Understanding official data sources

Final Report for the Low Pay Commission

Low Pay Commission (tender no. 14/5): Evaluation of official data sources

Felix Ritchie*, Damian Whittard*, Chris Dawson**

* University of the West of England, Bristol

** University of Bath

February 2014

Contents

Acknowledgements

We are grateful for the comments of the Low Pay Commission and secretariat, in particular Yi Zhang and Tim Butcher, and attendees at the LPC workshops. Yi's knowledge of the imputation code also proved invaluable.

The opinions expressed in this paper are those of the authors, and not the producers of the data or statistics, nor the LPC or ONS.

ASHE and some LFS data was accessed via the ONS VML (project 12016). Other LFS data and BHPS/USoc data was made available through the UK Data Service (project 72317). All statistical results presented are based on the authors' own calculations.

Corresponding author: Felix Ritchie, Bristol Business School, University of the West of England Bristol.

Email: felix.ritchie@uwe.ac.uk

Executive summary

The paper is split into three main sections: a review of existing data sources; an analysis of characteristics of the main reference datasets used for the production of research and official statistics; and an investigation into the imputation code used by the LPC to produce official statistics.

Part A: Official data sources

This section provides a qualitative summary of microdata resources and macroeconomic statistics, focussing on their availability and potential. The data is presented in three sections: core microdata, supplementary microdata and aggregate statistics. This section provides the LPC, and other potential users, with a simple overview of official data sources and their limitations with reference to low pay analysis.

This section is for information only; there are no recommendations for the LPC.

Part B: rounding behaviour and measurement error in microdata

The two most important data sources for the LPC are the employer survey, the Annual Survey of Hours and Earnings (ASHE), and the household-based Labour Force Survey (LFS). A report for the LPC (Fry and Ritchie, 2013) reported significant evidence for 'rounding' behaviour in both the employer and employee datasets. For employer data, 'rounding' behaviour indicates wages are not being set directly in relation to measurable productivity; other characteristics (such as the 'going wage' for a job) may influence wage setting. This may imply a greater degree of wage flexibility in the labour market, employers with limited resources who follow simple rules-of-thumb when setting wages, or both of these. In contrast, 'rounding' in employee data implies measurement error.

Analysis conducted in an earlier LPC report (Fry and Ritchie, 2013) demonstrated significant rounding behaviour, and indicated that this behaviour was predictable at a general level: given a minimum wage, the popularity of certain wages being reported by employers and employees could be done with a certain degree of confidence. This report attempted to unpick the rounding behaviour in both employer and employee datasets so that (a) employer behaviour can be better understood and (b) allowances can be made for measurement error. The main method of investigation was to repeat, many times, analysis on different combinations of years, datasets and variables to see what common themes emerged, if any.

For employers, rounding behaviour is predictable within limits. The smaller the firm, the more likely it is to round wages. The private sector is more likely to round wages, and collective bargaining reduces the likelihood of wage rounding. Most importantly, the occupation of the employee appears to be a more important factor than the industrial sector of the employer. Low-paying occupations are more likely to have wages set very tightly, irrespective of the sector they occur in. Thus a carer is less likely to have wages rounded, but for others working in the social care industry rounding is more likely; the implication is that non-carers working in the social care industry have more flexibility in their wages. This suggests that the LPC's resources are better focused on the occupation of low-wage employees rather than the industry they work in.

The analysis of the LFS, based on employee responses, shows that very little seems to determine measurement error; that is, it appears to be randomly distributed amongst households. There are some systematic effects (smaller firms, private sector) but as these also occur in ASHE this indicates accurate reporting of employer rounding behaviour, not employee error. Apart from employer effects already identified in ASHE, the only continually significant variables in LFS are those for proxy response and whether the respondent used his or her payslip in response to wage questions (but not other forms of documentation). In other words, the only things that reliably seem to be associated with measurement error are whether the person of interest filled in the form, and whether the hourly wage rate was checked by the respondent.

Despite the apparent negativity of these findings, this can be seen as positive news for the LPC. It is difficult to allow for measurement error when it systematically changes with variables of interest; but a genuinely random error is

statistically much easier to deal with. This provides support for the notion that ad hoc adjustments can be made to LFS statistics to bring them into line with the ASHE (for example, adjusting for the high numbers of employees reporting £6.00 wages in LFS when ASHE suggested many of these were being paid the NMW of £5.93). This is also a positive finding for low pay researchers carrying out the sort of multivariate analysis which the LPC typically commissions; and it explains the lack of a measurement error effect carried out in previous studies which explicitly tested for this (Fry and Ritchie, 2013; Le Roux et al, 2013).

Fry and Ritchie (2013) only studied ASHE and LFS; their results may be specific to those datasets, which have very different sampling qualities. In this paper, we carried out a similar analysis on two other datasets, BHPS and USoc, which have similar sampling and data characteristics to the LFS (there is no equivalent of ASHE). These support the view that household measurement error is largely random. However, they show that directly asking the respondent whether a wage rate is 'exact' or 'estimated' makes a notable difference to outcomes.

Recommendations to the LPC from this section are as follows:

- we do not feel that further analysis on 'rounding' issues is likely to provide sufficient new information to justify significant LPC funding
- LPC discussions with ONS over data quality should focus on the LFS and, in particular, the impact of the 'proxy' and 'documentation' variables; the 'exactness' check in BHPS and USoc may suggest a model
- findings on occupation suggest that LPC might usefully focus on its attention on the type of jobs being done, rather than the industry in which the workers operate; and on smaller, private firms in all industries.
- the apparent randomness of measurement error suggests that the LPC can place confidence in multivariate analysis using the LFS; in the case of descriptive statistics, the findings produced here would support ad hoc adjustment (for example to align with ASHE)

Part C: Imputation of earnings using LFS

ASHE is produced by ONS and used to calculate official statistics on earnings by both the LPC and ONS. However, some additional statistics required by the LPC (breakdowns by ethnicity, education, ethnicity and nationality; numbers of low-paid jobs) are not available in ASHE and are produced by the LPC from the LFS. A problem is that the preferred measure of earnings in the LFS, the stated wage, is missing in many cases, particularly for higher earners. This means that some of the statistics the LPC wishes to produce are based on very small numbers of recorded values.

A simple response is to scale up the LFS data so that the aggregate statistics are consistent with ASHE. An alternative is to use derived wage data (total earnings divided by hours) to supplement the stated wage data. Both of these make a number of strong assumptions about the distribution of the missing data.

An alternative strategy is to 'impute' the missing values using statistical models. About 2002/2003, ONS developed code to implement this strategy, using a relatively standard technique called 'multiple imputation' which was seen as best practice. ONS no longer uses this code as it only produces statistics from ASHE, but the LPC continues to use it as the LFS results are needed. The research team were asked to investigate whether the imputation code is still fit for purpose.

The team identified three potential errors with the data: not recognising the asymmetry caused by the existence of the NMW; small numbers of observations; and the choice of interventions available to the operator, which requires subjective judgements to be made. To understand whether theoretical problems occurred in practice, the team specified four alternative ways of carrying out the imputation, and reviewed the impact on aggregate statistics.

The analysis showed that the imputation method did matter. Subjective judgements about how outliers should be treated were relatively unimportant, but small adjustments to the underlying statistical model had a larger effect on descriptive statistics. These variations were largest in groups with very small numbers: ethnic minorities, disabled employees aged under 21, apprentices, for example. Whilst the code did not appear to generate imputed values

below the NMW, it did generate plausible wage distributions based on few or no data points. These were reflected in estimates for the 'bite' of the NMW and the number of low-paid jobs (NMW+5p) where ethnic minorities in particular saw their statistics change; in some cases the statistics seemed unlikely (10% of one population subgroup earning below the NMW, for example).

Analysing only stated wages produced nonsensical results due to the lack of high earners, so the technique of using derived hourly wage rates to fill in for missing stated wage rates was taken as the 'original' base case. This produced more variation than any of the imputation techniques, possibly because of a larger number of very low earners.

The analysis was hampered by problems with the code itself, which is difficult to update; does not allow alternative scenarios to be tested; does not allow original and imputed variable to be compared; produces inappropriate and excessive output; does not provide statistical information about the accuracy of estimates; and is very susceptible to errors in coding and use.

Overall, the conclusion is that the apparent methodological rigour of the imputation technique can generate statistics of spurious accuracy, but the lack of information about the estimates makes this impossible to quantify. On the other hand, the analysis clearly showed that simply scaling LFS results to ASHE statistics is difficult to justify: the sensitivity to the imputation method amongst population subgroups indicates differential impacts. Some form of imputation would therefore seem desirable, but the current coding does not have the flexibility or quality measures to make confident predictions.

Specific recommendations from this section are therefore

1. Imputation appears desirable in theory and practice but LPC should review
    a. whether the multiple imputation technique used is still the most appropriate
    b. whether the statistical model used is still appropriate
2. The current code is not fit for purpose; LPC should rewrite the code to
    a. simplify use and maintenance
    b. support scenario testing
    c. improve output, including quality measures
3. The revised code should be made available on the LPC website
4. Simply replacing missing stated wage rates with derived values should <u>not</u> be used in preference to imputing values

Part A: Official data sources

1. Introduction

The purpose of this section of the report is to provide a qualitative summary of the LPC's data sources.

The LPC's data sources (and uses thereof) have been extensively discussed by Knight (2010). Whereas Knight discussed the historical development of the data sources and their current use by the LPC, this report focuses on the availability and potential in the data sources. The aim is to provide a guide to researchers focusing on the relevance of the data source to LPC research. Opinions expressed in this report are those of the authors.

The microdatasets available through the UK Data Service are extensively documented there; and aggregate statistics, along with methodological notes, are accessible from the Statistics Hub. This report does not aim to duplicate the information provided there. This report should therefore be seen as complementary to Knight (2010), the UKDS website http://ukdataservice.ac.uk/, and the Statistics Hub www.statistics.gov.uk. As most data are produced by the Office for National Statistics (ONS), the site www.ons.gov.uk should also be consulted.

The focus in this report is on providing a simple overview of datasets available, along with their limitations and relevance to analysis of the low paid. The analysis at this stage is limited to public knowledge. Knight (2010) discusses private-sector sources which are omitted here, as these are out of scope of the project. This report covers some 'official' data sources which are not currently used by the LPC, and omits, at this stage, some of the sources listed by Knight. This reflects the team's decision to focus initially on the data easily available to government and academic researchers. The BIS Apprentice Pay Survey has not been covered in this report as a separate project is analysing this data.

The next section focuses on the 'core microdata' directly analysing those on the minimum wage. Section three reviews relevant datasets which are used by the LPC (or may have significant potential), either to supplement the direct analysis of the low paid or to provide more detail on the economic and social context. Section four reviews aggregate statistics, concentrating initially on those currently in use by the LPC.

Note on availability

All aggregates statistics are available on the internet. For microdata, the discussion below describes five levels of access to the data. These are as follows

Internet — All users; available for downloading from the specified source

End User Licence — Academics; available from downloading from the UK Data Service
http://ukdataservice.ac.uk/

Special Licence — UK academics, subject to meeting additional criteria; available for downloading from the UK Data Service
http://ukdataservice.ac.uk/

Secure facility — Available from one of the restricted access services:

- *SDS*: Secure Data Service (for UK-based academics only; remote access to UK academics from university sites)
  http://ukdataservice.ac.uk/
- *VML*: the ONS Virtual Microdata Laboratory (all other users; remote access from ONS sites and selected government offices)
  http://www.ons.gov.uk/ons/about-ons/who-we-are/services/virtual-microdata-laboratory/index.html
- *HMRC*: the on-site facility at HMRC offices (all researchers but projects must be of direct interest to HMRC)
  http://www.hmrc.gov.uk/datalab/
- *DWP*: the on-site facility established at the DWP offices (all users)
- *DWP-Essex*: the on-site facility at the University of Essex built to support access to more sensitive DWP datasets

http://ukdataservice.ac.uk/

- *Europe*: Access from the Eurostat Research Data Centre in Luxembourg

Service request Bespoke analysis may be carried out by the data holders on behalf of researchers

Select glossary for business data

Enterprise	A single business, typically the legal entity (for example, a chain of shops). See Ritchie et al (2012).
Enterprise group	An overarching structure comprising one or more enterprises; for example, a retail chain might have three enterprise within the enterprise group: the shops chain, a chain of petrol stations, and a financial services unit. See Ritchie et al (2012).
GOR	Government office region: nine English regions plus Wales, Scotland and NI.
IDBR reference	A unique number attached to all operating units (establishments and groupings) in the ONS register of businesses; allows business data to be linked across surveys.
Imputation	Using information from other sources (for example, similar-sized companies in the same sector) to fill in information which is not supplied on the original survey form.
Local unit	The establishment, the place where economic activity takes place; for example, a factory or shop. See Ritchie et al (2012).
Reporting unit	A collection of local units (see above) reporting statistics to the ONS. For 99% of businesses, this is the same as the enterprise (for example, a chain running a selection of shops). See Ritchie et al (2012).
Short- and long-form	Some survey respondents will get a 'long form', containing all information required. Others will get a 'short form' asking only a subset of questions, with the remaining information imputed from the answers by long-form respondents. Hence, there is scope for measurement error and collinearity in the microdata.
Statistical/selective editing	Automatic validation checks are carried out on all ONS data; under statistical editing, unlikely values are only flagged up for verification if changing the 'questionable' value to a more likely value is likely to have an impact on aggregate statistics. Integrity of the microdata is not a consideration , and so surveys with statistical editing applied are more likely to have unrecognised errors in data units with small values.

2. Core microdata

2.1 Annual Survey of Hours and Earnings (ASHE)

Common acronym:	• ASHE; NES for pre-2004 data
Available period:	• From 1975 onwards (New Earnings Survey, or NES) • From 2004 onwards (ASHE) • Data available approximately seven months after survey date
History:	• Created as New Earnings Survey in 1975, running annually • This was maintained as the NES Panel Dataset (NESPD) with individuals traceable across years by a common personal reference number; however, not all updates used the same reference number and so each panel (1975-1990, 1975-1991, 1975-1992 etc) was released, and should be treated, as a separate dataset • The annual NES data files were recovered by the VML team in 2003. These appeared to contain some more information than the NESPD. However, as each year file could have several versions, it is possible that the files selected by the VML team are not the 'final' or 'complete'. • In 2004 NES was replaced by ASHE with the same sample but additional questions on pensions and more focus on chasing up low earners; the estimation methodology was also changed; see Milton (2004). • There were further changes in 2007, due to the introduction of an occupational coding tool and a 20% cut in the survey size (ONS, 2007a); the 20% cut was subsequently reversed but it means that non-appearance in ASHE in 2007 must not be taken as evidence on non-employment • ONS created a pseudo-ASHE from the NES datafiles (Daffin, 2004); the main difference is to include weights calculated post hoc from the LFS. ASHE variables are included, but these are zero where the relevant data was not collected in the NES. • The LPC therefore views ASHE as four ASHE datasets, distinguished by changes in the sampling methodology, estimation methodology, or both: 1997-2003, 2004-2006, 2007-2010, 2011 onwards (the latter reflects the introduction of SIC 2010) • There is also an ASHE Panel (2002-); note that years 2002/2003 are based on NES data and so have some missing variables.
Type:	• Longitudinal, individual data • Annual
Sample:	• Employees only • Roughly 180,000 each year, sampled on basis of employment in second week of April each year • Target is 1% pure random sample of employees; achieved sample is nearer 0.8%. The same employees are counted each year, with retirees being replaced by new workers. Non-employment appears as absence from the data for that year.
Variables:	Earnings: Detailed information on hourly and totals earnings, including bonus incentives overtime and other pay; target hourly wage rates where those exist Hours: Detailed hours data Employment status: Employed only; within that, detail on FT/PT, apprentice, temporary/permanent, absence, whether in same job over one year; multiple jobs allowed Geography: Postcode (home and work) Industry: Five-digit SIC, current at time of data collection Occupation: Four-digit SOC, current at time of data collection

	Other personal: Gender, age in years Other: Pensions
Linkage:	• Potential for linking to DWP data through National Insurance (NI) number • Linked to ONS business data (2004 onwards) through IDBR reference number • Data can be linked across years by person ID and by job-based serial number
Data quality:	• Strong support for the accuracy of the earnings data, supplied by employer records; see Ormerod and Ritchie (2007a), Fry and Ritchie (2012) • Linking of individuals over time can be problematic because of the use of temporary NI numbers • Employer details are taken from IDBR home postcodes are known to be less accurate, and missing frequently
Strengths:	• Detailed hours and earnings information • The relevant NMW can be identified for all workers including apprentices • Large numbers mean very detailed analyses possible • Longitudinal element allows fixed-effect regression to be used • Linking to businesses allows firm characteristics to be included in earnings analysis • Identification of being in the same *job* allows this to be distinguished from being in the same firm
Weaknesses:	• Limited personal characteristics, particularly education; occupation is usually used as a proxy, or fixed-effects regression • No information on what people are doing when they are not included: they could be unemployed, self-employed, or out of the labour force for a period
Relevance for low pay analysis:	• Large data set with detailed earnings allowing inventive breakdowns into sub-groups • Concerns over coverage of the low-paid in the NES seem to have been largely addressed in the ASHE re-design • However, there are separate weights for use in low pay analysis to allow for residual under-representation
Current usage:	• Widely used by earnings researchers • Used extensively by government departments, particularly DWP and BIS • Primary source for LPC
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: All data • VML (approx. 7 months after survey) • SDS (one year after survey) *Service request*: Available to government users
Aggregate stats:	• Annuals hours and earnings series, with detailed breakdowns • Gender pay gap
Other:	• Combined with LFS to produce European Structure of Earnings Survey

2.2 Labour Force Survey/Annual Population Survey (LFS/APS)

Common acronym:	• LFS, APS
Available period:	• From 1992 onwards, quarterly (seasonal quarters up to 2009, calendar quarters from 2006) • Data available approximately six months after survey date for researchers (two months for government departments, including LPC)
History:	• Largely unchanged since late 1990s • Moved from seasonal quarters to calendar quarters in 2005/6
Type:	• Cross-section, individual • Some longitudinal element: respondents are surveyed for five quarters, with earnings information gathered in the first and last quarter
Sample:	• Roughly 60,000 per quarter • Random selection of households, then follow-up for four quarters • APS is boosted by additional one-off sample in Wales and Scotland
Variables:	Earnings: Stated wage, total earnings Hours: Target hours, actual hours Employment status: Employed, self-employed, unemployed, student, out of labour force – employment detail of FT/PT, temporary/permanent, apprentice Geography: Postcode Industry: Two-digit SIC up to 2010, five-digit SIC from 2011; all current at time of data collection Occupation: Four-digit SOC, current at time of data collection Education: Detailed qualifications and current status Health: Yes, including some detail on treatment/benefits and ability to work Ethnicity: Yes Disability: Yes Attitudinal vars: No Other personal: Proxy response, information provided with documentation (payslip, bank statement, other) Household vars: Relationship to household, gross income Other: Information about who filled in form and whether documentation was used
Linkage:	• No linkage currently possible (although see Ormerod and Ritchie (2007b)
Data quality:	• See ONS (2013e, updated quarterly) for current quality report • Response rates have fallen over time: from 65%-75% (q5-q1) in 2000 to around 45%-55% currently, with 'outright refusal' on the doorstep responsible for over 50% of the non-responses and 'non-contact' less than 20%. Large metropolitan areas appear to have lower response rates. • Proxy responses are over 70% for teenagers (16-19) • Fry and Ritchie (2013) highlighted problems with earnings data being rounded, particularly for those not referring to documentation. Their research was not clear whether this was a general trait or specific; the study showed little regional or industry differences but some occupational difference. • Ormerod and Ritchie (2007b) suggested overall distribution of hours and earnings was acceptable (ie similar to ASHE) but this is not relevant for highest earners and those near the NMW

	• High earners seem to be under-represented in the LFS • Concerns over whether information provided about employers is accurate eg LFS firms seem to be much smaller than ASHE/IDBR firms. Suspicion is that LFS respondents describe place of work rather than business.
Strengths:	• Rich set of personal characteristics • Large number of observations if viewed cross-sectionally (or using APS) • Quarterly data allowing response to NMW changes to be assessed
Weaknesses:	• Concerns over accuracy of some data, particularly earnings, although Fry and Ritchie (2013) suggest that this is very predictable • Repeated changes to classification (eg for ethnicity/nationality), and the use of year-specific variable names means analysis over several periods involves extensive recoding
Relevance for low pay analysis:	• Only data (save BHPS) which has both wage information and personal characteristics; hence necessary for analysis of effect of low pay on disabled, ethnic minorities, less educated etc
Current usage:	• Main source for labour market analysis in the UK by academics, especially data archive download • Also available in teaching set form • Used by LPC to connect low wages with personal characteristics • Contributes to European Labour Force Survey
Availability:	*Internet* — Not available *End User Licence* — Age limited to bands; government office region only *Special Licence* — Geography increased to LA, age in years *Secure facility (SDS, VML)* — All data (equivalent to within-government research version) in SDS and VML *Secure facility (Europe)* — Under review *Service request* — Not available
Aggregate stats:	• Hours and earnings by personal characteristics • Statistics on labour force: employment, unemployment, self employment
Other:	• Contributes to European Labour Force Survey, available as an End User Licence file with limited detail

2.3 British Household Panel Survey

Common acronym:	• BHPS
Available period:	• From 1991-2008/2009 (Waves 1-18) • Data available approximately 12 months after survey date
History:	• The BHPS was primarily funded by the ESRC with support from government departments. Since the start of the survey the University of Essex has been the Principal Investigator. • Wave 1 panel (1991) consists of 5,500 households (10,300 individuals) drawn from 250 areas of Great Britain. Additional samples of 1,500 households in each of Scotland and Wales were added to the main sample in 1999, and in 2001 a sample of 2,000 households was added in Northern Ireland. From 2009 onwards the BHPS sample has been merged into a much larger new longitudinal household study ('Understanding Society', USoc) with further widening of scope, including biosocial analysis.
Type:	• Longitudinal, individual • Annual
Sample:	• Basis for selection - a stratified random cluster sample of households was drawn from the population of British household postal addresses in Great Britain. • The same individuals are re-interviewed in successive waves. If individuals leave their original household, all adult members of their new households are interviewed. Children are interviewed once they reach the age of 16. From Wave 9 (1999), two additional booster samples were added to the BHPS for Scotland and Wales. At Wave 11 (2001) an additional sample from Northern Ireland (which forms the *Northern Ireland Household Panel Study* or NIHPS), was added. Waves 7-11, the BHPS provides data for the European Community Household Panel (ECHP). • The original sample of 5,500 households (comprising approximately 10,300 individuals) drawn from 250 areas of Great Britain was recruited in 1991. Follow-on rules establish the tracking of newly forming households involving originally-enumerated household members. Sample attrition rates in the BHPS are generally low and certainly comparable to those achieved in other similar household panels. As is typical with household panels the highest attrition rate of individuals was between Waves 1 and 2 (12%). Attrition between Waves 2 and 3 was 7% of the original individuals and subsequently averaged 2.4% of the original sample between waves. In common with nearly all previously published research using this data source, attrition is assumed to be a random event.
Variables:	Earnings: Hourly wage, annual income, overtime pay, performance related pay Hours: Detailed information on hours of work, overtime hours (paid/unpaid) Employment status: Self-employed, paid employment, unemployed, retired, family care, full time student, long term sick/disabled, maternity leave, government training scheme, or 'other'. Also information on second jobs. Information is available concerning the day the current labour force status begun. Geography: Postcode Industry: SIC and SIC(92) Occupation: SOC and SOC(2000) Education: Detailed lists of qualifications, highest educational attainment

	Health: Detailed information on subjective and objective health measures Ethnicity: Available in all waves, detailed breakdown. Country of Birth and Ethnic Group Membership. Disability: Available in all waves, detailed breakdown. Also information on how disability effects work. Attitudinal vars: Attitudes to work, the environment, gender roles Other personal: Gender, age in years Household vars: Household consumption, household income, household debt, home ownership Other
Linkage:	• Links to USoc (which comprises the BHPS panel cohort plus additional households); however, there have been some problems merging the two datasets • Possible links to *Northern Ireland Household Panel Study* (NIHPS) and the European Community Household Panel (ECHP)
Data quality:	• The BHPS provides high quality data to analyse change at the individual and household level for domains including: labour markets, income, savings and wealth, household and family organisation, housing, consumption, health, social and political values, education and training. More information at: https://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf
Strengths:	• Detailed information on individuals, including attitudes, labour market activity, health, finances and consumption • Individuals and households repeatedly re-interviewed • Can track changes in individuals and households
Weaknesses:	• Sample attrition (12% between waves 1 and 2)
Relevance for low pay analysis:	• Large data set with detailed earnings and hours data. • Extensive individual and household variables to allow inventive breakdowns into sub-groups • Able to track changes in individual, especially important to analyse changes individual reporting of earnings over time.
Current usage:	• Widely used by social scientists
Availability:	*Internet*: Data Archive ESDS *End User Licence*: Geography limited to GOR/ metropolitan area *Special Licence*: Not available *Secure facility*: All data at SDS *Service request*: Not available
Aggregate stats:	•
Other:	•

2.4 Understanding Society

Common acronym:	• USoc
Available period:	• From 2009-2012 (Waves 1-3). • Data available approximately 12 months after survey date
History:	• The Wave 1 panel (2009) started on 8th January 2009 and ended on the 7th of March 2011, consisting of 30,169 households (50,994, individuals) drawn from a representative probability sample of households from England, Scotland, Wales and Northern Ireland. This sample called the General Population Sample has various sub-groups, including; the Ethnic Minority Boost Sample and also a subset that includes participants from The British Household Panel Survey. Additionally, there is a separate survey called the Innovation Panel (IP), which is fielded in the year before the main survey in order to test varying measurement issues.
Type:	• Longitudinal - Data collection for a single wave is scheduled across 24 months. The same individuals are re-interviewed in successive waves. If individuals leave their original household, all adult members of their new households are interviewed. Household members aged 10-15 years are asked to complete a short self-completion youth questionnaire and children are interviewed once they reach the age of 16. • Annual
Sample:	• Number - The original sample of 30,169 households (comprising approximately 50,994 individuals) is drawn from England, Scotland, Wales and Northern Ireland. Data collection for a single wave is scheduled across 24 months. Sample attrition rates in USoc are generally low and certainly comparable to those achieved in other similar household panels, including the BHPS. The highest attrition rate of individuals was for the Ethnic Minority Boost Sample. Overall, for Waves 1 and 2, pairs of observations are available for 45,836 adults (47,282 adults if proxy and telephone interviews are included). • Basis for selection - The England, Scotland and Wales sample of the General Population Sample is a proportionately stratified, clustered sample of addresses selected from the Postcode Address File. For Northern Ireland an unclustered systematic random sample of addresses was selected from the Land and Property Services Agency list of domestic addresses.
Variables:	Earnings: Hourly wage, annual income, overtime pay, performance related pay. Hours: Detailed information on hours of work, overtime hours (paid/unpaid) Employment status: Self-employed, paid employment, unemployed, retired, family care, full time student, long term sick/disabled, maternity leave, government training scheme, something else. Also information on second jobs. Information is also available on the day the current labour force status begun. Geography: Country level data – England, Wales, Scotland and Northern Ireland. Industry: SIC Occupation: SOC and SOC(2000) Education: Detailed lists of qualifications, highest educational attainment. Health: Detailed information on subjective and objective health measures

	Ethnicity	Available in all waves, detailed breakdown. Country of Birth and Ethnic Group Membership.
	Disability	Available in all waves, detailed breakdown. Also information on how disability effects work.
	Attitudinal vars	Attitudes to work, the environment, gender roles,
	Other personal	Gender, age in years
	Household vars	Household consumption, household income, household debt, home ownership.
	Other	
Linkage	• Links to Understanding Society and also possible links to *Northern Ireland Household Panel Study* (NIHPS) and the European Community Household Panel (ECHP).	
Data quality:	• USoc provides a large high quality longitudinal data to analyse change at the individual and household level for domains including: labour markets, income, savings and wealth, household and family organisation, housing, consumption, health, social and political values, education and training.	
Strengths:	• Detailed information on individuals, including attitudes, labour market activity, health, finances and consumption. • Individuals and households repeatedly re-interviewed • Can track changes in individuals and households. • Sample size and ethnic boosts. • Low sample attrition between waves 1 and 2	
Weaknesses:		
Relevance for low pay analysis:	• Large data set with detailed earnings and hours data. • Extensive individual and household variables to allow inventive breakdowns into sub-groups • Able to track changes in individual, especially important to analyse changes individual reporting of earnings over time.	
Current usage:	• Widely used by social scientists	
Availability:	*Internet ()*	Data Archive ESDS
	End User Licence	
	Special Licence	
	Secure facility (SDS, VML)	
	Secure facility (Europe)	Not available
	Service request	Not available
Aggregate stats.	•	
Other	•	

3. Supplementary microdata

3.1 Annual Business Survey / Annual Respondents Database

Previously known as the

- Annual Census of Production (ACOP; 1974-96; production sector only)
- Annual Census of Construction (ACOC; 1991-96; construction only)
- 6 service trade inquiries (1994-96)
- Annual Business Inquiry (ABI; 1997-2009)

In microdata form available as the Annual Respondent's Database (ARD) and ABS

Common acronym:	• ARD / ABS, but sometimes also (mistakenly) ABI
Available period:	• From 1997 onwards in coherent form • Data available approximately 24 months after survey date
History:	• Production data available from 1974 (ACOP), construction from 1991 (ACOC); six surveys covering the service sector were run from 2004-2006. All datasets had a similar structure of three files per year, covering selected reporting units, non-selected reporting units, and local unit data • The production, construction and services surveys were brought together in 1997 to form the Annual Business Inquiry, with greater coherence across sectors • The survey was renamed the Annual Business Survey in 2009 to reflect the introduction of the Business Register and Employment Survey • The source data was combined (originally by researchers at Queen Mary College, Imperial College, Nottingham, the LSE and the IFS) to create the ARD from the ABI, ACOP and ACOC; see Barnes and Martin (2002) • It was redesigned in 2005 as ARD2 to provide a single set of variables across all years; see Robjohns (2006) • In 2011 a change in the VML team meant the ARD was no longer updated; the 2009 file is therefore the last extant ARD year at present, although some ABS files have been deposited in the VML and SDS • In December 2013, following concerns about the absence of updated ARD or ABS files, a meeting between academics, VML, SDS and the ABS team agreed that (a) all outstanding ABS files would be made available in the SDS, and (b) all should standardise on 'ABS' as the name for the source data files, with some other name (possibly still 'ARD') for the longitudinally consistent research file. In January 2014 VML agreed to re-assume responsibility for the ARD and deliver a development plan for the ARD by April 2014.
Type:	• Cross-section, businesses (including public and private sector); however, the sampling strategy means that large businesses are likely to be repeatedly sampled, allowing for a form of longitudinal analysis • Annual
Sample:	• Sampling at 'Reporting unit' (RU), approximating to a legally registered business; around 1% of enterprises are broken into more than one RU • Limited data is also available on 'local units' (LUs, sometimes referred to as establishments); around 5% of businesses have more than one LU • Stratified by size: 100% for largest companies (>500 employees), falling to around 0.5% for microbusinesses • Sample is pseudo-random selection from IDBR
Variables:	Earnings: Total wage bill only Hours: None Employment status: Number of employees

	Geography: Postcode of LU and head office of RU Industry: Five-digit SIC for RU; Five-digit SIC for each LU Other: Detailed financial information on businesses allowing productivity calculations to be made, including: • Turnover • GVA and its components • Labour costs • Investment
Linkage:	• Linkable to all other ONS business data through IDBR references
Data quality:	• High for large companies • Short- and long-form survey for small firms; not always clear where imputation has taken place • Subject to statistical editing, increasing risk of unverified values in small companies • Informal analysis by the VML team suggests that employment was accurately measured (compared to HMRC PAYE records); there was less correlation between ARD turnover and HMRC VATable turnover, but this may be due to different reporting periods • No independent verification possible of LU data since 2009 • Employment data is point-in-time (currently, September) whereas financial data is whole-year; therefore there exists scope for inappropriate per-capita calculations, particularly for seasonal businesses (note that data up to early 2000s had year-average employment) • 2009 ARD data has been reported 'unusable' by some academics. The arrival of this data coincided with the loss of institutional memory (changeover of the VML team and much of the ABS team) and with the introduction of ABS/BRES. This version was put together at short notice as a temporary measure by the old VML team, ONS' productivity team and the SDS, but has come under criticism. • See also Ritchie et al (2012) for a discussion of industry classification
Strengths:	• Detailed information on companies, including establishments • Large companies repeatedly sampled • Can track changes in performance of individual establishments, which may not be in the same business as the parent organisation
Weaknesses:	• Imputation and concerns over data quality of very small companies • Negligible probability of repeat observation in microbusinesses • Long release time
Relevance for low pay analysis:	• Widely used by researchers, often linked with other business data • Potential to compare productivity in detail in low-paying and other sectors • Can compare changes over time in larger companies
Current usage:	• Only used as aggregate productivity statistics (see below)
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: All data • SDS (to 2009 currently) • VML (unknown) *Service request*: Not available
Aggregate stats:	• Source of official (annual) estimates of productivity and GDP

3.2 Business Structure Database/Business Register and Employment Survey/Inter-departmental Business Register

Common acronym:	• BSD/BRES/IDBR
Available period:	• From 1997 onwards • Data available approximately one year after survey date
History:	• The IDBR was created in 1997 to facilitate ONS' data collection; it is a live register of businesses • The BSD is a longitudinally linked set of annual snapshots of the IDBR, made available to and designed for researchers • BRES is the annual survey which populates the IDBR with employment and industry details for enterprises and establishments
Type:	• Longitudinal database, on businesses • Annual
Sample:	• Constructed from administrative data augmented by survey • Approximately three million businesses in the IDBR in any one year; ONS estimated this counts for 50% of businesses but 99% of economic activity • All businesses with VAT, PAYE, registration at Companies House or who otherwise come to ONS's notice are included • BRES is stratified by size with large companies more likely to be sampled • For small companies (and new companies not yet surveyed) data for employment, turnover and type of activity is imputed from HMRC records and nearest-neighbour methods
Variables:	Geography: Postcode Industry: Five-digit SIC, current at time of collection, plus a 'consistent' SIC code mapping 92/03/07 Other: Turnover, birth and death date
Linkage:	• Links to all other ONS business data
Data quality:	• Exceptional for largest companies – ONS employs dedicated team to analyse those • Employment data believed to be accurate across all businesses – even if not surveyed, data is updated from HMRC PAYE records • Turnover is taken from VAT returns in most cases, so may disagree with ABS, for example • New companies are most likely to be identified by VAT returns, and this is used to estimate initial employment; hence, new and small companies more likely to have imputed rather than observed employment • Links between local units for very large companies may be unreliable: rather than being updated each year, they may be binned and a new set recreated with new reference numbers, giving the appearance of massive restructuring. ONS argues that this was a past problem. No recent analysis has been done.
Strengths:	• Comprehensive coverage at enterprise group, enterprise and local unit level
Weaknesses:	• Limited information • Companies are slow to die – they need to have no economic activity for two years before ONS removes them from the IDBR. Therefore, it is likely that the IDBR and BSD contain a number of ghost companies
Relevance for low pay analysis:	• Scope to study effect on company existence and employment levels at the establishment level
Current usage:	• BSD commonly used by VML/SDS researchers to study business demography
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Available to central/local government for research use

	Secure facility VML, SDS: All data Europe: Under review *Service request* Possible for government departments
Aggregate stats:	• ONS business demography data
Other:	• BRES is incorporated into the BSD (via IDBR). Assuming ARD development recommences, BRES will be the source for RU/LU employment. • However, it is likely it will continue to be generally released as a separate dataset to clarify lines of responsibility between the BRES and VML teams

3.3 Census/Controlled Access Microdata/Small Area Microdata/Samples of Anonymised Records

Common acronym:	• Census, CAMS, SARS
Available period:	• 2001 Census data for CAMS and SAM • 1991 and 2001 Census data for SARS • 2011 CAMS/SAM/SARS should be available 2014 (source: VML team)
History:	• The 1991 Census was used to produce the first SARS files. These were suitably anonymised files, managed by CCSR in Manchester and available to all academics • Confidentiality concerns meant that the 2001 SARS were much more restricted. As a results the CAMS were produced and made accessible through the VML. The level of CAMS is arguably similar to the 1991 SARS. • The SAM were produced at the same time to allow more geographical detail at the expense of other variables • Household SARS (H-SARs) were also produced, with changes in geography reflecting changes in confidentiality addressed in the individual SAR
Type:	• Sub-samples from a Census
Sample:	• SARS 1991: 2% (1.1m individuals) • CAMS/SARS 2001: 3% samples (1.75m individuals) • H-SAR 1991/2001: 1% (200,000 households, linked to 500,000 individuals) • SAM: 5% (3m individuals) • See http://www.ccsr.ac.uk/sars/gettingstarted/index.html
Variables:	Earnings: None Hours: None Employment status: Labour market status Geography: GOR in most 2001 data; local authority in CAMS and 1991 data Industry: Yes Occupation: NS-SEC and occupational coding Education: Yes Health: Yes Ethnicity: Yes Disability: Yes Attitudinal vars: Religion Other personal: Country of birth, age, gender Household vars: Relationship between members Other
Linkage:	• See ONS-Longitudinal Study, below • In theory, a Census 2011-ASHE 2011 link is feasible and desirable but no funds were available in 2013 and the Census link team has been dissolved.
Data quality:	• Most variables categorical so limited scope for eg rounding errors • Many datasets (eg LFS) are calibrated to the Census and so independent evaluation of data is not possible
Strengths:	• Large sample sizes • High quality data
Weaknesses:	• Limited geography • No earnings and limited work information
Relevance for low pay analysis:	• Enables evaluation of employment probabilities based on personal characteristics • If ASHE can be linked to the Census, then:

	○ ASHE data in 2011 can be analysed with all the personal characteristics of the Census – the ASHE and Census data collection periods are two weeks apart ○ ASHE data in others years can be augmented by the inclusion of time-invariant variables in the Census (eg completed higher education) or predictable ones (age of youngest child)
Current usage:	• SARS used for research and teaching • CAMS very limited use due to restrictions on access
Availability:	*Internet* Not available *End User Licence* SARS 1991 and 2001, H-SAR 1991, SAM *Special Licence* H-SAR 2001 *Secure facility* VML (onsite at ONS): CAMS *Service request* Not available
Aggregate stats:	• Major population statistics/Neighbourhood Statistics
Other:	• Provides weights for LFS, ASHE (via LFS), and other surveys • See also Longitudinal Study, below

3.4 Family Resources Survey

Common acronym:	• FRS
Available period:	• From 1993
History:	• The FRS has been running since 1979 but was redesigned in 1993 • Detailed datafiles were made available as Special Licence files but the data is now only accessible though the DWP 'safe room' at Essex; however, the range of variables was expanded as part of this move
Type:	• Cross-section of households • Head-of-household (HD), household (HH), within-household (WH)
Sample:	• 25,000 households • Multi-stage stratified random sample
Variables:	Earnings: Total earnings; extensive information on non-employment income (HD) Hours: Yes (HD) Employment status: Yes (HD) and some (WH) Geography: Local authority Industry: No Occupation: Social class via NS-SEC Education: Yes Health: Yes, for HD, WH Ethnicity: Yes Disability: Yes, for HD, WH Attitudinal vars: On spending habits/preferences Other personal: Extensive detail on financial situation/constraints Household vars: Assets and liabilities, size of HH and relationship Other
Linkage:	• Not linkable
Data quality:	• Not reviewed yet
Strengths:	• Extensive detail on family finances
Weaknesses:	• Not possible to directly identify minimum-wage households
Relevance for low pay analysis:	• Potential for studying low-income households – may be possible to make inference for households with low income in employment
Current usage:	• Used for studying family incomes constraints, particularly over the recession
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: DWP-room at UKDA only, since 2011 *Service request*: Not available
Aggregate stats:	• Statistics on household income, including welfare and other receipts
Other:	•

3.5 Living Costs and Food Survey

Common acronym:	• LCF
Available period:	• From 2001
History:	• The Family Expenditure Survey (FES) and National Food Survey (NFS) ran from 1957-2001 • These were combined in the Expenditure and Food Survey (EFS) in 2001 • In 2008 the EFS was replaced by the LCF, renamed as part of the abortive Integrated Household Survey project
Type:	• Cross section, household and individual
Sample:	• Interview (household and individual) and expenditure diaries • 5,000 households per year • Multi-stage stratified sample to generate population-weighted output
Variables:	Earnings: Earnings plus other income (individual and household; may be banded) Hours: No Employment status: Labour market participation Geography: GOR at present Industry: No Occupation: NS-SEC Education: No Health: No Ethnicity: Yes Disability: Can be inferred Attitudinal vars: No Other personal: Assets Household vars: Household income Other: ONS groups households into 'types' to simplify analysis; see Fry and Ritchie (2011) for example
Linkage:	• Not obviously feasible
Data quality:	• Data has been little used by microdata researchers (ESDS only lists six pieces of secondary research) and not in the VML, so there is little feedback to date
Strengths:	• Level of detail
Weaknesses:	• Relatively small number of households • No longitudinal component • No wage measures, only earnings, so difficult to identify the low-wage
Relevance for low pay analysis:	• Potential to study changing expenditure patterns of low-income households
Current usage:	• Negligible through VML; small number of users at UK Data Service
Availability:	*Internet*: Some teaching datasets with unrestricted access *End User Licence*: Yes, GOR only *Special Licence*: Not available *Secure facility*: SDS, VML: available, possibly down to postcode level *Service request*: Not available
Aggregate stats:	• Weighting for price indexes • Statistics on changing consumption, particularly food
Other:	•

3.6 Monthly Business Survey

Common acronym:	• MBS
Available period:	• Available 1997-2009 as MPI and MIDSS, 2010 onwards as MBS • Available approximately six months after the survey date
History:	• Collected as Monthly Production Inquiry (MPI) and Monthly Inquiry into the Distribution and Services Sector (MIDDS) 1997-2009 • Combined as the MBS from Jan 2010
Type:	• Rotating panel of businesses
Sample:	• Approximately 30,000 firms per month • Stratified by size and industry • Small firms are kept in the panel for no more than 15 months
Variables:	Earnings: None Hours: None Employment status: Numbers of staff Geography: Postcode Industry: Five-digit SIC Other: Limited financial variables enabling GDP calculations
Linkage:	• Linkable to other ONS datasets
Data quality:	• Believed to be high-quality on responses; see ONS (2012f) • However, when the Bank analysed MBS data versus MPI/MIDDS data there was a clear shift in the time series which it was not possible to account for; the ONS team suggested this was the result of different weights. It may also have been because the stratification bands changed when MBS was introduced. At all events, from a micro perspective there is an unresolved break in the data between MPI/MIDSS and MBS • Non-response data is imputed in MBS • Six-month delay before releasing data to researchers is ONS advice: this is enough time for most of the data to have come in and most of the queries to have been resolved. Note that some response can come in up to eighteen months late
Strengths:	• Frequency, sample size, timeliness in delivery to researchers
Weaknesses:	• Limited information
Relevance for low pay analysis:	• Potential to study demographic impacts on firms in more detail than is possible with BSD/ARD
Current usage:	• Bank of England only major user of microdata currently, for short-term modelling of economy
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility (SDS, VML)*: All data *Secure facility (Europe)*: Not available *Service request*: Not available
Aggregate stats:	• Quarterly GDP
Other:	• The 'Monthly Business Survey (Construction)' and 'Monthly Business Survey (Retail Sales)' are not part of the MBS; the rebranding is preparatory work

3.7 Monthly Wages and Salary Survey

Common acronym:	• MWSS
Available period:	• From 1999 • Data available approximately six months after survey date
Type:	• Cross-section but the stratified sampling means a longitudinal component for large businesses
Sample:	• 9,000 firms surveyed monthly • Stratified by size and industry
Variables:	Earnings: Company total, with bonuses and arrears separately identified; split by weekly-paid and salaried workers Hours: No Employment status: Number of employees, weekly- and monthly-paid Geography: Postcode Industry: Five-digit SIC at enterprise level Occupation: No Education: No Health: No Ethnicity: No Disability: No Attitudinal vars: No Other personal: No Household vars: No Other: Text information explaining particularly large movements in earnings; not currently retained by ONS
Linkage:	• Linkable to other ONS datasets
Data quality:	• Little used by microdata researchers so not much known • However, ONS is aware that, like other monthly data, there can be large amount of volatility, particularly as a result of bonus payments; see ONS (2005a)
Strengths:	• Frequency and timeliness
Weaknesses:	• Only average values available for each company • Unweighted microdata appears very volatile
Relevance for low pay analysis:	• Potential to see how firms react to introduction of the NMW in October; however, exploratory work for Fry and Ritchie (2013) found no relationship to low pay • Aggregate stats used in LPC analysis
Current usage:	• Very few, if any, current users
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: All data, VML and SDS *Service request*: Not available
Aggregate stats:	• Average Earnings Index/Average Earnings Ratio, Average Weekly Earnings
Other:	•

3.8 ONS Longitudinal Study

Common acronym:	• ONS-LS, ONSLS ,or just 'LS' in context
Available period:	• 1971, 1981, 1991 and 2001 Census; 2011 Census in 2014
History:	• The LS is created by linking a 1% sample of Census data to GP registrations, death data and cancer statistics; this allows inter-Census analysis. It has been running since the late 90s. 2001 data has been added. 2011 data is expected in 2014 • NI and Scotland, which both manage their own Census contributions, now also run longitudinal studies. The NILS covers approximately 28% of the NI population; the Scottish LS 5% of the Scottish population. The Scottish study in particular is being extensively linked to administrative records eg education.
Type:	• Longitudinal, individual
Sample:	• Census samples, augmented by administrative records • 1% random sample from each Census; England and Wales only
Variables:	Earnings: Imputed wage was created for the 2001 data from ASHE Hours: No Employment status: At the time of Census Geography: 'Super Output Area' Census time and in between Industry: None Occupation: Social class via NS-SEC Education: Yes Health: Yes, including date of death Ethnicity: Yes Disability: Yes Attitudinal vars: No Other personal: Age, gender Household vars: No Other
Linkage:	• Links to NI-LS (10% sample) and Scottish LS being investigated by ONS and others
Data quality:	• Generally same high-quality Census data and administrative data also of high accuracy • Migration between Censuses is problematic – very few out-migrants de-register from GPs, and internal migrants do not necessarily re-register until they need a doctor; these are more likely to be male, young
Strengths:	• Large numbers with potentially forty years of data in England and Wales • Good for identifying patterns of migration and observational health studies • For NI, very large proportion of coverage • For Scotland, linkage to a range of administrative data sources
Weaknesses:	• Limited range of variables between Censuses
Relevance for low pay analysis:	• Very large numbers make investigation of eg specific migrant groups practical, but the difficulty of identifying low-wage workers (no income data collected) means this is primarily useful for triangulation studies.
Current usage:	• Limited use due to access requirements • Use peaks after new Census data is added • Primary usage is amongst health researchers
Availability:	*Internet*: Not available

	End User Licence	Not available
	Special Licence	Not available
	Secure facility	VML only (ONS-LS); NISRA offices (NILS); GROS (Scottish LS)
	Service request	Free to academics; chargeable to governments
Aggregate stats:	• Used to produce ONS healthy life expectancy data	
Other:	• See (EW) http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/index.html (Scotland) http://sls.lscs.ac.uk/ (NI) http://www.qub.ac.uk/research-centres/NILSResearchSupportUnit/	

3.9 ProdCom / Producer Price Survey / Export and Import Price Surveys / Services Producer Prices

Common acronym:	• ProdCom • MIINPP (Monthly Inquiry for Index Numbers of Producer Prices) • MIINEP (Monthly Inquiry for Index Numbers of Export Prices) • MIINIP (Monthly Inquiry for Index Numbers of Import Prices) • QISPPI (Quarterly Inquiry for Services Producer Prices Index)
Available period:	• Prodcom: in theory from 1997 but only odd years have been used by researchers and so the data have not yet been systematically prepared for research use • PPI was made available in the VML for a test case but was not followed up; no other price data was requested; hence no data available at present due to lack of use • All data are accessible under the Statistics Act, but ONS may not hold back-series of the microdata in an easily retrievable form
History:	• Core Eurostat-mandated data collection subject to regular revision
Type:	• Price data; all have a longitudinal element by virtue of stratification
Sample:	• ProdCom: stratified sample from IDBR (21,500 annually; quarterly prior to 2005) • MNIIPP: sub sample of ProdCom providing prices only (4,000 monthly) • MIINEP and MIINIP: sample taken from HMRC exporter/importer records; updated annually (E:2,500 monthly I:2,600 monthly) • QISPPI: sample taken from IDBR; but several industries are covered by other bodies (eg Bank provides insurance industry SPPI) (2,200 quarterly)
Variables:	Geography: Postcode Industry: Five-digit SIC Other: Detailed product sales volume and prices
Linkage:	• To other ONS business data (but organised by product type, not industry)
Data quality:	• Data has been little used by microdata researchers and so there is little information on data quality from research users • ONS official quality report http://www.ons.gov.uk/ons/guide-method/method-quality/specific/business-and-energy/manufacturing/quality-and-methods/index.html • It's not clear how the export/import firms are chosen
Strengths:	• ProdCom has detailed product range
Weaknesses:	• Low numbers in services; likely to be problematic for analysis of low-wage business service companies
Relevance for low pay analysis:	• Potential for LPC to investigate whether NMW affects factory gate prices rather than retail prices, and where NMW increases are absorbed into the supply chain (would need to be linked to input-output tables) • But products not necessarily linked to industries (particularly 'low-wage')
Current usage:	• Very little by researchers, largely because researchers are more interested in companies than products
Availability:	*Internet*: Not available *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: VML/SDS - some ProdCOM available *Service request*: Not available
Aggregate stats:	• Producer Prices Index, Services Producer Price Index • Various National Accounts statistics, including supply-Use tables

3.10 Retail prices

Common acronym:	• No single acronym as data are multi-sourced
Available period:	• RPI indices – 1987 to 2013 • RPI weights – 1947 to 2013
History:	• Traditionally the data was collected by shoppers searching for a fixed basket of goods • Since the advent of the internet, more price data is collected centrally by ONS from the internet (typically large-ticket items such as holidays or white goods)
Type:	• Cross-section, price data, with clustering provided by shop IDs
Sample:	• Around 90,000 prices quotes collected monthly by shoppers; additional numbers collected centrally eg from internet
Variables:	Geography: Postcode Other: Price quotes
Linkage:	• No

Data quality:	• Data has been little used by microdata researchers and so there is little information on data quality from research users
	• Conceptually, users have raised concerns that
	  ○ the monthly collection round misses out short term effects
	  ○ the annually fixed weights don't allow for response to special offers such as BOGOF

Strengths:	• Detailed data available from the web • Large number of price quotes
Weaknesses:	• Complex multi-layered structure • Little experience in the research community
Relevance for low pay analysis:	• Potential to study how NMW wages raise retail prices • If 'low-wage families' could be identified in the Living Costs and Food Survey (which generates the RPI/CPI weights), it would in theory be able to produce a 'low wage CPI' measure
Current usage:	• Minimal – mostly for studying policy interventions such as the impact of alcohol prices
Availability:	*Internet*: Since 2011, at GOR level: http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/cpi-and-rpi-item-indices-and-price-quotes/index.html *End User Licence*: Not available *Special Licence*: Not available *Secure facility*: Available in detail but not cleaned (as of end 2012) *Service request*: Not available
Aggregate stats:	• CPI, RPI, Harmonised Index of Consumer Prices
Other:	•

3.11 Workplace Employment Relations Survey

Common acronym:	• WERS
Available period:	• 1980, 1984, 1990, 1998, 2004, 2011 • Data available approximately 24 months after survey date
History:	• Known as WIRS (Workplace Industrial Relations Survey) prior to 1998, when it underwent a major redesign into its current form • Microbusinesses were included in the 2004 survey but because of the cost were dropped from the 2011 survey • For the 2004 survey the WERS Information and Advisory Service (WIAS) provided advice and some service requests, with BIS and Acas providing support/info to other government departments and the business community, as covered by their policy mandates.
Type:	• Multi-level dataset: establishment data, plus questionnaires to managers, employee representatives and employees • Cross-section selection of establishments but with panel subset
Sample:	• 2,700 establishments in 2011 (of which 1,000 in panel from 2004) • Apart from panel, chosen as random sample from IDBR
Variables:	Earnings: Earnings (on employee questionnaire); Pay determination and systems (on management questionnaire); Impact of NMW on wage-setting Geography: Postcode Industry: Five-digit SIC Attitudinal vars: Extensive on perceptions of workplace, from 2004 onwards eg impressions of managers Other personal: None Household vars: None Other: Extensive on organisational systems eg flexible working, workplace representation etc
Linkage:	• Linked to ONS business data through IDBR reference (including ASHE)
Data quality:	• Unknown yet for 2011 but previous versions had reputation for high-quality data with the exception of the 1998 Financial Performance data • Subjective questions on financial performance ("are you more productive than competitors in the same industry") have been highlighted as problematic, particularly for the (self-evaluated) best and worst • 2004 Financial Performance data validated against ABI where possible
Strengths:	• Multi-level design, and attitudinal questions relating to conditions in the workplace
Weaknesses:	• Relatively small sample • Establishment level, so options to link with other data sources (mostly collected at the enterprise level) limited • Very small numbers agreed to their data being linked in the 1998 survey • No very small companies in 2011 data
Relevance for low pay analysis:	• Can be used to enhance ASHE with workplace characteristics • Can be used to study workplace characteristics in the 'low-wage industries' • Direct information on impact of NMW on wage policy
Current usage:	• 1998, 2004 increasingly popular with academic researchers; 2011 expected to follow trend • Reputation for being a 'difficult' dataset has limited use, and increased role of 'knowledge brokers', particularly in government; see Drew et al (2013)
Availability:	*Internet*: Not available

	End User Licence	1998 onwards
	Special Licence	Not available
	Secure facility (SDS, VML)	All data, linked to ONS data
	Secure facility (Europe)	Not available
	Service request	Currently only for government researchers
Aggregate stats:	• Primary source are WERS books: Kersley et al (2006) for WERS 2004 and van Wanrooy et al (2013) for WERS 2011, at present	
Other:	•	

3.12 Apprentice Pay Survey (ApPS)

BIS' Apprentice Pay Survey (2011-) is an important new source for LPC data. It contains a large amount of detail on the wages, hours, working conditions and training of apprentices, as reported by apprentices themselves. The 2011 survey is available for all the home nations (England, 6,140 respondents; Wales, 1,997; Scotland, 2041; NI, 842). The 2012 survey (England, 6,597 respondents; Wales, 1,817; NI, 640) omitted Scotland. Sampling rates were 40%-50% in Great Britain but due to the alternative data collection method in NI, response rates were 6%-7%.

As a relatively new dataset, experience with it is limited, particularly in the non-government research community. Higton et al (2012) produced detailed descriptive analyses of the 2011 data, noting the high rate of non-compliance and the importance of distinguishing between apprenticeship 'frameworks' and between experiences in the home nations. Although the 2011 survey covered all of the UK, they suggested that NI data might be of limited value given the small numbers and concerns over samples. Behling and Speckesser (2013) also studied the 2011 ApPS, but noted that the LFS or other datasets needed to be used for studying employment choices as this ,as the ApPS by design only covers those who have chosen to take up apprenticeships.

Higton (2013) reviewed the 2012 data, and considered that there was little statistical difference between that and findings from the 2011 (adjusted for the loss of Scottish data), despite the passage of 18 months between data collection points. Higton (2013) argues that this is because the policy changes in between survey periods had insufficient time to take effect, although it could also be that the dominant characteristics of the dataset (high non-compliance and framework effects) are not affected by the policy changes. For example, Drew et al (2014, forthcoming) suggest that the survey design is a key driver of the non-compliance findings.

As the ApPS is the subject of a separate review in the LPC's research program for the 2015 report (interim findings in Drew, Ritchie and Veliziotis, 2014, forthcoming), a more detailed review has been omitted from this report. Past reviews and future studies, when available, are all published on research pages of the LPC's website https://www.gov.uk/government/organisations/low-pay-commission.

3.13 Other data sources

Only the above data sources have been reviewed for this report. Other potential sources not considered include:

- HMRC profits and self-assessment data (HMRC)
- Job Seekers Allowance Data (DWP)
- The Life Opportunities Survey (DWP)
- The Work and Pensions Longitudinal Study (DWP)
- The Profits Survey (ONS)
- The Short-term Employment Survey (ONS)
- The Vacancy Survey (ONS)
- The Wealth and Assets Survey (ONS/DWP)

4. Aggregate statistics

4.1 Bankruptcy:

Description:	• The Insolvency service provides the official numbers of corporate and individual insolvencies in England and Wales • Publications include Insolvency Statistics, Individual Insolvencies by Region and Voluntary Arrangement Outcome Statistics
Source data:	• Derived from administrative sources (e.g. Accountant in Bankruptcy, Companies House, Insolvency Service) • ONS population estimates used for weighting
Type:	• Company and individual insolvency statistics are produced quarterly • Historic data are available back to 1960 in some cases • Case level data for individual insolvencies for England and Wales are available regionally and on an annual basis from 2000 to 2011
Accuracy, revisions, timeliness:	• Users consulted and statistics reviewed on an on-going basis • Data is seasonally adjusted • Insolvency statistics are published on the first Friday of the second month following the end of the quarter being reported • Revisions are made when inaccuracies in proceeds and systems have been identified and changes to classifications required
Breakdowns:	• Company insolvency and bankruptcy orders relating to the self-employed in England and Wales are presented by industry (SIC 2003) • Individual insolvencies are presented by region and local authority level • 2008-2011 tables are available by region, age and gender
Cautions:	• The introduction of debt relief orders in April 2009 has potentially created a break in the series in terms of levels and seasonality - there is currently not a long enough time-series to quantify the effect to seasonally • The data is not available on a SIC2007 basis • Statistics for insolvencies in Scotland and Northern Ireland are published but have some differences in definition • Prior to 20011 Q2 the data are based on the date the insolvency procedure was registered on the administrative recording system - after this date, in England and Wales, new cases are reported using the date of the court order
Use by LPC:	• Economic context
Availability:	• wwwinsolvencydirectbisgovuk/otherinformation/statistics/historicdata/HDmenuhtm • wwwonsgovuk/ons/taxonomy/indexhtml?nscl=Personal+Debt+and+Insolvencies
Special analyses:	• Individual Insolvencies Including Bankruptcies, England and Wales wwwonsgovuk/ons/rel/regional-trends/area-based-analysis/individual-insolvencies-including-bankruptcies--england-and-wales--2001-11/articlehtml • Characteristics of Individual Insolvencies Including Bankruptcies, England and Wales – www.ons.gov.uk/ons/dcp171766_279344pdf

4.2 Gross Domestic Product (GDP)

Description:	• GDP provides a measure of total economic activity in the UK • Gross value added (GVA) is equivalent to GDP plus taxes, less subsidies on products
Source data:	• Main sources are ONS Business Surveys • Monthly Business Surveys for quarterly data • Annual Business Inquiry for annual data • Supplemented by industry specific sources (e.g. ABI supplies insurance data)
Type:	• Quarterly and annual time-series

Accuracy, revisions, timeliness:	• See ONS (2012c) for current quality report
	• Quarterly data ±0.2% - quarterly data revised twice for each quarter and again for annual data. Example: if months are numbered M1-M12 then:
	  • M4 – first estimates for M1-M3 (output data only)
	  • M5 – second estimates for M1-M3 (output led but balanced to expenditure and income data)
	  • M6 – final quarterly estimates for M1-M3 (output led but balanced to expenditure and income data)
	  • M24 (approximately) revised to conform to annual data (expenditure led balanced through input output supply use framework)
	• See ONS(2012a) and (2012i) for more information on revisions and balancing

Breakdowns:	• Annual: by region, by industry and company size
Cautions:	• Annual regional breakdowns are available after 12 months with industry data available after 24 months. Regional breakdowns are based on number of employees and industry – data suffers for the 'commuting and head office' problem. • Quarterly data more affected by late returns by large companies • GDP estimates based on output, expenditure and income differ before being balanced
Use by LPC:	• Provides overall economic context for LPC decisions • Used to study performance of 'low-wage' industries
Availability:	• www.statistics.gov.uk
Special analyses:	• Some additional statistics produced as part of the annual production round but not released • Fully weighted analyses consistent with national data can be produced for a fee

4.3 Labour Market Outcomes

<table>
<tr><td>Description:</td><td>

Labour Market Statistics reflect the different aspects of the labour market and cover:

employment
unemployment and claimant count
economic inactivity
redundancies
earnings
jobs
vacancies
labour productivity
labour disputes

See ONS (2012b) for further detail

</td></tr>
<tr><td>Source data:</td><td>

Variety of administration and survey sources (mainly LFS and business surveys)
Every effort is made to ensure each series is internally consistent and directly comparable from start to finish
The claimant count is subject to distortions as a result to changes to the benefit rules

</td></tr>
<tr><td>Type:</td><td>

Monthly, quarterly and annually

</td></tr>
<tr><td>Accuracy, revisions, timeliness:</td><td>

See ONS (2011b) for current quality report
Most of the labour market estimates published by ONS are based on statistical samples and, as such, are subject to sampling variability
For each of the key series, 95 per cent confidence intervals are routinely published
The major component of labour market statistics is the LFS – please refer to the section on LFS for a review of its characteristics
The following list shows the time between publication from the end of the period for each of the main labour market estimates

<table>
<tr><td>LFS</td><td>approximately 44 days</td></tr>
<tr><td>workforce jobs</td><td>approximately 74 days</td></tr>
<tr><td>Ccaimant count</td><td>34 days</td></tr>
<tr><td>earnings</td><td>approximately 44 days</td></tr>
<tr><td>vacancies</td><td>approximately 40 days</td></tr>
<tr><td>labour disputes</td><td>approximately 44 days</td></tr>
</table>

Labour market statistics are published every month and include tables, text and charts
A fuller breakdown, including at the local level, is available through the NOMIS website www.nomisweb.co.uk
Coherence - a number of the series would, at first glance, seem to measure the same phenomena (claimant count and unemployment rate, workforce jobs and employment). There are, however, fundamental differences in the methodologies and therefore sometimes report contrasting results – see ONS (2012b) for more details.
All the main UK labour market series are seasonally adjusted
For the countries and regions of the UK only the main aggregates are seasonally adjusted
For smaller areas only non-seasonally adjusted data are available

</td></tr>
<tr><td>Breakdowns:</td><td>

The breakdown varies between each release
Generally data are published for the UK as a whole, and for each country and region of GB

</td></tr>
</table>

Cautions:	• Data comes from a variety of sources with differing start dates
Use by LPC:	• Short-term analysis on earnings and workers – variables include employment, unemployment, inactivity, self-employment, government training, apprentices, participation in full-time education, hours of work and redundancies • Where possible the data are disaggregated by gender, age, ethnicity, disability, region, industry, occupation, education, marital status and family background
Availability:	• http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Labour+Market
Special analyses:	• Labour market releases by ONS can be accessed via http://www.ons.gov.uk/ons/publications/all-releases.html?definition=tcm%3A77-21589

4.4 Prices (consumer): Retail Price Index (RPI)/ Consumer Price Index (CPI)

Description:	• RPI and CPI are currently ONS's preferred measures of consumer price inflation • The CPI is the main measure used for macroeconomic purposes of consumer price inflation • RPI is the most long-standing general measure of inflation and is used for a variety of purposes including the uprating of index-linked gilts • ONS have recently introduced two new measures for consumer price inflation – CPIH and RPIJ. CPIH extends the CPI to include a measure of owner occupiers' housing cost and RPIJ addresses a flaw in the RPI formula • The RPIJ is currently published as an experimental statistic and will remain so until it is assessed by the UK Statistics Authority in the summer of 2013. For more information on the developments to the indices see ONS (2013a, 2013b)
Source data:	• RPI/CPI - sample survey of more than 650 goods and services for which price movements are regularly monitored in approximately 20,000 outlets and 150 areas throughout the UK • There are three collection methods • Local collection – 20,000 shops in around 150 locations • Central shops – for some large chain stores there is a central pricing policy • Central collection – goods and services are the same for all UK residents or regional variation can be collected centrally (e.g. internet sales and utility costs) • Data used for the weights come from a variety of data sources (e.g. National Accounts, Living Cost and Food Survey, Annual Business Survey)
Type:	• CPI and RPI are produced monthly • For most products (excluding petrol and diesel, and car and house insurance premiums which take the monthly average price) the data refer to a specific point in time, usually the second or third Tuesday of every month • RPI series starts in 1947 • CPI series starts in 1996
Accuracy, revisions, timeliness:	• See ONS (2011c) for the current quality report • Annually chained linked series • Goods and services measured, outlets and locations are generally held constant within a single year, although they can be changed between years to be kept up to date and reflect changes in consumer taste • ONS use a variety of sophisticated approaches to select the sample in terms of location, outlet, selection of products and selection of items – further detail is available in ONS (2011c) • In the case of the RPI, once indices are published they are never revised - corrections are shown but future series continue to give the original figure with the correction attached • Since 1947 the RPI has been re-referenced on five occasions • CPI indices are revisable although past revisions have been minimal except of those arising from re-referencing • Coverage of CPI has been extended in stages (eg goods and services and population base) • Both RPI and CPI use various techniques to treat for issues of quality change: • Direct comparison • Direct quality adjustment • Imputation • Coherence – both measure the average change of consumer prices from month to month but differ in terms of coverage and methodology – for more information please refer to ONS (2003)
Breakdowns:	• Average change of month to month consumer prices in total and contribution by type of good and services

Cautions:	• RPI methodological short-comings mean that it is now being reassessed by the UK Statistics Authority to determine whether or not it merits continued designation as a National Statistic
Use by LPC:	• Important measure of how prices are changing for consumers
Availability:	• http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Price+Indices+and+Inflation
Special analyses:	• Price indices and inflation research conducted by ONS can be accessed via http://www.ons.gov.uk/ons/taxonomy/search/index.html?newquery=*&nscl=Price+Indices+and+Inflation&nscl-orig=Price+Indices+and+Inflation&content-type=publicationContentTypes&sortDirection=DESCENDING&sortBy=pubdate

4.5 Prices (industry): Producer Price Index (PPI)/ Services Producer Price Index (SPPI)

Description:	• The PPI for UK manufactures and SPPI for service providers is used to monitor inflation within business sector transactions • They are different from the RPI/CPI as they measure price changes between the business and household sector • PPI is constructed from a number of statutory monthly surveys of UK manufacturers - PPI has been in existence for over 100 years • SPPI captures the quarter on quarter change in the price received for services provided by businesses in GB - prices are provided for a limited range of service industries and at an aggregate level • SPPI was accredited in 2010 as a National Statistic

Source data:	• PPI is primarily based on the following surveys:
	  • Domestic PPI
	  • Export Price Indices
	  • Import Price Indices
	• SPPI is based on a quarterly price survey underpinned by two supplementary surveys (SPPI Recruitment Survey and SPPI Turnover Survey) and by administration data

Type:	• PPI is published monthly and leads with four PPI series:
	  • Gross Sector Output
	  • Net Sector Output (which excludes within-sector transactions)
	  • Gross Sector Input
	  • Net Sector Input
	• SPPI are published quarterly and lead with two SPPI series:
	  • Net Sector Output
	  • Gross Sector Output

Accuracy, revisions, timeliness:	• See ONS (2012g, 2012h) for the current quality reports for PPI and SPPI respectively **PPI** • Sample size of 4,000 domestic manufacturers providing 6,750 price quotes; 1,900 manufacturing exporters providing 3,800 price quotes; and 1,500 importers providing 2,500 price quotations • The Statistical Bulletin is published two weeks after the reference month • The price index numbers for current cost accounting are published one month after the reference month • PPI is subject to sampling error, non-sampling error, substitution bias (change of consumer behaviour) and quality change bias • Subject to revisions of the preceding two quarters (mainly due to late and revised responses) and revisions to seasonal adjustment factors **SPI** • Survey comprises 2,700 businesses and 6,500 price quotes per quarter • The sample covers GB and is selected from the IDBR • Since 2009 ONS has been working to improve the industrial coverage of the SPPI to meet user needs • Published approximately eight weeks after the end of the quarter to which it relates • Not seasonally adjusted • Subject to sampling error, non-sampling error, substitution bias (change of consumer behaviour) and quality change bias • Subject to revisions of the preceding two quarters (mainly due to late and revised responses) • Weights are updated every five years
Breakdowns:	• Sector and sub-sector breakdowns • Time-series available

Cautions:	• ONS are currently working to improve the industrial coverage of the SPPI • ONS currently produce a number of indices for new industries which are published as experimental statistics and sit alongside the regular SPPI bulletin
Use by LPC:	• Important measure of price changes of business transactions
Availability:	• http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Price+Indices+and+Inflation
Special analyses:	• Price indices and inflation research conducted by ONS can be accessed via http://www.ons.gov.uk/ons/taxonomy/search/index.html?newquery=*&nscl=Price+Indices+and+Inflation&nscl-orig=Price+Indices+and+Inflation&content-type=publicationContentTypes&sortDirection=DESCENDING&sortBy=pubdate

4.6 Productivity

Description:	• Labour productivity measures the amount of real economic output that is produced by a unit of labour input
	• The labour input can be measured in terms of:
	  • workers
	  • jobs
	  • hours worked
	• For a comprehensive overview of the productivity framework and methodology see ONS (2007b)

Source data:	• Quarterly National Accounts, for GVA • LFS micro dataset, for workers, filled jobs and hours worked • Workforce Jobs (WJF) and Short-Term Employers Survey (STES) for industrial allocation
Type:	• Quarterly. Regional estimates are produced annually
Accuracy, revisions, timeliness:	• See ONS (2012e) for current quality report • Quality of the statistics reflects the quality of the source data • Labour productivity industrial estimates include some service sector industry estimates which are on an experimental basis • Revisions are made in line with revisions to the output and labour input data sets • Quarterly data published three months after the end of the quarter • Annual data published 12 months after the reference period
Breakdowns:	• Whole economy unit labour costs index • Published at industrial level (SIC2007) for a number of sectors and sub-sectors • Regional productivity estimates in nominal terms and presented as indices where UK=100 • Market sector workers are calculated as LFS workers less public sector workers • Market sector hours are then calculated by multiplying market sector jobs by average hours from the LFS • These labour input series are used to calculate output per worker and output per hour using GVA for the market sector • Labour input series are seasonally adjusted before use in productivity estimates • Experimental data published includes market sector productivity and output per job and output per hour series for certain sub-sectors of services
Cautions:	• Quarter on quarter changes should be treated with extreme caution • There is difficulty in ensuring consistency of seasonal adjustment between the numerator and denominator • ONS use dual methodologies for estimating employee jobs by industry and estimating average hours by industry which leads to differences between published series
Use by LPC:	• Economic context
Availability:	• www.statistics.gov.uk
Special analyses:	• The underlying data for the labour productivity statistics are available to users on request

4.7 Profitability:

Description:	• Profitability compares, either at a company or a sector level, the economic gain or profit with the capital used to produce it - it is known as the 'rate of return' • ONS present profitability data on a gross and net basis • Gross and net in this context means before and after capital consumption (depreciation) • ONS is the only UK source collecting survey data on operating profits via the Quarterly Operating Profits Survey (QOPS) • ONS's quarterly Profitability of UK Companies Statistical Bulletin leads with net rate of return • Net estimates deduct accumulated capital consumption from profits and accumulated capital consumption from asset estimates • The UK's National Accounts publish profitability information on a gross basis (i.e. Gross Operating Surplus of UK Private Non-Financial Corporations, PNFCs) • Gross operating surplus of PNFCs consists of gross trading profits, plus income from rental of buildings, less inventory holding gains
Source data:	• QOPS • HMRC company profits data • Quarterly Capital Expenditure Survey
Type:	• Quarterly and annual time series
Accuracy, revisions, timeliness:	• Rates of return data for PNFC are revised due to benchmarking of the profits data from the QOPS to more definitive data from HMRC • Accuracy of the data are estimated to be high - profits data are benchmarked six monthly and the capital consumption and stocks data are now on a SIC 2007 basis • Standard errors for quarterly profits are currently being developed and will be published by the end of 2013 • Estimates and revision are consistent with Quarterly National Accounts **QOPS** • Sample of 1,650 businesses • A stratified random sample is created from the IDBR at the enterprise group level • There is no imputation in the survey • The survey is reviewed every three years • The data are published five to six weeks after the end of the quarter
Breakdowns:	• Aggregate information provided for PNFCs • Sector information given for manufacturing, services and continental shelf companies
Cautions:	• QOPS was reviewed and redesigned in 2003 • The number of matched pairs respondents in the first quarter of each year reduces due to some respondents dropping off the sample causing, on average, about 20% fewer matched pairs in this quarter than in remaining quarters • A new sample selection on a SIC2007 basis was created for first quarter, 2011
Use by LPC:	• Economic context
Availability:	• wwwonsgovuk/ons/rel/pnfc2/profitability-of-uk-companies/indexhtml
Special analyses:	• HM Treasury use profitability results for forecasting, analytical and briefing work on the economy wide output

4.8 Short-term Earnings Statistics: Average Weekly Earnings (AWE)

Description:	• Estimates of earnings and pay generally cover three main areas:
	  • basic pay
	  • overtime
	  • bonuses
	• Earnings statistics can be classified in to two categories:
	  • structural statistics
	  • short-term statistics
	• Structural statistics are more detailed and used to analyse trends in earnings over long periods - ASHE (prior to this the NES) is ONS's recommended source of structural employees pay level
	• Short-term statistics provide more up-to-date measures of earnings and earnings growth - the following section will focus on AWE
	• See ONS (2012b) for further detail on Labour Market Statistics

Source data:	• Based on the Monthly Wages and Salary Survey (MWSS) • AWE replaced the Average Earnings Index (AEI) as ONS's lead measure of short-term earnings growth in January 2010
Type:	• Monthly

