Wage & Employment Dynamics
THE WED PROJECT

METHODOLOGY PAPER

Longitudinal attrition in ASHE

Abstract

The Annual Survey of Hours and Earnings (ASHE) provides many of the UK's official earnings statistics. The survey operates on an annual 1% sample of employee jobs. However, the method of sampling - based on the final two digits of an employee's National Insurance number – means that records are linkable longitudinally. Many government and academic studies have utilised the dataset in this way. However, the longitudinal integrity of the ASHE sample has been the subject of little prior investigation, with the panel sample generally assumed free of any attrition biases that might compromise longitudinal analysis. We explore the validity of this assumption by comparing rates of year-on-year sample retention in ASHE with rates of employment retention estimated from a reference dataset (the Longitudinal Annual Population Survey). Our analysis confirms the existence of systematic patterns of longitudinal attrition in ASHE, which have the potential to introduce bias into longitudinal analyses of these data. We go on to construct longitudinal weights that correct for estimated attrition biases over adjacent years in ASHE. In an illustrative analysis, the application of these weights brings about a small widening of the distribution of individual wage growth.

John Forth, Van Phan, Lucy Stokes, Alex Bryson, Felix Ritchie, Damian Whittard and Carl Singleton

Version 1: 13th April 2022

Contents

1. Introduction 3
2. A framework for considering sample exit in ASHE 5
Route 1: Sample member moves out of scope 6
Route 2: Employee is not sampled 6
Route 3: Employer not traced 6
Route 4: Employer does not respond 6
Route 5: Repeated observations are not appropriately linked 7
3. Rates of sample exit in ASHE 8
4. A comparison of sample exit in ASHE and employment exit in the APS 10
5. Examining potential routes of longitudinal sample attrition 14
Route 2: Employee is not sampled 14
Route 3: Employer not traced 15
Route 4: Employer does not respond 15
Route 5: Repeated observations are not appropriately linked 17
6. Benchmarking patterns of sample exit in ASHE against the Longitudinal APS 19
Introduction 19
Method 19
Results 20
7. Constructing two-period weights to address attrition bias 24
Method 24
Results 25
8. Limitations and possible extensions 28
Extending the weighting adjustment to pairs of non-adjacent years 28
Addressing attrition bias in estimates of employment exit using ASHE 29
9. Summary and conclusions 30
Appendix A: ASHE sample design and survey administration 33
Appendix B: What other characteristics are associated with employment/sample exit? 37

1. Introduction

The Annual Survey of Hours and Earnings (ASHE), conducted by the Office for National Statistics (ONS), is a critical source of information on earnings of employees in the UK. Introduced in 2004, ASHE replaced the former New Earnings Survey (NES) as the ONS' main source of information on earnings. As well as forming a key source for ONS' labour market statistics, ASHE is widely used to inform policymaking and evaluation across a range of government bodies. It is used *inter alia* by the Low Pay commission (LPC) to monitor the impact of the minimum wage, by the Department of Work and Pensions (DWP) to analyse pension changes and by the Office for Manpower Economics to inform public sector pay reviews, among others.

The target population for ASHE is all employee jobs in the UK, including all industries and occupations.[1] The sample for the survey is drawn each year from the HMRC PAYE register. All PAYE-registered jobs held by employees with a National Insurance (NI) number ending in a particular two digits are selected from the register; if the employee holds multiple jobs on the register, all jobs held by that person are selected. The selected sample is therefore a one per cent, simple random sample of employee jobs. The survey itself is completed by employers, who are contacted by ONS in April each year (with surveys typically dispatched to employers in the second week of April) and asked to provide information about the employee's earnings and paid hours of work.[2] Although mandatory, the survey has an annual response rate of around two-thirds. Further details on the sample design and survey administration are provided in Appendix A; Stokes et al (2022) expands on these points and also considers the cross-sectional representativeness of ASHE.

The method of sampling for ASHE means that individuals with eligible NI numbers are selected into the issued sample in every year that they are in PAYE employment on the date of sample selection. In those years in which the employee's employer responds to the survey, the survey thus generates observations that can be linked over time via the employee's NI number, generating longitudinal (panel) data on employees and/or jobs. However, as the annual response rate is less than 100%, there is the potential for missing data in the panel dimension of ASHE, even for employees who remain in scope (i.e. in PAYE employment) in more than one year.

There is considerable interest in using the panel dimension of the data, and indeed a number of studies have already done so. These studies typically fall into one of two groups:

- One group of studies focuses on the subset of ASHE sample members who appear in the dataset in multiple years (typically two consecutive years) and uses this subset to make inferences about the wider population of persons who have employee status at both *t1* and *t2*. The focus of this literature is typically on the extent and determinants of wage progression within jobs and firms. Examples include Elsby et al. (2016) and Schaefer and Singleton (2019).
- A second group of studies focuses on the subset of ASHE employees who appear in the dataset at *t1*, but who do not appear at *t2*. This subset is used to make inferences about the wider

[1] ASHE covers the whole of the UK, but fieldwork for Great Britain and Northern Ireland are conducted separately, with ONS carrying out the survey for Great Britain and the Northern Ireland Statistics and Research Agency (NISRA) doing so for Northern Ireland. The data available for use within this paper relates to Great Britain only.
[2] https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualsurveyofhoursandearningsashe

population of employees who leave employee status. Examples include Dickens et al (2015) and Stokes et al (2017).[3]

In order for the first group of studies to arrive at unbiased estimates of wage progression and its determinants, they need to assume that those employees who appear in ASHE at both *t1* and *t2* are representative of those who are in PAYE employment at both time points. Similarly, in order for the second group of studies to arrive at unbiased estimates of the factors associated with leaving employee status, they need to assume that those employees who exit the ASHE sample between *t1* and *t2* are representative of those who exit PAYE employment. To the best of our knowledge, these assumptions remain untested. This is, potentially, an important omission, since non-random attrition in the panel dimension of ASHE could introduce bias into the estimates from either set of studies.[4]

In this paper, we examine the panel characteristics of the ASHE dataset, making inferences about the rates and correlates of longitudinal sample attrition. We use the term 'longitudinal sample attrition' to refer to the situation in which an employee who is in scope to the survey at *t1* and in-scope at *t2* (i.e. who remains in PAYE employment) is only observed at *t1* (see Box 1). One obvious reason for such sample attrition may be employer non-response.

Box 1: Definition of terms used in the paper

For all sample members in employment at t1 and observed at t1:

- *Sample exit* = not observed at t2
- *Employment exit* = not in PAYE employment and so out-of-scope at t2
- *Longitudinal attrition* = in PAYE employment and so in-scope at t2 but not observed at t2

Sample exit = employment exit + longitudinal attrition

Note: the definition of employment exit is based on PAYE employment (i.e. employee status); a person who exits PAYE employment for self-employment is considered to have 'exited employment' under this

Any analysis of longitudinal attrition must have some means of distinguishing patterns of attrition from patterns of employment exit. Ordinarily, one might make such inferences by referring to fieldwork data on whether each sample member is in-scope or out-of-scope to the survey each year, and whether in-scope units yield survey responses. However, ONS do not make such fieldwork data available to researchers. Accordingly, we rely on comparisons with a benchmark dataset. Specifically, we compare the rates and correlates of sample exit between pairs of years in ASHE with the rates at which people exit employee status between pairs of years, estimated from the Longitudinal Annual Population Survey (APS). The Longitudinal APS provides a good reference point as it is currently the

[3] There is arguably a third group, focusing on employer separations (that is, the end of a job with a particular employer) (e.g. Hijzen et al, 2010; Dickson and Papps, 2016). However, as an employer separation could result in either non-employment or another employee job, the issues are essentially the same as those faced by the two groups of papers already mentioned.
[4] Specifically, estimates based on longitudinal analysis of ASHE will be biased if the factors associated with longitudinal attrition are correlated with any of the characteristics or behaviours that one wishes to measure using the longitudinal dataset. In this instance, the factors associated with longitudinal attrition are said to be "non-ignorable" for the purposes of estimating the characteristic or behaviour in question.

largest publicly-available dataset capturing the annual labour market transitions of successive cross-sections of the GB population.

The comparison shows that an individual's probability of exiting the ASHE sample from one year to the next is around 17 percentage points higher than their probability of exiting employee status estimated from the APS (25 per cent rather than 8 per cent, on average). More importantly, however, the correlates of sample exit in ASHE differ from the correlates of employment exit estimated from the APS. For instance, younger employees, those on low wages and those working relatively few hours are more likely to exit the ASHE sample than one would anticipate, based on employment transitions seen in the APS. The implication is that longitudinal attrition in ASHE is non-random and the ASHE longitudinal dataset is not then fully representative of those employees who remain in, or leave, PAYE employment from one year to the next. This longitudinal attrition may be "non-ignorable" in any longitudinal analysis of ASHE, with the potential to bias any conclusions drawn from the data.

We use this analysis to lay the foundations for the construction of longitudinal weights that correct for estimated attrition biases in longitudinal samples of ASHE employees observed in two successive years. Such longitudinal samples are a typical basis for the analysis of individual wage progression. In an illustrative analysis, the application of these longitudinal weights brings about a small widening of the distribution of annual wage progression. These weights are made available via the WED project to all those seeking to undertake future longitudinal analysis of the ASHE data. We also propose a method of accounting for longitudinal attrition when using ASHE to study the determinants of employment exit across two successive years.

Methods of addressing longitudinal attrition over longer time-periods will be developed in a later phase of the WED project, once we gain access to PAYE data from HMRC's Real-Time Information system. These data will enable employment transitions over periods of more than one year to be identified.

The paper proceeds as follows. In Section 2, we set out a framework for considering sample exit in ASHE, indicating five mechanisms that may cause a sample member to depart from the ASHE sample between *t1* and *t2*. In Sections 3, 4 and 5, we examine annual rates of sample exit and look for evidence of the five mechanisms working in practice. In Section 6, we benchmark patterns of sample exit in ASHE against patterns of employment exit observed from the Longitudinal APS. Section 7 discusses the construction of two-period weights to address the apparent attrition biases in ASHE, whilst Section 8 discusses extensions and limitations. Section 9 concludes.

2. A framework for considering sample exit in ASHE

If we are to understand longitudinal attrition in the ASHE sample, we must first understand the nature of the ASHE survey design and the routes through which an employee may exit the sample.

In what follows, we focus on each employee's appearance in the *longitudinal* sample. We focus on the continued appearance of individuals, rather than the continued appearance of jobs; this is because some questions have been raised about the ability of the 'same job' variable in ASHE to reliably link jobs across years.[5] In practice, there are few employees with multiple jobs in ASHE (typically 2 per cent

[5] The WED project is constructing a job identifier that will better enable jobs to be tracked over time.

in any given year). Multiple job holders are included in our analysis, but their employment characteristics at any given point are described on the basis of the main job only (defined as the job with the highest hours).

Exit of an employee from the ASHE sample between *t1* and *t2* can arise through one of five routes:

1. Sample member moves out of scope to the survey between *t1* and *t2*
2. Sample member remains in scope at *t2* but is not sampled
3. Sample member remains in scope and is sampled at *t2* but cannot be traced
4. Sample member remains in scope, and is traced at *t2*, but employer declines to respond
5. Sample member remains in scope, is traced and employer responds at *t2*, but repeated observations on this sample member are not appropriately linked in the dataset.

Route 1 corresponds to a genuine exit from PAYE employment. Routes 2-5 correspond to different forms of sample attrition and have the potential to introduce bias into the longitudinal sample if not accounted for. We consider each of the five routes in turn.

Route 1: Sample member moves out of scope

The survey population extends to all employee jobs that are PAYE-registered (see above). Sample exit will then naturally occur where individuals leave PAYE employment to retire, become unemployed, or move into self-employment (as long as that person then exits the HMRC PAYE Register). We call this "employment exit" (see Box 1). As noted above, employment exit is a legitimate reason for someone to exit the ASHE sample, as that person has moved out of scope to the survey.

