

Improvements to the weighting methodology used in the Labour Force Survey longitudinal datasets

This article summarises the impact of changes made to the weighting methodology used in the Labour Force Survey longitudinal datasets.

Contact:
Bob Watson
labour.supply@ons.gov.uk
+44 (0)1633 455070

Release date:
21 January 2020

Next release:
To be announced

Table of contents

1. [Background](#)
2. [Methodology](#)
3. [Impact assessment](#)
4. [Conclusion](#)
5. [Next steps](#)
6. [Annex](#)

1 . Background

The Labour Force Survey (LFS) uses a rotating panel sample design, whereby each respondent is interviewed on a number of occasions across five consecutive quarters. This makes it possible to follow cases over time by linking together quarterly cross-sectional datasets to analyse gross flows between economic statuses. These longitudinal datasets are available across two consecutive quarters, and across all five quarters.

The sample used in these longitudinal datasets differs from that of the cross-sectional dataset. Respondents of working age are captured by the longitudinal datasets and so the age range is restricted to 16 to 64 years for both males and females. Those who were 15 years old in the first quarter, and 16 years old in the second are excluded from the longitudinal dataset as they are not part of the target population. They are, however, included in the cross-sectional dataset. This results in a difference in the inactivity series between the cross-sectional and longitudinal datasets.

The longitudinal sample also differs from the cross-sectional sample, as those with no information in either quarter are also removed. Respondents may also be lost between cross-sectional datasets through data cleaning, non-contact, or refusal. The sample size of the two-quarter longitudinal dataset is therefore smaller than that of the cross-sectional dataset, and the sample size of the five-quarter longitudinal dataset is smaller still. If the characteristics of those who remain in the sample and those who are lost differ, the estimates of gross flows may be affected. To account for this difference between the cross-sectional and longitudinal samples, the longitudinal datasets use a different weighting methodology to the cross-sectional datasets.

Changes are being made to the weighting methodology used in the LFS longitudinal datasets to improve flows estimates at lower levels of detail. These changes have been implemented from April to June 2012 to reduce the impact of any potential discontinuity on more recent estimates. Both the two-quarter and five-quarter datasets are subject to the weighting modification; however, the focus of this article is on the impact on two-quarter labour market flows estimates, as published in [Table X02](#).

2 . Methodology

Original methodology

The original methodology used to weight the longitudinal datasets was proposed by Clarke and Tate (1999). Weights for the longitudinal dataset are calculated and scaled to replicate the housing tenure distribution observed in the first quarter dataset. This largely mitigates the effects of attrition. The weights are then calibrated to known marginal totals of certain control variables. These variables are age, sex and region, based on mid-year population estimates derived from the census, and economic activity reported in the first and second cross-sectional quarters.

This methodology is adequate in maintaining consistency between the longitudinal and cross-sectional stock estimates for the three headline economic statuses (employed, unemployed, and inactive), albeit with differences because of smaller sample sizes in the longitudinal dataset. Additionally, at further breakdowns of employment such as full-time, part-time, employees, and self-employed, the overall trend is consistent between the longitudinal and cross-sectional datasets. The original weighting methodology successfully reduces non-response bias and the constraints used (tenure, age, sex, region, economic status) continue to be relevant.

Issues with original methodology

While the original methodology is successful in reducing non-response bias, uses relevant constraints, and achieves consistency with the cross-sectional estimates for the headline economic statuses, there are issues with the consistency of estimates at lower levels. Subcomponents of employment (such as full- or part-time employment, and employees or self-employed) reveal less consistency between the cross-sectional and longitudinal estimates. The self-employed and those working full-time are underestimated in the longitudinal datasets compared with the cross-sectional datasets. The lack of consistency at this level was determined to be problematic and potentially confusing for users, resulting in the need for revisions to the longitudinal weighting methodology.

Modified methodology aims

As the original methodology reduces non-response bias, and the constraints used are relevant, it was concluded that much of this methodology should be maintained. The modified methodology therefore aims to improve estimates at lower levels of detail by including employees and self-employed categories in the calibration of the longitudinal weights, addressing the underrepresentation of the self-employed in the longitudinal datasets.

