

Wage & Employment Dynamics

THE WED PROJECT



METHODOLOGY PAPER

Weighting for employer non-response in ASHE

Abstract

The Annual Survey of Hours and Earnings (ASHE) is based on a 1% sample of employee jobs and provides many of the UK's official earnings statistics. Weights are provided with the core dataset, which adjust the profile of the annual achieved sample such that it is representative of employees by gender, age, occupation and region. However, while the ASHE is based on a sample of employee jobs, the survey is completed by employers. Not all employers respond, and some do not respond quickly enough for their returns to be incorporated in the annual dataset. In this methodological note, we explore the characteristics of employers responding to ASHE. We show that certain types of organisations are over or under-represented in the achieved sample, even after applying the weights from the core dataset. We construct an adjustment to those existing weights that attempts to remove these biases. We discuss the implications for some headline estimates of earnings.

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

Contents

1. Introduction	3
2. Sampling, response and weighting in ASHE.....	3
Sampling.....	3
Response.....	4
Weighting.....	6
3. Which employers respond to ASHE?	6
4. Adjusting the existing ASHE cross-sectional weights.....	14
5. Implications of adjusted weights	17
6. Discussion and summary.....	22
Appendices.....	24

Acknowledgements

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.

The authors are grateful to ADR UK (Administrative Data Research UK) and the Economic and Social Research Council for funding this work (Grant No. ES/T013877/1).

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 the earnings of employees in the UK. As well as forming the basis for a range of ONS labour market statistics, ASHE is widely used to inform policymaking and evaluation across a range of UK government bodies in areas such as the National Minimum/Living Wage, pension provision and public sector pay (ONS, 2018).

The ASHE is based on a 1% sample of employee jobs, taken from Her Majesty's Revenue and Customs (HMRC) Pay As You Earn (PAYE) records. Employees are selected into the sample by virtue of the last two digits of their National Insurance number (NINOs). The survey itself is completed by their employer, who receives the form from ONS and is statutorily required to complete and return it.¹

The ONS makes cross-section weights available for each annual edition of the ASHE dataset. These adjust the profile of the achieved sample so that it is representative of the population of employee jobs in Britain in terms of gender, age, occupation and region. The nature of that weighting scheme is discussed further in Section 2. However, we show that certain types of organisation remain under or over-represented in the achieved ASHE sample, even after applying those weights. This has implications for the representativeness of the weighted sample, and so has the potential to bias descriptive statistics and model estimates derived from the survey, including UK National Statistics.

In this paper we look first at patterns of survey response. We then go on to assess whether the existing weighting scheme is dealing with any response biases which seem to be apparent in the profile of the achieved sample. Having shown that 'residual' response biases remain after applying the existing cross-section weights, we develop alternative cross-section weights which seek to remove these biases. We then go on to consider the implications for some statistics generated from the ASHE dataset.²

We develop and demonstrate alternative cross-section weights for the period 2009 to 2018 inclusive. We focus on this period to avoid prior years in which there were cuts to the sample size, and to have consistent variable classifications over time. In this paper, we focus our analysis of the implications of these new weights primarily on one year, 2018.

2. Sampling, response and weighting in ASHE

Sampling

The target population for ASHE is all employee jobs in the UK, including all industries and occupations. As noted above, the sample for the survey is drawn 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 an employee holds multiple jobs that are PAYE-registered, all jobs are

¹ ASHE covers the whole of the UK, but fieldwork for Great Britain and Northern Ireland is conducted separately. All of the analysis in this paper and the resulting weights relate to Great Britain only, as we do not have access to ASHE datasets for Northern Ireland.

² Response bias is not the only possible source of error in survey statistics. Measurement errors can also cause a survey estimate to depart from the true value of that statistic in the population. Weights which reduce the influence of response biases may also reduce the impact of measurement errors if they reduce the influence of sampled units where variables are more likely to be measured with error; however, the reverse also applies. The relationship between response biases and measurement errors is not something we consider in this paper.

selected. This is considered a one per cent, simple random sample of employee jobs. The survey itself is completed by employers and, in order to obtain their contact details, the sample drawn from the PAYE register is matched against the ONS' Inter Departmental Business Register (IDBR).

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).

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.

ASHE is conducted in April each year, with surveys typically dispatched to employers in the second week of April. 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).

Response

The issued sample for ASHE usually stands at around 260,000 employee jobs (ONS, 2011). But fewer employee jobs form the final achieved sample, typically standing at around 180,000; 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 (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). As discussed later in this note, the sample size was cut in 2007, before being restored in 2009, hence the lower figures in Table 1 for 2007 and 2008.

Table 1: 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%

Note: ONS estimates of jobs in March each year are obtained from the ONS Workforce Jobs summary series: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/workforcejobssummaryjobs01/current>

Some variations in response at the employer level have been documented elsewhere; 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 the 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. The stated aim of ONS is to encourage and assist employers to comply, rather than

penalise, where possible.³ Reminders to employers to complete the ASHE follow a set timetable each year, with 3 reminders sent to employers who respond via special arrangements (in June and July), and one reminder sent to all other employers (in June).

Weighting

Weights are provided with the core ASHE dataset. The variable *calwght* is the “standard” ASHE weight, enabling users to produce estimates that are representative of the population of UK employees.⁴

The ASHE data are weighted to UK population totals from the UK Labour Force Survey (LFS) based on classes defined by occupation, region, age and sex. There are two stages in constructing these weights, as described in ONS (2018).

Firstly, individual cases are assigned a design weight, on the basis of four strata, according to whether they are part of: (i) the original questionnaire despatch; (ii) one of the later supplementary surveys (part of the group who move jobs between sample selection and questionnaire despatch; (iii) the group who join the workforce after the original sample selection; or (iv) the group with special arrangements for response (where employers respond electronically). Cases in these groups are assigned different design weights because it is known that response rates vary across these strata.

The second stage involves post-stratifying responses to population estimates taken from the LFS across 108 post-strata, based on the cross- classification of:

- occupation (nine groups) – major groups from the 2010 Standard Occupational Classification
- age bands (three groups) – 16 to 21 years, 22 to 49 years and 50 years and over
- sex (two groups) – male and female
- region (two groups) – London and South East, and the rest of the UK

The resulting weight enables estimates to be produced from the ASHE data that are representative of the UK population in terms of gender, age, occupation and region.

3. Which employers respond to ASHE?

To explore which employers respond to ASHE, for each year of our analysis period, we match those employers that appear in the ASHE dataset to the Business Structure Database (BSD), using the enterprise reference (ENTREF). The BSD is an annual snapshot of the Inter-Departmental Business Register (IDBR) – a comprehensive list of UK business maintained by the ONS and used by government for statistical purposes. Employment information on the IDBR is updated periodically from administrative sources (HMRC PAYE and VAT records) and ONS surveys such as the annual Business Register and Employment Survey. The BSD represents our best available source of information on the **population** of employee jobs, which also contains information on the characteristics of the employers providing those jobs. The share of ASHE records that can be linked to the employing enterprise in the BSD is high across all years (98 per cent or more of ENTREFs in

³ <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/businesssurveys>

⁴ In addition, a low pay weight (*lpcalwght*) is also made available on the ASHE dataset, which reweights the dataset after excluding employees whose pay is affected by absence during the survey reference period. We focus solely on the standard weight within this paper.

ASHE match to a record in the BSD; the gap appears to be due to the differencing in timing of ASHE and the BSD snapshot). Match rates are presented in Appendix Table 1.

