Subject: Re: Weighting district-level data
Posted by Reduced-For(u)m on Mon, 28 Dec 2015 22:53:23 GMT
View Forum Message <> Reply to Message

Hi - sorry for the holiday-related delay. Some responses:

1. I think technically you would have to bootstrap the whole procedure (first and second stages) but, given that that is a little bit overboard, you should definitely cluster in the second stage.

2. I haven't math'd it out, but I think that if the weights cancel, you can just ignore them. If it doesn't make a difference, it doesn't make a difference.

3. You cluster regardless of your weighting choice - weighting is about getting a representative sample, clustering is about getting consistent standard error/p-value estimates.

4. You don't have to use "svyset", it is just one way to deal with it. You can just "reg Y X, cluster(district)" - or you can add your weights manually too (I also don't usually use svyset - only when trying to replicate DHS methods).

5. Your p-values will probably be a little too small if you don't bootstrap both stages, but it is probably not a huge deal. People ignore that all the time. You are just pretending that there is no measurement error in your observations (that is - pretending that your district means are actual observations and not estimates of district means). I mean - it is basically up to your field to decide what they will/won't accept, and I think 99% of smart researchers would not even consider this an issue - just me and some other sticklers.

Good luck - I think don't sweat it too much. But FYI - the bootstrap wouldn't be so hard really. You just place what you already do in a loop, use the "bsample" command (if Stata) at the beginning of the loop, and save the point estimate from each repetition. The standard deviation of your point estimates is then used as the standard error of your estimate (I mean, it is an estimate of your standard error, but that is all anyone ever has).