

Understanding Society User Support - Support #1228

Unbalanced panel weight

08/16/2019 11:04 AM - Louise Luo

| | | | |
|---|----------|--------------------|------------|
| Status: | Resolved | Start date: | 08/16/2019 |
| Priority: | High | % Done: | 100% |
| Assignee: | | | |
| Category: | | | |
| Description | | | |
| <p>Hello. I am working on regression models using fixed effect. It is a longitudinal analysis using all the available data at the level of adult main and proxy interview. The data I am using are drawn from both BHPS (w1-w18) and US (w1-w8). The following 2 questions are related to the survey weight for the unbalanced panel:</p> | | | |
| <p>1. In the case of a balanced panel, the survey weight is equal to the longitudinal weight of the last wave (suggested in the 2016 user guide). Are the survey weight of the unbalanced panel equal to the longitudinal weight of the last wave they participant (eg. the weight of an individual who participated in wave 1, 3 and 4 is equal to the longitudinal weight in wave 4, and the weight another individual participated in wave 2-6 is equal to the weight in wave6)?</p> | | | |
| <p>2. Which longitudinal weight should I use, n_indpxus_lw, n_indpxub_lw or other?</p> | | | |
| <p>Many thanks in advance.</p> | | | |

History

#1 - 08/19/2019 10:39 AM - Stephanie Auty

- Category set to Weights

- Assignee set to Olena Kaminska

- Private changed from Yes to No

#2 - 08/19/2019 12:06 PM - Olena Kaminska

Thank you for your question.

Generally you won't go wrong if you use the weight from the last wave in your analysis. For example if you use wave 8 of UKHLS - take lw weight from this wave - this weight will apply for all people. So, if you start with BHPS 1991 you would use '91' weight.

Now, there are other options to analyse the data which may increase sample size. To help you further I would need to know the following:

- what is your unit of analysis (people, events etc.)
- which population are you planning to represent?

Thank you,
Olena

#3 - 08/19/2019 03:16 PM - Louise Luo

Dear Olena,

Many thanks for your help.

Back to your question, the unit of my analysis is individual in England. I am investigating young graduates who are 20-34 years old and with at least a degree. I drew my sample from the indresp.dta file including young graduates participating in both proxy and main interview. Which weight is suitable for the case?

Kindly regards. Thank you again.

Louise

#4 - 08/20/2019 12:39 PM - Olena Kaminska

Louise,

So, I guess you don't care when a persons is 20, in 1991 or 2008 let's say. And I imagine you are doing a longitudinal analysis with them. So think about it in this way: we have a group of young people of 20 year old in 1991 - you will be using longitudinal information for them - so for these take the weight from the last wave in your analysis for them, and it should be '91' weight.

Then you may have a group of 20 year olds in 1992 - same. So, you will always use lw weight and it will be from the last wave of the information you use for this specific group. Make sure you can define group in substantive words, e.g. 20-year olds in 1992 etc. Essentially you will create a weight

that will have different values for each group.

Finally this may be helpful for you to decide on the zz part of the weight if the weight is described as w_XXxyzz_aa

zz part: what is the timeline of your research?

- Wave 6 onwards: ui_

- Between wave 2 and wave 6 or starting at wave 2 for longitudinal analysis: ub_

- Wave 1 only or starting at wave 1 for longitudinal analysis: us_

- Starting in 2001: 01_

- Starting in 1991: 91_

Hope this helps,
Olena

#5 - 08/20/2019 04:48 PM - Louise Luo

Hi Olena,

Thanks for your reply.

I am investigating young people aged 20 to 34 years old in each wave. It is a longitudinal analysis of the period from 1991 to 2017 (18+8 waves in total). Following your advice, I will definitely use '91' weight. However, there are two cases which I am still not sure about:

1. Person A is recorded from wave 1 to 8. Person A was 34 years old in wave 6. However, A is dropped from my sample since wave 7 because A is over 35 years old from wave 7. In the case, do I use A's weight in wave 8 or wave 6?
2. Person B is recorded from wave 1 to 6. For this person, weight in the wave 8 of UKHLS should be missing. In this case, I suppose I should use B's weight in wave 6?

Louise

#6 - 08/20/2019 05:23 PM - Louise Luo

Hi Olena,

Sorry for bothering you again.

I just found that '91' weight is only applied to those who participated in both BHPS and UKHLS. However, my sample also includes those who only participated in BHPS and only in UKHLS. Which kinds of weight are suitable for these individuals?

