

## Understanding Society User Support - Support #1794

### Weights

10/27/2022 10:34 AM - Caroline Kienast von Einem

<b>Status:</b>	Resolved	<b>Start date:</b>	10/27/2022
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>	Olena Kaminska		
<b>Category:</b>	Weights		
<b>Description</b> Hi team,  I have a question about applying survey weights: When I am applying the weights do I have to do this via specific commands e.g. svy prefix in stata or is it also possible to multiply my variable of interest with the weight to create a new weighted variable that I could then use alongside commands that cannot be combined with survey weights directly.  Thank you for your help.  Best wishes, Caroline			

### History

#### #1 - 10/27/2022 12:38 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 50

Hello,

If you want to correctly compute standard errors taking into account the sample design then you should use svyset. If you just want to apply weights then you can use the [pweight = name of weight variable] option that most Stata estimation methods allow. What is the estimation method you want to use which does not allow svyset or pweight option?

Best wishes,  
Understanding Society User Support Team

#### #2 - 10/27/2022 12:38 PM - Understanding Society User Support Team

- Private changed from Yes to No

#### #3 - 10/27/2022 12:59 PM - Caroline Kienast von Einem

Hi,

this is helpful thank you.

I am looking to use generalized estimating equations (in stata command: xtgee) but there does not seem to be a way to use weights. Unfortunately the multi-level approach using svy:melogit does not converge.

Best wishes,  
Caroline

#### #4 - 10/27/2022 01:08 PM - Understanding Society User Support Team

Stata help for xtgee says "iweights, fweights, and pweights are allowed; see weight. Weights must be constant within panel." This means you can use the "[pweight = name of your weight variable]" option but the weight must be constant within a panel, i.e., the weight cannot change for a person over time. So, you cannot use cross-sectional weights. But for longitudinal analysis the correct weight is the longitudinal weight from the last wave that you are including in your analysis, and so that will not change for an individual over time.

#### #5 - 10/27/2022 01:15 PM - Caroline Kienast von Einem

Hi,

thank you for your quick response.

I am hoping to do cross-sectional analysis. I am interested in differences between those moving home and those not moving. I am looking at the time

period between wave 3 and 6 and want to explore these difference prior moving (wave 3) and then post move (wave 6) in two regression models.

Could you help me understand why the weight within a panel would be constant for longitudinal weights but not cross-sectional?

Can you think of any alterntaive for me to perform a weighted analysis that accounts for the household clustering and lets me analyse predictors of the outcome (housemove yes/no).

Best wishes,  
Caroline

#### #6 - 10/27/2022 02:29 PM - Understanding Society User Support Team

Cross-sectional weights make the sample estimates generalizable to the population of that time period while longitudinal weights make the sample estimates generalizable to the intial population that has been followed over time. If you are doing a longitudinal analysis you are basically trying to estimate changes over time for the intitial population. The longitudinal weight basically accounts for bias due to design (unequal selection probability), initial wave non-response and all wave-on-wave non-response since then up until the last wave you are including in your analysis. So, it will be the same for each person over time.

Based on the information you have provided, it seems that you are going to use information before Wave 3 and after Wave 6, even if your key variable is moved or didn't move between Waves 3 & 6, you are including those who have responded in all waves. Which means you need to use longitudinal weight from the last wave you are including in your sample.

But if you are only using information collected in waves 3-6 and pooling that information, so not using any information about a person from their past or future interviews then you can use cross-sectional weights. There is more information about this in the Weighting FAQ ([https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/weighting\\_faqs.pdf](https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/weighting_faqs.pdf)).

Stata allows only longitudinal weights for xtgee as it is a panel data method. If you have more Stata related questions please check Statalist (<https://www.statalist.org/>) and see if someone has asked this question or you can post your own question. In the meantime I will assign this issue to our Survey Statistician and she may provide you with more help.

#### #7 - 10/27/2022 02:29 PM - Understanding Society User Support Team

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

#### #8 - 10/27/2022 04:27 PM - Caroline Kienast von Einem

Hi,

thank you very much!

