Abadie, Athey, Imbens and Wooldridge (2023)

When (and how) should you cluster standard errors?

Contents

Three Misconceptions

1. **Need for…**
   - ‘If there exists correlation between two units of a cluster’
2. **Harm of…**
   - ‘Cluster adjusting doesn’t hurt’

     => ‘If clustering changes standard errors then use them’
3. **Extent of…**
   - ‘You can either:
     A. Cluster adjust standard errors, or
     B. Use heteroskedasticity robust standard errors

     those are your two choices!’

Conventional Frameworks

1. **Model-based**
   - **E.g. Moulton (1986, 1987)**
     - Posits: random effects model
     - Many samples: drawn from each states' own variable distributions
     - Clusters: at the state level, pre-specified in model
     - Issue: Stand taken on error structure, but chance individuals sampled randomly
2. **Sampling-based**
   - Great for surveys, where originated - see Kish (1995)
   - Mechanical process:
     1. Select clusters randomly from infinite population
     2. Randomly sample from clusters
   - Issue: samples often drawn from *majority* of clusters

Design Aspects

- Design aspects are neglected in conventional clustering
  - Needed for treatment effects inference!
- Justifying error component structure very tough
- Mechanisms of sampling and assignment
  - Makes design explicit!
  - Avoids need for error structure

Three contributions

1. A novel framework
   - Clustering from a design perspective
2. Derivations fitting the framework
   - Reframing of Least-squares and Fixed-effects estimators
3. Proposals for new tools
   - Two new approaches to clustering standard errors

A New Clustering Framework

A Sequence of Populations

- Indexed k where k^{th} pop. has n_k units indexed $i = 1, \dots, n_k$.
- Partitioned into m clusters.
  - Unit i of pop. k belongs to cluster $m_{k,i} \in \{1, \dots, m_{k,i}\}$
- Two potential outcomes for each i:
  - $y_{k,i}(1)$ and $y_{k,i}(0)$
- Pop. ATE denoted by τ_k
- Stochastic treatment indicator $W_{k,i} \in \{0,1\}$
- Realised outcome is $Y_{k,i} = y_{k,i}(W_{k,i})$
- Randomly observe triple: $(Y_{k,i}, W_{k,i}, m_{k,i})$
- Inclusion in sample given by random variable $R_{k,i}$

A New Clustering Framework

The Sampling Process

- Independent of potential outcomes and assignments
- Two stages:

1. Clusters sampled, with probability $q_k \in (0,1]$
2. Units sampled, with probability $p_k \in (0,1]$

- Random sampling: $q_k = 1$
- Clustered sampling: $q_k < 1$

A New Clustering Framework

The Assignment Process

- Two stages also:

1. Clusters assigned treatments with probability $A_{k,m} \in [0,1]$
   - $A_{k,m}$ uniform in k, mean μ_k and variance σ_k^2
2. Units assigned treatments independently, also by $A_{k,m}$

- Random assignment:
  - $\sigma_k^2 = 0 \Rightarrow W_{k,i}$ random across clusters
- Clustered assignment:
  - $\sigma_k^2 = \mu_k(1 - \mu_k) \Rightarrow W_{k,i}$ same within clusters
- Partially clustered assignment:
  - $0 < \sigma_k^2 < \mu_k(1 - \mu_k) \Rightarrow W_{k,i}$ varies within clusters

Causal Cluster Variance (CCV)

- Conventional cluster variance estimator:
  - Roughly a sum over clusters of squared within cluster sums of residuals
    - Upward bias when within cluster residuals have non-zero means
- The CCV:
  - Corrects conventional cluster variance bias
    - Using an estimate of expected sum of products between regression errors and regressor values
  - Limitations:
    - Requires sufficient sample of clusters
    - And within-cluster variation in treatment assignment

Two-Stage Cluster Bootstrap (TSCB)

Algorithm 1. Two-Stage Cluster Bootstrap

Input:

