

Understanding Society User Support - Support #1394

Follow up on having a “unified” weighting variable for analyzing

08/11/2020 12:46 PM - Abigail Dumas

Status:	Resolved	Start date:	08/11/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		
Description			

History

#1 - 08/11/2020 12:53 PM - Abigail Dumas

Hello Alita,

I wanted to follow up on my query regarding having a “unified” weighting variable for analysing interviews from 2002 to 2018. I remember you mentioned that people interviewed from 2002 to 2008 (BHPS) are not the same ones interviewed from 2009 to 2018 (UKHLS). You said that we are essentially looking at two different UK populations since migration has had an impact on its composition. Has there been a resolution on this?

#2 - 08/11/2020 01:00 PM - Alita Nandi

- Status changed from New to Feedback
- Assignee changed from Alita Nandi to Abigail Dumas
- % Done changed from 0 to 50

Hi Abigail,

As this is a longitudinal survey, the same sample members are interviewed every year. Part of the BHPS sample, after the BHPS ended in 2018, continued to be interviewed from onwards 2010 as part of Understanding Society. You can check the PIDP values and you will find many respondents interviewed during 2002-2008 were also interviewed from 2010 onwards. But what I meant was that if you used the longitudinal weights `bw_indin01_lw` then the weighted estimates will be representative of the 2001 UK population while if you used the longitudinal weights `w_indinub_lw` then the weighted estimates will be representative of the 2010 UK population.

Hope this helps.

Best wishes,
Alita

#3 - 08/14/2020 05:18 PM - Alita Nandi

- % Done changed from 50 to 80
- Private changed from Yes to No

Hi Abigail,

Last time we spoke you said you wanted to estimate models using fixed effects using data from all 18 waves of BHPS and 9 waves of UKHLS and wanted to know which set of weights to use. This is the response based on advice from our survey statistician.

As NI was not included in the BHPS until 2001, if you want to look at the entire period 1991-2018, you should exclude NI from all analysis. Otherwise you should analyse BHPS 1-11 and rest of the waves separately, where your estimates for the first analysis will be representative of GB, for the latter... UK.

Including NI
BHPS Waves 11-18, use `br_indin01_lw`
UKHLS Waves 1-2, use `b_indinui_lw`
UKHLS Waves 2-6, use `f_indinub_lw`
UKHLS Waves 6-9, use `i_indinui_lw`

Excluding NI
BHPS Waves 1-11, use `bi_indin91_lw`
BHPS Waves 11-18, use `br_indin01_lw`
UKHLS Waves 1-2, use `b_indinui_lw`
UKHLS Waves 2-6, use `f_indinub_lw`
UKHLS Waves 6-9, use `i_indinui_lw`

Note Stata does not allow svyset with xtreg but it allows weights and clustering.
xtreg depvar indepvars [pw=weight], fe vce(cluster psu)

Stratification can be omitted - the CIs will be slightly conservative but otherwise correct. To include correct clustering you should have the following structure: wave information nested within pidp, then within hidp (unless longitudinal analysis is used; if so hidp should be omitted), and then within psu. You can omit any levels except for psu. You can add extra levels higher up too if you like. And the weight is of course should be appropriate to your analysis.

#4 - 08/14/2020 06:17 PM - Abigail Dumalus

Hello Alita,

I'll take note of these helpful details. Would it be easier if I can attach the respective weights before merging the waves together in one longitudinal data file? Or can I program this in a syntax on Stata after merging? Waves 2 and 6 have multiple (i.e. more than one) weights, since they overlap?

UKHLS Waves 1-2, use b_indinui_lw
UKHLS Waves 2-6, use f_indinub_lw
UKHLS Waves 6-9, use i_indinui_lw

I still need to drop wave 1 of UKHLS if I'm analysing waves 11-18 of the BHPS along with waves 2-9 of UKHLS longitudinally speaking, correct?
Thanks so much. I am sorry if I keep misunderstanding any of these.

#5 - 08/17/2020 08:07 AM - Alita Nandi

My suggestion would be to add the weights to the indresp files in before merging them. But remember to merge the relevant weight to all the waves it needs to be applied to. For example, attach br_indin01_lw to bl_indresp bm_indrep.... br_indresp.

#6 - 08/30/2020 11:38 PM - Abigail Dumalus

Hello Alita,

My understanding of your weighting assignment is that Waves 2 and 6 have multiple (i.e. more than one) weights, since they overlap (i.e. 1-2, 2-6, 6-9, seen below). Is this a wrong reading from my end?

UKHLS Waves 1-2, use b_indinui_lw
UKHLS Waves 2-6, use f_indinub_lw
UKHLS Waves 6-9, use i_indinui_lw

Isn't i_indinui_lw the "combined UKHLS+BHPS+IEMB longitudinal adult main interview weight" covering waves 7-9 of the UKHLS only? Isn't f_indinub_lw the "longitudinal adult main interview weight" covering waves 3-9 of the UKHLS? I am basing these from the variable search results on the Understanding Society website.