Route 2: Employee is not sampled

For the 2007 and 2008 surveys, the ASHE sample was cut by 20 per cent due to budget constraints. This clearly has consequences when trying to follow individuals over time in this period. This cut to the sample was focused on particular industries, targeting those which had shown the least variation in earnings (ONS, 2008). The sample was restored to its original size for the 2009 survey onwards. An employee who was observed in 2006 and remained in scope to the survey throughout this period may therefore drop out of the ASHE dataset simply as a result of the sample cut.

Route 3: Employer not traced

In order to trace the employer of the sampled employee, the sample taken from the HMRC PAYE register is matched against ONS's Inter-Departmental Business Register (IDBR) in order to obtain contact and address details for the employer. Attrition from *t1* to *t2* may therefore occur if the employer details are out of date at the point of sampling (for instance, if the employer has undergone a change of address, the survey form may not then be received). Attrition may also occur if the employee moves jobs between sampling and fieldwork and the survey forms are sent to the previous employer, rather than to the new employer.[6]

Route 4: Employer does not respond

While ASHE is a mandatory survey under the Statistics of Trade Act 1947, overall levels of non-response are estimated by ONS to sit at around 35% in a given year (see Appendix A). In practice, this means that around 65% of employers respond before the survey cut-off date, set by ONS in order to

[6] The sample is first drawn in January but, since 2004, a second sample has been drawn from the PAYE register in April in an attempt to address this issue (see Pont, 2007).

meet its deadlines for the production of annual statistics. If an employer responds on time at *t1*, but responds late or not at all at *t2*, this will lead to sample exit for the employee through temporal employer non-response. There may also be employers who typically respond whenever approached to complete the survey ("compliant firms"), and those who typically do not ("uncompliant firms"); sample exit due to non-response may thus also occur if an employee moves from a "compliant firm" to an "uncompliant firm". Compliance with official business surveys is typically positively associated with: employer size; public sector ownership; and administrative capacity. Furthermore, in ASHE, response is semi-automated for some large firms (so-called 'special arrangements', whereby the employer provides data electronically to the ONS rather than completing the survey form). Sample exit due to non-response may thus be more likely when employees move from larger to smaller firms, from the public to the private sector, and from firms with high levels of administrative capacity to those with less. Each of these factors may be correlated with average earnings or hours.

It should be noted at this point that ONS appear not to focus on preserving the longitudinal integrity of the sample. Their principal focus is the cross-sectional representativeness of ASHE. The panel element of the ASHE dataset is a convenient by-product of the ASHE sample design but the construction of longitudinal data is not the *raison d'etre* of the survey. The approach taken in the fieldwork for ASHE therefore differs fundamentally from a classic longitudinal survey (e.g. the UK Household Longitudinal Study), where explicit efforts are made to minimise longitudinal sample attrition by retaining original sample members over successive waves.

Route 5: Repeated observations are not appropriately linked

Finally, it is possible that an employee who remains in PAYE employment, and whose employer does respond to the survey, may not have their records linked longitudinally on the ASHE dataset. There would seem to be the potential for such linkage errors when employees are issued with temporary NI numbers, and then later given permanent NI numbers. Temporary NI numbers were a common occurrence prior before 2001, as employers were allowed to allocate new employees with a temporary number whilst awaiting the permanent number from HMRC. However, this practice is reported to have ceased in April 2001.[7] Linkage errors may also occur if there are year-specific errors in generating the unique personal identifier (PIDEN) that replaces the NI number on the publicly-available dataset.

In practice, it is not possible to distinguish categorically between all five forms of sample exit in ASHE, as no information is provided by ONS on the issued sample or fieldwork outcomes. What is more, to our knowledge, no efforts are made within ONS to track sample retention from one year to the next.

Our analysis of sample exit therefore relies primarily on seeking out circumstantial evidence in the data, and on comparing what is observable in the ASHE longitudinal data with what might be expected from patterns of employment exit estimated from a benchmark source (the Longitudinal APS).

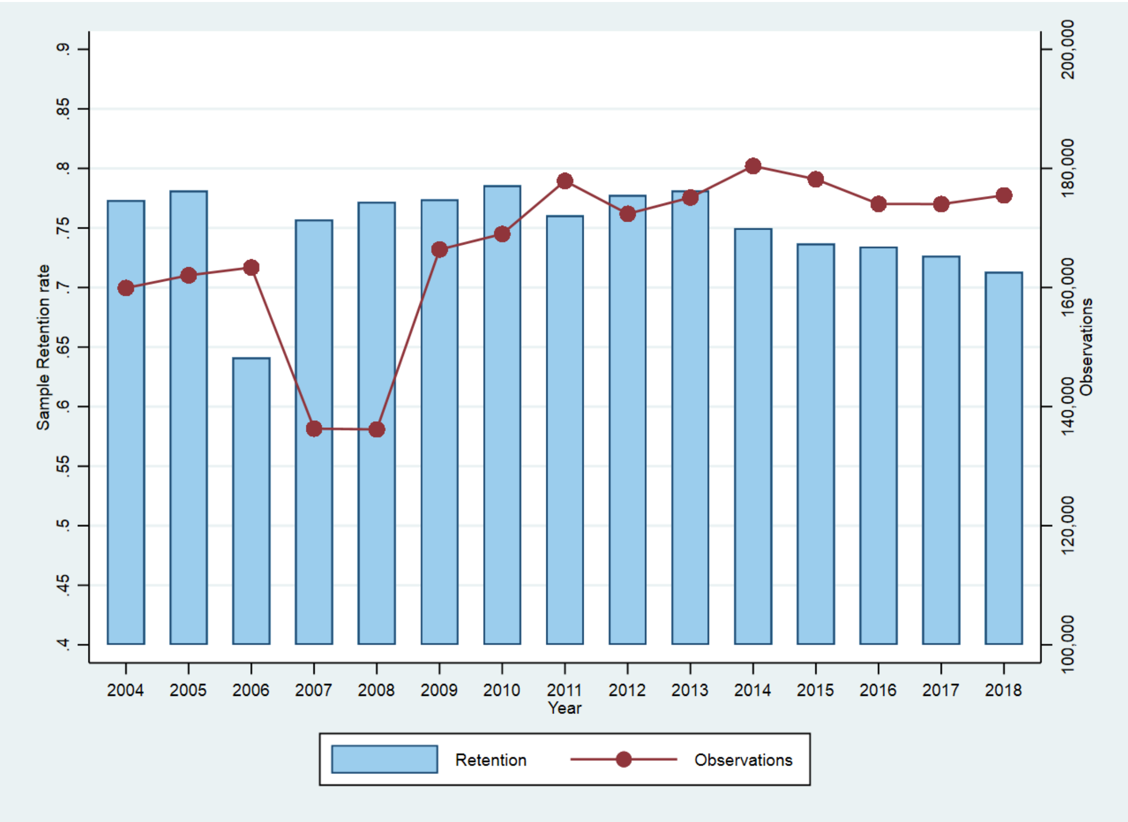
[7] See DWP (2013). We understand that the practice may have been temporarily reinstated in 2020, due to COVID placing restrictions on HRMC's ability to issue new, permanent NINOs. However, this occurs outside the period considered in this paper.

3. Rates of sample exit in ASHE

We start our analysis by simply exploring sample exit from one year of the survey to the next, that is: the share of employees observed in ASHE in year *t* who are also observed in ASHE in year *t+1*. Figure 1 shows the percentage of sample members retained in ASHE on this basis, covering the period from 2004, which is the year in which various methodological changes were made as part of the transition from the New Earnings Survey (NES) to ASHE (see Bird, 2004). The figure shows, for example, that 77 per cent of sample members from the 2004 ASHE dataset are present in the 2005 ASHE dataset. The retention rate fluctuates over time, but on average stands at around 75 per cent over the period 2004 to 2018. We see a drop in retention in 2006 to around 64 per cent, which coincides with the cuts to the sample made at this time. There are also signs of a monotonically declining retention rate from 2013 onwards.

The broad picture is of a relatively high level of sample exit from year *t* to year *t+1*, standing at around 25 per cent on average over this period, and at around 28-29 per cent in more recent years (Figure 1).

Figure 1: Sample retention rate for individuals from year t to year t+1 (unweighted figures)



Source: ASHE

In view of the substantial interest in understanding wage progression and employment retention over the medium-term, we also investigate sample exit rates over periods longer than one year. Table 1 shows the sample retention rate extending over different time horizons (*t+1*, *t+2*, …, *t+i*) for the sample observed in each year from 2004-2018. For example, when taking the year 2004 as a start year, we see that 77 per cent of sample members observed in year 2004 also appear in ASHE in year

2005, and 72 per cent of those observed in year 2004 also appear in ASHE in year 2006, and so on. On average, 75 per cent of employees observed in year *t* are also observed in year *t+1*, 68 per cent of employees observed in year *t* are observed in year *t+2*, 64 per cent are observed in year *t+3* and so on. Around half (49 per cent) of employees are observed 10 years later, and 39 per cent are observed 15 years later.[8]

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. This illustrates two things: first, retention rates drop with the length of the observation window; and second, retention rates are generally lower in more recent periods.

The figures presented in Table 1 do not require employees to have been observed in all intervening years. Such temporary disappearances may arise because of non-response or because of a period of non-employment. Restricting our attention to continual membership of the observed sample, we find that only six per cent of employees observed in 2004 are observed in all years to 2019 (see Table 2).

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

[8] Necessarily, the figures in the first column of Table 1 mirror those shown in Figure 1.

Table 2: Continuous 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	63.1	43.1	36.1	30.9	26.3	22.8	19.5	17.1	15.1	13.0	11.1	9.5	7.7	6.4
2005	78.1	52.4	43.1	36.5	30.8	26.6	22.6	19.9	17.4	14.9	12.7	10.9	8.8	7.3	
2006	64.1	51.3	42.8	35.8	30.7	26.0	22.7	19.8	16.9	14.3	12.2	9.9	8.1		
2007	75.7	61.5	50.8	43.3	36.3	31.5	27.4	23.2	19.6	16.6	13.5	11.1			
2008	77.2	62.3	52.5	43.6	37.6	32.5	27.5	23.1	19.5	15.9	13.1				
2009	77.4	63.7	52.2	44.4	38.0	32.0	26.7	22.4	18.3	15.1					
2010	78.6	62.8	52.6	44.7	37.2	30.9	25.8	21.0	17.3						
2011	76.0	62.1	51.9	42.7	35.1	29.1	23.4	19.2							
2012	77.8	63.6	51.6	42.0	34.6	27.7	22.6								
2013	78.1	61.9	49.5	40.3	32.1	26.1									
2014	75.0	58.4	46.6	36.9	29.7										
2015	73.7	57.1	44.6	35.4											
2016	73.4	56.1	43.9												
2017	72.7	55.2													
2018	71.3														
Average	75.1	59.4	48.1	40.1	33.9	28.9	24.6	21.0	18.0	15.3	12.9	10.7	8.8	7.5	6.4

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

4. A comparison of sample exit in ASHE and employment exit in the APS

As noted above, sample exit will naturally occur where individuals leave PAYE employment to retire, become unemployed, or move into self-employment (as long as that person then exits the HMRC PAYE Register). This is Route 1 in the framework set out in Section 2.

The WED project is in the process of linking ASHE to PAYE data from HMRC's Real-Time Information system, which will enable such transitions to be identified at the level of the individual sample member. However linking is not expected to take place until late 2022 and so, for now, we explore the extent to which sample exit in ASHE might be explained by employment exit through a comparison with the Longitudinal APS.

