Wage & Employment Dynamics
THE WED PROJECT

Why a QA workstream?

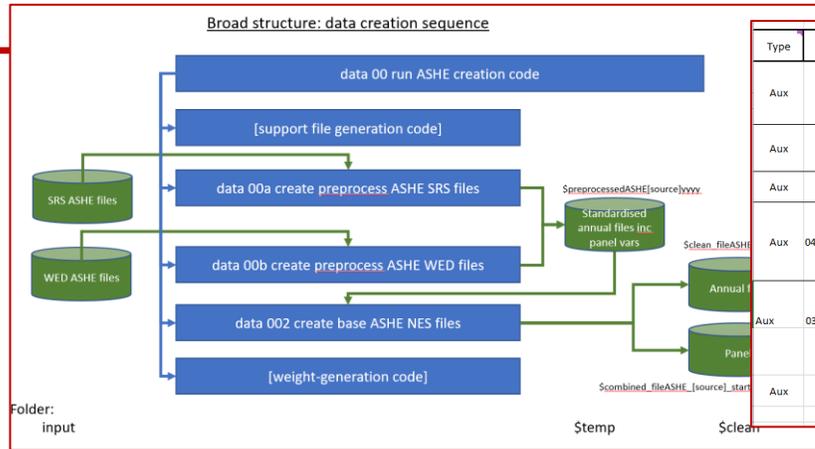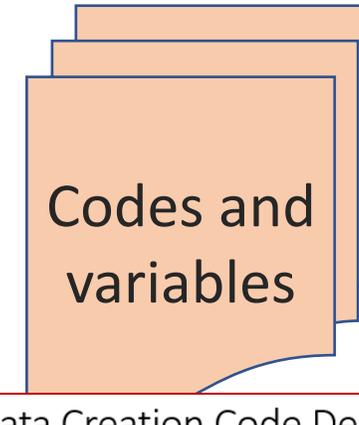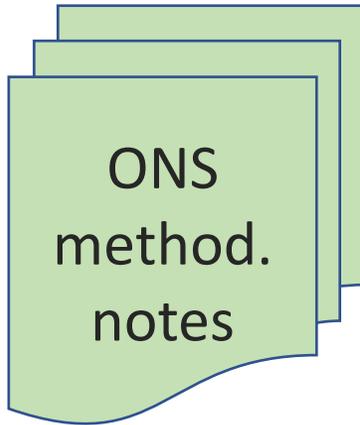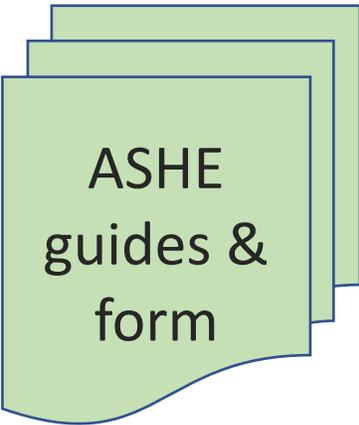ASHE: well-known but still many issues (and many surprises)

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 (*Questions for Louisa*) 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 fot 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	

Documentation - aim

- Collate what there is

- Create our own
  - Variables we've added
  - Data processing
  - QA and methodological notes

- Present in multiple ways

- Ideally, answer every question…

Documentation - ASHE

ASHE guides & form

ONS method. notes

Research papers

QA reports

Codes and variables

Quick start
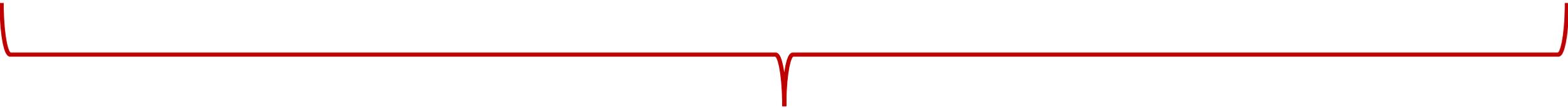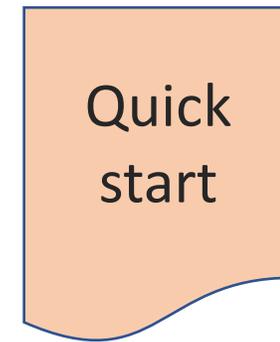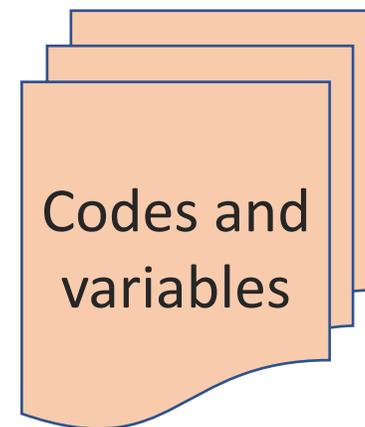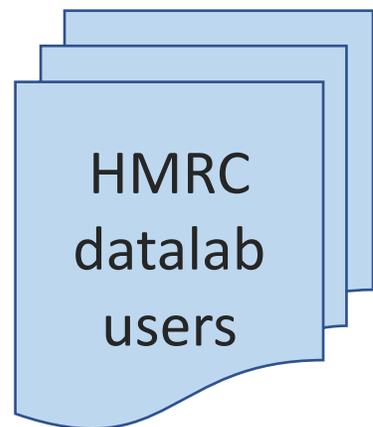
Timetable

