Nonrobustness of the conventional cluster–robust inference: with three robust alternatives

Yuya Sasaki[1]

[1]Vanderbilt University

2023 Stata Symposium

Abstract

- Cluster-Robust (CR) standard errors.

  e.g., `cluster()` and `vce(cluster)` in Stata®

- Empirical data often contain large clusters.
- Large clusters fail existing methods:
  - Standard root-G asymptotic normality fails.
  - Conventional assumptions fail.
  - Even self-normalized CLT can fail.
- Examples
- Three alternative robust methods.

This Presentation Will Answer

Q.1 Why can the cluster-Robust (CR) standard errors

e.g., `cluster()` and `vce(cluster)`, etc.

be non-robust?

Q.2 How can we check if it is okay to use them?

Short answer: `ssc install testout`

Q.3 When it is not okay to use them, what else we can use?

Three alternative robust methods

This Presentation Is Based On:

- Sasaki, Y. and Y. Wang (2022) Non-Robustness of the Cluster-Robust Inference: with a Proposal of a New Robust Method.
- Sasaki, Y. and Y. Wang (2023) Diagnostic Testing of Finite Moment Conditions for the Consistency and Root-N Asymptotic Normality of the GMM and M Estimators.
- Chiang, H.D., Y. Sasaki, and Y. Wang (2023) Cluster-Robust Inference Robust to Large Clusters.

Existing CR Methods

- Cluster-Robust (CR) standard errors
- Conventional assumptions:
  - Cluster size N_g = small (e.g., $N_g \leq \overline{N}$)
  - Number of clusters G = large

(e.g., White, 1984; Liang and Zeger, 1986; Arellano, 1987)

- More recent assumptions:
  - $\sup_g N_g^2/N \to 0$ where $N = \sum_{g=1}^{G} N_g$.

(e.g., Djogbenou, MacKinnon, and Nielsen, 2019; Hansen and Lee, 2019; Hansen, 2022)

Problem with the Existing CR Methods

- Common example: 51 states.
- Largest cluster = California ($\approx$10%).
- $\sup_g N_g^2/N \gg \sup_g N_g/N \approx 0.1 \gg 0$.
  - e.g., $\sup_g N_g^2/N \approx 82 \gg 0$ for PSID.
- Therefore, the assumptions of $\sup_g N_g^2/N \to 0$ will not hold.
- Existing CR methods of inference may fail.

Review of the CR Methods

- Clustered sample $\{\{(Y_{gi}, X'_{gi})'\}_{i=1}^{N_g}\}_{g=1}^{G}$.
- Notations
  - $U_g = (U_{g1}, \cdots, U_{gN_g})'$ for each $g \in \{1, \cdots, G\}$.
  - $X_g = (X_{g1}, \cdots, X_{gN_g})'$ for each $g \in \{1, \cdots, G\}$.
  - $Y_g = (Y_{g1}, \cdots, Y_{gN_g})'$ for each $g \in \{1, \cdots, G\}$.
- Linear model:

$$Y_{gi} = X'_{gi}\theta + U_{gi}, \qquad \mathbb{E}[U_g | X_g] = 0,$$

- OLS: $\widehat{\theta} = \left(\sum_{g=1}^{G} \sum_{i=1}^{N_g} X_{gi} X'_{gi}\right)^{-1} \left(\sum_{g=1}^{G} \sum_{i=1}^{N_g} X_{gi} Y_{gi}\right)$
$= \left(\sum_{g=1}^{G} X'_g X_g\right)^{-1} \left(\sum_{g=1}^{G} X'_g Y_g\right).$

Review of the CR Methods

- Common CR variance estimators:

$$\widehat{V}_{\hat{\theta}}^{\mathrm{CR}} = a_n \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1} \left(\sum_{g=1}^{G} \widehat{S}_g \widehat{S}_g'\right) \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1},$$

  where $a_n \to 1$ is a suitable finite-sample adjustment and $\widehat{S}_g = \sum_{i=1}^{N_g} X_{gi} \widehat{U}_{gi}$ with $\widehat{U}_{gi} = Y_{gi} - X_{gi}' \widehat{\theta}$.

- Also used is the jackknife variance estimator:

$$\widehat{V}_{\hat{\theta}}^{\mathrm{CR,JACK}} = \sum_{g=1}^{G} \left(\widehat{\theta}_{-g} - \widehat{\theta}\right) \left(\widehat{\theta}_{-g} - \widehat{\theta}\right)',$$

  where $\widehat{\theta}_{-g} = \left(\sum_{h \neq g} X_h' X_h\right)^{-1} \left(\sum_{h \neq g} X_h' Y_h\right)$ denotes the leave-one-cluster-out estimator.