The Longitudinal APS is a large household survey that is unusual in allowing us to estimate annual labour market transitions among cross-sectionally representative samples of employees. Data are available for each year from 2012 to 2017.[9] Further description of the Longitudinal APS is provided in Box 2. As described in Box 2, the APS dataset only contains individuals who have participated both in year *t* and year *t+1* of the APS; sample attrition is therefore unobservable to the analyst. However, the Longitudinal APS includes longitudinal weights that adjust for observable attrition biases (based on a methodology outlined by Clark and Tate, 1999). The availability of these longitudinal weights allows

[9] Data from the Longitudinal Labour Force Survey (LLFS), which is nested within the LAPS, is currently available for one further year (2018/19). However, the LLFS sample is much smaller than that provided by the LAPS, so we prefer to focus on the shorter time period covered by the LAPS.

us to use the Longitudinal APS to generate estimates that are representative of the population across successive years; we use these weights in all subsequent analyses reported below.

Box 2: The Longitudinal Annual Population Survey (LAPS)

The Annual Population Survey (APS) is a household survey gathering information on a range of topics relating to an individual's labour market behaviour. The topics covered include employment and unemployment, as well as housing, ethnicity, religion, health and education. The survey extends to all private households in Great Britain, with the exception of some communal households, though labour market information is only collected from those aged 16 and above.

The APS is not a stand-alone survey. One part of the sample comes from the Quarterly Labour Force Survey (QLFS); households provide data to the APS in the first quarter that they enter the QLFS sample (Wave 1) and also in the fifth and final quarter before they leave the QLFS sample (Wave 5), thus providing one observation to the APS in each calendar year. The remainder of the APS sample is compiled from a local sample boost (the Local Labour Force Survey or LLFS), which uses the same core questionnaire as the QLFS. Each household participating in the LLFS is surveyed once per year over four years.

The annual APS samples each provide observations on around 110,000 individuals in employee jobs. Around half of these respondents derive from the QLFS, with the remaining half deriving from the LLFS. As the description above indicates, the APS sample necessarily includes a longitudinal component, whereby sample members deriving from the QLFS can observed in two consecutive years, and sample members deriving from the LLFS in up to four consecutive years. The ONS uses this feature of the sample to make available two-year longitudinal APS datasets (hereafter, referred to as the Longitudinal Annual Population Survey or LAPS). For each respondent, these datasets provide one observation in calendar year *t* and a second observation in calendar year *t+1* (five quarters after the first observation). The panel dataset can then be used to observe labour market transitions (and other changes) by comparing the respondent's status at these two time points. Each two-year LAPS panel dataset typically offers an achieved sample of around 40,000 individuals who are in employee jobs in the first year.

Naturally, there is some sample attrition over successive waves of the QLFS and LLFS (see ONS, 2021a: Table 11), as well as some item-level non-response. For sample members deriving from the QLFS, missing data are imputed by using roll-forward methods, but typically only for Wave 5. This means that, if a respondent cannot be contacted for Wave 5, then the information that the respondent provided at Wave 4 is used to impute a response for the current interview. To minimise non-response bias, where there are other individuals in the household who can answer on behalf of the absent respondent, proxy responses are also collected. Responses from the annual boosts are not rolled forward due to time elapsed since the response.

continued

Box 2 continued

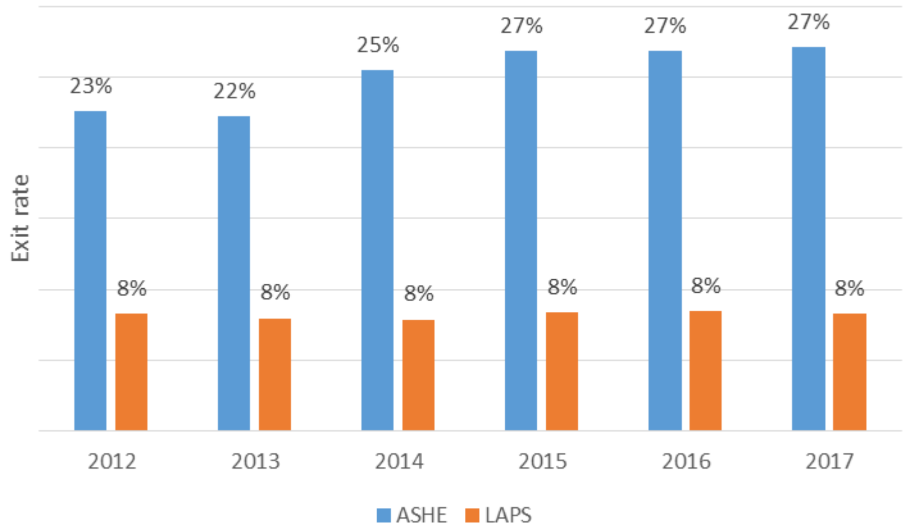
Sample attrition is addressed through the construction of longitudinal weights. The approach to weighting derives largely from a non-response study by Tate (1999) which showed that non-response was principally associated with age (higher for those aged 18-24) and housing tenure (higher for those in private rented accommodation). Economic activity and marital status were also found to play a smaller role (non-response being higher among the unemployed, those temporary employment and single persons). The derivation of weights (described by Clark and Tate, 1999) then involves two steps: (i) a set of 'prior weights' are first calculated and scaled to replicate the housing tenure distribution observed in year one; (ii) the weights are then calibrated to four sets of control totals based on the distributions of age, sex, region and economic activity in year t and t+1. Further information on the APS is provided by ONS (2012).

The weighted Longitudinal APS data allow us to estimate, in each year from 2012-2017, the share of all employees aged 16-68 in year *t* who are not in an employee job 12 months later in year *t+1* ("employment exit").[10] Estimated rates of employment exit from the Longitudinal APS are presented alongside rates of sample exit from ASHE in Figure 2. Overall, the Longitudinal APS indicates that, across the period 2012/13-2017/18, 8 per cent of employees in employment in year *t* were not in an employee job 12 months later. Among the 8 per cent who had exited employee status, one quarter (2 per cent) were self-employed, one quarter (2 per cent) were unemployed and half (4 per cent) were inactive (e.g. had retired, entered full-time education).

Across the same period, on average, 25 per cent of employees who appeared in ASHE in year *t* did not reappear in ASHE in the following year. This average rate of sample exit in ASHE is thus around 17 percentage points higher than the rate of employment exit estimated from the APS, implying that the majority of sample exit in ASHE is, in fact, due to sample attrition rather than true exit from PAYE employment. The comparison suggests that around one in every six employees in ASHE disappears from the sample between year *t* and year *t+1* due to sample attrition. This rate also appears to be increasing over time.

[10] We define this as follows: conditional on being in an employee job in year *t* (INECAC051=1), the person is observed within the longitudinal APS 12 months later, but is not in an employee job at that point (INECAC052≠1). We use APS data for calendar years (January-December); rates of employment exit vary across the year but, as we are summing across a 12-month period, we expect that the slight difference in sample period between the APS (Jan-Dec) and ASHE (Apr-Mar) is ignorable.

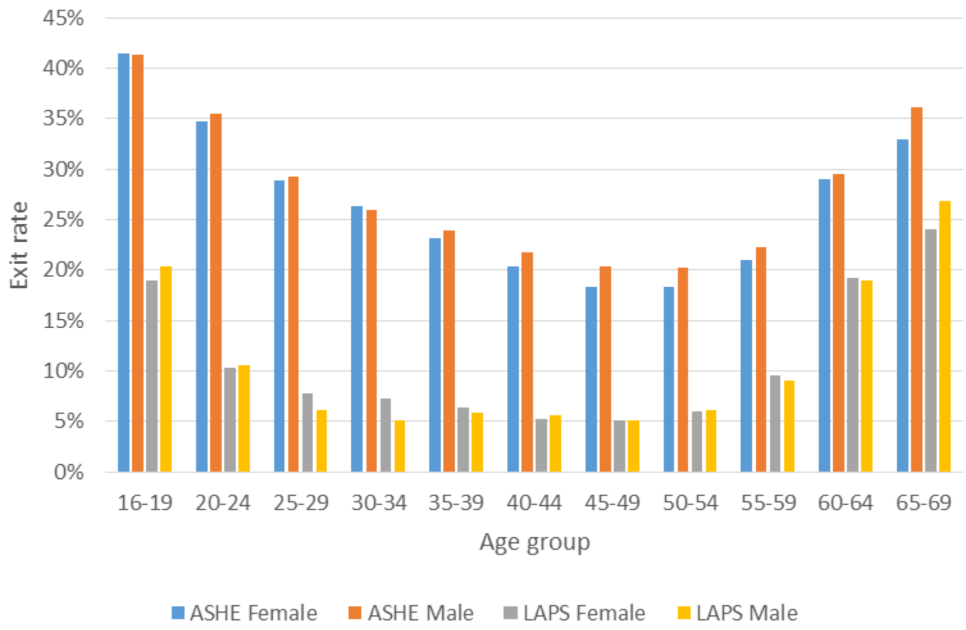
Figure 2: Employment exit in the Longitudinal APS and sample exit in ASHE, 2012-2017



As we would expect, the Longitudinal APS shows that there is a strong age-related component to employment exit, arising from the retirement of older workers and the entry into full-time education of young workers (Figure 3). The same broad age-related pattern is evident in ASHE, but the ratio of exit rates across the two sources is larger in the middle of the age range. Specifically, the sample exit rate in ASHE is around 1.5 times the employment exit rate in APS rate among older workers, but roughly double among young workers and larger by a factor of around four among the middle-aged. Thus, sample attrition is disproportionately affecting younger and middle-aged workers in ASHE. Figure 3 is also notable in showing higher rates of employment exit among female workers than male workers in the age range 25-39, possibly relating to maternity. There is no such difference in ASHE; in fact, rates of sample exit are generally higher for male employees in most age groups, indicating that sample attrition is more likely to affect men than women.

These patterns are indicative of substantial year-to-year sample attrition in the longitudinal ASHE data. This attrition is apparent across all age groups and among both genders, but appears to be stronger amongst middle-aged workers and, to a smaller extent, also among younger workers and amongst men. These patterns imply that employers with above-average shares of young or middle-aged workers, and above-average shares of male workers, are less likely to respond to ASHE, on average.

Figure 3: Employment exit in the Longitudinal APS and sample exit in ASHE, by age and gender, 2012-2017



5. Examining potential routes of longitudinal sample attrition

Having shown that rates of longitudinal sample attrition in ASHE are substantial, and non-random, we now go on to look for evidence of the causes. We consider each of the four potential routes of sample attrition discussed in the framework set out in Section 2.

