

Understanding Society User Support - Support #245

cross sectional hh weights in US w1/2/3

02/26/2014 01:37 PM - Ian Alcock

Status:	Closed	Start date:	02/26/2014
Priority:	Urgent	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		

Description

I am confused by the differences in the cross-sectional household weights available in the US a_hhresp b_hhresp and c_hhresp files. My understanding is this: in a_hhresp is a_hhdenus_xw which weights the households originating with Understanding Society (which comprise all households in this wave); in b_hhresp are b_hhdenbh_xw which weights the households originating with BHPS (and is set to 0 for households originating with Understanding Society) and b_hhdenus_xw which weights the households originating with Understanding Society (and is set to 0 for households originating with BHPS); in c_hhresp is c_hhdenub_xw which weights all households together, i.e. weights across households originating with BHPS and US. My questions: 1) Is my understanding correct? 2) If my understanding is correct, how do I weight all households in b_hhresp together (as I can do for households in c_hhresp), and how do I weight only the households originating in BHPS in c_hhresp (as I can do for households in b_hhresp). I want to do both of these things; I want to produce weighted quintiles of income in the previous month for the bhps originating households (so that the weighting increases their UK representativeness) in both b_hhresp and c_hhresp, and I want to produce weighted quintiles of income in the previous month for all available households (so that the weighting increases their UK representativeness) in both b_hhresp and c_hhresp, but I appear to be able to do only the former in b_hhresp and only the latter in c_hhresp. 3) What accounts for the difference in the cross-sectional household weights available in b_ and c_? Big Thank you in advance!

History

#1 - 02/28/2014 07:49 PM - Olena Kaminska

Ian,

Yes, your understanding is correct. There are a few comments here.

First, with wave 3 we also released a weight: b_hhdenub_xw (combined for UKHLS and BHPS) - this is in b_hhresp.dta.

Second, the weights are created from the perspective that an analyst may need to represent households in a particular wave. From wave 3 onwards we plan to provide only one weight - the combined one (for simplicity).

If I understand you correctly, your aim is to estimate quartiles and then use them in individual analysis. If you use BHPS dataset for individual analysis, you could still estimate quartiles using combined dataset. This estimate will be of quartiles in the population - you can then apply these to the dataset you need.

Let me know if you have further questions,

Olena

#2 - 03/02/2014 12:15 PM - Ian Alcock

Olena,

Thank you for drawing my attention to the delayed arrival of weight b_hhdenub_xw (for cross-sections of combined UKHLS and BHPS households) in file b_hhresp.dta as updated at the wave 3 release. This completely addresses one of the two 'missing weight' issues that I raised.

With regard to your reply to the other 'missing weight' issue that I raised, I have some comments and questions:

I note that weighted quintiles of households of BHPS origin (on income received, b_fihhmngrs) can be successfully derived using the cross-sectional weight designed to make households of BHPS origin representative of UK households (b_hhdenbh_xw):

```
xtile b_income bhps hh = b_fihhmngrs_dv [pw=b_hhdenbh_xw], nq(5)
```

The fact that weighted quintiles of the BHPS sample households are successfully derived is evidenced, I believe, by the weighted tabulation, showing the cut-points determine numbers of weighted observations in the 5 groups ranging from 1,230 to 1,237:

```
tab b_incomebhps hh [aw=b_hhdenbh_xw]
```

In contrast, the method you propose gives a very different outcome. I note that weighted quintiles of households of BHPS and US origin in combination (on income received, b_fihhmngrs) can be derived using the cross-sectional weight designed to make that combined BHPS/US sample representative of UK households (b_hhdenub_xw):

```
xtile b_incomebhushh = b_fihhmngrs_dv [pw=b_hhdenub_xw], nq(5)
```

I believe these weighted quintiles of this combined sample do not achieve weighted quintiles within the subsample of households of BHPS origin. This is evidenced, I believe, by the weighted tabulation using the weight designed to make those BHPS households representative, where the cut points determine weighted observations in the 5 groups ranging from 1,133 to 1,294, and by the weighted tabulation of the BHPS households using the weight designed to make all available households representative, where the cut points determine weighted observations in each group ranging from 1,111 to 1,360:

tab b_incomebhushh [aw= b_hhdenbh_xw]
tab b_incomebhushh [aw=b_hhdenub_xw] if b_hhorig >=3 & b_hhorig <=6

Does it matter, from an analytic point of view? I believe it does. Cross-sectional analysis of households from the BHPS incorporation sample requires the ability to establish these relativities within this sample. It seems to me that this was clearly the thinking behind the production of the b_hhdenbh_xw weight. (If I am mistaken about this, then I would be pleased to be informed of the reason why the b_hhdenbh_xw weight was produced for wave 2.) To the best of my understanding of the role of these cross-sectional household weights, I am concerned by your comment "From wave 3 onwards we plan to provide only one weight - the combined one (for simplicity)." Whilst I am sure you are correct that it is simpler, it prevents important forms of analysis. It is only with a 'c_hhdenbh_xw' that analysts can establish xtiles of households at "BHPS wave 20" as they can for BHPS wave 1-18, and also can for the bhps incorporation sample at US wave 2 (BHPS 'wave 19'). And thus my questions:

Why has the policy changed? Is it open to review and revision? Can a wave 3 equivalent to b_hhdenbh_xw, (a 'c_hhdenbh_xw') be derived? If so, how quickly can this be done, and how can I get it? Thank you for your help.

