

---

Subject: Re: Weighting district-level data

Posted by [amira.elshal.1@city.ac.uk](mailto:amira.elshal.1@city.ac.uk) on Thu, 24 Dec 2015 09:00:37 GMT

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

---

Dear Sir,

This is a kind reminder of my message below. I apologize for disturbing you again and I would appreciate your tolerance as I have a couple of more questions:

1- Do I still have to weigh the data? If I decide to follow Gary Solon in "What are We Weighting For" and argue that, with causal effects, I don't need to weight given that I carry out my analysis on district-level with district-level covariates (population size is one of the covariates I include), shall I cluster afterwards in the second stage (regression)? You have mentioned that "Clustering might not work here at all since you would lose the uncertainty generated in your first stage".

2- If I decide to weigh my data, how can I do so given that the numerator and the denominator cancel each other out in the first stage of calculating the percent of women in a district using modern methods?

3- Shall I use clustering in the second stage (regression) if I weighed my data in the first stage? Is clustering (in the second stage) related to my decision to weigh/NOT weigh data (in the first stage)?

4- If I have to cluster, shall I directly use `reg Y X, cluster(district)`? Or shall I first define PSU in the "svyset" command? I have not used the "svyset" command before.

5- I do not think I would be able to bootstrap the first stage by myself. Also, I think this will consume a lot of time. Do you think this is a necessary step in my case? To what extent will NOT bootstrapping affect the results given that I bootstrap standard errors in the second stage (regression)?

Once again, I would like to thank you for your guidance.

Kind regards,  
Amira

---