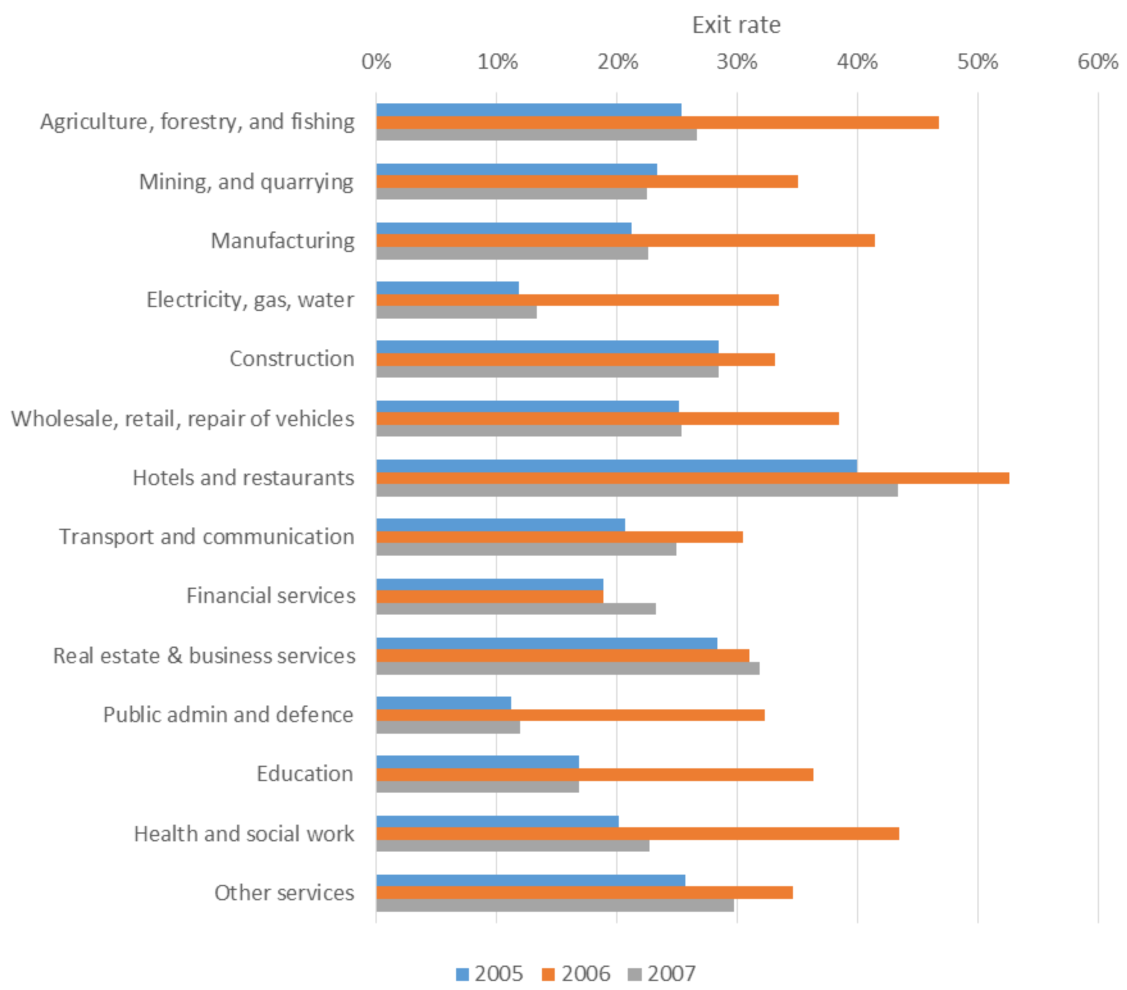
Route 2: Employee is not sampled

We can show the impact of the 20 per cent cut to the ASHE sample in 2007 by looking at sample exit rates in ASHE by industry around that time. The list of industries subject to the sample cut have not been widely published by ONS, and so we are unable to identify them precisely. However, Figure 4 shows sample exit rates by SIC(2003) industry sector for 2005, 2006 and 2007, with the impact of the sample cut being quite evident. Sample exit rates in Primary industries, Manufacturing, Public Administration, Health and Education increase by around 20 percentage points in 2006-7, before returning to their original levels in 2007-8. Other industry sectors see smaller increases in exit rates, indicating that the sample cut was only applied to some parts of these industries. Financial services and Real estate and business services appear unaffected.

Ideally, we would compare rates of sample exit in ASHE by industry over this period with rates of employment exit in APS in the same years. This would indicate whether any of the increase in sample exit rates in 2006-7 were, in fact, due to increased employment exit (which may be the case if industry-specific employment shocks occurred around the time of the sample cut). However, the Longitudinal APS is only available from 2012. ONS employment statistics shows stability in the numbers of jobs by industry sector in 2005, 2006 and 2007 (ONS, 2021b) and so it seems reasonable to assume that most, if not all, of the increases in sample exit seen here are the product of changes in the sample design.

Sample exit rates in ASHE therefore rose sharply in 2006 as a result of a change in the sample design, but the effects were felt more heavily in some parts of the sample than others. This was a temporary effect: the individuals who were removed from the sample in 2006 appear to have been reintroduced in 2008 (see Table 1). However, any analysis of the longitudinal ASHE over the period 2006-8 should take account of the fact that 20 per cent of sample members exit the sample for reasons unrelated to their probability of exiting employment (with the industry composition of the retained sample being skewed accordingly).

Figure 4: Sample exit rates in ASHE by industry sector, 2005-2007



Route 3: Employer not traced

Sample attrition may occur if the ONS cannot trace the employer of the sampled employee, which may occur (for instance) if the employer has undergone a recent change of address. However, we have no way of evidencing this and so we do not pursue this route any further.

Route 4: Employer does not respond

Employer non-response in ASHE may be of two basic types:

1. Non-response within an employment spell: The employer responds in year *t* but not in year *t+1*, even though employee continues in employment with the same employer

2. Non-response associated with the end of an employment spell: A person is employed by a responding employer in year *t* but moves jobs to be employed by a non-responding employer in year *t+1*.

In the first case, if non-response is occasional (i.e. the employer does not cease to participate in ASHE in perpetuity), one would expect to find sample exits in year *t* that are followed by the reappearance of the same employee in year *t+n*, employed by the same employer as in year *t* and with an employment start date that predates *t*. We explore this issue over the period 2007-2017, to avoid the sample cut described above. We find that around two-thirds of all sample exits are followed at some point by the reappearance of that person in ASHE. In around one-third of these cases (24% of all exits that happen between 2007 and 2017), we find that the exiting employee reappears with the same employer and an employment start date that predates the date of exit.[11] In most of these cases (20% of all exits), the employment start date is identical to that recorded at the point of exit. For around three-quarters of these (15% of all exits), the gap is only one year.

There could be a number of explanations for such gaps. The employer may opt to respond in some years and not others, perhaps because other tasks occasionally take precedence. Alternatively, the employer may respond quickly in some years and slowly in others, such that their return arrives with ONS before the cut-off date in year *t* and year *t+2,* but after the cut-off date in year *t+1*. A regression analysis of the occurrence of these gaps finds that, among all sample exits, such 'false exits' are more likely to affect female employees and those in middle-age. They are also more likely to affect employees in large firms than those in small firms, in manufacturing than services and in the public sector than in the private sector.

Hence, there are occasional gaps in individuals' employment spells, accounting for around one fifth of all exits in a given year; this equates to approximately one third of the gap in exit rates between ASHE and APS. These gaps do not occur at random and indicate patterns of sample attrition that are consistent with employer non-response.

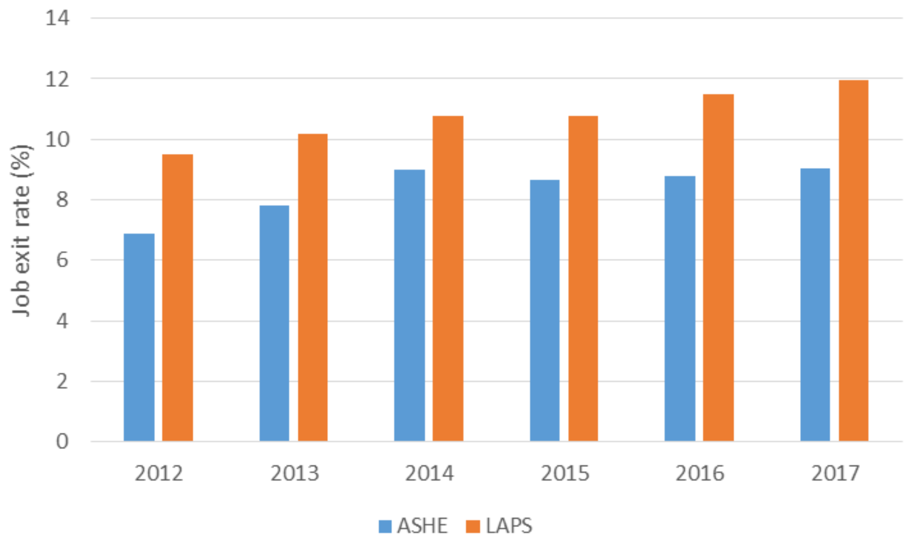
The gaps could feasibly by filled in via imputation, to improve the longitudinal integrity of the sample and reduce the bias in estimates of employment exit. However, one's ability to fill in such gaps necessarily becomes more limited as one approaches the end of the ASHE series (since an employee has more limited opportunities to reappear). Also, these in-filled records would have no pay or hours data, and so would be of little value in estimating wage progression.

The second type of non-response (occurring at the end of an employment spell), inflates the sample exit rate among job movers when compared with job stayers. If this type of non-response is occurring in ASHE, one would expect that the rate of job exit among those who remain in the sample over two consecutive years will be depressed (since job exiters are disappearing from the longitudinal sample at a higher rate than job stayers, all other things equal). A comparison of ASHE and LAPS over the period 2012-2017 finds that the share of ASHE employees who exit their job, when expressed as a share of those who remain in employment in year *t+1*, is around 8.4 per cent, whereas in LAPS it is 10.8 per cent.[12] The gap is evident in each of the six years (Figure 5). This is

[11] We use ENTREF to identify the employer and EMPSTA to identify the start date of the employment spell with that employer.
[12] In ASHE, we use EMPSTA to determine whether an employee remains with the same employer.

consistent with a scenario in which job changes that take employees from 'ASHE-compliant' employers to 'non-compliant' employers generate sample attrition through employer non-response.

Figure 5: Job exit rate among those remaining in an employee job, ASHE and LAPS, 2012-2017



Route 5: Repeated observations are not appropriately linked

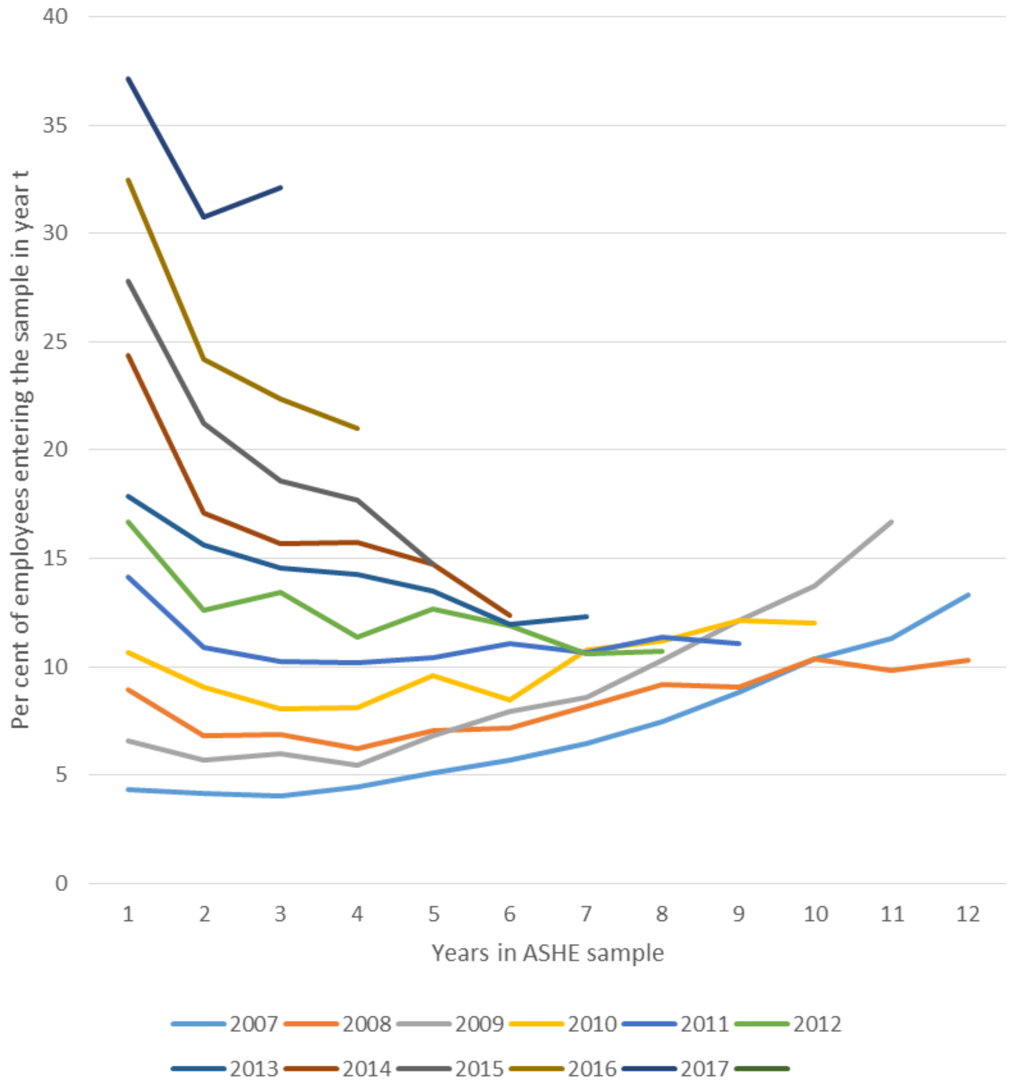
Linkage errors may occur if there are year-specific errors in generating the unique personal identifier (PIDEN) that replaces the NI number on the publicly-available dataset. If such errors are present in the data, one might expect to see PIDENs that appear for only one year. Such errors could, in fact, explain some of the gaps attributed to employer non-response under Route 4.

