Investigating attrition on the UK Labour Force Survey

Matthew Greenaway*, Andrea Lacey† (Office for National Statistics)

1 Introduction

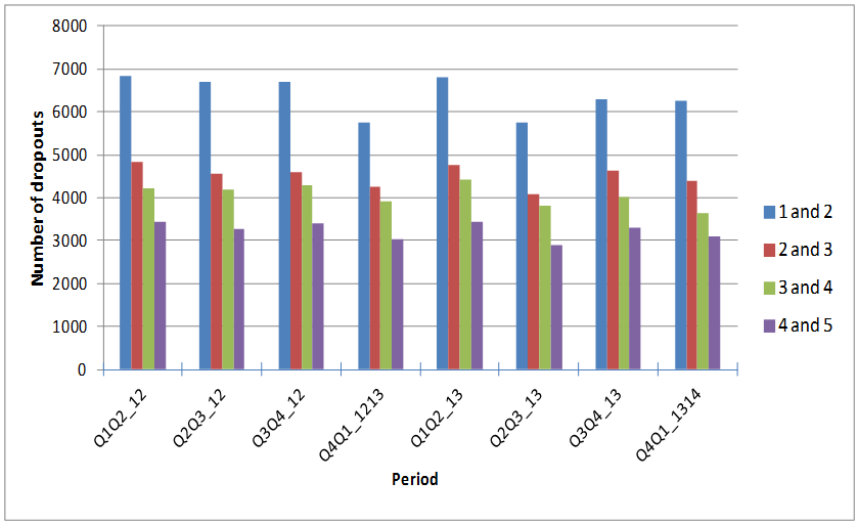
The UK Labour Force Survey (LFS) is a quarterly survey using a 'rotating panel' design - households, once they enter the sample, continue to be sampled for five consecutive quarters. In addition to non-response at first contact, rotating panel designs suffer from 'attrition', which we define as individuals not responding at waves 2-5, given that they have responded in a previous wave. Labour Force Survey attrition has been increasing over time, and a recommendation of the LFS National Statistics Quality review (NSQR) (ONS, 2014) was that an investigation be carried out into bias caused by attrition and ways of correcting this in LFS estimation.

Attrition is not a simple linear process of individuals responding at wave 1 and then gradually dropping out - individuals can drop out and then respond again at later waves. They can also enter the sample for the first time at waves 2-5, for example if new individuals move in to a sampled household, these individuals will be captured by the sample (and any individuals who move out will be dropped). We have limited this study to considering two consecutive LFS quarters and evaluating the individuals who drop out between these. We do not consider individuals who re-enter the sample at later waves.

2 Descriptive Analysis

The greatest number of dropouts is consistently between waves 1 and 2, with progressively fewer dropouts at later waves.

Figure 1: Dropout rates between waves.



A brief literature review suggests that around 30 characteristics have been found to be predictive of attrition and non-response. A selection of variables identified by this review are shown below, alongside the attrition rate between two given quarters for different groups.

*matthew.greenaway@ons.gsi.gov.uk
†andrea.lacey@ons.gsi.gov.uk

Table 1: Dropout rates for key variables (average = 25%) between Q2 and Q3, 2013.

Variable	**Highest dropout rate**	**Dropout rate**	**Lowest dropout rate**	**Dropout rate**
Age band	20-24	43%	65 +	15%
Tenure	Rented accommodation	33%	Owned outright	16%
Region	Inner/Outer London	33%	South West	21%
Household type	2 or more people	46%	1 person	21%
Labour market status	Unemployed	31%	Inactive	19%
Ethnicity	Mixed/Multiple	46%	White	21%
Marital status	Single	30%	Widowed	13%
No. family units	As the number of family units in a household,increases, the dropout rate increases			
Sex	No difference (both 25%)			
Time at address	As the length of time spent at the address, increases, the dropout rate decreases			

The impact of labour market status is particularly notable. However, it is important to emphasise that this table simply reports dropout rates for a number of variables independently.

3 Attrition Model

The variables identified as important by the exploratory analysis were used to model attrition in a number of periods. Six variables were identified as having a consistent and significant impact on attrition - household type, region, age, tenure, ethnicity and disability status. These six variables were used in the final attrition model.

Table 2: Wald statistics and p-values for one attrition model

Variable	Wald χ^2	P value
Region	124.3	<.0001
Tenure	294.8	<.0001
5 Year Age Bands	928.4	<.0001
Household Type	216.3	<.0001
Ethnicity	77.3	<.0001
Disability Status	41.2	<.0001

Although we noted in Table 1 that the dropout rate varies by labour market status, the labour market status variable is not significant in the model - if included it has a p-value of 0.877. This is consistent with earlier research on LFS attrition - in particular Clarke and Tate (1999) - and implies that the likelihood of an individual dropping out of the survey may not be directly influenced by their labour market status after region, tenure, age, household type, ethnicity and disability are controlled for.

