Understanding Society User Support - Support #1624

Weights for subsample

01/06/2022 02:49 PM - Ashley Burdett

Status:	Resolved		**Start date:**	01/06/2022
Priority:	Urgent		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Hello,

I am trying to estimate the fraction of people that transition to their first relationship (cohabitation or marriage) by age using the BHPS.

To do this I have constructed an unbalanced panel containing observations for individuals who have never had a relationship (marriage or cohabitation) before. Precisely I use observations for individuals that did not report a relationship in the marital history datasets but provided a full response to the wave 2 main survey. I also include observations for individuals that aged into the sample during the panel to increase my sample size.

I include observations for these individuals up until either they form their first relationship, they have a missing observation or the survey ends (2008).

Using this sample, I simply calculate the fraction of individuals observed at each age that transition to their first relationship at that given age.

My question is how do I appropriately incorporate weights into this analysis? I have tried numerous ways of approaching this problem and get very different results each time.

Many thanks in advance for your help.

All the best,

Ashley

History

#1 - 01/10/2022 09:57 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 - 01/10/2022 09:57 AM - Understanding Society User Support Team

- Status changed from New to In Progress

#3 - 01/10/2022 12:08 PM - Olena Kaminska

Ashley,

Thanks for your question. If you use wave 2 information than you will need to use longitudinal weight for each wave combination. You will use 91_lw weight.

Does this help?
Olena

#4 - 01/14/2022 11:58 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

- % Done changed from 0 to 80

#5 - 01/30/2022 09:50 AM - Ashley Burdett

Olena,

Thanks for your response and apologies for the delay.

Not sure this will resolve my problem. I am using observations from throughout the BHPS panel when individuals are aged between 24-28 which can occur in any wave. Is it appropriate to use the longitudinal weight from the final wave of the survey even if I am using observations from earlier on in the panel? I presume this will hurt my sample size, is there any way to preserve it?

Many thanks,

Ashley

#6 - 01/31/2022 11:06 AM - Olena Kaminska

Ashley,

Can you specify which information you use in each wave combination. Give me an example if a person is 24 in wave 2 of BHPS. Also, are you using only BHPS information or do you combine it with UKHLS data when it is available?

Thanks,
Olena

#7 - 01/31/2022 02:27 PM - Ashley Burdett

Olena,

I construct a panel using BHPS wave data only containing variables with information about age, marital status, sample origin, sample status and the various available weights.

To construct the sample I use for estimation, I select all single (never married/divorced) spells that are observed from their beginning i.e. I observe the individual in a relationship (marriage or cohabiting union) during the panel, but in the subsequent wave they report being single. Of these single spells, I drop all those in which the individual is not aged between 23-27 in the final observation of their last relationship, if the individual is not an OSM or if they are not part of the original bhps gb 1991 sample. For these single spells, I include all observations up until either the end of the panel, the individual has a missing observation or they form a new relationship (whichever happens first) - I have right censoring due to the panel ending or individuals missing an interview.

Using this sample, I estimate the transition rate to a new relationship by duration of the single spell as the simple ratio of the number of transitions at duration t over the total number of individuals that could have experienced a transition at duration t (lifetable calculation without adjustment). My question is how do I incorporate weights into this calculation?

I am stuck on the first step. In particular, I am unclear which is the appropriate weight to use. My observations come from throughout the panel and transitions can occur after any duration, thus it is unclear if longitudinal weights are appropriate. I have tried numerous ways to include both the x section and longitudinal weights, and in each instance I get very different results, none of which are similar to the estimates I obtain when I don't use any weights. Can you advise me on which weights are appropriate?

On a side note, I have utilized the USoc retrospective data to calculate the same statistic for the same period using the weights from wave 1 as a sense check. When I do this, I get similar results to the estimates I get when I use BHPS wave data and no weights (at least for short durations when my sample size isn't too bad in either dataset).

Hopefully, this provides you with enough information. Please let me know if anything is unclear.

All the best,

Ashley

#8 - 01/31/2022 02:40 PM - Olena Kaminska

Ashley,

Thank you for the details. So, for each person you are only interested in their life time period between 23 and 27.
But what do you specifically want to know: the proportion of single / married at the age of 23 (for example)? Or do you want to know how many of singles at 23 got married at 24 / by 27? It's improtant to differentiate as in the first example you should use xw weight and in the second - lw weight. Also, is there a specific reason you are only using BHPS data? Adding UKHLS data would give you higher sample size - as you are pooling you could use these together.