Computing the number of years across which each PIDEN appears in the data from 2007-2019, we find that one-in-seven (15 per cent) appear in only one year. Figure 6 shows, for each year in which a PIDEN first appears in the data within this 12-year window, the number of years that PIDEN is observed across this period. In most years, the share of PIDENs observed for only one year is high, relative to the share that appear for two, three or more years.[13]

Some of these PIDENs will relate to older workers who exit through retirement near the beginning of our observation window, or younger workers who enter employment for the first time near the end. However, even among males aged 30-44, who one might expect to have a reasonably continuous attachment to the labour market, we find that one in ten (11%) are observed for only one year across this 12 year period.

[13] The curve necessarily steepens in later years of entry because the window over which a PIDEN has the opportunity to be re-observed is shorter.

Figure 6: Number of years in ASHE sample, by year of entry, 2007-2017



If there were PIDEN errors, one might expect to be able to find a person with identical characteristics in the following year of the data. However, in only 1 per cent of sample exits are we able to identify a person in the following year of data with the same gender, firm ID (ENTREF), employment start date (EMPSTA) and workplace postcode (WPOST) as the person who exited the sample in the previous year. So whilst the relatively high rate of single-year PIDENs in ASHE is consistent with PIDEN errors, there is only weak evidence that these are, in fact, errors; instead, they appear to belong to individuals who have only a fleeting engagement with PAYE employment. One would need access to the underlying NINOs in order to verify with greater certainty whether there are errors in assignment of PIDENs over time.[14] But the current evidence suggests that PIDEN errors are not a major cause of sample attrition.

[14] Our forthcoming access to the HMRC RTI data will also provide a route to investigate this issue.

6. Benchmarking patterns of sample exit in ASHE against the Longitudinal APS

Introduction

The earlier discussion in Section 4 clearly indicated that annual rates of sample exit in ASHE are considerably higher than the rates at which people exit employment, estimated from the Longitudinal APS. More importantly, Section 4 provides at least some evidence that the correlates of sample exit in ASHE may differ from the correlates of employment exit estimated from the Longitudinal APS. This suggests that the raw ASHE longitudinal dataset may not provide a representative sample of those employee who remain in PAYE employment from one year to the next. If were to be the case, sample attrition in ASHE could be "non-ignorable", with the potential to introduce bias into longitudinal analyses of the ASHE data.

To investigate this issue more fully, we undertake a multivariate analysis to benchmark patterns of sample exit in ASHE against patterns of employment exit estimated in the Longitudinal APS across a wide range of observable characteristics. We then use this investigation to construct weights that seek to correct for estimated attrition biases in longitudinal samples from the ASHE dataset.

Method

Our approach involves, first, running a probit regression to estimate the probability of employment exit between year *t* and year *t+1* for each employee *i* in the Longitudinal APS:

$$Y^{*}_{APS} = \alpha + \beta X_{APS} + \varepsilon \qquad \text{(Eq. 1)}$$

where: Y^{*}_{APS} = 1 if the employee observed in the APS in year *t* is not in an employee job at *t+1* (zero otherwise); and X_{APS} is a vector of employee, job and employer characteristics that are measured here in the APS, but which are also common to ASHE. The regression is weighted using the APS longitudinal weights to account for sample selection and longitudinal attrition.

We use Eq. 1 to generate a predicted probability of employment exit - $\hat{p}(emp\_exit)$ - for each individual in ASHE. We then run the following OLS regression:

$$Y^{*}_{ASHE} - \hat{p}(emp\_exit) = \alpha + \beta X_{ASHE} + \varepsilon \qquad \text{(Eq. 2)}$$

where: Y^{*}_{ASHE} = 1 if the employee observed in ASHE in year *t* is not in the ASHE sample at *t+1* (zero otherwise); and X_{ASHE} is the same vector of employee, job and employer characteristics included in Eq. 1, but here measured in ASHE. This regression is weighted using the ASHE cross-sectional weights.[15]

The mean value of $Y^{*}_{ASHE} - \hat{p}(emp\_exit)$ provides an estimate of the rate of sample attrition in ASHE (i.e. extent to which the rate of sample exit in ASHE exceeds the rate that would be expected if employment exit were the only cause).

[15] We use the standard ASHE cross-sectional weight (calwght). We do not use the revised cross-sectional weight derived in Stokes et al (2022) as much of the work on longitudinal attrition described here pre-dated the derivation of that revised cross-sectional weight; however, we will seek to incorporate the revised cross-sectional weight in a future version of the paper.

In the regression of Y^*_{ASHE} - $\hat{p}(emp\_exit)$, the constant term α shows the estimated rate of sample attrition for an employee belonging to the reference category on all elements of X_{ASHE}; the coefficients β show the extent to which that rate of sample attrition varies across the elements of X_{ASHE}. If $\beta \neq 0$, this implies that the pattern of sample exit in ASHE associated with that element of X differs systematically from pattern of employment exit in APS associated with the same element of X. Such an outcome indicates the presence of attrition biases in the longitudinal ASHE sample. The approach is broadly equivalent to a forecast errors framework in which one seeks to identify the characteristics associated with departures from a predicted outcome.[16]

The analysis is run on pooled data from the period 2012-2017 – the period covered by published data from ASHE and the Longitudinal APS at the time of working.

Results

The results of Eq. 2 are shown in Table 3. Column [1] presents the coefficients from the regression model. Column [2] then presents F tests of the joint significance of each set of coefficients, so that the overall significance of a particular characteristic (e.g. age, occupation) can be assessed. These F tests indicate that there are statistically significant differences in relative exit rates across all characteristics. Differences by age group, tenure and year are shown to be particularly strong. The patterns suggest that employers with higher shares of male employees, higher shares of younger and middle-aged employees, and higher shares of employees with low tenure are less likely to respond to ASHE at *t2*, conditional on having responded at *t1*.

Table 3: Regression to estimate patterns of sample attrition in ASHE, 2012-2017

VARIABLES	[1]	[2]
Gender (Ref. Female):		
Male	0.012***	F(1)=148.85 p<0.01
	(0.001)	
Age group (Ref: 16-19)		
20-24	0.060***	F(10)=257.09 p<0.01
	(0.003)	
25-29	0.055***	
	(0.003)	
30-34	0.046***	
	(0.003)	
35-39	0.031***	
	(0.003)	
40-44	0.021***	
	(0.003)	
45-49	0.014***	
	(0.003)	
50-54	0.006*	
	(0.003)	
55-59	0.000	

[16] This approach is broadly equivalent to running separate regressions of Y_{APS} and Y_{ASHE} and then comparing the equality of the coefficients across the regressions (which can itself be achieved by pooling the APS and ASHE data and running a stacked regression in which all *X*s are interacted with a dummy variable identifying the ASHE sample members).

	(0.003)	
60-64	-0.018***	
	(0.004)	
65+	-0.048***	
	(0.004)	
Occupation (ref: Managers, directors and senior officials)		
Science, research, engineering and tech	-0.007***	
	(0.002)	
Associate professional and technical	0.014***	
	(0.002)	
Administrative and secretarial	-0.013***	
	(0.002)	
Skilled trades occupations	0.010***	
	(0.002)	F(8)=99.73 p<0.01
Caring, leisure and other service occupation	-0.006***	
	(0.002)	
Sales and customer service occupations	-0.031***	
	(0.002)	
Process, plant and machine operatives	-0.008***	
	(0.003)	
Elementary occupations	-0.030***	
	(0.002)	
Industry (ref: Sections A-E)		
F: Construction	0.024***	
	(0.003)	
G: Wholesale, retail, repair of vehicles	-0.002	
	(0.002)	
H; Transport, and storage	0.002	
	(0.002)	
I; Accommodation, and food service	0.074***	
	(0.003)	
J; Information, and communication	0.036***	
	(0.003)	
K: Financial and insurance activities	0.032***	
	(0.003)	F(14)=210.19 p<0.01
L: Real estate activities	0.028***	
	(0.004)	
M: Professional, scientific, and technical	0.036***	
	(0.002)	
N: Admin and support services	0.089***	
	(0.002)	
O: Public admin and defence	-0.010***	
	(0.003)	
P: Education	-0.005**	
	(0.002)	
Q: Health, and social work	0.006***	
	(0.002)	

R: Art, entertainment, and recreation	0.009**	
	(0.004)	
S: Other service activities	0.002	
	(0.004)	
Region of workplace (Ref: North East)		
North West	0.010***	F(10)=155.89 p<0.01
	(0.002)	
Yorkshire and Humberside	0.003	
	(0.002)	
East Midlands	0.012***	
	(0.003)	
West Midlands	0.020***	
	(0.002)	
South West	0.010***	
	(0.002)	
East of England	0.014***	
	(0.002)	
London	0.056***	
	(0.002)	
South East	0.023***	
	(0.002)	
Wales	-0.004	
	(0.003)	
Scotland	-0.003	
	(0.002)	
Sector of ownership (Ref. Private):		
Public	-0.015***	F(1)=112.07 p<0.01
	(0.001)	
Decile of real gross hourly pay (Ref: the lowest)		
2nd paydecile	-0.023***	F(9)=62.54 p<0.01
	(0.002)	
3rd paydecile	-0.028***	
	(0.002)	
4th paydecile	-0.027***	
	(0.002)	
5th paydecile	-0.034***	
	(0.002)	
6th paydecile	-0.047***	
	(0.002)	
7th paydecile	-0.045***	
	(0.002)	
8th paydecile	-0.048***	
	(0.002)	
9th paydecile	-0.051***	
	(0.003)	
10th paydecile	-0.039***	
	(0.003)	
Basic working hours (Ref: <=15)		

16-29	-0.030***	F(3)=129.64 p<0.01
	(0.002)	
30-47	-0.032***	
	(0.002)	
48 plus	-0.019***	
	(0.004)	
Tenure (Ref: <1 year)		
1-2 years	-0.027***	F(6)=536.50 p<0.01
	(0.002)	
2-5 years	-0.051***	
	(0.002)	
5-9 years	-0.074***	
	(0.002)	
10-20 years	-0.082***	
	(0.002)	
20 years or more	-0.085***	
	(0.002)	
Missing/invalid	-0.033***	
	(0.004)	
Workplace size (Ref: 1-24 employees)		
25-49	0.004**	F(4)=100.82 p<0.01
	(0.002)	
50-499	0.005***	
	(0.001)	
500+	-0.003*	
	(0.002)	
Missing	0.023***	
	(0.001)	
Year (Ref: 2012)		
2013	0.003**	F(5)=423.41 p<0.01
	(0.001)	
2014	0.032***	
	(0.001)	
2015	0.042***	
	(0.001)	
2016	0.044***	
	(0.001)	
2017	0.050***	
	(0.002)	
Constant	0.202***	
	(0.005)	
Observations	*1,054,135*	
R-squared	*0.03*	

Key to statistical significance in Column 1: *** p<0.01; ** p<0.05; * p<0.10. Robust standard errors in parentheses.

