**Calculating weights for the BIS Self-employed survey, an LFS follow up**

Katie Connolly

**Abstract**

ONS conducted and delivered results for a telephone survey of 1,500 self-employed individuals in a short space of time for the Department for Business, Innovation and Skills (BIS). The results of the survey would then be used to contribute to the evidence base of an independent review of self-employment being conducted by government.

Participants of the Labour Force Survey (LFS) are interviewed for 5 quarters about their employment status. This allows us to identify self-employed individuals. The list of LFS participants who consented to re-contact was used as a sampling frame to allow for efficient sampling. Households with one or more individuals identifying themselves as being self-employed were sampled and from these a single self-employed individual was interviewed. This sample design has been reflected in the weighting strategy.

This paper highlights how the innovative sampling methodology employed helped enable us to have “impact at pace” and describes the methodology used to calculate the weights.

1. **Background**

The number of self-employed people in the UK increased by around 0.5million between 2010 and 2015. In response to this BIS requested that ONS collects information about self-employed individuals in a short space of time.

Self-employed individuals are a diverse group of people with different circumstances. The purpose of the survey was to collect more detailed data on the experiences of the self-employed that were not available from other sources. This was to help understand what it is currently like to be self-employed in the UK.

The survey was funded by BIS and they expected the data to be a key element in developing the evidence base for a review of self-employment policy.

1. **BIS Self-employed Survey**

The field period for the survey was in 2015. It was a telephone interview allowing for information to be collected quickly and at a relatively low cost compared with other methods such as face-to-face interviews.

An initial question determining if the individual was self-employed was first asked. Those that said they had never been self-employed were excluded from the sample. Those that were still self-employed were asked further questions based on:

* different “types” of self-employment contained in the LFS statistics
* different routes of entry to self-employment
* choice between self-employment and working as an employee
* benefits of being self-employed
* challenges of self-employment
* how LFS self-employment relates to other measurements we make

Those that were no longer self-employed were included in the sample and asked questions such as:

* Why is it that you are no longer self-employed in your main job?
* When you were self-employed as your main activity, to what extent were the following a problem to you?

Those identifying themselves as no longer being self-employed made up approximately 13% of the unweighted sample.

1. **Sample Design for the BIS Self-employed Survey**

Since the purpose of the survey was to collect information on what it is currently like to be self-employed in the UK, the population of interest was defined as all of those identifying themselves as being self-employed in the UK. As questions were asked of people who were self-employed in 2014 but not necessarily at the time of the interview in 2015 this reduces to those who were self-employed in 2014.

For most social surveys the Postcode Address File (PAF) is used as a sampling frame from which addresses are sampled. This does not provide information about which addresses contain an individual who is self-employed. As a result a very large sample would have to be drawn to compensate for the addresses where there were no self-employed individuals present. This would be costly and increase the resources needed to conduct the survey.

However, the LFS collects information on employment from individuals. This can be used to target those that we know have identified themselves as being self-employed allowing for a more efficient sampling frame.

* 1. **Labour Force Survey**

The LFS is a quarterly survey collecting information about employment status, and has a systematic sample drawn from the PAF. This creates implicit geographical strata ensuring that there is a good geographical spread across the country. The sample is unclustered at the household level, but, as all individuals within a household are sampled, the sample is clustered at the individual level.

The LFS has a sample size of approximately 40,000 responding households in each quarter which equates to approximately 90,000 responding individuals. This provides a large number of individuals that may be considered for the sampling frame.

A rotating panel design is used in the sample design of the LFS. This has been illustrated in Figure 1.

Figure 1: Rotating panel diagram

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Q1** | **Q2** | **Q3** | **Q4** | **Q1** | **Q2** | **Q3** | **Q4** |
|  |   |   |  |   | W1 | W2 | W3 | W4 |
| **Panel 1** |   |   |   | W1 | W2 | W3 | W4 | W5 |
| **Panel 2** |   |   | W1 | W2 | W3 | W4 | W5 |  |
| **Panel 3** |   | W1 | W2 | W3 | W4 | W5 |   |   |
| **Panel 4** | W1 | W2 | W3 | W4 | W5 |  |   |   |
|  | W2 | W3 | W4 | W5 |   |   |   |   |
|  | W3 | W4 | W5 |   |   |   |   |   |
|  | W4 | W5 |   |   |   |   |   |   |
|  | W5 |   |   |   |   |   |   |   |

Figure 1 a rotating panel design. The columns represent quarters and the rows represent panels. A panel is sampled for five consecutive quarters each time increasing in wave number from wave 1 to wave 5. Panels that have been sampled for 5 quarters have been shaded in colours. In each quarter there are five different panels present, each in different waves. Each quarter has a new panel represented by wave 1 and loses a panel represented by wave 5 in the previous quarter.

Consecutive quarters have parts of the sample that are common. If only wave 5 from each quarter was to be considered there would be sample no members that are common between quarters as they are all from different panels. Wave 5 cases are used to create a wave 6 dataset that can be used as a sampling frame for follow-up surveys.

Four quarters (January-December) of waves 1 and 5 of the LFS along with some additional boost cases form the Annual Population Survey (APS). This is an annual dataset allowing for annual estimates, such as the total number of people who are self-employed, to be calculated.

