Impact of the Covid-19 Pandemic on LFS data

Salah Merad
Methodology Division, MDR, Office for National Statistics

Motivation

The labour Force Survey (LFS) is a rotating panel probability sample survey that is weighted to the UK resident population by age, sex and geography. The population totals for the groups used in weighting are obtained from population projections by single year of age in each local authority that are based on mid-year population estimates.

The data collection is mostly face-to-face in wave 1 and mostly via telephone in waves 2 to 5. Because of the Covid-19 pandemic, face-to-face interviewing was discontinued in mid-March 2020 and telephone interviewing was introduced to wave 1 cases at the end of March 2020. This was achieved by tele-matching wave 1 addresses and making a telephone portal available to contacted addresses. The change in data collection led to a substantial fall in response and an obvious impact on the composition of the responding sample. An example is household tenure: owner occupiers are more likely to have a landline and hence more likely to be tele-matched, making then over-represented in the sample compared with renting occupiers. To address the fall in response, the issued wave 1 sample was doubled from July 2020. To address potential bias from the change in sample composition, we introduced an additional control in relation to household tenure into the weighting – household tenure was found to be stable between 2018 and 2019. This has led to fairly important changes in employment estimates and reduced the observed gap between LFS estimates and HMRC data since the start of the pandemic. We are considering other factors, such as household composition, that could be added into the weighting to reduce further the bias in employment estimates.

Although the adjustment introduced into the weighting has improved the quality of the estimates at national level, some anomalies remain when considering estimates broken down by socio-demographic factors, in particular country of birth. LFS data indicated a big fall in the non-UK born population by Q3 2020, and a large increase in the UK born population – the latter is just a consequence of weighting to a fixed population size. It is unclear how much of this fall is real and how much of it is the result of differential non-response between the UK born and non-UK born populations.

We need to know the actual size of the population and its structure by country of birth since the start of the pandemic to address the quality of estimates by country of birth. The International Passenger Survey (IPS) used to be the main source for migration data but it was suspended in March 2020 because of the pandemic. As part of the transformation of population statistics, we will be using administrative data, including Home Office data, to estimate migration. However, estimates since the start of the pandemic are not yet available.

For more detail about the issue, see
https://www.ons.gov.uk/news/statementsandletters/theimpactofcovid19ontheuklabourmarketestimatesandpopulationflows

and https://blog.ons.gov.uk/2021/01/25/understanding-how-the-pandemic-population/

Therefore, we have attempted to use Real Time Information (RTI) data on employees from HMRC to obtain plausible rates of change in the non-UK born population and then derive revised population totals for use in weighting. RTI data is the only available source that is related to economic activity and is broken by nationality, which is a good proxy to country of birth. Nationality is also collected on LFS, but it is believed that it is subject to a relatively large measurement error compared with country of birth.

In Section A of this paper we present an analysis of the change in the population based on LFS data and the best estimates we can calculate, after controlling for known sources of bias, and explore sample attrition by country of birth. In Section B we present a preliminary proposal for the estimation of RTI-consistent adjusted population totals to use in weighting.

We are asking the panel for feedback on:

- Our assessment of the LFS-based estimates by country of birth
- Our approach to adjust the population totals post-pandemic using RTI data, and the introduction of a control on the population structure in the weighting to address differential non-response between UK born and non-UK born individuals

A. Analysis of Population Changes Estimated using LFS data

Estimates of the non-UK born population obtained using LFS data show a year-on-year fall from Q1 2020 (January 2020 - March 2020) onwards, with the magnitude of the fall increasing in subsequent quarters. In Q3 2020, the year-on-year fall is about 900K, which is equivalent to nearly 10% of the non-UK born population - see Table 1 below.

Table 1. Estimates of non-UK born using all LFS data

Quarter	2019	2020	Change
January - March	9,436,495	9,246,686	-189,809
April - June	9,413,650	8,968,317	-445,333
July - September	9,248,509	8,354,743	-893,766

To obtain more accurate estimates of change, we controlled for potential bias from imputation and the change in data collection since mid-March 2020, when face-to-face interviewing was suspended. We achieved this by excluding imputed cases and cases that joined the survey the change in data collection. Estimates from the weighted reduced dataset indicated that the main outflow occurred in Q2 2020, which resulted in nearly 10% year-on-year fall. The year-on-year change in Q1 2020 was in fact an increase of about 200K, which is similar to the level of annual net migration, whereas in Q3 2020 the year-on-year fall was lower than Q2 2020 by about 300K , which indicates a positive net flow during this quarter.