The effect of these differential patterns of sample exit will be to skew the longitudinal sample in ASHE away from the profile that would be expected on the basis of employment exit alone. In respect of age, for example, we could expect the longitudinal sample in ASHE to have an under-representation of middle-aged employees, due to the uneven effects of sample attrition.

None of the estimated coefficients in Table 3 are particularly large, and so the magnitude of these attrition biases may be small. However, their impact on estimates generated from the ASHE longitudinal sample (e.g. estimates of wage progression) is difficult to tell *a priori*. For this reason, we proceed in the Section 7 to discuss methods of removing any such biases for estimates based on the longitudinal ASHE sample.

Before doing so, it may be noted that the amount of variance explained in Table 3 is low. APS does include additional characteristics that are associated with employment exit, but which are not observed in ASHE (e.g. family circumstances). However, extending Eq. 1 to include these characteristics does not help in identifying possible attrition biases in ASHE, since characteristics observed in the APS alone cannot contribute to the estimation of *p̂(emp_exit)* for ASHE sample members. Similarly, one could extend Eq. 2 to include additional characteristics observed only in ASHE. But this has limited value as one cannot determine whether or not the estimated coefficients on these characteristics are net of any association with *p̂(emp_exit)*; in other words, the coefficients may not necessarily indicate attrition biases. Nevertheless, extended versions of Eq 1 and Eq 2 are presented in Appendix B for information. The additional variance explained by these models is relatively modest.

7. Constructing two-period weights to address attrition bias

Method

We use the analysis presented in Section 6 to construct longitudinal weights that address the attrition biases estimated to be present in samples of adjacent years in ASHE. We restrict our attention to adjacent years because our benchmark dataset (APS) does not provide data on the correlates of employment exit over a follow-up period of more than one year.

To construct the weights, we first run the probit regression of employment exit in the Longitudinal APS (Y_{APS}) (Eq. 1) and use the coefficients to generate the predicted probability of employment exit - *p̂(emp_exit)* - for each individual in ASHE. We then run an equivalent probit regression of sample exit in ASHE (Y_{ASHE}) and use the coefficients to generate the predicted probability of sample exit - *p̂(samp_exit)* - for each individual in ASHE. We run both regressions within year, to allow for time-variance in the correlates of employment exit and sample exit.

An 'attrition adjustment factor' adj_{it} is then constructed for each ASHE individual *i* in year *t* who survives in the sample to year *t+1*:

$$adj_{it} = \frac{1}{(1 - \hat{p}(samp\_exit))} * (1 - \hat{p}(emp\_exit))$$

The first term boosts the representation in the two-period sample of those most likely to exit ASHE between year *t* and *t+1* (essentially restoring the profile of the two-period sample to that of the full sample in year *t*). The second term then calibrates this adjustment to account for each individual's probability of exiting employment exit between year *t* and *t+1* (essentially reducing the size of the adjustment factor in proportion to the individual's likelihood of exiting employment).[17]

A final two-period longitudinal weight $watt_{it}$ is then constructed for each ASHE individual *i* in year *t* who survives in the sample to year *t+1* as:

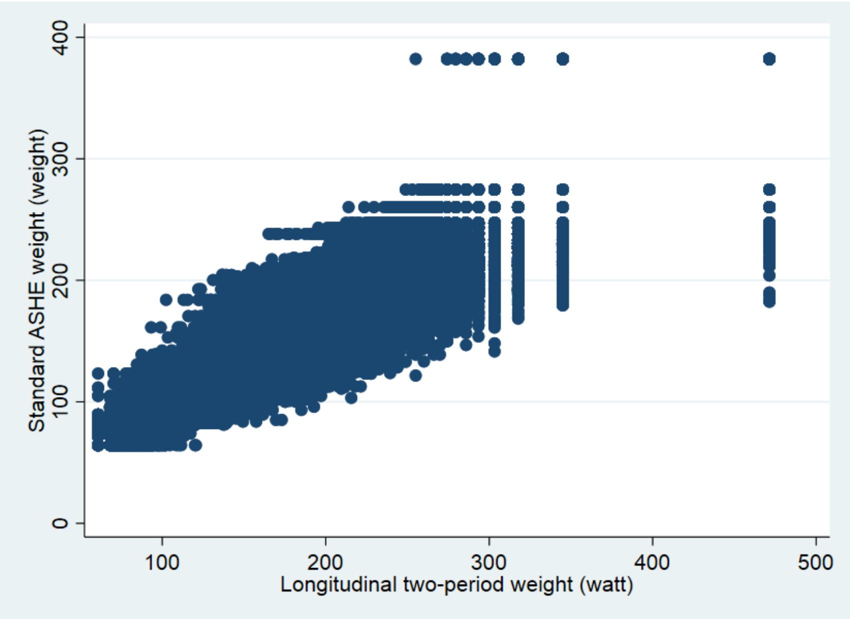
$$watt_{it} = weight_{it} * adj_{it}$$

where $weight_{it}$ is the original ASHE cross-sectional weight for individual *i* in year *t*.[18] The new weight is scaled so that $\sum watt_i = \sum weight_i$ within each year.

Results

The method described above generates a weight with the same mean and similar distribution to the original ASHE weight. A comparison of the two is provided in Figure 7.

Figure 7: Comparison of original ASHE weight and longitudinal two-period weight, 2012-2017



Note: sample members who appear in year *t* and *t+1*

The main effects of *watt* on the composition of the two-period sample when compared with *weight* are to induce small shifts in favour of younger workers, those with low tenure and those from the private sector (see Table 4).

To illustrate the impact on employment outcomes estimated from the ASHE longitudinal sample, Figure 8 shows the distribution of annual growth in individuals' wages under each weighting scheme. The impact here is minor, but application of the longitudinal weight does bring about a small widening of the distribution, raising the 75th percentile by around half a percentage point.

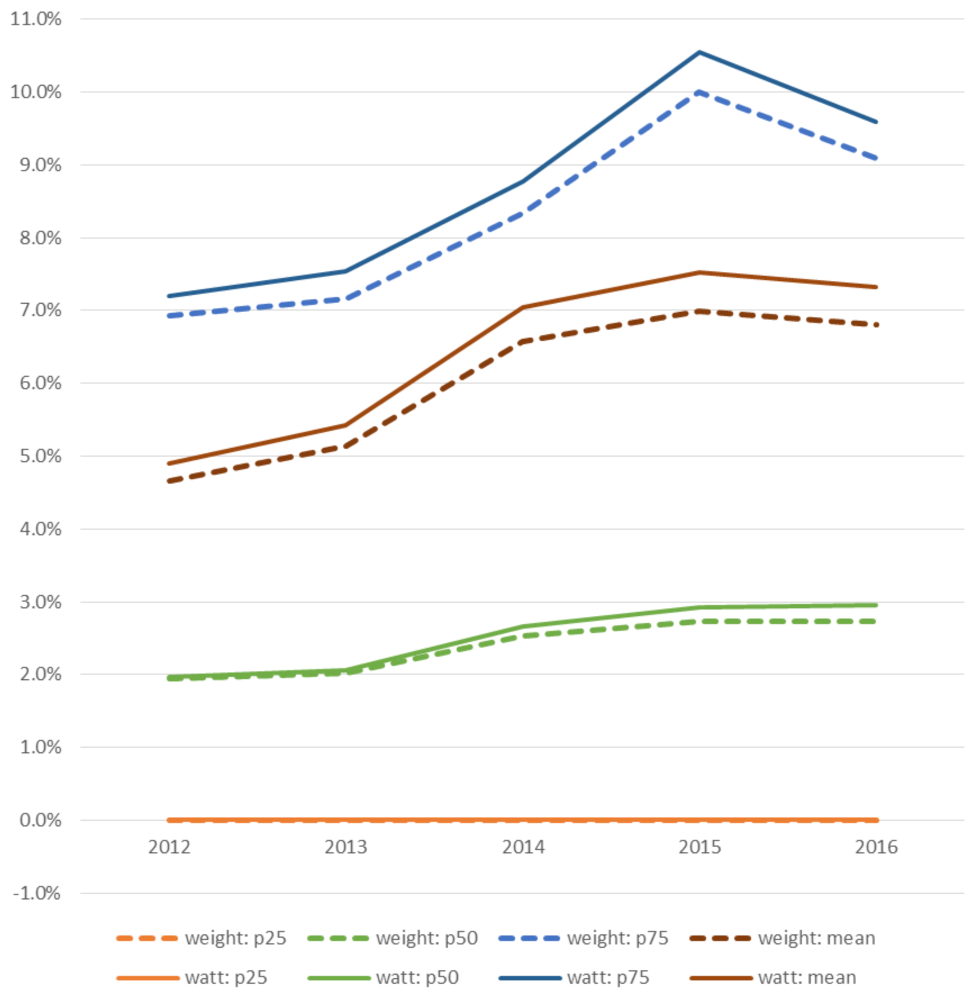
[17] The formula may also be reconfigured as $\hat{p}(emp\_stay)/\hat{p}(samp\_stay)$, i.e. the ratio of expected 'stayers' to observed 'stayers'. This helps to indicate that the largest adjustments will be for those cases where the probability of staying in the observed sample is much lower than the true rate of remaining in PAYE employment.

[18] See footnote 15 in respect of the ASHE cross-sectional weight.

Table 4: The profile of the ASHE two-period sample before and after longitudinal weighting adjustment (cell proportions)

Sample	**Observed in Year t and t+1**	**Observed in Year t and t+1**	**Difference: Col (2) – Col (1)**
Weights	Cross-sectional	Longitudinal	
	(1)	(2)	(3)
Gender of the employee:			
Female	49.8%	49.3%	-0.5%
Male	50.2%	50.7%	0.5%
Age of the employee:			
16-19	2.9%	3.2%	0.4%
20-24	7.8%	8.8%	1.0%
25-29	11.1%	11.9%	0.7%
30-34	11.7%	12.1%	0.4%
35-39	11.4%	11.3%	0.0%
40-44	12.6%	12.3%	-0.3%
45-49	13.8%	13.3%	-0.6%
50-54	12.5%	11.8%	-0.7%
55-59	9.3%	8.8%	-0.5%
60-64	4.8%	4.6%	-0.3%
65+	2.1%	1.9%	-0.1%
Job tenure:			
<1 year	14.0%	16.1%	2.0%
1-2 years	11.6%	12.3%	0.8%
2-5 years	21.2%	21.4%	0.3%
5-9 years	21.5%	20.7%	-0.9%
10-20 years	20.1%	18.7%	-1.4%
20+ years	9.8%	8.9%	-0.9%
Missing/ invalid	1.8%	2.0%	0.1%
Sector of ownership:			
Private	73.9%	75.6%	1.7%
Public	26.1%	24.4%	-1.7%

Figure 8: Distribution of annual growth in individuals' nominal gross hourly wages, 2012-2017, under different weighting regimes



8. Limitations and possible extensions

Extending the weighting adjustment to pairs of non-adjacent years

As noted earlier, the approach outlined above is limited by being applicable only to samples of individuals observed in adjacent years of data (*t* and *t+1*). This means that the weights cannot be used for analyses that seek to examine wage progression over 2, 3, 4 or more years.

In order to derive an equivalent set of longitudinal weights that accounts for sample attrition in samples of individuals who are observed more than one year apart, we would require a dataset where we observe *p̂(emp_exit)* over such periods. In other words, we would require a dataset where we can observe the probability that a sample member observed in year *t* is out of employment in year *t+i* (where *i=2,…,n*). These data do not exist, as far as we are aware.[19]

[19] The UK Household Longitudinal Study (Understanding Society) follows sample members over multiple years but, as we understand it, is not necessarily cross-sectionally representative beyond year 1 (2009). The LLFS boost to the Annual Population Survey notionally follows sample members over four years but, as far as we are aware, ONS do not provide weighted datasets for research use.

