Labour Force Survey weighting methodology

New population weights using PAYE Real-Time Information data to allow for changes to Labour Force Survey (LFS) collection methods, and changes to population movements in 2020. To be applied to LFS results from July 2021.

Contact:
Salah Merad
labour.market@ons.gov.uk
+44 (0)1633 455400

Release date:
17 May 2021

Next release:
To be announced

Table of contents

1 . Main points

- The population data used to produce labour market estimates are being updated to better reflect changes in international migration and other impacts as a result of the coronavirus (COVID-19) pandemic.

- A model has therefore been developed, using information from the payroll tax system, to provide improved population weights for labour market estimates from 2020.

- The model has been tested against existing data and will be applied in labour market publications from July 2021.

2 . Background (how we got to this point)

The coronavirus (COVID-19) pandemic has posed challenges for how we collect data and how we produce our population, international migration and labour market statistics. We have been conducting work to adapt our methods to better reflect the population changes that have taken place over the course of the pandemic and provide the best possible picture for the public and decision makers.

The population estimates used as part of the Labour Force Survey (LFS) statistics, predate the pandemic and so do not show the demographic and structural impacts. In addition, the profile of responders has changed as the LFS has moved from a mix of face-to-face and telephone interviews to fully telephone based. We reported in October 2020 differences included the proportions responding by characteristics such as ethnicity, disability status, nationality, country of birth and housing.

We introduced tenure weighting which addressed some of the issues, but we acknowledged further work was needed and we committed to introducing further reweighting to improve LFS estimates.

In March we published new data using the HM Revenue and Customs (HMRC) Real Time Information system, which is the source of our monthly employee payroll statistics. Linked to the Migrant Worker Scan, the data helps infer the nationality of employees. It showed smaller falls in employment by non-UK nationals than the LFS suggests (see appendix 2). To further address this issue (which stems from the change in the profile of responders mentioned earlier), we are introducing an additional control in the weighting in relation to the structure of the population by country of birth.

This article explains the method being taken to reweight the LFS, including the introduction of the new controls, alongside setting out our plan for implementing this in our labour market statistics published from July. Revised weights will be applied to LFS data from January to March 2020 onwards. Earlier data will still use weights consistent with the ONS population estimates and projections.

The outcome of a reweighted LFS should not be interpreted as a new estimate for population change nor does it replace official measures of population and migration. Rather it is used to address the issues set out previously, including better reflecting recent population changes since official population projections were last produced.

Using Pay As You Earn Real Time Information data

The challenges in producing official population and migration statistics since the pandemic have led to new innovative ways for measuring population change, using all available data sources. We have chosen to use HMRC Real Time Information (RTI) data to inform the reweighting of the LFS as these data are a complete coverage of payroll employees. When the RTI are linked to Migrant Worker Scan, we can also see different changes to levels of employment based on nationality. Whilst RTI data covers payroll employees, it will not cover those who are self-employed or not in employment. Therefore, we will continue to explore the best possible data sources to further improve our weights in future, alongside our work to transform population and migration statistics using administrative data.

3 . Overview of the reweighting method

To reduce the bias stemming from differential non-response, including between the UK born and non-UK born sub-populations, we need to obtain population estimates by country of birth (UK, EU, non-EU) to use in an additional calibration control for the weighting of Labour Force Survey (LFS) responses. We use Real Time Information (RTI) employee data from HM Revenue and Customs (HMRC) to obtain estimates of the EU and non-EU subpopulation.

Given that available data to estimate a statistical model are limited and a model based on data from the pre-pandemic may not be appropriate from 2020, we have developed a simple and robust method to estimate the population growth rates of the EU and non-EU sub-populations using RTI employee growth rates. This is based on two main assumptions:

- change in the population growth rate of the non-UK sub-populations is in the same direction as the change in their RTI employee growth rate

- the magnitude of change in population growth rate does not exceed that of change in RTI employee growth rate

The method involves adjusting the known population growth rate of a base period before the pandemic with the change in RTI employee growth rates adjusted by a specified factor. The estimator has been obtained in three steps:

1. we show that the change of the population growth rate between a base period, from the pre-pandemic period, and a period from 2020 is approximately proportional to the change in RTI employee growth rates between the same periods; the proportionality factor is unknown but it is shown that it is positive and less than 1

2. the average prediction error of the estimator with respect to the unknown proportionality factor is minimised by adjusting the RTI growth rates by a factor of 1/2

3. to improve the accuracy of the method, we adjusted the RTI employee growth rates by subtracting the RTI growth rates of UK nationals (as this accounts for background change in employment)

Using the RTI employee growth rate to estimate the non-UK population from 2020 is likely to lead to biased estimates as the population tends to change at a lower rate than employment. If the relative bias is constant over time, then it can be shown that the change in population growth rate is approximately equal to the change in RTI employee growth rates. However, when the relative bias varies over time, in the absence of information on the actual magnitude of the variation, it has proven optimal to adjust the RTI growth rates by half.

