Clustered variance estimates and fixed effects

Christopher A. Sims
Princeton University
sims@princeton.edu

February 22, 2021

Outline

These notes expand the discussion of fixed effects and clustering in the "Clustering, random effects, mixed models, sandwiches" notes.

- Clustering as an alternative to fixed effects
- Clustering combined with fixed effects
  - Coefficients of variables constant across groups get asymptotically almost correct, but biased, standard errors.
  - Clustered standard errors for coefficients that vary across groups are not available.

Clustering and fixed effects as alternatives

- If in our standard notation $y_{ig} = X_{ig}\beta + \nu_g + \varepsilon_{it}$ we are willing to assume

$$E[X'_{\cdot g}\varepsilon_{\cdot g}] = 0\,, \qquad (*)$$

  then we can treat the group shift variables ν_g as part of the error term.

- As we have already discussed, if we go further and assume $E[\nu_g \mid X_{\cdot\cdot}] = 0$, we can justify attempting to model the distributions of ε and ν apply GLS.

- But with the weaker assumption $(*)$, we can estimate with OLS and use a clustered covariance matrix for the estimates.

- Fixed effects estimation weakens the assumptions further, allowing correlation between ν_g and $X_{\cdot g}$

How to choose?

- All of the results on distribution for these estimates, other than Bayesian likelihood-based parameterized GLS, are asymptotic.
- OLS with group-clustered standard errors requires weaker assumptions than GLS, and fixed effects requires still weaker assumptions; but as we weaken the assumptions, the reasonableness of assuming the asymptotic theory applies in any given sample also weakens.
- The result is that the choice among these estimators is a matter of a priori judgment, depending on the substance of the application. The Bayesian response to this is that if we are going to be making a priori judgments based on substantive knowledge, it is better to make this explicit with a prior whenever that is possible.

Fixed effects and clustering on the same model

- Suppose we have some variables that have different coefficients in different groups, and others whose coefficients are constant across groups:

$$y_{ig} = Z_{ig}\gamma_g + X_{ig}\beta + \varepsilon_{ig} = \ . \qquad (1)$$

  A special case is where Z_{ig} is just a dummy variable for group g, constant with the group at 1 and 0 outside the group, in which case this becomes the standard fixed-effects model.

- If we estimate by OLS, then as $M \to \infty$ with group size n_g bounded, OLS estimates of β will be consistent, while those for γ_g, while unbiased, will not be consistent.

- If we naively apply the sandwich estimator for the variances of the OLS matrix, clustering by group, we will not get consistent estimators of the variances of the γ_g estimates.

The problem with the clustered covariance matrix when there are Z variables

The central factor in the sandwich covariance estimator is

$$\sum_{g=1}^{M} \begin{bmatrix} Z'_{\cdot g} \\ X'_{\cdot g} \end{bmatrix} \hat{\varepsilon}_{\cdot g} \varepsilon'_{\cdot g} \begin{bmatrix} Z_{\cdot g} & X_{\cdot g} \end{bmatrix} , \tag{2}$$

where $\hat{\varepsilon}_{\cdot g}$ is the vector of least squares residuals for group g.

Least squares residuals are algebraically orthogonal to all the right-hand-side variables in the regression. That means in this case that $Z'_{\cdot g}\hat{\varepsilon}_{\cdot g}$ is a $k \times 1$ vector of zeros, which in turn means that the first Mk rows and first Mk columns of the central matrix in the clustered sandwich covariance estimate are all 0. That makes the overall clustered covariance matrix singular, in all sample sizes, and hence not consistent.

How bad is that?

- The naively estimated clustered covariance matrix is singular, but all its elements are in general non-zero, because of the two side matrices in the $(X'X)^{-1}X'uu'X(X'X)^{-1}$ form. So there is no easily eyeballed problem with the estimate.

- However, only the entries in the covariance matrix corresponding to γ_g coefficients are unusable. The elements of the estimated covariance matrix corresponding to β parameters are almost correct.

Almost correct?

- If $W = [Z, X]$, the lower right block of the $(W'W)^{-1}$ matrix corresponding to the X variables is the inverse of the covariance matrix of the residuals $\tilde{X}$ from a regression of the $X's$ on the Z's. (This is an algebra fact that I won't try to derive here.)

- The $k \times k$ block of the central matrix in the clustered sandwich corresponding to the X's has expectation conditional on W, in the special case of $\varepsilon_{\cdot g} \sim N(0, \sigma_g^2 I)$, $\tilde{X}'\tilde{X}\sigma^2(1 - k/n_g)$

- This means that the estimated covariance matrix for β is in this special case, and assuming also constant n, downward biased by a factor $1 - k/n$

- If n does not grow, but M does, then this bias does not become small asymptotically, and if n is small, the bias can be substantial.

- It might be worth correcting for the bias, but we have a good estimate of the bias only for this special case where use of the sandwich is unnecessary in the first place.

Bootstrapping clustered regression

- If the numbeer of groups is large (as it has to be to justify the sandwich), one can get an estimated covariance matrix for the coefficients by a bootstrap that samples the groups, while always including all observations within a group when the group is in the bootstrap sample.

- However the appeal of OLS plus clustered standard errors is that they require few assumptions (other than the major assumption that the asymptotic theory applies to this sample) and are easy to compute.

- Bootstrapping is substantial computational work if the sample is large. Before taking this approach, one might consider approaches using MCMC on the likelihood or posterior, using a mixed model and/or modeling of time series correlation of residuals and/or modeling heteroskedasticity and/or including group means in the regression.

Reference

- Cameron and Miller (2015) covers many cases, including two-way clustering and a discussion of which commands in which packages give the "correct" bias-correction when combining fixed effects with clustered covariances for the parameters.

- They recommend using the $1 - k/n$ correction in the case we discussed above, without claiming that this generally eliminates bias.

- They do not discuss Bayesian approaches.

*

References

Cameron, A. C. and D. L. Miller (2015): "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50, 317–373.