After matching the two datasets by year, we identify which types of employers are more or less likely to respond to ASHE by creating a binary variable indicating whether each enterprise in the BSD appears in ASHE or not. We then compare this with the expected probability that an enterprise is selected to respond to ASHE. As ASHE is based on a 1% sample of employee jobs, the probability that an employer will be selected to respond to ASHE increases with the number of employees in its workforce. For each enterprise in the BSD, we calculate the expected probability that it is selected to appear in ASHE, based on the number of employees at the enterprise, as $1 - (0.99^{N_j})$, where N_j is the number of employees in enterprise j (Upward, 2007).⁵

We can then compare the ratio of actual response to the expected probability for any characteristic measured on the BSD. As an example, Table 2 shows the pattern of response by employer size (total number of employees) for 2018. The first column of Table 2 shows the expected probability of selection, whilst the second column shows actual outcomes. For example, enterprises with 5,000 or more employees would be almost certain of being selected into the issued sample for ASHE if it were a 1% random sample of employee jobs, but only 89 per cent of such enterprises appear in ASHE in 2018. The final column of Table 2 shows the ratio of actual response to expectations, which increases considerably with employer size.

⁵ As noted earlier, ASHE is a sample of employee jobs, rather than employees. Some employees have multiple jobs, and where these are jobs with different organisations, these will appear separately on the BSD. The BSD will however undercount employee jobs held by someone who has multiple jobs with the same employer. This should not have a substantive effect on our analysis as the share of all employee jobs that are multiple jobs is relatively small (5% in 2018); and of these, around 11% hold more than one job with the same employer- equivalent to around 0.5% of all employee jobs in 2018. These calculations are made based on the 91% of multiple jobs where both a person identifier and enterprise identifier were available.

Table 2: Actual versus expected response by employer size, 2018

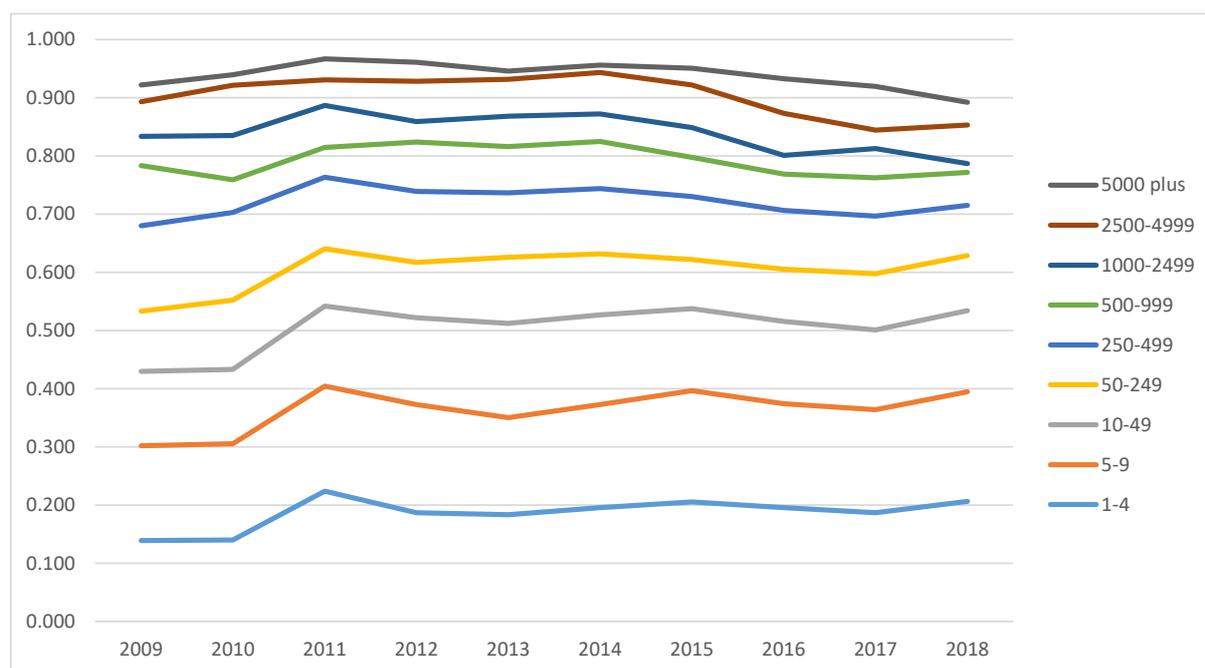
	(1)	(2)	(3)
	Expected probability of selection	Proportion of firms in the size band that appear in ASHE	Ratio: outcome/expected probability (2/1)
Employer size (total employees):			
1-4	0.016	0.003	0.206
5-9	0.064	0.025	0.395
10-49	0.174	0.093	0.534
50-249	0.591	0.372	0.629
250-499	0.961	0.688	0.715
500-999	0.998	0.770	0.772
1000-2499	1.000	0.787	0.787
2500-4999	1.000	0.853	0.853
5000 plus	1.000	0.892	0.892

Base: all enterprises with at least 1 employee as recorded in the BSD, located in England, Scotland or Wales. Probabilities are rounded to three decimal places and do not reach absolute certainty for any group.

Figure 1 plots the ratio of actual versus expected response, by employer size, for each year from 2009 to 2018. We can observe three main trends. Over the period 2009 to 2011, there is an improved yield from all sizes of firm, but this is particularly marked among firms with fewer than 2,500 employees. We then see broad stability over the period 2011 to 2014, and then a declining yield among firms with 500 or more employees over the period 2014 to 2018.⁶

⁶ We do not have information as to the reasons behind these trends: ONS does not publish response rates or the detail of how fieldwork processes may vary from year to year.

Figure 1: Actual versus expected response, 2009-2018



Base: all enterprises with at least 1 employee as recorded in the BSD, located in England, Scotland or Wales.

Table 3 presents, for 2018, the results of estimating a linear probability model in which the binary indicator of response (whether a BSD employer responds to ASHE or not) is regressed on the expected probability that they respond to ASHE, along with a set of employer characteristics: employer size, region, legal status, industry and firm age.

If employer non-response was random, we would not expect any of these employer characteristics to show statistically significant associations with our dependent variable, after controlling for the expected probability of response. In practice, we do observe patterns of non-response; the largest coefficients are observed for employer size (with the largest firms around 22 percentage points more likely to respond than the smallest firms) and legal status (with organisations in central government around 8 percentage points more likely to respond than private limited companies).⁷ Note that this analysis is intended as illustrative of the characteristics associated with response, rather than determining the actual adjustment ultimately used in deriving a set of alternative weights (discussed in Section 4).

⁷ We also checked the sensitivity of our regression results to additionally including a variable for the number of live local units; while this also showed a statistically significant relationship with response, we have some concerns over the accuracy of the variable and thus omit this from the main results presented here (the inclusion of this variable had a minimal impact on the r-squared, increasing it from 0.237 to 0.240 in 2018).

Table 3: Probability that a BSD employer appears in ASHE, OLS regression results, 2018

		Coefficient	Std. error
Expected probability		0.642***	(0.002)
Employer size (ref: 1-4 employees)	5-9	-0.011***	(0.000)
	10-49	-0.017***	(0.000)
	50-249	-0.013***	(0.002)
	250-499	0.061***	(0.003)
	500-999	0.120***	(0.003)
	1000-2499	0.134***	(0.004)
	2500-4999	0.186***	(0.006)
	5000 plus	0.216***	(0.005)
Region (ref: London)	NE	0.001***	(0.000)
	NW	0.000	(0.000)
	Yorks	0.001***	(0.000)
	EM	0.001***	(0.000)
	WM	0.000	(0.000)
	SW	0.001**	(0.000)
	East	-0.003***	(0.000)
	SE	0.002***	(0.000)
	Wales	0.000	(0.000)
	Scotland	0.000	(0.000)
Legal status (ref: private company)	Sole proprietor	0.000	(0.000)
	Partnership	0.000	(0.000)
	Public corporation	0.074***	(0.008)
	Central government body	0.079***	(0.002)
	Local authority	0.048***	(0.003)
	Non-profit making body	0.007***	(0.000)
Industry (SIC07) (ref: manufacturing)	Agriculture, forestry, fishing	-0.017***	(0.001)
	Mining and quarrying	-0.009***	(0.003)
	Electricity, gas	-0.012***	(0.002)
	Water supply	-0.001	(0.001)
	Construction	-0.006***	(0.000)
	Wholesale, retail	-0.005***	(0.000)
	Transport and storage	-0.005***	(0.000)
	Accommodation & food	-0.014***	(0.000)
	Information & communication	-0.005***	(0.000)
	Financial and insurance	-0.004***	(0.001)
	Real estate	-0.010***	(0.001)
	Professional, scientific, technical	-0.005***	(0.000)
	Admin and support	-0.008***	(0.000)