The pseudo r^2 for this model is consistently low, at around 8%. Although the model has identified a number of variables which have a significant impact on attrition, overall it still explains relatively little of the variation in attrition.

From studying the odds ratio, we note some key effects (holding other variables constant):

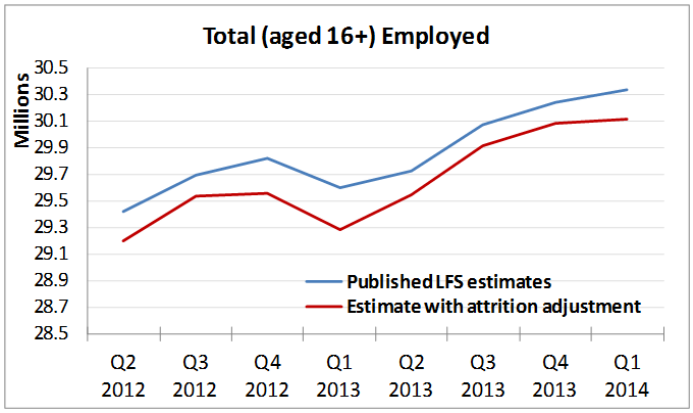
- Those who own their own home outright are considerably less likely to drop out of the survey, while those who rent are more likely to drop out
- Married couple households are less likely to drop out of the survey, while individuals in households containing multiple family units are much more likely to drop out
- Younger individuals are more likely to drop out of the survey
- White individuals are less likely to drop out of the survey
- The odds of dropping out of the survey vary substantially by region

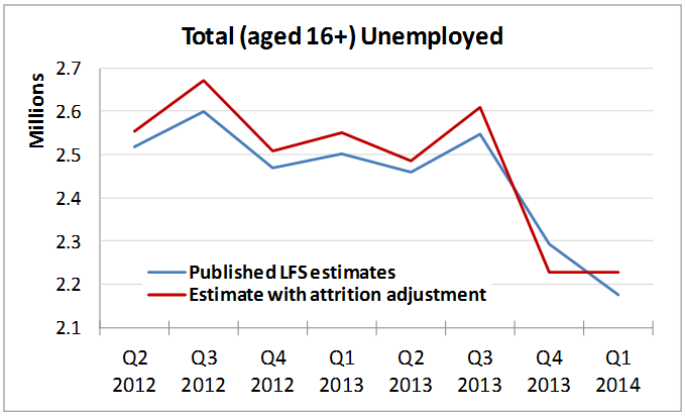
4 Sample-based weighting based on attrition model

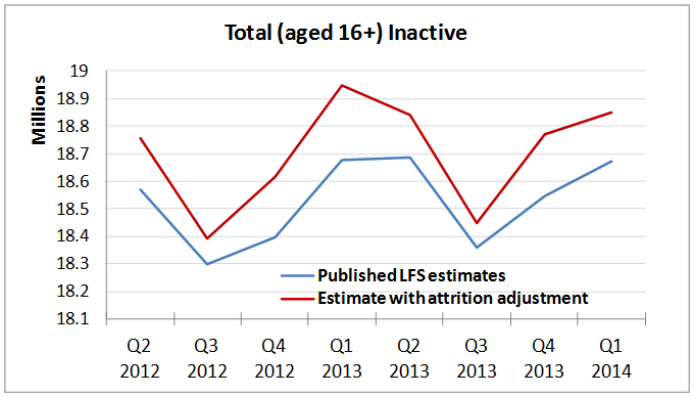
The LFS at present uses a 'population-based' weighting method, where each case is assigned a design weight based on the inverse of their probability of selection, and these weights are calibrated to known population totals. We apply a 'sample-based' adjustment to the design weights using attrition probabilities given by the model as described above, and calibrate these adjusted design weights in the same way as in current LFS estimation. Applying this adjustment ensures that those with a lower probability of staying in the survey get a larger weight, reducing attrition bias, although there will be an increase in standard errors.

Estimates for headline labour market totals under this new weighting scheme are shown in the graphs below. 'Published' estimates are provided for comparison, although it should be noted that these were calculated using ONS research datasets and were not seasonally adjusted, and so will differ from the official published figures.

Figure 2: Graphs showing total aged 16+ who are employed, unemployed and inactive, when an attrition adjustment is applied.







Employment estimates consistently dropped under an attrition adjustment, with inactivity and unemployment rising. For context, the 95% confidence interval for employment estimates is around plus or minus 150,000, meaning the impact on employment is larger than the confidence interval. This analysis suggests that attrition does have a notable and fairly consistent impact on headline LFS totals. Since the impact appears relatively stable, the effect on estimates of period-on-period change may be minor, although further analysis on a period of relative instability is needed to investigate further.

Running logistic regression as a part of the monthly LFS production system may not be practical. Estimating regression parameters using a single period and applying this across multiple periods produces approximately similar results, but this would not properly reflect changing patterns in attrition over time. We have therefore investigated other methods of mitigating attrition bias.