Maybe I should explain this better. Let's just focus on the first model to not make it even more complicated, I want to know if people who move are different from those who do not. I am looking cross-sectionally at wave 3 data as my baseline data BUT I am including a derived variable 'move' as my outcome which is based on variable 'X\_addrmov\_dv' from wave 4,5 and 6 to classify individuals as movers and non-movers (Those that move in the coming years after my baseline data is collected). Does what your saying mean should I instead include a longitudinal weight and/or include the weight of wave 6 rather than 3?

At the moment I am attempting my svy commands with c\_indinub\_xw - Another question would be if 'in' would be right thing to put for the derived addrmov\_dv variable as it is not entriely clear to me wether this includes proxys or selfcompletion asnwrs from the variable catalogue, or whether in for interview is the most applicable.

Thank you ever so much again.

Best wishes,  
Caroline

#### #9 - 10/30/2022 01:25 PM - Understanding Society User Support Team

- % Done changed from 50 to 80

So, you need to include all those who were interviewed in Wave 3, but you also need to observe them in Waves 4, 5 & 6 to make sure that they moved or not as the addrmov\_dv variable identifies moves between postcodes across any two consecutive waves. So, if someone missed Wave 5 interview the f\_addrmov\_dv will be missing. So, you will need to use the longitudinal weight from Wave 6.

About whether you should use f\_indinub\_lw or f\_indpxub\_lw, that depends on what other variables you will be including in your models.

If you only include variables that are available for proxy respondents then you can use f\_indpxub\_lw

If you only include variables that are available for non-proxy respondents but no self-completion questions then f\_indinub\_lw

If only include variables that are available for non-proxy respondents but also include self-completion questions then f\_indscub\_lw

#### #10 - 10/31/2022 09:52 AM - Caroline Kienast von Einem

Good morning,

thank you for your response this is very helpful.

When I run the svy:melogit command with the longitudinal weight I get the error: 'survey final weights not allowed with multilevel models; a final weight variable was svyset using the [pw=exp] syntax, but multilevel models require that each stage-level weight variable is svyset using the stage's

*corresponding weight() option'*

I wonder whether the way to resolve this would be to set it up in the svyset syntax as a 2-stage design using both individual and household level weights? In the Weighting FAQs document you talk about analyses at the individual level and analyses at the household level. However, wouldn't any analysis that is done at the individual level ought to allow for clustering by household? I understand the error from stata to pick this up but please let me know if I am reading that wrong or how best to proceed.

A second question is about the addrmov\_dv variable. As you mention, it requires two consecutive waves (compared to 'x\_adcts' which compares back to the last interview. Nevertheless if I look at both variables and consecutive respondents the answers are often very different e.g. adcts shows -8 (inapplicable) but addrmov\_dv shows 'mover' - This made me wonder how addrmov\_dv is derived and why these differences occur. Any input would be really appreciated.

Best wishes,  
Caroline

**#11 - 10/31/2022 10:43 AM - Caroline Kienast von Einem**

- File Capture.PNG added

**#12 - 11/01/2022 08:48 AM - Understanding Society User Support Team**

What are the different levels that you are considering in this multilevel model?

**#13 - 11/01/2022 08:50 AM - Understanding Society User Support Team**

Sorry I sent this too quickly. Are you considering household and individual levels in your multilevel model?

**#14 - 11/01/2022 08:54 AM - Caroline Kienast von Einem**

Good morning!

Households at the top level and individuals nested within the households at the lower level.

As I am not interested in change through time I have not included a lower level of observations nested within individuals - Does this change whether I should be using a longitudinal weight?

Also, reading through other forum entries it has sometimes been advised to include psu as the highest level - Is that correct? How would I do that?

**#15 - 11/01/2022 12:20 PM - Caroline Kienast von Einem**

Hi,

reading in more depth I am now sure that I will need to include weights at each level of my analysis if I want to use 'melogit' / multilevel analysis and that psu should be one of the levels. My questions about observations nested within individuals from above holds but so far I would assume 3 levels:

- Psu
- Household
- Individual

Q1: What weights would be appropriate to use? Could I use nominal weight of 1 for psu and then as discussed above 'f\_indinub\_lw' for the individual level?

Q1B: I know that household weights are not provided for longitudinal analyses - could you help me identify how to best create a weight for this level of my analysis?

Q2: Picking up from above - I am not really interested in change through time BUT I suppose I am following individuals from wave 3 to 6 to identify a move, so far I have not included a lower level of observations nested within individuals but would it make sense to? If I am not including observations within individuals (and therefore a time element) would it still make sense to use a longitudinal weight?