A Pitfall in the CR Inference

- Standard asymptotic argument:

$$\sqrt{G}\left(\widehat{\theta}-\theta\right)=\underbrace{\left(\frac{1}{G}\sum_{g=1}^{G}\Xi_g\right)^{-1}}_{\xrightarrow{p}\ Q}\underbrace{\left(\frac{1}{\sqrt{G}}\sum_{g=1}^{G}S_g\right)}_{\xrightarrow{d}\mathcal{N}(0,V)}\xrightarrow{d}\mathcal{N}(0,Q^{-1}VQ^{-1})$$

as $G\to\infty$, where

$$Q=\mathbb{E}[\Xi_g], \qquad \Xi_g=\sum_{i=1}^{N_g}X_{gi}X'_{gi},$$

$$V=\mathrm{Var}[S_g], \qquad S_g=\sum_{i=1}^{N_g}X_{gi}U_{gi}.$$

- This in particular requires $\mathbb{E}[\|S_g\|^2]<\infty$ among others.

Unbounded Second Moments of the Score

- Recall: we need $\mathbb{E}[\|S_g\|^2] < \infty$ where $S_g = \sum_{i=1}^{N_g} X_{gi}U_{gi}$.
- Short-hand notations:
  - Σ_g = the j-th coordinate of S_g.
  - Z_{gi} = the j-th coordinate of $X_{gi}U_{gi}$.

Theorem (Sasaki and Wang, 2022)

Suppose that (i) $\mathbb{E}[Z_{gi}^2|N_g] \in [C_1, C_2]$ for $0 < C_1 < C_2 < \infty$ a.s. for all i and g, and (ii) $Cov[Z_{gi}, Z_{gj}|N_g] \geq C > 0$ almost surely for all i, j, and g. Then, $\mathbb{E}[\Sigma_g^2] < \infty$ if and only if $\mathbb{E}[N_g^2] < \infty$.

On a Conventional Assumption for CR Inference

- H is *regularly varying* (RV) at infinity if it satisfies

$$\frac{1 - H(xt)}{1 - H(t)} \to x^{-\beta} \text{ as } t \to \infty$$

for any $x > 0$ and some constant $\beta > 0$.

- β = *tail exponent*, which measures the tail heaviness of H.

- RV is a mild condition satisfied by common distribution families.

On a Conventional Assumption for CR Inference

Theorem (Sasaki and Wang, 2022)

If N_g is i.i.d. across g and its distribution satisfies the RV condition with exponent $1 < \beta < 2$, then

$$\frac{\sup_{1 \leq g \leq G} N_g^2}{N} \to \infty \text{ with probability approaching 1 as } G \to \infty.$$

- Implication: the conventional assumption $\sup_g N_g^2 / N \xrightarrow{p} 0$ of CR inference is implausible if $\beta < 2$ (equivalently, $\mathbb{E}[N_g^2] = \infty$).

Non-Gaussian Limit

- For simplicity, consider $\theta = \mathbb{E}[Y_{gi}]$.
- Estimator $\hat{\theta} = N^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} Y_{gi}$
- Self-normalized CLT $\widehat{\mathbb{E}}[(\hat{\theta} - \theta)^2]^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$ if $\beta > 2$.
- Counter-examples of the self-normalized Gaussianity if $\beta < 2$.

(Sasaki and Wang, 2022)

Necessary and Sufficient Condition

- η is stable if there exists a sequence of i.i.d. random variables $\xi_1, \xi_2, \ldots$ and sequences of positive numbers $\{A_G\}_G$ and real numbers $\{D_G\}_G$ such that

$$\frac{\sum_{g=1}^{G} \xi_g - D_G}{A_G} \xrightarrow{d} \eta \quad \text{as } G \to \infty$$

- If η is stable, then $A_G = G^{1/\alpha} L(G)$ for an index of stability $\alpha \leq 2$ and a slowly varying function L.[1]

- If $\beta < 2$, then $\alpha = \beta$.[2]

[1] For any $x > 0$, $L(xt)/L(t) \to 1$ as $t \to \infty$.

[2] This relation holds under some regularity conditions on the tails.

Necessary and Sufficient Condition

- For simplicity of writings, let θ be a scalar.