Sample $(Y_{k,i}, W_{k,i}, m_{k,i})$

Fraction sampled clusters q_k

Number of bootstrap replications B

Stage 1:

1a: Create pseudo population by replicating each cluster $\frac{1}{q_k}$ times

1b: For each cluster in the pseudo population, calculate the assignment probability $\overline{W}_{k,m}$

1c: Create a bootstrap sample of clusters by randomly drawing clusters from the pseudo population from Stage 1a, where cluster m is sampled with probability q_k

1d: For each sampled cluster, draw an assignment probability $A_{k,m}$ from the empirical distribution of the $\overline{W}_{k,m}$ from Stage 1b

Stage 2:

2a: Randomly draw from the set of treated units in cluster m, $\lfloor N_{k,m}A_{k,m} \rfloor$ units with replacement, where $\lfloor N_{k,m}A_{k,m} \rfloor$ means the largest integer smaller than or equal to $N_{k,m}A_{k,m}$

2b: Randomly draw from the set of control units in cluster m, $\lfloor N_{k,m}(1 - A_{k,m}) \rfloor$ units with replacement

Calculations:

For the units in the bootstrap sample constructed in Stage 2, collect the values for $(Y_{k,i}, W_{k,i}, m_{k,i})$ and calculate the least-squares or fixed-effect estimator

Calculate the standard deviation of the least-squares or fixed-effect estimator (defined in Section V) over the B bootstrap samples

Two-Stage Cluster Bootstrap (TSCB)

- Put simply, there are two stages:
  1. Fraction treated drawn from distribution in data
  2. Samples of treatment and control taken as determined in stage 1

Simulations

- Artificial population constructed from census data
- Cluster across US states
- Five designs of constructed samples
- Asymptotic standard errors as derived:
  - $\sqrt{v_k}$ for least-squares
  - $\sqrt{\tilde{v}_k}$ for fixed-effects
- For the four standard errors of interest:
  - 10,000 simulations per sample, 100 bootstraps in each

Simulations

TABLE II

AVERAGE STANDARD ERRORS ACROSS SIMULATIONS

		$\sqrt{N_k}$ s.d.	$\sqrt{v_k}$	$\sqrt{\tilde{v}_k}$	Normalized standard error			
					Robust	Cluster	CCV	TSCB
Baseline design:								
$p_k = 1, q_k = 1,$	OLS	5.91	5.90		1.90	44.86	6.32	5.80
$\sigma_{\tau_k} = 0.120, \sigma_k = 0.057$	FE	2.34		2.32	1.90	44.63	2.31	2.29
Second design:								
$p_k = 0.1, q_k = 1,$	OLS	2.61	2.59		1.90	14.28	3.78	2.60
$\sigma_{\tau_k} = 0.120, \sigma_k = 0.057$	FE	1.95		1.95	1.90	14.21	1.95	1.94
Third design:								
$p_k = 0.1, q_k = 1,$	OLS	14.50	14.17		1.98	56.46	13.70	14.33
$\sigma_{\tau_k} = 0.480, \sigma_k = 0.206$	FE	12.14		11.89	2.13	56.79	11.61	12.07
Fourth design:								
$p_k = 0.1, q_k = 1,$	OLS	9.39	9.39		1.90	8.20	9.19	9.37
$\sigma_{\tau_k} = 0, \sigma_k = 0.206$	FE	2.04		2.04	2.04	1.97	2.04	2.09
Fifth design:								
$p_k = 0.1, q_k = 1,$	OLS	1.95	1.97		1.97	56.42	4.53	2.04
$\sigma_{\tau_k} = 0.480, \sigma_k = 0$	FE	1.91		1.94	1.94	56.42	1.96	1.90

Notes. $\sqrt{N_k}$ s.d. is the standard deviation of the estimators over the simulations, multiplied by the square root of the sample size. $\sqrt{v_k}$ is the square root of the asymptotic variance in equation (2). $\sqrt{\tilde{v}_k}$ is the square root of the asymptotic variance of the fixed-effect estimator in equation (16). The remaining four columns report average values of robust, cluster, CCV, and TSCB standard errors across simulations (multiplied by $\sqrt{N_k}$). p_k and q_k are the unit and cluster sampling probabilities, respectively. σ_{τ_k} is the standard deviation of the cluster average treatment effect. σ_k is the standard deviation across clusters of the treatment assignment probabilities.