* 1. **Wave 6 follow-up**

At the end of the final wave (wave 5) of the LFS those who have responded are asked for consent to being re-contacted. Those who have personally consented to re-contact are said to have given non-proxy consent. If an individual has not responded but another individual in the household did respond and provide consent to re-contact, the individual is said to have given proxy consent. Individuals with proxy or non-proxy consent can be included in wave 6 follow-up surveys.

* 1. **BIS Self-employed Sample Frame**

The sample frame used for the BIS Self-employed survey consisted of those who consented to re-contact following wave 5 in the four quarters of 2014. This included those that provided non-proxy consent in any of the quarters and those that provided proxy consent in either of the last two quarters. Those that provided proxy consent were included to ensure that the sample frame was large enough.

For the survey, contact information was required so records where this was unclear or missing were removed from the sampling frame. Any cases that had already been selected for another LFS follow-up survey such as the LFS dress rehearsal were also removed from the sample frame to reduce respondent burden.

As these individuals have previously answered the LFS, their employment status is known. For the sample frame only those who identified as being self-employed were included. This allowed us to target the sample to our population of interest creating an efficient sample of self-employed individuals to be selected.

* 1. **Sample Selection**

The sample was selected using the sample frame described in Section 3.3. From this frame all households were selected. If there was more than one individual identifying as being self-employed in a household the intention was that one individual was randomly selected. This was because it was thought that they were likely to have similar characteristics. However, in some cases more than one adult was selected and interviewed. This was accounted for in the calculation of the design weights.

A sample of 2,503 individuals was selected, of which 1423 responded and confirmed that they were previously self-employed. Weights were calculated for these cases.

1. **Weighting**

Weights are used to allow estimates of population characteristics to be calculated. In general, for ONS social surveys, design weights are first calculated, to account for the sample design, as the inverse of the probability of selection. Non-response adjustments may be made to reduce non-response bias. Calibration is then used to make the sample more representative of the population and estimates more precise. This ensures that the correct population totals and distributions, such as age, sex, region, are achieved.

Weights were calculated for those who responded and were still self-employed or who were no longer self-employed. Weights were not calculated for those who said that they’d never been self-employed.

* 1. **Design weights**

The January-December APS 2014 dataset contains the four quarters of LFS wave 5 data that we’ve used as a basis for our sample frame as described in Sections 3.1. and 3.3. The probability that a wave 5 case is in the APS is equal to the probability that a wave 5 case is in the corresponding quarter of the LFS. As a result, the design weights for each of the surveys are the same for these cases. It was decided to take the design weight from the APS dataset as a base design weight for the BIS Self-employed survey to save time and increase efficiency as they were all stored in one place.

As all households were selected the probability of selecting a household is equal to one. Therefore no adjustments to the design weights based on household selection were required. The APS design weights were scaled by the number of people that were self-employed in the household and the number of people selected within the household accounting for the probability of selecting a person within a household i.e.



* 1. **Non-response adjustments**

When calculating the weights for the BIS Self-employed survey, explicit sample based non-response adjustments were not made. This was because the weights were required in a short space of time and it was assumed that the main contributory factors to non-response would be age, sex and region which would be accounted for in calibration. Calibration would therefore reduce non-response bias.

* 1. **Population Totals**

The population of interest was those who identified themselves as being self-employed in the UK. As questions were asked of people who were self-employed in 2014 but not necessarily at the time of the interview in 2015 this reduces to those who were self-employed in 2014. These included individuals that were still self-employed in 2015 and those that were no longer self-employed in 2015. The sample would not capture those who had become self-employed since 2014 as none of these cases would have been present on the sample frame. As a result the weights were calibrated to the number of self-employed individuals in 2014. This means that all estimates relate to those who were self-employed in 2014. The total number of self-employed individuals in 2014 was estimated using data from the APS.

* 1. **Calibration**

Calibration was used to ensure that the weights summed to the correct population totals making the estimates representative of the population. This population was defined to be the number of self-employed individuals in 2014 as detailed in Section 4.3.

The intention was to use age, sex and region in some combination as calibration groups, but the practicality of this was limited by small sample sizes. As a result the following separate sets of marginal totals were used:

* Government Office Region (GOR) (11 groups)
* banded age (6 groups)
* sex by over/under 55 years (4 groups)

For these groups Merseyside was combined with the North West as a result of the small sample sizes and sex was split by over/under 55 years as a result of non-response patterns seen in the data.

Calibration using these groups illustrated some discrepancies in totals for small break downs but there wasn’t an obvious pattern underlying this. The number of calibration groups could not be increased due to the small sample size, therefore the best approximation has been made.

1. **Conclusion**

In response to the rise in the number of self-employed people in the UK BIS requested that ONS collects information about self-employed individuals in a short space of time. Using wave 6 of the LFS allowed for targeting self-employed individuals providing an efficient sample frame. Telephone interviews allowed for collection of information about those that were still self-employed and those that were no longer self employed. Using these methods allowed for selection of the sample and weighted responses within the time limitations provided.

The weights were calibrated to an estimate of the number of self-employed people in the UK in 2014 to account for limitations of the sampling frame not allowing for individuals who have become self-employed since 2014 to be sampled. This means that the estimates refer to those that were self-employed in 2014. Calibration was made difficult because of the small sample sizes for particular age groups. Discrepancies in break downs of the final calibrated weights showed no obvious distribution so the best approximation has been made.