Understanding Society User Support - Support #679

paired balanced panel in UKHLS/BHPS

12/08/2016 10:11 AM - Glenna Nightingale

Status:	Closed		**Start date:**	12/08/2016
Priority:	High		**% Done:**	100%
Assignee:	Glenna Nightingale			
Category:	Weights			

Description

Hello,
I am modelling data from both the UKHLS and BHPS and my dataset is constructed so that I can compare pairs of waves.
The format of my dataset when simplified has the following columns (jpeg is attached also):

1. ID (BHPS pid,UKHLS pidp)
2. var A at wave t
3. var A at wave t-1
4. Other covariates
5. Weights

1. I have used the cross sectional weight at wave t as the weight for each individual row. I have used the weights for the original sample until wave 8 then used those for inclusion of the boost of Scotland and Wales from wave 9 until wave 10, then used that for including Northern Ireland
2. I am however, wondering whether it is better to use a longitudinal weight for each row - considering the dataset as a paired balanced panel.
   What are your thoughts on this?
3. I have started considering constructing special weights for this - for comparing pairs of waves in the BHPS and UKHLS. Any suggestions on this are very welcome. Hoping to make my code available to others who might be interested

With best regards and thanks,
Glenna

History

#1 - 12/08/2016 10:12 AM - Glenna Nightingale

I have used the cross sectional weight at wave t-1 as the weight for each individual row. I have used the weights for the original sample until wave 8 then used those for inclusion of the boost of Scotland and Wales from wave 9 until wave 10, then used that for including Northern Ireland

#2 - 12/08/2016 10:17 AM - Victoria Nolan

- Status changed from New to In Progress

- Assignee changed from Olena Kaminska to Glenna Nightingale

- % Done changed from 0 to 10

- Private changed from Yes to No

Dear Glenna,

Many thanks for your enquiry - I will pass this on to our weighting team for a reply.

Best wishes, Victoria.

#3 - 12/08/2016 05:00 PM - Peter Lynn

Glenna,

I think that using the cross-sectional weight for either t or t-1 is reasonable (but not a mixture of them), for reasons that are explained in this note about pooloing: https://www.understandingsociety.ac.uk/support/issues/494.

Most of the issues discussed in that note are relevant to your analysis. The one additional point is that you are using pairs of waves, so you can think of this as pooled cross-sectional analysis with one extra step, which is that some of the cross-sectional respondents at each wave t are dropped from your analysis due to not being observed at t-1. You could therefore model this as an additional step of non-response and use the model-predicted probabilities to be observed at t-1 as an adjustment to the CS weight, i.e.:

Create the data set of all wave t respondents (pooled waves) and create a 0/1 indicator of whether or not they also responded at t-1 (i.e. whether the record can be used in your analysis). Model this indicator based on relevant respondent characteristics (e.g. a logistic regression). This will give you a

predicted probability for every respondent of the t-1 information being present. Call this P. You then need to adjust the wave t CS weight by multiplying it by 1/P for all the records that can be included in your analysis.

Hope that helps,

Peter

#4 - 12/14/2016 11:47 AM - Victoria Nolan

- Status changed from In Progress to Feedback

- % Done changed from 10 to 90

#5 - 01/04/2017 01:31 PM - Victoria Nolan

- Status changed from Feedback to Closed

- % Done changed from 90 to 100

Files

Presentation1.jpg	47.7 KB	12/08/2016	Glenna Nightingale