

Understanding Society User Support - Support #2075

Using UKHLS to look at trends across calendar months

03/13/2024 03:36 PM - James Laurence

Status:	Feedback	Start date:	03/13/2024
Priority:	Normal	% Done:	80%
Assignee:	Olena Kaminska		
Category:	Weights		

Description

Hi there,

I am interested in looking at calendar month trends in whether someone wants to move home or not (which is available in every wave): lkmov. Ideally, I would like to look at trends using all waves (1-13). However, if it is easier to look at trends from some other start point, e.g.. 2016 or 2017, then I am flexible. I am also flexible as to whether the BHPS sample is included or not. This will be cross-sectional analysis, so I hope to treat each calendar month as a cross-section (I won't be doing any longitudinal analysis).

I have been reading the helpful notes on 'Running analysis on a calendar year or month' (<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/how-to-use-weights-analysis-guidance-for-weights-psu-strata/>). However, I just had some questions and was hoping to see if where I'd got to so far looked right.

I have been using the w_month and wave variables to generate a new date variable of year-month. To capture calendar year, I have used the wave and w_month variables in the following manner:

```
gen year = 2009 if wave==1 & (month>0 & month<13)
replace year = 2010 if wave==1 & (month>12 & month<25)
replace year = 2010 if wave==2 & (month>0 & month<13)
replace year = 2011 if wave==2 & (month>12 & month<25)
replace year = 2011 if wave==3 & (month>0 & month<13)
...
replace year = 2021 if wave==13 & (month>0 & month<13)
replace year = 2022 if wave==13 & (month>12 & month<25)
```

To measure calendar month, I have recoded the w_month variable, combining the two monthly measures into one. So, in the w_month variable, it tells us whether someone was sampled in January in the year 1 sample or January in the year 2 sample. I've now combined these into a single category of whether someone was sampled in January. For example, 'jan yr1' and 'jan yr2' are now just 'jan'; 'feb yr1' and 'feb yr2' are now just 'feb', etc.

With these new calendar year and calendar month variables, I have now created a new measure of calendar year-month, which looks like this (I hope this is correct so far):

```
2009 Jan = 1
  2009 Feb = 2
  2009 Mar = 3
  2009 Apr = 4
  2009 May = 5
  2009 June = 6
  2009 July = 7
...
2022 June = 162
2022 July = 163
  2022 Aug = 164
  2022 Sep = 165
  2022 Oct = 166
  2022 Nov = 167
  2022 Nov = 168
```

I understand that whatever weight I choose to use I need to correct it due to Northern Ireland only being sampled in issue month 1-12 (and not 13-24). Therefore, I will apply the following adjustment to the weight (gen adj=1, replace adj=0.5 if w_country==4, gen weight=w_xxyyus_lw*adj 8) as outlined in the online notes.

However, where I've become a little lost is what weights to initially use. In the notes, it states due to exceptions in sample selection 'we recommend use of the us_lw weight in analysis'. Given my intention to look at calendar months up to wave 13, does this mean I should use the m_indpxus_lw weight? Is this the case, even if I just want to look at the data cross-sectionally (treat every calendar

month as a cross-sectional picture of lkmov)? Because it seems that if I use m_indpxub_lw then it substantially reduces the sample size (due to these longitudinal weights requiring someone to have participated in every wave). Is it possible to use the cross-sectional weights for my aims, while excluding the BHPS and IEMB, as is suggested that one needs to do for this kind of calendar month analysis in the online notes? Or, do I need to use longitudinal weights for my intended analysis?

I was also just trying to get my head around the issue of scaling discussed in the online notes: 'The weights provided are not designed directly for pooling data across waves as they are scaled to a mean value of 1.0 within each wave, and therefore produce different weighted sample sizes in each wave', under the section 'Pooling data from different waves for cross-sectional analysis.' Firstly, I just wanted to confirm this applies to my case of doing monthly trends?