5 Data brought forward

One feature of the current survey methodology which may help adjust for attrition is that data for individuals who drop out of the survey through circumstantial refusals or non-contacts (but not data for hard refusals) is rolled forward for one quarter only - referred to as 'data brought forward'.

To explore the impact that rolling data forward has on estimates, the data brought forward was removed and estimates were calculated and compared both to the current estimates and to the attrition-adjusted estimates calculated in section 4.

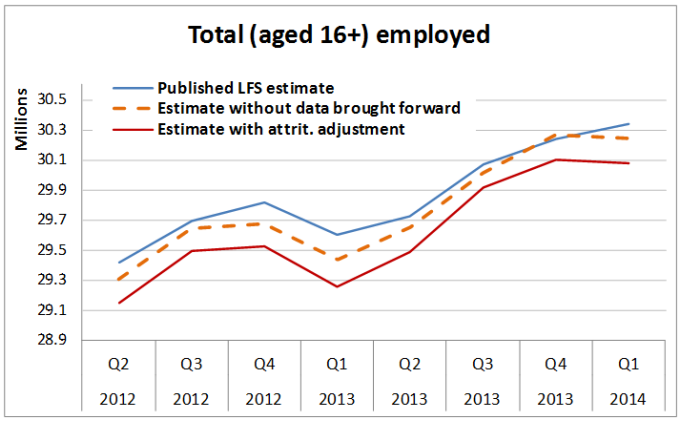


Figure 3: Graph showing total aged 16+ who are employed, when data brought forward is removed.

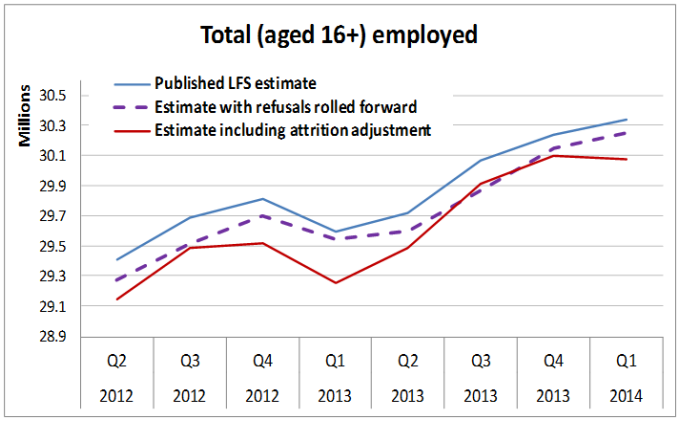


Figure 4: Graph showing total aged 16+ who are employed, when all refusals are rolled forward.

Rolling data forward actually has the opposite effect to adjusting for attrition - the estimates without data-brought forward are closer to the attrition-adjusted estimates. One possible cause of this is the practice for rolling data forwards for only circumstantial refusals and non-contacts, not hard refusals. The graph above compares the current estimates to estimates with all hard refusals brought forwards, again including attrition-adjusted estimates.

Rolling forward data for outright refusals in addition to those who are circumstantial refusals appears to bring the estimates closer to the attrition-adjusted estimates.

It is reasonably clear that the practice of rolling data forwards only for circumstantial refusals and non-contacts increases attrition bias. An alternative imputation method potentially either rolling no data forwards or rolling all data, including for outright refusals - would be preferable. This would need to be subject to further review.

6 Conclusion

Attrition appears to have a notable impact on the levels of key labour market estimates - applying an attrition adjustment based on a logistic regression model consistently decreases employment by more than the 95% confidence interval, with corresponding increases in inactivity and unemployment. This impact does appear to be fairly consistent, although further work on a longer time-span is needed.

The existing method of rolling forwards data appears to be increasing attrition bias, and will also reduce the surveys ability to detect short-term change. This method should be reviewed and replaced with an alternative imputation method.

7 Points for Discussion

- Do other countries have any recommendations on further evaluating attrition bias?
- The paper discusses the method of imputation used on the LFS, where cases are rolled forward. Whilst it is shown that the imputation method increases attrition bias, imputation can also be beneficial in

terms of increasing the overall sample size and the overlap created can also reduce the variance of change. Therefore there needs to be a tradeoff where imputation is concerned. What are the thoughts surrounding imputation? What imputation methods are other countries using, if any?

- Do any other countries have field work stratergies to deal with attrition bias and sample retention? Is there a way we can target those more likely to drop out between waves, perhaps though adaptive designs?

8 References

Ashworth, K., Merad, S., Weeks, A., and Fallows, A. (2013) Nonrespose weights for the Labour Force Survey? Results from the Census non-response link study available at http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/index.html accessed on 1/3/16

Clarke, P. S. and Tate, P. F. (1999) Production and Analysis of Longitudinal Data from the Labour Force Survey GSS Methodology Series no. 17, available at http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

Kanabar, R. (2013) Accounting for attrition in the 2 quarter UK Labour Force Survey, unpublished paper. Office for National Statistics (2014) Review of the Labour Force Survey National Statistics Quality Review: Series 2 report 1