**The 2016 Annual Business Survey: A Sample Optimisation**

Megan Pope, Jonathan Digby-North[[1]](#footnote-1)

**Abstract**

In terms of the number of variables collected, the Annual Business Survey (ABS) is the largest business survey run by the Office for National Statistics (ONS), sampling approximately 62,000 businesses in Great Britain. It covers the production, construction, services and distribution industries and provides important indicators of economic activity, including estimates of total turnover, the total value of purchases of goods, materials and services and approximate Gross Value Added. Much of the information it provides feeds directly into National Accounts.

The ABS sample was last optimised by Survey Methodology in 2010; over this time the target population has increased in number by around 25%; both due to natural evolution and the (first-time) introduction of certain single source (PAYE only) businesses into the target population. Other requirements included the selection of ‘Sharing Economy’ businesses and ensuring that all Reporting Units (RUs) within a multiple-RU enterprise were selected, to enable the delivery of robust enterprise level statistics to Eurostat in 2017. A survey re-design was therefore considered essential.

The method used to allocate the sample to strata was a ‘power’ allocation, which essentially seeks to simultaneously optimise the precision of estimates at both the overall and lower (e.g. strata) levels. The performance of the new allocation was assessed on multiple years of past data, enabling us to quantify the expected impact on the quality of the estimates at various levels of aggregation.

The chosen allocation selected 61,988 businesses in Great Britain – 53,897 in England and Wales and 8,091 in Scotland. The number of sampled 0-9 businesses dramatically decreased by around 20,000 businesses – this was distributed among the other employment size bands. On average, 80-90% of CVs across all domains remained the same or were improved upon using the new allocation.

**1. Background**

The ABS is the largest survey that ONS conducts in terms of the number of businesses selected and the number of variables collected. The ABS samples approximately 73,000 businesses in the UK - 62,000 in GB and 11,000 in NI - and collects information on around 600 different variables.

The ABS is predominantly a non-financial business survey covering the production, construction, distribution and services sectors. The ABS covers a small sector of the agriculture sector and samples the insurance and re-insurance industries within the Financial and Insurance sector.

The ABS collects information to understand the detailed structure, conduct and business performance across the UK. The ABS asks for information about turnover, expenditure in terms of employment costs, goods, materials and services, value of assets and stocks held and exports and imports. An approximate measure of Gross Value Added (GVA) is also published – this is used to indicate the amount that individual businesses, industries or sectors contribute to the overall economy. More information about the survey can be found in the ABS Technical Report[[2]](#footnote-2).

Information collected on the survey feeds into the UK National Accounts. This is a key source of data used in the compilation of input-output annual supply and use tables and is also a major contribution to the Blue Book which is used to inform HM Treasury in managing economic policy.

**2. Current Data Collection Method**

The ABS samples approximately 62,000 businesses in Great Britain and a further 11,000 businesses in NI. The sample is drawn from the Inter-Departmental Business Register (IDBR) – held and maintained by ONS. The IDBR holds information on over 2 million businesses in the UK and is the sampling frame for the majority of surveys run at ONS. Businesses are sampled for ONS surveys based on their Reporting Unit (RU). For the majority of the businesses in the UK, the RU is the head office or location to which the questionnaire is sent. For larger businesses, ONS often set up more than one RU to allow the business to report in a way which is easier for them. For example, if a business has a retail and manufacturing arm then it might be easier to report for the different parts of the business separately. Businesses are also split into more than one RU for legal purposes if they operate in both GB and NI.

The ABS has a very complex sample design. The population is stratified by country (England/Wales combined and Scotland) , 4/2-digit Standard Industrial Classification (SIC) and 6 employment size bands. This equates to around 4,000 different strata. By stratifying in this way, it allows us to sample businesses at very low levels of aggregations which in turn ensures that we are able to produce estimates at the class level.

NI are responsible for selecting their own sample and so this piece of work focuses on just the GB sample.

