

Understanding Society User Support - Support #1872

Pooling waves for longitudinal analysis

03/02/2023 04:13 PM - Spencer Thompson

Status:	Resolved	Start date:	03/02/2023
Priority:	Normal	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		
Description			
Hello,			
I am doing some analysis on informal care. I want to track people's incomes for several years after they start doing care work.			
So far I have used a single wave to set the cohort, e.g. I look at people who start care in wave 4 (t) and see what happens to their income in wave 5 (t+1), wave 6 (t+2), etc. However, I'd like to work with a larger sample and not rely on a single year. For example, I would pool people who start care in waves 4, 5, or 6, and treat all of these as t. For these respective cohorts, t+1 would then be waves 5, 6, 7, and so on.			
I have seen the documentation on adjusting weights in order to pool waves for cross-sectional analysis. How can I do this for my longitudinal analysis?			
Many thanks.			

History

#1 - 03/08/2023 10:56 AM - Spencer Thompson

Hi, I think I've figured this out, thanks.

I realised I don't need to change the longitudinal weights since all my observations are independent, i.e. separate people.

#2 - 03/08/2023 12:22 PM - Understanding Society User Support Team

- Category set to Weights

- Status changed from New to In Progress

- Assignee set to Olena Kaminska

- % Done changed from 0 to 10

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

#3 - 03/08/2023 12:22 PM - Understanding Society User Support Team

- Private changed from Yes to No

#4 - 03/08/2023 12:26 PM - Olena Kaminska

Dear Spencer,

This should answer your question:

Pooling can be cross-sectional or longitudinal. Theoretically, you will be combining 'separate samples of events / states' each of which will have the corresponding weight.

If you are just interested in events (and what happened at the same time / wave) you are looking at pooling cross-sectional information. For this create a new weights variable `new_weight`, and give it a value of the cross-sectional weight from each wave (e.g. `new_weight=a_indinus_xw` if `wave==1`; `new_weight=b_indinub_xw` if `wave==2` etc.)

Alternatively you may be interested in what happens before and / or after a particular event (e.g. studying work pattern for 3 years before birth of a first child and 3 years after for new mothers). In this situation you need to choose a longitudinal weight from the last wave in your analysis for each combination of waves (e.g. for birth at wave 3 where we observe waves 1-6, the weight will be `f_indinus_lw`; for birth at wave 4 with information in the model of waves 2-7 –it will be `g_indinub_lw` etc.).

Hope this helps,

Olena

#5 - 03/08/2023 01:54 PM - Spencer Thompson

Hi Olena,

Thanks, this is really helpful. This is indeed what I was trying to do:

Alternatively you may be interested in what happens before and / or after a particular event (e.g. studying work pattern for 3 years before birth of a first child and 3 years after for new mothers). In this situation you need to choose a longitudinal weight from the last wave in your analysis for each combination of waves (e.g. for birth at wave 3 where we observe waves 1-6, the weight will be f_indinus_lw; for birth at wave 4 with information in the model of waves 2-7 –it will be g_indinub_lw etc.).

However, why not just use the latest wave for everyone? To use your example - and assuming there are no more groups than what you mention, so overall we have data from wave 1 to wave 7 - why not just use the wave 7 weight for everyone?

If I take your approach, then if I want to look at studying patterns in the third year after the birth of a first child, I would be combining f_indus_lw (the first cohort) and g_indub_lw (for the second cohort), and similarly for the other time periods (2, 1, 0, -1, -2, -3). Wouldn't that be bad practice? Why not just use g_indub_lw for both cohorts?

#6 - 03/08/2023 02:42 PM - Spencer Thompson

P.S. - I should have clarified, I am not using wave 1 in my work. I mention this because in your example there was a distinction between indus and indub, but since I am only using waves 2 onwards it would only be indinub, if I've understood correctly.

To make it more concrete, I looking at people who start doing care work in waves 4, 5, and 6. I want to track their earnings over time, in years -2 (two years before starting care work), -1, 0 (year of starting care work), 1, 2, 3, 4, and 5 (five years after starting care work). So, for the cohort starting care work (t = 0) in wave 4, I need at least waves 2-9, for the cohort starting in wave 5 I need 3-10, and for the cohort starting in wave 6 I need 4-11.

So the question is whether I just use k_indinub_lw for all three cohorts (what I've done) or whether I use i_indinub_lw, j_indinub_lw, and k_indinub_lw respectively. I suppose the latter option would increase the sample as it would not be corrected for attrition in years that I don't require, but it would involve combining weights as I allude to above.

#7 - 03/09/2023 03:35 PM - Olena Kaminska

Dear Spencer,

Yes, you are completely right. Both approaches will be correct and should give you similar results. The second approach will give you higher sample size, so more statistical power to detect significant effects. In your particular situation the sample size difference is not very large as you use so many waves in your analysis, so you could choose the first option (only one weight), and only switch to the second option if you have a marginally significant result...

Hope this helps,
Olena

#8 - 03/22/2023 11:56 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

#9 - 11/30/2023 12:54 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 10 to 100