

Understanding Society User Support - Support #1979

Using UKHLS for financial year poverty rates

10/02/2023 05:40 PM - Sam Tims

Status:	Resolved	Start date:	10/02/2023																											
Priority:	Normal	% Done:	100%																											
Assignee:	Understanding Society User Support Team																													
Category:	Income																													
Description																														
Dear USOC team,																														
I am working on a project exploring the prevalence of mental health for people living in in-work poverty and have some questions on poverty rates and weighting that I hope you can help with. My analysis only uses UKHLS wave 3 onwards. My questions are below.																														
Thank you in advance, Sam																														
1. For this analysis I am interested in financial years rather than waves. In constructing my financial years I have followed the process set out in the user guide and in this forum (months 4 to 15 from wave n, months 16 to 24 from wave n-1 and months 1-3 from wave n+1). I am aware that I need to adjust my cross-sectional weights to account for attrition in a similar manner as is set out in box 1 of the weighting FAQ. I note that box 1 includes <code>sum ind [aw=b_indpxub_xw] if b_month>=1 & b_month<=12</code> but this old issue (https://iserredex.essex.ac.uk/support/issues/1221) implies the line should end with 24 instead of 12. Could I clarify which is correct?																														
2. As I am combining three waves in each financial year, I'm currently adjusting to the average weighted total of the three waves which I have seen suggested elsewhere in this forum. Previously I was adjusting by a factor such as <code>wave_1_total / wave_2_total</code> but switched to this method because of slight increases in sample size in some waves. Can I confirm this is an appropriate method for the adjustment? Please note that I have assumed the answer to question 1 is 24, and therefore I am accounting for the entire wave weight, rather than just the specific months I am combining:																														
<pre>gen count = 1 summarize count [aw = `wave_1'_hh_weight] local `wave_1'_total = r(sum) ... local avg = (`wave_1'_total' + `wave_2'_total' + `wave_3'_total') / 3 gen weight = weight * (`avg' / `wave_1'_total') if `wave'_month >= 16 & `wave'_month <= 24 replace weight = weight * (`avg' / `wave_2'_total') if `wave'_month >= 4 & `wave'_month <= 15 replace weight = weight * (`avg' / `wave_3'_total') if `wave'_month >= 1 & `wave'_month <= 3</pre>																														
3. For part of my analysis I am comparing poverty rates between this data and the FRS, which I expect to be similar but not identical. This is in part because of methodological differences and also because I do not have access to the Council Tax data in Understanding Society so my results will not be perfect. To generate my poverty rates I have followed worksheet 4 in the into to USOC using STATA. However my poverty rates from USOC are quite a bit below the FRS and also below figure 10 here (https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2019-08.pdf). Have other users reported a similar issue? I'd be happy to share my code if useful.																														
<table><thead><tr><th>Fin_Year</th><th>FRS</th><th>US</th></tr></thead><tbody><tr><td>2012/13</td><td>15.4%</td><td>12.9%</td></tr><tr><td>2013/14</td><td>15.2%</td><td>13.0%</td></tr><tr><td>2014/15</td><td>15.9%</td><td>13.1%</td></tr><tr><td>2015/16</td><td>16.3%</td><td>12.6%</td></tr><tr><td>2016/17</td><td>16.2%</td><td>12.5%</td></tr><tr><td>2017/18</td><td>17.1%</td><td>13.0%</td></tr><tr><td>2018/19</td><td>16.7%</td><td>13.7%</td></tr><tr><td>2019/20</td><td>17.8%</td><td>14.0%</td></tr></tbody></table>				Fin_Year	FRS	US	2012/13	15.4%	12.9%	2013/14	15.2%	13.0%	2014/15	15.9%	13.1%	2015/16	16.3%	12.6%	2016/17	16.2%	12.5%	2017/18	17.1%	13.0%	2018/19	16.7%	13.7%	2019/20	17.8%	14.0%
Fin_Year	FRS	US																												
2012/13	15.4%	12.9%																												
2013/14	15.2%	13.0%																												
2014/15	15.9%	13.1%																												
2015/16	16.3%	12.6%																												
2016/17	16.2%	12.5%																												
2017/18	17.1%	13.0%																												
2018/19	16.7%	13.7%																												
2019/20	17.8%	14.0%																												
4. Finally is there a single variable to flag if anyone in a household is in receipt of income-related benefits? I haven't spotted such a flag so have constructed my own but it would be good to make sure I haven't missed anything.																														