#3 - 03/03/2014 08:33 AM - Ian Alcock

Dear Olena,

Since writing, 20 hours ago, my further queries regarding the current policy on the derivation and publication of household level weights, and explaining in general terms my view about the apparent limitations this imposes on the research potential of the data, and my hopes for a rapid change in policy and practice, it has been put to me that there is an analogous, though even more problematic issue with the individual level weights. It is best if I summarise what I understand to be the case in this regard before asking a couple of further questions about this, since it may well be that I am confused about the available weights.

To the best of my understanding, the weights in the b_indresp.dta file allow the separate cross-sectional weighting of the BHPS incorporation sample, with the b_indinbh_xw weight for respondents to the main interview (with equivalents for inclusion of proxy respondents and the self-completion items), and also of the US sample, with the b_indinus_xw weight (and its parallel equivalents). But no weight for analysis of the combined sample of individuals. In contrast, the c_indresp.dta file has only weights for analysing the combined sample, c_indinub_xw (and its parallel equivalents), and no weights to allow the separate cross-sectional analysis of the BHPS and US samples.

Thus, to the best of my understanding, whereas the wave 2 data release offered weights for cross-sectional analysis of households at wave 2 only of the separate BHPS and US samples, the wave 3 data release corrected the omission of a weight for cross-sectional analysis of the combined sample of households at wave 2 by adding this, the analogous omission of a weight in the wave 2 data release for cross-sectional analysis of the combined sample of individuals at wave 2 was not corrected by adding this to the wave 3 data release.

And further, at the individual as well as at the household level, weights to allow the cross-sectional analysis of the separate samples at wave 3, as has been facilitated at wave 2, are omitted.

And thus I would like to ask the further questions: Is my understanding correct? If so, then please can a wave 2 equivalent to c_indinub_xw, (a 'b_indinub_xw') be derived and made available, and can wave 3 equivalents to b_indinbh_xw and b_indinus_xw (and their equivalents), ('c_indinbh_xw and c_indinus_xw' etc) be derived and made available? Also, please tell me the name of the person who leads on derivation and release of weights. Thank you for your help.

#4 - 03/18/2014 07:25 PM - Olena Kaminska

Thanks for your questions. The following should answer your questions above.

1. From statistical perspective by using sample one estimates something in the population. There is one population (per wave) regardless whether you are looking at BHPS, UKHLS or combined sample. Theoretically you can represent the population using any of the weights provided: longitudinal, cross-sectional for BHPS, cross-sectional for UKHLS or combined cross-sectional;
2. We recommend combined cross-sectional for two reasons: i) it has higher statistical power (higher sample size); ii) it is newer - updated in comparison to BHPS, and therefore includes recent immigrants (e.g. those moving to England since 1991); iii) it also to some extent includes new immigrants since 2010 (through TSMs). That's why we will continue combined cross-sectional weight.
3. It is extremely important that you use weights with correct sample. The simplest way is to not restrict through 'if' option to any sample - the weight will restrict to correct sample itself. But using a weight created for combined UKHLS+BHPS sample with only BHPS sample will indeed give wrong results.

Finally, it is relatively easy to create cross-sectional weights for BHPS - just use the standard weight-share method (look for it in google).

Thanks,
Olena

#5 - 03/20/2014 12:14 PM - Ian Alcock

Thank you for your reply. I note that you again recommend that I use the combined cross-sectional weight. I could explain why there are good reasons for me to carry out a weighted cross-sectional analysis of the BHPS sample. However, the possibility of this is acknowledged in the document "Weighting Strategy for Understanding Society" by Peter Lynn and Olena Kaminska, (Understanding Society Working Paper Series No. 2010 – 05) which states, "Each study population could be represented by just the new general population sample ("UKHLS-GPS") or the BHPS sample, or both. There are arguments for and against combining the two samples; the appropriateness of doing this may depend on the specific analysis being considered." (p.9)

On this basis the weighting strategy was stated by Peter Lynn and Olena Kaminska: "We therefore propose to produce weights for each sample separately and for the combination (from wave 2 onwards)". (p.9)

I think the policy proposed in 2010 was very sensible. One aspect of my questions which you have not addressed in your replies concerns how the envisioned strategy described in 2010 developed into the current practice, which I think can be fairly summarised as, 'Despite all earlier indications to

the contrary, we will not produce weights for each sample separately but only for the combination, and the demand that we fully anticipate from users for weights for the separate samples will be addressed with the suggestion that those users learn how to derive these weights themselves'. I asked about how that decision had come to be made because I would like to be able to argue for a revision of that policy. After all, I am sure you will agree that if the research potential of the data were in any instance limited by this 'available weights hurdle', that would be awful, given the many millions of pounds of public money which have been spent on acquiring the data, and given too how "relatively easy" it is to create weights for the separate samples, (especially, no doubt, for someone with expertise and experience in this area). I am sure that a lot of thought will have gone into the new "DIY weighting policy" as it is such a radical change from what was anticipated. The document "Weighting Strategy for Understanding Society" by Peter Lynn and Olena Kaminska, (Understanding Society Working Paper Series No. 2010 – 05) states "We believe it is inevitable that a large number of weight variables will be provided to users." (p.14)

Thanks, Ian

#6 - 03/20/2014 12:36 PM - Redmine Admin

- *Status changed from New to Closed*
- *% Done changed from 0 to 100*

Thanks for all the comments.

#7 - 11/10/2015 11:04 AM - Gundi Knies

- *Assignee set to Olena Kaminska*
- *Target version set to X M*