Developing a weighting adjustment for pairs of non-adjacent years will, however, be possible once we gain access to PAYE data from HMRC's Real-Time Information system (expected in the second half of 2022). These data will enable employment transitions over periods of more than one year to be identified.

Addressing attrition bias in estimates of employment exit using ASHE

A further limitation of the weights described in Section 6 is that they are currently derived only for sample members who remain in the *t1* sample at *t2*. So whilst this set of weights can be used to remove attrition bias from an analysis focused on people who remain in employment in both years (e.g. a study of wage progression), we have not yet developed weights to make the set of sample leavers representative of those employees who exit employment between t1 and t2. Such a weighting scheme would be needed if one were to use ASHE to describe the characteristics of people leaving employment in a given year. It is possible to construct such weights: one would use information from the APS on the characteristics of those exiting employee status along with information on the characteristics of sample leavers in ASHE, following an equivalent methodology to that outlined in Section 7. However, this remains for future work.

This would not, in itself, correct the classification errors that may compromise studies which seek to use ASHE to identify the correlates of employment exit (e.g. studies seeking to determine whether minimum wage employees are more or less likely to exit employment than higher-wage employees). Any sample members who exit ASHE through longitudinal sample attrition, rather than via true employment exit, are currently misclassified in such analyses.[20] Such classification errors can attenuate model coefficients, as well as potentially introducing bias to those coefficients if sample attrition is non-random (as we have shown it to be).

One does not currently know with certainty who is subject to these classification errors (i.e. who has exited the sample through attrition rather than true employment exit). We do have an estimate, which is: $\hat{p}(sample\_exit) - \hat{p}(emp\_exit)$. In cases where this value is relatively high, there is a high probability that the dependent variable in an employment exit equation based on ASHE is measured with error; in cases where this value is low, the dependent variable is more likely to be accurate. The inverse of this term might then be used to weight any analysis of employment exit in ASHE; this would have the effect of downgrading the influence on model coefficients of individuals with measurement errors in the dependent variable and upgrading the influence of individuals without measurement errors, thereby reducing the bias in any estimated coefficients. Such a weight could then be used to test the sensitivity of the estimated coefficients to attrition bias. However, the classification biases described here can also be addressed directly once we gain access to PAYE data from HMRC's Real-Time Information system, since these data will enable us to identify the true employment status of sample members in any given year. So again, this remains for future work.

[20] Specifically, the "employment exit" indicator for individuals subject to sample attrition will be coded TRUE (1) when it should be coded FALSE (0).

9. Summary and conclusions

Many government and academic studies have utilised the panel dimension of ASHE to conduct longitudinal analysis. However, the longitudinal integrity of the ASHE sample has been the subject of very little prior investigation, with the longitudinal sample generally assumed free of any attrition biases that might compromise analysis of outcomes measured over multiple years of data.

We explore the validity of these assumptions by comparing rates of year-on-year sample exit in ASHE with rates of employment exit estimated from a reference dataset (the Longitudinal APS). Our analysis confirms the existence of systematic patterns of longitudinal sample attrition in ASHE that have the potential to introduce bias into longitudinal analyses of these data.

The impact of these differential rates of attrition on the composition of the longitudinal sample are not dramatic. However, it is difficult to tell *a priori* whether they would be sufficient to alter any conclusions drawn from analysis of a dataset that made no efforts to account for them. We generate a longitudinal two-period weight that adjusts each two-period sample from 2012/13-2017/18 for the attrition biases that we estimate to be present, and use this new weight in an illustrative analysis of annual growth in individuals' wages. Here, we find that the application of the longitudinal weights brings about a small widening of the distribution.

These new, longitudinal two-period weights will be available through the ONS Secure Research Service in due course, as part of the WED project outputs, so that analysts can make use of them in their own analyses, or at least check the sensitivity of their results to the use of these weights.

There are a number of limitations to these weights.

One limitation is that the weights are derived only for sample members who remain in the sample, and so they cannot be used to correct analyses that use exit from ASHE as a proxy for exit from PAYE employment. Weights for sample leavers are feasible, however, and deriving such weights remains for future work.

A second limitation is that our analysis and weighting adjustments currently only cover the period 2012-2017. This is due to the absence of Longitudinal APS data outside of these years (at least at the time of working). We expect that datasets for later years will become available, at which point the weighting series can be extended.[21]

A third limitation is that the weights are derived only for individuals observed in adjacent years of data (*t* and *t+1*). This means that the weights cannot be used for analyses that seek to examine wage progression over 2, 3, 4 or more years. A fourth limitation is that we are unable to identify with certainty which sample members truly exit employment and which exit the sample due to longitudinal attrition. Addressing these two limitations will be possible, however, once we data on ASHE sample members from HMRC's Real-Time Information System (scheduled for the second half of 2022).

[21] Longitudinal five-quarter Labour Force Survey datasets are now available through to 2020, although they offer smaller samples than the APS.

References

Bird, D. (2004) “Methodology for the 2004 Annual Survey of Hours and Earnings”, Technical Report, *Labour Market Trends*, Office for National Statistics, pp. 457-464.

Clark P and Tate P (1999) *Methodological issues in the production and analysis of longitudinal data from the Labour Force Survey* (GSS Methodology Series No. 17)

Dickens R, Riley R and Wilkinson D (2015) “A re-examination of the impact of the National Minimum Wage on employment”, *Economica*, 82, 328: 841-864.

DWP (2013) *Response to Freedom of Information Request (DWP reference: VTR 4543)*. Available at: https://www.whatdotheyknow.com/request/178190/response/439137/attach/3/FOI%204543.pdf

Elsby M, Shin D and Solon G (2016) “Wage Adjustment in the Great Recession and Other Downturns: Evidence from the United States and Great Britain,” *Journal of Labor Economics*, 34(S1), 249-291.

NISRA (2019) Northern Ireland Annual Survey of Hours and Earnings, October 2019. Available at: https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NI-ASHE-Bulletin-2019.pdf

ONS (2002) *Review of Distribution of Earnings Statistics*. London: Office for National Statistics.

ONS (2007) *Changes to ASHE in 2007*, Office for National Statistics. Available at: https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningsashemethodologyandguidance/changeinashe07tcm77254755.pdf

ONS (2008) *Plans for the ONS Statistical Work Programme 2007-08*, London: Office for National Statistics.

ONS (2010) *Survey Control Reviews: Annual Survey of Hours and Earnings*. London: Office for National Statistics. Available at: https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningsashemethodologyandguidance/ashe2011tcm77248262.pdf

ONS (2011) *Summary Quality Report for the Annual Survey of Hours and Earnings (ASHE)*. Information Paper, London: Office for National Statistics.

ONS (2012) *Annual Population Survey (APS) QMI*. London: Office for National Statistics. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi [Last retrieved 9th April 2021]

ONS (2013) *Coverage and non-response errors in the Annual Survey of Hours and Earnings*. Information Paper, London: Office for National Statistics.

ONS (2017) *An overview of and comparison between Annual Survey of Hours and Earnings (ASHE) and Average Weekly Earnings (AWE): 2017*, London: Office for National Statistics.

ONS (2018) *Annual Survey of Hours and Earnings, Low pay and Annual Survey of Hours and Earnings pension results QMI*, Office for National Statistics. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningslowpayandannualsurveyofhoursandearningspensionresultsqmi

ONS (2019) "Analysis of job-changers and job-stayers", *Economic Review: April 2019*, London: Office for National Statistics. https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/compendium/economicreview/april2019/analysisofjobchangersandstayers

ONS (2021a) *Labour Force Survey Performance and Quality Monitoring Report: January to March 2021*, London: Office for National Statistics. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

ONS (2021b) *EMP13: Employment by Industry*. London: Office for National Statistics. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employmentbyindustryemp13

Pont, M. (2007) "Coverage and non-response errors in the UK New Earnings Survey", *Journal of the Royal Statistical Society Series A*, 170, Part 3, pp. 713–733.

Schaefer D and Singleton C (2019) "Cyclical labour costs", *European Economic Review*, Volume 120, Article 103317.

Stokes L, Bryson A and Forth J (2017) "Who fared better? The fortunes of performance-pay and fixed-pay workers through recession", *British Journal of Industrial Relations*, 55, 4: 778-801.

Stokes L et al (2022) "Cross-sectional representativeness of ASHE", WED Project, forthcoming at: www.wagedynamics.com

Tate P (1999) "Using longitudinally linked data from the British Labour Force Survey", *Survey Methodology*, 25: 99-104.

Appendix A: ASHE sample design and survey administration

The target population for ASHE is all employee jobs in the UK, including all industries and occupations.[22] All jobs held by employees with a National Insurance (NI) number ending in a particular 2 digits are selected from the PAYE register held by HMRC; if a sampled person holds multiple jobs on the register, all jobs held by that person are selected. This sample is therefore considered to be a one per cent, simple random sample of employee jobs (we discuss this further below). The survey itself is completed by employers, who are contacted by ONS after their contact details have been obtained by matching the sample drawn from the PAYE register to the ONS' Inter Departmental Business Register (IDBR).

ASHE is conducted in April each year (with surveys typically dispatched to employers in the second week of April).[23] The sample is first selected in January. A second extract is then taken in April, with the aim of identifying those individuals who may have become employees, or changed jobs, since the first extract was selected in January. The addition of this second extract represents one of the changes implemented with the introduction of ASHE, recognising that the former New Earnings Survey (NES) had been missing this group of employees who changed employer between sample selection and the survey reference date (Bird, 2004).[24]

The use of the PAYE register as the sampling frame means that some employee jobs will not be available for selection, as they are not present on the register. This can happen when an employer has no employees in their business earning above the threshold requiring the employer to register for PAYE (employers are legally required to operate PAYE if the earnings of any of their employees reach the NI Lower Earnings Limit, standing at £120 per week in 2020/21).[25,26] The ONS information paper on coverage and non-response errors in ASHE (ONS, 2013) describes the two sub-categories of employers who may not have a registered PAYE scheme:

- those where the employer is registered for VAT with HMRC, but not for PAYE (termed "VAT-only companies")
- those where the business is not registered for VAT or PAYE.

In 2004 and 2005, ONS did collect earnings data from VAT-only companies, but this was discontinued from 2006, as patterns were found to be similar to those of employees in companies who had registered PAYE schemes.[27]

---

[22] ASHE covers the whole of the UK, but fieldwork for Great Britain and Northern Ireland are conducted separately, with ONS carrying out the survey for Great Britain and the Northern Ireland Statistics and Research Agency (NISRA) doing so for Northern Ireland. The data available for use within this paper relates to Great Britain only.

[23]https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualsurveyofhoursandearningsashe

[24] ASHE replaced the New Earnings Survey (NES) from 2004, following the Review of Distribution of Earnings Statistics (ONS, 2002). A further change to the coverage of the survey was the introduction of supplementary surveys for VAT-only companies, although these were discontinued from 2006, as discussed below. Other changes introduced at this point, included weighting, questionnaire re-design and the introduction of imputation for item non-response. Our focus in this paper is on the period from 2004 onwards, and so we do not discuss the details of these differences between NES and ASHE here; further information can be found in Bird (2004).

[25] Other factors can require employers to register for PAYE too, see: https://www.gov.uk/paye-for-employers

[26] https://www.gov.uk/government/publications/rates-and-allowances-national-insurance-contributions/rates-and-allowances-national-insurance-contributions

[27] The other hypothetical scenario in which employee jobs could be missing from the PAYE register would be if an employer chooses not to include employees below the PAYE earnings threshold on its payroll. This practice has been documented in

The consequence of the above is that some employees with low earnings (whether due to lower levels of pay and/or working fewer hours) will not be present on the sampling frame. This affects the coverage of the survey even in cross-sectional analysis. In terms of following individuals over time, this can be problematic if an employee "disappears" from the sample where they have, for example, moved to a job with an employer who has not had to register for PAYE, as we cannot tell whether they have changed job, or if they have exited employment.

