Understanding Society User Support - Support #758

weights for pooled cross-sections over waves (a)-(f)

03/29/2017 06:32 PM - Nico Ochmann

Status:	Closed	**Start date:**	03/29/2017
Priority:	Normal	**% Done:**	100%
Assignee:	Nico Ochmann		
Category:	Weights		

Description

Hi there,

I am running hourly wage (constructed with w_paygu_dv) on a number of regressors in a pooled cross-section over all six waves. So far, I am using the whole sample based on GPS, EMBS, BHPS, IEMBS. I am not sure what kind of weights to use in this context given that I want to use all four samples. f_indinui_xw is available for all four for wave 6, so do I just go ahead and use that one? Any piece of advice would be terrific.
Thanks a lot!

History

#1 - 03/30/2017 09:59 AM - Victoria Nolan

- Status changed from New to In Progress

- Assignee changed from Victoria Nolan to Nico Ochmann

- % Done changed from 0 to 10

- Private changed from Yes to No

Dear Nico,

Many thanks for your enquiry, I am passing it on to our weighting team to look into.

Best wishes, Victoria

On behalf of the Understanding Society data user support team

#2 - 04/27/2017 01:19 PM - Peter Lynn

- Target version set to X M

- % Done changed from 10 to 50

Nico,

That would be a correct weight to use, in the sense that it will give population representation when using all 4 samples together. But note that your analysis will then only include people who participated at wave 6.

An alternative is for you to derive a new weight variable, which consists of f_indinui_xw for the wave 6 observations, e_indinub_xw for the wave 5 observations, and so on. See this note, which may help: #494.

Peter

#3 - 04/28/2017 07:53 PM - Nico Ochmann

Hi Peter,

I appreciate your reply very much. I have read your nice little note you coauthored with Olena. It is quite helpful. Let me first write this to make sure I properly understood your note. It seems to me that although wave 6 has been released, I do not get around this additional wrinkle of rescaling because I am pooling data from all six waves. Given that, I focus on Box 2 of your note. So, I generate for the years 2009-2015 strata_year and psu_year following your coding. For the outcome variable, I replace jbstat with paygu_dv and do the same for all seven years. Now and most important, I must derive the new weight variable given f_indinui_xw for wave 6 and e_indinub_xw for wave 5 etc. At this point I am not quite sure how to proceed and I would certainly appreciate it if you gave me a minor hint in one or two coding lines as to how to combine the original weights f_indinui_xw and e_indinub_xw (let's just stick with the two wave example) into one new weight variable. I looked at the online 'Intro to USoc using Stata' course, which is an excellent resource but it does not have any hints on weighting procedures beyond one wave.

If you happen to have any other resources with regard to my issue, please feel free to make any suggestions.

Thank you very much!

Nico

#4 - 05/02/2017 02:11 PM - Victoria Nolan

- Status changed from In Progress to Feedback

- Assignee changed from Nico Ochmann to Peter Lynn

- % Done changed from 50 to 60

#5 - 05/26/2017 04:43 PM - Peter Lynn

- % Done changed from 60 to 70

Nico,

When you pool the data, let's assume you add to each record a variable, wave, to indicate from which wave the record came. Then, you could create the weight with syntax like this:

```
ge newwgt= f_indinui_xw if wave==6
replace newwgt= e_indinub_xw if wave==5
```

Peter

#6 - 05/26/2017 04:43 PM - Peter Lynn

- Assignee changed from Peter Lynn to Nico Ochmann

#7 - 06/01/2017 01:08 PM - Nico Ochmann

Hi Peter,
thanks a lot for your reply. I pooled the data for all six waves and added for each record a wave variable. I then went ahead and generated my newwgt variable as follows:

```
gen newwgt = indinui_xw if wave==6
replace newwgt = indinub_xw if wave==5
replace newwgt = indinub_xw if wave==4
replace newwgt = indinub_xw if wave==3
replace newwgt = indinub_xw if wave==2
replace newwgt = indinus_xw if wave==1
```

Last but not least, I run logrealhourlywage on x1 x2 [pw=newwgt], cluster(pidp)
Is this reasonable or am I still completely off?
I might have to stop my at your seminar on weighting if I am doing this wrong.
Thanks Peter!

#8 - 06/07/2017 04:16 PM - Peter Lynn

Looks fine!

#9 - 06/13/2017 09:15 PM - Nico Ochmann

Thanks a lot Peter!

I appreciate your help and expertise very much.

Have a great week.

#10 - 06/15/2017 03:08 PM - Victoria Nolan

- Status changed from Feedback to Closed

- % Done changed from 70 to 100