A Practitioner's Guide to Cluster-Robust Inference

A. C. Cameron **D. L. Miller**

Presented by: Hasin Yousaf
Applied Reading Group

Example

DiD in State-Year Panel

$$y_{ts} = \gamma d_{ts} + z'_{ts}\delta + \alpha_s + \lambda_t + u_{ts}$$

Extent of the Problem (Cameron et al., 2008)

DGP with correlation within group and heteroskedastic errors

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + u_{ig} = \beta_0 + \beta_1(z_g + z_{ig}) + (\epsilon_g + \epsilon_{ig}) \quad (1)$$

with z_g, z_{ig}, and ϵ_g independent N(0, 1), and $\epsilon_{ig} \sim N(0, 9(z_g + z_{ig})^2)$

The DGP sets $\beta_0 = \beta_1 = 1$

Figure : Simulation Results for Rejection Rates of Test: $\beta_1 = 1$ (Cameron et al., 2008: Table 3)

Estimator #	Method	Number of Groups (G) 5	10	15	20	25	30
1	Assume i.i.d.	0.302 (0.015)	0.288 (0.014)	0.307 (0.015)	0.295 (0.014)	0.287 (0.014)	0.297 (0.014)
3	Cluster-robust	0.208 (0.013)	0.118 (0.010)	0.110 (0.010)	0.081 (0.009)	0.072 (0.008)	0.068 (0.008)

Simple Example

OLS with single, non-stochastic regressor

$$y_i = \beta x_i + u_i \qquad i = 1, ..., N \tag{2}$$

$$\hat{\beta} - \beta = \frac{\sum_i x_i u_i}{\sum_i x_i^2} \tag{3}$$

$$V[\hat{\beta}] = \frac{V[\sum_i x_i u_i]}{(\sum_i x_i^2)^2} = \frac{\sum_i \sum_j Cov(x_i u_i, x_j u_j)}{\sum_i x_i^2} \tag{4}$$

$$V[\hat{\beta}] = \frac{\sum_i x_i^2 V(u_i)}{\sum_i x_i^2} + \frac{\sum_i \sum_{j:i \neq j} x_i x_j Cov(u_i, u_j)}{\sum_i x_i^2} \tag{5}$$

Under no within-cluster correlation (equivalently N clusters): Second term in (5) is 0

Simple Example

Two Clusters with equi-correlated regressors and errors

$$V_{clu}[\hat{\beta}] = \underbrace{\frac{\sum_i x_i^2 V(u_i)}{\sum_i x_i^2}}_{V_{het}[\hat{\beta}]} + \frac{\sum_i \sum_{j:i\neq j} x_i x_j Cov(u_i, u_j) \mathbb{1}_{i,j\in C_k}}{\sum_i x_i^2} \quad (6)$$

Suppose $i = \{1, ..., N_1\} \in C_1$, and $i = \{N_1 + 1, ..., N\} \in C_2$

$$V[\hat{\beta}] = \frac{\sum_i x_i^2 V(u_i)}{\sum_i x_i^2} + \frac{\sum_{i\in C_1} \sum_{j:i\neq j \& j\in C_1} x_i x_j Cov(u_i, u_j)}{\sum_i x_i^2} + \frac{\sum_{i\in C_2} \sum_{j:i\neq j \& j\in C_2} x_i x_j Cov(u_i, u_j)}{\sum_i x_i^2} \quad (7)$$

Let $Corr(w_i, w_j) = \rho_w \quad if \quad i \in C_k \& j \in C_k \quad \forall \quad k$ where $z = \{u, x\}$,

$$V[\hat{\beta}] \approx [1 + \frac{N_1(N_1 - 1)}{N}\rho_x\rho_u + \frac{N_2(N_2 - 1)}{N}\rho_x\rho_u]V_{het} = \tau V_{het} \quad (8)$$

More Generally

Under equi-correlated regressors and errors

$$\tau_k \approx 1 + \rho_{x_k}\rho_u\left(\frac{V[N_g]}{\bar{N}_g} + \bar{N}_g - 1\right) \approx 1 + \rho_{x_k}\rho_u(\bar{N}_g - 1) \tag{9}$$

The result is exact if $\rho_{x_k} = 1$ for all regressors

The Variance inflation factor depends on:

1. Within-cluster correlation of the regressor
2. Within-cluster correlation of the error
3. The number of observations in each cluster

τ_k can be large despite small ρ_u. E.g., due to large group size (Moulton, 1990)

The Cluster-Robust Variance Matrix Estimate (CRVE)

$$y_{ig} = x'_{ig}\beta + u_{ig}, \qquad E[u_{ig}|x_{ig}] = 0 \tag{10}$$

$$E[u_{ig}u_{jg'}|x_{ig}, x_{jg'}] = 0 \text{ unless } g = g'$$

Stacking all observations in the g^{th} cluster,

$$y_g = x'_g\beta + u_g \tag{11}$$

$$\hat{\beta} = (X'X)^{-1}(X'Y) = (\sum_{g=1}^{G} X'_g X_g)^{-1}(\sum_{g=1}^{G} X'_g y_g) \tag{12}$$

$$V[\hat{\beta}] = (X'X)^{-1}[X'V[u|X]X](X'X)^{-1} = (X'X)^{-1}[B](X'X)^{-1} \tag{13}$$

$$B_{clu} = \sum_{g=1}^{G} X'_g E[u_g u'_g|X_g]X_g = \sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} x_{ig}x'_{jg}E[u_{ig}u_{jg}|X_g] \tag{14}$$

