

Understanding Society User Support - Support #2083

Creating a tailored longitudinal weight for unbalanced panel data

04/10/2024 05:53 PM - Isabelle Munier

Status:	In Progress	Start date:	04/10/2024
Priority:	Normal	% Done:	0%
Assignee:	Olena Kaminska		
Category:			
Description			
Hello,			
<p>I am performing a multilevel growth curve model, using STATA's mixed command, in which repeated measures of wages (level 1) are nested within individuals (level 2). I am using waves 6, 7, 8, 9, 10, 11, 12 and 13, since I'm particularly interested in migrant groups. The model that I have chosen is flexible in the sense that it allows for partially missing data and unbalanced panels. My aim is to estimate the effect of over-education on wages, and how this differs by migrant status and gender, over time. It is not a pooled cross-sectional analysis, since I'm interested in individual trajectories over time. I understand that I must use weights to correct for unequal selection probabilities, non-response and attrition, however, the longitudinal weights provided by Understanding Society require balanced panels. Since the mixed command in STATA is not compatible with the svyset command, and pweights must be positive to perform the analysis, there is no other way than dropping everyone that did not participate across all waves. In my case, if I use the appropriate longitudinal weight m_indinui_lw, my sample size is reduced from approximately 8000 respondents to 3000 respondents. As I am doing three-way interactions in my model, it requires sufficiently large sample sizes in each group. I see that my estimates from weighted analysis (using m_indinui_lw) and non-weighted analysis are fairly similar, however the estimates turn insignificant when applying weights, plausibly due to the small sample sizes in each group. I have looked through previous inquiries regarding weights and unbalanced panels here in the support forum, and looked through the training material on how to create your own tailored weights. However, I have not found a solution to my specific problem. In the example given in the Open Essex course, Module 5, it is shown how to create a longitudinal weight for responses at waves a, d and g specifically. In this case, one can predict probabilities of participating in wave a, d, and g conditional on non-zero weights in wave a (adding relevant covariates). In my case, the only requirement is that the individuals have participated in at least three waves, since Growth curve modeling requires a minimum of three repeated measures to estimate growth curves. When I have attempted to predict probabilities of participating in all my included waves conditional on non-zero weights in wave 6 I still end up with many 0 weights... would greatly appreciate any recommendations on how to proceed!</p>			

History

#1 - 04/11/2024 08:07 AM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee changed from Understanding Society User Support Team to Olena Kaminska
- 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 - 04/17/2024 03:01 PM - Olena Kaminska

Isabelle,

My understanding is that while your model allows for missingness, it does not correct for it. Is this right?

If the model corrects for missingness / or makes sensible assumptions about it, with which you are happy, you don't need to worry about nonresponse part due to attrition, and can start with wave 1 weight (or wave 2 weight for bigger sample size). The attrition correction will then be assumed to be accounted for by the model itself. No further adjustment would be needed.

If the model can't correct for attrition, you will have to do it yourself. My suggestion would be to use a tailored weight, and correct for attrition from wave 1 (or wave 2 for larger sample size) where 'response' will be all people who are included in your model. Did you look into our course already?: <https://www.understandingsociety.ac.uk/help/training/creating-tailored-weights/>

Hope this helps,
Olena

#3 - 04/19/2024 08:55 AM - Isabelle Munier

Hi Olena!

To my understanding, the multilevel model can handle missingness under the assumption that its missing at random (MAR). Thank you for your input. Just to clarify, my study period begins at Wave 6. You wrote that I should correct for attrition from Wave 1 or Wave 2, is this correct even when my study period starts at Wave 6 and ends at Wave 13? In which case, I should choose a cross-sectional weight as my start weight and adjust for mortality and emigration, correct?

Second, when you write that my "response variable" will be everyone included in my model, would that entail everyone in my final estimation sample? Meaning that the response variable would take the value of 1 if my sample equals 1 and zero otherwise?

I have looked at the course material, and it seems rather straight-forward how one corrects for attrition and non-response between specific waves, using inverse probabilities. For instance, having the response variable being equal to 1 if respondents have zero weights at the start of the observation and have given complete interviews at the waves of interest (e.g., a, d, and g). In my case, the definition of the response variable becomes more tricky. I could have a response variable being equal to 1 if respondents have zero weights in wave 6 and have given full interviews at wave 7,8,9,10,11,12 and 13. But I do not see how this weight would be any different from using the existing longitudinal weight for wave 13, `m_indinui_lw`.

Aside attrition, if I want to correct for unequal selection probabilities, can I use the UKHLS design weight `psnenuus_xd`?

#4 - 04/19/2024 01:21 PM - Olena Kaminska

Isabelle,

I suggest you start with wave 6 weight as it includes IEMB sample that has many immigrants and model from there. Starting with design weights would not be advisable unless you limit your analysis to GPS+EMB sample. You can then simply follow our example in creating your own tailored weight. Please always use longitudinal weight as your base weight. Cross-sectional weights are not suitable for this.

For response variable everyone if you estimation model will have a value of 1 (respondent) and 0 otherwise. My understanding is that you model does not require response in each wave, only some, so you should have larger sample with your tailored weight.

Hope this helps,
Olena