The large outflow seen in Q2 2020 could be driven in large part by differential attrition between the UK-born and non-UK born populations. Table 2 shows the attrition rates, defined as the percentage of responding individuals in Q1 2020 but did not respond in the following quarter, Q2 2020 - responders in their last wave in Q1 2020 were not included. For the UK born population, the overall

attrition rate was lower in 2020 than in 2019 by more than 3 percentage points, whereas for the non-UK born population it increased by nearly 2 percentage points. This pattern is seen in all age groups, but it was most extreme in the younger age group – however, this extreme pattern wasn't driven by students. We also found that EU born people showed a larger increase in attrition rate than non-EU born people (see Table 3).

Table 2. Attrition analysis by age in LFS

Age group	Country of birth	Jan-Mar 19 - Apr-Jun 19	Jan-Mar 20 - Apr-Jun 20	Difference
18-24	UK	49.7	47.6	-2.1
	Non-UK	56.7	65.6	8.9
18-24 excluding students	UK	51.0	46.7	-4.3
	Non-UK	57.0	65.5	8.5
25-74	UK	30.4	27.4	-3.0
	Non-UK	41.0	41.6	0.6
25-44	UK	42.5	38.4	-4.1
	Non-UK	45.8	45.8	0.0
45-74	UK	24.6	22.0	-2.6
	Non-UK	34.8	36.8	2.0
Under 75	UK	34.2	31.0	-3.2
	Non-UK	41.6	43.3	1.7

Table 3. Comparing attrition rates between EU and non-EU born populations

Age group	Country of birth	Jan-Mar 19 - Apr-Jun 19	Jan-Mar 20 - Apr-Jun 20	Difference	2020 counts
18-24	EU	61.1	68.6	7.4	120
	Non-EU	53.3	62.8	9.5	123
25-44	EU	46.4	47.7	1.3	578
	Non-EU	45.4	44.5	-0.9	713
45-74	EU	31.3	34.8	3.5	275
	Non-EU	36.7	38.0	1.3	611
25-74	EU	40.8	42.6	1.8	853
	Non-EU	41.2	41.3	0.1	1324
Under 75	EU	42.2	45.0	2.8	1136
	Non-EU	41.3	42.4	1.1	1622

The observed differential attrition could be caused by differential behaviour towards survey participation and/or by the fact that the non-UK born population was more likely to be absent from their home addresses – they could be in the UK staying elsewhere or outside the UK. The latter group could be away for a temporary period only and some may still be on the HMRC system – note the positive inflow of 300K in Q3 2020.

B. Using RTI estimates to estimate the non-UK born population

Even after controlling for known potential sources of bias in the estimates of the non-UK born population, the resulting estimates of change seemed rather implausible. Therefore, we needed to compare LFS estimates with estimates from another reliable source which is not subject to non-response bias. PAYE Real Time Information employee data by nationality from HMRC, which is an administrative source, were available for analysis.

We calculated the year-on-year growth rate in total employees for the quarters between Q4 2019 to Q3 2020 for EU nationals and non-EU nationals using RTI data. We also calculated equivalent growth rates in total employment by nationality using LFS data. As can be seen in Table 4, the growth rate of RTI EU employee total decreased from 2.6% to -4.6% between Q4 2019 and Q3 2020. On the other hand, LFS data yield a decrease in growth rate in EU employment from 1.6% to -16% over the same period. Estimates for non-EU nationals show a similar pattern but the difference between RTI data and LFS data is not as extreme.

Table 4. Comparing RTI-based and LFS-based growth rates for the EU and non-EU subpopulations

	RTI data		LFS data	
	Year-on-year percentage change in total employees		Year-on-year percentage change in total employment	
Period	EU	Non-EU	EU	Non-EU
Oct-Dec 2018 to Oct-Dec 2019	2.6	8.4	1.6	3.8
Mar-Jan 2019 to Mar-Jan 2020	0.5	7.1	-0.6	4.0
Apr-Jun 2019 to Apr-Jun 2020	-2.7	4.3	-9.1	1.7
Jul-Sep 2019 to Jul-Sep 2020	-4.6	2.4	-16.2	-4.8

Source: Labour Force Survey (LFS), Real-Time Information (RTI)

The obvious difference in the change in growth rates between LFS and RTI data indicates that LFS most likely suffers from residual differential non-response bias between the UK born and the non-UK born populations. The challenge is how to use RTI data to produce estimates of the size of the non-UK born population in each quarter since January 2020. RTI data are only available since 2014, which makes it difficult to derive a statistical model of population size using RTI data as covariates. We are thinking of deriving an expression of the change in population between a given quarter since January 2020 and a base quarter prior to the start of the pandemic, for example Q4 2019, in terms of change in RTI growth rates in the same period from plausible assumptions. Then, using RTI data and

migration data from the International Passenger Survey from 2014 and 2019, we could assess the estimator.

This would allow us to obtain estimates of population growth rate for the non-UK born population, which could then be used to obtain estimates of the size of this population. The size of the UK born population would be obtained by applying a growth rate from natural change. A calibration control in relation to the structure of the population would be added into the weighting to correct for differential non-response by country of birth.