Understanding Society User Support - Support #1892

Question regarding longitudinal weights

04/19/2023 04:58 PM - Johanna Pauliks

Status:	Resolved	**Start date:**	04/19/2023
Priority:	Normal	**% Done:**	100%
Assignee:	Olena Kaminska		
Category:	Weights		

Description

Dear support,

I've got another question regarding longitudinal weights in UKHLS. I'm using data from adult respondents (individuals) from wave 2 up to wave 10 and using the fixed-effects model (so I'm doing longitudinal analysis). According to the user guide, longitudinal weights are appropriate for this. But if I use the longitudinal weight from wave 10 I lose all cases of individuals who drop out of the sample in previous waves, even so they participated in all waves prior to this (let's say they participated from wave 2-8). This is a problem, as some of my subgroups are very small to begin with. Would it be possible to use the longitudinal weight from wave 10 for everyone who participated in all waves up to wave 10, the longitudinal weight from wave 9 for every respondent who participated until wave 9 and so on, or would this not be appropriate, and I need to create my own tailored weights?

Best regards
Johanna

History

#1 - 04/21/2023 02:22 PM - Olena Kaminska

Johanna,

It sounds to me that you may be pooling information from multiple waves, which depending on how you do it, you may need cross-sectional or different longitudinal weights. Have a look at this general advice below, as see if this may answer your question.

You can pool the data however you want. There are three most important points to keep in mind:
1. Always take into account clustering within PSUs with UKHLS data. Taking into account clustering within a person (in case you have multiple entries per person) is optional and could be used in addition to clustering within PSUs. This implies that you don't need to use multilevel models while pooling – you could use the standard svy command if this suits your purpose.
2. When pooling information from multiple waves, especially BHPS waves and UKHLS waves you need to apply additional scaling to weights in order for each wave to contribute a similar level as all others. See question 19 in this document for how to implement it.
3. Define your population carefully. Unlike unpooled analysis, where population definition is straightforward, we find that many users get confused with the population definition in the pooled analysis. A few examples follow presenting the population definition and the data structure:
- Events, e.g. hospitalization occurrences (staying in a hospital for over 24 hours) observed in GB between 1991 and 2009 and UK between 2009 and 2020. In this situation hospitalization variable would be created and data is pooled from all waves and all people observed in each wave between 1991 and 2020. Note, you are studying events, not people, in this situation.
- Event triggered situations, e.g. happiness upon marriage observed in UK between 2009 and 2020. If you study the state after marriage – you could pool all the observations after marriage in the data from all the time points. Your data will consists of all marriages and relevant observations following from all waves between 2009 and 2020. You are studying happiness following marriage, i.e. a state following an event (not people).
- A subgroup defined by a time point, e.g. 11 year olds living in UK between 2009 and 2020. You could pool information from 11 year olds from each wave and analyse them together in one model, which gives you more statistical power. In this situation you will have one observation per person as a person is 11 only once per lifetime (and wave).
- A subgroup defined by an event where event may happen multiple times, e.g. first year students studying in UK between 2009 and 2020. You could pool first year students from all the years we have in the study. Note, some people may have multiple occurrences of being a first year student. It then depends on your definition. If you want to study number of books read in a year by the first year students it may be appropriate to count all the multiple occurrences per person. In this situation you don't study people really but 'event triggered states'.
- Time variant state or characteristic, e.g. wellbeing observed in UK between 2009 and 2020. While wellbeing changes over time and it may be more appropriate to study it using a classic longitudinal analysing, there are situations, especially when studying very small subgroups, where pooling may add statistical power. In essence you are studying wellbeing states observed over a specific time period, (again not people). For this you just pool all the information on wellbeing from all the relevant waves.
- It does not make sense to study time-invariant states (e.g. eye colour) with pooled analysis. If you happen to do it, your effective sample size will not be any higher than in an unpooled analysis. So, technically there won't be any gain from pooling, and it would be easier and clearer to avoid it.

Pooling can be cross-sectional or longitudinal. Theoretically, you will be combining 'separate samples of events / states' each of which will have the corresponding weight.
If you are just interested in events (and what happened at the same time / wave) you are looking at pooling cross-sectional information. For this create a new weights variable new_weight, and give it a value of the cross-sectional weight from each wave (e.g. new_weight=a_indinus_xw if wave==1; new_weight=b_indinub_xw if wave==2 etc.)
Alternatively you may be interested in what happens before and / or after a particular event (e.g. studying work pattern for 3 years before birth of a first child and 3 years after for new mothers). In this situation you need to choose a longitudinal weight from the last wave in your analysis for each combination of waves (e.g. for birth at wave 3 where we observe waves 1-6, the weight will be f_indinus_lw; for birth at wave 4 with information in the model of waves 2-7 –it will be g_indinub_lw etc.).

#2 - 04/21/2023 07:22 PM - Johanna Pauliks

Dear Olena,

thank you very much for your in-depth answer! I'm interested in what happens after a particular event, so I'm going to do what you are suggesting. If it's not too much trouble, could you answer another question regarding PSUs? I was under the impression that accounting clustering within PSUs mainly affects the standard errors of my estimation. Am I correct in this, or do they also affect my point estimates?

Best regards
Johanna

#3 - 04/23/2023 11:10 PM - Understanding Society User Support Team

- Status changed from New to In Progress

- % Done changed from 0 to 80

#4 - 04/23/2023 11:10 PM - Understanding Society User Support Team

- Private changed from Yes to No

#5 - 04/24/2023 01:16 PM - Olena Kaminska

Johanna,

Yes, clustering changes only standard errors, and therefore confidence interval, and therefore p-values of group comparisons, for example. But it does not influence point estimates.

Hope this helps,
Olena

#6 - 05/10/2023 08:24 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

#7 - 11/30/2023 01:01 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100