	Aug 2021	Sep	Oct	Nov	Dec	Jan 2022	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
ASHE-IDBR			Initial clean, documented version Training/info sessions						Final version Coding delivered to ONS					
ASHE-Census			Delivery of HH files			Provisional docs/QA Training/info sessions Exploratory analysis?		Final documentation Available for secondary use						
HMRC			Delivery							Provisional docs/QA				Final docs Info sessions

Wage & Employment Dynamics

T H E   W E D   P R O J E C T

ASHE Cross-Sectional Weights

Contact: L.Stokes@niesr.ac.uk

ASHE cross-sectional weights

- Existing ASHE weights:
    - data are weighted to UK population totals from the UK Labour Force Survey, based on classes defined by occupation, region, age and sex
    - "standard" and "low pay" weights available
- ASHE is completed by employers – there may be differences in which types of employers respond, and which respond quickly...
- We are investigating employer response by linking ASHE to BSD:
    - explore characteristics of firms (in addition to size) that make them more/less likely to respond to ASHE (e.g. sector, industry)
    - aim to create a firm-level adjustment factor that we can then apply to the standard ASHE weight

Planned outputs:

- These alternative weights will ultimately be made available to users as part of WED versions of dataset
- With documentation/paper for users to explain motivation, derivation and implications

Disclaimer

"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."

Motivation

- ASHE sample contains a *de facto* panel element (sample selection each year based on same final digits of the NI number)

- Many have used the panel dimension of ASHE to make inferences about:
  - Wage progression within a job and its determinants (e.g. Elsby et al, 2016; Schaeffer and Singleton, 2019)
  - Job displacement or employment exit and its determinants (e.g. Dickens et al, 2015; Stokes et al, 2017)

- But ONS not focused on panel retention: their focus is cross-sectional representativeness

- Around 25% of those observed in ASHE at *t1* are not observed at *t+1*   Table

- Assumption to date: those who exit from ASHE between t1 and t+1 are representative of those who exit PAYE employment (i.e. longitudinal sample attrition is either random or ignorable).

- We investigate whether this assumption is reasonable by comparing rates of sample exit in ASHE with estimates of p(exit employment) obtained from the Two-Year Annual Population Survey (APS)

Sample retention rate across year-pairs

Table 1: Sample Retention rate in ASHE from year t to year t+i

Year (T)	T+1	T+2	T+3	T+4	T+5	T+6	T+7	T+8	T+9	T+10	T+11	T+12	T+13	T+14	T+15
2004	77.3	71.6	55.3	52.3	61.4	58.8	58.4	54.5	53.2	51.5	48.4	45.1	43.2	41.2	38.8
2005	78.1	59.0	55.1	64.4	61.2	60.8	56.6	55.3	53.7	50.3	46.9	44.9	42.8	40.2	
2006	64.1	58.6	67.8	64.0	63.3	58.9	57.4	55.7	52.2	48.6	46.5	44.3	41.8		
2007	75.7	71.2	66.3	65.4	60.8	59.1	57.4	53.7	50.0	47.9	45.4	42.8			
2008	77.2	70.2	68.5	63.3	61.6	59.6	55.7	52.0	49.7	47.1	44.5				
2009	77.4	73.1	67.0	64.8	62.5	58.2	54.3	51.9	49.3	46.5					
2010	78.6	70.7	67.9	65.3	60.7	56.5	54.1	51.9	48.9						
2011	76.0	71.5	68.1	63.1	58.5	55.9	53.2	50.2							
2012	77.8	72.4	66.6	61.4	58.6	55.5	52.4								
2013	78.1	70.3	64.4	61.1	57.9	54.5									
2014	75.0	67.5	63.3	59.7	56.2										
2015	73.7	67.1	62.4	58.5											
2016	73.4	66.3	61.6												
2017	72.7	65.6													
2018	71.3														
Average	75.1	68.2	64.2	61.9	60.2	57.8	55.5	53.1	51.0	48.7	46.3	44.3	42.6	40.7	38.8

Note: Shading is used to indicate the scale of sample retention: darker green for higher rates, amber for middling rates, then darker shades of red for lower rates.

Source: ASHE

Headlines

- Employees are three times more likely to exit ASHE yr-on-yr than they are to exit employment
  - ASHE sample exit rate = 0.25 p.a.
  - APS employment exit rate = 0.08 p.a.

  Sample attrition rate = 0.17 p.a.

- Employer non-response is a large part of this:
  - Sporadic non-response within a continuing employment spell: One-fifth (20%) of all sample exits -> sample member reappears in a later year, with the same employer and same empt. start date (often after a gap of one year).
  - Switch jobs from responding employer to non-responding employer: For those who appear in the sample in two consecutive years, share of employees who stay with the same employer is slightly higher than in APS (92% vs 89%)

- Panel attrition is non-random (and so may not be ignorable in longitudinal analysis of ASHE):
  - Relatively high among those: with low wages/hours; aged 20-44; in private-sector services; in Lon/SE
  - Higher in later years of ASHE

- We have used the ASHE/APS comparison to derive yr-to-yr weights to correct for observable attrition biases -> small widening of distribution of individuals' annual wage changes

- Our paper proposes methods to adjust for attrition bias over periods of 2+ years, and when employment exit is the dep. var. This is pending whilst we await news on HMRC PAYE data.

