Econometrics **October 26, 2023**

Topic 5: Two-Way Cluster-Robust (TWCR) Standard Errors

by Sai Zhang

Key points: The validity of Two-Way Cluster-Robust (TWCR) standard errors

Disclaimer: *This note is compiled by Sai Zhang.*

5.1 One-Way Clustering

First, consider the case of one-way clustering. The linear model with one-way clustering

$$y_{ig} = \mathbf{x}_{ig}\boldsymbol{\beta} + u_{ig}$$

where i denotes the ith of the N individuals in the sample, j denotes the gth of the G clusters, assume that

- $\mathbb{E}\left[u_{ig} \mid \mathbf{x}_{ig}\right] = 0$
- error independence across clusters: for $i \neq j$

$$\mathbb{E}\left[u_{ig}u_{jg'} \mid \mathbf{x}_{ig}, \mathbf{x}_{jg'}\right] = 0 \tag{5.1}$$

  unless $g = g'$, that is, errors for individuals within the same cluster may be correlated.

Grouping observations by cluster, get

$$\mathbf{y}_g = \mathbf{X}_g\boldsymbol{\beta} + \mathbf{u}$$

where $\mathbf{X}_g$ has dimension $N_g \times K$ and $\mathbf{y}_g$ has dimension $N_g \times 1$, with N_g observations in cluster g. Stacking over cluster, get the matrix form of the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

with $\mathbf{y}$, $\mathbf{u}$ being $N \times 1$ vectors, $\mathbf{X}$ being an $N \times K$ matrix. OLS estimator gives

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\sum_{g=1}^{G}\mathbf{X}_g'\mathbf{X}_g\right)^{-1}\sum_{g=1}^{G}\mathbf{X}_g'\mathbf{y}_g \tag{5.2}$$

then, by CLT, we have that $\sqrt{G}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma})$ where the variance matrix of the limit normal distribution $\boldsymbol{\Sigma}$ is

$$\left(\lim_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G}\mathbf{E}\left[\mathbf{X}_g'\mathbf{X}_g\right]\right)^{-1}\left(\lim_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G}\mathbf{E}\left[\mathbf{X}_g'\mathbf{u}_g'\mathbf{u}_g\mathbf{X}_g\right]\right)\times\left(\lim_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G}\mathbf{E}\left[\mathbf{X}_g'\mathbf{X}_g\right]\right)^{-1} \tag{5.3}$$

If the primary source of clustering is due to group-level common shocks, a useful approximation is that for the jth regressor, the default OLS variance estimate based on $s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ should be inflated by $\tau_j \simeq 1 + \rho_{x_j}\rho_u\left(\overline{N}_g - 1\right)$, where

- s is the estimated standard deviation of the error

- ρ_{x_j} is a measure of within-cluster correlation of x_j
- ρ_u is the within-cluster error correlation
- $\overline{N}_g$ is the average cluster size

It's easy to see the τ_j can be large even with small ρ_u (Kloek, 1981; Scott and Holt, 1982; Moulton, 1990). If assume the model for the cluster error variance matrices $\mathbf{\Omega}_g = \mathbb{V}\left[\mathbf{u}_g \mid \mathbf{X}_g\right] = \mathbb{E}\left[\mathbf{u}_g\mathbf{u}_g' \mid \mathbf{X}_g\right]$, and there is a consistent estimate $\hat{\mathbf{\Omega}}_g$ of $\mathbf{\Omega}_g$, we can estimate $\mathbb{E}\left[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g\right] = \mathbb{E}\left[\mathbf{X}_g'\mathbf{\Omega}_g\mathbf{X}_g\right]$ via GLS.