Theorem (Chiang, Sasaki, and Wang, 2023)

Suppose that

- $\{(X_g', S_g)\}_{g=1}^G$ *is i.i.d.,*
- $\mathbb{E}[N_g] \in (0, \infty)$*, and*
- $Q^{-1} S_g$ *and* Ξ_g *belong to the domain of attraction with an index of stability* $\alpha \in (1, 2]$.

$\sqrt{G}\left(\widehat{\theta} - \theta\right) \Big/ \widehat{V}_{\hat{\theta}}^{CR}$ *is asymptotically normal* *if and only if* $\alpha = 2$.

Wild Cluster Bootstrap

- Again, for simplicity of writings, let θ be a scalar.

Theorem (Chiang, Sasaki, and Wang, 2023)

Suppose that

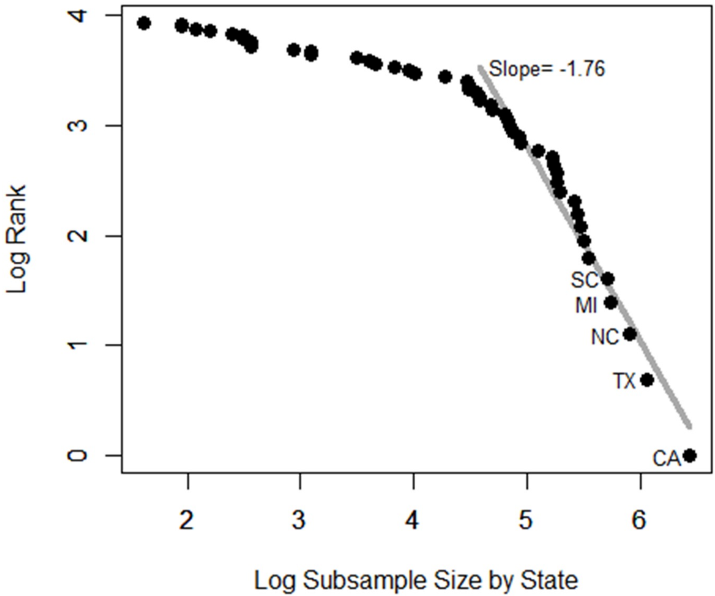
- $\{(X_g', S_g)\}_{g=1}^{G}$ *is i.i.d.,*
- $\mathbb{E}[N_g] \in (0, \infty)$, *and*
- $Q^{-1}S_g$ *and* Ξ_g *belong to the domain of attraction with an index of stability* $\alpha \in (1, 2]$.

The wild cluster bootstrap with Rademacher auxiliary random variables is inconsistent if $\alpha < 2$.

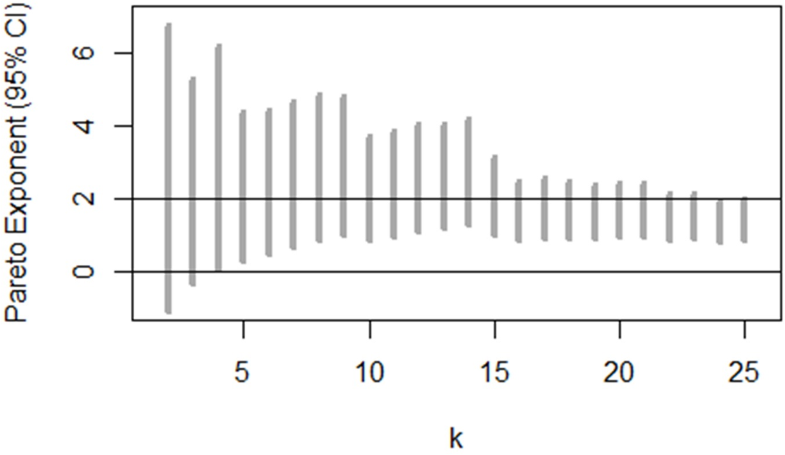
Example: 51 States

- 51 states: the most common example of cluster sampling.
- $\beta < 2$ (or equivalently $\mathbb{E}[N_g^2] = \infty$)?
- Male individulas in the 2019 wave of the PSID.
- $N = 4808$ and $G = 51$.
- Largest states
  - $N_4 = 629$ (California)
  - $N_{42} = 430$ (Texas)
  - $N_{32} = 366$ (North Carolina)
  - $N_{21} = 310$ (Michigan)
  - $N_{39} = 304$ (South Carolina)

