1. Clustered errors in OLS.
Consider estimation of the linear regression model

\[ y_i = x_i' \beta + u_i, \quad i = 1, ..., n, \]

where \( x_i \) is a \( k \times 1 \) vector of nonstochastic regressors, \( \beta \) is a \( k \times 1 \) parameter vector, and \( u_i \) is an error term. In matrix notation

\[ y = X \beta + u. \]

We consider individual cross-section data with an error term that includes both an individual-specific component, and a state-specific component common to all individuals from the same state, such as California. If these two error components are each i.i.d. with zero mean it follows

\[ E[u_i] = 0 \]
\[ \text{Cov}[u_i, u_j] = \begin{cases} 
\sigma^2 & i = j \\
\rho \sigma^2 & i \neq j \text{ and } i \text{ and } j \text{ in same state} \\
0 & i \neq j \text{ and } i \text{ and } j \text{ not in same state},
\end{cases} \]

where \( \rho \) is the cross-state correlation (more generally called the intraclass correlation).

Suppose there are \( p \) states. Define the \( n \times p \) matrix \( Z \) to be one of \((0, 1)\) indicators for membership in one of the \( p \) states. Then the \( ij^{th} \) element of \( ZZ' \) equals 1 if \( i \) and \( j \) in the same state and equals 0 otherwise.

Use the above information on \( u \) in answering the following.
(a) Verify that \( E[u_i u_j] = \sigma^2 (1 - \rho) I_n + \rho ZZ' \).
[i.e. verify this gives the same \( \text{Cov}[u_i, u_j] \) as above for \( i = j \) and for the two cases when \( i \neq j \)].

(b) Obtain the mean and variance of the OLS estimator \( \hat{\beta} = (X'X)^{-1}X'y \).

(c) State how to obtain a consistent estimate of \( \text{V}[\hat{\beta}] \) that does not depend on unknown parameters.

(d) Suppose all regressors are fixed within each of the \( p \) states, i.e. \( x_i = x_i^* \) for all \( i \) in state 1, ...., \( x_i = x_p^* \) for all \( i \) in state \( p \). Furthermore suppose that all \( p \) states have exactly the same number of observations \( m \) (so \( n = mp \)).

Then defining \( e_m \) to be an \( m \times 1 \) vector of ones and sorting by state

\[
X = \begin{bmatrix} x_1' \otimes e_m \\ \vdots \\ x_p' \otimes e_m \end{bmatrix} = \begin{bmatrix} x_1' \\ \vdots \\ x_p' \end{bmatrix} \otimes e_m = X^* \otimes e_m \text{ defining } p \times k \text{ matrix } X^*.
\]

and

\[
ZZ' = \begin{bmatrix} e_m e_m' & 0 & \cdots & 0 \\ 0 & e_m e_m' & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & e_m e_m' \end{bmatrix} = I_p \otimes e_m e_m'.
\]
Hence show that $X'ZZ'X(X'X)^{-1} = mI_k$ where $k$ is the number of regressors.
(Hint $(A \otimes B)' = A' \otimes B'$ and Then it can be shown that $X'ZZ'X(X'X)^{-1} = mI_k$.

(e) Use the result in part (b) to simplify your result in part (b) for $V[\hat{\beta}]$.

(f) Suppose $\rho = 0.02$ and $m = 200$ (so with 50 states there are 10,000 observations). What will be the extent of the error in reported standard errors and $t$—statistics based on the usual $\sigma^2(X'X)^{-1}$?

(g) What do you conclude about the validity of using standard OLS results if there is an unobserved state-specific effect in the error term?

2. Basic clustering for OLS

This question uses dataset mus09vietnam_ex2.dta from chapter 9 of Cameron and Trivedi (2009). The data are analyzed using program ass5s16.do

We focus on OLS regression of $\text{pharvis}$ on $\lnhhexp$ and $\text{illness}$

$\text{pharvis}$: number of direct pharmacy visits

$\lnhhexp$: log of household medical expenses

$\text{illness}$: number of illnesses

The data are individual level data with individuals clustered in household clustered in communes (a village).

(a) How many communes are there and what is the range of the number of individuals per commune?

(b) Are the regressors $\lnhhexp$ and $\text{illness}$ correlated within commune? Explain.

(c) Give the intracluster correlation (within commune) of $\lnhhexp$ and of the residual from OLS regression of $\text{pharvis}$ on $\lnhhexp$ and $\text{illness}$.

(d) Hence calculate how much larger you expect cluster-robust (cluster on commune) standard error to be than the default standard error for variable $\lnhhexp$ in this regression.

(e) Compare your answer in part (d) to that actually obtained by directly comparing the cluster-robust standard error to the default standard error.

(f) Do the cluster bootstrap standard errors differ much from the cluster-robust standard errors? Are you surprised?

3. Feasible GLS, FE and mixed estimators

Continue with the same data as previous example. Answer the following questions given the computer output.

(a) Should we use default standard errors for the fixed effects estimator?

(b) Should we use default standard errors for the random effects estimator?

(c) Has random effects estimation led to noticeable improvement in efficiency compared to OLS?

(d) Compare the first mixed model estimates with random effects model estimated using MLE (which additionally assumes normally distributed errors).

(e) Does the coefficient of $\text{illness}$ appear to vary across communes?
4. **Twoway cluster robust**

Continue with the same data as previous example. Here we cluster on both `commune` and `illdays` (the latter is used for illustrative purposes).

(a) Has additionally clustering on `illdays` changed the standard errors substantially? If so, how.

(b) Explain how `cgmreg.ado` computes the 2-way cluster-robust standard errors.

(c) Command `ivreg2` uses a slightly different way of combining the three components. Has it led to substantially different standard errors?

(d) In fact `pharvis` is a count model. Estimate using command `poisson pharvis lnhhexp illness` and obtain two-way cluster robust standard errors. You will need to do some coding.

5. **Bootstrap**

Now we use a reduced dataset with only ten communes. All bootstraps are cluster bootstraps with clustering on industry.

(a) Is there much difference between bootstrap confidence intervals obtained using the percentile method and those obtained using +/- 1.96 times the bootstrap standard error?

(b) Is there much difference between bootstrap confidence intervals obtained using BC and BCa methods (which are intended to provide an asymptotic refinement) and those obtained using the percentile method?

(c) Is there much difference between bootstrap confidence intervals obtained using the percentile-t method compared to those obtained using the percentile method?