So, this adjustment is set half-way between adding the change in RTI employee growth rates and making no change to the population growth rate of the base period. Fitting the model to pre-pandemic data, yields adjustment factors equal to 0.45 and 0.42 for the EU and non-EU sub-populations, respectively. This explains why the proposed method performs well using previous data. We expect an adjustment factor of around 0.5 if the employee part of the population is about half the total population and change in the size of the population is dominated by employees leaving or entering the UK. Therefore, if this has continued to hold approximately since 2020, then the prediction error of the estimates based on the proposed method should be small.

By adjusting the weights of the non-UK born respondents using the estimated population growth rates and adjusting the weights of the UK born respondents to reflect natural population change in the 2019 LFS datasets, we obtain estimates of the size of the whole population and by country of birth from 2020 for use in calibration. These modelled LFS estimates, which are informed by movements in non-UK nationals in the RTI employee figures, will then be the basis of reweighting the LFS figures to capture better the population changes over the course of the pandemic.

Also, to reduce the potential bias that stems from the change in the way households are contacted for the first time (use of tele-matching and a telephone portal instead of face-to-face interviewing), as the households that are successfully contacted may be different from the other households, we introduced a non-response adjustment using area level census data.

A detailed description of the method for the estimation of the non-UK born population can be found in the appendices.

4 . Comparison with existing data

The method was tested and evaluated using data from the pre-pandemic period and was found to produce estimates of year-on-year change that are close to previous long-term international migration (LTIM) estimates.

Real Time Information (RTI) data by nationality (UK, EU and non-EU) are available from July 2014 to December 2020 and net migration estimates from LTIM to the end of December 2019 were assessed. These data allow us to evaluate estimates of year-on-year change in the size of the population based on the proposed estimators against actual estimates of net migration over four years: 2016, 2017, 2018 and 2019. We carried out the comparison for the quarter October to December in each year.

Table 1 shows that using the adjusted RTI growth rates performs better in relation to all measures compared with the estimates based on RTI growth rates unadjusted for employee rates of UK nationals. The mean percentage deviation for the estimates based on the adjusted RTI growth rates is very close to 0 (negative 0.01%) whereas the mean deviation is only 2, which indicates that this estimator should be approximately unbiased. We can also see that the average and largest differences from the LTIM estimates are around 20,000 and 50,000 respectively, when the sampling errors of the estimates of year-on-year change in the Labour Force Survey (LFS) are about 70,000.

Using the adjusted RTI growth rates performs better than using the RTI growth rates because the former should be more correlated to population growth as it accounts for background change in employment. The formal proof can be found in Appendix 4.

Table 1: Comparing RTI-based estimates of year-on-year change with LTIM net migration estimates

Period	Country of birth	LTIM (000s)	Unadjusted RTI growth rates (000s)	Adjusted RTI growth rates (000s)
October - December 2016	EU	133	151	170
	Non-EU	179	126	156
October - December 2017	EU	99	79	92
	Non-EU	199	153	174
October - December 2018	EU	75	42	59
	Non-EU	205	172	198
October - December 2019	EU	50	31	56
	Non-EU	282	287	329
Mean percentage deviation from LTIM			-19.3	0.0
Mean percentage absolute deviation from LTIM			23.2	14.3
Largest percentage absolute deviation from LTIM			44.1	27.6
Mean absolute deviation from LTIM (in 1,000s)			28	21
Largest absolute deviation from LTIM (in 1,000s)			53	47

Source: Labour Force Survey (LFS), Real-Time Information (RTI), Long-Term International Migration (LTIM)

The evaluation described above was based on RTI growth rates at national level. RTI data are available by region, at the NUTS1 level, so we considered applying the method separately in each NUTS1 region or area of combined NUTS1 regions, where the NUTS1 regions are small. Table 2 shows the adjustments that would be applied to the population growth rate of the base quarter in different areas and in different rolling quarters. We used adjusted RTI employee growth rates in this calculation. We can see that the trend is the same in all areas but there are differences between the levels.

