

Understanding Society User Support - Support #1554

hsval - units

06/22/2021 07:49 AM - Edward Pinchbeck

Status:	Resolved	Start date:	06/22/2021																					
Priority:	Normal	% Done:	100%																					
Assignee:	Understanding Society User Support Team																							
Category:	Data inconsistency																							
Description																								
Hello,																								
I would like to use the hsval variable (About how much would you expect to get for your home if you sold it today?). However, the variable is often reported in different units (e.g. £, £/100,000) even within the same individual. See below for an actual example in the data.																								
<table><tr><th>pidp</th><th>wave</th><th>hvalue</th></tr><tr><td>XXXXXXX</td><td>a</td><td>165000</td></tr><tr><td>XXXXXXX</td><td>b</td><td>170000</td></tr><tr><td>XXXXXXX</td><td>c</td><td>150000</td></tr><tr><td>XXXXXXX</td><td>d</td><td>160</td></tr><tr><td>XXXXXXX</td><td>e</td><td>160000</td></tr><tr><td>XXXXXXX</td><td>f</td><td>170000</td></tr></table>				pidp	wave	hvalue	XXXXXXX	a	165000	XXXXXXX	b	170000	XXXXXXX	c	150000	XXXXXXX	d	160	XXXXXXX	e	160000	XXXXXXX	f	170000
pidp	wave	hvalue																						
XXXXXXX	a	165000																						
XXXXXXX	b	170000																						
XXXXXXX	c	150000																						
XXXXXXX	d	160																						
XXXXXXX	e	160000																						
XXXXXXX	f	170000																						
Is there any way to identify the units other than by manual inspection and educated guesswork? As you can imagine, not all cases are as straightforward as the one above.																								
Many thanks																								
Ted																								

History

#1 - 06/22/2021 09:19 AM - Understanding Society User Support Team

- Status changed from New to Feedback
- % Done changed from 0 to 80
- Private changed from Yes to No

We don't have any additional information that will help you to understand why some of the hsval reported are so low. I checked the distributions across waves, and the 1st percentile is around 50,000 for most waves, of which around half report values less than 10000. Number of cases with hsval<10000 ranges from 27-41 between waves 1-7 & 115-181 between waves 8-10. You will need to decide what to do with these cases. As the numbers are very low, you could consider dropping these cases or keeping them but using a 0-1 indicator variable for these cases to see if the coefficients you are interested in are different for these cases. Depending on what you are interested in, you could also categorise hsval into quantiles and then these cases will be included in the lowest quantile.

Best wishes,
Understanding Society User Support Team

#2 - 06/22/2021 10:44 AM - Edward Pinchbeck

OK. Thank you. It is true that there are few cases like my example above. However, there are many other cases where the units are otherwise truncated but which results in extreme implied changes between waves. While some of these may be genuine, I suspect that most are not. For example,

pidp	wave	hvalue
1224299891	19	350000
1224299891	20	350000
1224299891	21	350000
1224299891	22	325000
1224299891	23	35000
1224299891	24	350000
1224299891	25	380000

I can come up with some ways to deal with this, but thought it may be helpful to feed this back.
All best

Ted

#3 - 06/22/2021 11:39 AM - Understanding Society User Support Team

Thanks Ted. I see what you mean. I can see in this example it does look like they misreported hval in Wave 23 (UKHLS W5?). I will discuss this with the data team and see if they have any additional advice.

One way I can think of identifying such large changes that is likely due to reporting error is to check the cases where the person is still living at the same address. In those cases if the hval changes between waves by a % more than the national average for change in house prices, then that would be a red flag.

#4 - 06/22/2021 04:23 PM - Edward Pinchbeck

Fab, thanks. Yes Wave 23 is UKHLS W5.

#5 - 06/30/2021 07:57 AM - Edward Pinchbeck

Hello again. Just following this up: I wonder if the data team had any further advice on this?

Also, could you provide specific advice on how I can best tell whether individuals are still living at the same address. Would this be by using plnew for BHPS and origadd for UKHLS?

Many thanks
Ted

#6 - 06/30/2021 09:16 AM - Understanding Society User Support Team

Hello,

Please take a look at this User Forum post about the variables we provide to identify movers.

<https://iserrredex.essex.ac.uk/support/issues/1539>

Best wishes,
Understanding Society User Support Team

#7 - 07/08/2021 08:15 AM - Edward Pinchbeck

Thanks, that was helpful.

Regarding the hval variable, further inspection of the UKHLS and BHPS data suggests that several hundred (or more) of the entries for non-movers are likely misreported. Is there any process by which I could request that this field is investigated and/or cross checked against the source material?

Many thanks
Ted

#8 - 07/08/2021 12:43 PM - Understanding Society User Support Team

Thanks - I have passed on your request to the data team.

#9 - 08/05/2021 10:16 AM - Edward Pinchbeck

Hello. Thank you. When convenient, please could you let me know if the data team have been able to look at this yet?
All best
Ted

#10 - 08/05/2021 11:27 AM - Understanding Society User Support Team

Sorry for the delay in getting back to you. The data team have looked into this but their investigation revealed that while this soft check happens no variables are created - so a message appears for the respondent to check whether their reported value is correct and if not to go back and correct it.

#11 - 08/06/2021 08:42 AM - Edward Pinchbeck

Good morning. No problem.

So if I have understood correctly this soft check occurs when Hval < 50000 OR > 1000000 and seems to have been in place since wave 1 (2009-11). However, there is no other ex post check on, or validation of, the values reported.

I am sorry for the follow up but I wish to use this field in my research and I have some concerns that a non-trivial amount of the values contain errors despite the soft check. Many of these appear to be just missing (or extra) zeros e.g. my example above: pidp 1224299891 UKHLS W5 hval should be 350000 rather than 35000.

Is there any process by which I could request your data team investigate this further, ultimately with a view to correcting obvious errors or flagging more difficult cases, or is there nothing further you can do here?

Thanks,

Ted

#12 - 08/06/2021 11:18 AM - Understanding Society User Support Team

Thanks Ted. I understand. I have passed on your request to the data team. If you have written any code to automate the process of identifying such cases and would be happy to share it with us, I can pass that on to the data team as well.
Best wishes

#13 - 08/12/2021 03:17 PM - Edward Pinchbeck

Hi, sorry for the delay. I've been away.

After compiling the full BHPS and UKHLS data and retaining the HH head, I created wave and mover variables then used STATA code on the following lines:

```
gen ym = ym(intdaty_dv,intdatm_dv)
replace ym=ym(istrtdaty,istrtdatm) if mi(ym)
format ym tm
replace ym= tm(1991m9) if wave==1
gen hvalue= cond(hsval<. x%x hsval>0, hsval,.)
bys pidp (ym): gen d_hval=(hval-hval[_n-1])/hval[_n-1] if moved==0 & wave==wave[_n-1]+1
sum d_hval, det
bys pidp: egen maxdelta=max(d_hval)
bys pidp (ym): egen mindelta=min(d_hval)
br pidp hvalue d_hval wave moved if (maxdelta>=2|mindelta<-.5) & !mi(maxdelta)
```

#14 - 08/20/2021 01:19 PM - Understanding Society User Support Team

Thanks - I will pass this onto the data team

#15 - 10/12/2021 02:30 PM - Understanding Society User Support Team

- Assignee set to Understanding Society User Support Team

#16 - 10/12/2021 02:33 PM - Understanding Society User Support Team

- Status changed from Feedback to In Progress

#17 - 08/10/2022 04:39 PM - Understanding Society User Support Team

- Category set to Data inconsistency

- Status changed from In Progress to Resolved

- % Done changed from 80 to 100