	Public administration	-0.038***	(0.003)
	Education	0.002***	(0.001)
	Health and social work	0.001**	(0.000)
	Art, entertainment, recreation	-0.010***	(0.001)
	Other service activities	-0.006***	(0.001)
	Activities of households	0.000	(0.001)
Firm age (ref: 30 years or more)	1 year or less	-0.017***	(0.000)
	2-4 years	-0.016***	(0.000)
	5-9 years	-0.016***	(0.000)
	10-19 years	-0.013***	(0.000)
	20-29 years	-0.006***	(0.000)
Constant		0.013***	(0.000)
<i>R-squared</i>		0.237	
<i>N observations(enterprises)</i>		2,952,764	

Note: standard errors in parentheses. Statistical significance indicated as ***0.01 **0.05 *0.10

Base: all enterprises with at least 1 employee as recorded in the BSD, located in England, Scotland or Wales.

In Table 4, we go on to explore how the profile of employment in ASHE differs from that observed in the population of enterprises (as captured by the BSD). The value of this approach is that we can take account of the existing ASHE weights. For employer size, legal status and age, we show the percentage of employee jobs in each category as observed in the BSD (column 1), and in ASHE after applying the standard ASHE weights (column 2).⁸ The third column shows the difference between the ASHE weighted estimates and the BSD estimates.

⁸ Strictly speaking figures for the BSD show the percentage of employees rather than employee jobs, but as noted earlier, the percentage of employees holding multiple jobs with the same employer is very small. Unweighted estimates for ASHE are also presented in Appendix Table 2.

Table 4: Employer characteristics associated with response to ASHE, 2018

	(1)	(2)	(3)	(4)
	BSD (employee- weighted), per cent	ASHE weighted, per cent	ASHE minus BSD (2)- (1)	Squared error
				Mean Squared Error (MSE): 9.5
Number of employees (banded)				
1-4	11.6	4.3	-7.3	53.0
5-9	6.1	4.2	-1.8	3.3
10-49	15.0	14.3	-0.7	0.5
50-249	13.9	14.9	1.0	1.1
250-499	5.6	6.3	0.7	0.4
500-999	5.7	6.4	0.7	0.5
1000-2499	7.2	8.1	0.9	0.8
2500-4999	7.3	9.0	1.7	3.0
5000 plus	27.7	32.5	4.8	22.7
				MSE: 10.2
Legal status				
Company	72.1	65.3	-6.8	46.1
Sole proprietor	2.7	1.6	-1.0	1.1
Partnership	2.4	2.1	-0.3	0.1
Public corporation	0.6	0.6	0.0	0.0
Central government body	8.8	12.4	3.6	12.7
Local authority	6.6	9.6	3.0	9.0
Non-profit making body	6.8	8.5	1.6	2.7
				MSE: 20.6
Age of firm				
1 year or less	3.3	1.1	-2.2	4.8
2-4 years	8.7	3.8	-4.9	23.8
5-9 years	10.2	7.0	-3.2	10.2
10-19 years	13.7	11.7	-2.0	4.1
20-29 years	24.5	29.3	4.7	22.4
30 years plus	39.6	47.2	7.6	57.6
<i>N observations</i>	<i>2,952,764</i>	<i>178,264</i>		

Note: Figures for BSD are employee-weighted estimates, for those enterprises that are recorded as having at least one employee in the BSD. ASHE estimates exclude those employee jobs for whom it was not possible to match to an enterprise record in BSD (this affected 1 per cent of the original ASHE sample in 2018).

If employer response varied in a way that was already addressed by the existing ASHE weights, we would expect that the profile of employee jobs would look similar both in ASHE and BSD in Table 4. However, we can see that this is not the case. In the BSD, we can see that around 12 per cent of

employee jobs are in enterprises that have between 1 and 4 employees, while in ASHE, after weighting, employee jobs in enterprises of this size account for around 4 per cent. Conversely, employee jobs in larger enterprises are over-represented in ASHE; 33 per cent of employee jobs are in enterprises with 500 or more employees, compared with 28 per cent in the BSD. Jobs in larger firms are thus over-represented in the ASHE sample, even after applying the standard ASHE weights. Looking by sector, employee jobs in private sector enterprises (especially companies) are under-represented while employee jobs in public sector enterprises are over-represented, even after weighting. Employee jobs in older firms are over-represented in ASHE, while employee jobs in younger firms are under-represented.

The fourth column of Table 4 presents the squared error for each category of each variable, with the mean squared error (MSE) for each variable presented in bold at the top of each column. Here we are assuming that the BSD gives the true population profile, and we can use the squared error to tell us the magnitude of the “error” (or bias) in ASHE. The mean squared error allows us to compare the extent of bias across variables; with a smaller MSE indicating a smaller bias or error. The MSE has a value of around 10 for both workforce size and legal status, but is around double (20.6) for firm age. In interpreting these findings, it’s important to bear in mind that the figures in Table 4 are bivariate associations. Firm age is correlated with size and legal status, and hence the particularly pronounced differences by age seen in Table 4 are likely to be partly accounted for by size and legal status in the regression presented above.

We repeat this for additional employer characteristics (region and industry) in Appendix Table 2. While there are some differences between the BSD and ASHE in terms of the distribution of jobs across region and industry (jobs in Education and Health and Social Work are over-represented in ASHE compared to the profile observed in the BSD, for example), the magnitude of the bias is smaller than for those variables presented in Table 4.

In summary, response to ASHE appears to vary between different types of employer in ways that are not accounted for by the existing ASHE weights, which seek to make ASHE representative of the population of employee jobs in terms of gender, age, occupation and region, but which do not explicitly take into account the characteristics of employers.

4. Adjusting the existing ASHE cross-sectional weights

The analysis presented so far has shown that a number of employer characteristics are associated with response to ASHE in ways that are not accounted for by the existing ASHE weights.

To address this, we construct an adjustment to the standard ASHE weight (*calwght*), which aims to take account of these residual employer-related response biases. We construct the weighting adjustment by applying the ASHE weights and then undertaking a raking procedure to compute an adjustment factor which brings the weighted ASHE sample closer into line with the BSD profile along a small number of employer characteristics. We use the `-svycal, rake-` command within Stata, with control totals obtained from the BSD, following the approach set out by Valliant and Dever (2018: 59).

It is necessary to be relatively parsimonious in the variables we use for this purpose. A greater number of variables (and categories of variables) will reduce any sample bias, but is also likely to introduce additional variability into the weighting scheme, which will in turn impair the precision of the sample. Our aim is to find a weighting scheme which minimises sample bias without unduly inflating standard errors.

We focus on three variables in our adjustment: employer size (number of employees), legal status and firm age. Our choice is informed by the analysis presented in Table 4 (and Appendix Table 2), which shows those dimensions for which there is greatest deviation from the BSD profile.

In addition to being parsimonious in the number of variables entering the weighting scheme, we also trim the weights following the approach recommended in Valliant and Dever (2018: 157), that is capping the maximum and minimum weight values as equal to the median value of the weight, plus or minus three times the value of the interquartile range.