I still need to drop wave 1 of UKHLS if I'm analysing waves 11-18 of the BHPS along with waves 2-9 of UKHLS longitudinally speaking. Wouldn't it be easier to use i_indinus_lw for UKHLS Waves 2-9 as this is the "longitudinal adult main interview weight" covered in the indresp files for waves 2-9 of the UKHLS? Please let me know if I am on the right track.

As you have noted previously: if I used the longitudinal weights bw_indin01_lw then the weighted estimates will be representative of the 2001 UK population while if I used the longitudinal weights w_indinub_lw then the weighted estimates will be representative of the 2010 UK population. This tells me that it won't make sense to compare these distinct UK populations over time from 2002 to 2018. Does this mean that it is impossible to follow those from the 2001 UK population during the period between 2010 and 2018 since there is an entirely different set of individuals who comprise the 2010 UK population?

Thanks for you time and consideration!

#7 - 09/22/2020 12:35 PM - Alita Nandi

- Assignee changed from Abigail Dumalus to Alita Nandi

#8 - 09/23/2020 02:23 PM - Alita Nandi

- Assignee changed from Alita Nandi to Olena Kaminska

Sorry there was a typo: UKHLS Waves 1-2, use b_indinus_lw

And you can use i_indinub_lw for Waves 2-9.

And you can use i_indin01_lw for BHPS Waves 11-18+UKHLS Waves 2-9 for the BHPS sample only.

#9 - 09/25/2020 10:26 AM - Olena Kaminska

Abigail,

Just to add, if you are concerned about the differences in the populations, then you can use full samples but restrict to those who lived in this country in 1991 (or were born to people who lived here in 1991) or 2001.

And just to clarify Alita's last comment you can't use BHPS with BHPS weight alongside UKHLS with UKHLS weight - this will be wrong. You either should use all the data available with 'ub' or 'ui' weight or restrict your analysis to BHPS data and use BHPS weights.

Hope this helps,
Olena

#10 - 10/07/2020 06:06 AM - Abigail Dumalus

Hello Alita,

When I am restricting the analysis to Waves 11-18 of the BHPS, using `indin01_lw` as weighting variable results in zero weighted frequencies for life satisfaction. What am I misunderstanding here from what was recommended above? I have no problem with waves 2-9 of the UKHLS. Could it be that the respective `xx_indin01_lw` for waves 11-18 have been overwritten as missing when the `indresp` database files were merged? I was checking the data editor and found missing values for waves before wave 2 of the UKHLS. Kindly let me know what must I do to rectify this situation. Thanks so much.

Kind Regards,
Abigail

#11 - 10/08/2020 09:53 AM - Alita Nandi

- Assignee changed from Olena Kaminska to Abigail Dumalus

Hi Abigail,

It is not your mistake, it is mine. Very sorry. The names of the weight variables for BHPS have not yet been harmonised with the UKHLS naming conventions (it will be done at this release or the next). Once the names are harmonised the variable name I said will work. Basically for BHPS W11-18 the weight variable for longitudinal analysis is `br_lrwtuk1` (this will be renamed to `br_indin01_lw` later).

Again sorry about that.
Best wishes,
Alita

#12 - 10/08/2020 05:13 PM - Abigail Dumalus

Hello Alita,

If `br_lrwtuk1` is the weight variable for BHPS W11-18 for longitudinal analysis, then I have to generate a new weight variable that "unifies" this with `indin01_lw` in the meantime, don't I? As of now, using `indin01_lw` alone results in attrition of these 8 BHPS waves and the data starts from 2010 to 2018, covering UKHLS W2-9.

If I move back as far as 1991, does this mean that the respective `xx_indin91_lw` for BHPS waves before 2001 are also not yet harmonised with the `indin91_lw` variable? Can you please confirm that `lrwght` is the weight variable for longitudinal analysis for BHPS W2-18? If yes, does it make sense to "unify" this with `indin91_lw` variable for now?

#13 - 10/13/2020 04:46 PM - Alita Nandi

- Assignee changed from Abigail Dumalus to Olena Kaminska

#14 - 10/14/2020 11:30 AM - Olena Kaminska

Abigail,

The data for `indin01_lw` starts with 2001, and the data for `indin91_lw` starts in 1991. While the variable names are not harmonized between UKHLS and BHPS yet (they will be in the upcoming release) the values are harmonized - so the weights have identical concept - at the moment they are just under different names.

Hope this helps,
Olena

#15 - 10/14/2020 01:34 PM - Abigail Dumalus

Hello Olena,

Many thanks for your response. I tried generating a single variable to combine these two weighting variables. When I did, I used that to run the `xtreg` command with weighting but I keep getting this error: "weight must be constant within pidp". I used "`xtset pidp`" before this regression, by the way. Please let me know what I should be doing, instead, so that the aforementioned error is avoided. What am I missing here?