The underrepresentation of full-time workers is maintained despite the addition of these new constraints. While full- and part-time work could also be included as calibration constraints to further improve consistency, this would risk introducing greater variance in the weights, and therefore in the resulting estimates. A balance must be struck between the need for consistency with the cross-sectional dataset and the reduction of variance in the longitudinal estimates.

In addition to the changes to the weighting methodology described before, the number of age categories used in the weights have been reduced from 19 to 12. This is to reflect the fall in sample size over recent years resulting in smaller numbers of cases, which has particularly affected the younger age groups.

3 . Impact assessment

Figure 1: The effect of a change to the weighting methodology is most greatly observed in the self-employed levels

Difference between cross-sectional and longitudinal estimates of employees and self-employed based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019

Figure 1: The effect of a change to the weighting methodology is most greatly observed in the self-employed levels

Difference between cross-sectional and longitudinal estimates of employees and self-employed based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019



Source: Office for National Statistics – Labour Force Survey

The longitudinal estimates based on the modified weighting methodology are more consistent with the cross-sectional estimates than those based on the original weighting methodology. Improved consistency with the cross-sectional estimates is achieved for both employees and self-employed, but with a greater impact on the self-employed estimates.

Flows estimates

As the longitudinal datasets are most useful for examining transitions between labour market statuses across two quarters, the impact of this change in weights on the flows estimates is presented below.

Figure 2 shows the difference to estimates of flows from employment is minimal, with the absolute difference never exceeding 2.0%. For the three headline labour market statuses, the greatest difference in flows between the original and modified weights is 2.1%, and appears in the unemployed to inactivity series, shown in Annex Figure 5. Flows from inactivity are also shown in Figure 6 in the Annex to this paper.

Figure 2: The effect of the change to the weighting methodology on employment flows estimates is minimal

Difference in employment flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019

Figure 2: The effect of the change to the weighting methodology on employment flows estimates is minimal

Difference in employment flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019



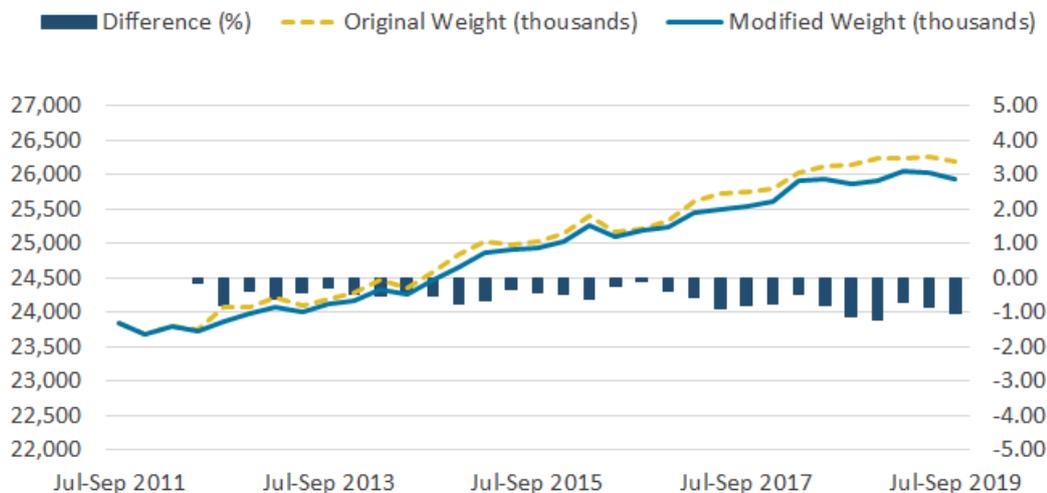
Source: Office for National Statistics – Labour Force Survey

When looking at type of employment, the most notable differences are in the estimates of self-employed flows. Figure 3 shows changes to the employee to employee series as a result of the modified weight reached 1.2%, while changes in the self-employed to self-employed series reached 7.4% (Figure 4).

Other series, including flows from self-employed to employee and vice versa, demonstrated changes that nearly reached 25%. These series can be seen in the Annex to this paper, Figures 7 and 8. This shows the impact of the modified weighting methodology is most greatly observed in series including self-employment, as expected, and had only a minimal effect on other series.

