11/2022

Econometrics/Labour Reading group at University of Warwick

WARWICK
THE UNIVERSITY OF WARWICK

"When should you adjust standard errors for clustering?" Abadie, Athey, Imbens and Wooldridge 2022

Presenter: Jiaqi Li

Mentor: Dr. Kenichi Nagasawa

Many thanks for helpful discussion with Prof. Wiji Arulampalam

Structure of today’s talk

- 1. Insights / caveats / take-away / for empirical research
- 2. Empirical examples.
- 3. Explanation with some details

Caution no.1 (when I think this paper is useful)

Asymptotic variance estimator when the number of cluster goes to infinity.

The setup with 52 states is probably okay.

The setup with 4 clusters (England, Welsh, Northern Ireland and Scotland) is probably not.

Binary regressor (such as treatment dummy)

No control variables.

Cross-section

Data are randomly chosen clusters and randomly chosen individuals.

Caution no.2 (when this paper is useful)

- **1. Shifting perspective**
  - Parameters of data generating processes
  - Potential outcome is stochastic.
- **2. Average treatment effects defined for the finite population at hand**
  - Such as the entire existing US residents.
  - Potential outcome is fixed.

Three motivating "empirical misconceptions"

- 1. Correlation in error structure within clusters -> The need for clustering adjustment for standard errors
  - (not entirely convinced)
- 2. There is no harm to do clustering
  - (if clustering SE making a difference, clustering should be the choice).
  - (confidence intervals may be unnecessarily conservative)
- 3. (Binary choice) either robust or clustering SE.
  - neither might be appropriate (new variance estimator is proposed "Causal cluster variance").

4. With cluster fixed effects, some people don't cluster.

fixed effects only absorb mean in the error term, but not variance heterogeneity across clusters.

Empirical example (log earnings and year of schooling)

- US Census: 2,632,838 observations, 52 clusters
- 1. Reg log(earnings) on

  state-average year of schooling

	Standard Errors					
Estimator $\rightarrow$		OLS			Fixed Effects	
Regressor $\downarrow$	EHW	CCV	LZ	EHW	CCV	LZ
State Ave $\hat{\lambda} = 1$	0.001	0.030	0.030	–	–	–

Empirical example (log earnings and year of schooling)

- US Census: 2,632,838 observations, 52 clusters

1. Regression of log earnings on state-level education indicator (year of schooling >12)

2. [Simulation]

Individual indicator is uncorrelated with cluster indicator.

	Standard Errors					
Estimator →	OLS			Fixed Effects		
Regressor ↓	EHW	CCV	LZ	EHW	CCV	LZ
State Ave $\hat{\lambda} = 1$	0.001	0.030	0.030	–	–	–
Random $\hat{\lambda} = 0.00004$	0.001	0.001	0.027	0.001	0.001	0.027

In the real data

the average education only differs slightly across states

This paper: neither is correct.

But correlation is there. So you need to do some adjustment for the standard errors.

So you don't need to do too much of adjustment for standard errors.

	Standard Errors					
Estimator →	OLS			Fixed Effects		
Regressor ↓	EHW	CCV	LZ	EHW	CCV	LZ
State Ave $\hat{\lambda} = 1$	0.001	0.030	0.030	–	–	–
Indiv Lev $\hat{\lambda} = 0.0103$	**0.001**	**0.005**	**0.027**	**0.001**	0.003	**0.028**
Random $\hat{\lambda} = 0.00004$	0.001	0.001	0.027	0.001	0.001	0.027

16

2. Set-up 1 – clustered sampling

Cross-sectional

Linear model

$$Y_i = X_i'\beta + \varepsilon_i$$

Least squares estimator:

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right)$$

How these things are implemented in practice.