Table 2: Percentage point adjustments to the October to December 2020 population growth rates by region, based on regional RTI data

Quarter ending	National		South East		London		North		Midlands		East England and South West		Scotland, Wales and Northern Ireland	
	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)
March 2020	-1.1	-0.7	-0.4	0.0	0.0	-0.1	-1.5	-0.3	-1.5	-0.7	-0.7	-0.2	-0.9	-0.7
April 2020	-1.3	-0.7	-0.2	-0.1	-0.4	-0.3	-2.1	-0.6	-2.0	-1.1	-0.9	-0.4	-0.8	-1.1
May 2020	-1.4	-0.8	-0.5	-0.4	-0.9	-0.4	-2.1	-0.5	-2.0	-1.0	-1.0	-0.3	-0.6	-0.8
June 2020	-1.7	-1.0	-0.8	-0.8	-1.2	-0.5	-2.0	-0.5	-1.8	-0.9	-1.1	-0.3	-0.6	-0.6
July 2020	-1.8	-1.3	-1.3	-1.4	-1.5	-0.7	-2.0	-0.6	-1.7	-0.9	-1.3	-0.5	-0.9	-0.6
August 2020	-2.0	-1.5	-1.3	-1.5	-1.6	-0.9	-2.2	-0.8	-1.8	-1.1	-1.4	-0.9	-1.2	-1.2
September 2020	-2.4	-1.7	-1.3	-1.6	-1.8	-1.0	-2.4	-0.9	-2.0	-1.3	-1.6	-1.1	-1.5	-1.6
October 2020	-2.8	-2.0	-1.5	-1.7	-2.3	-1.2	-2.8	-1.0	-2.2	-1.5	-1.9	-1.4	-1.8	-2.0
November 2020	-3.3	-2.4	-1.6	-1.8	-3.0	-1.6	-3.1	-1.2	-2.5	-1.7	-2.2	-1.7	-2.2	-2.6
December 2020	-3.7	-2.7	-1.9	-2.0	-3.8	-2.0	-3.5	-1.5	-2.7	-2.0	-2.5	-2.1	-2.7	-3.0

Source: Real-Time Information (RTI)

We applied the method using RTI data at national level and regional level and compared the national and regional estimates obtained using the two models with estimates of LTIM net migration. The results are shown in Table 3; they indicate that:

- the regional model performs better for regional estimates for both measures

- the regional model performs better for national estimates for one measure and nearly as well for the second measure

Table 3: Comparing the national and regional RTI-based models

Performance measure	Regional estimates		National estimates	
	National model	Regional model	National model	Regional model
Mean absolute deviation from LTIM (in 1,000s)	9	8	21	18
Largest absolute deviation from LTIM (in 1,000s)	57	23	47	48

Source: Labour Force Survey (LFS), Long-Term International Migration (LTIM), Real-Time Information (RTI)

5 . Impact on labour market estimates

The proposed method was applied to the Labour Force Survey (LFS) calendar quarters of 2020. Table 4 shows the year-on-year change in population totals of the UK born and non-UK born sub-populations and of the whole population that were obtained using the Real Time Information (RTI)-based method and the equivalent figures obtain from the published LFS datasets. We can see that in the first quarter, January to March 2020, there is little difference between the overall population totals, but the differences increase over the year. With regard to estimates by country of birth, the differences between the estimates are larger and increase over the year. It is clear that the non-UK born were under-represented in the LFS responses in 2020.

The figures represented in Table 4 do not reflect the official mid-year to mid-year change in official population estimates.

Table 4: LFS year-on-year change in population levels by country of birth

Period	RTI-based method			Labour Force Survey		
	UK	Non-UK	All	UK	Non-UK	All
January - March 2020	95,776	301,756	397,532	574,747	-191,826	382,921
April - June 2020	33,727	264,622	298,349	773,420	-394,105	379,315
July - September 2020	15,886	214,376	230,262	1,252,956	-879,753	373,203
October - December 2020	-1,983	112,044	110,061	1,568,810	-1,202,531	366,279

Source: Labour Force Survey (LFS), Long-Term International Migration (LTIM), Real-Time Information (RTI), ONS Population Projections

Table 5 compares estimates of levels and rates for economic activity over the calendar quarters of 2020. As expected, and as we have stated previously, there is limited impact on the rates as there are similar improvements in the estimation of the numerator and denominator used in the calculation of the rates. There is a large impact on the levels. Here we see the combined impact of non-response adjustment, to adjust for the change in mode of collection, and calibration, to adjust for the under-representation of the non-UK born respondents.