Log-Log Plot of the 51 State Clusters



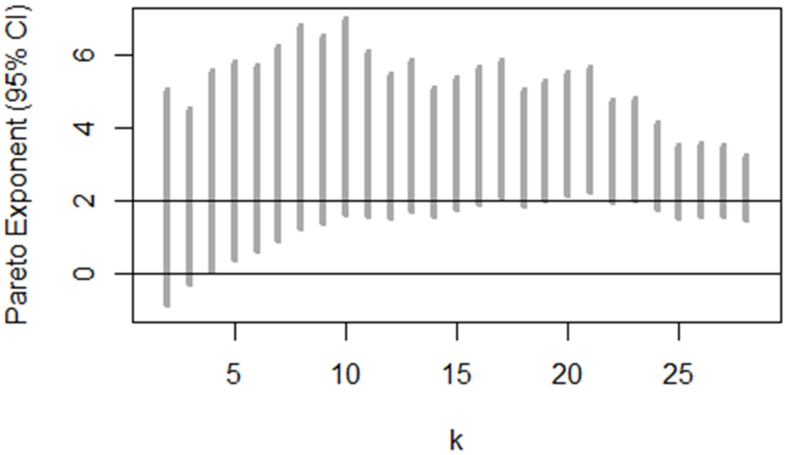
Hill Plot for the 51 State Clusters



- Implication: $\beta < 2$ cannot be ruled out.

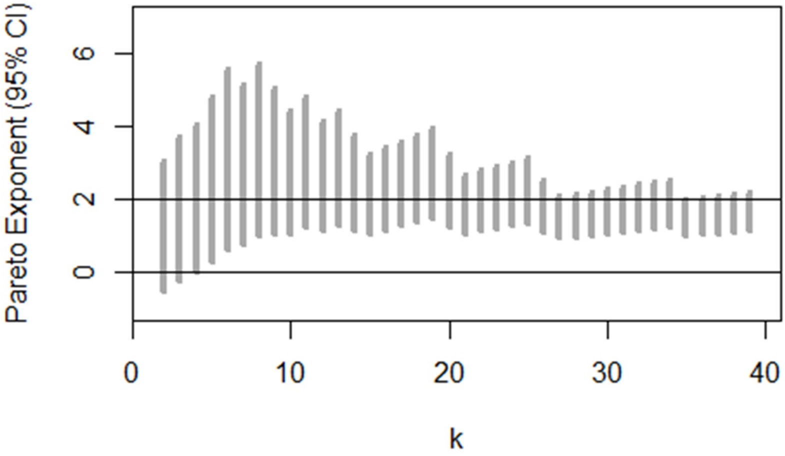
Three Examples from *Econometrica*, 2020 – #1

- State Clusters in the U.S.



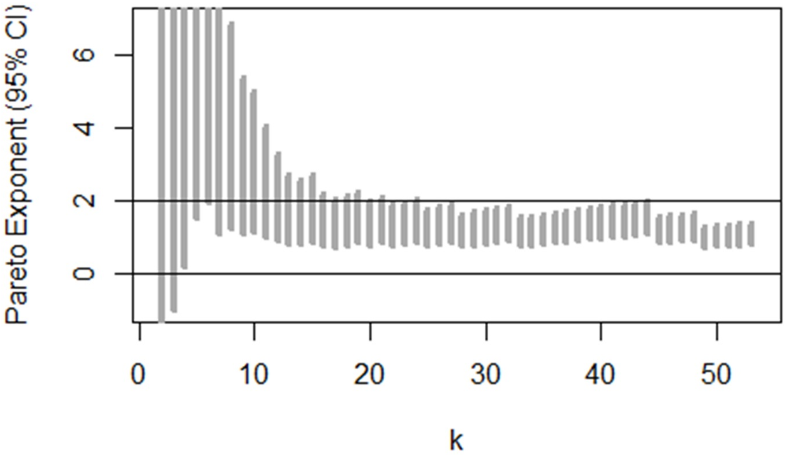
Three Examples from *Econometrica*, 2020 – #2

- Region Clusters in Russia



Three Examples from *Econometrica*, 2020 – #3

- Branch Clusters



Stata® Command for Dyagnostic Testing

Stata® command to test bounded second moments:

```
.  ssc install testout
```

Example usage:

```
.  testout y x1 x2 ...  [if] [in] [, iv(varlist)
                                        cluster(varname)]
```

This command is based on Sasaki & Wang (2023).

Alternative Robust Methods

- So far, we have seen that the commonly used methods

  e.g., `cluster()` and `vce(cluster)` in Stata®

  are not robust to large clusters, e.g., the 51 U.S. states.

