

Wage & Employment Dynamics

THE WED PROJECT

Quality assurance stream

Arusha McKenzie, Felix Ritchie and Damian Whittard



Why a QA workstream?

ASHE: well-known but still many issues (and many surprises)

Linked HMRC & Census: unknown quantities

For all datasets: need for QA integrated into documentation

Priority	Topic	Variable	Problem	Source	Checked by us?
1	Discrepancy between stated and actual	sjob ('same job')	data doesn't reflect this - diff jobs, diff company, diff location about 1/3 of time?	Ritchie (2018). Longitudinal analysis of low pay and wages: Is it feasible and useful? Earnings	partially
1	Discrepancy between stated and actual	Age	How consistent is it?	Ritchie NES QA files (released from SRS)	partially
1	Discrepancy between stated and actual	mjob	20% of the time biggest wage is not mjob or vice-versa	Ritchie (2018). Longitudinal analysis of low pay and wages: Is it feasible and useful? Earnings	yes- see mjob review paper
1	Discrepancy between stated and actual	BHRS	Full-time is defined as working over thirty basic hours in a week. But there are a tiny number of discrepancies in	Schaefer, D. and Singleton, C. (2020) Recent changes in British wage inequality: evidence	
1	Discrepancy between stated and actual	WGOR (1996-2002)	For 1996-2002, the work region of the employee is missing. We derive this ourselves consistent with the	Schaefer, D. and Singleton, C. (2020) Recent changes in British wage inequality: evidence	
2	Sampling and Non-response	employee; overtime hours, overtime pay,	high degree of non-responses	Bird, D. (2004). Methodology for the 2004 Annual Survey of Hours and Earnings. Office for National	
			Weights: do they vary and why? 2004 and 2006 where we observe exactly double the number of unique values	Email correspondence (<i>Questions for Louisa</i>) with FR, JT, LD (Jan 2020)	
2	Sampling and Non-response		They use data for 2002-2016; Appx p.1. States an ASHE response rate of 55%.	Sarah Jewell, Giovanni Razzu, Carl Singleton, Who Works for Whom and the UK Gender Pay Gap?,	
			How was response rate calculated? Is this what was expected?		
2	Sampling and Non-response		There is a shift in the size distribution of firms in the sample (towards large firms with 2000+ employees)	Schaefer, D. and Singleton, C. (2020) Recent changes in British wage inequality: evidence	
2	Sampling and Non-response		From 2005 and 2007, the sample size of the ASHE was reduced by twenty percent, with reductions targeted on	Schaefer, D. and Singleton, C. (2020) Recent changes in British wage inequality: evidence	
3	Sampling and Non-response		survey reference date (Apr) varies annually to avoid Easter	Milton, J. (200) New methodology for low pay estimates. Office for National Statistics.	
3	Sampling and Non-response		Appx p.1: The ASHE introduced some imputations, using similar matched 'donor' observations where responses	Schaefer, D. and Singleton, C. (2020) Recent changes in British wage inequality: evidence	

QA Aims and Objectives

Identify issues

Quantify problems

Create solutions/alternatives for users, if possible

Document everything

Examples of QA work (ASHE)

- Main job
- Same job
- Work location



Example 1: main job

Main job 'mjob' - ONS Definition

- $mjob=0$ if person has only 1 job.
- $mjob=0$ if person has more than 1 job, and this is not the main one
- $mjob=1$ if person has more than 1 job, and this is the main one

- The main job is one with "largest wage".

Example 1: why a known variable is needed

Correlations between measures of 'main job'

	Hourly Wage	Total Weekly Wage	Basic Weekly Wage	Hours Worked
Highest total weekly wage	0.0411			
	0.000			
Highest basic weekly wage	0.0441	0.9486		
	0.000	0.000		
Most hours worked	0.0094	0.1413	0.1288	
	0.410	0.000	0.000	
ONS-defined	0.0939	0.0944	0.0869	0.6817
	0.000	0.000	0.000	0.000

Example 1: main job

User outcome:

- created alternative variables
- stated our 'preferred' measure (hours)

Example 2: same job

- Dummy variable ('sjd') set to 1 if person has worked in the same **job** for more than 12 months.
 - Survey question
- ONS has also supplied data with a link to job in the previous year.
 - Used to create a persistent job number ('job_number').
- Is this reflected in the data?

Example 2: same job

	'sjd' Single jobs only	'job number'
same employer	97%	99%
same home COA	90%	90%
same work COA	90%	92%
same industry	94%	96%
same occupation	93%	95%

Example 2: same job

User outcome: have confidence in these

(*BUT* uncertainties of how these are created and used by ONS)



Example 3: work and work addresses

ASHE has home and work address – fantastic

But work address is pre-filled

Employers should change it if wrong - what if they don't?

