Clustered Standard Errors

Gov 2001: Quantitative Social Science Methods I

Week 12, Lecture 24

Spring 2026

For Today

Required Reading

- Aronow & Miller, §3.5 (pp. 135–141)
- Angrist & Pischke, §8.2 (pp. 221–232)

Today: Inference when observations are correlated within groups.

Roadmap

Part I: Why Clustering Matters

The Basic Issue

Standard OLS assumes: Errors are independent across observations.

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for } i \neq j$$

Reality: Often errors are correlated within groups.

- Students in the same classroom
- Voters in the same county
- Workers at the same firm
- Observations of the same person over time

Why Are Errors Correlated Within Groups?

Common shocks:

- Students in same class have the same teacher
- Workers at same firm face same management
- Voters in same county see same local news

Unobserved group-level factors:
Anything affecting the whole group that isn't in your model creates correlated errors.

Group members are more similar to each other than to outsiders.

Example: Class Size and Test Scores

Data: Students nested in classrooms.

$$\text{Score}_{ic} = \beta_0 + \beta_1 \text{ClassSize}_c + \varepsilon_{ic}$$

Student i in classroom c.

Issue:

- Teacher quality varies across classrooms
- Peer effects within classrooms
- Students in same class share these unobserved factors

$\Longrightarrow$ Errors ε_{ic} are correlated within classrooms.

Part II: The Clustering Problem

The "Effective Sample Size" Problem

If errors are correlated within groups:

Your sample is less informative than it appears.

- 1000 students in 50 classrooms $\neq$ 1000 independent observations
- Information comes from variation *between* classrooms
- Within-classroom variation is correlated

Extreme case: If ε_{ic} is identical within classrooms, you effectively have 50 observations, not 1000.

What Happens If You Ignore Clustering?

OLS point estimates: Still unbiased and consistent.

Standard errors: Typically **too small**.

- They assume n independent observations
- But effective sample size is smaller

Consequences:

- t-statistics too large
- p-values too small
- Confidence intervals too narrow
- Over-rejection of true null hypotheses

How Bad Can It Be?

Moulton (1990): Showed severe under-coverage of standard confidence intervals.

Example:

- 50 clusters, 20 observations each ($n = 1000$)
- Within-cluster correlation $\rho = 0.5$
- True rejection rate: 5%
- Actual rejection rate (ignoring clustering): $\approx$ 50%!

You reject the null 10 times more often than you should.

Part III: Cluster-Robust Standard Errors

The Solution: Cluster-Robust SEs

Idea: Allow for arbitrary correlation within clusters.

Assume:

- $\text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0$ if i, j in same cluster
- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ if i, j in different clusters

Don't restrict the correlation structure *within* clusters.
This is a generalization of robust (heteroskedasticity-consistent) SEs.

Cluster-Robust Variance Formula

$$\widehat{\text{Var}}_{CR}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g' \hat{\boldsymbol{\varepsilon}}_g \hat{\boldsymbol{\varepsilon}}_g' \mathbf{X}_g \right) (\mathbf{X}'\mathbf{X})^{-1}$$

where:

- G = number of clusters
- $\mathbf{X}_g$ = rows of $\mathbf{X}$ for cluster g
- $\hat{\boldsymbol{\varepsilon}}_g$ = residuals for cluster g

Key: The "meat" of the sandwich sums over *clusters*, not individual observations.

Intuition for Cluster-Robust SEs

Robust SEs: Use $\hat{e}_i^2$ to estimate $\text{Var}(\varepsilon_i)$.

Cluster-robust SEs: Use $\hat{\boldsymbol{\varepsilon}}_g \hat{\boldsymbol{\varepsilon}}_g'$ to estimate the full variance-covariance matrix within cluster g.

This captures:

- Heteroskedasticity across observations
- Correlation within clusters

Requirements for Validity

Cluster-robust SEs are consistent when:

1. Number of clusters $G \to \infty$
2. Clusters are independent of each other
3. (Some technical regularity conditions)

Warning: If G is small (say, < 50), cluster-robust SEs can be unreliable.

- Tend to under-estimate true variance
- Wild cluster bootstrap may help

Part IV: When and At What Level to Cluster

When to Cluster

Abadie et al. (2017) distinguish two reasons:

1. Sampling design:
If you sampled *clusters* (not individuals), cluster at that level.

- Sampled schools, then students within schools $\Rightarrow$ cluster by school

2. Treatment assignment:
Cluster at the level of treatment assignment.

- Treatment assigned at classroom level $\Rightarrow$ cluster by classroom

At What Level to Cluster?

General principle:
Cluster at the level where you think errors are correlated.

When in doubt, cluster more broadly:

- Clustering more narrowly can under-state SEs
- Clustering more broadly is conservative
- But need enough clusters for validity

Example:
Students in classrooms in schools in districts.
If unsure, cluster at school level rather than classroom level.

Clustering Examples

Setting	**Cluster Level**
Students in classrooms	Classroom
Voters in precincts	Precinct
Employees at firms	Firm
States over time (panel)	State
Countries over time (panel)	Country

Key question: At what level do unobserved factors create correlation?

Fixed Effects vs. Clustering

Fixed effects: Control for cluster-level confounders.
Add α_g for each cluster g.

Clustered SEs: Account for within-cluster error correlation.

These are different!

- Fixed effects remove bias from cluster-level confounders
- Clustering handles inference (standard errors)
- Often you need **both**

Including fixed effects does NOT eliminate the need to cluster.

Two-Way Clustering

Sometimes errors might be correlated along multiple dimensions.

Example: Trade flows between country pairs over time.

- Correlation within exporter-year
- Correlation within importer-year

Two-way clustering:

$$\widehat{\text{Var}}_{2way} = \widehat{\text{Var}}_{cluster1} + \widehat{\text{Var}}_{cluster2} - \widehat{\text{Var}}_{intersection}$$

(Advanced topic—we won't go deep here.)

Part V: Practical Guidance

Practical Decision Guide

1. **Are observations grouped in some way?**
   - No $\Rightarrow$ Robust SEs (no clustering)
   - Yes $\Rightarrow$ Continue
2. **Could errors be correlated within groups?**
   - Probably yes $\Rightarrow$ Cluster
3. **How many clusters?**
   - $G \geq 50$: Cluster-robust SEs should work well
   - $20 \leq G < 50$: Use caution; consider wild bootstrap
   - $G < 20$: Cluster-robust may be unreliable; be very cautious

Implementing in Software

R:

- `lm_robust(y ~ x, clusters = group, data = df)`
- `coeftest(model, vcov = vcovCL(model, cluster = ~group))`

Stata:

- `reg y x, cluster(group)`
- `reg y x, vce(cluster group)`

Python:

- statsmodels: `model.fit(cov_type='cluster', cov_kwds={'groups': df['group']})`

Summary

Key points:

1. Errors often correlated within groups (clusters)
2. Ignoring clustering $\Rightarrow$ SEs too small $\Rightarrow$ over-rejection
3. **Cluster-robust SEs** allow for within-cluster correlation
4. Cluster at the level of:
   - Sampling design, or
   - Treatment assignment
5. Need enough clusters (roughly $G \geq 50$) for reliability

Looking Ahead

Next week: Advanced Topics

- Variance weights in OLS
- Heterogeneous treatment effects
- Regression adjustment under unconfoundedness
- What OLS estimates when effects vary

When observations are grouped,
errors are often correlated within groups.

Ignoring this makes standard errors too small.

Cluster at the level of sampling or treatment assignment.