QA of Main Job Marker (mjob)

One-page summary

Issue/area of concern

ONS generates a dummy variable 'mjob' is supposed to indicate whether this job is the 'main job' if a person holds more than one job. It appears that this definition is based upon the job with the largest number of hours worked (but ONs state that this may be adjusted manually by the ASHE team). However, there are a substantial number of multiple job holders where this is not the case. In addition, the definition of 'mjob' is not helpful (if you have only one job, 'mjob' is set to 'false').

Findings

We have investigated alternative definitions of 'main job' based on different measures of hours and wages, and the correlation of these with the ONS measure. The relationship between these different measures is positive but not as strong would be expected; the relationship with weekly wage seems no better than other possible candidates for 'main job'. The closest relationship is between hours worked (most hours worked => more likely to be main job) but this is not outstanding.

There are also some small inconsistencies in the 'djob' marker which is supposed to flag up whether an employee has more than one job or not. These are small in number.

Changes made to data

We have added new variables to the dataset, with main job definitions based on different hours/earnings measure – see appendix for the list. The code is included in the Stata file.

We have also created a new 'mainjob_ons' which is set to 1 if the only job, or if multiple jobs and the ONS 'mjob' variable is set to 1.

Recommendations to researchers

We recommend using one of the new WED measures. At this stage we have no evidence to support one rather than another. In our own work, we prefer to use the hours measure as it seems more coherent.

Details of QA

Authors: Felix Ritchie, Arusha McKenzie, Van Phan, Damian Whittard, Lucy Stoke, John Forth, Alex Bryson and Carl Singleton

QA Code file: Main job analysis, P:\Working\code\sub_qa_main_job_analysis_v001.do

Variable creation code: can be found in the main ASHE creation code file.

This version: v06a December 2021

Detailed analysis

If a person has more than one job, ONS decides which is the person's 'main job' for the purpose of reporting. In email communication with the WED team, ONS confirmed (email to WED team, 1st December) "it is calculated on hours, but there may be a discrepancy as the ASHE team can change the mjob marker manually during validation/processing". It should be noted that it was thought by the research community until recently that this was based upon the job with the highest weekly wage. However, this appears to be incorrect.

Description of mjob as depicted in ASHE documentation/ONS communication

- Mjob should be set to 0 if the person has only one job
- Mjob should be set to 0 if the person has more than one job in that year but it is not the main job as determined by most hours
- Mjob should be set to 1 if and only if the person has multiple jobs and this job has the most hours
- Djob should be set to 0 if person does not have more than one job listed
- Djob should be set to 1 if person has one or more job listed

Note that the 'main job' criterion is not in the ASHE User Guide. It was widely assumed to be main wage but email communction from ONS (November 2021 to the WED team) stated that is was most hours but subject to manual adjustment by the ASHE team. We have no further information on the latter.

Problem(s):

- The way mjob is defined means a main job is not identified for all persons
- Preliminary analysis suggests
  - some double job jobs without a main job identified
  - multiple jobs as reported by ONS don't appear in the dataset
  - mjob isn't associated with highest hourly or weekly wage

QA Strategy:

- Test that djob works
  - Count the number of jobs an individual has in a year
  - Compare with djob
  - If necessary, construct a djob_ons variable that reflects ASHE data
- Construct an mainjob_ons variable that has a useful definition reflecting djob_ons and the fact that mjob_WED should be 1 if the individual only has one job in that year
  - mainjob_ons = 1 if individual has 1 job
  - mainjob_ons = 1 if individual has more than 1 job and ONS thinks this is the main job
  - mainjob_ons = 0 otherwise
- Check that mainjob_ons is
  - 0 for individuals with no multiple jobs
  - 1 for one and only one of the multiple jobs
- Within a year, for every individual with multiple jobs in the year but only those:
  - Construct alternative measures of 'main job' - different
    - Most hourly, weekly, total wage or some other wage
    - Hours spent on the job
    - If on temporary or permanent contract [tbd]
    - consider combinations eg permanent + biggest hourly wage [tbd]

- o   Tabulate these and compare with mainjob_ons
- Tabulate over years to see if there have been shifts over time
- Discuss results with ASHE team
- Make recommendations

Findings

Results here are for one year. Eyeballing suggests results similar for all the ASHE years

Are the duplicate/main job variables consistent on their own terms?

First test: is the djob variable (djob=1 if more than one job in a year) accurate? Test by counting the number of jobs observed in ASHE.

	Double job (as defined by counting no. of jobs)	
Double job (ONS- given)	N	Y
N	171,777	<10
Y	<10	7,626

Table 1 Double job variable, 2018: comparing ONS-defined vs WED counts; example year 2018, SRS dataset

There are very small variations, potentially to do with inclusion in the final ASHE sample in the SRS.

> **Proposal**: leave the ONS variable unchanged.

Second test: if we create a new variable where mainjob_ons=1 where this is the only job or the (ONS-defined) main job for multiple job-holders, how well does it compare to ONS' variable?