- Alternative methods of cluster-robust inference:

  1. CLT assuming weak within-cluster dependence.
  2. Reweighting observations to invoke CLT.
  3. Forget CLT and use subsampling.

  Details to follow in the next slide.

Alternative Robust Methods

1. Assume conditions for within-cluster CLT.

   Hansen (2007), Ibragimov and Müller (2010, 2016), Bester, Conley, and Hansen (2011), Canay, Santos, and Shaikh (2021), etc.

2. Reweighting observations to invoke CLT.

   Athey and Imbens (2017), Chandar, Hortaçsu, List, Muir, and Wooldridge (2019), Bai, Liu, Shaikh, and Tabord-Meehan (2022), Bugni, Canay, Shaikh, and Tabord-Meehan (2022), Sasaki and Wang (2022), etc. $\Leftarrow$ Details to follow

3. Forget CLT and use subsampling.

   Chiang, Sasaki, and Wang (2023). $\Leftarrow$ Details to follow

Reweighted Estimator

- Instead of

$$\widehat{\theta} = \left(\sum_{g=1}^{G} X_g' X_g \right)^{-1} \left(\sum_{g=1}^{G} X_g' Y_g \right),$$

- we propose

$$\widehat{\theta}^{\mathrm{WCR}} = \left(\sum_{g=1}^{G} N_g^{-1} X_g' X_g \right)^{-1} \left(\sum_{g=1}^{G} N_g^{-1} X_g' Y_g \right).$$

Reweighted Asymptotic Variance Estimator

- Instead of

$$\widehat{V}_{\hat{\theta}}^{\text{CR}} = a_n \left(\sum_{g=1}^{G} X_g' X_g \right)^{-1} \left(\sum_{g=1}^{G} \widehat{S}_g \widehat{S}_g' \right) \left(\sum_{g=1}^{G} X_g' X_g \right)^{-1},$$

- we propose

$$\widehat{V}_{\hat{\theta}}^{\text{WCR}} = a_n \left(\sum_{g=1}^{G} N_g^{-1} X_g' X_g \right)^{-1} \left(\sum_{g=1}^{G} N_g^{-2} \widehat{S}_g \widehat{S}_g' \right) \left(\sum_{g=1}^{G} N_g^{-1} X_g' X_g \right)^{-1}.$$

Reweighted Jackknife Estimator

- Instead of

$$\widehat{V}_{\hat{\theta}}^{\text{CR,JACK}} = \sum_{g=1}^{G} \left(\widehat{\theta}_{-g} - \widehat{\theta}\right)\left(\widehat{\theta}_{-g} - \widehat{\theta}\right)',$$

where $\widehat{\theta}_{-g} = \left(\sum_{h \neq g} X_h' X_h\right)^{-1} \left(\sum_{h \neq g} X_h' Y_h\right),$

- we propose

$$\widehat{V}_{\hat{\theta}}^{\text{WCR,JACK}} = \sum_{g=1}^{G} \left(\widehat{\theta}_{-g}^{\text{WCR}} - \widehat{\theta}^{\text{WCR}}\right)\left(\widehat{\theta}_{-g}^{\text{WCR}} - \widehat{\theta}^{\text{WCR}}\right)',$$

where $\widehat{\theta}_{-g}^{\text{WCR}} = \left(\sum_{h \neq g} N_h^{-1} X_h' X_h\right)^{-1} \left(\sum_{h \neq g} N_h^{-1} X_h' Y_h\right).$

Theoretical Justification

Assumption (Within-Cluster Conditions)

1. $\mathbb{E}\left[\|X_{gi}\|^4 | N_g\right] \in [C_3, C_4]$ for $0 < C_3 < C_4 < \infty$ almost surely for all i and g. 2. $\mathbb{E}\left[\|X_{gi}U_{gi}\|^2 | N_g\right] \in [C_5, C_6]$ for $0 < C_5 < C_6 < \infty$ almost surely for all i and g.

Assumption (Across-Cluster Conditions)

1. (N_g, X_g, Y_g) is i.i.d. across g. 2. $\mathbb{E}[N_g^{-1} X_g' X_g]$ is non-singular.

Theoretical Justification

Theorem (Sasaki and Wang, 2022)

Under the two assumptions,

$$\sqrt{G}(\widehat{\theta}^{WCR} - \theta) \xrightarrow{d} \mathcal{N}(0, V^{WCR})$$

as $G \to \infty$, *where*

$$V^{WCR} = (\mathbb{E}[N_g^{-1} X_g' X_g])^{-1} (\mathbb{E}[N_g^{-2} U_g' X_g' X_g U_g]) (\mathbb{E}[N_g^{-1} X_g' X_g])^{-1}.$$

Furthermore, we have $G\widehat{V}_{\hat{\theta}}^{WCR} \xrightarrow{p} V^{WCR}$ *as* $G \to \infty$.