We experimented with a number of alternative configurations of the key variables entering the weighting scheme, using different numbers of categories. Our preferred weight is presented in Table 5 (*wedwght*). The table includes the Kish design effect (*deff*) (Kish, 1965) for the adjusted weight, alongside that for the standard ASHE weight (*calwght*), to provide a measure of the variability of the weights. We also present the mean squared error for each employer characteristic, calculated based on the differences in profiles between ASHE and BSD across size, legal status, firm age, region and industry, after applying the adjusted weights to ASHE.⁹ The adjusted weight represents an improvement in MSE on all dimensions considered here, when compared against the original ASHE weighting. Other variants of the weighting scheme which did not perform as well as our preferred weight are presented in Appendix Table 3 for comparison.

The new weight is calculated from the sample for which a match was possible between ASHE and BSD. In order that we can work with the full ASHE sample, after applying the trimming, and rescaling the weights to have the same mean value as the standard ASHE weight, we use the value of the original weight variable (*calwght*) where it was not possible to construct an alternative weight (this affects less than one per cent of the ASHE sample).

⁹ Profiles by these employer characteristics for our preferred adjusted weight are presented in Appendix Table 4. The MSE is representing the difference between the BSD distribution and the re-weighted ASHE distribution.

Table 5: Design effects and mean squared errors, original and adjusted ASHE weight, 2018

	Original ASHE weight	Adjusted ASHE weight
	<i>calwght</i>	<i>wedwght</i>
<i>Size</i>		1-4; 5-49; 50-4999; 5000+
<i>Status</i>		Private; public and non-profit
<i>Firm age</i>		<10 years, 10-29 years, 30 +
MSE - size	9.5	1.4
MSE - status	10.2	0.4
MSE - age	20.6	2.5
MSE - region	1.6	0.9
MSE - industry	2.0	0.7
MSE (all)	8.8	1.2
Kish Deff	1.060	1.152

N=178,264 observations (unweighted).

In Appendix Table 4, we present the weighted profile of the ASHE sample by the same employer characteristics, using our adjusted weight (*wedwght*). Profiles by employer characteristics for ASHE are necessarily presented only for the sample which it was possible to match to BSD.

In developing the alternative weights, we have built on the original weighting variable, *calwght*, in a way that aims to remove substantive biases in terms of firm characteristics, rather than starting from scratch and incorporating both employee and employer characteristics within the raking scheme. To reassure ourselves that our adjusted weight does not inadvertently skew the weighted profile based on employee characteristics (especially for those characteristics targeted under the original weighting scheme), Table 6 compares the weighted profile under the original weight and our alternative weight for key employee characteristics. The unweighted profile is also shown for completeness. Columns 4 and 5 show the difference between the two weighted distributions and the mean squared error as a measure of the deviation between the two weighting schemes. Overall, we see that the weighted profiles are broadly similar under the two weighting schemes.

Table 6: Profile of ASHE sample by employee characteristics, original and alternative ASHE weights, 2018

	(1)	(2)	(3)	(4)	(5)
	Unweighted, per cent	ASHE weighted (calwght), per cent	ASHE adjusted weight (wedwght), per cent	Difference (3)-(2)	Squared error
Gender					MSE: 0.29
Female	52.71	49.79	49.25	-0.54	0.29
Male	47.29	50.21	50.75	0.54	0.29
Full-time					MSE: 3.17
Part-time	32.33	29.53	31.31	1.78	3.17
Full-time	67.67	70.47	68.69	-1.78	3.17
Age group					MSE: 0.19
16-19	3.65	3.66	4.05	0.39	0.15
20-24	8.76	8.57	9.04	0.47	0.22
25-29	11.74	11.82	12.1	0.28	0.08
30-34	11.76	12.1	12.15	0.05	0.00
35-39	11.16	11.58	11.5	-0.08	0.01
40-44	10.38	10.78	10.6	-0.18	0.03
45-49	11.84	12.11	11.79	-0.32	0.10
50-54	11.96	11.6	11.27	-0.33	0.11
55-59	9.96	9.57	9.32	-0.25	0.06
60-64	5.99	5.64	5.53	-0.11	0.01
65 plus	2.78	2.58	2.65	0.07	0.00
Occupation					MSE: 2.55
Managers, directors and senior official	7.68	9.92	10.22	0.3	0.09
Professional	17	21.05	18.81	-2.24	5.02
Associate professional and technical	11.93	14.44	13.67	-0.77	0.59
Administrative and secretarial	14.49	11.55	12.26	0.71	0.50
Skilled trades	6.28	7.3	7.94	0.64	0.41
Caring, leisure and other service	10.89	9.51	9.58	0.07	0.00
Sales and customer service	10.54	8.63	9.02	0.39	0.15
Process, plant and machine operatives	6.24	5.91	6.11	0.2	0.04
Elementary	14.93	11.69	12.39	0.7	0.49
<i>N observations (unweighted)</i>	<i>180,185</i>	<i>180,185</i>	<i>180,185</i>		

5. Implications of adjusted weights

We have seen from the above that the profile of employment according to employer characteristics is closer to that observed in the BSD once the new weights are applied. In this section we go on to explore the implications for estimates of the distribution and moments of gross hourly pay.

Table 7 shows mean hourly pay, along with standard errors, based on the standard ASHE weight and our adjusted weight.¹⁰ We also present the design effect for this particular variable. Our adjusted ASHE weight (*wedwght*) reduces the estimate of mean hourly pay by 2.5% per cent from £16.51 to £16.10 (a statistically significant difference) while the standard error of that estimate remains similar (increasing slightly from 3.91 to 3.92). The mean wage falls under the adjusted weighting scheme because this weighting scheme boosts the representation of small, young, private sector firms, which tend to pay below-average wages. The three percentiles shown in Table 7 (p25, p50 and p75) are also lower under the adjusted weighting scheme. However, they do not fall by equivalent degrees. In particular, the median has fallen by around 4 percentage points, whilst p25 has fallen by only 2.5 percentage points. This implies some compression in the lower-middle part of the wage distribution. One consequence is that the percentage of employees judged to be low paid (earning below two-thirds median hourly earnings: a point below p25 under either weighting scheme) falls by around 12% (or 2.1pp) from 17.8 per cent to 15.7 per cent; again, a statistically significant difference.

Table 7: Estimates of mean hourly earnings (pence), alternative ASHE weights, 2018, employees on adult rates, earnings not affected by absence

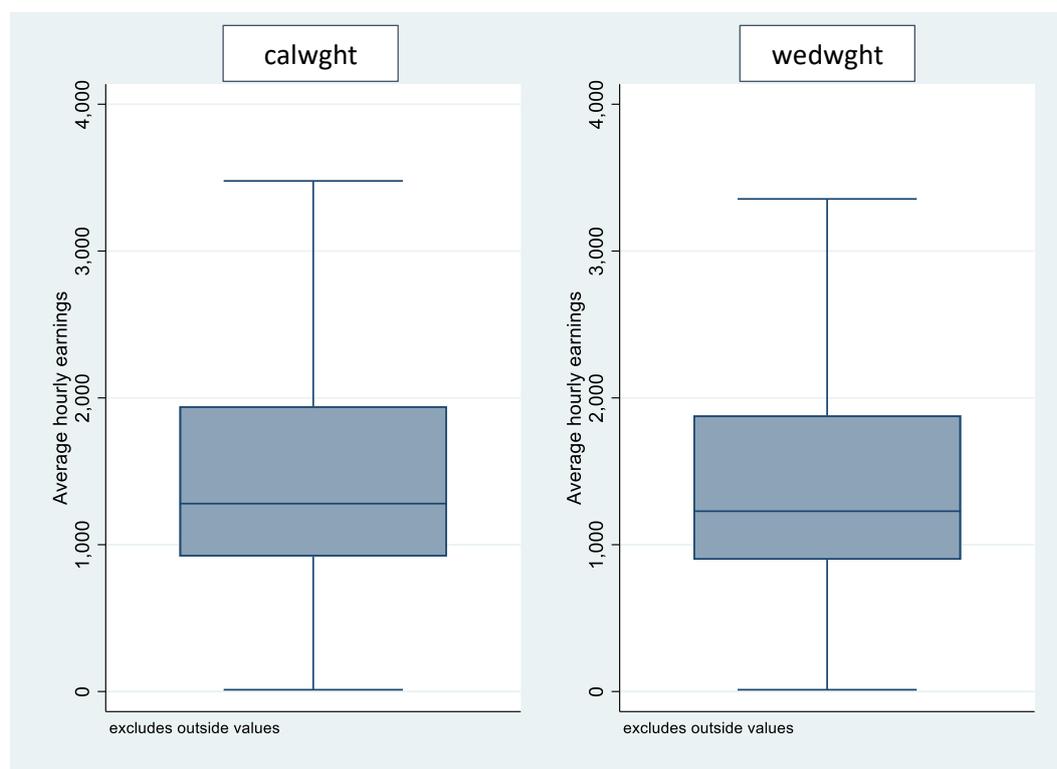
	Mean	Std. Err.	DEFF	Median	p25	p75	IQ range	2/3 median	% low paid
calwght	1651	3.91	1.22	1280	920	1943	1023	853	17.8
wedwght	1610	3.92	1.26	1228	897	1880	983	819	15.7

