

Understanding Society User Support - Support #1652

Should using longitudinal weights lead to a balanced panel?

02/05/2022 08:57 AM - Lucas Auer

Status:	Resolved	Start date:	02/05/2022
Priority:	High	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		
Description			
Dear Olena,			
<p>I have a question concerning the correct use of longitudinal weights. I want to perform longitudinal analysis in stata, starting in wave 6 and ending in wave 11, using information from the individual adult self-completion interview. I understand that I should therefore be using the weight k_indscui_lw.</p> <p>When reading in my data, I create a panel data set in long format, removing the wave prefix from the variables and instead introducing a wave variable called UKHLSwave. Subsequently, I create a new weight variable corresponding to the respective individual's value of k_indscui_lw for each observation (for later use in regression analysis) in the following way:</p> <pre>gen weight11_temp = indscui_lw if UKHLSwave == 11 by pidp: egen weight11 = max(weight11_temp)</pre> <p>My understanding from the weighting guidance (which I found very helpful – thanks a lot!) was that only individuals who gave a full interview at waves 6 through 11 should have a non-zero value of k_indscui_lw. In return, I expected to find my panel balanced for waves 6 through 11 once I condition on my new variable weight11 being non-zero and non-missing. However, when I do:</p> <pre>gen nonzeroweight = (weight11 > 0 & weight11 != .) tab nonzeroweight UKHLSwave if UKHLSwave >= 6</pre> <p>I get strictly increasing numbers of observations from wave 6 through wave 11.</p> <p>Can you please check if you can replicate my findings, and advise where I am doing/understanding something wrong? Otherwise, any insight into why it would be normal for the situation above to arise would be greatly appreciated.</p> <p>Many thanks, Lucas</p>			

History

#1 - 02/07/2022 11:36 AM - Understanding Society User Support Team

- Private changed from Yes to No

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days. While we will aim to keep to this response times due to the current coronavirus (COVID-19) related situation it may take us longer to respond.

Best wishes,
Understanding Society User Support Team

#2 - 02/08/2022 02:03 PM - Olena Kaminska

Lucas,

Thank you. Can you give me a little bit more details on your analysis. It's not clear why you are using long format.

If you were to use wide format and use a regression the weight choice would be correct (k_indscui_lw).

Best,
Olena

#3 - 02/08/2022 03:04 PM - Lucas Auer

Hi Olena,

Thanks so much for getting back to me. Using the data in long format seemed to make perfect sense to me, but I might be wrong. What I want to do is analyse how intrapersonal changes in the explanatory variables affect the dependent variable. To this end, I regress my dependent variable on a number of explanatory variables and controls, using individual-fixed effects (to account for unobserved factors that vary across individuals but not over

time) and wave-fixed effects (to account for unobserved factors that vary over time, but not across individuals). This is implemented in stata as follows, using Sergio Correia's `reghdfe` command:

```
reghdfe depvar expvars controlvars [pweight = weight11], absorb(pidp UKHLSwave) vce(cluster psu hidp pidp)
```

where `weight11` is defined as described in my original post, `absorb(pidp UKHLSwave)` implements the individual-fixed effects and the wave-fixed effects, and robust standard errors are cluster at `psu`, `hidp` and `pidp` level.

Using the data in long format (i.e. one observation per individual and wave) for this purpose to me seemed sensible – or at least innocuous. But like I say: I might have overlooked something. I hope this makes things clearer. Looking forward to hearing back from you!

Best,
Lucas

#4 - 02/08/2022 03:29 PM - Olena Kaminska

Lucas,

Thanks. Please request usersupport for the information on pooled data. Have a look and come back to me if you still have questions.

Best,
Olena

#5 - 02/08/2022 04:45 PM - Lucas Auer

Dear Olena,

Thanks for the swift reply. I have now requested the information on pooled data from user support via email which I will hopefully receive soon. Can I just summarise up to here: My current approach may be misguided in one way (how I prepare the data) or another (how I use the available weights), and you think that reading the information on pooled data will help me understand better how to prepare my data and/or what weights to use for my analysis. Is that correct?

Many thanks,
Lucas

#6 - 02/09/2022 12:11 PM - Olena Kaminska

Lucas,

Your approach seems fine. The choice of weights will be explained via email. Come back to me if you still have questions after reading it.

Best,
Olena

#7 - 02/09/2022 12:40 PM - Lucas Auer

Hi Olena,

Thanks for your reply. I was just typing up a follow-up:

I received the document yesterday and have given it a careful read. It appears that I would fall into the group of users who get confused when it comes to pooled analysis :-). So sorry about that!

Considering that I am interested in the contemporaneous effect of my explanatory variables (labour market status) on my dependent variable (life satisfaction), I realise now that I might want to perform – and most likely, without realising, might already be performing – pooled analysis. I guess the focus of my research falls somewhere between “event triggered situations” and a “time variant state or characteristic”. The only “longitudinal” element to my analysis are the individual-fixed effects that I use to rule out time-invariant unobserved factors.

All things considered, I am now using cross-sectional weights (`indscui_xw`) which I rescale as recommended in the pooling document to account for attrition over time (in my case from wave 6 to wave 11).

Up to here, everything now seems clear to me. Thank you very much for the quick help. In the “PS” below, I have attempted to spell out where my original question in the title of this issue came from. It includes a suggestion on how to rephrase a sentence in the section on “How to use weights” in the documentation. I hope it will be useful for you, but please feel free to ignore if, as I can imagine, there are more pressing matters at hand. Like I say – I am very grateful for your help up to this point and have found all questions answered that are relevant for me to move forward with my project.

Best,
Lucas

PS: With regards to my original question (which prompted me to the fact that something in my understanding of weighting in UKHLS data might be off): As far as I understand now, non-zero longitudinal weights will be assigned to individuals who, once they join the panel for the first time, take part in every subsequent wave. Once they miss a wave, their longitudinal weight will be zero/missing for this and any subsequent waves, even if they do take part again at a later point.

Suppose, for instance: An individual joins for the first time in wave 7, and also takes part in waves 8 and 9. Wave 10, the individual misses, before

taking part again in wave 11. If I have understood correctly, such an individual would receive a non-zero longitudinal weight for waves 8 and 9, and a zero longitudinal weight for wave 11.

If that is the case, my original misperception about longitudinal weights leading to a balanced panel had arisen from the example provided under the following link: <https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/how-to-use-weights> There, it currently says: "For example, there are 39,289 persons in the file h_indresp, but only 27,841 of these have a non-zero value of h_indinui_lw. These are the people who gave a full individual interview at all of Waves 6, 7 and 8."

To me, that sounded like an individual HAD to already be part of the sample at wave 6 to be eligible for a nonzero value for the longitudinal weight h_indinui_lw. I fear this could easily be interpreted in the wrong way by other users as well. Therefore, I would like to suggest rephrasing it to: "These are the people who, after giving a full individual interview for the first time at some point from wave 6 onwards, gave a full interview at all of the subsequent waves from that wave up to wave 8." This might make it a bit clearer that also an individual who was interviewed for the first time in wave 7, and is interviewed again in wave 8, will still receive a nonzero longitudinal weight for wave 8 (despite not being part of the sample of wave 6).

#8 - 02/09/2022 02:21 PM - Olena Kaminska

Lucas,

Thank you. You seem to be on a good track with your analysis.
And thank you for the suggestion on a better wording. I will try to resolve it.

Best wishes,
Olena

#9 - 02/21/2022 02:56 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 90

#10 - 06/06/2022 09:06 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 90 to 100