- Analytic Robust (EHW) variance estimator:

Define residuals: $\widehat{\varepsilon_i} = Y_i - X_i'\hat{\beta}$

$$\widehat{V^{EHW}} = \left(\sum_{i=1}^{N} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{N} \widehat{\varepsilon_i}^2 X_i X_i'\right)\left(\sum_{i=1}^{N} X_i X_i'\right)^{-1}$$

- Analytic cluster-robust (LZ) variance estimator (In Stata):

$$\widehat{V^{LZ}} = \left(\sum_{i=1}^{N} X_i X_i'\right)^{-1} \sum_{l=1}^{m} \left(\sum_{i:M_i=l} \widehat{\varepsilon_i} X_i\right)\left(\sum_{i:M_i=l} \widehat{\varepsilon_i} X_i\right)' \left(\sum_{i=1}^{N} X_i X_i'\right)^{-1}$$

stochastic

- Sample clusters: select cluster m to be in the sample with prob $\boldsymbol{q_k}$
- Sample units: sample units randomly for subpopulation of sampled clusters with prob p_k

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

V^{EHW} assumes we see all clusters ($q_k = 1$) or $\beta_{k,m} = \beta_k$ and $p_k \to 0$

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

In conventional set with $q_k \to 0$?

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

In conventional set with $q_k \to 0$, V^{LZ} is always valid.

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2} \left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2} \left(\beta_{k,m} - \beta_k\right)^2\right\}$$

In conventional set with $q_k \rightarrow 0$, V^{LZ} is always valid.

If there is no heterogenous treatment effect?

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

In conventional set with $q_k \to 0$, V^{LZ} is always valid.

If there is no heterogenous treatment effect, V^{LZ} is always valid.

Intuitively, in general, if we think we see a large fraction of the cluster of the population, V^{LZ} is not correct if there is heterogenous treatment effects.

Proposed new variance estimator:

$$\hat{V}^{CCV} = (1 - q_k)\hat{V}^{LZ} + q_k(1 - p_k)\hat{V}^{EHW} \approx V^{TRUE}$$

In practice, knowledge of q_k is key for variance estimator.

Do we see all the clusters or not? Value of q_k cannot be estimated from data.

$$V^{TRUE} = (1 - p_k)\sigma_{y,k}^2 + (1 - q_k)\left\{n_k p_k \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

$$V^{EHW} = \sigma_{y,k}^2$$

$$V^{LZ} = (1 - p_k)\sigma_{y,k}^2 + n_k p_k \left\{\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2}\left(\beta_{k,m} - \beta_k\right)^2\right\}$$

But if we have a random sample from a large population $q_k = 1$ $p_k \rightarrow 0$

V^{EHW} is correct and V^{LZ} is not correct (unnecessarily conservative)

Structural of today

- 1. Insights/ take-away for empirical research
- 2. Empirical examples.
- 3. Explanation with some details
  - 3.1 clustered sampling
  - **3.2 clustered sampling + clustered treatment assignment**

2. Set-up 2 – clustered sampling + clustered assignment

- $A_{k,m} \in [0,1]$ **is assignment probability of treatment in cluster m in population k.**
- $D_{k,m} \in \{0,1\}$ is binary treatment assignment for unit I in population k.

Critical term in the true variance

- $\boldsymbol{n_k p_k \left(E\left[A_{k,m}^2 (1 - A_{k,m})^2 \right] - q_k \left(E\left[A_{k,m}(1 - A_{k,m}) \right] \right)^2 \right) \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k^2} (\beta_{k,m} - \beta_k)^2}$
- **This term dominates the true variance unless**
- $\boldsymbol{q_k}$**=1 and** $\boldsymbol{V(A_{k,m}) = 0}$
- **Or** $\boldsymbol{\beta_{k,m}}$ **is constant**
- **(no need for clustering)**

Critical term in the true variance

- $\hat{V}_{CCV} = \hat{\lambda}\,\hat{V}_{LZ} + (1 - \hat{\lambda})\,\hat{V}_{EHW}$
- $\hat{\lambda} = 1 - q_k \times \frac{\hat{E}^2[\bar{D}(1-\bar{D})]}{\hat{E}[\bar{D}^2(1-\bar{D})^2]}$

Simulation

Table 1. Simulation result from Abadie et al. (2022)

Treatment assignment	True std	EHW	CCV	LZ
Random	0.0012	0.0013	0.0013	0.0171
High correlation	0.0267	0.0013	0.0253	0.0312
Partial	0.0048	0.0013	0.0048	0.0177

Conclusion

Clustering adjustments requires thinking about

- Population
- Sampling scheme
- Assignment mechanism

LZ cluster variance estimator can be too drastic when correlation is only partial.

Clustering adjustments for treatment assignment can be learned from data.

Clustering adjustments for sampling process relies on outside informaiton.