For further information in relation to the content of this presentation, please contact damian2.Whittard@uwe.ac.uk

Workplace Location Variable (preliminary findings)

- What's the problem?
  - ASHE workplace location variable may be systematically mismeasured
    - Workplace variable in ASHE form comes pre-filled (paper submissions)
  - If so, biased estimates potentially informing government policy – e.g. regional pay gaps
- How did we (indirectly) investigate the problem?
  - "Special arrangements" (2016-2018)
  - Link BSD and ASHE (technical challenges)
    - single/multisite and paper/electronic submissions
  - Compare proportion of head office employment (ASHE/BSD)
  - Compare distance travelled to work (ASHE – four groups)
- What's did the evidence show?
  - Greater proportion of employees are reported as working in the head office in the ASHE compared to BSD (multisite 80% compared to 61%)
  - Employees who work for multi-site organisations who provide a paper ASHE submission commute on average an additional 20km daily

- What did the evidence show (continued)?
  - For multisite companies making paper submissions, the distance travelled to work increases as:
    - Number of employees increased/ number of local units increased/ and employed in private sector (e.g. finance and law)
  - Regression analysis
    - Proportions at head office
      - Lower: **special arrangements**; size of company, London
      - Higher: **private sectors**; **number of local units**
      - Company structure and urban/rural classification less clear
    - Distance travelled to work (employees working for enterprises)
      - **Special arrangements** - live 10km/16% closer than those making paper submissions
      - Singlesites - live 6km/26% closer than multi-sites
      - Construction; finance/law; utilities **sectors** – live between 50% and 100% further away than public sector employees
      - For every additional 50 **local units**, employees lived approximately 5km further away from work
      - Those working in London lived about 5-10 km further away from work
  - Conclusion: strong circumstantial evidence of systematic measurement bias

- If conclusion accepted, what are the implications in the short-term?
  - Researchers need to take this in to account
    - Create sub-samples
      - Exclude multisite companies making paper submissions (97,000 observations/ 18,000 enterprises)
    - More nuanced approach – multisite companies making paper submissions conditioned on:
      - Number of local units
      - Sectors
      - Region
  - Include paper questionnaire dummy in regressions
  - Include number of local units as a new control variable
    - sector and region already included in many of the models
- What are the implications in the medium term?
  - Going forward there is potential to improve the quality of data
    - ONS to expand data collected using "special arrangements"
      - Number of local units
      - Sectors

"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."

Motivation

ASHE 2011
13/04/2011

+

Census 2011
27/03/2011

=

ASHE-Census linkage
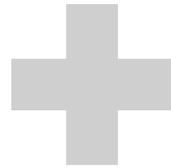
- ➢ Personal information
  - ▪ Age
  - ▪ Gender
  - ▪ Home postcode
- ➢ Employment information
  - ▪ Wage
  - ▪ Working hours
  - ▪ etc
- ➢ Employer information
  - ▪ Work postcode
  - ▪ etc

- ➢ Personal information
  - ▪ Age
  - ▪ Gender
  - ▪ Education Qualification
  - ▪ Country of Birth
  - ▪ Religion
  - ▪ Language ability
  - ▪ Ethnicity
- ➢ Family information

