design-based Uncertainty

- To visually explain the difference between traditional sampling-based uncertainty and design-based uncertainty
- Let's take a look at two tables from Abadie et al. (2020)
- R_i is an indicator of whether this observation is included in the sample

Design-based Uncertainty

- Sampling-based uncertainty

TABLE I

SAMPLING-BASED UNCERTAINTY (✓ IS OBSERVED, ? IS MISSING)

	Actual Sample			Alternative Sample I			Alternative Sample II			…
Unit	Y_i	Z_i	R_i	Y_i	Z_i	R_i	Y_i	Z_i	R_i	…
1	✓	✓	1	?	?	0	?	?	0	…
2	?	?	0	?	?	0	?	?	0	…
3	?	?	0	✓	✓	1	✓	✓	1	…
4	?	?	0	✓	✓	1	?	?	0	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…
n	✓	✓	1	?	?	0	?	?	0	…

Design-based Uncertainty

- Design-based uncertainty

TABLE II

DESIGN-BASED UNCERTAINTY (✓ IS OBSERVED, ? IS MISSING)

	Actual Sample			Alternative Sample I			Alternative Sample II			...
Unit	$Y_i^*(1)$	$Y_i^*(0)$	X_i	$Y_i^*(1)$	$Y_i^*(0)$	X_i	$Y_i^*(1)$	$Y_i^*(0)$	X_i	...
1	✓	?	1	✓	?	1	?	✓	0	...
2	?	✓	0	?	✓	0	?	✓	0	...
3	?	✓	0	✓	?	1	✓	?	1	...
4	?	✓	0	?	✓	0	✓	?	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
n	✓	?	1	?	✓	0	?	✓	0	...

Design-based Uncertainty

- Sampling-based uncertainty
  - Treatment is fixed, sampling observation is random
  - For non-sampled individuals, we cannot observe anything
  - Source of uncertainty: in each sample, we have different observations
- Design-based uncertainty
  - Treatment is random, sampling observation is fixed (e.g. all provinces in China)
  - For each individual, we only observe potential outcome in the realized status (but not counterfactual status)
  - Source of uncertainty: in each sample, we have different treatment status for each individual

Design-based Uncertainty

- Next, the authors construct a simple model and make the following four points:
  - 1. Show how design-based uncertainty affects the variance of the regression estimator
  - 2. Show White estimator remains conservative when we consider design-based uncertainty
  - 3. We can derive a finite-population correction for White estimator
  - 4. Discuss two sources of uncertainty and external/internal validity

Design-based Uncertainty

- Assume that we have a **finite** population of size n
- We randomly sample N from n
- $R_i \in \{0, 1\}$ as an indicator of whether i is sampled or not
- There is a random binary treatment regressor X_i
- n_1, N_1 are treated, n_0, N_0 are not treated
- We have observed and potential outcome as:

$$Y_i = Y_i^*(X_i) = \begin{cases} Y_i^*(1) \text{ if } X_i = 1, \\ Y_i^*(0) \text{ if } X_i = 0 \end{cases}$$

- Potential outcomes are assumed to be non-stochastic

Design-based Uncertainty

- We use bold letters to represent vector of the whole sample $(\mathbf{Y}, \mathbf{Y}_i^*(1), \mathbf{Y}_i^*(0), \mathbf{R})$
- We define three estimands as our proposed targets
  - Descriptive estimand: free of **R** and potential outcome (population mean difference) $\theta^{descr} = \frac{1}{n_1}\sum_{i=1}^{n} X_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - X_i) Y_i$
  - Causal estimand: parameter depending on potential outcome $\mathbf{Y}_i^*(1), \mathbf{Y}_i^*(0)$
    $\theta^{causal,sample} = \frac{1}{N}\sum_{i=1}^{n} R_i(Y_i^*(1) - Y_i^*(0))$
    $\theta^{causal} = \frac{1}{n}\sum_{i=1}^{n}(Y_i^*(1) - Y_i^*(0))$
- $\theta^{causal,sample}$ is the average causal effect of the current sample
- θ^{causal} is the average causal effect of the whole population