Proposition (Sasaki and Wang, 2022)

Under the two assumptions, $G\widehat{V}_{\hat{\theta}}^{\mathrm{WCR,JACK}} \xrightarrow{p} V^{\mathrm{WCR}}$ as $G \to \infty$.

Subsampling[3]

- Suppose that you do *not* want reweighting.

- Then, one cannot use the conventional critical value of 1.96.

- But one can still compute the valid critical value by subsampling (Chiang, Sasaki, and Wang, 2023).

[3] Precisely, it is a "'score' subsampling." We will call it 'subsampling' in short throughout this presentation.

Subsampling

- Let $b < G$ be a subsample size.
- Let $S_j \subset \{1, \ldots, G\}$ be a subsample s.t., $|S_j| = b$.
- For this j-th subsample S_j,

$$\widehat{\theta}_{b,j} = \left(\frac{G}{b}\right) \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1} \sum_{g \in S_j} X_g' Y_g$$

$$\widehat{V}^{\mathrm{CR}}_{\hat{\theta},b,j} = \left(\frac{G}{b}\right)^2 \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1} \left(\sum_{g \in S_j} \widehat{S}_{g,j} \widehat{S}'_{g,j}\right) \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1}$$

where $\widehat{S}_{g,j} = X_g'(Y_g - X_g \widehat{\theta}_{b,j})$.

Subsampling, Continued

- For simplicity of writings, let θ be a scalar.
- Set $M =$ a large number, e.g., 2,000.
- Generate such subsamples $S_1, \ldots, S_M \subset \{1, \ldots, G\}$ M times.
- Obtain M subsampled t-statistics $\frac{\widehat{\theta}_{b,1} - \widehat{\theta}}{\widehat{V}^{\mathrm{CR}}_{\hat{\theta},b,1}}, \ldots, \frac{\widehat{\theta}_{b,M} - \widehat{\theta}}{\widehat{V}^{\mathrm{CR}}_{\hat{\theta},b,M}}$.
- Its $(1-\alpha)$ quantile can be used as the critical value $\widehat{c}_{G,b}(1-\alpha)$.

Theoretical Justification

- Again, for simplicity of writings, let θ be a scalar.

Theorem (Chiang, Sasaki, and Wang, 2023)

Suppose that

- $\{(X_g', S_g)\}_{g=1}^G$ *is i.i.d.,*
- $\mathbb{E}[N_g] \in (0, \infty)$, *and*
- $Q^{-1}S_g$ *and* Ξ_g *belong to the domain of attraction with an index of stability* $\alpha \in (1, 2]$.

Then,

$$\mathbb{P}\left(\frac{\widehat{\theta} - \theta}{\widehat{V}_{\hat{\theta}}^{CR}} \leq \widehat{c}_{G,b}(1-\alpha)\right) \to 1 - \alpha.$$

Design

- Folllowing existing papers, we consider

$$Y_{gi} = \theta_0 + \theta_1 T_g + \sum_{j=1}^{K} \theta_j X_{g,j+1} + U_{gi}.$$

- T_g = cluster treatment (as in cluster RCT). 20% gets treated.
- N_g-variate random vectors, $(\tilde{X}_{g1j}, \cdots, \tilde{X}_{gN_gj})' \sim \mathcal{N}(0, \Omega)$, where $\Omega_{ii} = 1.0$ and $\Omega_{ii'} = 0.5$ if $i \neq i'$.

Four Alternative CR Standard Errors

Under cluster sizes $N_g \sim \lceil 10 \cdot \text{Pareto}(1, \beta) \rceil$, we compare:

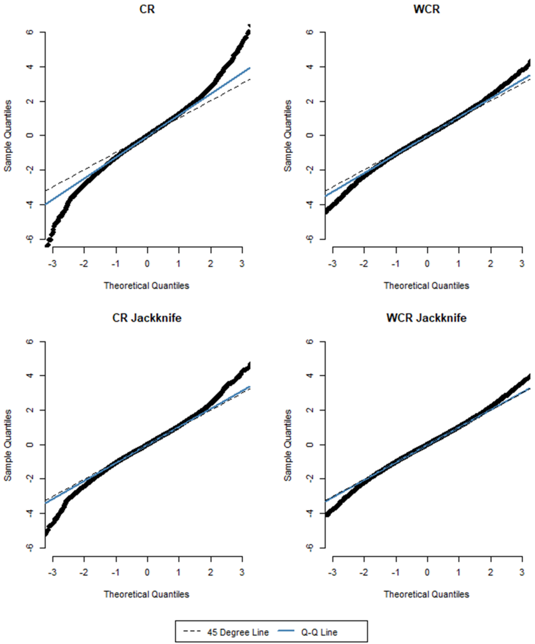
$$\text{(CR)} \qquad \left(\widehat{\theta}_1 - \theta_1\right) \Big/ \sqrt{\widehat{V}^{\text{CR}}_{\hat{\theta},11}},$$

$$\text{(CR Jackknife)} \qquad \left(\widehat{\theta}_1 - \theta_1\right) \Big/ \sqrt{\widehat{V}^{\text{CR,JACK}}_{\hat{\theta},11}},$$

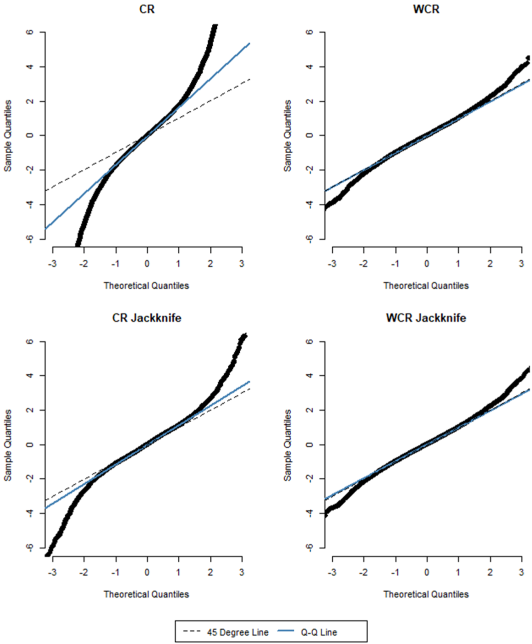
$$\text{(WCR)} \qquad \left(\widehat{\theta}^{\text{WCR}}_1 - \theta_1\right) \Big/ \sqrt{\widehat{V}^{\text{WCR}}_{\hat{\theta},11}},$$

$$\text{(WCR Jackknife)} \qquad \left(\widehat{\theta}^{\text{WCR}}_1 - \theta_1\right) \Big/ \sqrt{\widehat{V}^{\text{WCR,JACK}}_{\hat{\theta},11}}.$$

QQ-Plots of Self-Normalizd Statistics under $\beta = 2$



QQ-Plots of Self-Normalizd Statistics under $\beta = 1$



Monte Carlo Statistics

		Unweighted					Weighted		
		MSE	Rejection ($p = 0.050$)				MSE	Rejection ($p = 0.050$)	
K	β	$\widehat{\theta}_1$	$\widehat{V}^{\text{CR}}_{\hat{\theta},11}$	$\widehat{V}^{\text{CR,JACK}}_{\hat{\theta},11}$	WCB	SUB	$\widehat{\theta}^{\text{WCR}}_1$	$\widehat{V}^{\text{WCR}}_{\hat{\theta},11}$	$\widehat{V}^{\text{WCR,JACK}}_{\hat{\theta},11}$
0	2.00	(0.086)	0.111	0.086	0.067	**0.054**	(0.075)	0.087	0.072
	1.75	(0.091)	0.127	0.081	0.074	**0.053**	(0.073)	0.085	0.065
	1.50	(0.096)	0.144	0.084	0.090	**0.063**	(0.072)	0.080	0.066
	1.25	(0.111)	0.178	0.094	0.117	0.071	(0.071)	0.080	**0.067**
	1.00	(0.140)	0.234	0.086	0.158	0.104	(0.068)	0.068	**0.061**
1	2.00	(0.085)	0.111	0.072	0.065	**0.042**	(0.072)	0.085	0.066
	1.75	(0.088)	0.118	0.077	0.073	**0.047**	(0.072)	0.084	0.069
	1.50	(0.098)	0.141	0.082	0.085	**0.054**	(0.070)	0.078	0.062
	1.25	(0.109)	0.168	0.085	0.105	**0.065**	(0.068)	0.081	0.071
	1.00	(0.140)	0.238	0.092	0.154	0.097	(0.067)	0.070	**0.064**
4	2.00	(0.081)	0.109	0.070	0.064	**0.037**	(0.073)	0.086	**0.063**
	1.75	(0.087)	0.124	0.077	0.075	**0.043**	(0.072)	0.082	0.064
	1.50	(0.094)	0.147	0.084	0.089	**0.053**	(0.074)	0.088	0.072
	1.25	(0.108)	0.182	0.091	0.108	**0.062**	(0.072)	0.081	0.071
	1.00	(0.134)	0.244	0.096	0.152	0.086	(0.068)	0.078	**0.065**

Table: cluster sizes follow $N_g \sim \lceil \text{Pareto}(1, \beta) \rceil$.