- ➢ Value of ASHE-Census linkage
  - ▪ Returns to human capital
  - ▪ Wage gaps by gender, ethnicity, disability
  - ▪ Wage progression among migrants
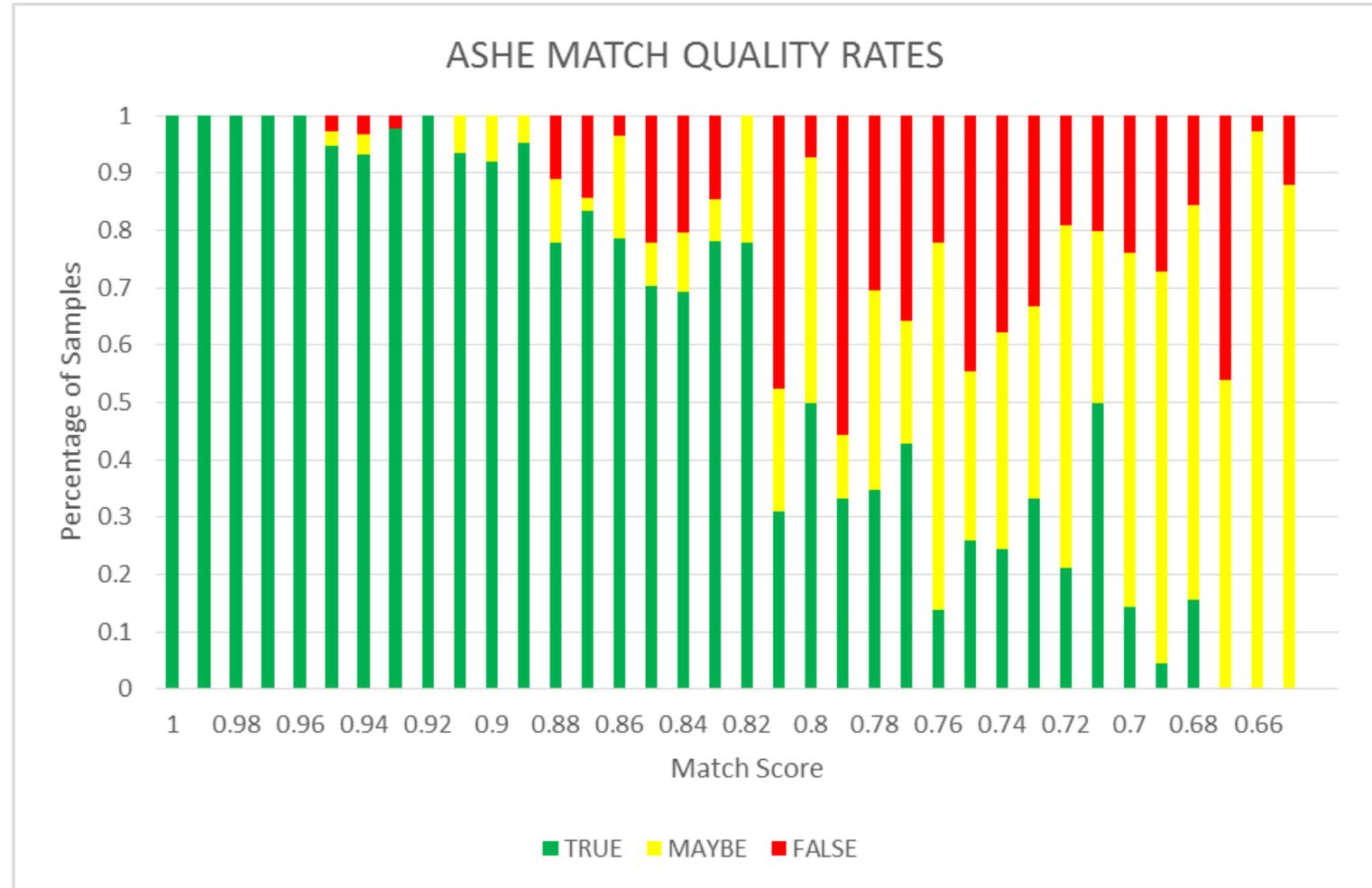 - ▪ Relevance of partner status

Linkage process

- ONS seek matches for all individuals with E&W resident addresses in ASHE 2010, 2011 or 2012

- Stage 1: Deterministic match on 46 combinations of name, gender, DOB and home postcode, for example:
  - Name + gender + DOB + postcode / Name + DOB + postcode