#16 - 10/14/2020 02:31 PM - Alita Nandi

- Assignee changed from Olena Kaminska to Abigail Dumalus

Hello Abigail,

If you are using xtreg you can only use one weight for the entire set of waves - as it is the same set of people you are following over that period. So, if you are using xtreg for BHPS wave 1 to UKHLS W9 - then use `i_indin91_lw` and carry it over to all previous waves. But if you are using xtreg for analysing data from BHPS waves 1 -10, then use `j_lrwtuk1` from BHPS W10 and apply that to all waves prior. Basically, you need to decide which sample you are going to analyse. You can analyse separate sets of samples - but as we discussed earlier you will need to analyse them separately.

#17 - 10/14/2020 02:43 PM - Abigail Dumalus

Hello Alita,

I am using xtreg for W11-18 of BHPS along with W2-9 of UKHLS, so I am looking at the "harmonisation" of `lrwtuk1` for the former and `indin01_lw` for the latter set of waves. I combined this under one variable but I keep getting the error above. Do I need to drop all weights variables first then merge these two weights with the longitudinal data files from BHPS W11-18 to UKHLS W2-9? I do understand that we are following the same people from 2001 to 2018. It is just frustrating that this xtreg error keeps emerging. Is there a better way around this challenge? Thanks so much for your consideration.

#18 - 10/14/2020 02:51 PM - Alita Nandi

Option 1: Append individual level files (`indresp`) from W11 of BHPS until W18, and W2 UKHLS until W9 into a long format file. Then (many to one) merge `i_indin01_lw` from `i_indresp` onto this long file.

Option 2: Append individual level files (`indresp`) from W11 of BHPS until W18 into a long format file. Then (many to one) merge `br_lrwtuk1` from `br_indresp` onto this long file. Repeat the same with UKHLS W2-W9 UKHLS `indresp` files & `i_indin01_lw` from `i_indresp`. Analyse the two parts separately.

Preference is for Option 1.

#19 - 10/14/2020 03:15 PM - Abigail Dumalus

Hello Alita,

For option 1, does this mean I can disregard the weight variable `lrwtuk1` for BHPS W11-18 and simply focus on `indin01_lw`? As you mentioned before, `lrwtuk1` corresponds with the aforementioned BHPS waves. Wouldn't I be getting zero weights for these BHPS waves when using `indin01_lw` alone when merging from the latest `indresp` data file? My understanding is that I am analysing the same people from 2001 onwards. I mentioned in a previous update that I am getting zero weights for 2001 to 2008 when I look at `indin01_lw`. Is my understanding mistaken?

#20 - 10/14/2020 03:38 PM - Alita Nandi

- *File UserForum_Issue1394_public.do added*

Hi Abigail - I am providing a Stata syntax file to show you what I mean.

#21 - 10/14/2020 04:18 PM - Abigail Dumalus

Hello Alita,

Thanks so much for the syntax file that you sent. I may be misunderstanding what you said here:

"The names of the weight variables for BHPS have not yet been harmonised with the UKHLS naming conventions (it will be done at this release or the next). Once the names are harmonised the variable name I said will work. Basically for BHPS W11-18 the weight variable for longitudinal analysis is `br_lrwtuk1` (this will be renamed to `br_indin01_lw` later)."

My understanding is that using `indin01_lw` indicates that we are following the 2001 UK population until wave 9 of the UKHLS. How should I consider `lrwtuk1` in relation to `indin01_lw` then? Is `lrwtuk1` representative of another UK population? I apologise that I am still confused why using `indin01_lw` results in zero weights for the period covering W11-18 of BHPS. I am deeply sorry for misunderstanding this one bit.

#22 - 10/14/2020 04:24 PM - Alita Nandi

No same population. This is how longitudinal weights in this survey are computed: Initial wave cross sectional weight is multiplied by non-response weight between two consecutive waves. So, `br_lrwtuk1` = bk cross-sectional weight X non-response weight to correct for non-response between Wave 11 & 12 X non-response weight to correct for non-response between Wave 12 & 13.....non-response weight to correct for non-response between Wave 17 & 18

So, if you were to use the 2001 BHPS sample and

- follow them until Wave 18, use `br_lrwtuk1`
- follow them until UKHLS W9 then use `i_indin01_lw`
- follow them until BHPS Wave 12, use `bl_lrwtuk1`
- follow them until UKHLS W2, use `b_indin01_lw`

.....

#23 - 10/13/2021 11:20 AM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*

- *Assignee deleted (Abigail Dumalus)*

- *% Done changed from 80 to 100*

Files

UserForum_Issue1394_public.do	1.03 KB	10/14/2020	Alita Nandi
-------------------------------	---------	------------	-------------