Accuracy, revisions, timeliness:	• See ONS (2011a) for current quality report
	• AWE accuracy is subject to sampling error (±- 0.5 percentage points excluding bonuses), unit response error and changes in the classification of individual or groups of businesses- see MWSS for more information
	• Published six-seven weeks after the end of the month
	• Unadjusted estimates are revised and finalised 10 – 11 weeks after the end of the month
	• Seasonally adjusted estimates are revised in line with the unadjusted estimates, with two additional considerations:
	  • The last five months seasonally adjusted back data are revised
	  • Seasonal adjustment parameters are revised annually and can lead to revisions of the whole time series

Breakdowns:	• AWE uses current employment weights (earnings growth between months is adjusted for changes in the composition of the workforce) and provides estimates in pounds • The two methodological improvements identified above differentiate the AWE from the AEI • Information is provided on the whole economy and by sector and industry • Separate estimates are produced for wage and employment contributions to AWE growth

Cautions:	• AWE only gives earnings per employee. For example if the number of paid hours worked increases (assuming there is no change to the pay rate) the AWE will record an increase in average earnings.
	• In October 2010 AWE moved to SIC2007 from previous SIC2003 classification
	• All time series has been reweighted on the SIC2007 basis, however, this may have caused increased sample errors between August 2010 and July 2011
	• AWE is only available from 2000
	• Changes in classification have led to apparent discontinuities in AWE - two notable examples are:
	  • banks that received Government financial report and were reclassified to the public sector in July 2009 - this led to higher average rates of public sector pay and lower rates of private sector pay
	  • sixth form and further education colleges were reclassified to the public sector in June 2010 - this caused a fall in the average rates of public sector pay but had little impact on the private sector
	• MWSS does not sample firms with less than 20 employees - AWE uses information from ASHE to impute values for employee earnings from small

	firms • AWE mean earnings tend to be lower than the equivalent ASHE figures
Use by LPC:	• AWE is used to compare results between short-term indicators and also with long-term structural indicators
Availability:	• http://www.ons.gov.uk/ons/rel/awe/average-weekly-earnings/index.html
Special analyses:	• Earnings research conducted by ONS can be accessed via http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Earnings

4.9 Short-term Earnings Statistics: Index of Labour Cost per Hour (ILCH)

Description:	• Short-term statistics provide more up-to-date measures of earnings and earnings growth than structural statistics - the following section will focus on ILCH • See ONS (2005b) for further detail
Source data:	• ILCH is an experimental statistic provided by ONS to meet EU regulation • It is mainly drawn from MWSS • Hours worked is drawn from the LFS
Type:	• Quarterly
Accuracy, revisions, timeliness:	• See ONS (2012d) for current quality report • Users have commented on the volatility of the ILCH, particularly for the lowest industry aggregates • ASHE provide ILCH estimates of pay for employees of small business and for Northern Ireland • Confidence intervals are not published as it is constructed from a number of data sources • ILCH data on wages and salaries are subject to revision - late and revised responses are updated the following quarter and methodological changes are applied to keep the statistics in-line with the underlying MWSS data • Seasonal adjustment parameters are updated annually and can lead to revisions to the entire time-series • The ILCH moved to SIC2007 in October 2010 • The index is published approximately ten weeks after the end of the quarter
Breakdowns:	• ILCH is a wider measure than AWE as it includes non-wage costs such as NI and pension contributions, sickness, maternity and paternity payments, and additional benefits in kind
Cautions:	• ILCH gives estimates of per hour actually worked, however they are still experimental statistics and subject to further development
Use by LPC:	• As ILCH are still experimental and in development, historically LPC have placed limited emphasis on them
Availability:	• www.statistics.gov.uk
Special analyses:	• Earnings research conducted by ONS can be accessed via http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Earnings

4.10 Work Force Jobs (WFJ)

Description:	• The number of people with jobs is measured by the LFS although it can also provide an estimate of the number of jobs by adding main and second jobs together • ONS's preferred measure of short-term job change, however, is WFJ as it directly provides estimates of the number of jobs in the economy • See ONS (2012b) for further detail
Source data:	• WFJ is a compound measure made up of various sources including employer surveys (STES and ABI), LFS and various administrative sources
Type:	• WFJ is produced on a quarterly basis and are published in both the national and regional labour market statistical bulletins

Accuracy, revisions, timeliness:	• See ONS (2013d) for current quality report
	• A fundamental redevelopment was undertaken in 2010; main changes were:
	  • conversion to SIC2007
	  • redesign and new estimation methods for STES
	  • benchmarking of employee Jobs to the Annual Business Inquiry and the removal of discontinuities from previous benchmarking exercises
	  • revisions to inputs including public sector employment and LFS
	  • review of seasonal adjustment
	  • changes to WFJ publication tables
	• See Barford (2010) for further information
	• The employee jobs components of WFJ are derived from 32,800 businesses questioned in ONS's STES
	• Stratified random sample by industry and employment drawn from IDBR
	• All reporting units (RU) with employment above a threshold (varying by industry) are included in the sample
	• The RU employment data are apportioned between their local units (LU) using the ratio of LU to RU in order to produce more refined employment estimates by region and by industry
	• Its main strengths are the data are comparable over its whole duration (from 1959) and the accuracy of industrial information
	• There are limitations on the disaggregation of the industrial breakdown and there is no information on geographies below NUTS1
	• The time lag between publication and reference period is 11-12 weeks
	• Quarterly revisions are applied to previous quarters
	• Annual revisions of the previous two years are made at Q3
	• WFJ are benchmarked to ABI

Breakdowns:	• Estimates cover GB businesses registered for VAT/PAYE, classified to SIC2007 • Covers all major industry groups, such as production, construction, distribution, service trades and various other groups in SIC2007 • Splits are available by gender and full-time/part-time
Cautions:	• As data come from multiple sources, the risk is heightened that discontinuities in source data can lead to reduction in the quality of the outputs • ONS moved from a matched pairs estimator to a point-in-time estimator which has removed the bias of the matched pairs estimator but increased the volatility for small and medium sized businesses – see Barford (2010).
Use by LPC:	• Industrial breakdowns - WFJ is mainly compiled from surveys of business and is ONS's preferred source of statistics on jobs by industry as it provides a more reliable industry breakdown than LFS • As the industrial breakdown is limited, however, for a more detailed industrial analysis employee jobs series should be used
Availability:	• www.statistics.gov.uk
Special analyses:	• Earnings research conducted by ONS can be accessed via http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Workforce+Jobs

Part B: Rounding behaviour and measurement error in microdata

5. Introduction

Fry and Ritchie (2013) identified, using the Annual Survey of Hours and Earnings (ASHE), a preference amongst employers to 'round' hourly wages, for example to the nearest 50p or to 10p over the minimum wage. They also noted that the respondents to the Labour Force Survey (LFS) were liable to 'round' estimates of their earnings, leading to error in the measurement of the number of low paid.

These results were presented at a national level, but low wages are not uniformly distributed amongst occupations, industries and individuals. It is useful to know whether rounding behaviour is universal, or if it differs for the low paid. For example, if employer rounding is likely amongst low-paying organisations, this suggests that those organisations have more flexibility to set wage rates and are not constrained by the NMW. On the other hand, if there is less rounding amongst employee responses for a social group, this might indicate greater awareness of minimum wages amongst this group. This section therefore considers the universality of rounding behaviour.

One flaw of Fry and Ritchie (2103) is that it only considered ASHE and LFS; therefore it was not possible to definitely state that one type of behaviour is 'employer' or 'employee', however plausible it might be, because the findings could be a unique product of each survey design. As a partial remedy, this section augments ASHE and LFS analysis with results from the BHPS/USoc. This household survey shares many characteristics with the LFS, and so can provide evidence as to whether LFS findings are common to other surveys.

This part is organised as follows:

- Section 6 considers whether rounding behaviour is a product of hourly wages or of specific datasets, or whether it is a more universal characteristic of human behaviour
- Section 7 looks at the simple probability of rounding behaviour, to examine broad support for the hypothesis that rounding might be more common in some groups
- Section 8 carries out repeated regression analysis to separate out specific effects
- Section 9 considers whether the same factors are at work in rounding of weekly wages
- Section 10 summarises the findings of Part B
- Section 11 makes recommendations to the LPC

Note on 'rounding'

Throughout this section, reference is made to rounding in both datasets. For employer data, 'rounding' behaviour implies a lower level of constraint in the labour market and the possibility of transfers of a labour surplus from employers to employees. This therefore has policy implications for setting the NMW. In contrast, 'rounding' in the employee data implies measurement errors in the data, as it has been demonstrated that the pattern of rounding in employer and employee data differs substantially. This affects the LPC's ability to identify the impact of the NMW.

Hence 'rounding' reflects two different behaviours of interest to LPC. For simplicity the following discussion focuses on 'rounding' per se as the characteristic under observation, but it should be borne in mind that this is being used as a proxy for labour market conditions in the employer data and measurement error in employee data.

Note on types of wages discussed

This section discusses hourly wage rates, weekly wages and annual salaries. For clarity reference are always to hourly wage rates unless otherwise specified.

Finally, ASHE wage data is based on the 'derived' wage using the criteria defined by the LPC (including incentive and other payments but excluding overtime). LFS wage data is based on the 'stated' wage unless otherwise stated. For BHPS/USoc, both a 'derived and 'stated' wage are used. The contention of Fry and Ritchie - that employer data is

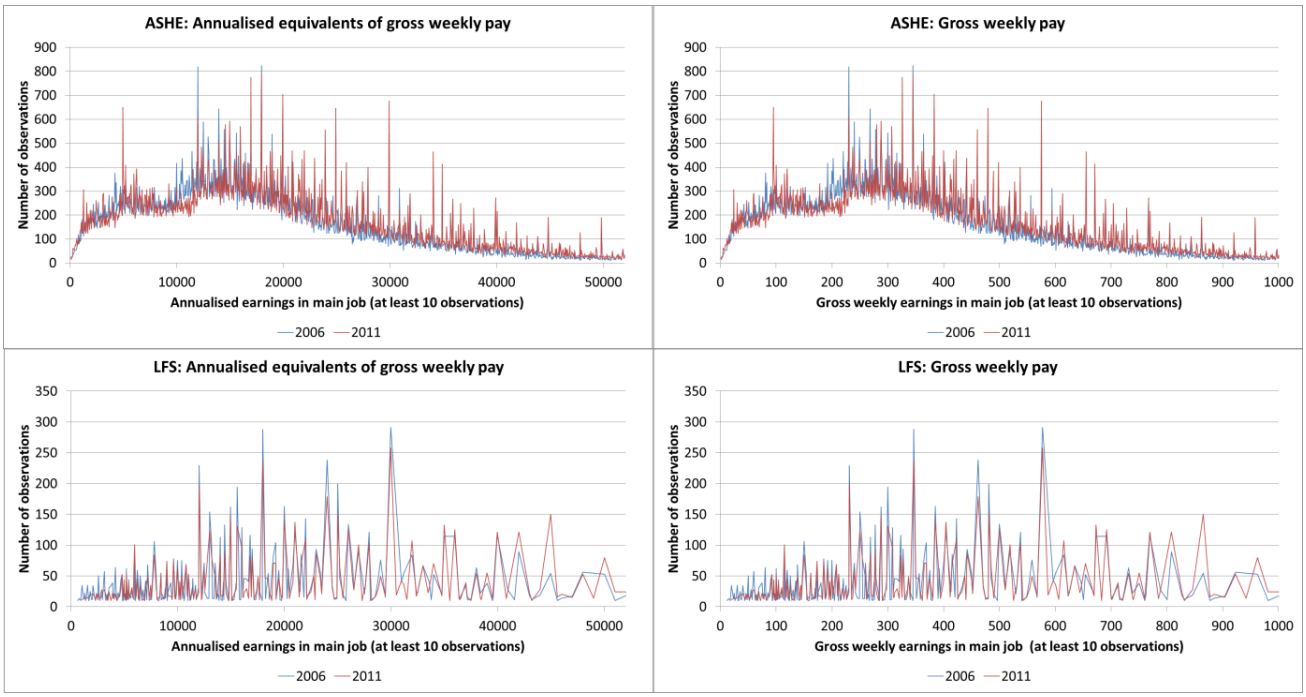
accurate and rounding reflects wage-setting behaviour, whereas the employee data is less accurate and rounding indicates a level of measurement error- is maintained here.

6. Is the rounding of responses a universal characteristic?

6.1 Gross earnings and respondent behaviour

Fry and Ritchie (2013) concentrated on hourly wage rates but noted that it might be worth considering whether total wages are also being rounded by employers or employees; this would suggest that rounding is a fundamental human behaviour. Both LFS and ASHE have several options for calculating annual wages; asking for information directly, and/or calculating annual wages from weekly ones (or vice versa). For consistency, an annualised wage based upon 52 times the weekly wage was used[1]. Figure B1 presents data for two years, 2006 and 2011. Only pay values with at least ten 10 survey responses are displayed.

Figure B 1 Gross earnings distributions in ASHE and the LFS

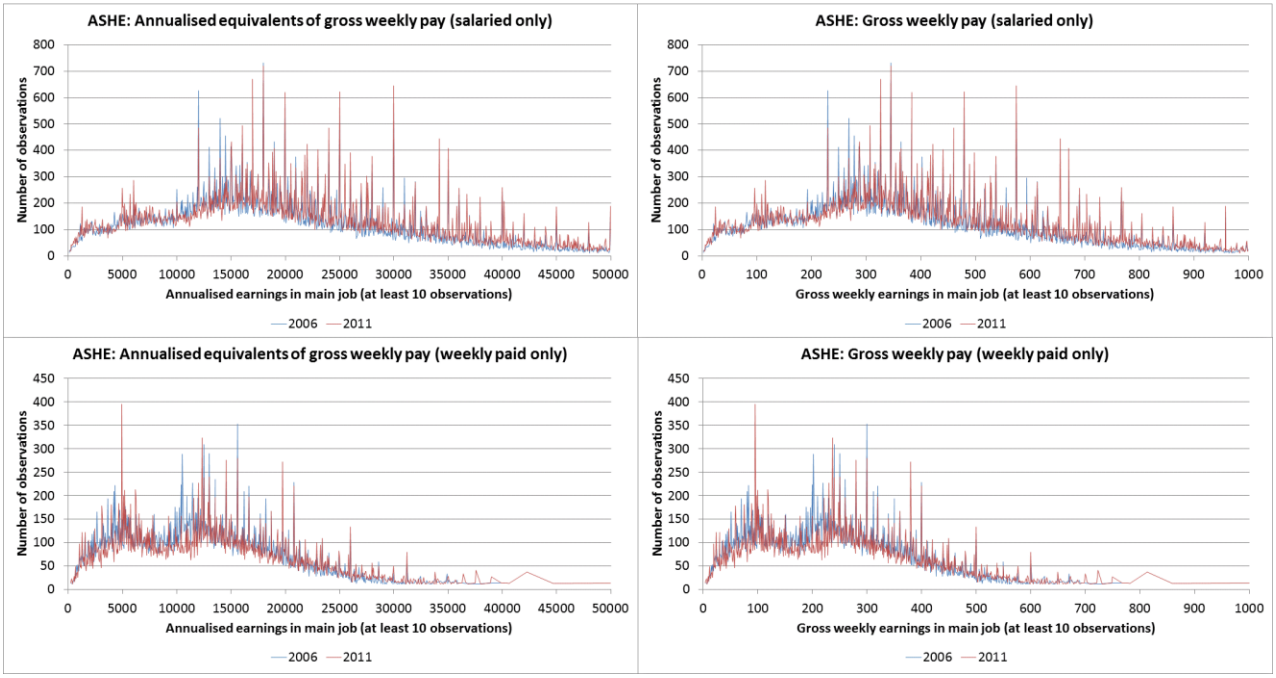


Both data sources show that gross earnings are more likely to be concentrated at certain points, such as £30,000 annual salary. However, for the low paid, the distributions are relatively smooth in the employer data, with popular exception of £100/week (£5200p.a.), and demonstrate the classic skewed shape of earnings distributions. Although wage rates are often rounded, Fry and Ritchie (2013, section 4.2) showed that the distribution of hours is more evenly distributed. This would suggest that, for the low paid, rounding on the part of the employers is more likely to take place in the context of the specific wage.

Figure B2 plots the ASHE responses by the salaried (defined as those whose salary reflects a calendar month) and the weekly paid (those whose pay covers a one-, two-, three- or four-week period):

[1] ASHE and LFS use slightly different values to translate between weeks and years: 52.167 and 52.143.

Figure B 2 ASHE earnings distributions by payment period



The weekly paid distribution is more skewed to the left confirming the assumption that, on average, those which earn less are more likely to be paid weekly. The weekly ASHE data provide some evidence of rounding to major focus points. In both 2006 and 2011 gross earnings of £250, £300 and £400 are within the ten most common observations. Of the ten most frequently observed levels of gross earnings, in 2006 eight were rounded (all to 10s) and in 2011 six were rounded (five to 10s and one to 5).

There is marginally less evidence of rounding of gross earnings of those paid a salary. In 2006, of the top ten most commonly observed levels of gross earnings, only three were rounded; this increases to five in 2011. For those that were paid a salary, the most frequent total wage rate in both years is £17,940 (£345). This maybe just a statistical quirk but could provide some evidence of little movement in total earnings, and hence wage rates, over this period.

An examination of the ASHE data provides limited support for the relevance of pay period. For example, Table B1 shows ASHE data for potentially two 'peak' week earnings:

Table B 1 Focus points and pay period (ASHE)

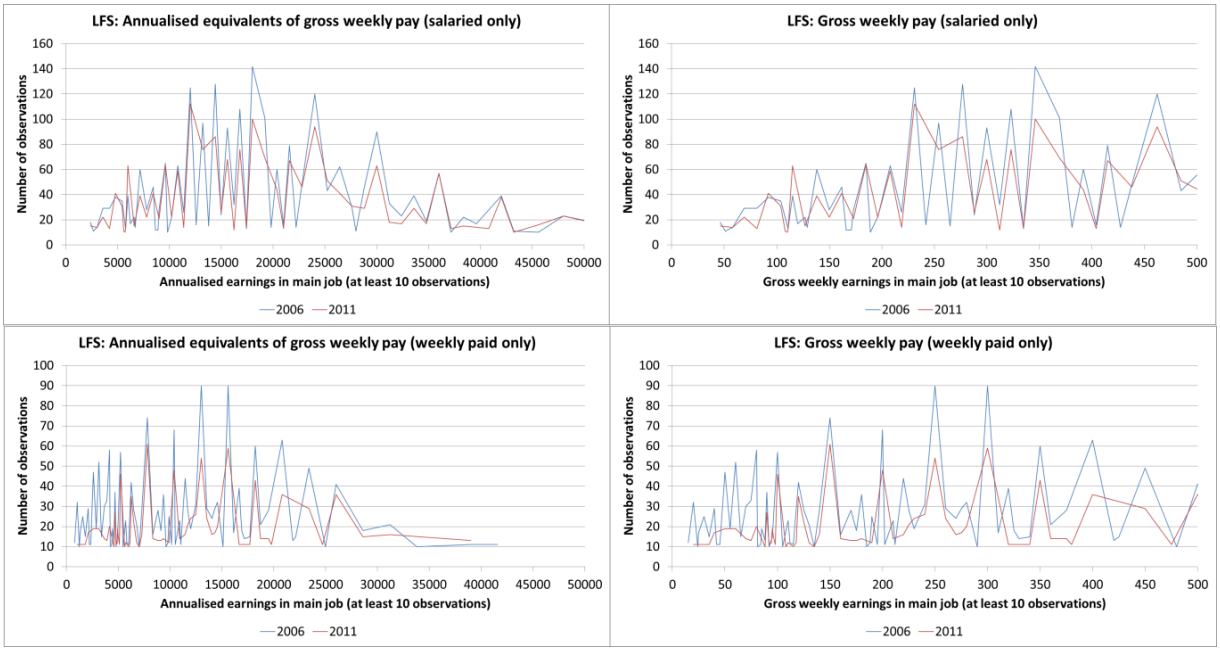
	Paid £100/week (£5,200 pa)		Paid £115/week (£6,000 pa)	
	2006	2011	2006	2011
Number of observations	351	408	282	372
Weekly-paid only	162	211	85	86
Proportion weekly-paid	46%	52%	30%	23%

In contrast to the prior expectation that a higher proportion of observations earning £100 per week (£5,200 pa) would have come from those paid weekly, Table B1 shows that, of the 351 and 408 individuals that earned £100 week, the proportions were almost equally split between those that were paid weekly and those that were salaried.

As expected, there is some evidence of a higher proportion of salaried workers rounding to the focus of point of £6,000 pa (£115 week). In 2006, 70 per cent of observations at this major focus point were from salaried workers; this rose to 77 per cent in 2011.

The employee response data in the LFS shows marginally more tendency to report at focus points. Figure B3 repeats the analysis above and breaks the LFS responses into the salaried and the weekly paid:

Figure B 3 LFS earnings distributions by payment period



The weekly paid seem to report more 'round' weekly wages (£150, £200, £250 etc); the regularity of these focus points is similar to that found in wage rate reporting. For the salaried there is less rationale for the pattern of observations.

An examination of the LFS data supports the relevance of pay period to employees. For example, Table B2 shows LFS data for the same two 'peak' week earnings as in Table B1 above:

Table B 2 Focus points and pay period (LFS)

	Paid £100/week (£5,200 pa)		Paid £115/week (£6,000 pa)	
	2006	2011	2006	2011
Number of observations	62	52	73	101
Weekly-paid only	57	46	11	10
Proportion weekly-paid	92%	88%	15%	10%

The LFS data for 2006/2011 show that 62 people claim to be paid £100/week, but 73 claim £115 per week, which approximates to £6,000p.a. If the salaried workers are separated, it is clear that most of those reporting £100 are weekly-paid, and most of those reporting £115 are salaried.

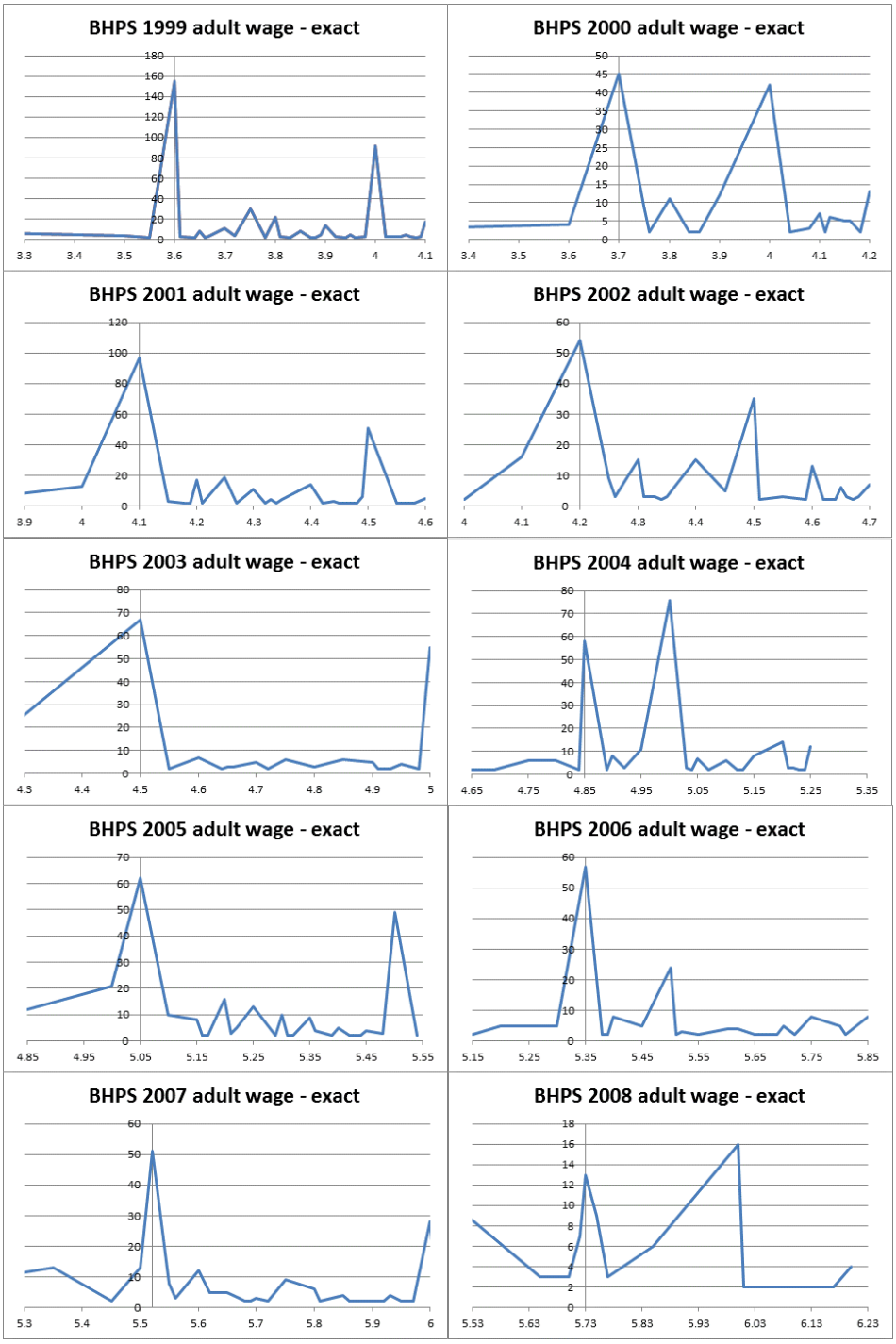
6.2 Rounding in the BHPS: like ASHE, like the LFS, or something else?

The British Household Panel Survey (BHPS) is widely used in academic research in the UK. Like ASHE, it has a longitudinal component – the same individuals have been followed from 1991 until 2008. Like the LFS, it has a wide range of personal and household characteristics, it includes individuals at all levels of labour market engagement, and it measures stated, derived and gross wages.

A natural question is whether the LPC could make more use of the BHPS. Equally, the household-based BHPS may be able to usefully inform the debate about the quality of the LFS data. This project examined BHPS through the same quality prism applied to the ASHE and LFS. To this end, the programs and analyses used on those two datasets for this report and Fry and Ritchie (2013) have been adapted and applied to the BHPS. At this stage, the key question is: does the BHPS look more like LFS, more like ASHE, or a separate entity?

A key element in identifying the rounding issue in ASHE and LFS was comparing wage distributions of stated and derived wages. Figure B4 below shows the sampling frequencies around the NMW back to 1999 and up to 2008; the y-axis crosses the x-axis at the minimum wage.

Figure B 4 Near-NMW wage distributions in the BHPS



An important feature of the BHPS is that it asks the respondent to identify whether a wage is an 'exact' amount or an estimated one; from the survey form:

> E34. What is your hourly rate of pay for your basic hours of work? WRITE IN AMOUNT PER HOUR IF EXACT AMOUNT NOT KNOWN ENTER APPROXIMATE AMOUNT AND CODE `Estimated amount' BELOW

The 'exact' wage is reported here. It is clear that BHPS shows the same strong preference for focus points as the LFS; but unlike the LFS, there appear to be relatively few observations below the NMW. This is even the case in 2005 and 2007 when the NMW was set just above a focus point, and when the error in the LFS is most visible. Note that the 'estimated' wage shows much more variation, and less definition at the minimum wage.

This suggests the BHPS might be a more accurate measure of household responses, and possibly may give an insight into the LFS when the differences between exact and estimated wages are reviewed. However, consider the data

needed to create the blobograms used in Fry and Ritchie (2013), which map derived wages and stated wages around the NMW in each year; see Table B3:

Table B 3 Popular stated-derived wage combinations in the BHPS

Stated	Derived	Year	Counts
410	407	2002	2-4 obs
410	410	2001	2-4 obs
410	411	2001	5-9 obs
420	411	2002	2-4 obs
420	412	2002	2-4 obs
420	419	2002	2-4 obs
420	420	2002	2-4 obs
420	425	2002	2-4 obs
430	429	2002	2-4 obs
450	450	2003	15
450	458	2003	2-4 obs
450	462	2003	2-4 obs
485	488	2004	2-4 obs
485	500	2004	2-4 obs
500	488	2004	2-4 obs
500	500	2004	17
500	500	2005	5-9 obs
505	500	2005	5-9 obs
505	505	2005	2-4 obs
505	506	2005	2-4 obs
535	535	2006	2-4 obs
550	550	2006	2-4 obs
550	550	2007	2-4 obs
552	538	2007	2-4 obs
552	544	2007	2-4 obs
552	550	2007	2-4 obs

As for the LFS, focal points dominate the results, irrespective of the NMW. The cause of this is that many stated wage variables are missing, and those that are present differ significantly from the derived wage (calculated from earnings over hours). If the stated wage rate is set equal to the derived rate when missing then the NMW values dominate. This differs from ASHE, where imputing a stated rate from the derived rate makes little difference to the distribution of identified wages.

6.3 Summary: the universality of rounding

Gross earnings paid by employers do show some of the same tendency as wage rates for some wages to be more popular than others; but there is little evidence of the strongly predictable and regular focus points found in wage rates. Rounding would therefore appear to be carried out at a wage rate specific to the pay period, which makes rounding of hourly wages more relevant to the low paid than other workers.

Employer responses show evidence of rounding of gross earnings. This rounding is more pronounced for the weekly paid workers at the lower end of the earnings distribution, whereas for salaried workers rounding, although less prominent, is dispersed throughout the earnings distribution.

Employee responses appear to show strong, persistent and predictable rounding of reported weekly earnings, but only for those who are paid weekly. As the LFS distributions differ from the ASHE distributions, which are believed to be more reliable, this suggest that predictable measurement error found in wage rates is also found in gross earnings and is concentrated amongst the lower paid.

Finally, the distribution patterns have changed little between 2006 and 2011, implying that the rounding effect amongst the low paid is persistent and independent of the level of wages. Again, this accords with previous findings on wage rates.

Analysing the BHPS shows that these results appear in all the earnings datasets considered. Surprisingly, the BHPS appears to have more in common with ASHE, in that a derived wage rate is both more common and seems to give more plausible outcomes for wages paid (as opposed to employees' perceptions of wages paid). However, this may arise because the way that wage data is collected it the BHPS allows one to distinguish between estimated and exact responses.

7. Rounding behaviour and sub-national low pay measures

7.1 Method

In this section we consider whether the overall tendency for rounding, either in wages or in survey responses, can be associated with the specific characteristics of the low paid. For example, Fry and Ritchie (2013, Table 18) reported that making ad hoc adjustments for rounding effects in the LFS, based upon observations of the ASHE distribution of wages, appeared to reduce the likelihood of minimum wages being reported for medium-sized firms.

Fry and Ritchie's analysis was limited to simple characteristics because of need for consistent analysis over several years. This meant that industry and occupation effects were not investigated. This analysis corrects that by incorporating the low-paying industries and occupations identified by the LPC at 4-digit (SOC) occupation and five-digit (SIC) industry level.

2011 is the first year SOC2010 and SIC2007 codes are available for both datasets. However, this does not allow pre- and post-recession effects to be measured. In addition, adult and youth NMWs were just below focus points in 2011, which is known to stimulate different responses compared to years where the NMW is at or above a focus point. Unless otherwise stated, ASHE and LFS data is used from 2002 to 2012. Only LFS Q2 is used; wage reporting around the NMW varies over the year, but using Q2 means the data collection period is consistent with ASHE reporting.

The LPC has identified specific groups of low-paid industries and occupations. Accordingly, the data was analysed using only the LPC detailed definitions, with other observations identified as 'non-low-pay industry/occupation'. For classifications prior to SOC2010 and SIC2007, a concordance table was used, included in Annex xxx.

7.2 Simple rounding probabilities for ASHE and LFS

Table B4 tabulates the simple rounding probabilities for ASHE and LFS data. Numbers are broken down by region, firm size, NMW age group, full-time, and whether the individuals are in low-paying industries or occupations, as defined by LPC. Each category is further broken down by 'low-paid', defined as whether the individuals earn within 50p of the relevant minimum wage. LFS-only markers (ethnicity, education and disability) were not included at this stage as the aim was to contrast employer and employee responses.

Data are presented for both absolute values of wages, and wage rates relative to the appropriate NMW. Rather than presenting numbers, the table below highlights whether the observed frequencies are more likely than expected, with the darker the colour the more unlikely that this rounding is statistically random. All rows have observations in the thousands.

Table B 4 Raw rounding probabilities for ASHE and LFS

			ASHE (derived)														LFS (stated)													
			Factor in absolute wage											Factor in relative wages				Factor in absolute wage											Factor in relative wages			
		Low paid?	2	3	4	5	6	7	8	9	10	25	50	5	10	25	50	2	3	4	5	6	7	8	9	10	25	50	5	10	25	50
Overall	All											x	x					o		x	x			x		XX	XX	XX	x	x	o	x
	<NMW+50p	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x		
	>=NMW+50p	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
Region	North East	No										o	x					o		x	x			x	-		XX	XX	XX	x	x	o	x
	North West	No										o	x					o	-	x	x			x	-		XX	XX	XX	x	x	x	x
	Yorks & Humber	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
	East Midlands	No										x	x					o		x	x		-	x			XX	XX	XX	x	x	x	x
	West Midlands	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
	South West	No										x	x					o		x	x			x			XX	XX	XX	x	x	x	x
	East	No										x	x					o		x	x			x	-		XX	XX	XX	x	x	x	x
	London	No										o	x					o		x	XX		-	x	-		XX	XX	XX	x	x	x	x
	South East	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
	Wales	No										o	x					o		x	x			x			XX	XX	XX	x	x	o	x
	Scotland	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
	North East	Yes				o					x	x	XX			-				o	x	o					XX	XX	XX	x	x		
	North West	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x	-	
	Yorks & Humber	Yes				o					x	x	XX			-					x	o		o			XX	XX	XX	x	x		
	East Midlands	Yes				o					x	x	XX			-					x	o		o			XX	XX	XX	x	x		
	West Midlands	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x	-	
	South West	Yes				o					x	XX	XX			-				o	x	o		o			XX	XX	XX	x	x		
	East	Yes				o					x	XX	XX			-				o	x	o	-	o			XX	XX	XX	x	x		
	London	Yes				o					x	XX	XX			-				o	XX	o		x			XX	XX	XX	x	x		
	South East	Yes				o					x	XX	XX			-				o	x	o		o			XX	XX	XX	x	x		
	Wales	Yes				o					x	x	XX			-					x	o		o			XX	XX	XX	x	x		
	Scotland	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x		
Firm size	0-9 emp	No				o					x	XX	XX		o		o	o		x	XX			x	-		XX	XX	XX	x	x	x	x
	10-49 emp	No				o					o	x	XX					o		x	x			x			XX	XX	XX	x	x	x	x
	50-249 emp	No										x	x					o		x	x			x			XX	XX	XX	x	x	o	x
	250+ emp	No											o					o		x	x			x			XX	XX	XX	x	x	o	x
	0-9 emp	Yes				x	o		o		x	XX	XX	o		-				o	XX	o		o			XX	XX	XX	x	x		
	10-49 emp	Yes				x					x	XX	XX	o	o	-					x	o		o			XX	XX	XX	x	x		
	50-249 emp	Yes				x					x	XX	XX			-					x	o		o			XX	XX	XX	x	x	-	
	250+ emp	Yes				o					o	x	x			-				o	x	o		o			XX	XX	XX	x	o		
Age band	Adult	No										o	x					o		x	x			x			XX	XX	XX	x	x	o	x
	Youth	No				o					o	x	XX		o					o	x			x			XX	XX	XX	x	XX	x	x
	Adult	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x		
	Youth	Yes				x					x	XX	XX					o		o	XX	o					XX	XX	XX	x	x	o	
Is in LP industry	No	No										o	x					o		x	x			x			XX	XX	XX	x	x	x	x
	Yes	No									o	x	x					o	-	x	x			x			XX	XX	XX		o	-	
	No	Yes				o					x	x	XX			-					x	o					XX	XX	XX	x	x		
	Yes	Yes				o					x	XX	XX			-				o	x	o	-	x			XX	XX	XX			-	
Is in LP occupation	No	No										o	x					o		x	x			x			XX	XX	XX	x	x	o	x
	Yes	No									o	x	x					o		x	x			x			XX	XX	XX	x	x	x	x
	No	Yes				o					x	XX	XX			-					x	o		o			XX	XX	XX	x	x		
	Yes	Yes				o					x	x	XX			-				o	x	o		o			XX	XX	XX	x	x		
Full time	No	No									o	x	x					o		x	x			x			XX	XX	XX	x	x	x	x
	Yes	No										o	x					o		x	x			x			XX	XX	XX	x	x	o	x
	No	Yes				o					x	XX	XX			-				o	x	o		o			XX	XX	XX	x	x		
	Yes	Yes				o					x	x	XX			-					x	o		o			XX	XX	XX	x	x		

Key

-	Observed frequency less than 75% of expected
o	Observed frequency 50% higher than expected
x	Observed frequency twice as high as expected
XX	Observed frequency four times as high as expected

For ASHE, absolute wages are rounded to 10p, 25p and 50p at a consistently higher rate than probability would suggest. However, this is most prominent for those earning within 50p of the NMW. This outcome occurs in all groups. Apart from this factor, most groups show a similar amount of rounding for all outcomes. Therefore, it appears that employers are most likely to round wages at low wage levels, irrespective of the characteristics of the job or worker. Employer rounding also seems to focus on absolute values, rather than setting to, for example, “10p over the NMW”. This contradicts Table 2 of Fry and Ritchie which showed a relative effect of 5p and 10p for small firms. The above table demonstrates that small firms do seem to offer wages rounded relative to the NMW, but only for the lowest-paid workers.