  - Full name / surname only
  - Literal name / soundex
  - Full postcode / postcode district only

- If unmatched -> Stage 2: Probabilistic match, including SIC and SOC

- If unmatched -> Stage 3: Determ. match on: gender, DOB, home pcode, work pcode and SIC/SOC (awaiting results)

- Each matched record has a 'score' (0-1) based on the completeness of the matchkeys

First-run: clerical checks
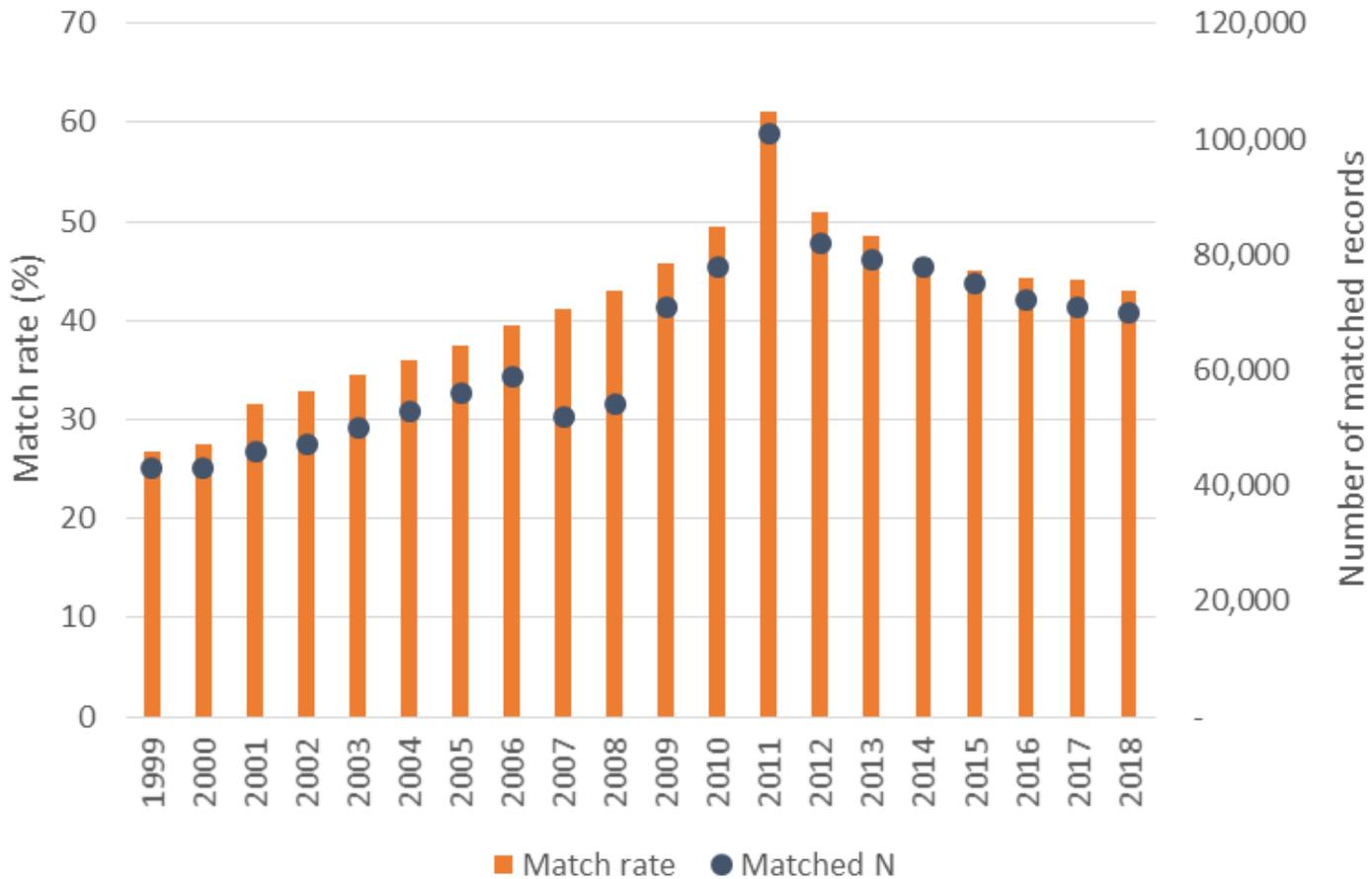
- Checks on 895 matched records with match score >=0.65