CRVE: Properties

$$B_{clu} = \sum_{g=1}^{G} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} x_{ig} x'_{jg} E[u_{ig} u_{jg} | X_g] \quad (15)$$

The $V[\hat{\beta}]$ is bigger when:

1. Regressors within cluster are correlated
2. Errors within cluster are correlated
3. N_g is large
4. The within-cluster regressor and error correlations are of the same sign (usual case)

CRVE: Estimation

$$V[\hat{\beta}] = (X'X)^{-1}[\sum_{g=1}^{G} X_g' \hat{u_g} \hat{u_g}' X_g](X'X)^{-1} = (X'X)^{-1}[\hat{B_{clu}}](X'X)^{-1} \tag{16}$$

Consistent as $G \to \infty$

Can be used also if $N_g \to \infty$ (Hansen, 2007)

Problem with "few" clusters

Finite-sample modifications: STATA uses $c\hat{u}_g$ instead of $\hat{u}_g$, where

$$c = \frac{G}{G-1}\frac{N-1}{N-K} \approx \frac{G}{G-1}$$

Does not fully eliminate downward bias – No Clear Best Correction (More on this next week)

Cluster-Bootstrap Variance Matrix Estimate

Do the following steps B times:

(1) form G clusters by resampling with replacement G times from original sample
(2) Compute the estimate of β, denote b^{th} bootstrap result as β_b. Then,

$$\hat{V}_{clu,boot}[\hat{\beta}] = \frac{1}{B-1}\sum_{b=1}^{B}(\beta_b - \bar{\hat{\beta}})(\beta_b - \bar{\hat{\beta}})'$$

Feasible GLS

Suppose $\Omega_g = E[u_g u_g' | X_g]$, then:

$$\hat{\beta}_{FGLS} = (\sum_{g=1}^{G} X_g' \hat{\Omega}_g^{-1} X_g)^{-1} (\sum_{g=1}^{G} X_g' \hat{\Omega}_g^{-1} y_g) \tag{17}$$

where $\hat{\Omega}$ is a consistent estimator of Ω.

$$\hat{V}_{clu}[\hat{\beta}_{FGLS}] = (X'\hat{\Omega}^{-1}X)^{-1} [\sum_{g=1}^{G} X_g' \hat{\Omega}_g^{-1} \hat{u_g} \hat{u_g}' \hat{\Omega}_g^{-1} X_g] (X'\hat{\Omega}^{-1}X)^{-1} \tag{18}$$

It requires: $Corr(u_g, u_h) = 0 \quad$ if $g \neq h$, and $\quad G \rightarrow \infty$

Feasible GLS: Efficiency Gain

Efficiency gain may be modest

For equi-correlated errors, balanced clusters and $x_{ig} = x_g$, FGLS $=$ OLS

No clear guide when FGLS leads to considerable improvements in efficiency

Cluster-Specifc Fixed Effects (CSFE)

$$y_{ig} = x_{ig}'\beta + \alpha_g + u_{ig} = x_{ig}'\beta + \sum_{h=1}^{G} \alpha_g dh_{ig} + u_{ig} \qquad (19)$$

It can be estimated by:

1. Least Squares Dummy Variable (LSDV)
2. Fixed Effects (Within) Estimator

Consistent if $G \to \infty$ or $N_g \to \infty$

Controls for limited form of endogeneity of the regressors

CSFE: Practicalities I

Does not fully control for within-cluster error correlation, so CRVE should still be used

CRVE valid provided $G \to \infty$ and N_g is small (Arellano, 1987)

CRVE can also be used when $N_g \to \infty$ (Hansen, 2007)

If cluster-robust SE are used and cluster sizes are small, then inference should be based on within estimator SE

Within estimator leads to correct finite-sample correction

CSFE: Practicalities II

The need for cluster-specific FE can be conducted by Hausman Test

$$T_{Haus} = (\hat{\beta}_{1;FE} - \hat{\beta}_{1;RE})\hat{V}^{-1}(\hat{\beta}_{1;FE} - \hat{\beta}_{1;RE})'$$

Since above requires strong assumptions (β_{RE} fully efficient under null),

Wooldridge (2010) proposes implementing:

$$y_{ig} = x'_{ig}\beta + \bar{w}'_g\gamma + u_{ig}$$

where $\bar{w}_g = N_g^{-1}\sum_{i=1}^{G} w_{ig}$

What to Cluster Over

Bias-Variance Trade-off: Larger Clusters have less bias but higher variability

1. Broadest clustering level at which errors could be correlated within cluster. Alternatively, no need to cluster over group within which errors are potentially uncorrelated

2. Better estimation of $\hat{V}_{clu}[\hat{\beta}]$ is achieved when G gets large. Cluster size should not be very large so that we do not have "few" clusters

What to Cluster Over II

No formal test to address this trade-off

Consensus is to be conservative and avoid bias

In practice, can use progressively higher levels and stop when little change in SE

For samples, one should at least cluster at PSU; and WLS with sampling weights

Clustering is Not Always Necessary

Example

Key regressor is randomly assigned within cluster