The classes “is in a low-paying occupation/industry” should be interpreted with caution, as the 4-digit SOC2000 and 5-digit SIC2003 definitions were constructed by the authors by working backwards from the current definitions. Nevertheless, it is interesting that the likelihood of rounding is still dominated by the characteristics of the specific job rather than industry-/occupation-wide.

The employee data from LFS tells a very different story. Absolute wages rounded to 10p, 25p and 50p are much more likely in all cases. Factors of 4p and 8p seem common, and relative distance from the NMW is important (the latter may be a result of the NMW often being set at 5p/10p levels during this period). Finally, there seems no relationship between rounding and being low paid. In summary, there is very little agreement between ASHE and LFS over rounding other than the general observation that it does occur.

7.3 Simple rounding probabilities for BHPS

It was noted above that the BHPS shares some characteristics with both employer and employee data. Reproducing the raw rounding probabilities for the BHPS reinforces this opinion: see Table B5 below. The statistics for the derived wage variable show more similarity to ASHE, whereas the stated variable approximates the LFS proportions.

Table B 5 Raw rounding probabilities for the BHPS

			BHPS (derived)														BHPS (stated)													
			Factor in absolute wage											Factor in relative wages				Factor in absolute wage											Factor in relative wages			
		Low paid?	2	3	4	5	6	7	8	9	10	25	50	5	10	25	50	2	3	4	5	6	7	8	9	10	25	50	5	10	25	50
Overall	All											o	x							o	x			o		XX	XX	XX	x	x	x	x
	<NMW+50p	Yes									o	x	XX								x			o		XX	XX	XX	x	x		
	>=NMW+50p	No			o	x			o		XX	XX	XX	x	x	x	XX			o	x			o		XX	XX	XX	x	x	x	XX
Region	North East	No										o	x					o		o	x	o		o		XX	XX	XX	x	x	XX	XX
	North West	No										x	x							o	x			o		XX	XX	XX	x	x	x	XX
	Yorks & Humber	No										o	x							o	x			o		XX	XX	XX	x	x	x	XX
	East Midlands	No										o	x							o	x					XX	XX	XX	x	x	x	x
	West Midlands	No										x	x							o	x			o		XX	XX	XX	x	x	x	XX
	South West	No										x	x				o			o	x			o		XX	XX	XX	x	x	x	XX
	East	No										o	x							o	x	o		o		XX	XX	XX	x	x	x	XX
	London	No										o	x							o	x			o		XX	XX	XX	x	x	x	x
	South East	No										o	o							o	x			o		XX	XX	XX	x	x	x	x
	Wales	No										x	x								x					XX	XX	XX	x	x	x	x
	Scotland	No										o	x							o	x	o		o		XX	XX	XX	x	x	x	XX
	North East	Yes						o	-			x	x						-		XX		-			XX	XX	XX	x	o		
	North West	Yes				o					o	x	XX	o						o	XX	o	-	o		XX	XX	XX	x	o	o	
	Yorks & Humber	Yes					-				x	x	XX	o							x	o		o	o	XX	XX	XX	x	XX	o	
	East Midlands	Yes		-							o	x	XX		o	-		o		o	x	o		o	o	XX	XX	XX	x	o		
	West Midlands	Yes						-				x	XX		-	-					x	o	-			XX	XX	XX	x	x		
	South West	Yes								-	o	x	XX					o		o	x			x		XX	XX	XX	x	x		
	East	Yes										x	XX			o					x		-			XX	XX	XX	x	x	o	
	London	Yes										x	x							o	x					XX	XX	XX	x	x	o	
	South East	Yes										x	x		-						x				o	XX	XX	XX	x	x		
	Wales	Yes					-					o	x			-		o		o	x	o		x		XX	XX	XX	x	x	-	
	Scotland	Yes									o	x	XX			-					x			o		XX	XX	XX	x	x		
Firm size	0-9 emp	No									o	x	XX			o	o	o		o	XX			o		XX	XX	XX	x	x	x	XX
	10-49 emp	No									o	x	x							o	x			o		XX	XX	XX	x	x	x	XX
	50-249 emp	No										o	x							o	x	o		o		XX	XX	XX	x	x	x	XX
	250+ emp	No																		o	x			o		XX	XX	XX	x	x	x	x
	0-9 emp	Yes									o	x	XX					o		o	XX	o		o	o	XX	XX	XX	x	x		
	10-49 emp	Yes									o	x	XX								x	o				XX	XX	XX	x	x	o	
	50-249 emp	Yes										x	x			-				o	x			o		XX	XX	XX	x	x		
	250+ emp	Yes										o	o		-						x		-			x	XX	XX	x	x	-	
Age band	Adult	No										o	x							o	x			o		XX	XX	XX	x	x	x	XX
	Youth	No									o	x	XX		o	o	o			o	x					XX	XX	XX	x	XX	XX	x
	Adult	Yes										x	x								x			o		XX	XX	XX	x	x		
	Youth	Yes									o	x	XX		o			o		o	XX	x	-	o	x	XX	XX	XX	x	XX	x	
	No	No										o	x							o	x			o		XX	XX	XX	x	x	x	XX
Is in LP	Yes	No									o	x	x				o			o	x			o		XX	XX	XX	x	x	x	XX
industry	No	Yes										x	x							o	x	o	-	x		XX	XX	XX	x	x		
	Yes	Yes									o	x	XX								x					XX	XX	XX	x	x		
	No	No										o	x							o	x			o		XX	XX	XX	x	x	x	XX
Is in LP	Yes	No									o	x	x							o	x					XX	XX	XX	x	x	x	XX
occupation	No	Yes									o	x	x								x			o		XX	XX	XX	x	x		
	Yes	Yes								-	o	XX	XX							o	x	o		o		XX	XX	XX	x	x	x	
Full time	No	No									o	x	XX				o			o	x			o		XX	XX	XX	x	x	x	XX
	Yes	No										o	o							o	x			o		XX	XX	XX	x	x	x	XX
	No	Yes									o	x	XX								x	o	-			XX	XX	XX	x	x		
	Yes	Yes										x	x								x			o		XX	XX	XX	x	x		

Key

-	Observed frequency less than 75% of expected
o	Observed frequency 50% higher than expected
x	Observed frequency twice as high as expected
XX	Observed frequency four times as high as expected

7.4 Summary of raw rounding probabilities

Overall, the employer data suggest that rounding is slightly more common for the low-paid (those earning less than 50p over the NMW); this is largely manifested in higher likelihood of wages at multiples of 10p, 25p and 50p. However there seems very little 'relative' wage effect. In contrast LFS respondents are much more likely to report rounded wages, but this does not seem to be related to whether they are low paid or not. BHPS respondents show characteristics of both datasets: derived wages follow the ASHE pattern, stated wages the LFS pattern.

8. Regression analysis of rounding probabilities

The above analysis only considers each category in isolation. Regression analysis was used to consider interactions between variables.

A regression analysis is normally carried out to test a particular model or hypothesis. Because the aim here was to study general characteristics of the dataset which might be useful knowledge in a range of circumstances, the decision was taken to run a variety of analyses and thus to identify, if possible, common themes arising.

To keep the task manageable, the analysis was limited probabilistic models of whether or not rounding was happening, as this was the phenomenon we wished to explain. It is quite possible that a very different approach (for example, trying to estimate wage regressions adjusting for rounding) would throw up different answers. However, the results found here are consistent with the findings from le Roux et al (2013) and Fry and Ritchie 92013).which did try some alternative specifications.

A series of probability models over various periods were run initially for ASHE and LFS. LFS analysis was not run prior to 2009 because of the difficulty of reconciling industry and occupation codes. Subsequently these models have been applied to the British Household Panel Survey (BHPS) and Understanding Society (USoc) datasets. These results are compared and contrasted to the ASHE and LFS datasets in the following section.

8.1 Method

Analysis was carried out on all those earning below and up to £1 over the NMW, and for adults and youth workers only. Variables were created for whether the wage was rounded in absolute levels, or relative to the appropriate NMW for that worker. For presentation purposes and ease of interpretation, only the regressions on absolute levels are presented in the main section. The regressions based on wages relative to the appropriate NMW are included in the appendices. Rounding is defined for these regressions as having a factor of 5p, 10p, 25p or 50p. Note that, prior to 2009, the NMW itself was rounded to at least 5p in every year except 2007.

The set of regressions is summarised in Table B6 below.

Table B 6 Hourly wage rounding: regression summary table

Regression test purpose		Included variables						Numbers of observations									
	Years	All observations wage < NMW+100p	All observations above NMW NMW<wage<NMW+100p	LPI codes	LPI yes/no marker	LPO codes	LPO yes/no marker	ASHE, derived wage, absolute	ASHE, derived wage, relative	LFS, stated wage, absolute	LFS, stated wage, relative	BHPS, derived wage, absolute	BHPS, derived wage, relative	BHPS, stated wage, absolute	BHPS, stated wage, relative	USoc, stated wage, absolute	USoc, derived wage, absolute
Pre-recession	2004 - 2007	yes		yes		yes		105,565				3,332	1,718	2,094	1,679		
	2004 - 2007		yes	yes		yes		82,392	82,392			1,718		1,670			
Post recession	2008 onwards	yes		yes		yes		114,077		18,520						5,967	13561
	2008 onwards		yes	yes		yes		88,553	88,553	13,760	13,760					4,508	2416
Single year with	2008 only	yes		yes		yes		24,054				770	344	475	362		
rounded NMW	2008 only		yes	yes		yes		18,769				344		362			
Focal point just	2011 only	yes		yes		yes		32,642		5,433						1,379	2991
above NMW	2011 only		yes	yes		yes		25,188	25,188	3,850	3,801					1,019	551
Focal point just	2012 only	yes		yes		yes		32,299		1,465							
below NMW	2012 only		yes	yes		yes		24,062	24,062	1,016	954						
Rounding to 5p	all years	yes		yes		yes		219,642	170,945	18,520	13,760	4,102	2,073	2,573	2,044	5,967	13561
Rounding to 10p	all years	yes		yes		yes		219,642	170,945	18,520	13,760	4,102	2,064	2,580	2,035	5,967	13561
Rounding to 25p	all years	yes		yes		yes		219,642	170,945	18,520	13,760	4,099	2,036	2,573	2,021	5,967	13561
Rounding to 50p	all years	yes		yes		yes		219,642	170,945								
Broad impact of	all years	yes			yes		yes	219,642		18,520		2,076		2,040		4,508	13561
LPI or LPO	all years		yes		yes		yes	170,945	170,945	13,760	13,760	3,662	1,885	2,212	1,757	2,175	2416
Not in LPO	all years	yes			yes			98,975	74,812	3,439	2,738	444	190	346	288	3,792	5706
In LPO	all years	yes			yes			120,667	96,133	15,081	11,022	1,637	871	650	541	1,478	7855
Not in LPI	all years	yes					yes	80,269	62,175	5,132	4,023	2,469	1,205	1,924	1,503	4,477	6086
In LPI	all years	yes					yes	139,373	108,770	13,388	9,737	1,283	591	744	527	1,480	7475
Firm size 0-9	all years	yes		yes		yes		25,059		5,155		591		509		943	3404
employees	all years		yes	yes		yes		17,599	17,599	3,480	3,480	1,528	764	976	749	2,475	444
Firm size 10-49	all years	yes		yes		yes		27,250		6,938		759		747		1,834	5130
employees	all years		yes	yes		yes		20,181	20,181	5,112	5,112	582	314	367	306	1,077	917
Firm size 50-249	all years	yes		yes		yes		25,741		3,725		301		308		866	2606
employees	all years		yes	yes		yes		19,300	19,300	2,925	2,899	684	381	460	430	872	503
Firm size 250+	all years	yes		yes		yes		141,432		2,544		379		430		784	2355
employees	all years		yes	yes		yes		113,762	113,762	2,136	2,136	1,060	531	877	714	1,954	467
Retail sector only	all years	yes				yes		69,321		5,672		531		714		1,565	2976
	all years		yes			yes		58,616	58,616	4,361	4,361	367	179	230	189	501	685
Social care sector	all years	yes				yes		12,698		1,704		179		155		406	734
only	all years		yes			yes		10,196	10,192	1,330	1,337						175
Childcare sector	all years	yes				yes		1,216		471							
only	all years		yes			yes		931	930	368	364						
panel	all years	yes		yes		yes						4,107	2,088	2,581	2,045		

Notes:
"All years": 2004-2011 (ASHE); 2009-2012 (LFS); 2004-2008 (BHPS); 2009-2012 (Usoc)
"LPO" - low-paying occupation
"LPI" - low-paying industry

The first two columns detail the various probability model specifications estimated and time period covered. Six regression approaches were taken.

The first set of regressions estimates the aggregate probability of rounding for pre- and post-recession as well as for specific years; the second set estimates the probability of rounding to specific factors (5p, 10p etc.); the third set considers whether simple yes/no markers for low-paying industry and occupation are useful; the fourth set estimates whether the likelihood of rounding is structurally different for those working in industries or occupations where low pay is frequent; the fifth set of regressions are estimated for company specific factors (i.e. size and specific low paying sectors); the final estimation uses a random-effects panel model estimation to control for unobserved heterogeneity. Due to modelling and processing constraints the authors were only able to run a random effects model for the BHPS dataset.

Columns three to eight identify a number of alternative model specifications tested including: all wages below the NMW and up to £1 above minimum wage; for wages above the NMW and up to £1 above; low-paying industry (11 industries) and occupation codes (14 occupations); and low-paying industry or occupation markers (yes/no). For presentation purposes only the results using wages below and above the minimum wage are presented in the main text - results just for those wages above the minimum wage are presented in Appendix xxx.

The third and final set of columns (nine to eighteen) identify which specific datasets have been analysed, whether the wage used in the regression is a derived wage or stated wage and whether the regression has been estimated for absolute and/or relative wage (results for the relative wage are not reported in the main text but are presented in Appendix xxx). Where there is a number recorded in a particular cell, this indicates that this model specification has been estimated – the figure identifies the number of observations in the regression. For example, the first two rows tells us a probability model was estimated for the pre-recession period 2004-2007, the regression included wages below and up to £1 over the national minimum wage (first row), as well as wages up to £1 above the minimum wage but excluding those under it (second row), and included identifiers for 11 low paying industries and 14 low paying occupations – this specification was run on ASHE using a derived wage in absolute terms (with between 82,000 and 106,000 observations) and BHPS for both derived and stated wage and in both absolute and relative terms, with 1,600 to 3,300 observations depending on the particular model.

In addition all regressions included the following variables: the relevant NMW for that worker, whether the NMW is rounded, basic working hours, gender, age, whether full-time, firm size (default is very large), public sector and region (default is NW England). Additional specific variables were included for specific datasets where available - i.e. unionisation (ASHE , BHPS and USoc), and disability, ethnicity, education (LFS, BHPS and USoc).

In addition, measures on the accuracy of the wage variable were included for LFS, BHPS and USoc; no equivalents are available for ASHE. As noted above, the BHPS/USoc marker for 'exact' wages seems a good indicator of ASHE-like wage distributions (by assumptions, more accurate ones). For the LFS, Ormerod and Ritchie (2007a) and Fry and Ritchie (2012) found that knowing whether the data came from proxy responses and was based on documentation affected the likelihood of rounding. In contrast, Fry and Ritchie (2013) found that only the use of documentation mattered, suggesting that proxy and direct respondents both tend to 'round' numbers in the same way.

The results are presented in the tables below. Because of the large number of outputs, and because the coefficients from a probability regression do not have a direct interpretation, only a schematic view is presented here. Significant and positive results are presented as "+" in green, with a darker green and more plusses indicating a more significant result; negative results are in red. A zero indicates an insignificant result[2].

8.2 ASHE

Table B7 below records the ASHE results from the regression analysis of factors affecting rounding for absolute wages. The regression results reported in the appendices record that rounding relative to the minimum wage appears less common than absolute rounding.

[2] Significance levels are calculated at 1%, 5% and 10%, meaning that, if a factor is significant at 1%, there is only a 1% chance that that result would have occurred by chance, and a 99% chance that this is a genuine result – assuming the model specification is correct.

Table B 7 Regression analysis of factors affecting rounding (ASHE)

ASHE absolute wage rate (derived wage)		Factors determining likelihood of rounding					Likelihood of rounding to specific values				Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		
Regression type:		Pre-recession	Post-recession	Single year (2008)	Single year (2011)	Single year (2012)	Rounding to 5p	Rounding to 10p	Rounding to 25p	Rounding to 50p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Childcare
Wage influences	wage	---	o	o	+++	+++	---	---	---	---	---	---	---	---	---	+++	o	--	---	---	--	++
	NMW	+++	---	---	---	---	+++	+++	--	++	+++	o	+++	+++	++	---	--	o	+++	+++	+++	o
	if NMW rounded		+++	o	o	o	+++	+++	o	o	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
	basic working hours	---	o	--	o	o	---	---	---	---	---	---	---	o	---	o	o	--	---	---	o	-
Personal characteristics	if female	---	--	o	o	o	---	---	---	---	---	---	---	---	---	o	---	---	---	++	o	o
	age	---	o	--	o	+++	---	---	---	---	---	---	o	---	---	---	o	--	---	---	--	o
	if full_time	---	---	++	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	o	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	+++
	firm size 10-49 employees	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	+++
	firm size 50-249 employees	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	+++
	if public sector	---	---	--	---	---	---	---	---	---	---	---	---	---	--	---	---	---	---	o	o	o
Region (default = NW)	NE England	+++	o	o	--	o	++	o	--	--	+	o	+++	o	+++	+	o	o	o	o	o	o
	Yorks	++	o	o	o	o	++	+++	o	o	++	o	+++	+++	o	o	o	o	++	o	o	o
	East Midlands	o	o	o	o	o	o	o	o	o	o	o	o	+++	o	+	o	o	o	o	--	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	-	+++	o	o	o	o	o	o	o	--	o
	SW England	---	++	o	o	o	o	++	+++	+++	o	o	o	o	o	+++	o	o	-	o	-	o
	East England	--	+++	o	+++	o	o	+	+++	+++	++	o	++	o	++	o	o	--	++	o	-	o
	London	---	o	o	o	+++	---	--	++	++	o	o	o	o	o	---	--	o	o	---	--	o
	SE England	---	+++	++	+++	+++	o	o	+++	+	o	o	o	o	o	--	o	o	o	---	o	o
	Wales	o	+	o	o	o	+	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Scotland	-	o	o	o	+++	o	o	---	--	o	--	o	--	o	o	o	o	o	o	o	o
	NI																					
If in low-paying industry (default = not LPI)	Agriculture	o	+++	o	+++	+++	+++	+++	o	o						++	++	o	+			
	Food processing	o	+++	o	+++	o	+++	o	---	--						+++	+++	o	o			
	Textiles	o	o	-	-	o	o	o	---	---						+++	o	o	o			
	Retail	---	o	---	o	o	---	---	---	---						+++	+++	++	---			
	Hospitality	+++	+++	+++	+++	++	+++	+++	+++	+++						+++	+++	+++	+++			
	Cleaning	+++	+++	+++	+++	+++	+++	+++	+++	+++						+++	+++	+++	+++			
	Social care	+++	+++	-	+++	o	+++	++	---	---						+	+++	+++	o			
	Childcare		o		o	o	o	o	o	o						o	o	o	---			
	Leisure	+++	+++	+	+++	o	+++	+++	+++	+++						+++	+++	+++	+++			
	Hairdressing																					
	Employment agencies		+++		+++	+++	+++	+++	+++	+++						+++	+++	+++	+++			
If in low-paying occupation (default = not LPO)	Agriculture	o	--	-	o	--	o	o	o	o						o	o	o	o	o	o	
	Food Processing	+++	---	--	---	--	++	o	-	o						o	+++	+++	-	o	o	
	Textiles	+	-	-	o	o	o	o	o	++						o	+	-	o	o	+++	
	Retail	+++	---	o	---	+	--	---	---	---						+++	+++	+++	---	o	---	o
	Hospitality	---	o	o	o	o	o	o	+++	+++						---	o	o	++	o	o	o
	Cleaning	+++	o	o	--	---	+	+++	---	o						++	o	o	o	o	o	o
	Social Care					o																
	Childcare	---	--	--	o	-	---	---	---	---						o	--	--	-	o	---	++
	Leisure	---	--	o	--	-	---	--	---	-						o	--	o	---	o	o	++
	Hairdressing	o	--	o	o	o	--	o	--	o						+++	---	o	o	o		
	Office Work	---	---	o	o	--	---	--	o	o						o	---	---	o	---	o	
	Non-food Processing	+++	-	o	--	o	+++	o	--	o						+++	+++	o	o	o	o	
	Storage					o																
	Transport	+	--	o	-	-	o	o	o	++						+++	o	o	---	o	o	
	Unionised	---	---	---	o	+++	---	---	---	---	---	---	---	---	---	o	---	---	---	---	---	---
NMW rounded to X' pence (default = 5p)	NMW rounded to 10p							+++														
	NMW rounded to 25p								+++													
	NMW rounded to 50p									+++												
Low paying (Y/N marker)	Is LPI										+++	+++	+++									
	Is LPO										---			+++	---							

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

"LPO" - low-paying occupation
"LPI" - low-paying industry

The ASHE data shows a clear result that company size matters: smaller firms are more likely to round wages. A rationale for this may be that rounding is a consequence of employers wanting to simplify wages to avoid administrative burden. This finding is robust to all model specifications.

In terms of low-paying industries, the ASHE data shows that hospitality, cleaning, leisure industries, and possibly employment agencies and agriculture, are more likely to round wages. Social care also seems more likely to round wages although the results are more sensitive to the specification.

The retail trade is less likely to round wages. This may be because this industry is dominated by a small number of very large companies who have the systems to manage complex wage scales effectively. This argument is supported

by the likelihood of rounding results estimated separately by company size (in the penultimate section of results, towards the right of the table). For all retail companies with less than 250 employees, they are more likely to round wages, whereas companies employing over 250 employees are less likely. Therefore we can conclude that it is retail firms with over 250 employees which are driving the overall effect for the retail sector; the overall result that retail is less prone to rounding is due to the dominance of very large firms in that sector.

As noted in Fry and Ritchie (2013), the results confirm that the public sector is less likely to round. The effects appear unaffected by the recession years.

In terms of low paying occupations generally, rounding seems less likely to occur, although the results are sensitive to model specification. Occupational variations, however, seem to weaken with the onset of the recession. There is a juxtaposition of the results for low paying industries which are generally likely to round, and low-paying occupations, which generally do not.

In terms of specific occupations, a number seem to show relatively consistent results. Childcare, leisure and office work seem to round less. These need to be interpreted in combination with the other variables. Hence, for example, leisure as an industry may round wages, but the leisure workers themselves are less likely to have their wages rounded. This can be interpreted as within the industry itself there seems less market constraint, whereas front-line leisure workers are more tightly priced and likely to be paid at their marginal product.

It is not clear that there is any strong regional effect; even the impact of working in London is sensitive to the regression specification.

The 'unionised' variable in ASHE reflects whether an employee is covered by a collective bargaining agreement, not whether that particular individual is in a union or not. There is a strong 'unionised' effect; collective bargaining seems to reduce the likelihood of wages being rounded. This may be evidence of tighter markets or tight negotiations. However, the public sector also does not round wages; as unions are much more active in the public sector this may, in some part, reflect the predominance of unions representing public sector employees in collective bargaining agreements.

Personal characteristics appear to play a significant role, but given that ASHE represents employer preferences, any interpretation must be cautious. For example, it is not clear why female workers appear consistently less likely to have their wages rounded, but this may be results of gender differences in employment not being fully addressed by other variables. Wages for individuals are more likely to be rounded if they work longer hours or are employed full-time. This may be explained by the fact that total cost to the employer increases as individuals work longer hours. As full-time jobs outnumber part time, on average, by about three to one in the UK, this implies that employers may focus more closely on pricing labour at its marginal product for the group of workers which account for the greatest share of its labour costs. The full-time result is sensitive to the recession period. This may reflect the changing structure of the labour market during this period which moved towards more flexible working patterns.

The setting of the NMW to focus points (rounded to 10p, 25p and 50p) has a strong positive effect on the likelihood of rounding in both absolute and relative values.

These regressions assume that the industry/occupation effects are marginal; that is, the basic relationship is the same for everybody, but those working in particular industries, for example, see their rounding probabilities adjusted. It could be argued that the relationship between the job, the employee and the likelihood of rounding is structurally different for those working in industries or occupations where low pay is frequent. Regressing the likelihood of rounding by industry and occupation allows more flexibility in these relationships. These regressions include a simple yes/no markers for industry and occupation; separate regressions then split the sample based on whether in a low-paying industry or occupation.

Employer size, working in the public sector, firm characteristics and unionisation remain significant, while region appears to show some more consistent results. Working in one of the low-paying industries always leads to more

likelihood of rounded wages, whereas working in a low-paying occupation makes rounding less likely *unless* the employer is not in one of the low-paying industries. This suggests there is some flexibility in wage setting in low-paying industries whereas the market for low-paying occupations seems to be tighter. When employed in a low paying industry and low paying occupation, occupation seems the dominant factor as wages are less likely to be rounded.

Overall, the ASHE results suggest that there are certain workplace characteristics that are strongly and frequently associated with the rounding of wages by employers. The probability of rounding increases the smaller the size of the company, for some low paying industries and when the NMW is rounded itself. The probability of rounding decreases as hours increase, for certain personal characteristics, working in the retail sector, working in low paying occupations and being part of a collective bargaining agreement.

8.3 LFS

Table B8 presents LFS results for the equivalent regression (2009 onwards). The LFS regression uses the preferred stated wages measure as opposed to derived wages which is the preferred measure in ASHE.

Table B 8 Regression analysis of factors affecting rounding (LFS)

LFS absolute wage rate (stated wage)		Factors determining likelihood of rounding			Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		
Regression type:		Post-recession	Single year (2011)	Single year (2012)	Rounding to 5p	Rounding to 10p	Rounding to 25p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Childcare
Wage influences	wage	+++	+++	+++	+++	+++	++	+++	o	+++	o	+++	+++	+++	++	o	+++	+++	o
	NMW	---	---	---	---	---	---	---	o	---	-	---	---	---	---	o	---	---	o
	if NMW rounded	+++			+++	+++	o	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	o
	basic working hours	-	-	o	-	---	--	--	o	-	o	--	o	o	o	o	o	o	o
Personal characteristics	if female	-	-	-	-	o	---	o	o	--	o	-	-	o	o	-	o	o	o
	age	+++	++	o	+++	+++	o	+++	o	+++	+++	++	++	++	++	o	+++	o	o
	if full_time	o	o	o	o	+	o	o	o	o	o	+	o	o	o	o	o	o	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+++	+++	o	+++	+++	+++	+++	+++	+++	+++	+++					+++	o	o
	firm size 10-49 employees	+++	++	o	+++	+++	+++	+++	+++	+++	+	+++					+++	o	o
	firm size 50-249 employees	o	o	o	o	o	o	++	+	+	o	+++					++	o	o
	if public sector	o	++	o	o	o	o	o	o	o	o	o	-	---	o	+	o	o	o
Region (default = NW)	NE England	--	o	o	--	--	-	--	---	o	-	--	--	o	o	o	o	--	o
	Yorks	--	--	o	--	-	--	--	o	--	-	o	o	---	o	o	o	o	o
	East Midlands	o	o	o	o	--	o	o	--	o	--	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	--	o	-	o	o	o	o	o	o	o	o
	SW England	+	o	o	+	o	+++	+	o	+	o	o	+++	o	o	o	o	o	o
	East England	+	o	o	+	o	+	+	o	++	o	o	o	o	o	o	o	o	o
	London	+++	+++	++	+++	++	+++	+++	o	+++	++	+++	+++	o	+	++	o	o	o
	SE England	+++	++	o	+++	++	+++	+++	o	+++	o	+++	++	o	o	o	o	o	o
	Wales	o	--	o	o	o	o	o	o	o	--	o	o	--	o	o	o	o	o
	Scotland	o	o	o	o	o	o	o	-	o	o	o	o	o	o	o	o	o	o
	NI																		
If in low-paying industry (default = not LPI)	Agriculture	o	o	o	o	o	o						o	o	o	o			
	Food processing	--	---	o	--	o	o						o	o	o	o			
	Textiles	o	o	o	o	o	o						o	o	o	o			
	Retail	---	---	o	---	---	---						---	---	--	---			
	Hospitality	o	o	o	o	o	o						o	o	o	o			
	Cleaning	o	o	o	o	o	o						o	o	o	o			
	Social care	---	--	---	---	---	---						o	-	--	o			
	Childcare	o	o	o	o	o	o						o	o	o				
	Leisure	-	o	o	-	--	---						o	o	o	o			
	Hairdressing																		
	Employment agencies	o	o	o	o	o	o						o	o	o	o			
If in low-paying occupation (default: not LPO)	Agriculture	o	o	o	o	-	o						o	o	o	o	o	o	o
	Food Processing	---	o	---	---	---	---						o	o	--	o	o		
	Textiles	--	o	o	--	o	o						o	o	-	o	o		
	Retail	---	o	o	---	---	---						o	---	--	o	o	o	
	Hospitality	---	o	o	---	---	---						-	o	--	o	o	o	o
	Cleaning	---	--	o	---	-	-						o	o	--	o	o	o	o
	Social Care	o	o	++	o	o	o						o	o	o	o		o	-
	Childcare	o	o	o	o	o	o						o	o	o	o		o	o
	Leisure	o	o	++	o	o	o						o	o	o	+			
	Hairdressing	o	o	+	o	o	o						o	+	o		o		
	Office Work	o	o	o	o	o	o						o	o	o	+	o	o	
	Non-food Processing	--	o	o	--	-	--						o	o	o	o	o		
	Storage	--	o	-	--	--	---						o	--	o	o	o		
	Transport	-	o	o	-	o	o						o	o	o	o	o		
Disability (default: none)	DDA disabled and work-limiting	o	o	o	o	--	o	o	o	o	o	o	o	o	o	o	o	o	o
	DDA disabled only	--	o	o	--	o	o	---	o	---	o	--	o	o	--	o	o	o	o
	Work-limiting disabled	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Ethnicity (default: white)	Mixed	o	o		o	+	o	o	o	o	o	o	o	o	o	o	o	o	
	Asian	o	o	o	o	o	o	o	o	o	o	o	+	o	o	o	o	o	o
	Black	o	o	o	o	o	o	o	o	o	++	o	o	o	o	o	o	o	o
	Chinese	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
	Other	+++	o	o	+++	+++	o	+++	o	+++	o	+++	++	o	o	++	o	o	
Education (default: A*-C GCSE or equivalen	Degree or equivalent	o	o	o	o	++	+	o	++	o	o	o	+	o	o	o	o	o	o
	Higher education	o	++	o	o	o	o	o	++	o	o	o	++	o	o	+	o	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	++	o	o	o	o	o	o
	Other qualifications	o	o	o	o	o	--	o	o	o	o	o	o	o	o	o	o	o	o
	No qualification	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Other	Proxy	+++	+++	++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	++	o
	With doc. (payslip)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-
	With doc. (bank)	++	+	o	++	o	o	++	o	++	o	o	o	++	o	o	+		
	With doc. (other)	---	--	o	---	---	-	---	---	o	---	o	---	o	o	o	o	o	o
NMW rounded (default = 5p)	NMW rounded to 10p																		
	NMW rounded to 25p						+												
Low paying (Y/N marker)	Is LPI							---	---	---									
	Is LPO							---			---	o							

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

"LPO" - low-paying occupation
"LPI" - low-paying industry

THE LFS results are less clear than the ASHE findings. Fewer results are significant. The regularly significant factors are; the size of the company; retail and social care industries; proxy response (given by a third party from the household); and whether responses supported by documentation.

The likelihood of rounding by size of company is consistent with the findings from ASHE, as is the likelihood of rounding in the retail industry. These results may represent true effects rather than implying measurement error.

The hypothesis here is wages supported with documentation will result in less rounding. This is partially true: individuals with payslips are less likely to round. However, a bank record increases rounding, if anything. This may be because bank documentation records gross salary payments and the individual still goes through the same rounding operation of dividing wage by hours as identified in Fry and Ritchie (2013). 'Other documentation' reduces the likelihood of rounding, but as it is not entirely clear what this covers, this is difficult to assess.

As expected, a third party response is more likely to report rounded wages and is further evidence of measurement error. This contradicts Fry and Ritchie (2013) who found no proxy effect, and who therefore argued that 'humans are human' and so a proxy is a likely to round as the employee. These results suggest that the employee is still more accurate than the proxy.

Given that the rounding is hypothesised to be a consequence of personal psychology, it is not surprising that many of the firm characteristics do not have a discernible impact on rounding by employees. Of more concern is that few of the personal characteristics seem to be robust to different specifications. The LFS, as contrasted with ASHE, allows the inclusion of a number of additional personal characteristics; in this analysis we have included disability, ethnicity and education. Disappointingly none of these personal characteristics provide any additional clear evidence to the sources of rounding error.

The likelihood of rounding by industry and occupation allows us to assess whether the likelihood of rounding is structurally different for those working in industries or occupations where low pay is frequent. For the LFS, employees seem to be less likely to round if in one of the low-paying industries or occupations. This contradicts the ASHE results that low-paying industries are *more* likely to round wages and is evidence of measurement error.

8.4 BHPS

Compared to ASHE and LFS, for this analysis the BHPS does have a number of drawbacks. First, the BHPS is missing more categories. Some of this is because of small numbers. However, the BHPS only holds 3-digit occupation and 4-digit industry codes. Therefore the BHPS can only approximate some of the LPC definitions.

Second it is clear from the regression summary table that there are relatively few observations. The BHPS is one sixth the size of a single wave of the LFS; the LFS is collected four times a year, and boosted for Scotland and Wales. Missing stated wages have been imputed from derived wages where feasible. Therefore the quality of the BHPS data might be high, but the usability may be limited for such a restricted perspective as wages in the vicinity of the NMW.

FInally, there is a problem of timing. The BHPS is collected in the last two quarters of the year; it straddles the introduction of the new NMW. It is known that some companies anticipate the introduction of the NMW as the date approaches, and some delay updating wages. In addition, employees questioned in October about wages may legitimately be reporting for a period when the previous NMW was relevant. For ASHE and the LFS these problems are avoided by using data from the Spring quarter. In theory it is feasible to correct for this by using the interview date and pay period, but this is not straightforward.

For simple tabulations BHPS appeared to be more akin to ASHE. This does not hold for analytical outcomes. Consider the regression results shown earlier, which now are generated for the BHPS and displayed in Tables B9 and B10.