N=168,742 observations (unweighted). Low paid is defined as below two-thirds of the median wage.

Figure 2 presents box plots for hourly pay, with the left hand chart showing this using the standard ASHE weight, and the right hand chart showing this using our adjusted weight, *wedwght*.

¹⁰ We use the variable “he”, which is average hourly earnings for the reference period (gross pay divided by total paid hours worked). Our estimates are based on the sample used for publication of ASHE estimates (employees on adult rates of pay, with earnings not affected by absence), but results look very similar when using the full sample.

Figure 2: Distribution of hourly pay, 2018, box plots, standard ASHE weight and adjusted weight (wedwght), employees on adult rates, earnings not affected by absence



Note: The boundaries of the box show the 25th and 75th percentiles; the line within the box shows the median; the boundaries of the upper (lower) whisker shows the value that is 1.5 times the inter-quartile range above the 75th percentile (below the 25th percentile). Outside values are excluded.

Figures 3a and 3b shows the impact on average hourly earnings over our analysis period of 2009 to 2018, for both mean and median earnings respectively. We can see that while the adjusted weight brings down the estimates of average earnings in all years, it has not affected the trajectory of average earnings, which remains similar under both sets of weights. It does however, have more of an impact on the percentage of workers classified as low paid (defined here as below two-thirds of median earnings), with this percentage falling more steeply over time, particularly from around 2014 onwards, when applying the adjusted weight (Figure 4).

Figure 3a: Mean hourly pay, 2009-2018, standard ASHE weight and adjusted weight, employees on adult rates, earnings not affected by absence

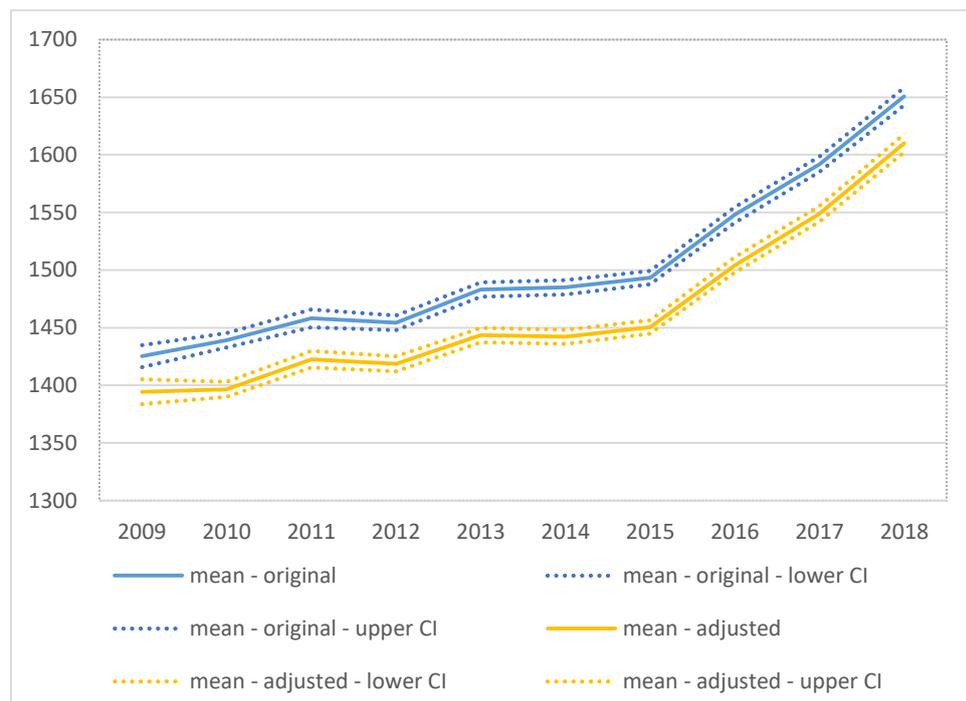


Figure 3b: Median hourly pay, 2009-2018, standard ASHE weight and adjusted weight, employees on adult rates, earnings not affected by absence

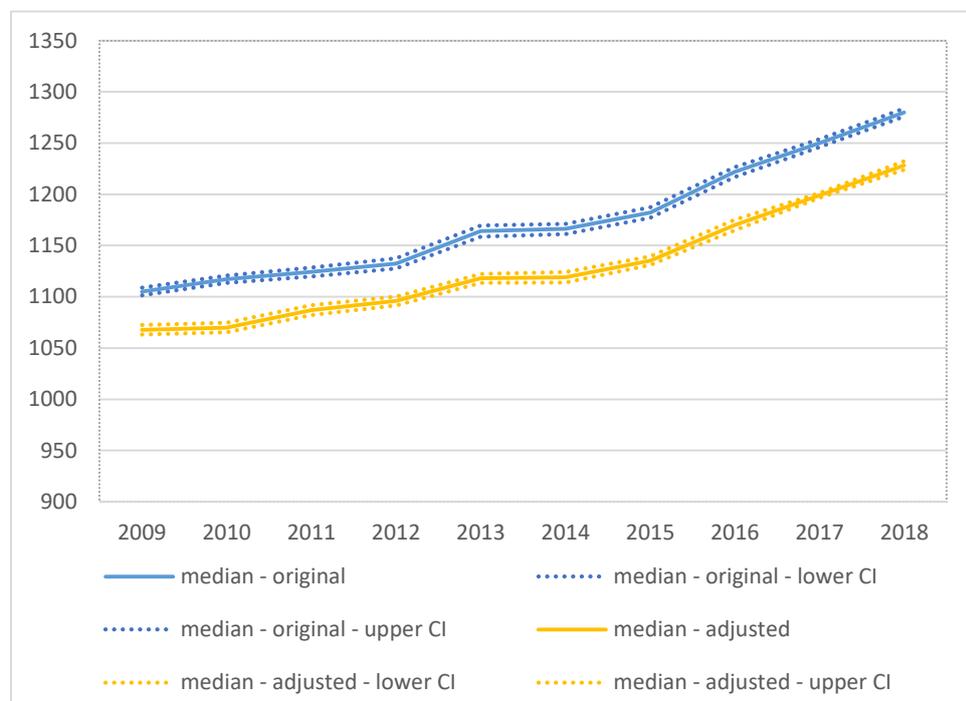
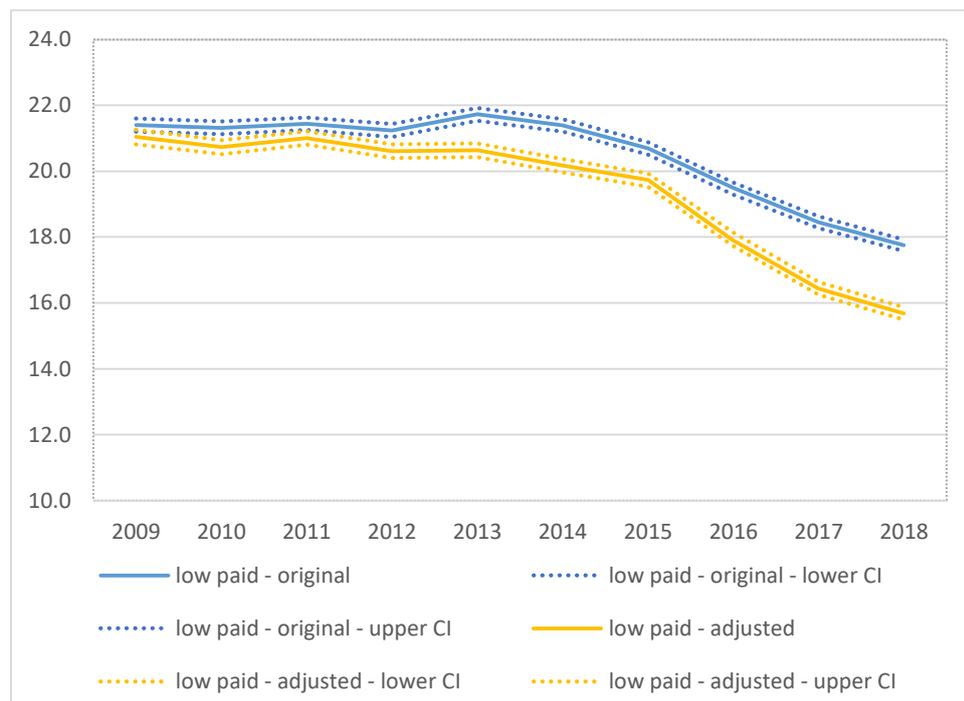