In this sample of 895:

- 90% of matches with a score >=0.82 are found to be 'true'

- 82% of all 'true' matches have a score >=0.82

ASHE MATCH QUALITY RATES

Percentage of Samples

Match Score: 1 0.98 0.96 0.94 0.92 0.9 0.88 0.86 0.84 0.82 0.8 0.78 0.76 0.74 0.72 0.7 0.68 0.66

■ TRUE ■ MAYBE ■ FALSE

First-run: linkage outcomes

- Around 60% of records in ASHE 2011 can be matched to Census 2011

- Matched records = those with match score >=0.82

- Data includes match score

- 79% of all matched records have a match score of 1.0

- Match less likely for women than men (odds ratio 0.76)

- Less likely for those in London (0.66) or SE (0.89)

Wage & Employment Dynamics

THE WED PROJECT

Distribution of Ethnic Gender Pay Gaps

Contact: c.a.singleton@reading.ac.uk, van4.phan@uwe.ac.uk

"The Distribution of Ethnic Pay Gaps in England and Wales: The Role of the Firm"

Motivation

- Large ethnic wage gap literature for Britain based exclusively on household survey data
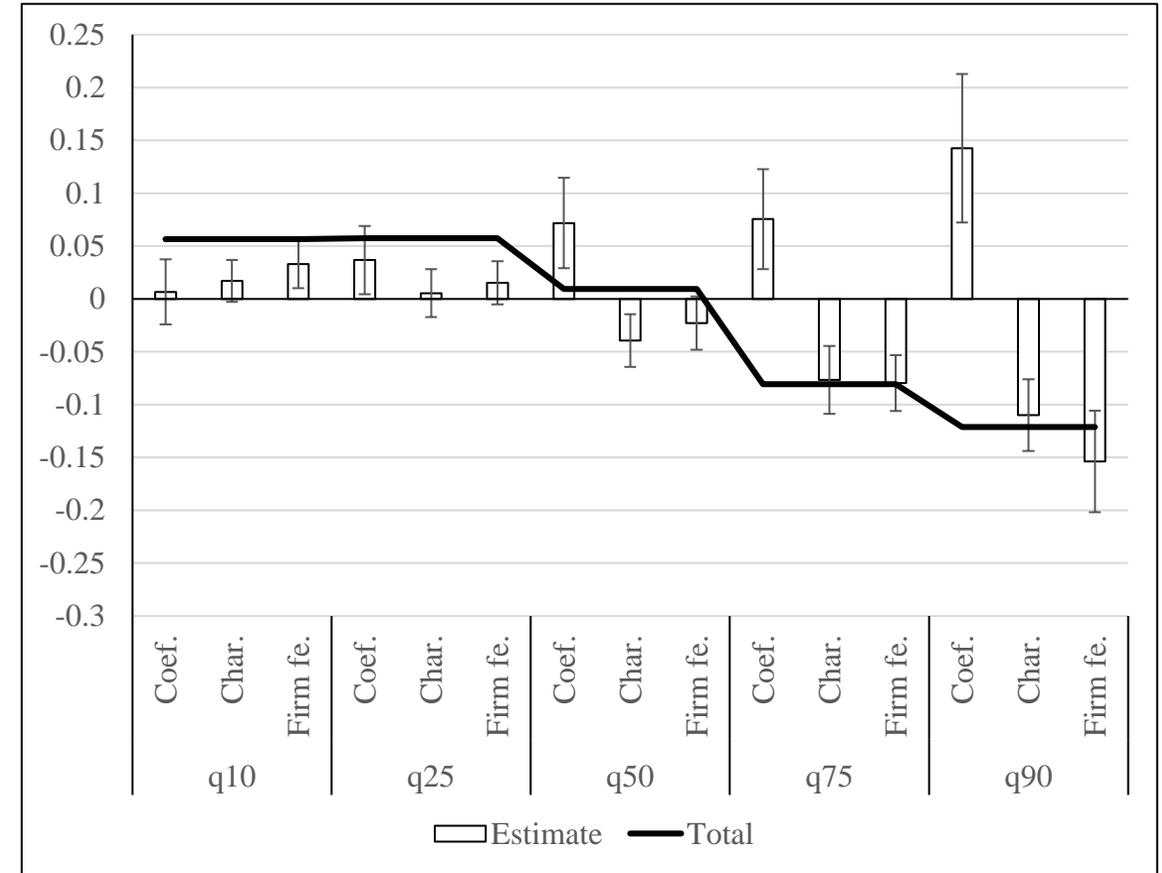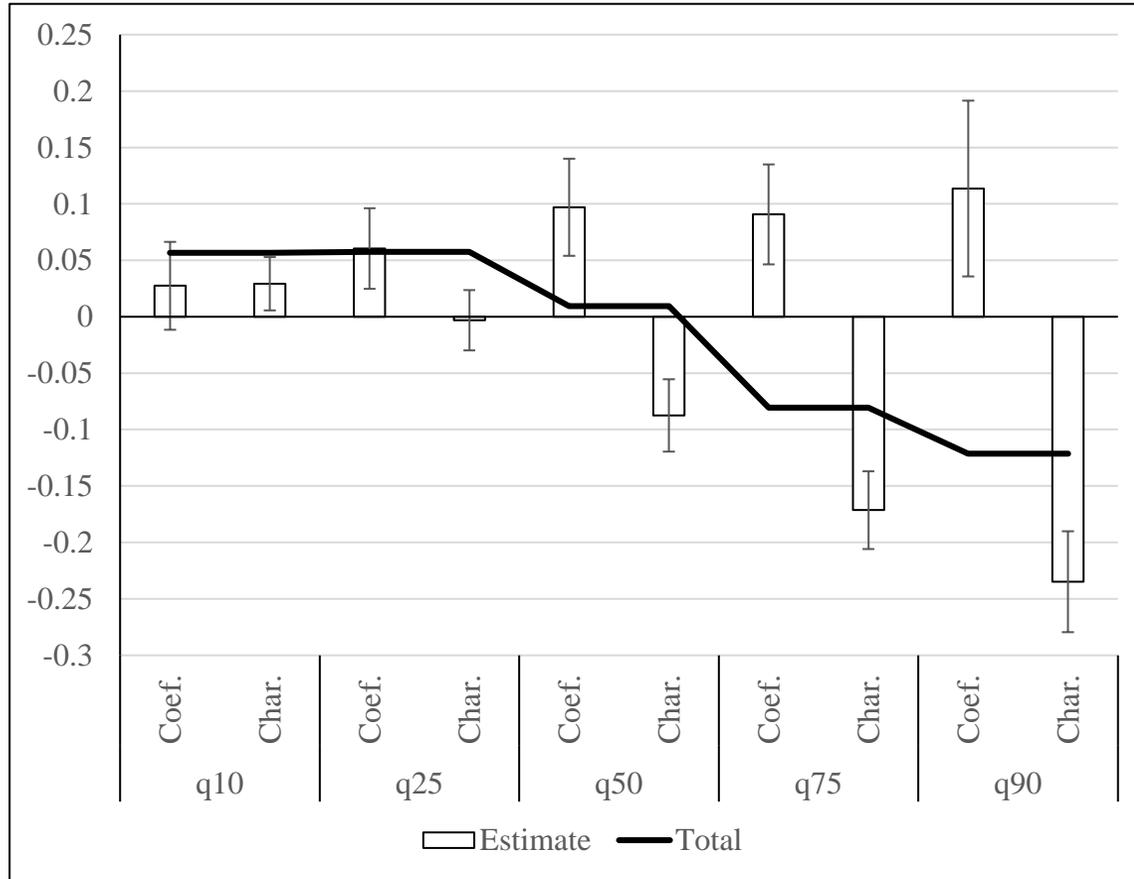- The firm is an omitted variable in these studies