Table B 9 Regression analysis of factors affecting rounding (BHPS stated wage)

BHPS absolute (stated wage)		Factors determining likelihood of rounding		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		Panel
Regression type:		Pre-recession	Single year (2008)	Rounding to 5p	Rounding to 10p	Rounding to 25p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Panel - Random effects
Wage influences	wage	- - -	+ + +	-	+ +	+ + +	-	o	o	o	o	o	- - -	o	o	o	o	-
	NMW	o	- - -	o	- - -	- - -	o	o	o	o	o	- -	o	o	o	o	o	o
	if NMW rounded	+ +	o	+ + +	o	o	+ + +	+ + +	+ + +	o	+ + +	+ + +	+ + +	o	+ + +	+ + +	+	+ + +
	basic working hours	o	o	o	o	o	o	+	- -	o	o	+ +	o	+ + +	-	o	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	- - -	o	o	o	o	o	o	o	o	o
	age	o	o	o	+ +	o	o	o	o	o	o	o	o	o	o	o	o	o
	if full_time	o	o	o	o	o	o	o	o	o	o	- -	o	- - -	+ +	o	o	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +					+ + +	o	+ + +
	firm size 10-49 employees	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ +	+ +	+ + +					+ + +	o	+ + +
	firm size 50-249 employees	+ + +	+ +	+ + +	+ + +	+ + +	+ + +	+ + +	o	o	+ + +					+ + +	o	+ + +
	if public sector	o	o	o	o	- - -	o	o		-	o	o	-	o	o	o	o	o
Region (default = NW)	NE England	o	- -	o	-	o	o	o	o	-	o	o	o	o	o	o	o	o
	Yorks	o	o	o	o	o	o	o	o	- -	o	o	o	o	o	o	o	o
	East Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	-	o	- - -	o	o	o	- -	o	o	o	o
	SW England	o	o	o	o	o	o	o	o	o	o	o	+ +	o	o	o	o	o
	East England	o	-	o	o	- -	-	- -	o	-	o	o	o	- -	o	- -		o
	London	o	o	o	+ +	+	o	o		o	o	o	o	o	o	o	o	o
	SE England	o	o	+	+ + +	+ +	o	o	o	o	+ +	+	+	o	o	o	o	o
	Wales	o	- - -	-	o	o	-	-	o	-	o	o	o	- - -	o	o	o	o
	Scotland	o	o	o	o	o	o	o	-	- -	o	o	o	o	o	o	o	o
	NI																	
If in low-paying industry (default = not LPI)	Agriculture	-	o	o	-	o						o	o	o				o
	Food processing	o	o	o	o	o							o	o	- -			o
	Textiles	o	o	o	o	o							o	o	o			o
	Retail	- - -	- -	- - -	- - -	- - -						o	- - -	o	- - -			- - -
	Hospitality	o	o	o	- - -	- -						o	o	o	o			o
	Cleaning	o	o	o	o	o						o	o	o	o			o
	Social care	o	o	o	o	- - -						o	o	o	-			o
	Childcare																	
	Leisure	-	o	-	- -	- - -						- -	o		o			o
	Hairdressing																	
	Employment agencies																	
If in low-paying occupation (default = not LPO)	Agriculture	o	o	o	o	o							o					o
	Food Processing																	
	Textiles																	
	Retail	o	o	o	o	o						o	o	- -	+ +	o		o
	Hospitality																	
	Cleaning																	o
	Social Care																	
	Childcare																	
	Leisure	o	o	o	o	o						o	o	o	o	o		o
	Hairdressing	+ + +	o	+ +	o	o						o	o	o				+ +
	Office Work	o		o	o	o						o				-		o
	Non-food Processing	o		o	o	o							- - -					o
	Storage																	
	Transport																	
Disability (default: none)	DDA disabled and work-limiting																	
	DDA disabled only																	
	Work-limiting disabled																	
Ethnicity (default: white)	Mixed																	
	Asian	o	o	o	o	o	o	o	o	o	o		o	o	o	o		o
	Black				o													o
	Chinese																	
	Other	o		o	o	o	o	o			o		o		o	o		o
Education (default: A*-C GCSE or equivalent)	Degree or equivalent	o	o	o	o	o	o	o	+	o	o	o	o	o	+	o	o	o
	Higher education	+ +	o	+ +	+ +	o	+ +	o	+ +	+	o	o	o	o	+	o	o	+ +
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Other qualifications	o	o	o	o	o	o	-	o	- -	o	o	o	o	o	- -	o	o
	No qualification	o	o	o	o	o	o	o	+ +	o	o	o	o	o	+ +	o	o	o
Other	Exact	- - -	- - -	- - -	- - -	- - -	- - -	- - -	o	- - -	- - -	- -	- - -	- - -	- - -	- - -	o	- - -
	Unionised	- - -	o	- - -	- - -	- - -	- - -	- - -	-	-	- - -	- - -	o	- - -	o	o	o	- - -
NMW rounded (default = 5p)	NMW rounded to 5p																	
	NMW rounded to 10p				o													
	NMW rounded to 25p					+ + +												
Low paying (Y/N marker)	Is low-paying industry						- - -	- - -	o									
	Is low-paying occupation						o			o	+							

Table B 10 Regression analysis of factors affecting rounding (BHPS derived wage)

BHPS absolute (derived wage)		Factors determining likelihood of rounding		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		Panel
Regression type:		Pre-recession	Single year (2008)	Rounding to 5p	Rounding to 10p	Rounding to 25p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Panel - Random effects
Wage influences	wage	++	++	+++	++	+++	+++	++	++	o	+++	++	++	o	o	+++	o	+++
	NMW	---	-	---	---	---	---	---	o	-	---	--	-	-	o	---	o	---
	if NMW rounded	o	o	o	o	o	o	o	o	o	+++	o	+	o	o	+	o	o
	basic working hours	o	o	o	-	o	o	o	o	o	o	o	-	o	o	o	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	age	+++	o	++	+++	o	++	++	o	o	++	o	+++	+	o	++	o	++
	if full_time	o	o	o	o	---	o	o	o	o	o	o	o	--	o	o	o	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+++	o	+++	+++	+++	+++	++	+++	o	+++					+++	o	+++
	firm size 10-49 employees	o	o	o	o	++	+	o	+++	o	+++					+++	o	o
	firm size 50-249 employees	+	o	o	o	++	+	o	+	o	++					++	o	o
	if public sector	o	o	o	o	---	o	o	+	o	o	o	o	---	o	o	o	o
Region (default = NW)	NE England	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Yorks	o	--	--	o	-	--	--	o	o	--	o	---	o	o	o	o	--
	East Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	SW England	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	East England	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o
	London	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	SE England	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o
	Wales	o	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o
	Scotland	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o
	NI																	
If in low-paying industry (default = not LPI)	Agriculture	o	o	o	o	o						o	o	o				o
	Food processing	++	o	+	++	o							o	o	++			+
	Textiles	o	o	o	o	o							o	o	o			o
	Retail	--	--	---	---	---						o	o	--	--			--
	Hospitality	o	o	o	o	o						o	+	o	o			o
	Cleaning	o	o	o	o	o						o	o		o			o
	Social care	o	o	o	o	o						o	o	o	o			o
	Childcare																	
	Leisure	o	o	o	--	o						o	o	o	o			o
	Hairdressing																	
	Employment agencies																	
If in low-paying occupation (default = not LPO)	Agriculture	o	o	o	o	o						o	o			o		o
	Food Processing																	o
	Textiles																	
	Retail	o	o	o	o	o						o	+	o	o	o		o
	Hospitality																	
	Cleaning																	o
	Social Care																	
	Childcare																	o
	Leisure	o	o	o	+	o						+	o	o	o			o
	Hairdressing	o	o	+	+++	++						++	o	o				+
	Office Work	o	o	o	o	o						o	o	o		o		o
	Non-food Processing	o	o	o	o	o						o	o	o				o
	Storage																	
	Transport	o		o	o									o				o
Disability (default: none)	DDA disabled and work-limiting																	
	DDA disabled only																	
	Work-limiting disabled																	
Ethnicity (default: white)	Mixed																	
	Asian	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+	o	o
	Black	o		o	o	o	o	o			o		o					o
	Chinese	o	o	o	+	o	o	o			o	o						o
	Other	o	o	o	o	o	o	o		o	o	o	o			o		o
Education (default: A*-C GCSE or equivalent)	Degree or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Higher education	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Other qualifications	-	o	-	o	o	-	o	o	o	-	o	--	o	o	o	o	o
	No qualification	o	-	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Other	Exact	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Unionised	---	--	---	---	---	---	---	o	---	---	--	---	o	o	o	o	---
NMW rounded (default = 5p)	NMW rounded to 5p																	
	NMW rounded to 10p				o													
	NMW rounded to 25p					o												
Low paying (Y/N marker)	Is low-paying industry						o	o	o									
	Is low-paying occupation						+			o	++							

The BHPS does provide some supporting evidence for ASHE and LFS: small firms are more likely to round, and less likely to round if working in the retail sectors or covered by a collective bargaining agreement (unionisation).

However, there are notable differences between the stated and derived wage outcomes; for example, the regional and industry outcomes are quite different. Using the stated wage produces stronger results but, as noted above, this seems to be a lower quality variable, and there are fewer valid responses.

One interesting feature is the 'exact wage variable', which asks whether the wage is exact or estimated by the respondent. For the stated wage this significantly reduces the chance of rounding. Hence it plays a similar role to the use of pay slips in the LFS data, but it could be argued that this is a more relevant question as it allows the researcher to adjust for the accuracy of information whether documentation is used or not.

Overall, the preliminary conclusion on the BHPS is that in some statistical respects it bears more similarity to ASHE than the LFS, despite the collection method. This suggests that it might be helpful in determining the nature of rounding behaviour by employers, particularly with respect to disadvantaged employees. However the timing and small numbers limits its direct application to understanding the impact of the NMW.

8.5 Understanding Society (USoc)

Difficulties with linking BHPS and USoc meant that the initial analysis concentrated on the BHPS as it had several years' worth of data to analyse. The BHPS finished in 2008, and so the lessons that can be drawn from it may be out of date, particularly with respect to labour market behaviour during the recession.

USoc, the BHPS' successor, has a much larger sample size and revised industry, occupational and ethnic classifications. However, the USoc dataset is still in its infancy and under development. This has meant that a number of variables were unavailable (e.g. region, unionisation) and therefore a restricted regression specification has been reported. The analysis has been applied to all three waves of the USoc (2009-2012). Results for 2012 have not been reported individually due to the small number of observations.

As USoc is an extension to its predecessor BHPS, the expectation is that the wage distribution for derived wages will be more closely aligned with ASHE than with LFS. The regressions were run for both the derived wage and the stated wage and the results are presented in tables B11 and B12.

Table B 11 Regression analysis of factors affecting rounding (USoc stated wage)

USoc absolute (stated wage)		Factors determining likelihood of rounding		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	Rounding to 5p	Rounding to 10p	Rounding to 25p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care
Wage influences	wage	+++	+++	+++	o	+++	++	++	o	o	++	+	+++	o	o	o	o
	NMW	--	---	--	o	--	-	-	o	o	--	---	-	o	o	o	o
	if NMW rounded	+++		+++	+++	o	+++	+++	+++	+++	+++	+++	+++	+++	o	+++	++
	basic working hours	---	o	---	o	-	--	--	o	-	-	--	o	o	--	o	---
Personal characteristics	if female	o	o	o	o	o	o	o	o	o	o	o	o	o	o	-	o
	age	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+
	if full_time	++	-	++	o	+	++	+	+	o	++	++	o	o	o	o	++
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+++	+++	+++	+++	+++	+++	+++	+++	++	+++					+++	++
	firm size 10-49 employees	+++	o	+++	++	o	+++	+++	+++	o	+++					++	++
	firm size 50-249 employees	o	o	o	o	o	+	o	o	o	++					o	+
	if public sector	---	--	---	---	-	---	o	---	---	-	---	--	o	--	o	o
If in low-paying industry (default = not LPI)	Agriculture	o		o	o	o						o	o				
	Food processing	---	o	---	o	---						--	o	-	o		
	Textiles	o	+++	o	+	o							o	o			
	Retail	---	---	---	---	---						--	---	---	---		
	Hospitality	---	o	---	--	-						-	o	o	---		
	Cleaning	--	o	--	o	o						o	o	o	o		
	Social care	o	o	o	-	---						o	o	o	-		
	Childcare	o	o	o	o	o						o	+				
	Leisure	o	o	o	o	o						o	--	o	o		
	Hairdressing																
	Employment agencies	o	+	o	o	++							o		o		
If in low-paying occupation (default = not LPO)	Agriculture	o	o	o	o	o						o	o	o	o	o	
	Food Processing	o	o	o	o	o						o	o	o	o	o	o
	Textiles	o	-	o	o	o						o	o	-	o	o	
	Retail	o	o	o	o	o						o	o	o	o	o	
	Hospitality																
	Cleaning	o	o	o	o	o						o	o	o	o	++	o
	Social Care																
	Childcare	o	o	o	o	+						o	o	o	o	o	
	Leisure	o	o	o	o	o						o	o	o	o		
	Hairdressing	o	o	o	o	o						o	o	o	o	o	
	Office Work	o	o	o	+	++						o	++	o	o	o	
	Non-food Processing	o	+	o	o	o						o	o	o			o
	Storage																
	Transport	+++	o	+++	o	o						o	+++	o	o	o	
Disability (default: none)	DDA disabled and work-limiting																
	DDA disabled only	o	o	o	o	o	o	o	o	o	o	--	o	o	o	o	o
	Work-limiting disabled																
Ethnicity (default: white)	Mixed	o	o	o	o	o	o	o	+	o	+	o	o	o		o	o
	Asian	++	o	++	o	o	++	+	+	++	o	+	o	o	o	o	+
	Black	+	o	+	o	o	+	o	o	o	o	o	++	o	o	o	o
	Chinese	o		o	o	o	o	o	o		o	o	o	o	o	o	
	Other	o	o	o	o	o	o	o	o		o	o	o	o	o	o	
Education (default: A*-C GCSE or equivalent)	Degree or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Higher education	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o
	GCE, A-level or equivalent	o	o	o	+	o	o	o	o	o	o	o	++	o	o	+	o
	Other qualifications	+	o	+	++	o	++	o	o	o	++	o	++	o	o	o	
	No qualification	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Other	Exact	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Low paying (Y/N marker)	Is low-paying industry						---	---	---								
	Is low-paying occupation						--			o	--						

Table B 12 Regression analysis of factors affecting rounding (USoc derived wage)

USoc absolute (derived wage)		Likelihood of rounding			Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	Abs single year (2012)	Rounding to 5p	Rounding to 10p	Rounding to 25p	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care
Wage influences	wage	o	o	-	o	o	o	o	o	o	o	o	o	o	- -	o	o	o
	NMW	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	if NMW rounded	o			o	o	++	o	o	o	o	o	o	o	o	o	o	+++
	basic working hours	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o	o
	age	o	o	o	o	- -	o	o	o	-	o	o	-	o	o	o	- -	o
	if full_time	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	-	o	o	-	-	o	-	o	- -	o	o					o	-
	firm size 10-49 employees	o	o	o	o	o	o	o	o	-	o	o					-	o
	firm size 50-249 employees	o	o	o	o	o	o	o	o	-	o	o					o	o
	if public sector	++	+	o	++	o	o	+++	+	+	++	++	+++	o	o	o	o	o
If in low-paying industry (default = not LPI)	Agriculture	o	o		o	o	o						o	o	o			
	Food processing	o	o		o	o	o						+	o	o	o		
	Textiles	o	o		o	o	o						o	o	o	o		
	Retail	o	o	o	o	o	o						+	o	o	o		
	Hospitality	++	+	-	++	o	-						o	o	o	o		
	Cleaning	o	o		o	o	o						o	- -	o	++		
	Social care	o	o		o	o	o						o	o	o	o		
	Childcare	o	o	o	o	o	+						o	o	o			
	Leisure	o	o		o	o	o						o	o	o	o		
	Hairdressing																	
	Employment agencies	o	o		o	o	o						o	o	o	o		
If in low-paying occupation (default = not LPO)	Agriculture	o	o		o	o	o						o	o	o	o	o	
	Food Processing	o	o	o	o	o	o						o	o	o	o	-	
	Textiles	o	o		o	o	o						o	o	o	o	o	
	Retail	o	o	o	o	o	o						o	o	o	o	-	
	Hospitality																	
	Cleaning	o	o	o	o	+	+						o	o	o	+	o	o
	Social Care																	
	Childcare	o	o	o	o	o	o						o	o	o	o	o	o
	Leisure	++	o		++	o	+						o	+++	o	o		++
	Hairdressing	o	o		o	o	o						o	o		o		
	Office Work	o	o	o	o	o	o						o	o	o	o	-	o
	Non-food Processing	o	o	o	o	o	o						o	o	o	o		o
	Storage																	
	Transport	+	o		+	++	o						o	+	o	o	o	o
Disability (default: none)	DDA disabled and work-limiting																	
	DDA disabled only	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o	o	o
	Work-limiting disabled																	
Ethnicity (default: white)	Mixed	o	o		o	o	o	o	o	o	o	o	o	o	+	o	o	o
	Asian	o	o		o	o	o	o	o	o	o	o	o	++	o	o	o	o
	Black	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Chinese	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	
	Other	o	o		o	o	o	o	o	o	-	o	o	o	o	o	o	o
Education (default: A*-C GCSE or equivalent)	Degree or equivalent	o	o		o	o	o	o	o	- -	o	-	o	o	o	o	- -	o
	Higher education	o	o		o	o	o	o	o	o	o	-	o	o	o	o	o	-
	GCE, A-level or equivalent	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Other qualifications	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	No qualification	+	o		+	+	o	+	o	+	+	o	o	+	o	o	o	o
Other	Exact	o	o	-	o	o	+	o	o	o	o	o	o	o	o	o	o	o
Low paying (Y/N marker)	Is low-paying industry							o	o	o								
	Is low-paying occupation							o			o	o						

The results from the derived wage in Table B12 provide very little evidence of factors determining the likelihood of rounding and therefore the following analysis focusses purely on the stated wage, Table B11.

The stated wage, although potentially of poorer quality than the derived wage, provide some support for the evidence in ASHE, LFS and BHPS. There is continued evidence that wages for smaller firms and more likely to be rounded, whereas public sector and retail sector wages are less likely to be rounded.

As with the BHPS, specifying an exact wage reduces the chance of rounding. This is potentially an important conclusion for the Low Pay Commission. If LFS data collection methods encouraged exact responses and included this variable in their survey, the LFS dataset could potentially be of more use to the LPC and low pay analysts.

8.6 Summary of regression analysis on hourly wages

The regression analysis suggested that there are some workplace characteristics which help to define whether wages will be rounded or not. This matters as 'rounding' is in effect a proxy for labour market power. If a company sets wages above the minimum and seemingly based on social norms or a desire for simplicity, this implies that the company has 'room to manoeuvre'. In these circumstances the NMW has less impact compared to a company which is rigidly following an NMW or NMW-plus policy.

There are some odd results. LFS, BHPS (stated) and USoc (stated) all report less likelihood of rounding when working in a low-paying industry. This result was not internally consistent within BHPS and USoc and contradicts the ASHE finding that working in low paying industry increases the probability of rounding wages.

The LFS results are disappointing on the face of it. As one of the aims of this project was to develop tools to address measurement error, the lack of evidence for the source means that it is problematic to systematically correct. Fewer (consistently) significant responses make it difficult to identify factors influencing rounding, and hence whether low wages (or breaches of the NMW) in particular cases are genuine or reporting error. The analysis does not suggest there are clear sources for reporting error. This is challenging for the LPC, as it does not allow corrections to be made to population statistics. Ironically, however, the lack of significant results is good from an analytical perspective; it suggests that the likelihood of rounding is genuinely random, which is statistically much easier to deal with compared to response bias.

There are of course several recurring significant factors in the LFS. Some are not important for measurement error, as they are also reflected in ASHE (and therefore are assumed to represent true values). Of the remainder, it is clear that accounting for measurement error requires the researcher to acknowledge the importance of proxy responses and payslips – but possibly little else.

Although USoc has a larger sample size and revised industrial, occupational and ethnicity codes, it is still a dataset under development which poses some considerable challenges for the users. Theoretically it should be possible to match the BHPS with USoc although this is not straightforward and has not been achieved in this analysis. Combining the two data sources will enable a richer analysis to be undertaken. If USoc, like its predecessor BHPS, begins to resemble ASHE then it could be a very useful resource to the LPC in understanding the nature of rounding behaviour by employers. As the size of the sample is much greater than its predecessor it may prove to be of real value in understanding the impact of the NMW.

Perhaps the most interesting finding from the BHPS and USoc is the inclusion of the exact variable which reduces the likelihood of rounding wages. If a true result, this provides evidence that the inclusion of such a question in the LFS could provide valuable information when analysing the dynamics of low-pay.

9. Weekly wages

9.1 Method

To investigate whether the issues identified in the previous section were relevant across all earnings or just limited to the hourly wage (which is more likely to affect low-paid workers), the team was asked to look at weekly wages. A subset of the regressions of the previous section was repeated, but using weekly wages and with rounding associated with £5, £10, £25 and £50 boundaries. In this analysis, all employees are included (previous analysis only used those near, but not on, the minimum wage). Given the findings above, all years were analysed together, with 2011 2012 only being analysed as test cases. Relative wages were not analysed, as there is no obvious value against which weekly wages could be compared. Table B13 below shows the numbers of observations and the combinations of regressions studied, and results for ASHE and the LFS are detailed in Tables B14 and B15.

Table B 13 Weekly wage rounding: regression summary table

Regression test purpose	Years	LPI codes	LPI yes/no marker	LPO codes	LPO yes/no marker	N. observations: ASHE, derived wage, absolute	N. observations: LFS, stated wage, absolute
Single year	2011	yes		yes		171,774	36,013
	2012	yes		yes		165,125	9,374
Rounding to 5p	all years	yes		yes		1,406,222	125,179
Rounding to 10p	all years	yes		yes		1,406,222	125,179
Rounding to 25p	all years	yes		yes		1,406,222	125,179
Rounding to 50p	all years	yes		yes		1,406,222	
Broad impact of LPI or LPO	all years		yes		yes	1,406,222	125,179
Not in LPO	all years		yes			921,949	87,730
In LPO	all years		yes			484,273	37,449
Not in LPI	all years				yes	1,006,270	89,535
In LPI	all years				yes	399,952	35,644
Firm size 0-9 employees	all years	yes		yes		107,760	23,923
Firm size 10-49 employees	all years	yes		yes		157,498	36,003
Firm size 50-249 employees	all years	yes		yes		164,681	29,921
Firm size 250+ employees	all years	yes		yes		975,682	34,614
Retail sector only	all years			yes		195,864	15,020
Social care sector only	all years			yes		48,592	4,795
Childcare sector only	all years			yes		3,748	1,557
Weekly paid	all years	yes		yes		430,827	28,235
Salaried	all years	yes		yes		975,395	96,944

Notes:
"All years": 2004-2012 (ASHE); 2009-2012 (LFS)
"LPO" - low-paying occupation
"LPI" - low-paying industry

Table B 14 Regression analysis of factors affecting rounding (ASHE derived wage, weekly equivalent)

ASHE absolute (derived wage)				Likelihood of rounding to specific values				Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector			By payment period	
Regression type:		Single year (2011)	Single year (2012)	Rounding to £5	Rounding to £10	Rounding to £25	Rounding to £50	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Childcare	Weekly Wage	Salaried
Wage influences	wage	+++	+++	---	---	---	---	---	--	---	---	---	---	---	o	---	---	--	o	---	---
	weekly paid	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	++	o		
	basic working hours	o	--	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	++	+++	+	o	--	+++	o
Personal characteristics	if female	---	-	---	---	---	---	---	---	---	---	---	---	---	---	o	---	o	-	---	---
	age	o	o	o	o	---	-	o	o	o	---	+++	+++	o	o	--	o	o	o	---	o
	if full_time	-	o	--	-	---	---	--	---	o	---	o	o	o	--	--	o	o	o	-	o
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	o	-	+++	+++
	firm size 10-49 employees	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	-	+++	+++
	firm size 50-249 employees	+	++	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	o	o	+++	o
	if public sector	+++	++	+++	+++	+++	+++	+++	+++	+++	+++	o	--	o	++	++	o	o	o	---	+++
Region (default = NW)	NE England	o	o	o	o	-	--	o	o	o	o	o	---	o	o	o	o	o	o	o	o
	Yorks	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	East Midlands	o	o	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o	o
	SW England	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o	o
	East England	o	o	o	o	+++	+	o	o	o	+	o	o	++	o	o	o	o	---	+++	o
	London	o	o	+	+++	+++	+++	++	o	++	o	+++	+++	+++	+++	o	o	o	o	+++	o
	SE England	o	o	o	o	+++	+	+	o	++	o	o	+++	++	o	o	o	o	o	+++	o
	Wales	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--
	Scotland	--	o	---	---	---	---	---	---	o	---	o	---	-	o	o	o	o	o	--	o
	NI																				
If in low-paying industry (default = not LPI)	Agriculture	o	o	-	--	o	--						---	o	---	o				---	o
	Food processing	o	o	--	-	---	---						o	++	o	o				-	o
	Textiles	-	-	-	---	o	o						--	o	o	o				---	o
	Retail	---	o	---	---	--	---						+++	++	o	---				o	o
	Hospitality	o	o	o	---	---	o						o	o	o	o				---	+
	Cleaning	+++	o	+++	o	o	o						o	+	+++	+++				+++	++
	Social care	o	o	---	---	---	---						---	--	o	o				---	o
	Childcare	--	---	---	---	---	---						---	-	o	o				---	-
	Leisure	o	o	o	o	o	---						---	o	o	o				---	o
	Hairdressing																				
	Employment agencies	o	--	---	---	-	---						---	o	o	o				--	o
If in low-paying occupation (default = not LPO)	Agriculture	--	o	-	---	---	---						---	--	o	o	+	o		--	o
	Food Processing	o	o	---	---	---	---						--	---	o	o	o	o		---	o
	Textiles	o	o	o	o	---	o						o	---	--	+++	o	o		---	o
	Retail	-	---	---	---	---	---						---	---	--	o	---	o	o	---	--
	Hospitality	---	o	---	---	---	---						---	---	---	o	-	o	-	---	o
	Cleaning	o	o	---	---	---	---						---	---	--	o	o	o	o	---	o
	Social Care	---	--	---	---	---	---						---	---	---	---	o	--	o	---	--
	Childcare	o	++	--	--	---	---						---	o	-	o	o	o	o	---	o
	Leisure	o	o	---	-	-	-						--	o	o	o	o	o	o	--	o
	Hairdressing	o	o	+++	++	o	o						o	o	o	o	o	++		--	o
	Office Work	o	o	---	---	---	---						---	--	o	++	o	o	o	---	o
	Non-food Processing	o	o	---	---	---	---						---	---	---	o	o	+		---	o
	Storage	o	-	---	---	---	---						---	---	o	o	o	o	o	---	o
	Transport	+++	+++	+++	+++	o	o						---	o	+++	++	+++	o		+++	+
Low paying (Y/N marker)	Unionised	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	o	-	+	---	o
	Is low-paying industry							---	---	--											
	Is low-paying occupation							---			---	---									

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

Table B 15 Regression analysis of factors affecting rounding (LFS stated wage, weekly equivalent)

LFS absolute (stated wage)		Factors determining likelihood of		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector			By payment period	
Regression type:		Single year (2011)	Single year (2012)	Rounding to £5	Rounding to £10	Rounding to £25	All industries and occupations	All industries but NOT in LPO	All industries but in LPO	All occupations but NOT in LPI	All occupations and in a LPI	0-9 employees	10-49 employees	50-249 employees	250+ employees	Retail	Social Care	Childcare	Weekly Wage	Salaried
Wage influences	wage	o	o	o	---	+++	o	o	++	o	o	---	o	+	++	--	+	o	+++	o
	weekly paid	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++		
	basic working hours	++	o	+++	+++	+++	+++	o	+++	o	+++	+++	+++	o	o	+++	o	o	+++	o
Personal characteristics	if female	o	o	---	o	+++	---	o	---	o	---	-	-	o	o	o	o	o	---	o
	age	o	o	o	---	o	o	o	-	o	---	o	o	o	o	o	o	o	o	o
	if full_time	++	+++	+++	+++	+++	+++	+	o	o	+++	+++	o	o	o	o	o	o	+++	--
Firm (default: 250+ employees, private sector)	firm size 0-9 employees	o	o	o	o	---	o	o	++	o	+					++	o	-	o	o
	firm size 10-49 employees	o	o	o	o	---	o	o	+	o	o					o	o	o	o	o
	firm size 50-249 employees	o	o	o	o	---	o	o	o	o	o					++	o	--	o	o
	if public sector	o	o	o	o	+++	o	o	+	--	++	o	o	o	o	o	+	o	o	-
Region (default = NW)	NE England	o	o	o	o	o	o	o	--	o	o	o	o	o	o	o	o	o	---	o
	Yorks	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	East Midlands	o	--	o	o	o	o	o	o	o	o	o	o	o	o	++	o	o	o	o
	West Midlands	o	o	o	-	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o
	SW England	-	-	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	East England	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o
	London	o	-	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+++	o
	SE England	o	o	o	o	+	o	o	o	o	+	o	o	o	o	o	o	o	o	o
	Wales	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Scotland	o	--	o	o	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o
	NI																			
If in low-paying industry (default = not LPI)	Agriculture	o	o	o	o	o						o	o	o	o				++	o
	Food processing	o	o	o	o	o						o	++	o	o				o	o
	Textiles	o	o	o	o	o						o	o	o	o				o	o
	Retail	o	o	---	---	o						o	--	o	---				---	o
	Hospitality	o	o	-	---	---						--	o	o	o				o	o
	Cleaning	-	o	---	---	-						o	--	---	o				-	--
	Social care	o	o	o	---	o						o	o	o	o				o	o
	Childcare	o	+	o	o	++						o	+	o	+++				o	+
	Leisure	o	o	-	o	o						o	--	o	o				o	o
	Hairdressing																			
	Employment agencies	o	o	o	o	o						o	o	o	o				++	o
If in low-paying occupation (default = not LPO)	Agriculture	o	o	o	o	o						o	o	o	--	o	o	o	o	o
	Food Processing	o	o	+	+	---						o	o	++	o	+++	o		o	o
	Textiles	--	o	--	o	o						o	o	o	--	o			--	o
	Retail	o	o	---	---	---						---	o	--	o	o	o	o	--	o
	Hospitality	--	o	o	o	---						o	o	o	o	o	o	-	o	o
	Cleaning	o	o	--	---	---						---	---	o	o	o	o	o	---	+++
	Social Care	o	o	o	o	+++						o	o	o	o		++	o	o	o
	Childcare	o	o	o	-	-						o	o	o	o	o	o	o	o	o
	Leisure	o	o	++	o	o						o	+	o	o	o	o		o	++
	Hairdressing	o	o	+++	o	--						o	+	o	o	o			+++	o
	Office Work	o	o	o	++	o						o	+	o	o	o	o	o	o	o
	Non-food Processing	o	o	o	+	o						o	o	o	o	o	o		o	o
	Storage	o	o	o	o	-						o	o	++	o	++	o		o	o
	Transport	o	o	+++	+++	o						o	++	o	o	++	o		+	o
Disability (default: none)	DDA disabled and work-limiting	o	o	o	o	o	o	o	o	o	---	o	-	o	o	-	o	+	---	o
	DDA disabled only	--	o	-	o	o	-	o	---	o	--	o	---	o	o	--	--	o	---	o
	Work-limiting disabled	o	o	o	o	o	o	o	o	+	--	o	o	o	+	o	o	o	o	o
Ethnicity (default: white)	Mixed	o	o	o	--	o	o	o	o	o	o	-	o	o	o	o	o		o	o
	Asian	++	o	o	o	o	o	o	+	o	+++	++	o	o	o	o	o	o	+++	o
	Black	o	o	o	o	++	o	o	o	o	o	o	o	o	o	o	o	o	+	o
	Chinese	o	o	o	--	o	o	o	o	o	o	o	o	o	o	o	o	o	++	o
	Other	o	o	++	+	++	++	o	o	+	o	o	o	o	++	o	o	o	++	o
Education (default: A*-C GCSE or equivalen	Degree or equivalent	o	o	o	o	--	o	o	o	o	o	+	o	o	o	o	o	o	+++	o
	Higher education	+++	o	+++	o	+	+++	+++	o	+++	+	++	o	o	+++	o	o	+	o	+++
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+	o	o
	Other qualifications	o	o	o	+++	+++	o	o	o	o	+	o	+	o	o	o	++	+	o	o
	No qualification	o	o	o	+++	o	o	o	o	o	++	o	o	++	o	o	++	o	o	o
Other	Proxy	o	o	+++	+++	o	+++	o	+++	o	+++	+++	o	o	o	++	o	o	+++	---
	With doc. (payslip)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--
	With doc. (bank)	o	o	o	--	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	With doc. (other)	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o
Low paying (Y/N marker)	Is low-paying industry						---	--	o											
	Is low-paying occupation						o			o	o									

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

For ASHE data, firm size is again a good predictor, with smaller firms more likely to round weekly wages. Interestingly, the public sector also seems more likely to round weekly wages. Most of the low-paying occupations are less likely to round weekly wages. Being paid a weekly wage rather than an hourly rate increases the probability of rounding; and when separate regressions are run on weekly paid and salaried employees (in the last two columns of Table B13), these results persist for the weekly paid but there is no systemic impact on salaried workers.

These results from ASHE appear to be strongly significant, but for most columns there are a very large number of observations. The lack of significant results from ASHE for 2011 and 2012, despite having over 160,000 observations, may indicate that these findings are less robust then they appear.

For the LFS, the only consistent influences on rounding appear to be whether wages are paid weekly, and whether payslips were checked. Other than that, there is little effect, even when estimating separately for the weekly paid and salaried.

9.2 Summary of regression analysis on hourly wages

Section 6.1 noted that rounding appears to be associated with pay period. The ASHE results suggest that the rounding of weekly pay is more likely in some low-paying jobs and in smaller firms, but only for those with a weekly pay period; for the salaried there is little evidence of systematic rounding. For LFS data, there is almost no evidence of evidence of rounding effects, with the exception that using a pay slip reduces rounding. This reinforces the argument of section 8, that rounding in household surveys appears to be randomly distributed amongst the working population.

10. Quality and rounding in the LPC microdata: summary

This project can be seen as a natural extension of Fry and Ritchie (2013). That project identified rounding behaviour in employers and employees; this project has sought to clarify where that arises.

For employers, the consistency of results across a range of analytical methods suggests that rounding behaviour is to some degree predictable. As rounding is seen as a proxy for flexibility in setting wages, this indicates that there are certain identifiable parts of the labour market where wages are set according to characteristics other than measurable productivity. This result also carries over to weekly wages and salaries: rounding occurs at the level of the pay period, so weekly wages are more likely to be rounded if the employee is paid weekly, for example.

This rounding is most evident in small firms and private sector companies; other results are less clear, although it seems as if some sectors are more likely to round than others. The occupation of the employee appears to be a more important factor than the industrial sector of the employer: low-paying occupations are more likely to have wages set very tightly, irrespective of the sector they occur in.

The reasons for wage flexibility are not understood. Small firms might pay round wages because they directly negotiate with employees who the owner knows personally. Alternatively, it may be that small firms do not have the necessary management information and hence rely upon ‘rules of thumb’. Both hypotheses are consistent with the results presented here. However the findings from the retail sector (that big chains tie down wages tightly whereas small firms are likely to round) may weight the argument towards the ‘management information’ hypothesis.

The LFS findings on measurement error in employee responses identified in Fry and Ritchie (2013) are reinforced here. The analysis appears to lead to a dead end, in that very little seems to determine the error; but this can be seen as good news for researchers as random variation is statistically easier to deal with than error correlated with variables of interest. There is some evidence to support this: Fry and Ritchie (2013) noted that manual rounding adjustments in the LFS made little difference to a simple econometric model; and Le Roux et al (2013)’s more detailed analysis showed that while rounding may affect summary statistics, it appears to have much less effect on multivariate analyses.

It may also be good news for the LPC, in that it justifies some crude adjustment of data. For example, in 2012 many more employees report wages of £6.00 instead of £6.08 than is expected; these are likely to be erroneous. On the basis that these errors are random, LPC can adjust LFS data to reflect ASHE proportions whilst being reasonably confident that this will not overly bias results.

The BHPS and USoc information provide interesting supporting evidence; they largely support the view that household measurement error is largely random, at least in the context of variables of interest to analysts. They do however provide strong evidence of the value of having a check on whether respondents' answers are 'exact' or not. It could be argued that this is a more useful measure than the LFS questions on proxy and documentation; those two clearly matter in terms of quality indicators, but directly probing the accuracy of the variable might be more useful.

11. Recommendations on use of ASHE, LFS, BHPS and USoc microdata

Fry and Ritchie (2013) recommended further analysis to uncover the roots of rounding behaviour. On the basis of results presented here, we do not feel that further analysis is likely to provide sufficient new information to justify significant LPC funding.

Instead, the impact of the 'proxy', 'documentation' variables in the household surveys suggest that LPC should pursue the issue of LFS data collection with ONS; this would seem, to the authors, to be the single most important way to improve data quality. The 'exactness' check in BHPS and USoc may suggest a model.

The findings on occupation suggest that LPC might usefully focus on its attention on the type of jobs being done, rather than the industry in which the workers operate; and on smaller, private firms in all industries.

Finally, the apparent randomness of measurement error in the LFS (with the exception of the 'payslip' variable) suggests that the LPC can place confidence in multivariate analysis using the LFS; in the case of descriptive statistics, ad hoc adjustment (for example to align with ASHE) may be justified.

Part C: Imputation in the LFS

12. Background

12.1 The purpose of imputation

The authors were asked to look into the SPSS imputation code. This is applied to the LFS microdata and with a view to estimating missing wage observations. LPC analysis is carried out on the post-imputation dataset.

The code was developed by ONS and the University of Southampton around 2002/3. At the time of development, concerns over the coverage of the low paid in the New Earnings Survey (NES, ASHE's predecessor) meant that low pay estimates were based on combined NES and LFS estimates. The multiple imputation (MI) method was a novel development – a review by the authors in 2006 identified this as the only extant example of this technique in use at ONS. ONS no longer uses the imputation code, relying instead upon ASHE data for most analyses and non-imputed LFS data for specialist studies.

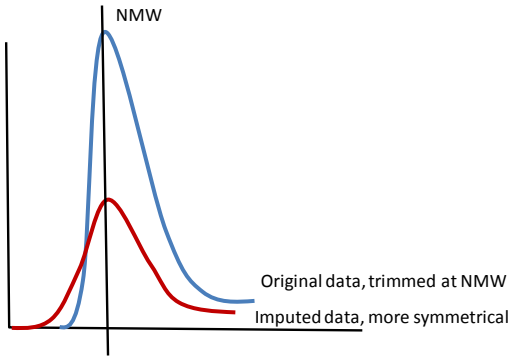
The LPC values the imputation procedure as a necessary corrective procedure to the raw LFS data. LFS data is used to provide low pay estimates for those groups which ASHE cannot identify: by education, disability, nationality or ethnicity. It is known that the LFS tends to produce lower numbers of the low paid than ASHE, which is seen as the more reliable estimate. However, LFS estimates for population subgroups cannot simply be scaled up, as it is thought that certain groups (for example, those without qualifications) are disproportionately likely to be low paid. The imputation procedure was developed to allow gaps in the LFS to be filled on a more statistically sound basis.