Cluster-robust variance matrix estimate consider

$$\hat{\mathbb{V}}\left[\hat{\boldsymbol{\beta}}\right] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{g=1}^{G}\mathbf{X}_g'\hat{\mathbf{u}}_g\hat{\mathbf{u}}_g'\mathbf{X}_g\right)(\mathbf{X}'\mathbf{X})^{-1} \tag{5.4}$$

where $\hat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{X}_g\hat{\boldsymbol{\beta}}$. This estimate is consistent if

$$G^{-1}\sum_{g=1}^{G}\mathbf{X}_g'\hat{\mathbf{u}}_g\hat{\mathbf{u}}_g'\mathbf{X}_g - G^{-1}\sum_{g=1}^{G}\mathbb{E}\left[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g\right] \xrightarrow{\text{p}} \mathbf{0}$$

as $G \to \infty$. An informal presentation of Eq.(5.4) is to rewrite the central matrix as

$$\hat{\mathbf{B}} = \sum_{g=1}^{G}\mathbf{X}_g'\hat{\mathbf{u}}_g\hat{\mathbf{u}}_g'\mathbf{X}_g = \mathbf{X}'\begin{bmatrix}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1' & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \hat{\mathbf{u}}_2\hat{\mathbf{u}}_2' & & \vdots\\ \vdots & & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & & \hat{\mathbf{u}}_G\hat{\mathbf{u}}_G'\end{bmatrix}\mathbf{X} = \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^G\right)\mathbf{X} \tag{5.5}$$

where $\otimes$ denotes element-wise multiplication. The (p,q)th element of this matrix is

$$\sum_{i=1}^{N}\sum_{j=1}^{N}x_{ia}x_{jb}\hat{u}_i\hat{u}_j \cdot \mathbf{1}\left(i, j \text{ in the same cluster}\right)$$

with $\hat{u}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$.

$\mathbf{S}^G$ is an $N \times N$ indicator matrix with $\mathbf{S}^G_{ij} = 1$ only if the ith and jth observation belong to the same cluster: it zeros out a large amount of $\hat{\mathbf{u}}\hat{\mathbf{u}}'$ (asymptotically equivalently, $\mathbf{u}\mathbf{u}'$), specifically, only $\sum_{g=1}^{G}N_g^2$ out of $N^2 = \left(\sum_{g=1}^{G}N_g\right)^2$ terms are not zero (sub-matrices on the diagonal). Asymptotically

- for fixed N_g, $\frac{1}{N^2}\sum_{g=1}^{G}N_g^2 \xrightarrow{G\to\infty} 0$
- for balanced clusters $N_g = N/G$, $\frac{1}{N^2}\sum_{g=1}^{G}N_g^2 = \frac{1}{G} \xrightarrow{G\to\infty} 0$

A strand of literature popularizes this method:

- Liang and Zeger (1986): in a generalized estimatin equations setting
- Arellano (1987): fixed effects estimator in linear panel models
- Hansen (2007): asymptotic theory for panel data where $T \to \infty$ in addition to $N \to \infty$ (or $N_g \to \infty$ in addition to $G \to \infty$ in the notation above).

5.2 Two-Way Clustering

Now, consider the case of two-way clustering,

$$y_{i,gh} = \mathbf{x}'_{i,gh}\boldsymbol{\beta} + u$$

where each observation may belong to **two** dimension of groups: group $g \in \{1, \cdots, G\}$ and $h \in \{1, \cdots, H\}$, and for $i \neq j$

$$\mathbb{E}\left[u_{i,gh}u_{j,g'h'} \mid \mathbf{x}_{i,gh}, \mathbf{j}, \mathbf{g'h'}\right] = 0 \tag{5.6}$$

unless $g = g'$ or $h = h'$, that is, errors for individuals within the same group (along either g or h) may be correlated.

Cluster-robust variance matrix estimate extending the one-way clustering case, keep elements of $\hat{\mathbf{u}}\hat{\mathbf{u}}'$ where the ith and jth observations share a cluster in **any** dimension, then similar to Eq.(5.5)

$$\hat{\mathbf{B}} = \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{GH}\right)\mathbf{X} \tag{5.7}$$

here $\mathbf{S}^{GH}$ is an $N \times N$ indicator matrix with $\mathbf{S}^{GH}_{ij} = 1$ only if the ith and jth observation share any cluster, the (p, q)th element of this matrix is

$$\sum_{i=1}^{N}\sum_{i=1}^{N} x_{ia}x_{jb}\hat{u}_i\hat{u}_j \cdot \mathbf{1}\left(i, j \text{ share any cluster}\right)$$