Three testable hypotheses:

- ASHE vs IDBR: more people working at head office in ASHE
- ASHE ttw distances - should be further for multi-site companies
- Firms electronically submitting don't see these issues

Example 3: work and work addresses

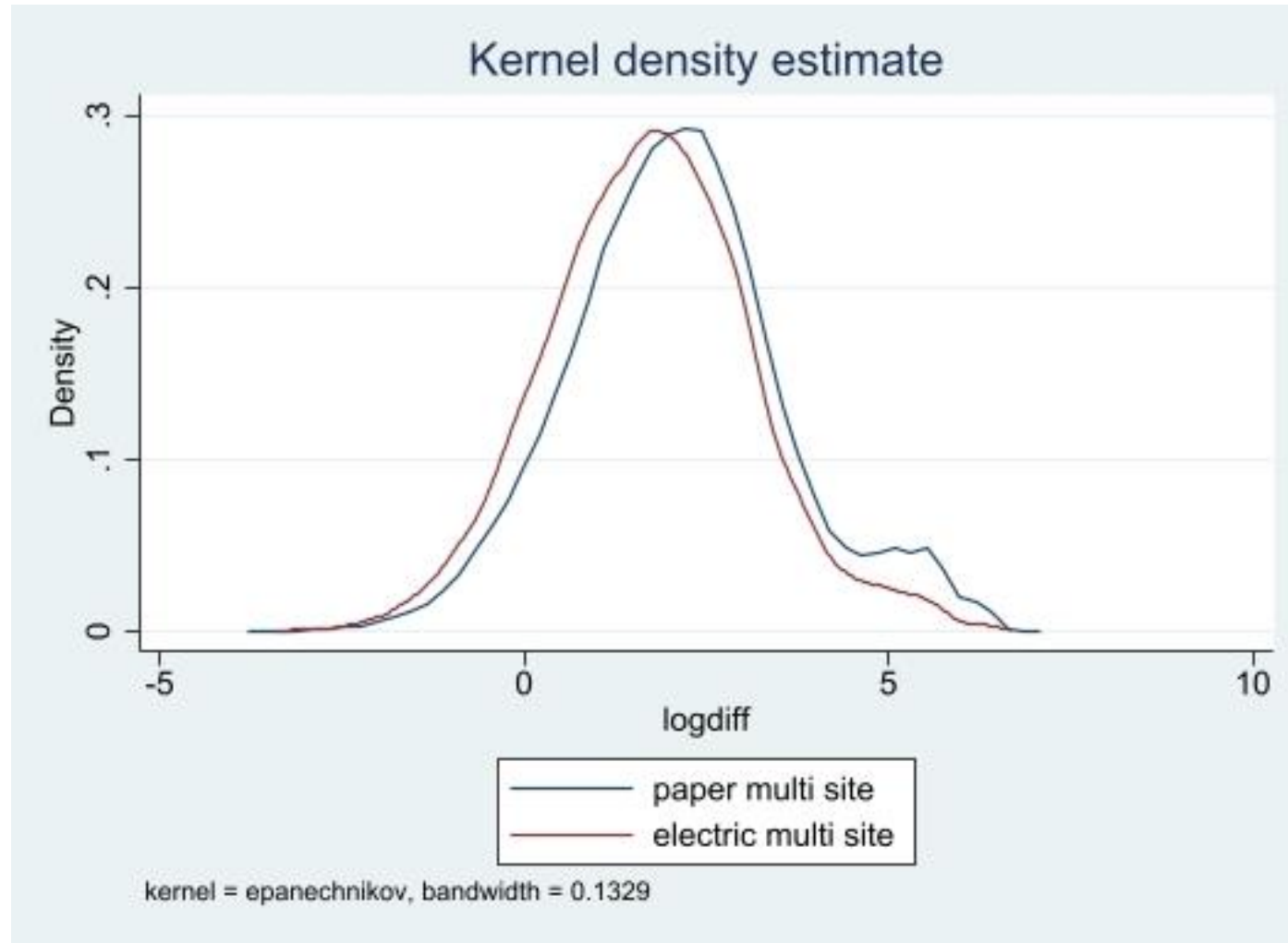
Table 4: Distance between recorded employee work and home addresses in the ASHE (in kilometres), 2018

Organisation type	Observations	Average	10th percentile	50th percentile	90th percentile
Single-site paper	58,113	16.6	0.5	5.7	29.6
Multi-site, paper	99,916	28.0	1.2	7.8	54.2
Single-site, electronic	233	8.5	0.6	5.3	19.3
Multi-site, electronic	25,200	16.6	0.9	5.6	31.1

Example 3: work and work addresses

KDE of the log travel-to-work distance (straight line)

Employees at multi-site organisations only



Example 3: work and work addresses

All 3 hypotheses supported (tabular and regression analysis)

Effect size varies by sector, number of LUs, etc but often significant

No apparent way to check

Example 3: work and work addresses

User outcome:

- No analytical solution - yet
- Documentation: warn (detailed analysis by sector, size etc)

All of this is going into our revised documentation

- ASHE first version due this coming Sept

Similar process for Census/ASHE and HMRC/ASHE, but

- many more unknowns
- starting from scratch

hence goals are ‘provisional’ documentation

- Census: provisional for Jan 2022
- HMRC: provisional May 2022

Thank you for listening

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates."