Figure 4: Per cent low paid, 2009-2018 standard ASHE weight and adjusted weight, employees on adult rates, earnings not affected by absence



Note: “Low paid” defined here as below two-thirds of median earnings

We also explore the impact of the alternative weights on a simple regression of log hourly earnings on a set of conditioning variables, namely gender, age, occupation, full-time/part-time status, and employer characteristics (employer size, industry, firm age, region, legal status).¹¹ To test whether the alternative weight induces changes to the results of this regression, we run a standard wage equation under each weight in turn.

We test the equality of each coefficient in the model separately. Table 8 presents results for those variables where we observe statistically significant differences in coefficients when applying the alternative set of weights; full models results are presented in Appendix Table 5 (which also includes results from an unweighted regression for comparison). The first column presents results from a regression using the standard ASHE weight, and the second column presents results from using the adjusted weight. Here we can see that there are statistically significant differences in the coefficients for three of the occupational categories, managers, professional and skilled trades occupations (as indicated by the final column of Table 8). Some statistically significant differences were also observed in terms of employer size.

¹¹ Those observations above the 99.5th percentile and below the 0.5th percentile for hourly earnings are excluded from the models.

6. Discussion and summary

The existing ASHE datasets contain a set of weights which adjust the profile of the achieved sample so that it is representative of the population of employee jobs in terms of gender, age, occupation and region. The analysis presented in this paper has shown that certain types of organisation remain under or over-represented in the achieved ASHE sample for Great Britain, even after applying those weights. For example, we observe that employees in private sector enterprises (especially companies) are under-represented in ASHE, while those in public sector organisations are over-represented. Having shown that such ‘residual’ response biases remain after applying the existing cross-section weights, we develop an alternative set of cross-section weights which seek to reduce these biases.

We consider the implications of these alternative weights for some key statistics generated from the ASHE dataset. The alternative weight brings down estimates of average earnings (hourly pay) in all years considered (2009 to 2018), but it does not affect the trajectory of average earnings, which remains similar under both the alternative and original weights. We see more of an impact on the percentage of workers classified as low paid, with this percentage falling more steeply, particularly from around 2014, under the alternative weight. When applied within a regression framework, exploring the determinants of hourly earnings, we see that, while the broad associations between hourly earnings and employee and employer characteristics remain the same under both sets of weights, there are some statistically significant differences in the coefficients associated with employer size, and for some occupational groups.

These alternative cross-sectional weights (constructed for years 2009 to 2018 inclusive) will be made 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. The code to derive the weights will also be made available to users.

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.
- Hausman (1978) Specification tests in Econometrics. *Econometrica* 46(6): 1251-71.
- Kish L (1965) *Survey Sampling*, New York: Wiley
- 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/changeinash e07tcm77254755.pdf>
- ONS (2010) Survey Control Reviews: Annual Survey of Hours and Earnings. Available at: <https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningsashemethodologyandguidance/ash e2011tcm77248262.pdf>
- ONS (2011) *Summary Quality Report for the Annual Survey of Hours and Earnings (ASHE)*. Information Paper, Office for National Statistics.
- ONS (2013) *Coverage and non-response errors in the Annual Survey of Hours and Earnings*. Information Paper, 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/annualsurveyofhoursandearningslowpayandannualsurveyofhoursandearningspensionr esultsqmi>
- 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.
- Upward, R. (2007) *Linking the ASHE and the BSD*, University of Nottingham, mimeo.
- Valliant. R. and Dever, J.A. (2018) *Survey Weights: A Step-by-Step Guide to Calculation*, Stata Press

Appendices

Appendix Table 1: Match rates, ASHE entrefs to BSD

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Matched	48559	50309	52844	51460	52,136	55,612	58,698	57,289	57,551	62,615
Unmatched	358	332	955	808	749	903	894	867	847	993
Total	48917	50641	53799	52268	52885	56515	59592	58156	58398	63608
Matched %	99.3	99.3	98.2	98.5	98.6	98.4	98.5	98.5	98.5	98.4

Base: all enterprises (entrefs) in ASHE

Appendix Table 2: BSD and ASHE population profiles, selected employer characteristics: 2018

	(1)	(2)	(3)	(4)	(5)
	BSD (employee-weighted), per cent	ASHE unweighted, per cent	ASHE weighted, per cent	ASHE minus BSD (3)-(1)	Squared error
Number of employees (banded)					MSE: 9.5
1-4	11.6	4.4	4.3	-7.3	53.0
5-9	6.1	4.3	4.2	-1.8	3.3
10-49	15.0	14.3	14.3	-0.7	0.5
50-249	13.9	14.8	14.9	1.0	1.1
250-499	5.6	6.2	6.3	0.7	0.4
500-999	5.7	6.4	6.4	0.7	0.5
1000-2499	7.2	8.1	8.1	0.9	0.8
2500-4999	7.3	8.9	9.0	1.7	3.0
5000 plus	27.7	32.7	32.5	4.8	22.7
Legal status					MSE: 10.2
Company	72.1	65.9	65.3	-6.8	46.1
Sole proprietor	2.7	1.8	1.6	-1.0	1.1
Partnership	2.4	2.2	2.1	-0.3	0.1
Public corporation	0.6	0.5	0.6	0.0	0.0
Central government body	8.8	12.0	12.4	3.6	12.7
Local authority	6.6	9.4	9.6	3.0	9.0
Non-profit making body	6.8	8.2	8.5	1.6	2.7
Age of firm					MSE: 20.6