Design-based Uncertainty

- When estimating θ^{descr}, we do not care about design-based uncertainty
  Nothing about treatment or potential outcome
- When estimating $\theta^{causal,sample}$, we do not care about sampling-based uncertainty
  Nothing about sampling process (given current sample)
- When estimating θ^{causal}, we do care about both types of uncertainty

Design-based Uncertainty

- To estimate these estimands, we use a simple OLS regression of Y_i on X_i to have:

$$\hat{\theta} = \frac{1}{N_1}\sum_{i=1}^{n} R_i X_i Y_i - \frac{1}{N_0}\sum_{i=1}^{n} R_i (1 - X_i) Y_i$$

- Sampling-based uncertainty comes from the randomness of $\mathbf{R}$
- Design-based uncertainty comes from the randomness of $\mathbf{X}$
- We further assume that both sampling and treatment assignment are random

Design-based Uncertainty

- It is shown that:

$$E[\hat{\theta}|\mathbf{X}, N_1, N_0] = \theta^{descr}$$
$$E[\hat{\theta}|\mathbf{R}, N_1, N_0] = \theta^{causal,sample}$$
$$E[\hat{\theta}|N_1, N_0] = \theta^{causal}$$

- Conditioning on treatment, θ is unbiased for descriptive estimand
- Conditioning on sampling, θ is unbiased for causal sample estimand
- Conditioning on none of them, θ is unbiased for causal estimand

Design-based Uncertainty

- We define the population variances as follows:

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i^*(x) - \frac{1}{n}\sum_{j=1}^{n} Y_j^*(x)\right)^2, \text{ for } x = 0, 1$$

$$S_\theta^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i^*(1) - Y_i^*(0) - \frac{1}{n}\sum_{j=1}^{n}(Y_j^*(1) - Y_j^*(0))\right)^2$$

- S_x^2 is the variance of potential outcomes for population
- S_θ^2 is the variance of treatment effect for population

Design-based Uncertainty

- Based on the defined population variance, we can derive three variances

$$V^{total}(N_1, N_0, n_1, n_0) = var(\hat{\theta}|N_1, N_0) = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{n_0 + n_1}$$

$$V^{sampling}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1}\left(1 - \frac{N_1}{n_1}\right) + \frac{S_0^2}{N_0}\left(1 - \frac{N_0}{n_0}\right)$$

$$V^{design}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}$$

- Now let's analyze them one by one

Design-based Uncertainty

- V^{total} is the total variance, considering both sampling-based and design-based uncertainty: $var(\hat{\theta}|N_1, N_0)$
- It is the variance we want to capture in inference for causal estimator
- $V^{sampling}$ is the variance from only sampling-based uncertainty, by conditioning on treatment assignment: $E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0]$
- It is the variance in inference for descriptive estimator
- V^{design} is the variance from only design-based uncertainty, by conditioning on current sample: $E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0]$
- It is the variance in inference for causal sample estimator

Design-based Uncertainty

- We have the following expressions of variances

$$V^{total}(N_1, N_0, n_1, n_0) = var(\hat{\theta}|N_1, N_0) = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{n_0 + n_1}$$

$$V^{sampling}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1}\left(1 - \frac{N_1}{n_1}\right) + \frac{S_0^2}{N_0}\left(1 - \frac{N_0}{n_0}\right)$$

$$V^{design}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}$$

  - 1. Generally, $V^{sampling}$ and V^{design} cannot be ranked, depending on the sampling rates $\frac{N}{n}$. A very large sampling rate means a very small $V^{sampling}$.

Design-based Uncertainty

- We have the following expressions of variances

$$V^{total}(N_1, N_0, n_1, n_0) = var(\hat{\theta}|N_1, N_0) = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{n_0 + n_1}$$

$$V^{sampling}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1}\left(1 - \frac{N_1}{n_1}\right) + \frac{S_0^2}{N_0}\left(1 - \frac{N_0}{n_0}\right)$$

$$V^{design}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}$$

  - 2. When $n \to \infty$, $V^{sampling} = V^{total}$
    If the population is infinite, then design-based uncertainty is ignorable and traditional inference for causal estimand (without considering design-based uncertainty) is fine