We noted above that ASHE is considered to be based on a one per cent sample of employee jobs. Based on the 2011 ASHE quality report (ONS, 2011), the issued sample for ASHE usually stands at around 260,000 employee jobs. But fewer employee jobs form the final achieved sample, typically standing at around 180,000 but varying year-on-year; thus below one per cent of employee jobs.

While completion of the survey is mandatory under the Statistics of Trade Act 1947, inevitably not all employers respond. Analysis of 2004 data reported in Pont (2007), found that "good data" were collected for 68% of the issued sample (noting that a substantive proportion of other questionnaires returned related to individuals exempt from the survey, and that some questionnaires were not useable as a result of insufficient quality). An ONS review of ASHE in 2010 indicated that the anticipated yield for ASHE (based on the latest survey at the time) stood at 63% of employee jobs, or 55% if considered in terms of individual respondents (ONS, 2010). Our own calculations estimating the number of responses to ASHE as a percentage of ONS estimated jobs in March/April of each year suggest response may have been falling in more recent years, standing at around 60% between 2016 and 2019 (Table 1).[28]

the past, when it was noted to be concentrated in the hotel and catering industry (ONS, 2013). However, our understanding is that employers registered for PAYE are now required to include all employees, irrespective of their earnings.

[28] Note that response rates for the NI ASHE stand at around 90 per cent (NISRA, 2019).

Table A1: ASHE responses as % of ONS estimated employee jobs and employees, authors' calculations

Year	% employee jobs	% employees (approximate)
1997	0.62%	
1998	0.64%	
1999	0.63%	
2000	0.61%	
2001	0.61%	
2002	0.61%	
2003	0.62%	
2004	0.61%	0.67%
2005	0.61%	0.68%
2006	0.61%	0.68%
2007	0.50%	0.56%
2008	0.50%	0.55%
2009	0.62%	0.67%
2010	0.64%	0.67%
2011	0.67%	0.70%
2012	0.64%	0.68%
2013	0.65%	0.68%
2014	0.65%	0.70%
2015	0.63%	0.68%
2016	0.60%	0.66%
2017	0.59%	0.66%
2018	0.59%	0.66%
2019	0.58%	0.65%

Source: ASHE and ONS Workforce jobs series

Some reasons for variation in response have been documented; for example, those employers with special arrangements (for electronic completion) are more likely to respond to the survey, which led to employers with such arrangements being treated as a separate stratum for the purpose of weighting (ONS, 2007). ONS (2013) also report that where employers do not respond to the survey; they typically do not respond for any of their employees. Thus most variation in response seems likely to be across employer, rather than within employer, although in some instances, employers are found to respond for a subset of their selected employees.

In practice, there are limits to the time and resources available to pursue employers to return questionnaires. Pont (2007) reports on the results of two intensive follow-up exercises run in 2003 and 2004, which did yield additional responses, but also demonstrated that even with this additional chasing, there remains a “hardcore” of employers who do not respond. There are, to our knowledge, no published figures on the number of employers prosecuted for not responding to ASHE. Information on completion of ONS business surveys more generally indicates that the ONS Enforcement Unit deals with thousands of cases of non-completion per year, but that few of these reach court or result in prosecution.[29] ONS note that their aim is to encourage and assist employers to comply, rather than penalise, where possible.

The design of the sample for ASHE, by virtue of the fact that it samples individuals based on NI numbers ending in a particular 2 digits, means that the sample has a longitudinal element; these individuals are in scope for the survey every year providing they are in employment at the time. This longitudinal element is potentially a valuable feature of the survey (see existing uses cited earlier). However, while the ASHE sample forms a panel by design, there appears to be little in the administration of the survey to preserve the integrity of the panel, by either investigating or addressing non-random attrition.

Users of ASHE also need to bear in mind that a number of changes to ASHE have been implemented over time, including changes to questionnaire design and the approach to cross-sectional weighting[30]; a complete list can be found in ONS (2017). Of particular relevance for this paper however, is that for the 2007 and 2008 surveys, the sample was cut by 20 per cent. This clearly has consequences when trying to follow individuals over time in this period. This cut to the sample was focused on particular industries, targeting those which had shown the least variation in earnings (ONS, 2008). The sample was restored to its original size for the 2009 survey onwards.

[29] https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/businesssurveys

[30] Including the implementation of the 2010 Standard Occupational Classification in the 2011 survey, which affected the calculation of weights this makes use of information on occupation.

Appendix B: What other characteristics are associated with employment/sample exit?

In this Appendix, we seek to identify the extent to which the analyses presented in Section 6 are limited by the range of common covariates in ASHE and LAPS. We do this by adding to the LAPS and ASHE models some further covariates that are found in only one of the two datasets. In ASHE, we add the loss of pay marker. In LAPS, we add a range of individual and household characteristics (e.g. marital status, home owner, degree-level qualification, temporary contract, and children).

The results are shown in Table B1. In line with expectations, employment exit in LAPS is less common among home owners than renters, more common among those on temporary employment contracts than those on permanent contracts, and more common among those with very young children than among those with no children. In ASHE, sample exit is more common among those who have experienced a loss of pay in Year t. These regressions do not indicate whether patterns of exit would differ between the surveys along any of these dimensions .

Table B1: Marginal effects – Probability of employment exit in LAPS and ASHE (2012-2017)

VARIABLES	LAPS	ASHE
	(1)	(2)
Gender (Ref. Female):		
Male	0.000	0.014***
	(0.002)	(0.001)
Age group (Ref: 16-19)		
20-24	-0.022***	0.016***
	(0.005)	(0.003)
25-29	-0.036***	0.001
	(0.005)	(0.003)
30-34	-0.041***	-0.009***
	(0.005)	(0.003)
35-39	-0.037***	-0.022***
	(0.005)	(0.003)
40-44	-0.041***	-0.037***
	(0.005)	(0.003)
45-49	-0.040***	-0.047***
	(0.005)	(0.003)
50-54	-0.025***	-0.043***
	(0.005)	(0.003)
55-59	0.006	-0.014***
	(0.005)	(0.003)
60-64	0.063***	0.053***
	(0.005)	(0.003)
65+	0.077***	0.067***
	(0.005)	(0.003)
Occupation (ref: Managers, directors and senior officials)		
Science, research, engineering and tech	-0.014***	-0.020***
	(0.003)	(0.002)
Associate professional and technical	-0.015***	0.001
	(0.003)	(0.002)
Administrative and secretarial	-0.027***	-0.037***
	(0.004)	(0.002)
Skilled trades occupations	-0.009**	0.001
	(0.004)	(0.002)

Caring, leisure and other service occupation	-0.014***	-0.018***
	(0.004)	(0.002)
Sales and customer service occupations	-0.021***	-0.052***
	(0.004)	(0.002)
Process, plant and machine operatives	-0.016***	-0.020***
	(0.004)	(0.002)
Elementary occupations	-0.017***	-0.041***
	(0.004)	(0.002)
Industry (ref: Sections A-E)		
F: Construction	0.030***	0.053***
	(0.004)	(0.003)
G: Wholesale, retail, repair of vehicles	-0.008**	-0.012***
	(0.003)	(0.002)
H; Transport, and storage	-0.003	-0.002
	(0.004)	(0.002)
I; Accommodation, and food service	0.002	0.070***
	(0.004)	(0.002)
J; Information, and communication	0.012**	0.046***
	(0.005)	(0.003)
K: Financial and insurance activities	-0.002	0.031***
	(0.005)	(0.003)
L: Real estate activities	-0.002	0.026***
	(0.007)	(0.004)
M: Professional, scientific, and technical	0.006	0.038***
	(0.004)	(0.002)
N: Admin and support services	0.000	0.082***
	(0.005)	(0.002)
O: Public admin and defence	0.003	-0.009***
	(0.004)	(0.003)
P: Education	0.004	0.000
	(0.004)	(0.002)
Q: Health, and social work	-0.001	0.005**
	(0.004)	(0.002)
R: Art, entertainment, and recreation	0.004	0.020***
	(0.006)	(0.003)
S: Other service activities	0.014**	0.018***
	(0.006)	(0.003)
Region of workplace (Ref: North East)		
North West	0.002	0.012***
	(0.004)	(0.002)
Yorkshire and Humberside	-0.002	0.001
	(0.004)	(0.003)
East Midlands	-0.003	0.008***
	(0.004)	(0.003)
West Midlands	-0.003	0.016***
	(0.004)	(0.003)
South West	-0.007**	0.002
	(0.004)	(0.003)
East of England	-0.002	0.011***
	(0.004)	(0.003)
London	0.014***	0.066***
	(0.004)	(0.002)
South East	0.001	0.023***

	(0.004)	(0.002)
Wales	0.003	-0.003
	(0.004)	(0.003)
Scotland	-0.002	-0.008***
	(0.004)	(0.003)
Sector of ownership (Ref. Private):		
Public	-0.014***	-0.028***
	(0.003)	(0.002)
Decile of real gross hourly pay (Ref: the lowest)		
Pay missing	-0.004	
	(0.003)	
2nd paydecile	-0.013***	-0.033***
	(0.004)	(0.002)
3rd paydecile	-0.018***	-0.043***
	(0.004)	(0.002)
4th paydecile	-0.027***	-0.050***
	(0.004)	(0.002)
5th paydecile	-0.029***	-0.058***
	(0.004)	(0.002)
6th paydecile	-0.029***	-0.070***
	(0.004)	(0.002)
7th paydecile	-0.038***	-0.073***
	(0.004)	(0.002)
8th paydecile	-0.032***	-0.071***
	(0.004)	(0.002)
9th paydecile	-0.023***	-0.066***
	(0.004)	(0.002)
10th paydecile	-0.011***	-0.042***
	(0.004)	(0.002)
Basic working hours (Ref: <=15)		
16-29	-0.022***	-0.058***
	(0.003)	(0.002)
30-47	-0.054***	-0.093***
	(0.003)	(0.001)
48 plus	-0.031***	-0.058***
	(0.004)	(0.003)
Tenure (Ref: <1 year)		
1-2 years	-0.018***	-0.048***
	(0.003)	(0.002)
2-5 years	-0.028***	-0.080***
	(0.003)	(0.001)
5-9 years	-0.038***	-0.115***
	(0.003)	(0.001)
10-20 years	-0.043***	-0.133***
	(0.003)	(0.002)
20 years or more	-0.027***	-0.126***
	(0.003)	(0.002)
Missing/invalid	-0.009	-0.044***
	(0.010)	(0.003)
Workplace size (Ref: 1-24 employees)		
25-49	-0.017***	-0.016***
	(0.003)	(0.002)
50-499	-0.020***	-0.017***

	(0.002)	(0.001)
500+	-0.027***	-0.030***
	(0.003)	(0.002)
Missing	-0.012*	0.012***
	(0.007)	(0.001)
Year (Ref: 2012)		
2013	-0.003	-0.001
	(0.003)	(0.002)
2014	-0.005*	0.028***
	(0.003)	(0.001)
2015	-0.000	0.041***
	(0.003)	(0.001)
2016	0.002	0.045***
	(0.003)	(0.002)
2017	0.001	0.049***
	(0.003)	(0.002)
Additional covariates specific to each survey:		
Married	-0.001	
	(0.002)	
Home owner	-0.010***	
	(0.002)	
Education degree	-0.002	
	(0.002)	
Temporary contract	0.060***	
	(0.003)	
Any children under 2 years old	0.028***	
	(0.003)	
Any children aged 2-4 years old	0.004	
	(0.003)	
Any children aged 5-9 years old	0.002	
	(0.002)	
Any children aged 10-15 years old	-0.003	
	(0.002)	
Pay loss marker		0.058***
		(0.002)
Obs.	237,890	1,054,135
Pseudo R2	*0.0856*	*0.0537*
Pseudo R2 in the absence of additional covariates	*0.0769*	*0.0529*

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1