1 year or less	3.3	1.0	1.1	-2.2	4.8
2-4 years	8.7	3.9	3.8	-4.9	23.8
5-9 years	10.2	7.0	7.0	-3.2	10.2
10-19 years	13.7	11.6	11.7	-2.0	4.1
20-29 years	24.5	29.2	29.3	4.7	22.4
30 years plus	39.6	47.3	47.2	7.6	57.6
Region					MSE: 1.6
NE	3.3	3.5	3.4	0.1	0.0
NW	9.9	10.1	9.8	-0.1	0.0
Yorks	7.9	8.9	8.6	0.7	0.4
EM	7.2	7.8	7.4	0.2	0.0
WM	8.6	8.8	8.5	-0.1	0.0
SW	10.2	10.8	10.5	0.3	0.1
East	21.0	17.2	18.3	-2.7	7.3
London	14.1	13.5	14.1	0.1	0.0
SE	7.3	7.7	7.6	0.3	0.1
Wales	3.4	3.8	3.7	0.2	0.1
Scot	7.2	8.1	8.1	0.9	0.8
SIC 2007 Section level					MSE: 2.0
Agriculture, forestry, and fishing	0.9	0.6	0.5	-0.4	0.1
Mining and quarrying	0.2	0.2	0.2	-0.1	0.0
Manufacturing	7.9	9.3	9.4	1.5	2.1
Electricity, gas, air cond. supply	0.7	0.6	0.7	0.0	0.0
Water supply, sewerage, waste	0.7	0.7	0.6	-0.1	0.0

Construction	4.9	3.6	3.8	-1.1	1.2
Wholesale, retail, repair of vehicles	15.3	16.3	15.3	0.0	0.0
Transport and storage	4.4	4.4	4.1	-0.3	0.1
Accommodation and food service	7.9	6.4	6.0	-2.0	4.0
Information and communication	4.3	3.2	3.8	-0.5	0.2
Financial and insurance activities	3.5	3.3	3.6	0.1	0.0
Real estate activities	1.8	1.3	1.4	-0.5	0.2
Professional, scientific, and technical	8.1	6.5	7.1	-1.0	0.9
Admin and support services	9.3	6.5	6.0	-3.3	10.9
Public admin and defence	3.3	4.5	4.7	1.4	2.0
Education	10.2	13.6	13.9	3.8	14.3
Health and social work	12.1	15.1	14.8	2.7	7.3
Art, entertainment, and recreation	2.4	2.2	2.2	-0.2	0.0
Other service activities	2.0	1.8	1.8	-0.2	0.0
Activities of households as employers	0.2	0.1	0.1	-0.1	0.0
N observations (unweighted)	2,952,764	178,264	178,264		

Appendix Table 3: Design effects and mean squared errors, original and alternative ASHE weights, 2018

	Original ASHE weight	Alternative adjusted weights				
	calwght	wedwght	nrakewt1	nrakewt2	nrakewt3	nrakewt5
<i>Variables used in weighting</i>						
<i>Size</i>		1-4; 5-49; 50-4999; 5000+	1-49; 50-999; 1000+	1-4; 5-49; 50-999; 1000-4999; 5000+	1-4; 5-49; 50-4999; 5000+	1-4; 5-49; 50-4999; 5000+
<i>Status</i>		Private; public and non-profit	Private; public and non-profit	Private; public and non-profit	Private; public and non-profit	Company; sole proprietor/partnerships; public and non-profit
<i>Firm age</i>		<10 years, 10-29 years, 30 +				
MSE - size	9.5	1.4	6.4	1.2	1.1	1.3
MSE - status	10.2	0.4	0.1	0.3	0.3	0.1
MSE - age	20.6	2.5	9.4	9.3	9.3	9.4
MSE - region	1.6	0.9	1.0	0.9	0.9	0.9
MSE - industry	2.0	0.7	0.9	0.8	0.8	0.8
MSE (all)	8.8	1.2	3.5	2.5	2.5	2.5
Kish Deff	1.060	1.152	1.123	1.126	1.128	1.126

N=178,264 observations (unweighted).

Appendix Table 4: BSD and ASHE population profiles, original ASHE weight and adjusted ASHE weight, 2018

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	BSD (employee- weighted), per cent	ASHE weighted, per cent	Difference (2)-(1)	Squared error	ASHE adjusted weight, per cent	Difference (5)-(1)	Squared error
Number of employees (banded)				MSE: 9.5			MSE: 1.4
1-4	11.6	4.3	-7.3	53.0	8.8	-2.8	7.8
5-9	6.1	4.2	-1.8	3.3	5.2	-0.9	0.8
10-49	15.0	14.3	-0.7	0.5	16.4	1.4	1.9
50-249	13.9	14.9	1.0	1.1	14.9	1.0	1.0
250-499	5.6	6.3	0.7	0.4	5.9	0.3	0.1
500-999	5.7	6.4	0.7	0.5	5.9	0.2	0.0
1000-2499	7.2	8.1	0.9	0.8	7.1	-0.1	0.0
2500-4999	7.3	9.0	1.7	3.0	7.2	0.0	0.0
5000 plus	27.7	32.5	4.8	22.7	28.7	1.0	0.9
Legal status				MSE: 10.2			MSE: 0.4
Company	72.1	65.3	-6.8	46.1	71.2	-0.9	0.8
Sole proprietor	2.7	1.6	-1.0	1.1	2.8	0.1	0.0
Partnership	2.4	2.1	-0.3	0.1	2.5	0.1	0.0
Public corporation	0.6	0.6	0.0	0.0	0.4	-0.2	0.0
Central government body	8.8	12.4	3.6	12.7	10.1	1.3	1.7
Local authority	6.6	9.6	3.0	9.0	6.9	0.3	0.1
Non-profit making body	6.8	8.5	1.6	2.7	6.2	-0.6	0.4
Age of firm				MSE: 20.6			MSE: 2.5
1 year or less	3.3	1.1	-2.2	4.8	1.9	-1.4	1.9
2-4 years	8.7	3.8	-4.9	23.8	6.6	-2.1	4.3

5-9 years	10.2	7.0	-3.2	10.2	11.2	1.0	1.1
10-19 years	13.7	11.7	-2.0	4.1	12.7	-1.1	1.1
20-29 years	24.5	29.3	4.7	22.4	26.7	2.1	4.5
30 years plus	39.6	47.2	7.6	57.6	41.0	1.4	1.8
Region				MSE: 1.6			MSE: 0.9
NE	3.3	3.4	0.1	0.0	3.4	0.2	0.0
NW	9.9	9.8	-0.1	0.0	9.8	0.0	0.0
Yorks	7.9	8.6	0.7	0.4	8.5	0.7	0.4
EM	7.2	7.4	0.2	0.0	7.5	0.3	0.1
WM	8.6	8.5	-0.1	0.0	8.6	0.1	0.0
SW	10.2	10.5	0.3	0.1	10.7	0.5	0.2
East	21.0	18.3	-2.7	7.3	18.2	-2.9	8.2
London	14.1	14.1	0.1	0.0	14.2	0.2	0.0
SE	7.3	7.6	0.3	0.1	7.7	0.4	0.2
Wales	3.4	3.7	0.2	0.1	3.6	0.2	0.0
Scotland	7.2	8.1	0.9	0.8	7.7	0.5	0.3
SIC 2007 Section level				MSE: 2.0			MSE: 0.7
Agriculture, forestry, and fishing	0.9	0.5	-0.4	0.1	0.6	-0.3	0.1
Mining and quarrying	0.2	0.2	-0.1	0.0	0.1	-0.1	0.0
Manufacturing	7.9	9.4	1.5	2.1	9.5	1.5	2.4
Electricity, gas, air cond. Supply	0.7	0.7	0.0	0.0	0.7	0.1	0.0
Water supply, sewerage, waste	0.7	0.6	-0.1	0.0	0.7	0.0	0.0
Construction	4.9	3.8	-1.1	1.2	4.4	-0.5	0.2
Wholesale, retail, repair of vehicles	15.3	15.3	0.0	0.0	16.1	0.8	0.7
Transport and storage	4.4	4.1	-0.3	0.1	4.2	-0.2	0.0
Accommodation and food service	7.9	6.0	-2.0	4.0	7.0	-0.9	0.9