$\hat{\mathbf{B}}$ can also be presented in one-way cluster-robust fashion:

$$\hat{\mathbf{B}} = \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{GH}\right)\mathbf{X} = \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{G}\right)\mathbf{X} + \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{H}\right)\mathbf{X} - \mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{G\cap H}\right)\mathbf{X} \tag{5.8}$$

where $\mathbf{G}^{GH} = \mathbf{G}^{G} + \mathbf{G}^{H} - \mathbf{G}^{G\cap H}$, with

- $\mathbf{G}^{G}$: $\mathbf{G}^{G}_{ij} = 1$ only if the ith and jth observation belong to the same cluster $g \in \{1, 2, \cdots, G\}$
- $\mathbf{G}^{H}$: $\mathbf{G}^{H}_{ij} = 1$ only if the ith and jth observation belong to the same cluster $h \in \{1, 2, \cdots, H\}$
- $\mathbf{G}^{G\cap H}$: $\mathbf{G}^{G\cap H}_{ij} = 1$ only if the ith and jth observation belong to **both** the same cluster $g \in \{1, 2, \cdots, G\}$ and the same cluster $h \in \{1, 2, \cdots, H\}$

then, similar to one-way clustering case,

$$\begin{aligned}\hat{\mathbb{V}}\left[\hat{\boldsymbol{\beta}}\right] &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{G}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &+ \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{H}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &- \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\hat{\mathbf{u}}\hat{\mathbf{u}}' \otimes \mathbf{S}^{G\cap H}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\end{aligned} \tag{5.9}$$

that is,

$$\hat{\mathbb{V}}\left[\hat{\boldsymbol{\beta}}\right] = \hat{\mathbb{V}}^{G}\left[\hat{\boldsymbol{\beta}}\right] + \hat{\mathbb{V}}^{H}\left[\hat{\boldsymbol{\beta}}\right] - \hat{\mathbb{V}}^{G\cap H}\left[\hat{\boldsymbol{\beta}}\right] \tag{5.10}$$

each of Eq.(5.10) can be separately computed by OLS of $\mathbf{y}$ on $\mathbf{X}$, with variance matrix estimates $\hat{\mathbb{V}}$ based on

i clustering on $g \in \{1, 2, \cdots, G\}$
ii clustering on $h \in \{1, 2, \cdots, H\}$
iii clustering on $(g, h) \in \{(1, 1), \cdots, (G, H)\}$

Practical considerations It is required to know what *ways* will be potentially important for clustering, which can be tested via checking the dimension of correlations in the errors. There are several ways to test

- estimate sample covariances of $\mathbf{X}'\hat{\mathbf{u}}$ within dimensions, test the null that the **average** of such covariances is 0: rejecting this null is sufficient (not necessary) to reject the null of no clustering (White, 1980)
- for **small samples**, Eq. (5.4) is baised downwards. This is corrected (in Stata) by replacing $\hat{\mathbf{u}}_g$ with $\sqrt{c}\hat{\mathbf{u}}_g$, where $c = \frac{G}{G-1}\frac{N-1}{N-K} \simeq \frac{G}{G-1}$. For two-way clustering (Eq. 5.8), there are 2 ways of correction:
  - choose correction terms for each of the 3 components:
  $$c_1 = \frac{G}{G-1}\frac{N-1}{N-K}, c_2 = \frac{H}{H-1}\frac{N-1}{N-K}, c_3 = \frac{I}{I-1}\frac{N-1}{N-K}$$
  with I being the number of unique clusters determined by $G \cap H$
  - choose a constant terms for all components:
  $$c = \frac{J}{J-1}\frac{N-1}{N-K}$$
  with $J = \min(G, H)$
- $\hat{\mathbb{V}}\left[\hat{\boldsymbol{\beta}}\right]$ might have negative elements on the diagonal (Eq. 5.10)

References

Manuel Arellano. Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics*, 49(4):431–434, 1987.

Christian B Hansen. Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics*, 141(2):597–620, 2007.

Teunis Kloek. Ols estimation in a model where a microvariable is explained by aggregates and contemporaneous disturbances are equicorrelated. *Econometrica: Journal of the Econometric Society*, pages 205–207, 1981.

Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Brent R Moulton. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The review of Economics and Statistics*, pages 334–338, 1990.

Andrew J Scott and D Holt. The effect of two-stage sampling on ordinary least squares methods. *Journal of the American statistical Association*, 77(380):848–854, 1982.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838, 1980.