A note on computation

Joe Ritter

Carter, Schnepel and Steigerwald (2017) define the approximate effective number of clusters for a linear combination of coefficients by

$$G^{*A} = \frac{G}{1 + \Gamma^A} \quad \text{where} \quad \Gamma^A = \frac{\widehat{\text{Var}}(\gamma_g)}{\bar{\gamma}_g^2},$$

$$\gamma_g = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}_g'\mathbf{\Lambda}\mathbf{X}_g)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a},$$

where $\mathbf{X}_g$ contains the data for cluster g, $\mathbf{\Lambda}$ is the assumed variance matrix of the errors, and $\mathbf{a}$ is a selection vector. This package enables only selection vectors for single coefficients (e.g., $[0, 0, 1, 0, 0, 0]'$).

Carter et al. suggest setting $\mathbf{\Lambda}$ to a matrix of ones, which implies, conservatively, that errors within a cluster are perfectly correlated. However, this choice is a problem if the regression includes cluster fixed effects: either $\mathbf{X}$ is the outcome of a within transformation or explicitly includes columns of fixed effects. In the former case, the problem is easy to see. The means of $\mathbf{X}_g$ should be zero, but, *computationally,* they are tiny nonzero numbers. Therefore, $\mathbf{X}_g'\mathbf{\Lambda}$ should be zero, but, computationally, it is filled with tiny nonzero numbers. This ultimately results in a tiny, computationally unstable value for $\bar{\gamma}^2$. Not surprisingly, when cluster fixed effects are columns of $\mathbf{X}$ the same problem crops up. Therefore, `effClust` sets the off-diagonal elements of $\mathbf{\Lambda}$ to 0.999 by default, which makes no appreciable difference in other cases.

When clusters are very large, the $\mathbf{\Lambda}$ matrix is huge, so direct implementation of the matrix formula can be a problem. The package therefore follows the approach outlined in MacKinnon, Nielson, and Webb (2022), equation (38). However, it does not force $\rho = 0$ when there are cluster fixed effects as recommended there. MacKinnon et al. recommend $\rho = 0$ when there are cluster fixed effects on the grounds that the fixed effects typically remove all of the intra-cluster correlation in the equi-correlated case. But since equi-correlation in $\mathbf{\Lambda}$ is just a convenience, not a maintained hypothesis, judgement about the value of ρ is left to the user.

References

Andrew V. Carter, Kevin T. Schnepel, and Douglas G. Steigerwald, "Asymptotic Behavior of a t-test Robust to Cluster Heterogeneity," *The Review of Economics and Statistics,* October 2017, 99(4). `https://doi.org/10.1162/REST_a_00639`.

James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb, "Leverage, Influence, and the Jackknife in Clustered Regression Models: Reliable Inference Using summclust," QED Working Paper 1483, Queen's University (2022). https://www.econ.queensu.ca/research/working-papers/1483.