Information and communication	4.3	3.8	-0.5	0.2	4.1	-0.2	0.0
Financial and insurance activities	3.5	3.6	0.1	0.0	3.7	0.2	0.0
Real estate activities	1.8	1.4	-0.5	0.2	1.5	-0.4	0.1
Professional, scientific, and technical	8.1	7.1	-1.0	0.9	8.0	-0.1	0.0
Admin and support services	9.3	6.0	-3.3	10.9	6.9	-2.4	5.8
Public admin and defence	3.3	4.7	1.4	2.0	3.4	0.1	0.0
Education	10.2	13.9	3.8	14.3	11.1	0.9	0.8
Health and social work	12.1	14.8	2.7	7.3	13.6	1.5	2.2
Art, entertainment, and recreation	2.4	2.2	-0.2	0.0	2.2	-0.3	0.1
Other service activities	2.0	1.8	-0.2	0.0	2.1	0.1	0.0
Activities of households as employers	0.2	0.1	-0.1	0.0	0.3	0.1	0.0
N observations (unweighted)	2,952,764	178,264			178,264		

Appendix Table 5: Wage regressions, unweighted, standard and alternative ASHE weights, 2018, full results

		Unweighted			Original weight			Alternative weight			Stat sig diff	
		Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.	Orig weight vs unweighted	Alternative weight vs unweighted
Age	Age	0.004	***	0.000	0.005	***	0.000	0.005	***	0.000	***	***
Gender	Male	0.082	***	0.002	0.085	***	0.002	0.080	***	0.002		
Full-time	Full-time	0.065	***	0.002	0.066	***	0.002	0.066	***	0.002		
Occupation	Managers, directors and senior officials	0.695	***	0.004	0.689	***	0.004	0.668	***	0.004		***
	Professional occupations	0.704	***	0.003	0.695	***	0.004	0.681	***	0.004		***
	Associate professional and technical	0.456	***	0.004	0.452	***	0.004	0.447	***	0.004		
	Administrative and secretarial	0.218	***	0.003	0.214	***	0.004	0.216	***	0.004		
	Skilled trades	0.220	***	0.004	0.224	***	0.004	0.213	***	0.004		
	Caring, leisure and other service	0.085	***	0.004	0.082	***	0.004	0.079	***	0.004		
	Sales and customer service	0.047	***	0.004	0.046	***	0.004	0.040	***	0.004		
	Process, plant and machine operatives	0.064	***	0.004	0.061	***	0.005	0.058	***	0.005		
Employer size	5-9	0.039	***	0.006	0.046	***	0.006	0.030	***	0.005		
(ref: 1-4)	10-49	0.089	***	0.005	0.109	***	0.005	0.094	***	0.004	*	
	50-249	0.131	***	0.005	0.158	***	0.005	0.141	***	0.004	**	
	250-499	0.150	***	0.006	0.177	***	0.006	0.163	***	0.005	**	
	500-999	0.161	***	0.006	0.190	***	0.006	0.175	***	0.005	**	
	1000-2499	0.171	***	0.006	0.201	***	0.006	0.184	***	0.005	**	

	2500-4999	0.172	***	0.006	0.201	***	0.006	0.182	***	0.005	**	
	5000 plus	0.163	***	0.005	0.190	***	0.006	0.172	***	0.005	**	
Region	NE	-0.048	***	0.005	-0.051	***	0.005	-0.056	***	0.005		
(ref: London)	NW	-0.056	***	0.003	-0.060	***	0.004	-0.063	***	0.004		
	Yorks	-0.066	***	0.004	-0.071	***	0.004	-0.074	***	0.004		
	EM	-0.049	***	0.004	-0.055	***	0.004	-0.061	***	0.004		
	WM	-0.044	***	0.004	-0.046	***	0.004	-0.052	***	0.004		
	SW	-0.029	***	0.003	-0.029	***	0.004	-0.030	***	0.004		
	East	0.068	***	0.003	0.071	***	0.003	0.069	***	0.003		
	SE	-0.053	***	0.004	-0.057	***	0.004	-0.057	***	0.004		
	Wales	-0.072	***	0.005	-0.076	***	0.005	-0.080	***	0.005		
	Scot	-0.034	***	0.004	-0.038	***	0.004	-0.042	***	0.004		
Legal status	Sole proprietor	0.008		0.007	0.014	*	0.008	0.012	**	0.006		
(ref: Company)	Partnership	0.012	**	0.006	0.022	***	0.006	0.019	***	0.006		
	Public corporation	0.078	***	0.012	0.070	***	0.012	0.075	***	0.014		
	Central government body	0.067	***	0.004	0.066	***	0.004	0.072	***	0.005		
	Local authority	0.039	***	0.005	0.033	***	0.005	0.042	***	0.006		
	Non-profit making body	0.037	***	0.004	0.030	***	0.004	0.032	***	0.005		
Industry	Agriculture, forestry, and fishing	-0.119	***	0.011	-0.137	***	0.012	-0.124	***	0.011		
(ref: Manufacturing)	Mining and quarrying	0.237	***	0.022	0.238	***	0.022	0.234	***	0.023		
	Electricity, gas, air cond. supply	0.153	***	0.011	0.147	***	0.011	0.156	***	0.010		
	Water supply, sewerage, waste	-0.010		0.011	-0.014		0.011	-0.017		0.011		
	Construction	0.034	***	0.005	0.033	***	0.005	0.028	***	0.005		
	Wholesale, retail, repair of vehicles	-0.147	***	0.004	-0.145	***	0.004	-0.141	***	0.004		
	Transport and storage	0.036	***	0.005	0.035	***	0.005	0.033	***	0.005		
	Accommodation and food service	-0.228	***	0.005	-0.240	***	0.005	-0.237	***	0.005		
	Information and communication	0.062	***	0.006	0.068	***	0.005	0.070	***	0.005		

	Financial and insurance activities	0.159	***	0.006	0.181	***	0.006	0.182	***	0.006		
	Real estate activities	-0.068	***	0.008	-0.069	***	0.008	-0.065	***	0.008		
	Professional, scientific, and technical	0.036	***	0.004	0.044	***	0.004	0.043	***	0.004		
	Admin and support services	-0.114	***	0.004	-0.108	***	0.005	-0.100	***	0.005		
	Public admin and defence	-0.088	***	0.006	-0.086	***	0.006	-0.091	***	0.007		
	Education	-0.099	***	0.005	-0.089	***	0.005	-0.093	***	0.006		
	Health and social work	-0.166	***	0.005	-0.161	***	0.005	-0.158	***	0.005		
	Art, entertainment, and recreation	-0.208	***	0.007	-0.216	***	0.007	-0.213	***	0.007		
	Other service activities	-0.131	***	0.007	-0.138	***	0.007	-0.137	***	0.007		
	Activities of households as employers	-0.007		0.024	0.017		0.027	-0.005		0.018		
Firm age	1 year or less	0.012		0.009	0.016	*	0.009	0.017	**	0.007		
(ref: 30 years plus)	2-4 years	-0.006		0.005	-0.008		0.005	-0.007	*	0.004		
	5-9 years	-0.004		0.004	-0.004		0.004	-0.004		0.004		
	10-19 years	-0.003		0.003	-0.005		0.003	-0.007	**	0.003		
	20-29 years	0.008	***	0.002	0.007	***	0.002	0.005	*	0.003	***	
	Constant	6.564	***	0.007	6.513	0.008	***	6.542	***	0.007	***	***
	R-squared			0.508		0.511				0.505		
	Adjusted R-squared			0.508		0.511				0.505		
	N observations (unweighted)			166,243			166,243			166,243		

Notes: Statistical significance indicated as ***significant at 1 per cent, **significant at 5 per cent, * significant at 10 per cent.