There have been a number of changes to both the sample and the population since the ABS was last optimised in 2010. Due to natural evolution the business population has grown in size and the business area have had to make manual changes to the sample in past years to adhere to the dispatch target. For ABS 2016 the population will also include for the first time, certain single source (PAYE only) businesses. Taking this into account, the ABS population will have increased by 23% since 2010, from 1.8 to 2.2 million businesses. If the current sample design was applied to the new population, the target dispatch count would be breached by nearly 11,000 businesses.

In addition to these changes, there are also a number of requirements that must be met for the 2016 ABS. These will be explained in detail in Section 3.

**3. Sample Requirements**

There were a number of requirements, both old and new, for the 2016 ABS.

Requirement 1 – Businesses with high employment must be sampled

All businesses with high employment should be included in the ABS sample. In general, these businesses contribute significantly to the overall estimates and so it is important that these are captured in the survey. The sampling scheme for ABS remains the same throughout the majority of industries, however there are some exceptions. For some industries within the services sectors, employment is often extremely high but turnover is very low. In these industries the threshold for the census stratum is increased.

Requirement 2 – Businesses with low employment but high turnover must be sampled

As well as ensuring that businesses with large employment are always included in the sample, the ABS also uses inclusion markers to force businesses with low employment but high turnover into the sample. It is important that these returns contribute to the overall estimate, but we do not want these businesses to represent other similar businesses in terms of industry and size as they are unique in terms of their own characteristics.

Requirement 3 – RUs within a multiple-RU enterprise must be sampled

In order to comply with European regulations, ONS must provide enterprise level data to Eurostat in October 2017. As mentioned in Section 2, most enterprises (businesses) have just one RU and therefore it is simple to produce an enterprise level estimate for these. For enterprises with more than RU, it is important that we have data on all of the RUs to ensure that we can produce a total for the enterprise.

Requirement 4 – Businesses in the Sharing Economy

In response to recommendations set out in the Bean Review, and in order to obtain a better understanding of businesses in the Sharing Economy, National Accounts requested that a small number of businesses be added to the ABS sample. These businesses are likely to have low employment but very high turnover. A small list of 100 businesses were added to the sample,.

Requirement 5 – The Scottish sample should include at least 8,100 businesses

The Scottish Government provides funding for additional questionnaires to be dispatched. For ABS 2016 it was agreed that the total sample size in Scotland should be at least the same as the level obtained in 2010 of 8,100.

Requirement 6 – Businesses with 0-9 employment should remain similar in number

After the 2010 allocation the number of businesses with 0-9 employment increased dramatically, which provided difficult for ONS when response chasing. When allocating the sample, the number of businesses of this size should not increase from this level.

These six requirements were taken into consideration when allocating the sample for ABS 2016.

**4. Methods**

**4.1 Power Allocation**

A power allocation (Sarndal, Swensson and Wretman, 1997) was used to reallocate the ABS sample. The power allocation tries to find a middle ground between optimising for the variance at both the overall level (for example like a Neyman allocation) and the strata levels. For ABS this is appropriate because estimates are published down to the class level (4-digit SIC).

The formula for a power allocation is as follows:



where the population size in a stratum,  , is raised to some power, *a*, and  is the estimate of the population variance in that stratum. Note that this formula is the same as that for the Neyman allocation, except for this exponent *a*. What this allocation does is effectively reduce the impact large population sizes have on the sample size that is allocated to a stratum. This means that more sample can be allocated to the smaller strata and the result is a compromise of the quality of estimates at the overall and strata levels, depending on the value of *a* chosen (as a-->0 the allocation optimises for the strata level and when a=1 this becomes a Neyman allocation).

Another way (Sarndal, Swensson and Wretman, 1997) to look at this type of allocation is to rewrite it as:



where  is some average size measure of the elements in strata *h*. The size measure was chosen to be the average (median) turnover in the universe for each stratum.

