Referee Report for "When should you adjust standard errors for clustering"

Zihan Zhang

Research Question

In which cases we should apply cluster standard errors? And how do we better use cluster standard errors?

Goals of the paper

This paper addresses a key question in empirical research, when and how do we cluster standard errors in regressions. It proposes two new estimators based on analytical derivation and bootstrap method to help empirical economists to identify the causality with regressions more precisely.

Formally, we use clustered standard errors for allowing some correlation on observations within clusters, which breaks the classical assumption of i.i.d. However, in applications, there are still several questions that do not have a clear answer. For example, why do we adjust standard errors on some clusters but not others? Is the cluster variance estimator valid when we have large fraction of clusters in the population? And in what settings does the choice of whether and how to cluster make a difference. Answering those questions with a conventional framework typically requires researchers have an inference on the data generation process and the structures of errors. In this paper, they proposed a new framework based on sampling mechanism and had advantages of having no need to assume the error component structure.

The theory part of this paper demonstrates that robust variance estimator typically underestimates the variance, and cluster variance estimator tends to be super conservative. Exceptions are when the expected fraction of clusters in the sample is small, or when there are almost no heterogeneous average treatment effect across clusters.

Key insights: When should we use cluster standard errors? And What are the performances of two estimators proposed by this paper (CCV and TSCB), comparing to the conventional clustered estimator?

- When units are randomly sampled from a large population and treatment is randomized at the individual level, there is no need to use cluster standard errors, even if there is within-cluster correlation in outcomes. Clustering can result in extremely conservative standard errors, leading to super wide confidence intervals.

- When treatment is assigned at the cluster level and the sample is randomly drawn from or includes the entire population.

    - When the assignment is fully clustered, in other words, all individuals within a cluster receive the same treatment. Conventional clustered standard errors is good enough.

    - When the assignment is partially clustered, in other words, there is treatment variation within clusters, some individuals receive treatment while others do not. Conventional clustered standard error is conservative. CCV and TSCB can substantially reduce standard errors and more close to the (hypothetical) true values.

- When samples are draw randomly from the population and then sampling individuals from those clusters (cluster sampling). We should apply clustered standard errors. If the fraction of sampled clusters is small, conventional cluster standard errors work well. However, if cluster sizes are large and there s treatment variation within clusters, CCB and TSCB should be applied. They can reduce estimated standard errors.

- In general. We have no need for cluster standard errors when treatment is randomized at the individual level. However, when treatment is assigned at cluster level, or there is a design of cluster sampling, cluster standard errors should be applied. And, CCV and TSVB standard errors can provide better accuracy in most cases.