Figure 3: The effect of the change to the weighting methodology on the employee to employee series is minimal

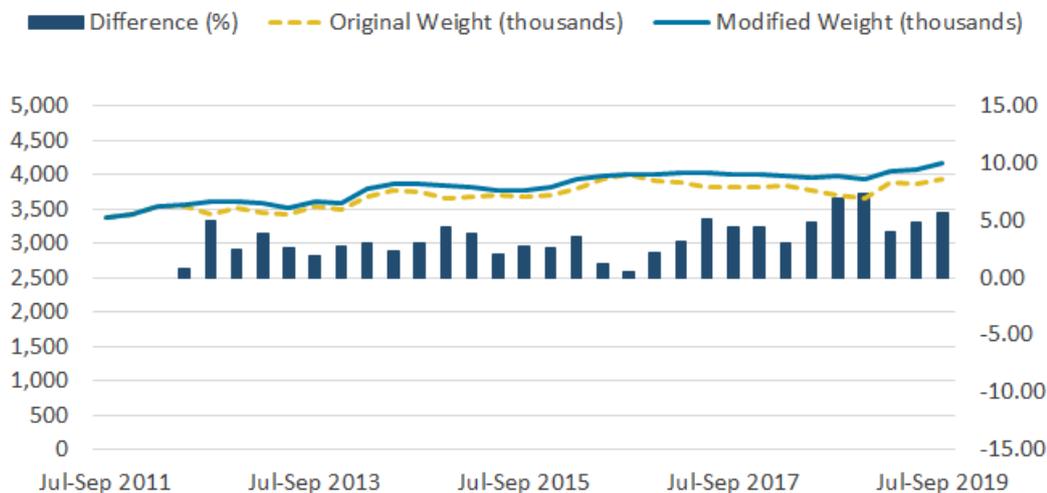
Employee to employee flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, July to September 2011 to July to September 2019



Source: Office for National Statistics – Labour Force Survey

Figure 4: The change to the weighting methodology is more evident in the self-employed to self-employed series

Self-employed to self-employed flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, July to September 2011 to July to September 2019



Source: Office for National Statistics – Labour Force Survey

4 . Conclusion

The methodology used to produce weights for the longitudinal LFS datasets has been reviewed to achieve greater consistency with the cross-sectional LFS estimates. While the original methodology produces consistent estimates of economic statuses at higher levels (employment, unemployment, and inactivity) and consistent trends for smaller breakdowns (full-time, part-time, employees, and self-employed), the estimates for these smaller breakdowns are not consistent.

To balance the needs of improving this consistency, while reducing the potential for increased variance in the estimates, the weighting methodology has been altered to include employees and self-employed as constraints to calibrate the weights, alongside unemployment and inactivity. As expected, this change has affected the estimates of self-employed people to a greater extent than other estimates, both in terms of the levels and the flows between employment statuses. Despite large changes in the self-employed estimates, these have improved the consistency of the estimates of self-employed people in the longitudinal datasets with comparable estimates from the cross-sectional datasets.

5 . Next steps

All series featured in this article are available in the [accompanying dataset](#). In the next release of Table X02 on 18 February 2020, the latest data (October to December 2019) will be based on the modified methodology and all series will be revised back to April to June 2012 on the same basis.

The two- and five-quarter longitudinal datasets containing the modified weights will be made available following the publication of [Table X02](#) in February 2020.

