Understanding Society User Support - Support #1903

Weighting guidance for intergenerational analysis

05/16/2023 05:34 PM - Esme Lilly

Status:	Resolved		**Start date:**	05/16/2023
Priority:	Normal		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Hi, I am doing research into the role of education in intergenerational income mobility, and I had some questions on tailoring my own weights.
My sample restrictions are individuals with atleast one income observation aged 25+, born between 1975-1995 with atleast one observed parental household income aged 0-16, restricted to only one observation per household (youngest eligible sibling) from **the original and booster BHPS samples** . Children are matched to their parents through the alterego file. My sample spans all 30 waves of the BHPS & USoc.

My main variables of interest are parental household income and offspring individual income, and the highest educational attainment of the offspring. To measure income, I take an average of income observations during aged 25+ and parental income observations aged 0-16, making it a pooled cross-section.
As my sample is already quite small based on these sample restrictions (N=1,392), I want to retain as many observations as possible. As my analysis is not longitudinal, the longitudinal weight provided is sub-optimal for me because I want to include respondents even if they haven't responded in the most recent wave or have missed waves. I am not interested in change over time.

I have been reading the FAQs and other questions on the forum to help with tailoring my own weights, but I still had a few questions at the different stages of tailoring my own weights. I followed the advice from [#658](#) and on the online course and weighting FAQ.
First, to create my base weight, I took the cross-sectional weight for everyone with atleast one income observation aged 25+, born between 1975-1995, as this is the sub-population.

1. 1) Is it ok that the cross-sectional weights I take are USoc cross-sectional weights, despite me only including offspring from the BHPS samples?
2. 2) Why do some people have a cross-sectional weight of zero in the BHPS and then a non-zero weight in USoc? If I'm trying to maximise my sample, can I take the first weight they have observed past the age of 25 that is non-zero?

However, my colleagues suggested that I should actually do the weighting based on parental characteristics. That is to say, the base sample should be all parents who responded in the original and booster samples, and then model the nonresponse based on the substantial attrition of how many offspring have an income observation aged 25+. My problem with this being that the base weight wouldn't account for potential parents who went on to have children but left the survey beforehand. My questions are:

1. 3) Is it ok to use people who answer the survey 25+ as the base weight, or would it be better to use people who had children from the original and booster samples as my base weight? What would the difference between these two be?

Finally, after rescaling the weights and taking mortality into account, I adjust for non-response. I fit a logistic regression model and tested the following predictors: sex, region, race, migrant status, quintile of income and education, and the wave they first responded over the age of 25. Variables of income, education and region are from the wave they are first observed over 25. When excluding the first wave they respond over 25, all other predictors are highly significant, but when including it, they are all less significant and sex is not significant.

History

#1 - 05/17/2023 01:42 PM - Olena Kaminska

Esme,

Thank you for your question. To answer it you would need to clarify the population that you want to represent, and units of analysis.
You describe: My sample restrictions are individuals with atleast one income observation aged 25+, born between 1975-1995 with atleast one observed parental household income aged 0-16, restricted to only one observation per household (youngest eligible sibling) from the original and booster BHPS samples .

Instead, please think of the subpopulation that exists outside our study - in the greater UK population. Which people do you want to represent?
In simple words, we can't represent people who have at least one observation of a kind in UKHLS, because such people do not exist outside UKHLS. So, there aren't weights for such representation. Once your population is identified, I could help you with the choice of weights.
In short, if your analysis is cross-sectional, it is unlikely you need tailored weights. But if you are using some information from multiple waves (parents income reported earlier, for example, with the current youth information), your analysis is then longitudinal.

With more information I should be able to advice you,

Olena

#2 - 05/17/2023 04:46 PM - Esme Lilly

Dear Olena,

Thanks for your swift response.

The population that I want to represent are people born 1975-1995.

My research is not cross-sectional, as I am using information from multiple waves. That is, parents income reported earlier, and also their own income reported across multiple waves. However I can't use the longitudinal weights because it would reduce my sample to around 250 people, and I don't need to exclude people if they haven't answered in every preceding wave.

Many thanks,

Esme.

#3 - 05/19/2023 12:21 PM - Olena Kaminska

Esme,

Thank you for this clarification. You will need longitudinal weights. It may be best for you to create you own tailored weights. Please follow our online course on how to do it: https://www.understandingsociety.ac.uk/help/training/online/creating-tailored-weights .

Hope this helps,
Olena

#4 - 05/19/2023 01:40 PM - Esme Lilly

Hi Olena,

Thanks. I have followed the online course, however the course could not provide me with the answers on how to tailor the weights for my situation, which is what led me to my original question.

I came to my above solution from the online course and issues #658 and #985 , which suggest that for my base-weight I take the cross-sectional weight for people born 1975-1995 from the first wave they turn 25.

My questions related to that base=weight and whether I've applied the weights correctly.

Many thanks,

Esme.

#5 - 05/19/2023 01:43 PM - Olena Kaminska

Esme,

Cross-sectional weight is a wrong one as a base weight, as we don't follow TSMs by design. You should use a longitudinal (can be enumeration) weight as your base weight. Using a weight when a person turns 25 is a good idea. You then represent 25 year olds and what happened to them thereafter.

Hope this helps,
Olena

#6 - 05/19/2023 02:29 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 80

- Private changed from Yes to No

#7 - 05/22/2023 02:22 PM - Esme Lilly

Hi Olena,

Thanks. I have done so using enumeration weights for my sample from the year they turn 25. However, now 330 of my sample have a weight of 0, severely reducing my sample size.

Is there any other way to do this so that there are less people with a weight of 0, to maximise my sample size but still make sense? E.g. using enumeration weight from an earlier age?

Thanks for your help,

Esme.

#8 - 05/22/2023 02:46 PM - Olena Kaminska

Esme,

First, check if 330 people are OSMs. If they are TSMs they should not have any weight. And yes, if you want, you could use an earlier base weight, for example an issue weight to wave 2.

Hope this helps,
Olena

#9 - 05/22/2023 04:55 PM - Esme Lilly

Hi Olena,

Only 24 of the 330 people are TSMs. I've tried using issue weights from wave 2, but unfortunately, it leads to a similar amount of people excluded either through a weight of 0, or not having an issue weight from wave 2. If there's no other way, I suppose I'll just have to use this smaller sample size?

Thanks,

Esme.

#10 - 05/23/2023 01:30 PM - Olena Kaminska

Esme,

Yes, I think you've done your best.

Olena

#11 - 10/24/2023 09:21 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100