

Information paper

Coverage and non-response errors in the Annual Survey of Hours and Earnings

1 Quality and reliability of the estimates

The Annual Survey of Hours and Earnings (ASHE) aims to provide high-quality statistics on the structure of earnings for various industrial, geographical, occupational and age-related breakdowns. However, the quality of these statistics varies depending on various sources of error.

Sampling error results from differences between a target population and a sample of that population. Sampling error varies partly according to the sample size for any particular breakdown or 'domain'. Indications of the quality of ASHE estimates are provided in the form of coefficients of variation (cv). The coefficient of variation is the ratio of the standard error (se) of an estimate to the estimate, expressed as a percentage. Generally, if all other factors are constant, the smaller the cv the higher the quality of the estimate. Tables of cvs corresponding to estimates are published alongside the estimates themselves.

It should be noted that at low levels of disaggregation high coefficients of variation imply estimates of low quality. For example, for an estimate of £400 with a cv of 10%, the true value is likely to lie between £321.60 and £478.40. This range is given by the estimate $\pm 1.96 \cdot se$. Where these ranges for different estimates overlap, interpretation of differences between the relevant domains becomes more difficult.

In addition to sampling error, ASHE statistics are also subject to **non-sampling errors**. For example, there are known differences between the coverage of the ASHE sample and the target population (i.e. all employee jobs). A discussion on coverage errors is given below. Further, non-response bias may also affect ASHE estimates. This may happen if the jobs for which respondents do not provide information are different to the jobs for which respondents do provide information. Similarly, a section on non-response is provided below.

Finally, ASHE results tables do not account for differences in the composition of different 'slices' of the employee workforce. For example, figures for the public and private sectors include all jobs in those sectors and are not adjusted to account for differences in the age, qualifications or seniority of the employees or the nature of their jobs, all factors which may affect how much employees earn.

2 Coverage errors

Most of ASHE's 1% sample of employees is extracted from Her Majesty's Revenue and Customs (HMRC) Pay As You Earn (PAYE) system by reference to the last two digits of an employee's National Insurance number. Survey data are collected from employers, who provide information about each eligible employee. The allocation of National Insurance numbers to people is random for the last two digits; therefore the sample can be considered to be a simple random sample. Some of the returned sample is from employers who supply data electronically for all employees within their system who are in scope. Such responses would therefore include a small number of new employees who are not yet recorded in HMRC's files, and exclude those who have left the business since the sample file was extracted.

The use of the PAYE frame gives easy access to the population members of interest and allows a longitudinal aspect easily and cheaply, but there is a need to look at all employees across the whole economy. This includes part-time workers and workers on low rates of pay. In particular, precise estimates are needed of the number of people earning low hourly rates in order to make accurate estimates of those being paid below the national minimum wage.

The coverage issue arises because not all employees are registered on PAYE schemes, and PAYE information from HMRC is the sampling frame for the survey. Employers have a legal obligation to operate PAYE on the payments made to employees if their earnings reach the National Insurance Lower Earnings Limit. For the tax year 2012/13 this is £107 per week, £464 per month or £5,564 per year. Since the sampling frame from HMRC does not include employees whose employers have not registered PAYE schemes there is some frame undercoverage. Frame undercoverage could mean that the estimates which are generated by the survey are biased. The issue for ASHE results is that the employees who are not covered could have different characteristics from those in the sampling frame obtained from HMRC. There are two sub-categories of these employers: one where the employer has registered for VAT with HMRC but not for PAYE (so-called VAT-only companies) and the other where the company has not registered for VAT nor for PAYE with HMRC. The latter case can be seen to contain a very small number of employees since the majority of businesses in this category consist of working proprietors only with no employees, but the former case has the potential to contain a greater number of employees. ONS collected earnings information from these VAT-only companies for two years but found that their employees' earnings characteristics were similar to those from companies who had registered PAYE schemes, and therefore discontinued the surveys from 2006.

A different type of coverage issue is that of so-called off-PAYE employees. This is where the employer has raised a PAYE scheme for employees, but chooses not to place certain employees onto the scheme. This could be because the employees are low paid (due either to a low hourly rate of pay or to a small number of hours worked). It is considered that this phenomenon of off-PAYE employees occurs largely in the hotels and catering sector and with very little frequency elsewhere. Again, the interest is in the fact that the employees are likely to be paid low rates of pay. The study which ONS carried out into employees in this category showed that it was very difficult to get meaningful information from businesses that followed this practice.

Other coverage issues which are dealt with satisfactorily in ASHE are those of the employees who change jobs between the date of the first PAYE extract file obtained from HMRC in January and the survey reference date in April (described in more detail in the non-response section below), and those employees who start jobs for the first time after the date of the first extract file and in time for the survey reference date. These cases are handled by taking a second PAYE extract file in April and identifying new entrants into the labour market and people who have moved jobs by comparisons with the January extract.

3 Non-response errors

In addition to these coverage errors, further biases may arise because of non-response. Two main types of non-response exist.

First, there is genuine non-response where employers fail to respond to the survey questionnaire for one or more of their employees. Generally, non-response is for the whole employer, but there are some cases of employers responding for only a partial set of their employees.

Second, the employer may not be able to respond for a particular individual for a variety of reasons, perhaps, for example, because the employee no longer works for the employer. These are known as exemptions since the employer is exempt from completing the survey questionnaire for the employee in question. In some cases, the exempt employee was out of scope — for instance, because the person had left the labour market — so these 'returns' are simply ignored. The largest exemption category is where the employee was no longer with the same employer on the survey date and was assumed to have taken another job. This situation arises largely through delays in updating the administrative systems that provide the sampling frame; the frame is extracted in January, about three months before the survey date, but even the information on that extract is subject to delays in the administrative systems. The consequences of losing these employees from the sample are potentially large because people changing employers are likely to have different characteristics from those who remain with the same employer. Therefore the second extract file is used in order to ensure the coverage, and hopefully response, for these employees, as described above in the last paragraph of the coverage section.

4 Conclusion

Like any survey, there are imperfections in ASHE which ought to be considered before results are used. However, measures have been put in place to reduce non-sampling errors (such as coverage errors) where possible. Furthermore, ONS has made other improvements over the years to questionnaire design, and to editing and imputation methods, and the combined effect of these improvements has resulted in an annual earnings survey which is widely respected and used in the UK and further afield.