History

#1 - 10/05/2023 08:46 AM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee set to Olena Kaminska
- % Done changed from 0 to 10
- 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.

Best wishes,
Understanding Society User Support Team

#2 - 10/09/2023 12:11 PM - Olena Kaminska

Sam,

Thank you for your question. I understand you want to scale your weight. The answer below may help you with this, and I suggest you follow this simpler example.

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 bwtot=r(sum_w)
sum ind [aw=d_indpxub_xw] if wave=4
gen dwtot=r(sum_w)
sum ind [aw=f_indpxub_xw] if wave=6
gen fwtot=r(sum_w)
sum ind [aw=h_indpxub_xw] if wave=8
gen hwtot=r(sum_w)
```

```
replace weightscaled=d_indpxub_xw*(bwtot/dwtot) if wave=4
replace weightscaled=f_indpxub_xw*(bwtot/fwtot) if wave=6
replace weightscaled=h_indpxub_xw*(bwtot/hwtot) 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
```

#3 - 10/09/2023 12:11 PM - Olena Kaminska

- Assignee changed from Olena Kaminska to Understanding Society User Support Team

#4 - 10/09/2023 02:13 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback
- % Done changed from 10 to 50

Hello Sam,

About your third question: after you construct the weights as described by Olena, could you please confirm if you have used the same definitions in constructing the share of households below the poverty line (60% of median net equivalised hh income) as in the paper, described in section 6.2?

About your fourth question: The variable, w_fihhmsben_dv, shows the social benefit income that is received by the household. You could use this to produce a flag variable. Is this what you are interested in?

Best wishes,
Alita

#5 - 10/11/2023 02:32 PM - Sam Tims

Hi Both,

Olena, thank you for your help. I have adjusted my weights as instructed and I'm getting the same weighted total in each set of waves.

Alita, On question 4 the variable fihhmnsben_dv contains all benefits, which while useful for some purposes, is not the same as income-related benefits. E.g. DLA & PIP are independent of someone's earnings/income. But that is ok, I can construct the flag variable I am interested in.

On poverty rates, I am still finding that the USOC data is on average around 2 percentage points lower than the USOC results in figure 10. I have created financial year versions of the USOC but I have not removed recent migrant households from the FRS as I am interested in this difference between the datasets. So I believe my methodology is broadly the same as outlined in 6.2, except for which income variable has been used. I have used fihhmnet1_dv as I do not currently have access to the special license version of the data. I imagine 6.2 uses fihhmnet3_dv (post Council Tax) to match the HBAI approach. Is it likely that using fihhmnet3_dv (via the special license) would sufficiently close the gap?

Thanks again,
Sam

#6 - 10/13/2023 09:01 AM - Understanding Society User Support Team

Thanks for the response Tim. I will pass your questions on to our income team.
Best wishes,
Alita

#7 - 10/17/2023 02:11 PM - Understanding Society User Support Team

- % Done changed from 50 to 80

We have heard back from the income team. Their response:

You could use the datafile "income.dta" to see who reports the specific benefits you are interested in.

If you want to compare the two distributions, yes if you don't deduct council tax you will have different numbers from the ones in the paper and that could explain the difference. You should also check that you are using the same deflators and the same rescaling of the equivalence scales.

Hope this helps.
Best wishes,
Alita

#8 - 11/30/2023 01:32 PM - Understanding Society User Support Team

- Category set to Income

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100