

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 i th of the N individuals in the sample, j denotes the g th of the G clusters, assume that

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

$$\mathbb{E}[u_{ig}u_{jg'} | \mathbf{x}_{ig}, \mathbf{x}_{jg'}] = 0 \quad (5.1)$$

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

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

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

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

If the primary source of clustering is due to group-level common shocks, a useful approximation is that for the j th regressor, the default OLS variance estimate based on $s^2(\mathbf{X}'\mathbf{X})^{-1}$ should be inflated by $\tau_j \approx 1 + \rho_{x_j}\rho_u(\bar{N}_g - 1)$, 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
- \bar{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 $\mathbf{\Omega}_g = \mathbb{V}[\mathbf{u}_g | \mathbf{X}_g] = \mathbb{E}[\mathbf{u}_g \mathbf{u}_g' | \mathbf{X}_g]$, and there is a consistent estimate $\hat{\mathbf{\Omega}}_g$ of $\mathbf{\Omega}_g$, we can estimate $\mathbb{E}[\mathbf{X}_g' \mathbf{u}_g \mathbf{u}_g' \mathbf{X}_g] = \mathbb{E}[\mathbf{X}_g' \mathbf{\Omega}_g \mathbf{X}_g]$ via GLS.

Cluster-robust variance matrix estimate consider

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

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

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

as $G \rightarrow \infty$. A strand of literature popularizes this method:

- citetliang1986longitudinal: in a generalized estimatin equations setting

References

- 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.