Simulations

TABLE III

COVERAGE RATES ACROSS SIMULATIONS

		Coverage of 95% confidence interval					
		$\sqrt{v_k}$	$\sqrt{\tilde{v}_k}$	Robust	Cluster	CCV	TSCB
Baseline design:							
$p_k = 1, q_k = 1,$	OLS	0.949		0.467	1.000	0.971	0.947
$\sigma_{\tau_k} = 0.120, \sigma_k = 0.057$	FE		0.950	0.893	1.000	0.947	0.942
Second design:							
$p_k = 0.1, q_k = 1,$	OLS	0.951		0.846	1.000	0.996	0.952
$\sigma_{\tau_k} = 0.120, \sigma_k = 0.057$	FE		0.950	0.944	1.000	0.950	0.948
Third design:							
$p_k = 0.1, q_k = 1,$	OLS	0.947		0.208	1.000	0.960	0.950
$\sigma_{\tau_k} = 0.480, \sigma_k = 0.206$	FE		0.941	0.284	1.000	0.918	0.948
Fourth design:							
$p_k = 0.1, q_k = 1,$	OLS	0.952		0.308	0.905	0.966	0.952
$\sigma_{\tau_k} = 0, \sigma_k = 0.206$	FE		0.952	0.951	0.932	0.951	0.955
Fifth design:							
$p_k = 0.1, q_k = 1,$	OLS	0.952		0.953	1.000	1.000	0.959
$\sigma_{\tau_k} = 0.480, \sigma_k = 0$	FE		0.954	0.955	1.000	0.957	0.949

Notes. Average coverage rates across simulations for nominal 95% confidence intervals based on the standard errors of Table II.

Implications for Practice

- No cluster sampling?
  - Unit-level attributes corr. with treatment effects?
    - Use Abadie et al. (2020) methods!
  - Random sample from large population?
    - Clustering will be ‘conservative’
  - Large sample and heterogeneous treatment effects?
    - Robust will also be conservative

Implications for Practice

- Clustered assignment?
  - Perfectly?
    - CCV not going to improve much, TSCB not useable
  - Partially?
    - If treatment assignment varies within clusters…
      - CCV and TSCB can improve things

Implications for Practice

- Cluster sampling?
  - If q_k -> 0, or
  - if q_k large, clusters large, but p_k small
    - Use standard clustering!
  - If clusters large and treatment effect varies
    - CCV and TSCB can improve things

Conclusions

- Attention shift from modelling data generating process
  - Toward design thinking
- Sampling not clustered?
  - Cluster standard errors at assignment level!
- Sampling and assignment random?
  - DO NOT CLUSTER!
- Xu (2019) goes into nonlinear world with same framework

References

- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, “When should you Cluster Standard Errors?”, QJE, 138 (2023), 1-35.
- Kish, Leslie, *Survey Sampling* (Hoboken, NJ: Wiley-Interscience, 1995).
- Moulton, Brent R., “Random Group Effects and the Precision of Regression Estimates,” *Journal of Econometrics*, 32 (1986), 385–397.
- Moulton, Brent R., “Diagnostics for Group Effects in Regression Analysis,” *Journal of Business & Economic Statistics*, 5 (1987), 275–282.
- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, “Sampling-Based versus Design-Based Uncertainty in Regression Analysis,” *Econometrica*, 88 (2020), 265–296.
- Xu, Ruonan, “Asymptotic Properties of M-estimators with Finite Populations un- der Cluster Sampling and Cluster Assignment,” Rutgers University Working paper, 2019.