Best,
Olena

#9 - 01/31/2022 03:54 PM - Ashley Burdett

Olena,

Thanks for getting back to me so quickly.

Not necessarily, individuals could be in my sample until they are in their forties if they became single in wave 2 at age 27 and remain in the sample (observed to be single in every wave) until late in the panel.

Specifically, I am hoping to obtain an estimate for the transition rate to a new relationship if an individual becomes single when they are "prime-age" (24-28 years old).

Regarding including UKHLS data, I haven't combined the datasets so far because my paper is a criticism/extension of an existing paper that only uses BHPS data. I, therefore, base my analysis on the same data source. Thus far, I have only looked at the retrospective UKHLS data as a sense check because it is more straightforward to include weights.

Many thanks,

Ashley

#10 - 02/01/2022 11:44 AM - Olena Kaminska

Ashley,

It sounds to me as a classic pooled analysis where you are not interested in representing people, but rather representing events: a change from being single to being in a relationship. Can I just clarify - if a person becomes single three times in their 20s - do you see it as three separate observations? If so, read our FAQ on pooled analysis (not yet published, so email a request to usersupport).

If you are interested in a change and you derive it from the information from one wave (where you may learn that a person is in a new relationship from one wave questionnaire) you can use xw weights. If you need one previous wave and the current wave to get your information then use lw weight from the current wave. Also look at predictors, other variables in the model. If all the information is always from the current wave - use xw weights. If you use any longitudinal information (past or future) use the weight from the last wave of an observation.

Hope this helps,
Olena

#11 - 02/01/2022 02:45 PM - Ashley Burdett

Thanks, Olena.

I think, given your question, that I am interested in a pooled analysis (if an individual becomes single twice between age 24 and 28 they will have two single spells in my sample). I've therefore requested the document you suggested and will follow up here if there are any questions once I've read it through.

All the best,

Ashley

#12 - 02/01/2022 02:54 PM - Olena Kaminska

Ashley,

Ok, just one more question: are you using survival analysis? If so, you would use the weight at the first observation observation instead of the last and let survival analysis to correct for attrition.

Best,
Olena

#13 - 02/03/2022 02:04 PM - Ashley Burdett

Olena,

I'm still waiting for the pooled analysis document, but I was wondering if your last comment applies to lifetable methods? Is it sufficient to use wave 1 weights for my individuals and proceed with the standard lifetable calculations?

Many thanks,

Ashley

#14 - 02/03/2022 02:27 PM - Olena Kaminska

Ashley,

I haven't used lifetable calculations, so I can't say. It depends on how the analysis treats missing data. Survival analysis can use truncated data and nonmonotone data and makes specific assumptions about missingness - thus taking attrition into account within the analysis. You would need to check those aspects with your method. If unsure using our longitudinal weights will always give you a correct result (though with smaller sample size).

Hope this helps,
Olena

#15 - 02/04/2022 10:00 AM - Ashley Burdett

Olena,

I have now reviewed the pooled analysis document and it seems the document recommends that I use scaled x section weights. I tried to implement the code in the document last night and got very different estimates compared to when I don't use weights. The estimates are also very sensitive to the range of ages I include in the calculation. For example, If I also include 23 years olds the estimates decrease significantly. This is raising concerns to me about the robustness of what I am doing.

In addition, I also tried using the first wave weights, as you recommended for survival analysis. In that case, the estimates only change slightly compared to when I don't use weights and are similar to the estimates I get when I compute the same estimates using the retrospective USoc data over the same period.

As a consequence, I am unsure about the best way to proceed. My instinct is to opt for the estimates I get when I use the first wave weights because they are corroborated with the USoc estimates. Do you have a recommendation?

Many thanks,

Ashley

#16 - 02/04/2022 10:32 AM - Olena Kaminska

Ashley,

You can use only the first wave weights with survival analysis.
I am still not sure whether your analysis is cross-sectional or longitudinal. If you need to observe a person before or after (and not just at one point in time) then you should not use cross-sectional weights. Especially for analysis of new couples it will give you very skewed estimates. If unsure, just use _lw weights.
The change in estimate with scaling is probably driven by population changes over time - that's why scaling in this situation is important.
And adding 23 year olds to analysis changes estimates because your 23 year olds are very different on the estimate than others. So this should reflect the true population situation.

Hope this helps,
Olena

#17 - 02/04/2022 11:45 AM - Ashley Burdett

Olena,

Thanks very much for your help. I have now implemented the longitudinal weights as you suggested and my estimates seem more reasonable.

Many thanks,

Ashley

#18 - 04/30/2022 01:50 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100