Understanding Society User Support - Support #1868

Use of weights for analysing job quality in UKHLS, Waves 4, 6, 8 and 10

02/25/2023 04:00 PM - Thomas Stephens

Status:	Resolved		**Start date:**	02/25/2023
Priority:	Normal		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Good afternoon,

I have a few questions about the weights to use for some analysis of job quality which I'm carrying out using Understanding Society. I have read another very useful support response on this (see: https://iserredex.essex.ac.uk/support/issues/1739), but this still gives rise to some further questions.

I'm planning on carrying out two distinct types of analysis for my research. Although working conditions data is available in every other wave, note that I exclude wave 2 from my analysis, for reasons I expand on below:

- **Descriptive statistics of changes over time across every other wave,** ie comparing Wave 4 vs. 6 vs. 8 vs. 10...;

- **Analysis of pooled data from these waves**, to understand the relationship between job quality and various other individual and household characteristics across all of Waves 4, 6, 8, 10.

I have the following questions about which weights to use, and the weighting process in general:

1. Will I have to use two different weights for these two types of analysis? My understanding is that the existing indinub_xw weight (ie just removing the wave prefix) would suffice for the first type of analysis (as per my reading of the Harmonised BHPS user guide, p. 25, https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/bhps-harmonised-user-guide.pdf), but that I will have to do weight rescaling for the pooled analysis to avoid under-representing respondents from later waves. Is this correct?

2. Although I only analyse at every other wave, I have created some new indicators by looking back at data from the wave immediately prior to the respondent's wave (eg I use Wave 3 data to establish whether respondents in Wave 4 have been continuously employed for >1 wave or >2 waves, wave 5 for wave 6, wave 7 for wave 8, etc...). Does this have any bearing on the weight I should rescale to for the pooled analysis?

3. For the above reason, I exclude wave 2 data, as the relevant questions weren't asked in wave 2. Am I correct in assuming that if my pool starts at Wave 4, this means I need to re-scale to Wave 4 rather than Wave 2, using a variation (albeit in R rather than Stata, as that's what I'm using...) of the code you give here: https://iserredex.essex.ac.uk/support/issues/1739? Are there any other issues I need to be aware of?

4. I won't be analysing changes based on calendar years; I'll be keeping respondents in their waves. My reading is that I therefore don't have to carry out the adjustments you outline in p. 10 of your weighting FAQs: https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/weighting_faqs.pdf. Is this correct?

5. I haven't seen any discussion of seasonality in the user forum or weighting FAQs. Ie if one wave happens to over-represent people interviewed in later seasons where labour market statistics might be different. Could I check whether your weights account for this?

Many thanks in anticipation.

Best wishes,

Tom

History

#1 - 03/02/2023 02:43 PM - Olena Kaminska

Tom,

Thank you for your questions.

1. Yes, different weights. If you use multiple waves in your analysis (e.g. comparing wave 4 to 6) you will have to use a longitudinal weight. For pooled analysis xw weight seems to be sufficient. Rescaling is advisable.
2. No bearing on rescaling, but this means you need a longitudinal weight;
3. Really doesn't matter which number you rescale to - you just need the same weighting total for each wave after rescaling to ensure that each wave contributes the same amount of information to the total analysis;
4. No adjustments for calendar years needed. Also, see below for the advice on simpler rescaling.
5. No, weights do no specifically take seasonality into account - they are for the analysis of the whole wave. While nonresponse may differ over seasons, in our weights it is the overall effect that is accounted for. Seasonality, and therefore nonresponse related to it, would be important if you study different months of the year separately. The approach would be similar to the one described for a calendar year in FAQs, and if you are concerned about seasonality of nonresponse and it's influence on your estimates, there would be a possibility to create your own tailored weights to account for this effect specifically. This would not be necessary though if you are planning to use the whole wave in your analysis (instead of splitting it by seasons).

Hope this helps,
Olena

#2 - 03/03/2023 02:12 PM - Thomas Stephens

Olena,

Many thanks, this is very helpful. However just three clarifications / further queries ...

First, apologies but I can't seem to see the "advice on simpler rescaling" you mention in your response to Q4 - could you forward/resend?

Second, I think I have been unclear in how I explained my first type of analysis ("Descriptive statistics of changes over time across every other wave, ie comparing Wave 4 vs. 6 vs. 8 vs. 10...;"), hence why you've recommended a longitudinal weight here.