Design-based Uncertainty

- We have the following expressions of variances

$$V^{total}(N_1, N_0, n_1, n_0) = var(\hat{\theta}|N_1, N_0) = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{n_0 + n_1}$$

$$V^{sampling}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1}\left(1 - \frac{N_1}{n_1}\right) + \frac{S_0^2}{N_0}\left(1 - \frac{N_0}{n_0}\right)$$

$$V^{design}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}$$

  - 3. Consider estimating θ^{descr} or θ^{causal}:
    When population is finite, V^{total} and $V^{sampling}$ are overstated if we think it is infinite
    $V^{total}(N_1, N_0, \infty, \infty) - V^{total}(N_1, N_0, n_1, n_0) = \frac{S_\theta^2}{n_0+n_1} \geq 0$,
    $V^{sampling}(N_1, N_0, \infty, \infty) - V^{sampling}(N_1, N_0, n_1, n_0) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} \geq 0$

Design-based Uncertainty

- We have the following expressions of variances

$$V^{total}(N_1, N_0, n_1, n_0) = var(\hat{\theta}|N_1, N_0) = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{n_0 + n_1}$$

$$V^{sampling}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{X}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1}\left(1 - \frac{N_1}{n_1}\right) + \frac{S_0^2}{N_0}\left(1 - \frac{N_0}{n_0}\right)$$

$$V^{design}(N_1, N_0, n_1, n_0) = E[var(\hat{\theta}|\mathbf{R}, N_1, N_0)|N_1, N_0] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} - \frac{S_\theta^2}{N_0 + N_1}$$

  - 4. Consider estimating $\theta^{causal,sample}$:
    When population is finite, V^{design} is fine even if we think it is infinite
    $V^{design}(N_1, N_0, \infty, \infty) = V^{design}(N_1, N_0, n_1, n_0)$
    Relative sample size does not affect variance conditional on current sample

Design-based Uncertainty

- In practice, we usually use White estimator of the variance matrix
- It is calculated without considering design-based uncertainty[1]

$$\hat{V}^{w} = \frac{\hat{S}_1^2}{N_1} + \frac{\hat{S}_0^2}{N_0}, \text{ where } \hat{S}_1^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{n} R_i X_i \left(Y_i - \frac{1}{N_1}\sum_{i=1}^{n} R_i X_i Y_i \right)^2$$

- It is unbiased for V^{total} when n is infinite
- The small population bias is $E[\hat{V}^{w} | N] - V^{total} = S_\theta^2 / n$

[1] $\hat{S}_0^2$ is defined analogously

Design-based Uncertainty

- We can see that if we ignore design-based uncertainty in inference
- It is fine if we have a small sample compared with a massive population
- Like you have a CFPS dataset to represent all families in China
- But the positive bias will become large if we have a large sample size compared with a limited population
- Like you have a province-level regression
- In this case, traditional variance estimation can be too large and too conservative
- Because you ignore the fact that you already have a large part of the population

Design-based Uncertainty

- But fortunately, we can derive a bias-corrected estimator
- By taking into consideration
  - You have a large sample relative to a small population
  - You have uncertainty in treatment assignment
- The derivation of this estimator is technical
- Read Abadie et al. (2020) if you are interested

Clustered Standard Errors: Motivating Example

- Next, let's consider the clustering issue
- Many scholars claim that smaller classes are better
- What is the impact of class size on students' achievement?
- Hard to identify using observational data (selection problem)
- STAR is a RCT to answer this question

Clustered Standard Errors: Motivating Example

- It involves 11,600 children in TN
- Kids are randomly assigned to two kinds of classes
  (1) Small class with 13-17 children; (2) Regular class with 22-25 children
- Then we can identify the treatment effect of class size
- One assumption we always make is i.i.d.
- However, students in the same class are of course not independently sampled
- What will happen if we have correlations at class/school/district... level?