The MI technique is straightforward in principle. Where a stated hourly wage rate exists, a regression model is used to generate an equation for wages. This is used to generate a predicted hourly wage and residual for all observations, so that missing observations can be ranked. Missing observations are then augmented with ten 'donor' observations above and below in the order list of observations, the so-called 'nearest neighbours'. The dataset weights are adjusted to ensure that the MI does not create additional jobs overall.

12.2 Potential concerns

There are three major concerns with the process.

- *Symmetric imputation*: imputation processes assume a smooth distribution for the characteristics under study. However, the NMW is an on-off marker, and so imputation which changes the dispersion of the data affects the number of low paid even if means are unchanged. The LFS MI method uses donors from above and below the missing values, and so potentially will add weight to the number of low paid; for example:



- *Manual intervention*: the imputation process requires the operator to make subjective decisions about how outliers are to be dealt with, if at all, and the identification of outliers is ad hoc. Hence there is scope for the number of low paid to vary by operator.
- *Small numbers*: MI requires a large number of observations. For the analysis of low pay in the areas where the LPC is particularly interested and for which the LFS is the only feasible source (detailed breakdowns by

age, ethnicity, qualification and disability), the numbers involved are very small. Imputed results may therefore be sensitive to the specific breakdown.

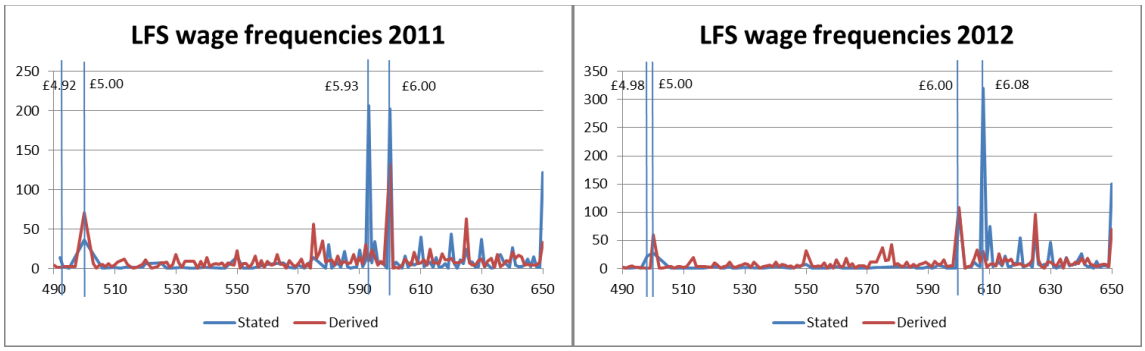
In addition, the imputation process only produces imputed values; it does not generate appropriate confidence intervals. Therefore, there is no statistical information on which to judge the quality of the derived estimates.

It should also be noted that the imputation code is large and complex (some 10,000 lines of code spread across seventeen programs). The code is written for a specific dataset, meaning variable names have to be edited. Much of the code is duplicated (for example, for main and second jobs, or teenage and youth analysis) rather than managed within loops with varying parameters. Recoding of source variables (for example, to generate occupational groups or to set the NMW) is dispersed throughout the code. Finally, although the code files are numbered, there appears to be no version control and data files are re-used. The program suite therefore has a large possibility of error.

12.3 Rounding in the LFS 2011/2012 and the implications for imputation

Before studying the effect of imputation, we briefly review the 2011 and 2012 LFS data. Figure C1 below shows numbers stating wage rates, and wage rates derived from stated earnings.

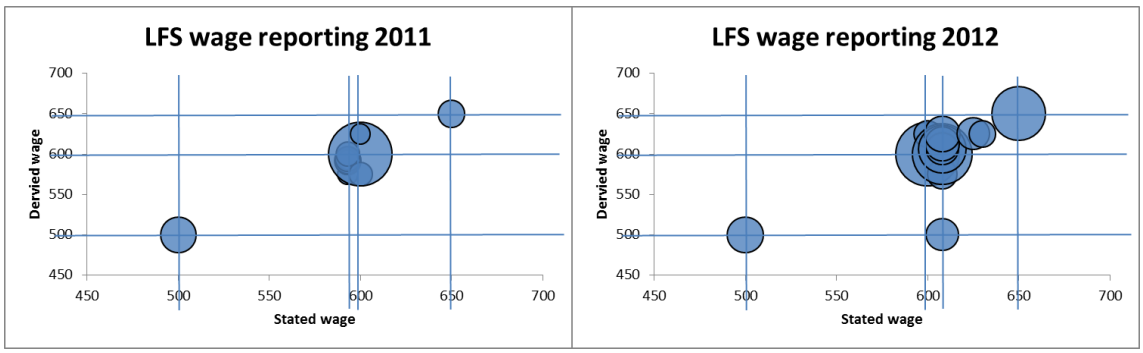
Figure C 1 LFS wage frequencies, 2011 and 2012



As noted in Fry and Ritchie (2013), the stated wage appears to be more accurate than the derived wage. It is concentrated at the NMW (for both youth and adult employees), whereas the derived wage tends to cluster around focus points such as £5.00 and £6.00.

Where both stated and derived wages are observed, the difference is more striking. Employing the blobograms of Fry and Ritchie (2013), where the size of the bubble represents the numbers of observations (minimum five observations), gives Figure C2.

Figure C 2 LFS stated vs derived wage, 2011 and 2012



In 2011, those who reported both stated and derived wages overwhelmingly reported both to be £6.00. In 2012 there was more variation but almost all derived wages came out as £6.00. This is similar to 2006 when stated wages were almost equally located at either £5.00 or the adult NMW of £5.05, but almost all derived wages came out as £5.00 (Ormerod and Ritchie, 2007a, Figure 4; Fry and Ritchie, 2012, Figure 7B).

The relevance is that both wages are used in the imputation procedure. The stated wage has the most influence, and is used to provide 'donor' wages; but where the stated wage is absent or an unlikely value, the derived wage contributes to the predicted value. In 2011 the most popular derived wage was above the most popular stated wage, but in 2012 this situation was reversed. Differences are therefore expected between the two years.

13. Method of investigation

13.1 Choice of treatments

Rather than an extensive examination of the code, the outputs of the MI procedure were examined under four variants:

- The original dataset was maintained – no imputation was done (listed as 'original' on the graphs, below)
- Imputation was carried out, but without outliers from the wage regression removed ('imputed')
- Imputation was carried out, with regression outliers having wage values set to zero to avoid their use as donors ('imputed with treatment')
- The regression equation used to generate the 'nearest neighbours' was altered ('adjusted regression')

The 'imputed with treatment' outcome should be closest to the figures used by the LPC in their reports. However, because there is an element of judgement in the determination of outliers and in the choice of treatment, some differences may occur – as will be seen later.

Because of the difficulty of recoding for previous years, the analysis was only run on 2011 and 2012, Q1, using the current code supplied by the LPC. In 2011 both youth and adult NMWs (£4.92 and £5.93) were both just below major focus points, whereas in 2012 the youth rate (£4.98) was just below a focus point and the adult rate (£6.08) just above.

It proved impossible to separate the weighting from the output process in the SPSS code, so that the 'original' (unadjusted) data appeared to have produce higher rates of low paid. This is because the higher-paid are much more likely to report a weekly or annual wage. Hence, stated hourly wages are concentrated amongst the low paid, and any analysis which only takes account of stated wages will be analysing a very low-wage population. This meant a number of ad hoc adjustments had to be made when dealing with the 'original' data, as described below.

13.2 Regression modelling

At the heart of any imputation is a regression, a statistical model aiming to identify which are the 'nearest neighbours' (that is, observations with most similar characteristics) so that appropriate 'donors' can be found for missing values. To test whether the imputation was sensitive to this core regression, it was decided to make a small set of changes rather than building up a new regression model from scratch. The rationale was that the purpose was to determine the sensitivity of the imputation to the regression specification, rather than find the best possible regression as an aim in itself; this would in any case have been of limited value, as an inspection of regressions from different periods showed that the 'best' model varied over time.

The LPC imputation regression was remodelled by

1. removing the regional markers
2. adding an interactive female-married term

The latter is justified economically and statistically: it is relatively well-established that being married is associated with different wage effects for men and women. The loss of regional effects was justified in that these often appeared to be relatively insignificant, and there was less logic in expecting regional variation compared to, say, missing data spread over industries or occupations.

Table C1 below shows the variables included in both the original regression and the adjusted one. Results are from running the regression on 2011 and 2012 Q1 data, main jobs only. The differences in variables are highlighted.

Table C 1 Alternative imputation regression specifications

	2011 Q1						2012 Q1					
	Original			Adjusted			Original			Adjusted		
	Coeff.	Sig.		Coeff.	Sig.		Coeff.	Sig.		Coeff.	Sig.	
Log hourly earnings	0.425	.000	***	0.430	.000	***	0.416	.000	***	0.418	.000	***
Log hourly earnings, squared	0.134	.000	***	0.136	.000	***	0.142	.000	***	0.144	.000	***
Additions to basic pay	-0.022	.009	***	-0.021	.012	**	-0.024	.009	***	-0.021	.022	**
Age	0.006	.000	***	0.007	.000	***	0.009	.000	***	0.009	.000	***
Age, squared	0.000	.003	***	0.000	.002	***	0.000	.000	***	0.000	.000	***
Months continuously employed	0.000	.000	***	0.000	.000	***	0.000	.000	***	0.000	.000	***
Female	-0.011	.191		-0.009	.352		-0.025	.005	***	-0.009	.391	
Industry 1	-0.023	.588		-0.019	.656		-0.062	.123		-0.065	.105	
Industry 2	0.023	.425		0.023	.442		0.049	.066	*	0.051	.059	*
Industry 3	-0.014	.312		-0.017	.222		0.019	.185		0.014	.349	
Industry 4	0.039	.037	**	0.040	.032	**	0.084	.000	***	0.077	.000	***
Industry 5	-0.053	.000	***	-0.052	.000	***	-0.042	.001	***	-0.042	.001	***
Industry 6	0.026	.093	*	0.025	.098	*	0.033	.053	*	0.035	.042	**
Industry 7	-0.005	.730		-0.003	.833		0.038	.007	***	0.040	.004	***
Industry 9	-0.029	.073	*	-0.028	.079	*	-0.008	.636		-0.007	.672	
Paid by hour or day	-0.572	.000	***	-0.580	.000	***	-0.598	.004	***	-0.598	.004	***
Married	0.008	.250		0.010	.367		0.015	.041	**	0.038	.001	***
Part-Time	-0.017	.024	**	-0.015	.045	**	-0.027	.001	***	-0.025	.002	***
Qualification level 1	0.074	.000	***	0.043	.000	***	0.055	.000	***	0.045	.001	***
Qualification level 2	0.060	.000	***	0.027	.040	**	0.040	.003	***	0.027	.044	**
Qualification level 3	0.034	.000	***	-0.034	.000	***	0.014	.144		-0.013	.157	
Qualification level 5	-0.019	.076	*	-0.051	.000	***	-0.020	.080	*	-0.031	.010	**
Qualification level 6	-0.040	.001	***	-0.075	.000	***	-0.018	.176		-0.031	.023	**
Pay period	-0.033	.000	***	-0.035	.000	***	-0.036	.000	***	-0.037	.000	***
Region 1	-0.033	.039	**				-0.024	.167				
Region 10	-0.056	.001	***				-0.014	.435				
Region 11	-0.036	.009	***				0.003	.843				
Region 12	-0.024	.251					-0.025	.252				
Region 2	-0.036	.005	***				-0.027	.045	**			
Region 3	-0.028	.031	**				-0.012	.357				
Region 4	-0.030	.030	**				-0.023	.115				
Region 5	-0.059	.000	***				-0.029	.051	*			
Region 6	-0.020	.129					-0.011	.480				
Region 7	0.012	.437					0.044	.010	***			
Region 9	-0.015	.263					-0.013	.352				
>25 employees in organisation	0.032	.000	***	0.031	.000	***	0.034	.000	***	0.034	.000	***
Occupational group 1	0.078	.001	***	0.078	.001	***	0.059	.013	**	0.054	.021	**
Occupational group 2	0.228	.000	***	0.225	.000	***	0.212	.000	***	0.208	.000	***
Occupational group 3	0.075	.000	***	0.075	.000	***	0.081	.000	***	0.080	.000	***
Occupational group 5	0.061	.000	***	0.059	.000	***	0.010	.534		0.011	.518	
Occupational group 6	-0.035	.011	**	-0.037	.007	***	-0.026	.074	*	-0.028	.058	*
Occupational group 7	-0.048	.001	***	-0.052	.000	***	-0.046	.002	***	-0.048	.002	***
Occupational group 8	0.000	.984		-0.002	.915		-0.043	.011	**	-0.046	.007	***
Occupational group 9	-0.053	.000	***	-0.056	.000	***	-0.073	.000	***	-0.073	.000	***
Last pay is same as usual	-0.254	.000	***	-0.249	.000	***	-0.280	.000	***	-0.286	.000	***
Teenager NMW band	-0.107	.000	***	-0.103	.000	***	-0.132	.000	***	-0.133	.000	***
Youth NMW band	-0.039	.023	**	-0.033	.053	*	-0.011	.562		-0.011	.560	
Not in permanent post	0.025	.033	**	0.027	.023	**	0.029	.031	**	0.031	.020	**
Ever work overtime	0.040	.000	***	0.041	.000	***	0.054	.000	***	0.053	.000	***
Usual gross pay, if being paid it	0.120	.000	***	0.118	.000	***	0.140	.000	***	0.143	.000	***
Female and married				-0.003	.804					-0.035	.012	**
Adjusted R-squared	0.766			0.764			0.701			0.701		

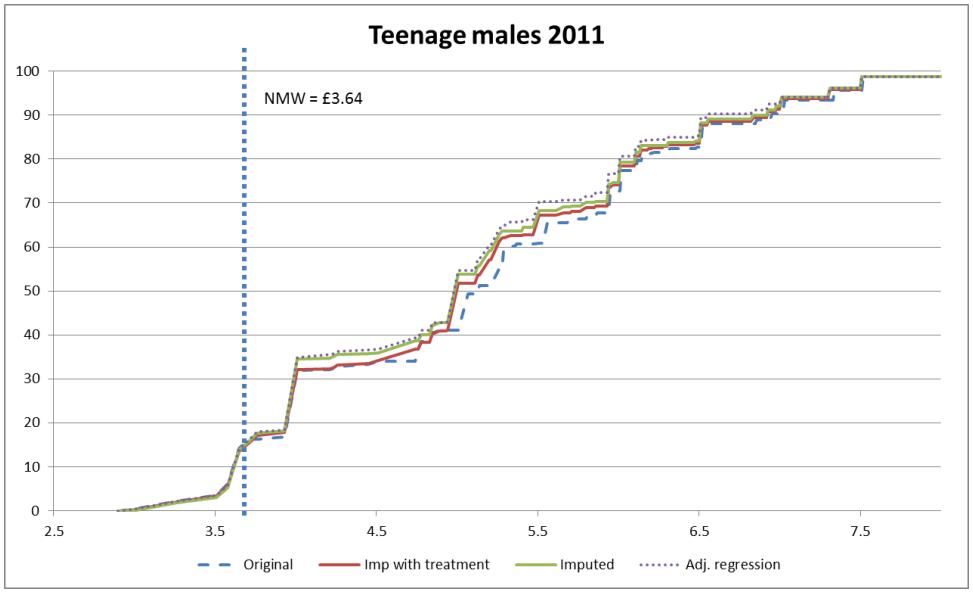
Overall the performance of the regressions with and without the adjustments is similar. Making the adjustments appears to have the most effect on the estimated coefficients for qualifications, which is sensible in the context of the literature on gender pay effects and qualifications.

14. Impact of imputation on wage distributions

14.1 Calculating the wage distributions

The main output generated by the code is a set of "1p tables", which contain cumulative numbers earning below each 1p wage level. The analysis of the tables allowed us to see whether the observed distributions were sensitive the imputation methods. An example is shown below in figure C3.

Figure C 3 Example finding from imputation runs



This graph shows the cumulative numbers being paid a particular wage; for example, around 15% of teenage males were being paid at or below the relevant NMW of £3.64 in 2011. The 'original' line has no imputation; the 'imputed with treatment' adjusts for unlikely values from the regression; the 'imputed' takes the regression nearest neighbours without any treatment for exceptional values; and the 'adj. regression' uses the alternative regression specification, again without treatment for extreme values.

These proportions are calculated with respect to the total working population. As noted above, the lack of stated wages in the source data means that the population for the 'original' distribution appears to be far smaller, and concentrated in the low-wage sector; hence, proportions covered by each wage are much higher than in the imputed data. To produce results consistent with the imputed data, the unadjusted results were rescaled to the average denominator in the three sets of adjusted results, so that the 'original' data show the same population proportions being paid below £10.00 as the imputed estimates[3]. Thus the 'original' series should not be taken as the actual LFS proportions. The key element is the shape of the distributions, particularly around the NMW.

For clarity, the NMW lines are not depicted on all graphs below. The axes have been adjusted to focus on detail while keeping the scale constant within age groups. For teenage and youth workers, the left hand scale is from 0%-100%; for adults, the scale is from 0%-70%. The horizontal scale runs from £2.50-£8.00 (teenagers), £3.50-£8.00 (young workers) and £4.50-£7.00 for adults.

It should be noted that although some graphs have very small numbers of observations, this reflects the small number eligible for the imputation process, not the original source data. For example, in one case, only one record is usable for modelling despite the fact that at least five other workers have the same characteristics apart from a valid

[3] An alternative would have been to use the wage rate created from derived wages (earnings divided by hours). This was rejected as the aim here was to see how much 'real' information was being used.

stated wage rate; but all have valid derived wages. Again, this emphasises that the imputation code can run on a very small subset of admissible observations.

Only selected individual results are presented below, for illustrative purposes; the full set of results for 2011 and 2012 is available from the LPC. Note that the data here uses both first and second jobs for multiple job holders, in line with LPC calculations.

14.2 Overall impacts

Figures C4 and C5 show the overall cumulative rates for those earning below a specified value, for males and females and by NMW age band, in 2011 and 2012. Overall, imputation and the specific outlier treatment appear to have little effect, in that the break points are similar even if the specific rates are slightly different.

Figure C 4 Overall wage proportions, 2011

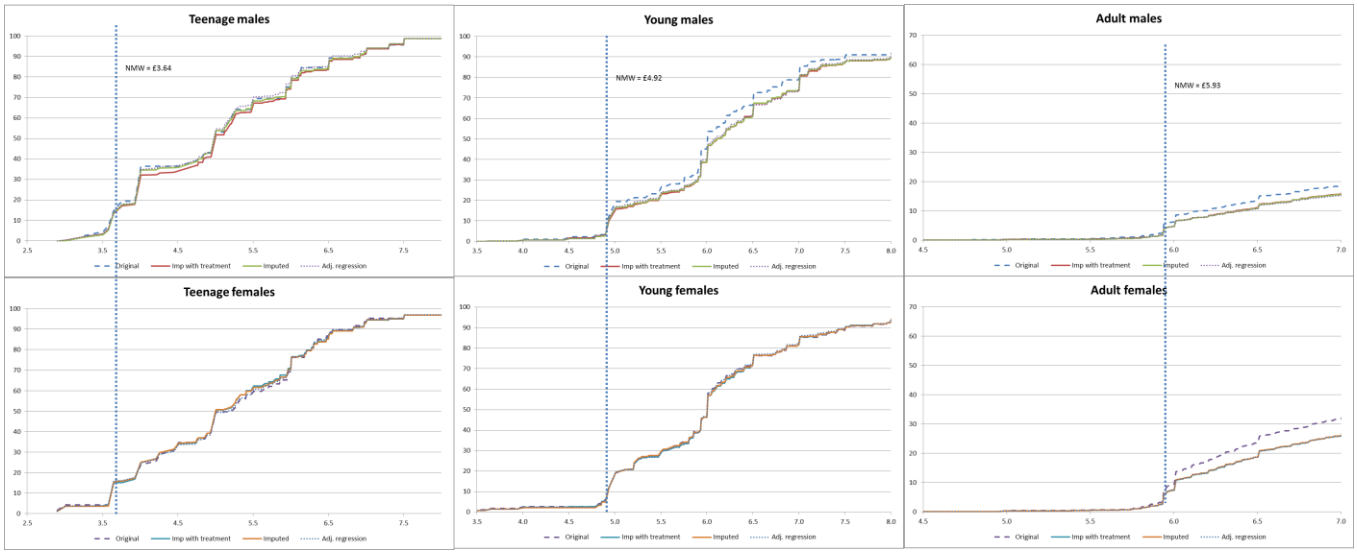
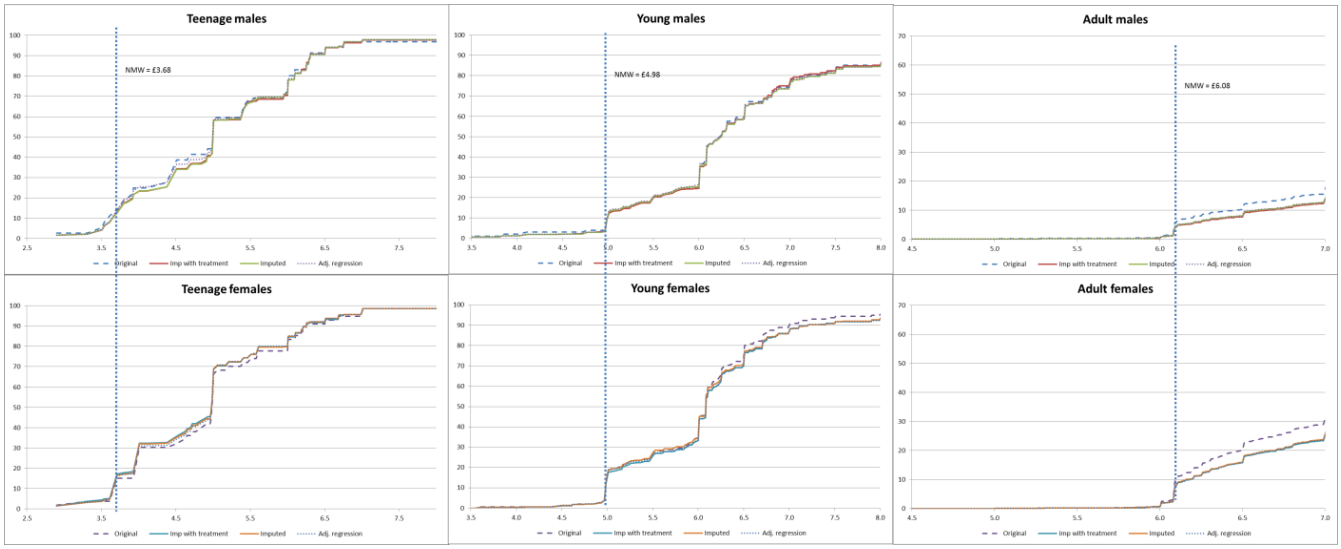


Figure C 5 Overall wage proportions, 2012



For teenagers and young workers, there is negligible difference between the original and imputed rates around the NMW; above the NMW (and at all wages for adults) imputation appears to reduce the proportions of the low paid. The population weights for the 'original' data are constrained to give the same proportion of the population earning below £10 as in the imputed distributions; this therefore suggests that the imputed data identifies fewer low-paid adults at the bottom of the distribution. The LPC's formal analysis consistently shows that imputed LFS data produces a higher proportion of adult minimum wage workers than the ASHE data over time; however, the imputation methodology has reduced the differences between hourly earnings estimates from the LFS and ASHE.

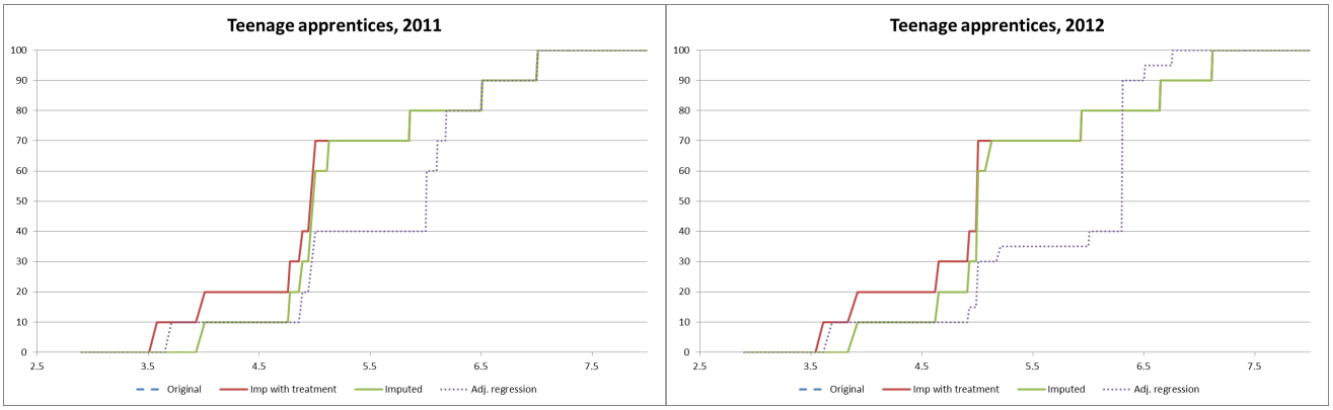
Apart from this, the impact of imputation does not appear significant, positive or negative. There is little difference between outcomes in 2011 and 2012, despite the presence of 'underestimation' and NMW workers in 2011 and 'overestimation' of adults in 2012. This is a recurrent theme: that the imputation replicates the distribution regardless of whether that distribution is the 'true' one or not. Therefore, measurement error around the NMW is not corrected by the imputation, but nor is it made notably worse.

If the impact of imputation overall does not seem great, nor is the importance of the type of imputation. All three variations produce similar results, in both years. Thus the imputation procedure appears relatively robust.

14.3 Analysis of subgroups

The distributions are less robust when sub-divisions of the data are analysed; small numbers become important, and some distributions are based upon no valid information at all. Consider the breakdown of qualifications into seven categories (NQF levels 2-4, apprentice, below NQF2, other and unqualified). Figure C6 depicts the number of apprentices in each year as an example.

Figure C 6 Proportions, teenage apprentices



The imputed distributions suggest significant numbers of apprentices, and present a convincing distribution for them suggesting that all of them earn £7.00 or less. However, there are no teenage apprentices in either year with valid stated hourly wages in the original data, and this distribution is therefore based wholly on either older workers or teenagers who are not apprentices. A similar result is found for teenagers born outside the UK. The distributions are plausible, but it is difficult to argue that they are substantially based on real data.

A similar story holds for disability. In neither period were there any valid stated wage rates for teenagers with work-limiting disabilities, or young workers defined as disabled under the DDA; but in 2012 the latter were given distributions. It is not immediately clear what has led to this result, especially as, within each year, the three imputation methods stepped the distribution in similar but slightly different ways.

The education breakdowns show very similar patterns over the two years; this may be because education is correlated closely with many of the other variables in the regression, and so 'nearest neighbours' will be of similar education groups. However, the migration breakdowns (by UK/non-UK born) show a similar consistency, and this is less easy to rationalise. For the other breakdowns, the large numbers in the 'adult' groups mean that these stay similar across years, but the teen/youth subgroups show much more variation.

One area where imputation repeatedly and noticeably makes a difference is in ethnicity. Minority ethnic groups are often low in number, particularly for younger workers. Hence, for example, distributions for Chinese workers under 21 are almost entirely imputed, and so the difference between the adjusted regression and the other imputations is noticeable. This is also found in the other ethnic groups, although not consistently across sub-groups or years (the adjusted regression appears to produce more variation in 2012).

14.4 Summary of analysis of distributions

It was suggested above that there were three potential errors with the data: not recognising the asymmetry caused by the existence of the NMW, small numbers of observations, and the choice of interventions available to the operator.

The symmetry of the MI code is demonstrated in the replacement of stepped functions with smoother distributions. One notable result stands out: in all the cases with very small numbers, the imputation procedure does not generate new observations below the relevant NMW. This is likely to be the result of the small numbers of individuals, so that the probability of getting very low wages is negligible. In contrast, where there are many observations, the MI process does impute new observations with wages below the NMW.

Overall, the two imputation methods using the same regression show similar outcomes, with occasional small variations; the choice of whether to intervene manually appears to make relatively little substantive difference. Imputation using a different regression specification shows more variation. The 'original' is less smooth, as it has fewer observations.

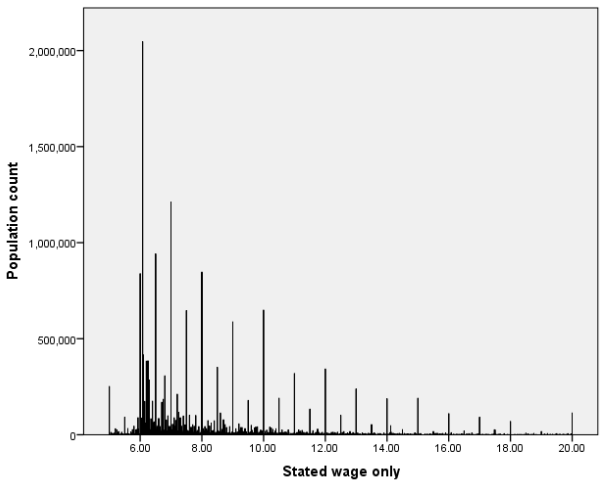
Where there are large numbers of observations, the imputation method makes little difference. Small numbers (for example, in ethnic groups) increase the variation. For youths and teens, the small numbers across combined domains (for example young/disabled) mean that most of the estimates are sensitive to the estimation method; for adults, imputed distributions are relatively stable.

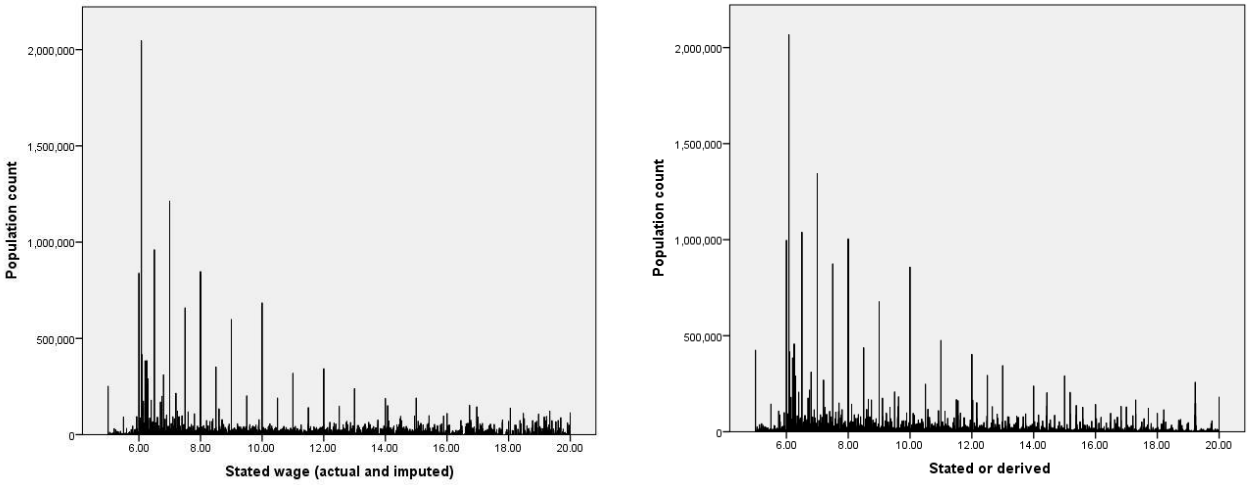
15. Impact of imputation of LPC statistics

The LPC uses the imputed code to generate a number of measures of impact, specifically those concerning the 'bite' of the NMW and estimates of the number of jobs paid below the NMW. The team were asked to consider the impact of different imputation methods on these statistics.

For this section, the 'original' data is constructed by taking the stated hourly wage rate or, if no stated wage is available, using the derived hourly wage rate. The rationale for this can be clearly seen in Figure C7.

Figure C 7 Distribution of stated, derived and imputed wages in the LFS





This depicts distributions for 2012 data showing stated wages only, hour wages when missing stated wages are replaced by the derived wages (the 'original' data) and wages imputed using the model closest to the LPC method. It is clear that stated-wages-only significantly under-estimates the number of higher paid, whereas replacing missing stated wages with the derived wage is a fair approximation of the imputed wage distribution. Using this stated-plus-derived combination also has some justification in that the derived wage plays an important role in the imputation regression. Replacing missing stated wages with derived wages is also a popular workaround for researchers.

15.1 Imputation and the NMW 'bite'

The 'bite' of the NMW is its value relative to a point on the earnings distribution, measuring how much of the workforce is affected by the NMW. For example, supposing the NMW is £6.00 and the 25th percentile's wage is £7.00 (25% of the population earn £7.00 or under), then the bit is 85.7% (£6/£7). A 'bite' of 100% at median implies half of the workforce in that group are earning at or below the minimum wage.

The bite is calculated using four quarters of the NMW year (Oct-Dec, Jan-Mar, Apr-Jun, Jul-Sep). Hence, LFS data may show lower values than ASHE which only measures earnings in April, by which time wages largely caught up with the NMW changes (see Ormerod and Ritchie, 2007a). In the analysis below, wages are calculated under different imputation strategies. 'Published' is the value estimated by LPC and used in its 2012/13 reports. 'Original' is stated-plus-derived, as described above. 'Imputed' is fully imputed. 'Adj. regression' is the model with the adjusted imputation regression. The 'imputed' value should be closest to the published value, although the scope for human intervention in the imputation process can lead to some differences even if the 'imputed' and 'published' calculations are the same in principle. 'Imputed' is therefore taken as the base, rather than 'published', as the interventions made to generate the 'original' and 'adj. Regression' values are variations on the 'imputed' estimate.

Table C2 describes the wages at the median; that is, the lower-earning half of the population earns this wage or less:

Table C 2 Median wages under alternative models

		Medians													
		2011							2012						
		Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*Imp vs adj*	Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*Imp vs adj*
Sex	Male	11.49	11.50	11.34	11.45	*15*	*16*	*11*	11.37	11.55	11.49	10.83	*-12*	*6*	*-66*
	Female	9.30	9.35	9.21	9.23	*9*	*14*	*2*	9.38	9.50	9.34	8.92	*3*	*16*	*-42*
Ethnicity	White	10.33	10.40	10.35	10.26	*-2*	*5*	*-9*	10.31	10.50	10.29	9.88	*2*	*21*	*-41*
	Ethnic Minority	9.98	10.00	10.00	9.79	*-2*	*0*	*-21*	10.00	10.00	9.92	9.17	*8*	*8*	*-75*
Quals	Any qualification	10.70	10.71	10.66	10.61	*4*	*5*	*-5*	10.57	10.83	10.61	10.00	*-3*	*23*	*-61*
	No qualification	7.00	7.00	6.98	6.98	*2*	*2*	*-1*	7.03	7.00	7.00	7.00	*3*	*0*	*0*
Disability	No disability	10.50	10.50	10.45	10.40	*5*	*5*	*-5*	10.43	10.63	10.43	9.99	*0*	*20*	*-44*
	Disability	9.50	9.61	9.44	9.50	*6*	*17*	*6*	9.57	9.63	9.54	9.12	*3*	*9*	*-42*
Migration	UK Born	10.43	10.43	10.39	10.34	*3*	*4*	*-5*	10.37	10.58	10.35	9.95	*2*	*23*	*-40*
	Non-UK Born	9.62	9.61	9.50	9.47	*12*	*11*	*-3*	9.77	9.75	9.66	9.00	*11*	*9*	*-66*
Nationality	UK Nationality	10.50	10.50	10.46	10.40	*4*	*4*	*-6*	10.43	10.63	10.43	10.00	*1*	*21*	*-43*
	Non-UK Nationality	9.00	8.97	8.95	8.89	*5*	*2*	*-6*	9.04	9.15	9.12	8.67	*-8*	*3*	*-45*
Age	21 - 54	10.40	10.38	10.35	10.25	*5*	*3*	*-10*	10.33	10.50	10.30	9.87	*3*	*20*	*-43*
	55 - 64	9.93	10.06	10.00	10.00	*-7*	*6*	*0*	10.01	10.28	10.00	9.62	*1*	*28*	*-38*
Overall		10.28	10.33	10.27	10.19	*1*	*6*	*-8*	10.29	10.48	10.26	9.84	*3*	*22*	*-42*
Detailed ethnicity	White	10.33	10.40	10.35	10.26	*-2*	*5*	*-9*	10.31	10.50	10.29	9.88	*2*	*21*	*-41*
	Black	9.89	10.00	10.07	9.97	*-19*	*-7*	*-10*	9.50	9.63	9.50	8.92	*0*	*13*	*-58*
	Indian	11.17	11.11	11.49	11.60	*-32*	*-38*	*11*	11.55	12.03	12.00	10.12	*-46*	*3*	*-188*
	Pakistani	8.47	8.11	8.40	8.30	*7*	*-29*	*-11*	9.17	8.25	8.35	8.12	*82*	*-10*	*-23*
	Bangladeshi	7.59	7.50	7.50	7.49	*9*	*0*	*-2*	7.83	7.50	7.51	6.80	*33*	*-1*	*-71*
	Other non-white	10.03	9.96	9.59	9.58	*44*	*37*	*-1*	10.00	10.00	10.00	9.36	*0*	*0*	*-64*

The first four columns for each year are the wages estimated at the median. The next three columns compare each of the modelled wages with the imputed value. Particularly large or small changes or highlighted in blue (below the imputed value) or red (above). A darker colour means more variation (5p-10p difference, and over 10p difference).