Contribution

- Examine ethnic wage gaps across the earnings distribution using payroll data linked to 2011 Census
- First use of linked ASHE-Census data
  - Census provides ethnicity previously lacking from ASHE
  - With other co-variates important in estimating wages, e.g., education, family
- Illustrative and experimental
  - Some work still to be done
  - But provides important insights into
    - The firm's role in the ethnic wage gap in England and Wales across the wage distribution
    - What can be done with the linked ASHE-Census data (not only Ethnicity)
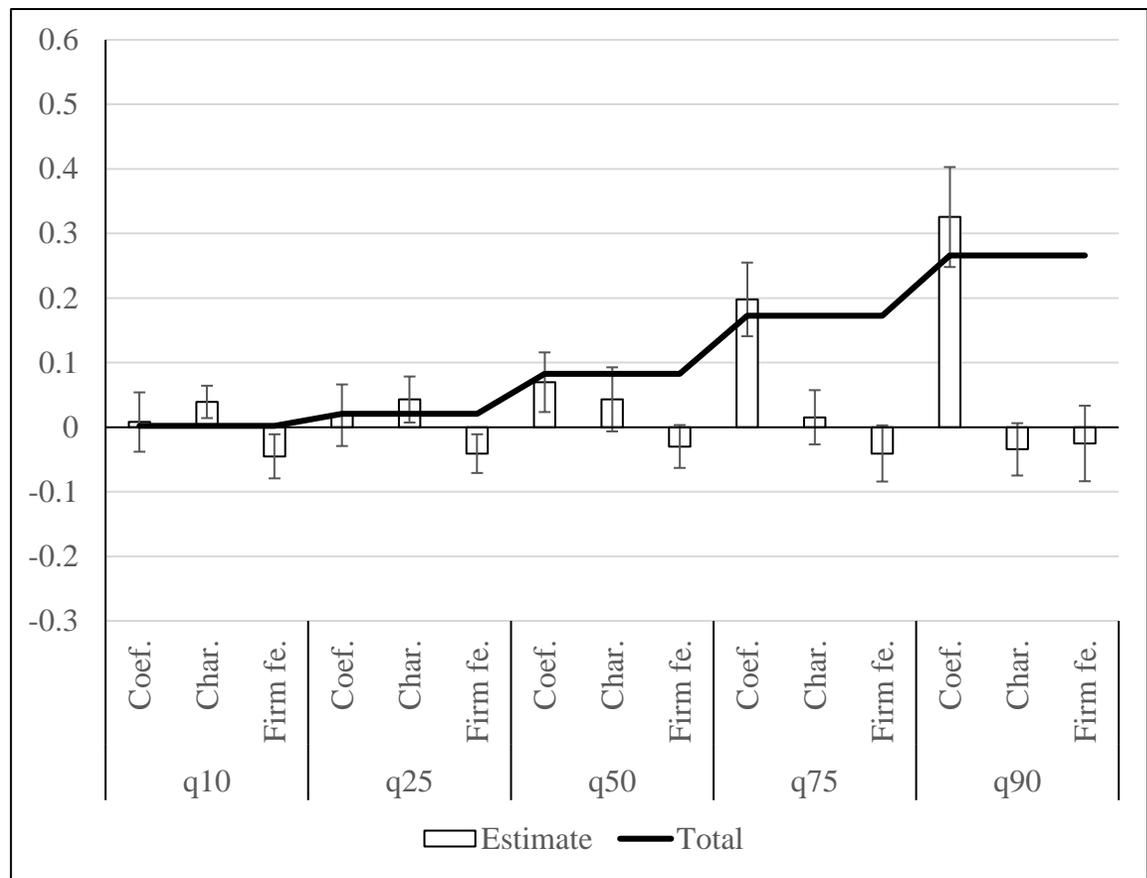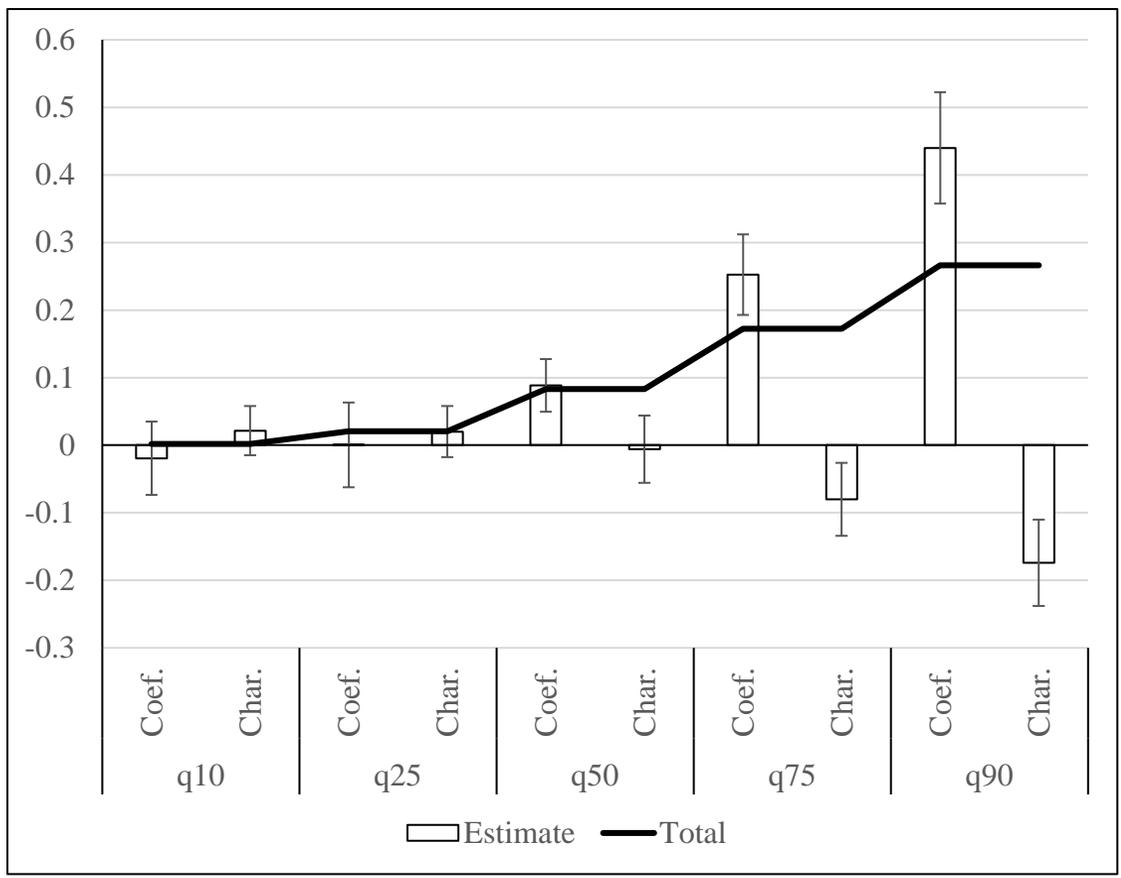
DECOMPOSING THE DISTRIBUTION OF ETHNIC WAGE GAPS
Without firm-specific wage effects vs. ***With*** firm-specific wage effects