Clustered Standard Errors: Motivating Example

- The short answer is: we may underestimate the standard error
- Let's see why it is and how to fix this issue

Clustered Standard Errors: Setting

- Let's go on with the STAR experiment
- Consider the following regression for student i in class g:

$$y_{ig} = \beta_0 + \beta_1 x_g + e_{ig}$$

- y_{ig} test score; x_g class size (randomly assigned); e_{ig} error term
- This is a special case when x is fixed at g level (same treatment for the whole class)
- Test scores in the same class tend to be correlated (Same environment, teacher...)

Clustered Standard Errors: Setting

- Thus, we give up i.i.d. assumption and assume that for student i and j:

$$E[e_{ig}e_{jg}] = \rho_e \sigma_e^2 > 0$$

- ρ_e is the error intraclass correlation, σ_e^2 is the error variance
- Assume that we can decompose error into

$$e_{ig} = \nu_g + \eta_{ig}, \quad \nu_g \perp\!\!\!\perp \eta_{ig}$$

- We assume that ν_g captures all within class correlations ($\eta_{ig} \perp\!\!\!\perp \eta_{jg}$)
- Also assume homoskedasticity for both ν_g and η_{ig}
- Then we can prove that

$$\rho_e = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\eta^2} \tag{1}$$

- Intraclass correlation is the share of intraclass uncertainty in the total uncertainty

Clustered Standard Errors: Setting

- Equation (1) is called "intraclass correlation coefficient"
- Homework: Derive equation (1) from the previous setting

Clustered Standard Errors: Bias and Moulton Factor

- Let $V_c(\hat{\beta}_1)$ be the conventional OLS variance, $V(\hat{\beta}_1)$ be the correct variance
- Assume we have classes with equal size n, then

$$\frac{V(\hat{\beta}_1)}{V_c(\hat{\beta}_1)} = 1 + (n-1)\rho_e$$

- We call this Moulton factor
- $n, \rho_e \uparrow \Rightarrow$ Bias of conventional variance $\uparrow$
- Larger n means fewer groups $\Rightarrow$ less information
- Homework 2: What will happen if $\rho_e = 1$? (Answer in MHE)

Clustered Standard Errors: Bias and Moulton Factor

- Previous setting assumes fixed x_g within each group
- Let's see Moulton factor in a more general case when x_{ig} can vary across i in the same group

$$\frac{V(\hat{\beta}_1)}{V_c(\hat{\beta}_1)} = 1 + \left[\frac{V(n_g)}{\bar{n}} + \bar{n} - 1 \right] \rho_x \rho_e \tag{2}$$

- $\bar{n}$ is average group size; $V(n_g)$ is variance of group sizes; ρ_x is intraclass correlation of x_{ig}

Clustered Standard Errors: Bias and Moulton Factor

- In general, bias from within class correlation is larger when
  - (1) Average group size $\uparrow$
  - (2) Variance of group size $\uparrow$
  - (3) Intraclass correlation of treatment x_{ig} $\uparrow$
  - (4) Error intraclass correlation $\uparrow$
- The implication of (3)
  - Bias can be very large in the fixed group treatment x_g case
  - No need to cluster anything if the assignment is totally random for every individual
- The implication of (4): Naturally, no bias when $\rho_e = 0$

Clustered Standard Errors: Fix the Bias

- Now we know that std error estimation can be biased when we have correlation within classes
- What we should do? Several methods are available
  - (1) Use Moulton factor equation (2) to correct
    Not that good: error structure assumptions (homoskedasticity)
  - (2) Recommended: Liang and Zeger (1986) clustering estimator
    Generally consistent as number of groups $\rightarrow \infty$ (In stata, use option *cluster*)
  - (3) Running group-level regressions $\bar{y}_g = \beta_0 + \beta_1 x_g + \bar{e}_g$ using WLS (group size as weights)
    Better finite-sample properties, but x_g has to be group-fixed
  - Other methods: Block bootstrap, MLE...

Clustered Standard Errors: Choosing Cluster Levels

- How to choose the level of clustering?
- In STAR experiment, why not boy/girl, black/white/asian...?
- Clustering in more dimensions/higher level gives you larger std errs
- Is that OK to always cluster in more and more dimensions (be conservative)? NO. You can be too conservative $\Longrightarrow$ Overestimate std err
- Similarly, not always good to cluster at higher and higher level

Clustered Standard Errors: Choosing Cluster Levels

- This is because when you cluster in more and more dimensions
- Or at higher and higher level
- Your effective sample size compared with effective population becomes larger and larger
- As Abadie et al. (2020) has shown, it leads to overestimation of the std err
- For example, you have data of 10,000 firms in 20 provinces
- 10,000 can be a very small proportion of all firms in mainland China
- When you cluster at province level, effective sample rate becomes 20/31!