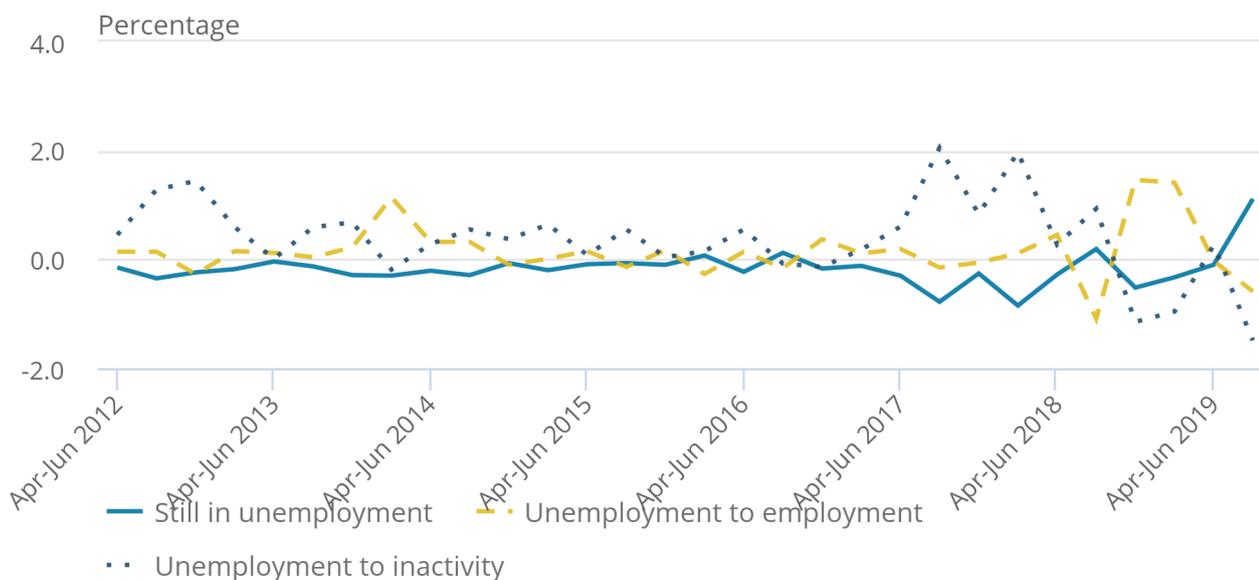
6 . Annex

Figure 5: The effect of the change to the weighting methodology on unemployment flows estimates is minimal

Difference in unemployment flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019

Figure 5: The effect of the change to the weighting methodology on unemployment flows estimates is minimal

Difference in unemployment flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019



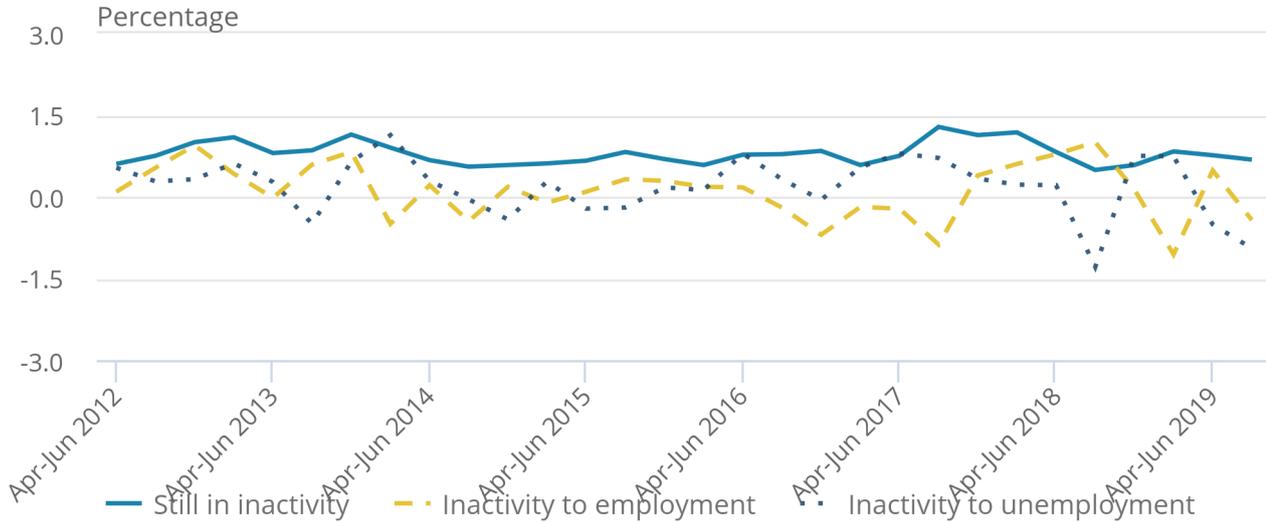
Source: Office for National Statistics – Labour Force Survey