To clarify: I'm carrying out cross-sectional time series analysis, ie creating a cross-section of each wave and seeing how the job quality scores, percentage deprived, etc change in each cross section. I'm not doing any longitudinal analysis: ie what is the score of person X in wave 4, person X in wave 6, etc. The UKHLS user guide seems to advise a cross-sectional _xw weight for my kind of analysis - eg see the Stata code on p.25, which happens to mention the exact weight I use:
https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/bhps-harmonised-user-guide.pdf.

Taking all this together, my understanding - based on what you've said - is I should therefore use the same weight for all of my analysis (indinub_xw), but with rescaling for the pooled data.

Finally, you're correct to say that I won't be studying months of the year separately, only wave-by-wave or pooled waves. My question was more about the implications for time series analysis if, say, people in Wave 4 are more likely to be interviewed later in the year than people in Wave 6, etc. My initial analysis suggested that there is quite a bit of wave-by-wave difference in when people are interviewed, and I suspect this will be especially difficult for waves during the pandemic, should I analyse them? However, I'm fine with your conclusion regardless.

Many thanks once again for your advice.

Best wishes,

Tom

Olena Kaminska wrote in #note-1:

> Tom,
>
> Thank you for your questions.
>
> 1. Yes, different weights. If you use multiple waves in your analysis (e.g. comparing wave 4 to 6) you will have to use a longitudinal weight. For pooled analysis xw weight seems to be sufficient. Rescaling is advisable.
> 2. No bearing on rescaling, but this means you need a longitudinal weight;
> 3. Really doesn't matter which number you rescale to - you just need the same weighting total for each wave after rescaling to ensure that each wave contributes the same amount of information to the total analysis;
> 4. No adjustments for calendar years needed. Also, see below for the advice on simpler rescaling.
> 5. No, weights do no specifically take seasonality into account - they are for the analysis of the whole wave. While nonresponse may differ over seasons, in our weights it is the overall effect that is accounted for. Seasonality, and therefore nonresponse related to it, would be important if you study different months of the year separately. The approach would be similar to the one described for a calendar year in FAQs, and if you are concerned about seasonality of nonresponse and it's influence on your estimates, there would be a possibility to create your own tailored weights to account for this effect specifically. This would not be necessary though if you are planning to use the whole wave in your analysis (instead of splitting it by seasons).
>
> Hope this helps,
> Olena

#3 - 03/03/2023 02:43 PM - Olena Kaminska

Tom,

Yes, sounds like cross-sectional weight will be suitable. After wave 6 it will be 'ui' instead of 'ub'.
You don't need rescaling if you are not pooling information into one model (e.g. separate cross-sectional analysis for each wave).

Seasonality should not be a problem with UKHLS - the interviews are carried over 2 years in all seasons. There are sampling months that are issued in each months of this 24 months period. Covid didn't create a problem with this regard, so most definitely use those waves.

Rescaling:
In pooled analysis and sometimes in other types of analysis you may need to apply an additional scaling to our weights. Our weights have a mean of 1 in each wave, which means that if combined in a pooled analysis the waves with smaller sample size will have a smaller contribution in your analysis. This includes BHPS waves and later waves (as sample size decreases with attrition). Ideally, when combining events / states over 30 years (for example) you want each year to have the same importance. To ensure this follow this example to calculate an additional scaling for your weights. For example, you are looking at job quality and therefore are pooling information from wave 2, 4, 6 & 8 as these are the waves when the questions are asked. Here is how to create a scaled weight for this analysis.

```
ge weightscaled=0
replace weightscaled=b_indpxub_xw if wave=2

ge ind=1
sum ind [aw=b_indpxub_xw] if wave=2
gen bwtdtot=r(sum_w)
sum ind [aw=d_indpxub_xw] if if wave=4
gen dwtdtot=r(sum_w)
sum ind [aw=f_indpxub_xw] if if wave=6
gen fwtdtot=r(sum_w)
sum ind [aw=h_indpxub_xw] if if wave=8
gen hwtdtot=r(sum_w)

replace weightscaled=d_indpxub_xw*(bwtdtot/dwtdtot) if wave=4
replace weightscaled=f_indpxub_xw*(bwtdtot/fwtdtot) if wave=6
replace weightscaled=h_indpxub_xw*(bwtdtot/hwtdtot) if wave=8
```

You can double check by looking at the sum of ind with weightscaled for each wave – it should be the same.
```
sum ind [aw=weightscaled] if wave==2
sum ind [aw=weightscaled] if wave==4
sum ind [aw=weightscaled] if wave==6
sum ind [aw=weightscaled] if wave==8
```

#4 - 03/08/2023 12:22 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 80

- Private changed from Yes to No

#5 - 11/30/2023 12:53 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100