Subject: Re: Weighting district-level data
Posted by Reduced-For(u)m on Fri, 18 Dec 2015 20:37:46 GMT
View Forum Message <> Reply to Message

Interesting. Quick replies:

1 - cool. My guess is that since it is usually 1 PSU going to 1 district, and all observations in some PSU have the same weight, that it is basically the numerator/denominator canceling each other out. But while we are on that - since it is a fraction, and since that fraction is an estimate of the population proportion, fewer observations means a less good estimate of the actual proportion. Clustering might not work here at all since you would lose the uncertainty generated in your first stage - you might have to bootstrap both stages. I'll leave that to you to decide how far you want to go down that rabbit-hole, but you may want to account for the uncertainty in your "observations" because those are estimates themselves.

2 - it is nice to help (smiley face)

3 - diff district_indicator, t (treated) p (t), cluster(District) ... I think the problem is the extra "," after p(t). Delete that, and I think it will run.

4 -I see what you are doing. That is interesting, and I could see how it would work. But you should also know that you are not necessarily comparing "apples to apples" anymore. Suppose PSU 1 is in District 1 in round 1. Then, in Round 2, District 1 contains PSU 397 (which wasn't sampled in round 1). Then you are using different people from different towns/areas to define the same "district" variable. How many PSUs per district do you have, on average? I ask because with many, I'd think a law of large numbers might apply and you'd be fine. But with only 1 or 2 PSUs per round, much of the difference across time within district is going to be do to sampling variation and not do to real changes. Again, in theory, with many N(obs), G(groups) and T(periods) you are OK, but in finite numbers you are asking a whole lot of the data.