Alternative Robust Methods

1. Assume conditions for within-cluster CLT.

   Hansen (2007), Ibragimov and Müller (2010, 2016), Bester, Conley, and Hansen (2011), Canay, Santos, and Shaikh (2021), etc.

2. Reweighting observations to invoke CLT.

   Athey and Imbens (2017), Chandar, Hortaçsu, List, Muir, and Wooldridge (2019), Bai, Liu, Shaikh, and Tabord-Meehan (2022), Bugni, Canay, Shaikh, and Tabord-Meehan (2022), Sasaki and Wang (2022), etc.

3. Forget CLT and use subsampling.

   Chiang, Sasaki, and Wang (2023).

Pros and Cons of the Alternative Robust Methods

	0. Conventional	1. Panel Cluster	2. Reweighting	3. Subsampling
Invariance in Estimand e.g., under Heterogeneous Treatment Effects	◎	◎	× Estimand And Estimate Change	◎
Allows for Strong Within-Cluster Dependence	◎	× Weak Dependence Is Assumed	◎	◎
Allows for Heavy-Tailed Cluster-Size Distributions	× Two+ Moments Are Required	◎	◎	○ Allowed Under Power Law

0. Conventional methods: White (1984), Liang and Zeger (1986), Arellano (1987), etc., e.g., `cluster()` and `vce(cluster)` in Stata.
1. Panel cluster robust methods: Hansen (2007), Ibragimov and Müller (2010, 2016), Bester, Conley, and Hansen (2011), Canay, Santos, and Shaikh (2021), etc.
2. Reweighting: Athey and Imbens (2017), Chandar, Hortaçsu, List, Muir, and Wooldridge (2019), Bai, Liu, Shaikh, and Tabord-Meehan (2022) Bugni,Canay, Shaikh, and Tabord-Meehan (2022), Sasaki and Wang (2022), etc.
3. Subsampling: Chiang, Sasaki, and Wang (2023).

Summary

- Cluster-Robust (CR) standard errors.

  e.g., `cluster()` and `vce(cluster)` in Stata®

- Empirical data often contain large clusters.

- Large clusters fail existing methods:
  - Standard root-G asymptotic normality fails.
  - Conventional assumptions fail.
  - Even self-normalized CLT can fail.

- Examples

- Alternative robust methods – pros and cons.

We Answered The Following Questions

Q.1 Why can the cluster-Robust (CR) standard errors

e.g., `cluster()` and `vce(cluster)`, etc.

be non-robust?

Q.2 How can we check if it is okay to use them?

Q.3 When it is not okay to use them, what else we can use?

Detailed answers to follow in the next 3 slides.

Question #1

Q.1 Why can the cluster-Robust (CR) standard errors

e.g., `cluster()` and `vce(cluster)`, etc.

be non-robust?

A
- Presence of large clusters, e.g., California.
- Formally, $\alpha < 2$ or $\beta < 2$ fails the CR standard errors.

Question #2

Q.2 How can we check if it is okay to use them?

A
- Visually, use the log-log plot.
- Formally but still visually, use the Hill plot.
- More formally, conduct the test by the `testout` command.
  Installation: `ssc install testout`

Question #3

Q.3 When it is not okay to use them, what else we can use?

A
1. Panel cluster methods that assume within-cluster CLT;
2. Reweighting; or
3. Subsampling.

Be aware of the pros and cons.

Links

- Stata® command `testout` for dyagnostic testing:

```
.  ssc install testout
```

  This command is based on Sasaki & Wang (2023).

- The reweighting method is Sasaki & Wang (2022).

- The subsampling method is Chiang, Sasaki, & Wang (2023). Theoretically, this subsampling method strictly dominates `cluster()` & `vce(cluster)` in terms of the generality of the required conditions.

My Stata Webpage

- The current slides and other resources are made available on my Stata webpage:

https://sites.google.com/site/yuyasasaki/Home/stata

Thanks!
Yuya Sasaki