Clustered Standard Errors: Choosing Cluster Levels

- Thus, two issues remains
  - How to choose cluster level reasonably?
  - How to incorporate design-based uncertainty?
- Abadie et al. (2023) considers clustering as a sampling/design problem
- Cluster level depends on how you get your samples/assign your treatment
- It comes from the basic idea of Abadie et al. (2020)
- You have to consider both sampling-based and design-based uncertainty
- This is more to the core of the clustering problem

Clustered Standard Errors: Choosing Cluster Levels

- There are three misconceptions they want to clarity
- 1. The need for clustering hinges on the presence of a correlation between residuals
  - No. The essence is the clustering of sampling or treatment assignment
  - Even if students' scores are correlated within classroom, there is no need to cluster when sampling and treatment are totally random
- 2. No harm in using clustered std err when they are not required
  - Confidence intervals will be unnecessarily conservative
- 3. Researchers either fully adjust for clustering by using Liang and Zeger (1986) or not do that at all
  - Not really. They propose a new estimator CCV/TSCB to correct for large effective sample rate in clustering

Clustered Standard Errors: Choosing Cluster Levels

Here are some empirical suggestions from Abadie et al. (2023)

- 1. If sampling and treatment are both random
  - Do not cluster!
  - In this case, if sample represents a large fraction of the population, even White estimator is too conservative (Abadie et al., 2020)
- 2. If random sampling but clustered treatment assignment
  - Cluster at the treatment level.
  - In the fuzzy design case, using CCV/TSCB estimator

Clustered Standard Errors: Choosing Cluster Levels

- 3. If clustered sampling, random treatment assignment
  - Cluster at the sampling level, if you have small fraction of sampled clusters or small fraction of sampled units within each cluster
  - This is specifically important in panel data analysis
  - Do not cluster in other cases
- 4. If clustered sampling, clustered treatment assignment
  - Cluster at the higher level to be conservative

Clustered Standard Errors: Choosing Cluster Levels

- Let us go over two practical examples
- Case 1: (*Sampling cluster*) Some household/firm survey will
  - (1) Randomly select 50/300 cities in China
  - (2) Randomly select 100 households in each sampled city
- It gives you a natural stratified data set
- Just cluster at city level (in general, first sampling stage level)
- Case 2: (*Treatment cluster*) STAR assigns treatment at class level
- Then just cluster at class level

Clustered Standard Errors: DID and Serial Correlation

- One special case we must underscore is panel data analysis
- When using panel data, we usually employ time variation for identification
- You draw people, but not people in a specific year $\Rightarrow$ serial correlation
- You are drawing samples/assign treatment clustered at individual level
- Thus, DID gives a natural clustering structure of error
- One-level-up principle:
  Cluster at individual/province/city level, but NEVER individual-year/province-year/city-year level!!

Conclusion

- Today we discuss two nonstandard standard error issues
  - When sample is large compared with population
  - When errors are not i.i.d. but clustered
- In the first issue, we claim that we need to consider both sampling-based and design-based uncertainty
- Using traditional inference will have too large and conservative std err

Conclusion

- In the second case, we find that not adjusting for cluster will generate a too small std err
- We can use LZ estimator to fix it (consistent as #groups$\rightarrow \infty$)
- Clustering at higher level is not always good
- Clustering comes from either clustered sampling or clustered treatment
- Cluster at the first sampling stage, or treatment assignment level
- Do NOT cluster if you have a totally random sample and random treatment
- In DID, cluster one level up to take care of the serial correlation

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2020. "Sampling-based versus Design-based Uncertainty in Regression Analysis." *Econometrica* 88 (1):265–296.

———. 2023. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (1):1–35.

Liang, Kung-Yee and Scott L Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1):13–22.