The Neyman allocation macro was used to implement the power allocation, to ensure certain constraints could be imposed, such as minimum sample sizes and taking a census of all large businesses. The strata level estimates of the population variances were set to:



This then means that as the usual Neyman allocation has  , we were effectively performing a power allocation with the macro. The following constraints were applied to the allocation:

* median estimates of population variance were calculated using the average turnover on the IDBR.
* census stratum were fully enumerated by setting the sample size to the population size.
* all strata with fewer than 5 businesses in the population were fully enumerated.
* for those strata with employment sizeband 100-999 (select services industries), the sampling fraction was fixed at 50%

The power allocation macro was run twice after checking the outputs from the first run. After the first run, strata with sampling fractions larger than 0.75 were fully enumerated and all others were constrained to have a maximum sampling fraction of 0.5. The allocation was then run again. This is to avoid oversampling, where businesses are dropped by the survey and then picked up again shortly in the next couple of years.

**4.2 Quality of the estimates**

To test the performance of the chosen allocation, it was applied to the past 3 years of data (2012-2014) and then compared with the current allocation. This enabled us to assess the impact on the quality of the estimates (CVs). The following rules were followed:

* if the stratum was fully enumerated in the allocation output then it would be fully enumerated in testing
* if the allocated sample in a cell was larger than the population size in that year then that cell was be fully enumerated in testing
* otherwise the stratum sample size was set to the allocated sample size

Note: the CVs when using the new allocation were calculated assuming 100% response, however current CVs include non-response.

**5. Results**

In total, 54,000 businesses were selected in England and Wales and a further 8,091 businesses were selected in Scotland.

Figure 5-1 shows the difference in the current ABS sample size per SIC section and the new proposed sample size. Many of the sample counts remain similar; however there are some larger differences in some of the sections. The sample size in the manufacturing sector has increased from 7,800 to nearly 14,000 businesses. The sample size in the construction sector has decreased from 7,600 to around 4,400, and the sample in the professional, scientific and technical sector has decreased from around 7,800 to 5,200.

**Figure 5-1
Comparison between current and new sample sizes per SIC section**



The new allocation was then applied back to previous years of data to assess the quality of the estimates at a section, division, group and class level. On average, 80-90% of CVs across all domains remained the same or had improved using the new allocation. For example, in 2014, 90% of the class level CVs remained the same or had improved using the new allocation – with some of the CVs improving by up to 40%.

One of the requirements for the sample was to ensure that the amount of businesses with 0-9 employment did not exceed the current number. The new allocation dramatically reduces the number of businesses with 0-9 employment by nearly 20,000. The sample taken from strata with 0-9 employment has been redistributed among the other employment size bands and has been used to fulfil the additional survey requirements.

**Table 5-2
Comparison between current and new sample sizes per SIC section**

Analysis was also performed to see what impact the new allocation had on CVs by section and employment size band. CVs for the 0-9 size band generally increased due to the sample reduction, however CVs for other size bands improved or remained the same.

**6. Conclusions and Recommendations**

The ABS sample has been allocated using a power allocation – an allocation that aims to find a middle ground between optimising for the variance at the overall level and the strata levels. The new allocation takes into account the changes in the ABS population and fulfils a number of requirements set out by ONS and the business area. The allocation samples 61,988 businesses – 53,897 in England and Wales and 8,091 in Scotland. For the majority all the domains investigated, the quality of the estimates has remained the same or has improved using the new allocation. The ABS sample was selected at the end of December 2016 and provisional results will be published in November 2017.

**7. References**

Sarndal, C,, Swensson, B., and Wretman, J. (1997). “Model Assisted Survey Sampling”. (Springer: New York)

1. Megan Pope, ONS, megan.pope@ons.gov.uk; Jonathan Digby-North, ONS, jonathan.digby-north@ons.gov.uk; [↑](#footnote-ref-1)
2. <https://www.ons.gov.uk/ons/guide-method/method-quality/specific/business-and-energy/annual-business-survey/quality-and-methods/abs-technical-report.pdf?format=hi-vis> [↑](#footnote-ref-2)