Figure 6: The effect of the change to the weighting methodology on inactivity flows estimates is minimal

Difference in inactivity flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019

Figure 6: The effect of the change to the weighting methodology on inactivity flows estimates is minimal

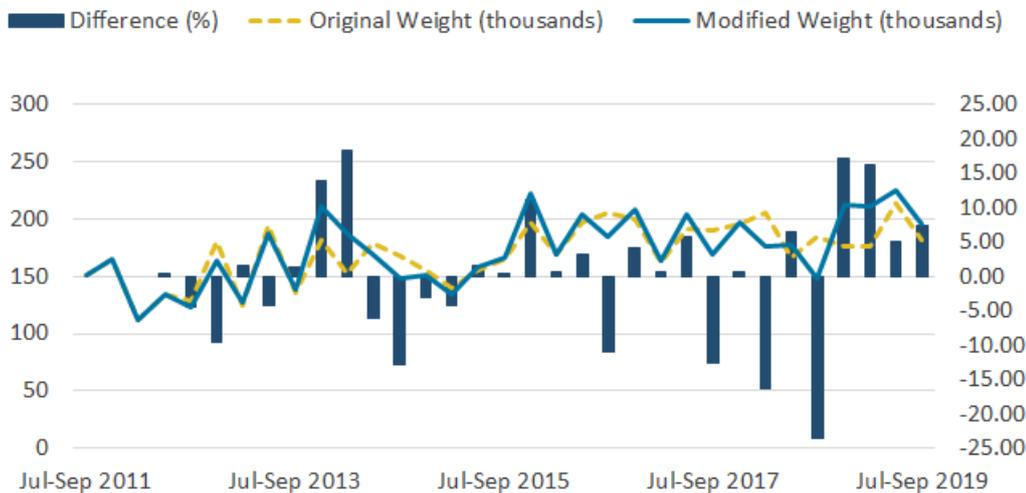
Difference in inactivity flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, April to June 2012 to July to September 2019



Source: Office for National Statistics – Labour Force Survey

Figure 7: The effect of the change to the weighting methodology on the employee to self-employed series is large

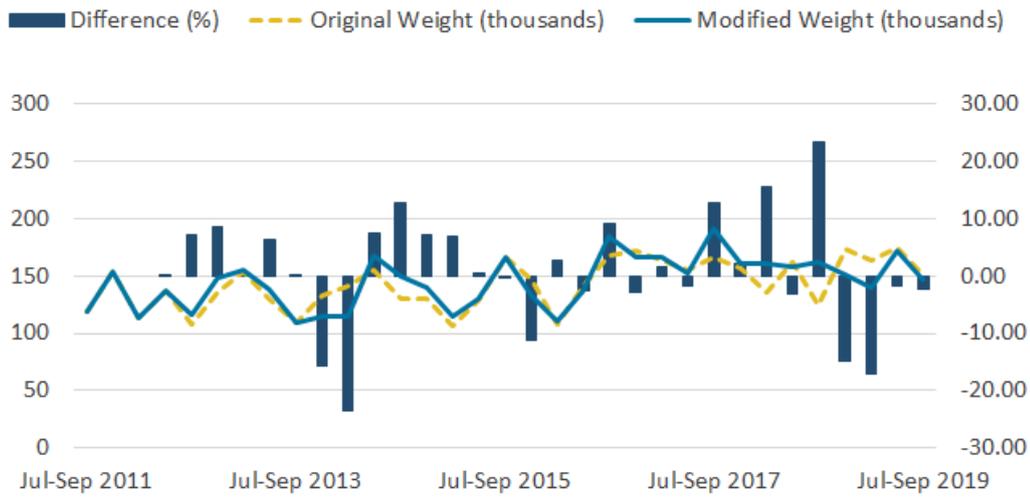
Employee to self-employed flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, July to September 2011 to July to September 2019



Source: Office for National Statistics – Labour Force Survey

Figure 8: The effect of the change to the weighting methodology on the self-employed to employee series is also large

Self-employed to employee flows estimates based on the original and modified weighting methodologies, UK, not seasonally adjusted, July to September 2011 to July to September 2019



Source: Office for National Statistics – Labour Force Survey