As expected, the imputed and published values are generally similar; the differences are measured in pennies. However, there is much more variation in small groups: the disabled, non-UK born, and ethnic minorities. This is not entirely surprising, as the small numbers mean a single missing observation can make a large difference to the step in wages between employees.

The difference between 2011 (2010Q4-2011Q3) and 2012 (2011Q4-2012Q3) is striking. The latter shows much more variation between the imputation methods, with the 'adjusted regression' continually showing a lower median and 'original' showing a higher one. The latter might be explained by the wider variation in derived wages used to fill in the gaps in the distribution; but it is not at all clear why changing the imputation regression should, in effect, create noticeably larger numbers at the bottom of the wage distribution. The distributional analysis of the previous section did show more variation on the adjusted regression, but not obviously always below. Of more concern is that in 2012 the imputation methods affect the median wages even in the larger groups (white, qualified, no disability etc). Again, it is not clear why. A reasonable hypothesis might be that it is the result of the adult NMW being just above or below a focus point, but it is not obvious for the above table that adjusted regression estimates are any more or less likely to produce median wages on focal points.

Table C3 presents the same information for the bottom decile (the lowest-earning tenth of the working population), and the impact of a much more compressed distribution is easily seen:

Table C 3 10th percentile wages under alternative models

	10th percentile													
	2011							2012						
	Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*Imp vs adj*	Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*Imp vs adj*
Male	6.63	6.25	6.55	6.55	*8*	*-30*	*0*	6.70	6.39	6.67	6.50	*3*	*-28*	*-17*
Female	6.18	6.00	6.15	6.15	*3*	*-15*	*0*	6.30	6.09	6.28	6.23	*1*	*-19*	*-5*
White	6.36	6.05	6.34	6.33	*2*	*-29*	*-1*	6.48	6.20	6.45	6.36	*3*	*-25*	*-10*
Ethnic Minority	6.15	5.93	6.10	6.10	*5*	*-17*	*0*	6.29	6.08	6.30	6.20	*-1*	*-22*	*-10*
Any qualification	6.45	6.11	6.40	6.40	*5*	*-28*	*0*	6.50	6.25	6.50	6.41	*0*	*-25*	*-9*
No qualification	5.93	5.80	5.93	5.93	*0*	*-13*	*0*	6.08	6.00	6.08	6.08	*0*	*-8*	*0*
No disability	6.36	6.04	6.32	6.31	*4*	*-28*	*-1*	6.50	6.20	6.46	6.35	*4*	*-26*	*-11*
Disability	6.20	6.00	6.19	6.18	*1*	*-19*	*-1*	6.31	6.10	6.30	6.27	*1*	*-20*	*-3*
UK Born	6.38	6.08	6.35	6.35	*3*	*-27*	*0*	6.50	6.25	6.50	6.40	*0*	*-25*	*-9*
Non-UK Born	6.07	5.93	6.04	6.05	*3*	*-11*	*1*	6.23	6.08	6.23	6.18	*0*	*-15*	*-5*
UK Nationality	6.38	6.08	6.35	6.36	*3*	*-27*	*0*	6.50	6.25	6.50	6.40	*0*	*-25*	*-10*
Non-UK Nationality	6.04	5.93	6.00	6.00	*4*	*-7*	*0*	6.20	6.08	6.15	6.13	*5*	*-7*	*-2*
21 - 54	6.31	6.00	6.29	6.28	*3*	*-29*	*-1*	6.45	6.17	6.41	6.31	*4*	*-24*	*-10*
55 - 64	6.37	6.10	6.40	6.40	*-3*	*-30*	*0*	6.50	6.26	6.50	6.46	*0*	*-24*	*-4*
	6.32	6.00	6.30	6.30	*2*	*-30*	*0*	6.46	6.20	6.43	6.34	*3*	*-23*	*-9*
White	6.36	6.05	6.34	6.33	*2*	*-29*	*-1*	6.48	6.20	6.45	6.36	*3*	*-25*	*-10*
Black	6.31	6.05	6.28	6.28	*3*	*-23*	*0*	6.41	6.09	6.41	6.38	*0*	*-32*	*-3*
Indian	6.25	5.93	6.15	6.15	*10*	*-22*	*0*	6.40	6.15	6.38	6.25	*2*	*-23*	*-13*
Pakistani	6.00	5.80	6.00	6.00	*0*	*-20*	*0*	6.10	6.00	6.10	6.10	*0*	*-10*	*0*
Bangladeshi	5.95	5.80	6.00	6.00	*-5*	*-20*	*0*	6.08	5.83	6.08	6.08	*0*	*-25*	*0*
Other non-white	6.10	5.93	6.05	6.05	*5*	*-12*	*0*	6.30	6.08	6.28	6.23	*2*	*-20*	*-5*

At the 10th percentile there is little variation in 2011 between imputation methods, with the exception of the 'original' wages. Ormerod and Ritchie (2007b) noted that the derived rate in the LFS seemed to show a more 'infeasible' distribution and this may be reflected in the results here. In 2012 the adjusted regression shows more variation; again this seems to be more prevalent in the larger groups.

Tables C4 and C5 give the bite at the median and the bottom decile. As before, negative differences are in blue and positive ones in red, with the boundary between light and dark set at 0.5% and 1% difference. The reason for these limits is that the bite is a sensitive issue, with increase or decreases being scrutinised as evidence of whether the NMW is having more or less effect in the labour market. Hence, a small variation may be enough to trigger concerns about LPC recommendations.

Table C 4 Median bite under alternative models

		Median bite													
		2011							2012						
		Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*imp vs adj*	Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*imp vs adj*
Sex	Male	51.6	51.6	52.3	51.8	*-0.7*	*-0.7*	*-0.5*	53.5	52.6	52.9	56.1	*0.6*	*-0.3*	*3.2*
	Female	63.8	63.4	64.4	64.2	*-0.6*	*-1.0*	*-0.2*	64.9	64.0	65.1	68.2	*-0.2*	*-1.1*	*3.1*
Ethnicity	White	57.4	57.0	57.3	57.8	*0.1*	*-0.3*	*0.5*	59.0	57.9	59.1	61.5	*-0.1*	*-1.2*	*2.4*
	Ethnic Minority	59.4	59.3	59.3	60.6	*0.1*	*0.0*	*1.3*	60.8	60.8	61.3	66.3	*-0.5*	*-0.5*	*5.0*
Quals	Any qualification	55.4	55.4	55.6	55.9	*-0.2*	*-0.2*	*0.3*	57.5	56.1	57.3	60.8	*0.2*	*-1.2*	*3.5*
	No qualification	84.7	84.7	84.9	85.0	*-0.2*	*-0.2*	*0.1*	86.5	86.9	86.9	86.9	*-0.3*	*0.0*	*0.0*
Disability	No disability	56.5	56.5	56.7	57.0	*-0.3*	*-0.3*	*0.3*	58.3	57.2	58.3	60.9	*0.0*	*-1.1*	*2.6*
	Disability	62.4	61.7	62.8	62.4	*-0.4*	*-1.1*	*-0.4*	63.5	63.1	63.7	66.7	*-0.2*	*-0.6*	*3.0*
Migration	UK Born	56.9	56.9	57.1	57.4	*-0.2*	*-0.2*	*0.3*	58.6	57.5	58.7	61.1	*-0.1*	*-1.3*	*2.4*
	Non-UK Born	61.7	61.7	62.4	62.6	*-0.8*	*-0.7*	*0.2*	62.2	62.4	62.9	67.6	*-0.7*	*-0.6*	*4.6*
Nationality	UK Nationality	56.5	56.5	56.7	57.0	*-0.2*	*-0.2*	*0.3*	58.3	57.2	58.3	60.8	*0.0*	*-1.1*	*2.5*
	Non-UK Nationality	65.9	66.1	66.2	66.7	*-0.3*	*-0.1*	*0.4*	67.3	66.4	66.6	70.1	*0.6*	*-0.2*	*3.5*
Age	21 - 54	57.0	57.1	57.3	57.9	*-0.3*	*-0.2*	*0.6*	58.9	57.9	59.0	61.6	*-0.2*	*-1.1*	*2.6*
	55 - 64	59.7	58.9	59.3	59.3	*0.4*	*-0.4*	*0.0*	60.7	59.1	60.8	63.2	*-0.1*	*-1.7*	*2.4*
Overall		57.7	57.4	57.8	58.2	*-0.1*	*-0.4*	*0.4*	59.1	58.0	59.3	61.8	*-0.2*	*-1.2*	*2.5*
Detailed ethnicity	White	57.4	57.0	57.3	57.8	*0.1*	*-0.3*	*0.5*	59.0	57.9	59.1	61.5	*-0.1*	*-1.2*	*2.4*
	Black	60.0	59.3	58.9	59.5	*1.1*	*0.4*	*0.6*	64.0	63.1	64.0	68.1	*0.0*	*-0.9*	*4.1*
	Indian	53.1	53.4	51.6	51.1	*1.5*	*1.8*	*-0.5*	52.7	50.5	50.7	60.1	*2.0*	*-0.1*	*9.4*
	Pakistani	70.0	73.1	70.6	71.5	*-0.6*	*2.6*	*0.9*	66.3	73.7	72.8	74.9	*-6.5*	*0.9*	*2.1*
	Bangladeshi	78.2	79.1	79.1	79.2	*-0.9*	*0.0*	*0.2*	77.6	81.1	81.0	89.4	*-3.4*	*0.1*	*8.4*
	Other non-white	59.1	59.5	61.8	61.9	*-2.7*	*-2.3*	*0.1*	60.8	60.8	60.8	64.9	*0.0*	*0.0*	*4.1*

Table C 5 10th percentile bite under alternative models

		10th percentile bite													
		2011							2012						
		Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*imp vs adj*	Published	Original	Imputed	Adj. regression	*Imp vs pub*	*Imp vs orig*	*imp vs adj*
Sex	Male	89.4	94.9	90.5	90.6	*-1.1*	*4.3*	*0.0*	90.7	95.1	91.1	93.5	*-0.4*	*4.0*	*2.4*
	Female	96.0	98.8	96.4	96.4	*-0.5*	*2.4*	*0.0*	96.6	99.8	96.8	97.6	*-0.2*	*3.0*	*0.8*
Ethnicity	White	93.3	98.0	93.6	93.7	*-0.3*	*4.4*	*0.1*	93.8	98.1	94.2	95.6	*-0.4*	*3.9*	*1.4*
	Ethnic Minority	96.4	100.0	97.2	97.2	*-0.8*	*2.8*	*0.0*	96.7	100.0	96.5	98.1	*0.2*	*3.5*	*1.6*
Quals	Any qualification	91.9	97.1	92.7	92.7	*-0.8*	*4.3*	*0.0*	93.5	97.3	93.5	94.9	*0.0*	*3.7*	*1.3*
	No qualification	100.0	102.2	100.0	100.0	*0.0*	*2.2*	*0.0*	100.0	101.3	100.0	100.0	*0.0*	*1.3*	*0.0*
Disability	No disability	93.2	98.2	93.8	93.9	*-0.6*	*4.3*	*0.1*	93.5	98.1	94.1	95.7	*-0.5*	*4.0*	*1.7*
	Disability	95.6	98.8	95.8	96.0	*-0.2*	*3.0*	*0.2*	96.4	99.7	96.5	97.0	*-0.2*	*3.2*	*0.5*
Migration	UK Born	92.9	97.5	93.4	93.4	*-0.4*	*4.2*	*0.0*	93.5	97.3	93.6	95.0	*-0.1*	*3.7*	*1.4*
	Non-UK Born	97.6	100.0	98.1	98.0	*-0.5*	*1.9*	*-0.1*	97.6	100.0	97.6	98.4	*0.0*	*2.4*	*0.7*
Nationality	UK Nationality	92.9	97.5	93.3	93.3	*-0.4*	*4.2*	*0.0*	93.5	97.3	93.5	95.0	*0.0*	*3.7*	*1.5*
	Non-UK Nationality	98.2	100.0	98.8	98.8	*-0.7*	*1.2*	*0.0*	98.1	100.0	98.9	99.2	*-0.8*	*1.1*	*0.3*
Age	21 - 54	93.9	98.8	94.4	94.5	*-0.4*	*4.5*	*0.1*	94.3	98.5	94.9	96.4	*-0.5*	*3.7*	*1.5*
	55 - 64	93.1	97.2	92.7	92.7	*0.4*	*4.6*	*0.0*	93.5	97.1	93.5	94.1	*0.0*	*3.6*	*0.6*
Overall		93.8	98.8	94.1	94.1	*-0.3*	*4.7*	*0.0*	94.1	98.1	94.6	95.9	*-0.5*	*3.5*	*1.3*
Detailed ethnicity	White	93.3	98.0	93.6	93.7	*-0.3*	*4.4*	*0.1*	93.8	98.1	94.2	95.6	*-0.4*	*3.9*	*1.4*
	Black	93.9	98.0	94.4	94.4	*-0.5*	*3.6*	*0.0*	94.9	99.8	94.8	95.3	*0.0*	*5.0*	*0.5*
	Indian	94.9	100.0	96.4	96.4	*-1.5*	*3.6*	*0.0*	95.0	98.9	95.3	97.3	*-0.3*	*3.6*	*2.0*
	Pakistani	98.8	102.2	98.8	98.8	*0.0*	*3.4*	*0.0*	99.7	101.3	99.7	99.7	*0.0*	*1.7*	*0.0*
	Bangladeshi	99.7	102.2	98.8	98.8	*0.8*	*3.4*	*0.0*	100.0	104.3	100.0	100.0	*0.0*	*4.3*	*0.0*
	Other non-white	97.2	100.0	98.0	98.0	*-0.8*	*2.0*	*0.0*	96.6	100.0	96.8	97.6	*-0.3*	*3.2*	*0.8*

The variation in earnings has relatively little effect on the bite either at the median or the bottom decile. The only consistent impact comes from the 'original' data inthe bottom decile. Augmenting hourly wages with derived wages leads to the bite typically increasing by around 4%. However, as noted above, in 2012 the adjusting the imputation regression has a more notable impact. It increases the bite by around 3% at the median, and 1.5% at the bottom decile.

Several times the 'original' results generate a 10th percentile bite of over 100%, suggesting either widespread non-compliance or measurement error. If concerns over the accuracy of the derived wage variable in the LFS are justified, this would favour the latter explanation and argue against the use of a simple derived-unless-stated wage imputation.

In summary, these results suggest that imputation does affect wages at different points of the distribution, and hence the bite. Replacing missing stated hourly wages with derived wages has the largest impact, consistently under-estimating wages and over estimating the bite compared to the preferred method. Perhaps of more concern is the finding that the impact of regression adjustments is not consistent over time: sometimes it has very little impact, sometimes much more, and it is not immediately obvious why.

15.2 Imputation and the number of low-paying jobs

The LPC produces estimates of the number of workers in low paid jobs; that is, those being paid at no more than 5p over the appropriate minimum wage. Table C6 below reproduces the published figures for 2012, plus data calculated using the different imputation methods for 2012 Q2. The last column shows the values imputed over the four quarters of NMW year 2012.

Table C 6 Low-paid workers, published and estimated

	LPC		Own calculations			
	ASHE	LFS	Original	Imputed	Adj. Reg.	Imputed 4Q
Male	4.2	5.5	7.9	3.9	3.9	4.4
Female	6.3	8.9	11.2	6.8	6.6	7.5
White			9.4	5.3	5.2	5.8
Ethnic Minority		7.7	11.0	5.5	5.3	7.5
Any qualification			8.8	4.9	4.8	5.3
No qualification		18.1	21.7	12.0	12.5	16.1
No disability			9.4	5.2	5.2	5.8
Disability		9.2	10.3	5.8	5.8	7.0
UK Born			9.2	5.1	5.1	5.5
Non-UK Born			11.3	6.6	6.3	8.6
UK Nationality			9.1	4.9	4.9	5.4
Non-UK Nationality		8.9	12.5	8.3	7.9	9.8
<18	14.2	12.3	20.2	8.7	8.7	11.6
18-20	15.5	19.8	25.8	18.7	18.3	14.3
21+			8.7	4.7	4.7	5.5
Older	8.0		22.2	9.1	14.2	13.3
Working age	5.2	7.2	8.6	5.3	4.6	5.4

There is a large variation in these figures. LPC estimates from the LFS, while larger than ASHE, are consistent over time and generally show higher rates for low-paying jobs. The calculations shown here, using the same data source as above but not the LPC code, shows much more variation and are not consistently higher than ASHE. Nevertheless, it is clear that the two imputation methods produce broadly similar numbers except when considering the age of

workers. As before, supplanted stated wages with derived wages produces clear outliers. The LPC calculations are done in Q2 to provide comparability with ASHE. As the last column shows, looking at the number of low-paid jobs over a NMW year produces a higher estimate as wages increase over the year and non-compliance falls.

16. Discussion of imputation findings

Imputation is hard to validate, given that its purpose is to replicate information that does not exist. The LPC is keen to retain the imputation code, given that (1) LFS is the only source of information on the subgroups of interest and very small numbers in each group mean that gaps become significant (2) in groups where LFS and ASHE are comparable, LFS appears to generate notably different results. Not using imputation means that, while the proportions in sectors are based only on 'known' data, they are more likely to be volatile from year to year.

Initially it should be noted that the imputation code is, in statistical terms, ancient. This does not necessarily make it unfit for purpose, but it suggests there is scope for improvement in the light of changes to data collection and new statistical knowledge. For comparison, ONS' methodology guides suggest that processes be reviewed every five years or so.

Whether the imputation adds to the understanding of low pay is a moot point.

The changes introduced by the full imputation process are relatively small but can have disproportionate effects. The LFS is a large dataset in the context of UK surveys, but it is still small compared to the population. A figure for the number of low paid is an estimate, with the accuracy of that estimate determined by the sampling method, the response rate and the weighting procedure. The use of imputed values implies that confidence intervals should be increased by the predictive error of the imputed values. MI seeks to reduce this predictive error by taking an average across several different possible candidates. This works well when there are many possible donors (for example, estimating a missing value for a white middle-aged office worker) but can increase the error when there are very few possible candidates (17-year olds with higher education certificates, for example).

Each working individual in the LFS with valid income data is weighted to represent around 2,300 workers in the UK (the weight for all LFS respondents to represent the UK population is around 230). Single observations generated by the imputation procedure can therefore produce seemingly significant numbers. Although not calculated as part of this analysis, the suspicion is that any improvement in mean statistics due to MI is outweighed by the increase in uncertainty surrounding those estimates. In other words, the MI may be introducing spurious accuracy, particularly for the special cases of interest to the LPC.

Of more concern is the quality of the regression model itself. This is estimated over the whole dataset, and therefore approximates well to individuals with common sets of characteristics. However, the cases of interest to LPC may not share those characteristics. For example, 'married' is one of the explanatory variables in the prediction regression, but there are no married white individuals of either sex under the age of 19 in one of the quarters studied(with or without valid wage rates), and none for other ethnic groups under 21. The coefficient associated with marriage is therefore dominated by the wage profiles of older white workers.

In theory the linear regression model used allows for this: all coefficients are estimated "all other things being equal". If the regression model is correct, then rare combinations of characteristics do not matter as the independence of the variables guarantees an accurate estimate of the mean effect. However, this assumption of independent variables may not be correct.

Apart from removing (frequently, but not always) insignificant region markers, the only change to create the 'adjusted regression' was to add a variable interacting gender and marriage. As noted in the Interim Report, it is a standard result in labour economics that the impact of marriage, children and other family-related markers is closely tied to the gender of the respondent. Adding the interactive term was thus justified on economic grounds, and it was (frequently, but not always) significant at the 1% level.

It is perhaps not surprising that a variable addressing an element of family structures should have most impact when the data are broken down into ethnic subgroups. It is of more concern that such a simple change produced a notable effect. One could envisage additional variables that could be associated with the sub-groups; for example, adding language skills (or whether an EU citizen) might be expected to produce different estimates for migrant groups. As the characteristics of particular interest to the LPC (ethnicity, education, disability, migrant status, age) all enter the regression model independently, it is feasible that imputation results are much more sensitive to the regression model.

As noted in the Interim Report, there is always an element of judgment in designing a regression model. The criticisms could be levelled at any regression model; what is important is whether this makes any difference to the outcome. However, the core regression model does not appear to have been changed since it was developed, save for changes in the definition of variables, and the evidence presented here suggest that the specification is important.

17. Summary on imputation

It is clear that the method of imputation does affect estimates. This is true for studying distributions, calculating the bite, or identifying the number of low-paid workers. The difficulty is in interpreting the results: there is no 'correct' answer, results are sensitive to interventions based on personal judgement, and repeating estimates is an extremely laborious process, dissuading sensitivity analysis.

An alternative to the MI process might be to simply replace missing stated wages with derived values. The evidence presented here suggests this is not helpful. The 'original' results tend to disagree with other results, and produce odd results such as 10% of the population earning below the minimum wage. These findings should also concern other researchers using the LFS and employing the derived-unless-stated model.

The imputation process does not seem to 'create' employees below the minimum wage, even in a year such as 2012 when there is a large amount of misreporting of wages at the £6.08 category at £6.00, although if there are observations below the NMW these can be replicated in the imputed data; this tends to happen where there are many observations to start with.

The imputation process does affect both the calculations of the bite (particularly at the 10th percentile and for sparsely-sample groups such as ethnic minorities) and estimates of jobs below the minimum wage. Of some concern is that it is difficult to say exactly how the imputation process matters. For example, the above results showed that varying the imputation regression had very little impact in 2011, but a much larger effect in 2012, and it is not clear why.

18. Recommendations on imputation

While the imputation code is well commented, it is not fit for current purpose: the code is difficult to update; does not allow alternative scenarios to be tested; does not allow original and imputed variable to be compared; produces inappropriate and excessive output; does not provide statistical information about the accuracy of estimates; and is very susceptible to errors in coding and use.

The MI method itself is old, but it is still popular. A sensible strategy may be to have the method reviewed by an MI expert, even if informally.

Specific recommendations are therefore

5. The code be rewritten to simplify use and maintenance
    a. Redundancy is removed by the use of loops and macros
    b. The code can be run on multiple years
    c. The code stores all relevant variable definitions associated with each year
6. The code is rewritten to support scenario testing

   a. Editing of a single parameter file should be all that is necessary to test multiple scenarios, including alternative regression specifications
   b. Choices of sub-groups is controlled by editing the parameter file
   c. Decisions over outliers etc should be controlled by yes/no options, as well as to include second jobs
   d. The code should not require human intervention other than editing the parameter file
   e. Creating properly-weighted 'unimputed' (possibly derived-unless-stated) data is an option
7. Output is more appropriate
   a. Temporary files are cleaned up
   b. LPC tables of rates/levels are automatically generated
   c. Output file names are specific to runs
8. Statistical information is improved
   a. If possible, estimates of numbers of low paid should be produced with confidence intervals
   b. LPC may consider whether a 'mean of estimates' approach (reporting results from the scenario testing, rather than choosing a 'preferred estimate') is more appropriate for reporting.
9. Review the MI method, specifically to consider
   a. Whether 'best practice' recommends an alternative approach
   b. How confidence intervals could be produced
   c. Whether the 'hard edge' of the NMW could be incorporated
10. Make the code available on the LPC website
   a. Encourage others to confirm/test/develop the code
11. Simply replacing missing stated wage rates with derived values should not be used in preference to imputed values

References

Barford N. (2010) "Revisions to Workforce Jobs", *Economic and Labour Market Review*. ONS. September. www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--9--september-2010/index.html

Barnes, M. and Martin R. (2002), "Business Data Linking: An Introduction". *Economic Trends* no. 581, April http://www.ons.gov.uk/ons/rel/elmr/economic-trends--discontinued-/no--581--april-2002/business-data-linking--an-introduction.pdf

Behling F. and Speckesser S. (2013) *An impact analysis of the introduction of the Apprenticve Rate of the National Minimum Wage.* Research report for the Low Pay Commission. February. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/226951/National_minimum_wage_apprentice_rate_impact_anaylsis.pdf

Daffin C. (2004) *Annual Survey of Hours and Earnings: An analysis of historical data 1998-2003*. ONS, mimeo http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/annual-survey-of-hours-and-earnings/annual-survey-of-hours-and-earnings/an-analysis-of-historical-data-1998-2003.pdf

Drew H., King A. and Ritchie F. (2013) *Impact Evaluation: Workplace Employment Relations Survey and European Social Survey; Final report* . ESRC http://www.esrc.ac.uk/_images/WERS_ESS_tcm8-26047.pdf

Fry S. and Ritchie F. (2011) 'Impact of the recession on household expenditure' in Family Spending 2011 (Ch. 5), Office for National Statistics http://www.ons.gov.uk/ons/rel/family-spending/family-spending/family-spending-2011-edition/family-spending-2011-pdf.pdf

Fry S. and Ritchie F. (2012) 'Issues in the measurement of low pay, 2010'; University of the West of England Economics Working paper no. 1210 http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1210%20Issues%20in%20the%20Measurement%20of%20Low%20Pay%202010.pdf

Fry S. and Ritchie F. (2013) *Behavioural aspects of the National Minimum Wage*. Research report for the Low Pay Commission. February. http://www.lowpay.gov.uk/lowpay/research/pdf/Behavioural_Aspects_of_the_National_Minimum_Wage_for_publication.pdf

Higton, J. (2013) *Apprenticeship pay survey 2012: research findings*. Research paper no. 121. Department for Business Industry and Skills. October. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/246419/bis-13-p137-apprenticeship-pay-survey-2012-research-findings.DOC

Higton J., Hirst A., Klahr R., Lefaucheux M., and Salmon C. (2012) *An assessment of the introduction of the Apprentice Rate*. Research report for the Low Pay Commission. February. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/227559/LPC-IpsosMORICambridgePolicyConsultantsIntroduction_of_the_Apprentice_RateFINAL.pdf

Higton J., Kaur-Ballagan K., Navin Shah J. and Medien K (2012) *Apprentice Pay Survey 2011*. Research paper no. 64. Department for Business, Industry and Skills. March. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32286/12-p137-apprenticeship-pay-survey-2011.pdf

Knight J. (2010) *Data sources and concepts explained*. Low Pay Commission report.

Kersley B., Alpin C., Forth J., Bryson A., Bewley H., Dix G. and Oxenbridge S. (2006). *Inside the Workplace: Findings of the 2004 workplace employment relations survey.* Routledge. London.

Milton J (2004), 'New methodology for low pay estimates'. http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/annual-survey-of-hours-and-earnings/low-pay-estimates/new-methodology-for-low-pay-estimates-in-2004.pdf

ONS (2003) *The New Inflation Target: the Statistical Perspective.* ONS. http://www.ons.gov.uk/ons/rel/cpi/consumer-price-inflation/new-inflation-target---full-text/new-inflation-target---executive-summary.pdf

ONS (2005a) *Methodology for the Annual Earnings Ratio.* Paper prepared for the Methodology Advisory Committee. http://www.ons.gov.uk/ons/guide-method/method-quality/advisory-committee/2001-2004/fifth-meeting/methodology-for-the-average-earnings-ratio.pdf

ONS (2005b) *The New Experimental Index of Labour Costs per Hour.* ONS. http://www.ons.gov.uk/ons/rel/lms/labour-market-trends--discontinued-/volume-113--no--8/the-new-experimental-index-of-labour-costs-per-hour.pdf

ONS (2007a) *Changes to ASHE in 2007*. ONS http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/annual-survey-of-hours-and-earnings/annual-survey-of-hours-and-earnings/changes-to-ashe-in-2007.pdf

ONS (2007b) *The ONS Productivity Handbook: A Statistical Overview and Guide.* ONS. www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/productivity-measures/productivity-handbook/the-ons-productivity-handbook--a-statistical-overview-and-guide.pdf

ONS (2011a) *Quality and Methodology Information Sheet- Average Weekly Earnings.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-statistics/quality-and-methodology-information-for-average-weekly-earnings.pdf.

ONS (2011b) *Summary Quality Report for Labour Market Data Releases.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-statistics/new-component.pdf.

ONS (2011c) *Quality and Methodology Information Paper – Inflation – CPI and RPI-.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-statistics/quality-and-methodology-information-for-average-weekly-earnings.pdf.

ONS (2012a) *Balancing the Three Approaches to Measuring Gross Domestic Product, 2012*. ONS. http://www.ons.gov.uk/ons/dcp171766_273489.pdf

ONS (2012b) *Guide to Labour Market Statistics.* ONS. http://www.ons.gov.uk/ons/rel/lms/labour-market-guidance/guide-to-labour-market-statistics/guide-to-lm-statistics.html

ONS (2012c) *Quality and Methodology Information Sheet- GDP.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economy/summary-quality-report-for-gross-domestic-product--gdp-.pdf

ONS (2012d) *Quality and Methodology Information Sheet- Index of Labour Costs per Hour.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/quality-and-methodology-information-reports-by-theme/labour-market/quality-and-methodology-information-for-index-labour-costs.pdf.

ONS (2012e) *Quality and Methodology Information Sheet- Labour Productivity.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economy/quality-and-methodology-information-for-labour-productivity.pdf

ONS (2012f) *Quality and Methodology Information Sheet- MBS.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-and-energy/monthly-business-survey.pdf

ONS (2012g) *Quality and Methodology Information Sheet- Producer Price Index.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economy/quality-and-methodology-information-for-producer-price-indices--ppi-.pdf.

ONS (2012h) *Quality and Methodology Information Sheet- Service Producer Price Index.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/business-statistics/summary-quality-report-for-services-producer-price-indices.pdf.

ONS (2012i) *Why is GDP Revised? ONS*. www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/2011-present/why-is-gdp-revised-/why-is-gdp-revised---download.pdf

.ONS (2013a) *Introducing the New CPIH Measure of Consumer Price Inflation*. ONS. http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/introducing-the-new-cpih-measure-of-consumer-price-inflation.pdf

ONS (2013b) *Introducing the New RPIJ Measure of Consumer Price Inflation*. ONS. http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/introducing-the-new-cpih-measure-of-consumer-price-inflation.pdf

ONS (2013c)*Quality and Methodology Information Sheet- GDP.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economy/summary-quality-report-for-gross-domestic-product--gdp-.pdf

ONS (2013d) *Quality and Methodology Information Sheet- Workforce Jobs.* ONS. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/labour-market/summary-quality-report-for-workforce-jobs.pdf

ONS (2013e) *Performance and Quality Monitoring Report LFS Jan-Mar 2013.* Office for National Statistics; updated quarterly. http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-force-survey/pmq-report-january-to-march-2013.pdf

Ormerod C and Ritchie F (2007a) 'Issues in the Measurement of Low Pay', Economic and Labour Market Review, Volume 1, No. 6, pp.37-45 http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--6--june-2007/issues-in-the-measurement-of-low-pay.pdf

Ormerod C. and Ritchie F. (2007b) "Linking ASHE and LFS: can the main earnings sources be reconciled?", *Economic & Labour Market Review*, v1:3 pp 24-31 http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--3--march-2007/linking-ashe-and-lfs--can-the-sources-be-reconciled-.pdf

Ritchie F., Thomas A. and Welpton R. (2012) *What is a manufacturing job?* Bristol Business School Working Papers in Economics no. 1218 http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1218.pdf

Robjohns J. (2006) "ARD2: the new Annual Respondents Database" *Economic Trends* no. 630, May http://www.ons.gov.uk/ons/rel/elmr/economic-trends--discontinued-/no--630--may-2006/ard2--the-new-annual-respondents-database.pdf

Le Roux S., Lucchino and Wilkinson D. (2013) *An investigation into the extent of non-compliance with the National Minimum Wage.* Research report for the Low Pay Commission. February. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/227481/NIESR_Report_LPC12_Final_March_2013.pdf

Taylor C., James G. and Pring P. (2011) "The development of the Monthly Business Survey", *Economic and Labour Market Review*. ONS. February. http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--2--february-2011/economic---labour-market-review.pdf

van Wanrooy B., Bewley H., Bryson A., Forth J., Freeth S., Stokes L. and Wood S. (2013) *The 2011 Workplace Employment Relations Study: FIRST findings*. Department for Business, Innovation and Skills https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175479/13-535-the-2011-workplace-employment-relations-study-first-findings1.pdf

Appendix A: regression results from all datasets with wages above the NMW

ASHE absolute (derived wage)		Factors determining likelihood of rounding				by industry and occupation	Likelihood of rounding by company size				Likelihood of rounding by sector		
	Regression type:	Abs pre-recession	Abs post-recession	Single year (2008) when abs=rel	Abs single year (2011) exact LPc defs	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Childcare
Wage influences	wage	o	---	-	---	---	---	---	---	---	---	---	-
	NMW	---	+++	---	+++	+++	o	+	+++	o	+++	+++	o
	if NMW rounded		---	---		o	--	o	++	o	--	++	o
	basic working hours	---	o	---	o	---	o	o	o	---	---	o	-
Personal characteristics	if female	---	--	o	-	---	o	---	---	---	o	o	o
	age	---	-	--	o	---	---	---	---	---	---	---	o
	if full_time	---	---	o	---	---	---	---	---	---	o	o	o
Firm characteristics (default size 250+)	firm size 0-9 employees	+++	+++	+++	+++	+++					+++	+++	+++
	firm size 10-49 employees	+++	+++	+++	+++	+++					+++	+++	+++
	firm size 50-249 employees	+++	+++	+++	+++	+++					+++	+++	++
	if public sector	---	---	---	---	---	---	--	---	---	o	o	
Region (default = NW)	NE England	+	o	o	o	o	++	o	o	o	o	o	o
	Yorks	+++	o	o	o	+++	o	o	o	+++	o	o	o
	East Midlands	o	o	o	o	o	+	o	o	o	o	-	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	--	o
	SW England	o	o	o	o	+	+++	o	o	o	o	o	o
	East England	o	+	o	++	+++	o	o	--	+++	o	o	o
	London	---	o	o	o	o	---	o	o	o	---	-	o
	SE England	---	+++	o	+++	o	---	o	o	++	--	o	o
	Wales	+	+	o	o	++	++	o	o	o	o	o	o
	Scotland	o	o	o	o	o	o	o	o	o	o	o	o
	NI												
If in low-paying industry (default = not low-paying ind)	Agriculture	o	+++	o	+++		+++	+++	o	+			
	Food processing	o	+++	o	+++		++	+++	o	o			
	Textiles	o	o	o	o		+++	o	o	o			
	Retail	---	-	---	o		+++	++	++	---			
	Hospitality	+++	+++	+++	+++		+++	+++	+++	+++			
	Cleaning	+++	+++	+++	+++		+++	+++	+++	+++			
	Social care	+++	+++	o	+++		+++	+++	+++	o			
	Childcare		o		o		o	o	o	o			
	Leisure	+++	+++	+++	+++		+++	+++	++	+++			
	Hairdressing												
	Employment agencies		+++		+++		+++	+++	+++	+++			
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	---	o	--		-	o	o	o	o	o	
	Food Processing	+++	-	o	o		o	+++	o	--	o	o	
	Textiles	o	o	o	o		o	o	--	o	o		
	Retail	o	---	o	---		+++	+++	+++	---	--	--	o
	Hospitality	o	o	o	o		---	o	o	+++	o	o	o
	Cleaning	o	o	o	o		++	--	o	o	o	o	o
	Social Care												
	Childcare	---	---	---	--		o	---	o	--	o	---	o
	Leisure	---	---	o	---		o	-	o	---	o	o	++
	Hairdressing	o	o	o	+		+++	o	o	o	++		
	Office Work	---	---	-	o		o	---	---	o	---	o	
	Non-food Processing	+++	o	o	o		+++	+++	o	o	o	o	
	Storage												
	Transport	o	-	o	o		++	o	o	---	---	o	
	Unionised	---	---	---	---	---	o	---	---	---	---	--	o
Low paying (Y/N marker)	Is low-paying industry					+++							
	Is low-paying occupation					---							

