Understanding Society User Support - Support #2251

jbsic07_cc Missing Data Query

05/13/2025 05:10 PM - Michael Francis

Status:	Feedback	Start date:	05/13/2025
Priority:	Urgent	% Done:	50%
Assignee:	Understanding Society User Support Team		
Category:	Data inconsistency		
Description			
Hi there			

I am using a combination of SIC and SOC codes to delineate between various employment and Covid groups. However, I've noticed that the proportion of individuals who have responded to wave n but have not been assigned a SIC code (coded -8) has increased massively. For example, if you check the number of NA (inapplicable,-8) responses (respondents who have an interview date) in wave n it comes to 25,885 compared to only 14,241 in wave m, 14,797 in wave I and 14,693 in wave k?

I am going to use the 'samejob' variable to see if this can help but please could somebody check if this is an actual mistake in the data or processing of the data itself - it also shows on the website variable page as having a lot more missing data in wave n (inapplicable, -8), for both the general jbsic07_cc and the special licence version.

If someone could point to a better solution that would also be super helpful!

Many thanks in advance,

Michael Francis

History

#1 - 05/14/2025 11:37 PM - Understanding Society User Support Team

- Status changed from New to Feedback
- % Done changed from 0 to 50
- Private changed from Yes to No

Hello Michael,

For a similar issue, please refer to Issue #2248, which you can access here: https://iserredex.essex.ac.uk/support/issues/2248.

I hope this information is helpful.

Best wishes, Roberto Cavazos Understanding Society User Support Team

#2 - 05/15/2025 11:37 AM - Michael Francis

- File clipboard-202505151137-58dmi.png added

Hi Roberto,

Thanks for the reference - this clarifies what the issue is. However, this does not really resolve the issue in my case, as the UK Govt's definition of key workers is a cross-reference matrix of SOC2010 and SIC2007. I am also using the HRMC definition of furloughed industries, which only uses SIC 2007. Without an accurate representation of SOC/SIC this could possibly make the group estimates unstable.

In terms of a solution, would you recommend using the samejob flag? This seems to produce quite similar estimates to the previous wave once weighted, see image attached:

clipboard-202505151137-58dmi.png Alternatively, I could use jbsic07_1_cc and just look at the first job for wave n?

Another option would be to consider, as you explained, highlighting the individuals with multiple jobs and only using the first job for their sic classification? It would be good to get some guidance on a specific method as it is not possible to reliably use the latest wave (n) otherwise, and I'm looking at the recovery since Covid, so it's really the crux of my project.

Also, in order to externally validate the study results, e.g., with the LFS, it is necessary for the UKHLS SIC codes to be accurate and based on the same definition as the current job definition in the previous waves a-m.

Many thanks in advance,

Michael Francis

Files

clipboard-202505151137-58dmi.png

296 KB 05/15/2025

Michael Francis