A couple of examples:
White men vs. Indian men
White men vs Black Caribbean men

White men vs. Indian men

Reflections so far

- Ethnic wage gaps vary substantially across the wage distribution and across ethnic groups and by gender
  - Perhaps no surprise there
  - In some cases lower gaps at the bottom end of the distribution are consistent with minimum wages limiting discriminatory behaviour
- Firm-specific wage effects …
  - … explain substantial parts of the distribution of ethnic wage gaps
  - … generally lead to smaller estimated contributions from other traits (because those are linked to firm entry)

WED 2: HMRC

- Research Ready Data
  - PAYE data: transitions between surveys
  - Self-assessment data: broaden the coverage
- Linked through encrypted National Insurance Number
  - Project signed off by HMRC/DWP NINO Policy Board
  - All variables signed off by HMRC
  - NINO Encryption complete
  - ONS currently linking data (expected shortly)
  - Following QA data should be available between June and September 2022
- Strategic Research
  - Quality assurance
  - Calibration of earnings survey data
  - *Longitudinal analysis of labour market transitions (*potentially delayed till WED3)

Communications

- Stakeholder Updates
  - Academic Stakeholder Group
    - Carl Singleton joined the WED team (Summer 2021)

  - Government Stakeholder Group
    - HMRC representation(September 2021)

  - WED Public Engagement Strategic Group
    - Formed August 2021
    - Supports ADR-UK's wider aim of meeting societal needs by consulting with organisations that shape public policy and deliver public services

- Chance for public input on attitudes towards administrative data
  - Opportunity of understanding of what is important to go both ways
  - Ethically sound

- Further Benefits for WED
  - How to engage with public
  - Raise awareness and increase public acceptability

- WED PESG Members
  - Charitable foundation: Joseph Rowntree Foundation
  - Trade Union: TUC
  - Equality and Diversity: Business Disability Forum
  - Recruitment Company: Indeed

- # Website/Social Media Development
  - www.wagedynamics.com
  - Twitter: @Wage_Dynamics

  - Autumn Publications
    - On an ad-hoc basis (as and when complete)
    - Notes and blog pieces on research findings (October 2021)

  - Future Developments
    - Documentation
    - Quality Assurance documents
    - Training sessions & ASHE quick start user guide

Future Plans: Wage and Employment Dynamics 3 (WED 3)

- WED3: complete the ambitious programme of developing a world-class, research ready, wage and employment data spine
  - DWP data (linked to the ASHE population) held within ONS
  - Census2021 linked to the existing ASHE-Census2011 data
  - LEO and/or HESA data
  - Migration Worker Scan data held at ONS
- Flexible and cautious delivery timetable is planned
  - Feasibility study ahead of data delivery
- Expression of interest submitted to ADR England - September 17th (initial decision due: mid-October)

	Timetable															
	2022			2023				2024				2025				2026
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
HMRC continuation																
Census 2021/ Scottish Census																
HESA/LEO																
Migration																
DWP																
Feasibility studies																
Fixed support																

Wage & Employment Dynamics

THE WED PROJECT

ESRC Working Lives Grant Proposal

Contact: c.a.singleton@reading.ac.uk

ESRC – Transforming Working Lives Call (£720k – Submitted today, for 3 years from April 2022)

- Building research capacity using the new linked datasets - adding Simonetta Longhi (Uni. of Reading), Rebecca Riley (KCL) & Post-doc
- Addressing relevant and potentially impactful research questions
- Using the longitudinal ASHE-Census-BusinessData

- Will address the following major research areas:
  - A. Do firms continue to share their rents with workers and, if so, under what conditions?
  - B. Who is really paying for workplace pension plans (and auto-enrolment)?
  - C. How has labour mobility changed over time and what are its implications for wage formation?
  - D. How much does who works for whom explain wage and progression gaps for minority workers?

- Embedding Policy Engagement Throughout, for example:
  - Strategic Advisory Group including DWP, TUC, EHRC, LPC

Thank you