

---

Subject: Re: Use of svyset when standard errors are already clustered

Posted by [Reduced-For\(u\)m](#) on Sat, 12 Aug 2017 22:32:10 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

I think I see what both of you are saying:

For Tom (DHS) the nature of the data collection/sampling is such that the only "built in" issues of correlated standard errors (from a sampling standpoint) is in the PSU, and hence the DHS suggests (tightly) "clustering" in some way on PSU.

For Ale, the question isn't about data-collection design but about econometric/statistical design, where in month-year of birth probably determines some "treatment assignment" and then the whole "correlated treatment assignment" issue arises... if "treatment" is assigned by the cohort, any within-cohort correlation in Eps gets blown up by the serial-correlation within-group in T (whatever the treatment assignment variable of interest is). This is a version of the problem in Bertrand et al "how much should we trust diff in diff?" paper.

Clustering on one won't fix the other, because they are not nested (there is not a "higher" group to cluster on that nests the other groups. An obvious way to deal with both concerns would be to use the Cameron, Gelbach and Miller approach to multi-way clustering. Then you could deal with the spatial auto-correlation in the error term the survey design generates, and the cohort level auto-correlation in X that the identification strategy generates (and risks blowing up any within-cohort correlations in the error term).

I'd also advise you to google about birth date measurement in the DHS (it isn't as good as we'd like) and consider that this may (I have no idea) pose a threat to your identification strategy.

---