And secondly, if so, from what I can see, the syntax kindly provided is intended to produce an accurate weight to look at the variable jbstat for the calendar year 2011, using months 13-24 of wave 2 and 1-12 of wave 3. At the end, we get the weight variable weight2011, to use for weighting calendar year 2011. In my situation, I would like to do a longer running trend of values of lkmov by months. Would I need to create these weights for each calendar year I look at? So, for 2014, I would need to create a new cross-sectional weight using e_indpxub_xw and f_indpxub_xw (waves 5 and 6). For 2015, I would need to create a new cross-sectional weight using f_indpxub_xw and g_indpxub_xw (waves 6 and 7). For 2016, I would need to create a new cross-sectional weight using g_indpxub_xw and h_indpxub_xw (waves 7 and 8). And to follow this all the way to my last calendar year. Then, to look at monthly trends, treating the data as pooled cross-sectional, I would have my data in long-format and have a new weight variable made up of all these new calendar year weights I've created?

I was also wondering if it would be possible to include monthly lkmov data from the calendar year 2022 (using wave 13 of the UKHLS mainstage). As I understand things, previous calendar years (e.g., 2018) are composed of samples from two waves (waves 9 and 10 of the mainstage). However, for the calendar year of 2022, it is only composed of the sample from wave 13. Is it still possible to look at calendar month trends in lkmov for 2022? If so, would I need to make other sample restrictions to the other calendar years, for example, drop the IEMB sample from the trends? And would I need to make other adjustments to the weights? Or, is it not possible yet to look at monthly trends until wave 14 comes out? I think from the online notes this is mentioned: 'The analysis sample is only representative when all 24 monthly samples are combined in equal measure.' Does this point refer to my question?

I am also interested in potentially looking at quarterly trends (Jan-Mar, Apr-Jun, etc.), instead of monthly trends (using the x_quarter variable). To do so, can I take the same approach as above? So, create a new time variable which is years divided into quarters (e.g., 2013 Jan-Mar, 2013 Apr-Jun, 2013 July-Sep, 2013 Oct-Dec, 2014 Jan-Mar, 2014 Apr-June...2022 Jul-Sep, 2022 Oct-Dec). Do I need to do anything different with the weights?

I hope this all makes sense.

Thanks so much in advance.

James

History

#1 - 03/13/2024 04:11 PM - Understanding Society User Support Team

- Category set to Weights
- Status changed from New to In Progress
- Assignee changed from Understanding Society User Support Team 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 - 03/14/2024 12:03 PM - Olena Kaminska

James,

Thank you for your question. From your description I find that you are technically not using 'calendar' month but 'sample' month, which makes it much easier for weighting. This means that while some people (late respondents) answer interviews later, you are including them in January months if they are sampled in January month. Also, you are conducting cross-sectional analysis. In this situation:

1. You should always combine two sample years (soe January in year 1 with January in year 2). In this situation you won't need correction for NI (so don't divide it by 2), and you can use IEMB and BHPS sample. So basically use our data as it is (don't do any selection on samples), and use one of our weights. You can use ub or ui weights (but the same weight for both sample years).
2. Technically you could use longitudinal weights, but there is no such need for your analysis. I suggest xw weights.
3. On scaling. Scaling is important if you are analysing all data in one model, e.g. multilevel model. In such situation you want to avoid some years

contributing more than others, and scaling is necessary. But scaling is not needed if you just want to graph a time trend, i.e. if you want to estimate separate proportions for each time period. Scaling won't make any difference in this situation and won't be needed.

4. Calendar year 2022 should be find as (I think) uses wave 12 year 2, and wave 13 year 1. I believe wave 13 is released. But calendar year 2023 is trickier in terms of weighting. I suggest you use calendar year release (which we are working on at the moment), and this will be shaped and have weights ready for you. Calendar year release precedes full mainstage release, but includes questions only from core questionnaire. Weighting is tricky if you want to include calendar months based on year 1 only, due to uneven sampling in year 1. Correction for NI, and potential exclusion of some samples may be advisable in this situation.

5. Quarterly trends will follow the same logic, and if you base it on sample months weighting is much more straightforward, and everything mentioned above applies.

Hope this helps,
Olena

#3 - 03/15/2024 03:38 PM - James Laurence

Hi Olena,

Thank you for the detailed reply. It's really helpful. Apologies for mixing up calendar month and sample month but you are absolutely right that I will be using sample month.

Best wishes,

James

#4 - 03/26/2024 10:58 AM - Understanding Society User Support Team

- *Status changed from In Progress to Feedback*

- *% Done changed from 10 to 80*