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

LFS absolute (stated wage)		Factors determining likelihood of rounding			By industry and occupation	Likelihood of rounding by company size				Likelihood of rounding by sector		
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	Abs single year (2012)	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Childcare
Wage influences	wage	- - -	- - -	+ + +	- - -	- - -	- - -	- -	o	o	o	- - -
	NMW	+ + +	o	- -	+ + +	+ +	+ +	o	o	o	o	+ +
	if NMW rounded	o			o	o	o	o	o	o	o	- -
	basic working hours	-	o	o	-	o	o	o	o	o	o	o
Personal characteristics	if female	- - -	-	-	- - -	- -	-	o	o	o	o	o
	age	+	o	o	o	o	o	+ +	o	+ + +	o	o
	if full_time	o	o	-	o	o	o	o	o	o	o	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	+ + +	+ + +	o	+ + +					+ + +	o	o
	firm size 10-49 employees	+ + +	+ + +	o	+ + +					+ + +	o	o
	firm size 50-249 employees	o	o	o	+ + +					+ +	o	o
	if public sector	- - -	o	- - -	- - -	- - -	- - -	o	o	o	- -	- -
Region (default = NW)	NE England	o	o	-	o	o	o	o	o	o	o	o
	Yorks	- -	o	o	- -	o	- - -	o	o	o	o	o
	East Midlands	o	o	o	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	o
	SW England	+	o	o	+	+ + +	o	o	o	o	o	o
	East England	o	o	o	o	o	o	o	o	o	o	-
	London	o	o	o	o	o	o	o	+	o	o	o
	SE England	o	o	o	o	o	o	o	o	o	o	o
	Wales	o	o	o	o	o	o	o	o	o	o	+
	Scotland	o	o	o	o	o	o	o	o	o	- -	o
	NI											
If in low-paying industry (default = not low-paying ind)	Agriculture	o	o	o		o	o	o	o			
	Food processing	o	o	o		o	o	o	o			
	Textiles	o	o	-		o	o	o				
	Retail	- - -	- - -	o		- - -	- -	- - -	- - -			
	Hospitality	+ + +	o	o		+ +	+ +	o	o			
	Cleaning	o	o	o		o	o	o	o			
	Social care	- - -	- -	- -		o	- - -	- -	o			
	Childcare	o	o	-		o	o	o				
	Leisure	o	o	- -		o	o	o	o			
	Hairdressing											
	Employment agencies	o	o	o		o	o	o	o			
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o			o	o	o	o	o		o
	Food Processing	-	o	- - -		o	o	o	o	o		
	Textiles	-	o	o		o	o	-	o	o		
	Retail	- - -	o	o		o	- - -	-	o	o	o	
	Hospitality	- -	o	o		- -	-	o	o	o	o	o
	Cleaning	o	o	o		o	o	o	o	+ +	o	o
	Social Care	o	o	o		o	o	o	o		+	- -
	Childcare	o	o	o		o	o	o	o		o	o
	Leisure	o	o	+		o	o	o	o			
	Hairdressing	o	+ +			o	o	o		o		
	Office Work	o	o	o		o	o	o	+	+ + +	o	
	Non-food Processing	o	o	o		o	o	o	o	o		
	Storage	o	o	o		o	- -	o	o	+		
	Transport	o	o	o		o	o	o	o	o		
Disability (default = no dis.)	*Disability**	+	o	o	+	o	o	o	o	o	o	o
	DDA disabled and work-limiting	o	o	o	o	o	o	o	o	o	o	o
	DDA disabled only	o	o	o	o	o	+	o	o	o	o	o
	Work-limiting disabled											
Ethnicity (default = white)	*Ethnicity**	o	o	+ +	+	o	o	o	o	o	o	o
	Mixed	o	o		o	o	o	o	o	o	o	
	Asian	o	o	o	o	o	o	o	o	o	o	o
	Black	o	o	- -	o	o	o	o	o	o	o	o
	Chinese	o	o	o	o	o	o	o	o	o	o	
	Other	+	o		+	o	o	o	+	o	o	
Education (default = A*-C) GCSE or equivalen	*Education**	o	o	o	o	o	o	o	o	o	o	o
	Degree or equivalent	o	o	o	o	o	o	o	o	o	o	o
	Higher education	o	o	o	o	o	o	o	o	o	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o
	Other qualifications	o	o	o	o	o	o	o	o	o	o	o
	No qualification	o	o	o	o	o	o	o	o	o	o	o
Other	Proxy	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	o
	*Documentation**	+ +	+	o	+ +	+ +	o	o	o	o	o	+ +
	With doc. (payslip)	o	o	o	o	o	o	o	o	o	o	+
	With doc. (bank)	+ +	o		+ +	o	o	o	o	+ +		
	With doc. (other)											
NMW rounded (default = 5p)	NMW rounded to 10p											
	NMW rounded to 25p											
Low paying (Y/N marker)	Is low-paying industry				- - -							
	Is low-paying occupation				- -							

BHPS absolute (stated wage)		Likelihood of rounding		By industry and occupation	Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs pre-recession	Single year (2008) when abs=rel	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care
Wage influences	wage	o	o	o	+	o	o	o	o	o
	NMW	o	o	o	- -	o	o	o	o	o
	if NMW rounded	o	o	+	o	o	o	+ +	+ + +	o
	basic working hours	o	o	o	o	o	+ +	-	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o
	age	o	o	o	o	o	-	o	o	o
	if full_time	o	o	o	o	o	- - -	+ +	o	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	+ + +	+ + +	+ + +					+ + +	o
	firm size 10-49 employees	+ + +	+ + +	+ + +					+ + +	o
	firm size 50-249 employees	+ +	+ + +	+ + +					+ +	o
	if public sector	o	o	o	o	o	o	o	o	o
Region (default = NW)	NE England	o	- -	o	o	o	o	o	o	o
	Yorks	o	o	o	o	o	o	o	o	o
	East Midlands	o	o	o	o	o	o	o	o	o
	West Midlands	o	- -	- -	o	- -	- -	o	o	- -
	SW England	o	o	o	o	o	-	o	o	o
	East England	o	- -	- -	o	o	- -	o	- -	
	London	o	o	o	o	o	o	o	o	o
	SE England	o	o	o	o	o	o	o	o	o
	Wales	o	- - -	- -	o	o	- - -	o	o	o
	Scotland	o	o	o	o	o	o	o	o	
	NI									
If in low-paying industry (default = not low-paying ind)	Agriculture	- -	o		-	o	o			
	Food processing	o	o			o	o	-		
	Textiles	o					o			
	Retail	- -	o		o	- -	o	- - -		
	Hospitality	o	o		o	o	o	o		
	Cleaning	o	o		o	o	o	o		
	Social care	o	o		o	o	o	-		
	Childcare									
	Leisure	o	o		- - -	o		o		
	Hairdressing									
	Employment agencies									
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o			o				
	Food Processing									
	Textiles									
	Retail	o	o			o	-	+ +	o	
	Hospitality									
	Cleaning									
	Social Care									
	Childcare									
	Leisure	o	+ +		o	o	o	o	o	
	Hairdressing	+ +	o		o	o	o			
	Office Work	o			o					
	Non-food Processing									
	Storage									
	Transport									
Disability (default = no dis.)	DDA disabled and work-limiting									
	DDA disabled only									
	Work-limiting disabled									
Ethnicity (default = white)	Mixed									
	Asian	-	o	o		o	o	o	o	
	Black									
	Chinese									
	Other	o		o		o		o	o	
Education (default = A*-C)	Degree or equivalent	o	o	o	o	o	o	o	o	o
	Higher education	+ +	+	+ + +		o	o	+	o	
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o
	Other qualifications	- -	o	-	o	o	o	o	- -	o
GCSE or equivalent)	No qualification	o	-	o	o	-	o	o	o	o
Other	Exact	- - -	- - -	- - -	o	- - -	- - -	- - -	- - -	o
	Unionised	- - -	- -	- - -	- - -	o	- -	o	o	- -
NMW rounded (default = 5p)	NMW rounded to 5p									
	NMW rounded to 10p									
	NMW rounded to 25p									
Low paying (Y/N marker)	Is low-paying industry			-						
	Is low-paying occupation			o						

BHPS absolute (derived wage)		Likelihood of rounding		By industry and	Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs pre-recession	Single year (2008) when abs=rel	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care
Wage influences	wage	+ + +	o	+ + +	+ + +	+	o	+ +	o	o
	NMW	- - -	o	- - -	- - -	-	- -	o	o	o
	if NMW rounded	o	o	o	o	o	o	o	o	o
	basic working hours	o	o	o	o	o	o	o	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o
	age	+ + +	o	+ + +	o	+ + +	+ +	o	+ + +	o
	if full_time	o	o	o	o	o	o	o	o	-
Firm charactoristics (default size 250+)	firm size 0-9 employees	+ +	o	+ +					+ + +	o
	firm size 10-49 employees	o	o	o					+	o
	firm size 50-249 employees	o	o	o					o	o
	if public sector	o	o	o	o	o	- - -	+	o	o
Region (default = NW)	NE England	o	o	o	o	o	o	o	o	o
	Yorks	o	o	- -	o	- - -	o	o	- - -	o
	East Midlands	o	+ +	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	-	o
	SW England	o	o	o	o	- -	o	o	o	o
	East England	o	o	o	o	- -	o	o	o	o
	London	- -	o	-	o	o		o	o	o
	SE England	o	o	o	o	- -	o	+ +	-	o
	Wales	o	o	o	o	o	o	o	-	o
	Scotland	o	o	o	o	o	o	o	-	o
	NI									
If in low-paying industry (default = not low-paying ind)	Agriculture	o	o		-	o	o			
	Food processing	o	o			o	o	+ +		
	Textiles	o	o			o	o			
	Retail	o	- -		o	o	o	o		
	Hospitality	o	o		o	+	o	o		
	Cleaning	o	o		o	o		o		
	Social care	o	o		o	o	o			
	Childcare									
	Leisure	o			o	o	o	o		
	Hairdressing									
	Employment agencies									
If in low-paying occupation (default = not low-paying occ)	Agriculture	o			o					
	Food Processing									
	Textiles									
	Retail	o	o		o	o	o	o	o	
	Hospitality									
	Cleaning									
	Social Care									
	Childcare									
	Leisure	+ +	o		+	o	o			
	Hairdressing	+	o		o	o				
	Office Work	o	o			o	o		o	
	Non-food Processing	o	o		o	o				
	Storage									
	Transport	o					o			
Disability (default = no dis.)	DDA disabled and work-limiting									
	DDA disabled only									
	Work-limiting disabled									
Ethnicity (default = white)	Mixed									
	Asian	o	o	o		o	o	o	o	
	Black									
	Chinese									
	Other									
Education (default = A*-C)	Degree or equivalent	o	o	o	o	o	o	o	o	o
	Higher education	o	+ +	o	o	o	o	o	o	o
	GCE, A-level or equivalent	o	o	o	o	+ +	o	o	o	o
	Other qualifications	-	o	o	o	o	o	o	- -	o
GCSE or equivalent)	No qualification	o	o	o	-	o	o	o	- -	o
Other	Exact	o	o	o	o	o	-	o	o	o
	Unionised	- - -	o	- - -	- -	- - -	o	o	o	-
NMW rounded (default = 5p)	NMW rounded to 5p									
	NMW rounded to 10p									
	NMW rounded to 25p									
Low paying (Y/N marker)	Is low-paying industry			o						
	Is low-paying occupation			o						

USoc absolute (stated wage)		Likelihood of rounding		By industry and occupation	Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care
Wage influences	wage	- - -	-	- - -	o	-	-	o	o	o
	NMW	+ +	o	+ +	o	o	o	o	o	o
	if NMW rounded	o		o	o	+	o	o	o	o
	basic working hours	- -	o	-	-	o	o	- -	o	- - -
Personal characteristics	if female	o	o	o	+	o	o	o	-	+
	age	o	o	o	o	o	o	o	+	o
	if full_time	+ +	o	+	o	o	o	o	o	+ +
Firm charactoristics (default size 250+)	firm size 0-9 employees	+ + +	+ + +	+ + +					+ + +	+ +
	firm size 10-49 employees	+ + +	o	+ + +					+ + +	+ + +
	firm size 50-249 employees	o	o	+					o	+ +
	if public sector	- - -	- -	- - -	- - -	- -	- -	-	o	o
Region (default = NW)	NE England									
	Yorks									
	East Midlands									
	West Midlands									
	SW England									
	East England									
	London									
	SE England									
	Wales									
	Scotland									
	NI									
If in low-paying industry (default = not low-paying ind)	Agriculture	o				o				
	Food processing	- -	o			o	- -	o		
	Textiles	o				o	o			
	Retail	- - -	- - -		- - -	- - -	- -	- - -		
	Hospitality	-	o		- -	o	o	-		
	Cleaning	o	-		o	o	o	o		
	Social care	-	o		o	o	o	- -		
	Childcare	o	o		o	+ +				
	Leisure	o	o		o	o	o	o		
	Hairdressing									
	Employment agencies	o				o		o		
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o		o	o	o	o	o	
	Food Processing	o	o		o	-	o	o	o	o
	Textiles	o	o		o	o	o	o	o	
	Retail	o	o		o	-	o	o	o	
	Hospitality									
	Cleaning	o	o		o	o	+	o	+ +	o
	Social Care									
	Childcare	o	o		o	o	o	o	o	
	Leisure	o			o	o	o	o		
	Hairdressing	o	o		o	o		o	o	
	Office Work	o	o		o	+	o	o	o	
	Non-food Processing	o	o		o	o	o			o
	Storage									
	Transport	+ +	o		o	+ +	o	o	o	
Disability (default = no dis.)	DDA disabled and work-limiting									
	DDA disabled only	o	o	o	- -	o	o	o	o	o
	Work-limiting disabled									
Ethnicity (default = white)	Mixed	o	o	o		o	o		+	o
	Asian	+ +	o	+	+	o	o	o	o	o
	Black	o	o	o	o	o	o	o	o	o
	Chinese	o		o	o	o	o	o	o	
	Other	o		o	o	o		o	o	
Education (default = A*-C)	Degree or equivalent	o	o	o	- -	o	o	o	o	o
	Higher education	o	o	o	o	o	o	- -	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o
	Other qualifications	o		o	- -	o	o	o	o	
GCSE or equivalent)	No qualification	o	o	o	o	o	o	o	o	o
Other	Exact	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
	Unionised									
NMW rounded (default = 5p)	NMW rounded to 5p									
	NMW rounded to 10p									
	NMW rounded to 25p									
Low paying (Y/N marker)	Is low-paying industry			- - -						
	Is low-paying occupation			- - -						

USoc absolute (derived wage)		Likelihood of rounding		By industry and occupation	Likelihood of rounding by company size				Likelihood of rounding by sector	
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	All industries and occupations	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care
Wage influences	wage	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	o
	NMW	+ + +	+ + +	+ + +	+ +	+ + +	+	+ +	+ + +	o
	if NMW rounded	+ + +		+ + +	o	+ + +	o	+ + +	+ +	+ + +
	basic working hours	-	o	-	- -	o	o	-	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o
	age	o	o	o	o	o	o	o	o	o
	if full_time	o	o	o	o	o	o	o	o	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	o	o	o					o	- - -
	firm size 10-49 employees	o	o	o					- -	- -
	firm size 50-249 employees	- -	o	- -					- -	- -
	if public sector	o	o	o	+ +	o	o	o	o	o
Region (default = NW)	NE England									
	Yorks									
	East Midlands									
	West Midlands									
	SW England									
	East England									
	London									
	SE England									
	Wales									
	Scotland									
	NI									
If in low-paying industry (default = not low-paying ind)	Agriculture	o			+	o				
	Food processing	o	o		o	o		o		
	Textiles	o			o	o				
	Retail	o	o		o	o	o	o		
	Hospitality	o	o		o	o	o	o		
	Cleaning	o	o		o	o	o			
	Social care	o	o		o	o	o	+ +		
	Childcare	o	o		o	o				
	Leisure	o	o		o	o	o	o		
	Hairdressing									
	Employment agencies	o	o				o			
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o			o				
	Food Processing	o	o		o	-	o	o	-	
	Textiles	o	o		o	o		o	o	
	Retail	o	o		o	o	o	o	-	
	Hospitality									
	Cleaning	o	o		o	o	o	o	o	o
	Social Care									
	Childcare	o	o		o	-	o	o		
	Leisure	o	o		o	o	o	o		
	Hairdressing	o				o				
	Office Work	o	o		o	o	o	o	-	o
	Non-food Processing	o			o			o		
	Storage									
	Transport	o	o		o	o	o	o	o	
Disability (default = no dis.)	DDA disabled and work-limiting									
	DDA disabled only	o	o	o	o	o	o	o	o	o
	Work-limiting disabled									
Ethnicity (default = white)	Mixed	o		o		o	o	o	o	
	Asian	o	o	o	o	+	o	o	o	o
	Black	o	o	o	o	o	o	o	o	o
	Chinese									
	Other	o	o	o	o	o				o
Education (default = A*-C)	Degree or equivalent	+ + +	o	+ + +	+	o	o	o	o	o
	Higher education	o	+	o	o	o	o	o	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	-	o
	Other qualifications	o		o	o	o	o		o	
GCSE or equivalent)	No qualification	+	+	+	o	+ +	o	o	o	o
Other NMW rounded (default = 5p)	Exact	o	o	o	o	o	o	- - -	- -	o
	Unionised									
	NMW rounded to 5p									
	NMW rounded to 10p									
	NMW rounded to 25p									
Low paying (Y/N marker)	Is low-paying industry			o						
	Is low-paying occupation			o						

Appendix B: Regression results, relative wages only

Note that for relative wages, only wages above the NMW are considered.

ASHE relative (derived wage)		Factors determining likelihood of rounding			Likelihood of rounding to specific values				Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		
Regression type:		Abs pre-recession	Abs post-recession	Abs single year (2011) exact LPc defs	Abs rounding to 5p	Abs rounding to 10p	Abs rounding to 25p	Abs rounding to 50p	All industries and occupations	All ind but NOT in a low paying occ	All ind but in a low paying occ	All occ but NOT in a low paying ind	All occ and in a low paying ind	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Childcare
Wage influences	wage	o	+++	+++	+++	+++	+++	+++	++	+++	o	+++	--	+++	+++	o	o	---	---	o
	NMW	---	+++	+++	---	---	---	+++	---	---	---	---	---	o	o	o	---	---	o	o
	if NMW rounded		+++		+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
	basic working hours	---	o	o	---	o	o	o	---	o	---	o	---	--	o	o	---	o	o	o
Personal characteristics	if female	---	o	o	---	--	o	o	---	---	---	---	-	o	--	---	--	o	o	o
	age	---	o	o	---	---	--	--	---	---	---	---	---	o	---	--	---	---	---	o
	if full_time	---	o	o	---	-	o	o	--	---	o	---	o	o	---	--	o	o	-	o
Firm characteristics (default size 250+)	firm size 0-9 employees	+++	+++	---	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	o
	firm size 10-49 employees	+++	+	---	+++	+++	+++	+++	+++	+++	+++	+++	+++					+++	+++	o
	firm size 50-249 employees	+++	o	o	+++	+++	o	++	+++	+++	+++	+++	+++					+++	+	o
	if public sector	---	o	++	---	---	o	--	---	---	--	---	o	---	o	--	---	o	-	o
Region (default = NW)	NE England	+	o	--	o	o	o	-	o	o	o	o	o	o	o	o	o	o	o	o
	Yorks	+++	o	o	+++	o	+	o	+++	+	+	+++	o	o	o	-	+++	o	o	---
	East Midlands	o	o	o	o	o	o	o	o	o	o	+	o	o	o	o	o	o	o	o
	West Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o	o	o
	SW England	o	o	o	o	o	o	o	o	o	o	o	o	++	o	-	o	o	---	o
	East England	o	--	--	o	o	o	o	o	o	o	o	o	o	-	---	o	-	--	o
	London	---	+++	+	o	---	---	---	o	o	+	o	o	---	--	--	++	o	---	o
	SE England	---	o	o	-	--	o	o	o	-	o	o	o	-	o	---	o	o	---	o
	Wales	+	o	o	o	o	++	o	o	+	o	o	o	o	o	o	o	o	o	o
	Scotland	o	---	---	--	-	o	o	--	-	-	--	o	o	o	--	--	o	o	o
	NI																			
If in low-paying industry (default = not low-paying ind)	Agriculture	o	o	o	o	o	o	o						o	+	o	o			
	Food processing	o	o	o	o	+++	o	o						+	o	o	---			
	Textiles	o	o	++	o	o	o	o						o	o	o	o			
	Retail	---	o	o	---	--	o	o						+++	++	o	---			
	Hospitality	+++	-	---	+++	+++	+++	+++						+++	+++	+++	+++			
	Cleaning	+++	o	---	+++	+++	+++	+++						++	+++	+++	+++			
	Social care	+++	+++	---	+++	+++	+++	o						o	+++	+++	+++			
	Childcare		+++	o	o	o	o	o						o	+++	++	o			
	Leisure	+++	+	o	+++	+++	+++	+++						+++	++	+++	+++			
	Hairdressing																			
	Employment agencies		o	---	++	+++	o	o						o	o	o	o			
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o	o	o	o	o	o						o	o	o	o	o	o	
	Food Processing	+++	+++	+++	+++	o	+	o						o	+++	+	+++	o	o	
	Textiles	o	o	o	o	o	o	o						o	o	--	o	o		
	Retail	o	o	+	o	o	+	o						+++	o	++	o	o	--	
	Hospitality	o	++	o	o	o	o	o						-	o	o	o	o	o	o
	Cleaning	o	++	+++	o	o	+++	o						+	o	o	o	+++	o	o
	Social Care																			
	Childcare	---	--	+	---	-	o	o						o	--	--	-	o	--	o
	Leisure	---	o	o	--	o	o	o						o	o	o	--	o	o	o
	Hairdressing	o	o	o	o	o	o	++						o	o	-	o	o		
	Office Work	---	o	o	---	--	o	o						o	---	o	o	-	o	
	Non-food Processing	+++	o	o	+++	+++	++	o						o	+++	o	o	o		
	Storage																			
	Transport	o	o	o	o	o	o	o						++	o	o	o	o	o	
	Unionised	---	---	o	---	---	---	---	---	---	---	---	---	o	--	---	---	---	o	o
NMW rounded to X' pence Default = 5p	NMW rounded to 10p					+++														
	NMW rounded to 25p						+++													
	NMW rounded to 50p							+++												
Low paying (Y/N marker)	Is low-paying industry								+++	+++	++									
	Is low-paying occupation								---			+++	---							

--- Negative, significant at 1%
-- Negative, significant at 5%
- Negative, significant at 10%
0 Not significant

\+ Positive, significant at 10%
++ Positive, significant at 5%
+++ Positive, significant at 1%

LFS relative (stated wage)		Factors determining likelihood of rounding			Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		
Regression type:		Abs post-recession	Abs single year (2011) exact LPc defs	Abs single year (2012)	Abs rounding to 5p	Abs rounding to 10p	Abs rounding to 25p	All industries and occupations	All ind but NOT in a low paying occ	All ind but in a low paying occ	All occ but NOT in a low paying ind	All occ and in a low paying ind	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Childcare
Wage influences	wage	+++	+++	-	+++	+++	+++	+++	o	+++	o	+++	o	+++	+	o	o	o	o
	NMW	+++	+	o	+++	++	o	+++	o	+++	o	+++	+	+++	o	o	++	o	++
	if NMW rounded	+++			+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
	basic working hours	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Personal characteristics	if female	o	++	++	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o
	age	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	if full_time	o	o	o	o	o	o	++	o	+	o	+	o	o	o	o	o	o	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	+++	o	o	+++	+++	o	+++	+++	+++	o	+++					+++	o	o
	firm size 10-49 employees	+++	o	o	+++	+++	o	+++	++	+++	o	+++					+++	o	o
	firm size 50-249 employees	o	o	o	o	o	o	o	o	o	o	o					o	o	o
	if public sector	o	o	o	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o
Region (default = NW)	NE England	o	o	++	o	o	o	o	o	o	o	o	o	o	o	o	- -	o	++
	Yorks	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++
	East Midlands	+	o	o	+	o	o	+	o	+	o	o	++	+	o	o	o	o	++
	West Midlands	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++
	SW England	o	o	+	o	o	o	+	o	++	++	o	o	+	o	o	o	o	+
	East England	o	o	o	o	o	o	o	o	o	+	o	o	o	o	o	o	o	+
	London	++	o	++	++	o	o	++	o	+	o	o	o	+	++	o	++	o	o
	SE England	++	o	++	++	o	o	++	o	o	o	o	o	++	o	o	o	o	o
	Wales	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+++
	Scotland	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	NI																		
If in low-paying industry (default = not low-paying ind)	Agriculture	o	+	o	o	o	o						++	o	o	o			
	Food processing	o	o	o	o	o	o						o	o	+	o			
	Textiles	- -	o		- -	- - -	o						o	o	o				
	Retail	o	o	o	o	o	o						o	o	o	o			
	Hospitality	o	o	o	o	o	o						o	o	o	o			
	Cleaning	o	o	o	o	o	o						o	o	o	o			
	Social care	o	o	o	o	o	o						o	o	o	o			
	Childcare	o	o	o	o	o	++						o	o	++				
	Leisure	o	o	o	o	o	o						o	o	o	o			
	Hairdressing																		
	Employment agencies	o	o		o	o	o						o	o	o	o			
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o		o	o	o						o	o	o	o	o	o	
	Food Processing	o	+	o	o	o	o						o	o	o	o	o		
	Textiles	o	o		o	o	o						o	o	o	o	o		
	Retail	- -	o	o	- -	- - -	o						o	o	-	o	o	o	
	Hospitality	o	-	o	o	o	+						- -	o	o	++	o	o	o
	Cleaning	o	o	o	o	o	o						o	o	o	o	o	o	o
	Social Care	o	o	o	o	o	o						o	o	o	o		o	o
	Childcare	o	o	o	o	o	-						-	o	o	o			o
	Leisure	- -	o		- -	o	-						- - -	o	o	o		o	
	Hairdressing	- -			- -	-	o						- -	o			o		
	Office Work	o	o	o	o	o	o						o	o	o	o	o	o	
	Non-food Processing	o	++	o	o	o	o						o	o	o	++	++		
	Storage	o	o		o	-	o						-	-	o	o	o		
	Transport	o	o	o	o	o	+						o	o	o	o	o		
Disability (default = no dis.)	*Disability**	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++
	DDA disabled and work-limiting	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++
	DDA disabled only	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++
	Work-limiting disabled																		
Ethnicity (default = white)	*Ethnicity**	-	o	o	-	- -	o	-	o	o	o	o	o	- -	o	o	o	o	o
	Mixed	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o		
	Asian	++	o	o	++	+	+	++	++	+	++	o	++	+	o	o	o	o	o
	Black	++	o		++	+++	+++	++	o	+	o	+	o	+++	o	o	o	o	o
	Chinese	o	o	o	o	o	o	o	o	o	o	o	o	o		o	o	o	
	Other	o	o		o	+	o	o	+	o	o	o	o	o	o	o	-	o	
Education (default = A*-C) GCSE or equivalent)	*Education**	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Degree or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Higher education	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Other qualifications	o	o	o	o	o	o	o	o	o	o	o	- -	o	o	o	o	o	o
	No qualification	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
Other	Proxy	o	- -	o	o	++	-	o	o	o	o	o	o	o	o	o	o	o	o
	*Documentation**	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	++	o
	With doc. (payslip)	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	- -	+	o
	With doc. (bank)	o			o	o	o	o	o	o	o	o	o	o	o	o	o	++	
	With doc. (other)																		
NMW rounded (default = 5p)	NMW rounded to 10p																		
	NMW rounded to 25p						+++												
Low paying (Y/N marker)	Is low-paying industry							o	o	o									
	Is low-paying occupation							o			o	o							

BHPS relative (stated wage)		Factors determining likelihood of rounding		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		Panel
Regression type:		Abs pre-recession	Single year (2008) when abs=rel	Abs rounding to 5p	Abs rounding to 10p	Abs rounding to 25p	All industries and occupations	All ind but NOT in a low paying occ	All ind but in a low paying occ	All occ but NOT in a low paying ind	All occ and in a low paying ind	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Panel - Random effects
Wage influences	wage	-	o	--	o	+++	--	--	o	o	-	o	-	o	o	o	o	-
	NMW	---	o	---	---	-	---	---	o	-	---	-	---	o	o	o	o	---
	if NMW rounded	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
	basic working hours	o	o	o	o	o	o	o	o	o	o	o	o	++	o	o	+	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	age	--	o	o	o	o	o	-	o	o	o	o	o	o	o	o	o	o
	if full_time	o	o	o	o	o	o	o	o	o	o	o	o	--	o	o	-	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	+++	o	+++	o	+	+++	+++	o	o	+++					+++	++	+++
	firm size 10-49 employees	+++	o	+++	o	+	+++	+++	o	o	+++					+++	+++	+++
	firm size 50-249 employees	++	o	+	o	o	++	++	o	o	+++					+++	+++	+
	if public sector	o	++	o	o	o	o	o	o	o	o	+	o	o	o	o	o	o
Region (default = NW)	NE England	+	o	+	+	o	+	o	o	o	+++	++	o	o	+	o	+	o
	Yorks	o	o	o	+	+	o	o	o	---	++	+	o	o	o	o	o	o
	East Midlands	o	o	o	o	o	o	o	o	o	+	o	o	-	o	o	+++	o
	West Midlands	o	o	o	o	o	o	-	o	--	o	o	o	--	o	o	o	o
	SW England	o	o	o	++	o	o	o	o	o	o	o	o	o	o	o	o	o
	East England	o	o	o	o	o	o	o	o	--	o	o	o	---	o	o	+	o
	London	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	SE England	o	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o
	Wales	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Scotland	o	o	o	o	o	o	o	--	--	o	o	+	--	o	o	o	o
	NI																	
If in low-paying industry (default = not low-paying ind)	Agriculture	-	o	-	o	o						o	o	o				o
	Food processing	o	o	o	o	o							o	o	o			o
	Textiles	o		o	o									o				o
	Retail	--	o	--	o	o						o	--	o	---			--
	Hospitality	o	o	o	o	o						o	o	o	o			o
	Cleaning	o	o	o	o	o						o	o	o	o			o
	Social care	o	o	o	o	o						o	o	++	o			o
	Childcare																	
	Leisure	o	o	o	o	o						o	o		o			o
	Hairdressing																	
	Employment agencies																	
If in low-paying occupation (default = not low-paying occ)	Agriculture	o	o	o	o	+							o			o		o
	Food Processing																	
	Textiles																	
	Retail	o	o	o	o	o						o	o	o	++	o		o
	Hospitality																	
	Cleaning																	o
	Social Care																	
	Childcare																	
	Leisure	o	o	o	o	o						o	o	o	o			o
	Hairdressing	++	o	o	o	o						o	o					o
	Office Work	o		o									o					o
	Non-food Processing	o		o		o						o						o
	Storage																	
	Transport																	
Disability (default = no dis.)	DDA disabled and work-limiting																	
	DDA disabled only																	
	Work-limiting disabled																	
Ethnicity (default = white)	Mixed																	
	Asian	o	o	o	o	o	o	o	-	o	o		o	++	o	o		o
	Black	o		o	o	o	o	o			o							o
	Chinese																	
	Other	o		o	o	o	o	o			o		o		o	o		o
Education (default = A*-C)	Degree or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+	o	o
	Higher education	++	o	+	o	o	+	+	o	o	o	o	o	++	o	o	o	+
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	o	o	o	o	++	o	o	o	o
	Other qualifications	-	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
GCSE or equivalent)	No qualification	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o	o	o
Other	Exact	---	o	---	o	o	---	---	o	-	---	o	--	--	---	---	o	---
	Unionised	---	o	---	o	o	---	---	o	--	--	---	o	o	o	o	--	---
NMW rounded (default = 5p)	NMW rounded to 5p																	
	NMW rounded to 10p				+++													
	NMW rounded to 25p					+++												
Low paying (Y/N marker)	Is low-paying industry						o	-	o									
	Is low-paying occupation						o			o	+							

BHPS relative (derived wage)		Factors determining likelihood of rounding		Likelihood of rounding to specific values			Likelihood of rounding by industry and occupation					Likelihood of rounding by company size				Likelihood of rounding by sector		Panel
Regression type:		Abs pre-recession	Single year (2008) when abs=rel	Abs rounding to 5p	Abs rounding to 10p	Abs rounding to 25p	All industries and occupations	All ind but NOT in a low paying occ	All ind but in a low paying occ	All occ but NOT in a low paying ind	All occ and in a low paying ind	Firm size: 0-9 employees	Firm size: 10-49 employees	Firm size: 50-249 employees	Firm size: 250+ employees	Retail	Social Care	Panel - Random effects
		BHPS_r_0408_detail	BHPS_r_08on_detail	BHPS_r_detail_r5	BHPS_r_detail_r10	BHPS_r_detail_r25	BHPS_r_indyn_occyn	BHPS_r_indyn_xLPo	BHPS_r_indyn_LPo	BHPS_r_occyn_xLPi	BHPS_r_occyn_LPi	BHPS_r_size1	BHPS_r_size2	BHPS_r_size3	BHPS_r_size4	BHPS_r_retail	BHPS_r_soc_care	BHPS_r_panelRE
Wage influences	wage	+++	o	+++	+++	+++	+++	+++	++	++	++	++	o	o	o	o	o	+++
	NMW	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	if NMW rounded	o	o	++	o	o	++	++	o	+	o	+++	+	o	o	o	o	++
	basic working hours	o	o	o	o	o	-	o	- - -	o	-	o	o	-	o	o	o	o
Personal characteristics	if female	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	age	++	o	++	o	o	++	++	o	o	++	o	++	o	o	o	o	++
	if full_time	o	o	o	o	o	o	o	++	o	o	o	o	o	o	o	o	o
Firm charactoristics (default size 250+)	firm size 0-9 employees	++	o	+	o	o	+	o	++	o	+					+++	o	+
	firm size 10-49 employees	++	o	+	o	o	+	o	++	o	+					o	o	+
	firm size 50-249 employees	o	o	o	o	o	o	o	o	o	o					o	o	o
	if public sector	o	o	o	o	o	o	o	o	o	o	o	o	- -	o	o	o	o
Region (default = NW)	NE England	o	- -	- -	o	o	- -	-	o	- -	o	-	-	o	o	o	-	-
	Yorks	- -	- - -	- - -	- -	o	- - -	- - -	-	- -	- -	o	- - -	o	o	-	o	- - -
	East Midlands	o	-	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	West Midlands	o	-	- -	o	o	- -	- -	o	- -	o	o	o	o	o	o	o	- -
	SW England	o	o	o	o	o	o	o	o	o	o	o	- -	o	o	o	o	o
	East England	o	o	- -	o	o	- -	- -	o	-	o	o	- -	o	o	o	o	- -
	London	-	o	-	o	o	-	o	o	-	o	o	o	o	o	o	o	-
	SE England	o	o	o	o	o	o	o	o	-	o	o	- -	o	o	o	o	o
	Wales	o	- -	-	o	o	- -	- -	o	- - -	o	o	-	o	o	o	o	-
	Scotland	o	-	- -	o	o	- -	- -	o	o	-	o	- -	o	o	o	o	- -
	NI																	
If in low-paying industry (default = not low-paying ind)	Agriculture	o	o	o	o	o						- -	++	o				o
	Food processing	o		o	o	o							o	o	++			o
	Textiles	o	o	o	o	o							o	o				o
	Retail	o	o	o	++	o						o	o	o	o			o
	Hospitality	o	o	o	o	o						o	o	o	o			o
	Cleaning	o	o	o	o	o						o	o		o			o
	Social care	+	o	+	++	o						o	o	o				+
	Childcare																	
	Leisure	o	++	o	+	o						o	o	++	+			o
	Hairdressing																	
	Employment agencies																	
If in low-paying occupation (default = not low-paying occ)	Agriculture	o		o		o						o	o					o
	Food Processing																	o
	Textiles																	
	Retail	o	o	o	o	o						o	o	o	o	o		o
	Hospitality																	
	Cleaning																	o
	Social Care																	
	Childcare																	
	Leisure	++	o	++	o	+						+	o	o				++
	Hairdressing	+	o	o	o	o						o	o	o				o
	Office Work	o	o	o	o	o							o	o	o	o		o
	Non-food Processing	o	o	o	o	o						+	o					o
	Storage																	
	Transport	o		o	o									o				o
Disability (default = no dis.)	DDA disabled and work-limiting																	
	DDA disabled only																	
	Work-limiting disabled																	
Ethnicity (default = white)	Mixed																	
	Asian	o	o	o	o		o	o		o	o		o	o	o	o		o
	Black																	o
	Chinese																	o
	Other																	o
Education (default = A*-C) GCSE or equivalent)	Degree or equivalent	o	o	o	o	++	o	o	o	+++	o	o	o	++	o	o	o	o
	Higher education	-		-	-	o	-	- -	o	o	o	o	o	o	o	o	o	-
	GCE, A-level or equivalent	o	o	o	o	o	o	o	o	++	o	o	++	+	o	- - -	o	o
	Other qualifications	- -	-	- -	o	o	- - -	- -	o	o	- - -	-	o	o	- -	- - -	o	- -
	No qualification	o	o	o	- -	o	o	o	o	o	- -	o	o	o	- -	- -	o	o
Other	Exact	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Unionised	- - -	o	- - -	- -	o	- - -	- - -	o	- -	- - -	- -	- - -	o	o	o	- - -	- - -
NMW rounded (default = 5p)	NMW rounded to 5p																	
	NMW rounded to 10p				+++													
	NMW rounded to 25p					+++												
Low paying (Y/N marker)	Is low-paying industry						++	++	o									
	Is low-paying occupation						o			o	o							