Best wishes,  
Caroline

**#16 - 11/01/2022 01:07 PM - Olena Kaminska**

Caroline,

Firstly, I can confirm that you need a longitudinal weight. This is because those people who moved but dropped out between waves 3 and 6 are nonrespondents, and their correction is present in longitudinal wave 6 weight, but not in cross-sectional weight 3 weight. Any cross-sectional weight will be wrong for this analysis as you take information from multiple waves (3 to 6) to learn about their move.

Your best option would be to use [pweight=...] instead of svyset and svy: commands. Try to use longitudinal weight from wave 6 for a weight - you should not be getting an error message.

Clustering is taken into account as a level in multilevel models. Make use to specify PSU at a higher level. Think very carefully about household level - how do you define if some people in a household move and others stay versus full household moving? Use a household level from one wave (wave 3?), if you decide to keep it. We generally think that households are not a longitudinal concept and are generally avoided in a longitudinal analysis.

Generally you don't need to specify different weights at different levels (unless this is related specifically to your research questions, and from my reading this isn't your case).

Hope this helps,  
Olena

**#17 - 11/02/2022 07:34 AM - Caroline Kienast von Einem**

Hi Olena,

Thank you!

Do you know whether I would need to specify running [pweight=...] instead of svyset and svy commands as:

```
melogit move [pweight=f_indinub_lw] || psu:  
or  
melogit move [pweight=f_indinub_lw] || psu: || hipd:
```

The first runs well, the second as been trying to converge since yesterday ~2pm.

I understand your concerns about households. I would define both, those moving with entire household as well as those splitting from a household, as movers. Because so many moves are occurring within households (icc: 0.999) I think it is important to include this dependency in the model to not inflate the associations with moving. Would a 'normal' multi-level model that includes households as a level allow movers to split from the other members to become a mover?

Oh it supises me that you say it is not necessary to specify different weights at different levels - I thought all multilevel models require specification of all clustering levels.

Always appreciate your help.

Best wishes,  
Caroline

**#18 - 11/02/2022 12:04 PM - Olena Kaminska**

Caroline,

On the specific melogit command it is best to check Stata manual.

Your non-convergence problem is likely due to small size of household clusters (are you using only adults)? Consider studying single households separately from multiple-person households - this may help convergence.

Theoretically to specify change in household belonging you could use a cross-classified multilevel model.

You are right that multilevel models need specification of all levels of clustering. This nevertheless does not mean that you need to specify different weights at different levels to run such models.

Hope this helps,  
Olena

**#19 - 11/04/2022 09:55 AM - Caroline Kienast von Einem**

Hi Olena,

I will check the manual yes thanks!

I am only using adukts yes so some of my household clusters are very small (include only one person). Is there another correction for this other than studying single households separately from multiple-person households?

Another question for specifying the complex study design, do I need to set a finite population correction?

Best wishes,  
Caroline

**#20 - 11/04/2022 12:14 PM - Olena Kaminska**

Caroline,

Yes, there can be other reasons why a model does not converge.  
You don't need to specify finite population correction as UK population is very large.

Thanks,  
Olena

**#21 - 11/04/2022 01:17 PM - Caroline Kienast von Einem**

Hi,

sorry maybe I wasn't clear I was more intersted in ways to *account* for these small households in my model.

Amazing. One last question is then about scaling the weights. In most forum entries I have read it seems to be adviced that this is not necessary if one is only using UKLHS data but then I have also been reading some external papers (like the one below) which really emphasis the need for scaling. What would be your recommendation?

Best wishes,  
Caroline

Carle, A. C. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. BMC medical research methodology, 9(1), 1-13.

**#22 - 11/04/2022 01:23 PM - Olena Kaminska**

Caroline,

Literature on multilevel modelling will help you with your first question.

You can skip scaling if you are using waves 3 to 6 as variation in sample size is not so significant.

Hope this helps,  
Olena

**#23 - 10/24/2023 09:44 AM - Understanding Society User Support Team**

- *Status changed from Feedback to Resolved*
- *% Done changed from 80 to 100*

**Files**

Capture.PNG	18.1 KB	10/31/2022	Caroline Kienast von Einem
-------------	---------	------------	----------------------------