Understanding Society User Support - Support #1421

top coding in income components

10/07/2020 11:05 AM - Luca Giangregorio

Status:	Resolved	Start date:	10/07/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:			

Description

Hi there,
sorry to bother during these unfortunate times.
I'm working with UKHLS 1-3 to build a 3-years income average at the individual level.
However, I have some doubts in the income definitions. Specifically, the monthly net income (fimnnet_dv) is the sum of six components. Looking at the section about top-coding, though, only some of these components are effectively top-coded (w_paynu_dv, w_seearnnet_d and w_j2paynet_dv and the w_fiyrinvinc_dv).
The doubt I have is the following: considering that w_fiyrinvinc_dv/12 - i.e. the monthly income from savings and investments is automatically top-coded, I guess the values above the 8,333£ in the w_fimninvnet_dv are due to the other components, right?
For example, in wave 3 the maximum value of 349999.9£ it should derive from other sources than savings and investments (like private pensions, other property rents, etc.).

In this case, is it harmful to top-code the final total monthly income given that not all components are effectively top-coded and there are some very extreme values?

Thanks for your help and support

History

#1 - 10/08/2020 03:22 PM - Alita Nandi

- Assignee set to Luca Giangregorio

- % Done changed from 0 to 50

- Private changed from Yes to No

Hello,

Thanks for your query. I have passed on your concern to the income team.

About your question, "In this case, is it harmful to top-code the final total monthly income given that not all components are effectively top-coded and there are some very extreme values?" The answer depends on what you are trying to do in your analysis. For example, if you are using income as an explanatory variable then you could use income quantiles, then the very high incomes (including top-coded ones) will be in the same highest income category. Whether only components or the total income is top-coded will not matter.

But note that top-coding is done only to reduce disclosure risks. In many analyses this is not a problem. But if in your analysis precise income values are needed, then you should apply for the Special License version of the data where these variables are not top-coded.
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6931

Best wishes,
Understanding Society User Support

#2 - 10/08/2020 05:21 PM - Luca Giangregorio

Alita Nandi wrote:

> Hello,
>
> Thanks for your query. I have passed on your concern to the income team.
>
> About your question, "In this case, is it harmful to top-code the final total monthly income given that not all components are effectively top-coded and there are some very extreme values?" The answer depends on what you are trying to do in your analysis. For example, if you are using income as an explanatory variable then you could use income quantiles, then the very high incomes (including top-coded ones) will be in the same highest income category. Whether only components or the total income is top-coded will not matter.
>
> But note that top-coding is done only to reduce disclosure risks. In many analyses this is not a problem. But if in your analysis precise income values are needed, then you should apply for the Special License version of the data where these variables are not top-coded.
> https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6931

Best wishes,
Understanding Society User Support

Alita Nandi wrote:

Hello,

Thanks for your query. I have passed on your concern to the income team.

About your question, "In this case, is it harmful to top-code the final total monthly income given that not all components are effectively top-coded and there are some very extreme values?" The answer depends on what you are trying to do in your analysis. For example, if you are using income as an explanatory variable then you could use income quantiles, then the very high incomes (including top-coded ones) will be in the same highest income category. Whether only components or the total income is top-coded will not matter.

But note that top-coding is done only to reduce disclosure risks. In many analyses this is not a problem. But if in your analysis precise income values are needed, then you should apply for the Special License version of the data where these variables are not top-coded.
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6931

Best wishes,
Understanding Society User Support

Thanks for your prompt reply.

I will use income as my dependent variable for conditional and unconditional quintile regressions. I was guessing that if some of these values - like top 0.5% - are already very extremes using the top-coded version, with the full-version the problems of right-skewness can only be exaggerated and using quintiles of income as outcomes, these should still be within the same quintile as you were pointing out.
In this sense I was asking if is it harmful or not.

Thanks again for your support!

#3 - 10/13/2020 04:42 PM - Alita Nandi

- % Done changed from 50 to 80

Hi Luca - Yes, if you use the non-topcoded values, there will be more extreme values. But please note that we do not top-code the data because the distribution is skewed, but rather because the disclosure risks are higher for those with extreme values of income.

#4 - 10/13/2020 04:43 PM - Alita Nandi

- Status changed from New to Feedback

#5 - 10/13/2021 10:56 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- Assignee deleted (Luca Giangregorio)

- % Done changed from 80 to 100