Many thanks in advance.

Louise

#7 - 08/21/2019 10:22 AM - Alita Nandi

- Assignee changed from Olena Kaminska to Louise Luo

- % Done changed from 0 to 50

#8 - 08/21/2019 02:22 PM - Olena Kaminska

- Assignee changed from Louise Luo to Olena Kaminska

- % Done changed from 50 to 0

Louise,

'91' is the zz part of the weight. To choose it correctly please refer to my earlier message.

Generally it is wrong to pick one subsample (e.g. UKHLS without BHPS) - our weights are designed to combine these from the time UKHLS starts. There should not be theoretical substantive situation when you'd want to separate these. '91' is the weight for a longitudinal analysis that starts before 2009. We also have '01' weight for analysis that starts since 2001 but before 2009 etc. Each of our weights includes all people who can be included (i.e. all relevant subsamples with appropriate adjustments).

I understand that you want to use unbalanced analysis. We don't have a weight for this. So you would need to start with one of our weights and correct for the leftover nonresponse yourself. I suggest that you start with the weight from the first wave of your analysis. Then you can adjust for nonmonotone attrition yourself, and multiply this adjustment by our weight.

Hope this helps,
Olena

#9 - 08/27/2019 03:38 PM - Stephanie Auty

- Status changed from New to In Progress

- % Done changed from 0 to 70

#10 - 08/30/2019 11:13 AM - Louise Luo

Olena,

Thanks a lot. This is a great help!

Unbalanced panel weight might be more suitable for my research. However, I am not quite sure what you mean by "Then you can adjust for nonmonotone attrition yourself". I suppose it might not be an easy task to calibrate the weight on my own, but if you have any further instructions on how to compute a new weight for the unbalanced analysis, pls don't hesitate to send it to me.

Louise

#11 - 09/02/2019 02:19 PM - Olena Kaminska

Louise,

Yes, we have a general guidance for such situations. Please see below:

Creating your own tailored weight

When?

Suppose you wish to carry out longitudinal analysis of responses to questions that were included at waves 1, 4 and 7. Your analysis base is therefore sample members who completed an individual interview at each of those three waves (let's assume that your survey questions of interest were not all included in the proxy questionnaire and that you therefore cannot include proxy responses in your analysis).

One option would be to use the wave 7 longitudinal weight for the wave 1 sample, i.e. `g_indinus_lw`. However, this weight is only defined for sample members who gave a full personal interview at all seven waves, thus 18,510 persons have this weight, whereas 20,390 responded at waves 1, 4 and 7 (so, 1,880 of those who responded at waves 1, 4 and 7 must have failed to respond at one of waves 2, 3, 5 or 6). Using this weight for your analysis would therefore cause almost 10% of your potential analysis sample to be dropped from the analysis. This reduction in sample size will cause a modest reduction in the precision of your analysis (increase in standard errors). The effect will be rather small, and you may well be willing to accept this slight reduction in sample size, unless you are producing estimates for very small population subgroups. But if you want to be able to include all 20,390 respondents in your analysis, you will need to derive your own weight.

How?

First, identify the (smallest) hierarchically-superior sample for which weights have been provided. In this example case, it is the wave 1 responding sample. For this sample, the weight `a_indinus_xw` has been provided. This will serve as your "base weight", to which you will make an adjustment tailored to your analysis sample.

Next, fit a conditional model (e.g. logit) of response to your wave-combination of interest. In the example case, the base for the model would be all wave 1 responding OSMs (i.e. OSMs with a non-zero value of `a_indinus_xw`) and the dependent variable would be a 0/1 indicator of whether they also responded at both wave 4 and wave 7 (and removing from the base any known to have died or emigrated before wave 7). Predictor variables in the model can be anything relevant observed at wave 1. The model will give you a predicted probability for every wave 1 respondent of responding also at waves 4 and 7. Call this π .

Now, to make the adjustment to your base weight you simply multiply `a_indinus_xw` by $1/\pi$ for all the cases in your analysis sample.

Hope this helps,

Olena

#12 - 03/02/2021 03:55 PM - Understanding Society User Support Team

- Status changed from In Progress to Resolved

- % Done changed from 70 to 100

#13 - 08/05/2021 01:58 PM - Understanding Society User Support Team

- Assignee deleted (Olena Kaminska)