Hello

I am doing my first paper as PhD student and I am using panel data for Egypt - KR file for 2005, 2008, and 2014.
I append them in one file and I extract all the childern above two years old.
I am testing the realtionship between maternal employment and child health (stunting, wasting, underweight, and overweight).

The Analysis Approach: 1- OLS 2- IV-Instrument variable.

I usually start my analysis by add svy: in the begining, which is conducted by:
**********************svy*************
gen strata=v023
gen psu=v021
gen sampwt=v005/1000000
svyset psu, strata(strata) weight(sampwt) vce(linearized) singleunit(centered)
.

.
.***************************STATA command********************
.***************Analysis***************
****OLS*****
svy: regress stunted motheremployed child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates))
svy: regress wasted motheremployed child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates))
svy: regress overweight motheremployed child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates))
svy: regress underweight motheremployed child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates))

*****IV******
******IV-unweighting****
ivregress 2sls stunted   child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates)) (motheremployed= instrument), robust first
.

.
*******IV- wieghting*******
.
ivregress 2sls stunted child characteristics mother characteristics partner characteristics Household characteristics (fixed effect (year) (governorates)) (motheremployed= instrument) [pweight = sampwt], vce(cluster v021) first

.... *************************************

My question here is relted to the result that I got from in IV approach, the result are differ between weighting and unweighting IV appreoach

IV- weighting
the variable mother employment 0/1 - is significat at 5% in(wasted- in IV weighted regresion)
IV- not weighting
the mother employment 0/1 - is significat at 5% in(stunted -wasted-overweight- in IV unweighted regresion)

So I am wondering that
1- why the result is differ between weighting and unweighting.
and 2- In my paper can I include the unweighting regression or it yields unseginificant result.

Thank you and I appreciate your guidance.

Subject: Re: Weighting in IV approach
Posted by Bridgette-DHS on Tue, 28 Mar 2023 13:19:10 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

We recommend always using weights, and also the adjustments for clustering and stratification, which have been discussed in many forum posts. The estimates change slightly when you use weights, and the standard errors change slightly when you adjust for clustering and stratification. The weights tend to increase the standard errors and occasionally a coefficient will be significant when the weights are not used and non-significant when the weights are used. When this happens it is especially important to make the adjustments for clustering and stratification (as you are doing with svyset).

I hesitate to describe successive surveys in the same country as panels. The surveys are independent cross-sections. The individuals, households, and clusters are different in every survey.

You are on your own when it comes to including instrumental variables in your analysis. I hope you are confident in making the required assumptions. I am not familiar with "ivregress" but if it does not allow svy then I'd be very cautious in the interpretation of borderline significance. "Cautious" means interpreting a borderline value as "not significant", even if it's disappointing.

Subject: Re: Weighting in IV approach
Posted by Alanood on Tue, 28 Mar 2023 21:35:40 GMT
View Forum Message <> Reply to Message

Thank you for your reply.

could you please clarify more " I'd be very cautious in the interpretation of borderline significance. "Cautious" means interpreting a borderline value as "not significant", even if it's disappointing." ?

Regarding the reason of using panel, it is because i want to test the imapct of maternal employment and child health in multiple years, and include years fixed effect in the regression, Even though there are not the same group of people interviewd, its still interesting to test for it.

I have no issue with including "svy" with OLS, but when i used ivreg "svy" not allowed.
I tried to include:

gen strata=v023
gen psu=v021
gen sampwt=v005/1000000

[pweight = sampwt], strata vce(cluster psu)
.
however STATA does not allow "ivreg" with strata (v023)!!
I only can include weight and psu.
[pweight = sampwt], vce(cluster psu)

Any way regarding the weighting, I found two views: the svy (weight+psu+strata),
1- Some mentioned that If we have the svy information in the dataset, we should always use it.
"To avoid any bias in the coefficients from the over-sampled population and adjust the standard errors for the survey clustering effect (Deaton, 1997), all the regression analyses have corrected for the survey design, i.e. the sampling weight, the cluster, and the strata were all taken into account."

2-The other view as found in this reply "
https://userforum.dhsprogram.com/index.php?t=tree&goto=2 6509&#page_top[/url] " and in this paper "   https://www.nber.org/system/files/working_papers/w18859/w188 59.pdf " by Gary Solon,Steven J. Haider, and Jeffrey Wooldridge, 2013,WHAT ARE WE WEIGHTING FOR?.

they stated that
" If you want a population average, you have to use the weights. That is a general truth about

representative sampling and the sampling structure of the DHS>
But, if you want a causal estimate, it gets a little murkier. If you believe (read: assume) that every person, regardless of their characteristics, will have the same response to some causal input, then you do not need to weight your regressions, because it doesn't matter who was in the sample."
.
.

In conclusion:
1- I am thinking now, Is the option of using "svy" in the analysis up to the author ? or this will yield biased results.!!
2- Also If I have to use the weight+psu+strata in the analysis, Do you know how to include strata in the IV regression command.?

  *******************************command********************** *******

here is the command for IV weighted and Unweighted.
weighted IV
ivregress 2sls ....(outcome variable) (controls) (fixed effect) .... (endogenous variable = instrument)[pweight = sampwt], vce(cluster v021) first

unweighted IV
ivregress 2sls ....(outcome variable) (controls) (fixed effect) .... (endogenous variable = instrument), robust first

****

Thank you

Subject: Re: Weighting in IV approach
Posted by Alanood on Tue, 28 Mar 2023 23:40:58 GMT
View Forum Message <> Reply to Message

Hi again

After I have reviewd the both views of using and not using the svy in the analysis, I think i will go with the first view; to use weight, psu,and strata.

I only faced one issue I mentioned it earlier: strata option is not allowed with "ivreg 2sls" only the sample weight and (cluster psu) is working,

Is there any way to fix this issue and get the "ivreg" result with account for "strata" option. ??? or I should accept the stata results with no strata ?

Thank you

Subject: Re: Weighting in IV approach
Posted by Bridgette-DHS on Wed, 29 Mar 2023 12:47:24 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

Some estimation commands do not allow for a full svyset adjustment. If ivreg does not allow strata, then I'd recommend proceeding without that part of svyset.

There have been other posts on whether or not to use weights. If you do not use weights, then all estimates, not just for the means, but all coefficients, will be biased toward the subpopulations that have been over-sampled and away from the subpopulations that have been under-sampled. Comparisons between surveys in different countries or at different times in the same country will be affected.

Some people, mainly economists, do not like weights because they tend to increase standard errors. If you want to align with that view, you are certainly free to do so.

There is actually a hybrid option, to use the weights for the estimates of coefficients, but not use the weights for the estimates of standard errors. However, this gets messy and is not practical.

I recommend the full svyset adjustment, including weights, but I recognize that some people do not like weights. It's up to you and whoever will review your analysis.

Subject: Re: Weighting in IV approach
Posted by Alanood on Thu, 30 Mar 2023 10:37:12 GMT
View Forum Message <> Reply to Message

Thank you for your reply and clarifications.

I have much clearer vision, and Your advice will be taken into consideration.