	Main job (WED definition – main if only job or mjob)	
Main job (ONS – defined only if djob=1)	N	Y
N	4,673	171,780
Y	32	3,700

Table 2 Main job variable: comparing ONS-defined vs WED counts

Finding: there are inconsistencies but not many, once you allow for the definition of 'mjob'. The inconsistencies come in because of the inconsistencies in djob. We assume these inconsistencies arise because of the aforementioned manual adjustment.

> **Proposal**: recommend researcher use the WED variable mainjob_ons

How well does the main job definition match to its own supposed definition and to alternatives?

Consider tabulating the mainjob_ons variable against a dummy constructed from looking at actual weekly wages (gpox, but others produce similar results; taken from 3 years to show similar results):

ONS definition	weekly pay definition		
	not main job	main job	Total
2010			
not main job	1,833	457	2,290
main job	291	1,903	2,194
2014			
not main job	2,365	484	2,849
main job	349	2,317	2,666
2018			
not main job	1,992	379	2,371
main job	293	1,921	2,214

Table 3 Main job variable: comparing ONS-defined vs hexo; double jobs only; 2018, SRS dataset

Clearly there is a substantial mismatch, and it persists.

We can postulate several ways the main job might be defined:

- On hourly pay (most valuable job per hour)
- On weekly basic pay (highest expected contribution to income)
- On weekly total pay (highest contribution to income in reference week)
- On basic hours (largest expected time commitment)
- On total hours (largest time commitment in survey week)
- On whether permanent or temporary (assuming permanence is desirable)
- On whether the job is part-time or full-time

We construct variables for most of these except the last two. It is possible to have multiple permanent jobs so this is an insufficient criterion on its own. Similarly, the person may have multiple part-time or full-time jobs (as this is defined by hours).

It is possible that multiple criteria work best eg highest weekly wage and permanent. We have not explored this, as it is not clear how to deal with eg a job with the highest weekly wage but temporary compared to a lower wage but permanent.

A correlation matrix shows that no measure is outstanding, in terms of its correlation with other measures or the ONS definition (there are high correlations between variants on the same measure eg hours or weekly wage, but these are to be expected). P-values in italics; correlations significant at 5% in bold.

	Main job defined by highest…							
	Hourly wage	Hourly wage (LPC def)	Weekly pay (gpay)	Weekly pay (gpox)	Hours, then hourly wage	Total hours	Basic hours	Tenure
Hourly wage (LPC defn)	**0.895** *0.000*							
Weekly pay (gpay)	**0.180** *0.000*	**0.198** *0.000*						
Weekly pay (gpox)	**0.153** *0.000*	**0.174** *0.000*	**0.838** *0.000*					
Hours then hourly wage	0.025 *0.098*	**0.058** *0.000*	**0.833** *0.000*	**0.688** *0.000*				
Total hours	**0.029** *0.047*	**0.062** *0.000*	**0.840** *0.000*	**0.696** *0.000*	**0.967** *0.000*			
Basic hours	0.001 *0.951*	**0.033** *0.027*	**0.795** *0.000*	**0.687** *0.000*	**0.931** *0.000*	**0.942** *0.000*		
Tenure	**0.030** *0.044*	**0.034** *0.020*	**0.138** *0.000*	**0.127** *0.000*	**0.137** *0.000*	**0.141** *0.000*	**0.143** *0.000*	
ONS defn	**0.108** *0.000*	**0.132** *0.000*	**0.707** *0.000*	**0.597** *0.000*	**0.672** *0.000*	**0.668** *0.000*	**0.671** *0.000*	**0.212** *0.000*

Table 4 Correlation between stated main job and derived measures; multiple job holders only, 2018, SRS dataset

There are no compelling statistical reasons to compare one measure over another. Theoretical arguments could be made eg "hours is a better measure as total time is limited; more hours worked in one job always reduces the hours available for a second job; we would expect the same marginal utility for all jobs, and so the jobs where MU is reached at more hours implies a higher initial utility (assuming decling MU)". But there is no hard evidence here to support such an argument

Proposal: include all the potential main_job_xxx wages in the core dataset and let users choose

New variables added to the dataset

These variables are created in the core ASHE data creation code. Note that all of these are defined such that main_job_x = 1 if this is the *only* job *or* this is the main job on the specified criterion, and main job_x = 0 if the person has multiple jobs and this is not the main one on the criterion.

Variable name	Definition: 'variable =1 (main job) iff (only job) or (…)'
main_job_hourly	Most hourly paid (hexo)
main_job_hourly_LP	Most hourly paid, wage used by LP
main_job_weekly_gpay	Max weekly paid (gpay)
main_job_weekly_gpox	Max weekly paid (gpox)
main_job_LP	Most hours and (if tied) most hourly pay on LP calculation
main_job_hours	Most time-consuming in total hours (thrs)
main_job_basic	Most basic hours (bhr)
main_job_tenure	Longest tenure observed (start_date)
main_job_ons	Main job defined by ONS if multiple jobs