Table 5: Comparing published and Real Time Information-based reweighted Labour Force Survey estimates for 16 to 64 age group

Quarter	Economic activity (ILODEFR)	Labour Force Survey Levels (000s)	Labour Force Survey Rates (%)	RTI-based method Levels (000s)	RTI-based method Rates (%)	Differences Levels (000s)	Differences Rates (%)
January - March 2020	Employed	31,601	76.3	31,589	76.3	-12	0.0
	Unemployed	1,337	4.1	1,331	4.0	-6	-0.1
	Inactive	8,457	20.4	8,484	20.5	27	0.1
April - June 2020	Employed	31,416	75.8	31,261	75.6	-154	-0.2
	Unemployed	1,363	4.2	1,332	4.1	-31	-0.1
	Inactive	8,641	20.9	8,777	21.2	135	0.3
July - September 2020	Employed	31,186	75.3	31,006	75.0	-180	-0.3
	Unemployed	1,593	4.9	1,711	5.2	118	0.3
	Inactive	8,662	20.9	8,635	20.9	-28	0.0
October - December 2020	Employed	31,082	75.0	30,874	74.8	-207	-0.2
	Unemployed	1,714	5.2	1,750	5.4	36	0.2
	Inactive	8,664	20.9	8,676	21.0	12	0.1

Source: Labour Force Survey (LFS), Real-Time Information (RTI), Long-Term International Migration (LTIM), ONS Population Projections

6 . Implementation plan

We plan to implement these changes and recalibrate the weights of all Labour Force Survey (LFS) and Annual Population Survey (APS) datasets, that include data from January 2020 onwards, from July 2021. The publication plan is as follows:

1. LFS person datasets (including LFS weekly), 15 July 2021

2. LFS two-quarter longitudinal dataset, 17 August 2021

3. APS person datasets, 18 August 2021 (aggregate estimates to be published in September)

4. LFS household datasets by early October 2021

5. APS household datasets by early October 2021

7 . Future developments

Alongside the next steps set out here for labour market statistics, we are continuing to work closely with our colleagues across the Government Statistical Service to bring together all available data to deliver the best possible insights on the population and migration.

In June, we will publish official mid-year population estimates for mid-2020, which provide a detailed age and sex profile of the population. We are also continuing our work to transform population and migration statistics using administrative data and are refining our methods for producing Admin-Based Migration Estimates (ABMEs), which we will report on later this year. In 2022, we are planning the provisional release of Census 2021 data which will provide further detail on the population of England and Wales.

Our work across labour market, population and migration statistics will therefore continue to evolve as we bring in new data sources and develop our methods and approaches. As further sources for estimating the size of the population become available, for example the data from Census 2021, the performance of the LFS reweighting model set out in this article will be assessed and a further reweighting may take place if needed.

As stated, in July 2021, we are planning to revise the weighting process for the Labour Force Survey, including taking into account changes in the non-UK population from available administrative data. In August we will be doing the same with the APS, which are used to inform our non-UK population stock estimates. We will then subsequently publish our population of the UK by country of birth and nationality output.

We will continue to provide regular updates on our plans and are always keen to receive feedback. Please contact us at labour.market@ons.gov.uk with any comments.

8 . Appendix 1 - Detailed method

In this section we derive in detail the estimator presented in the previous sections. We start by presenting some notation.

Notation

Let E_{OD19} and N_{OD19} denote the number of employees from Real Time Information (RTI) and the size of the EU subpopulation in the quarter October 2019-December 2019.

Let E_{OD18} and N_{OD18} denote the equivalent quantities for the same quarter one year earlier.

Similarly, let E_{JS20} and N_{JS20} denote equivalent quantities for the quarter July 2020-September 2020 and E_{JS19} and N_{JS19} the equivalent quantities for the same quarter one year earlier.

Let the year-on-year growth rate of the EU subpopulation between OD18 and OD19 be denoted by:

$$\gamma_{OD19,EU} = \frac{N_{OD19}}{N_{OD18}} - 1 \qquad (1)$$

Similarly, let the year-on-year growth rate of the population between JS19 and JS20 be denoted by:

$$\gamma_{JS20,EU} = \frac{N_{JS20}}{N_{JS19}} - 1 \qquad (2)$$

Let the corresponding RTI employee growth rates be denoted by:

$$\theta_{RTI,OD19,EU} = \frac{E_{OD19}}{E_{OD18}} - 1,$$

and

$$\theta_{RTI,JS20,EU} = \frac{E_{JS20}}{E_{JS19}} - 1$$

Derivation of the estimators

The change of population growth rate between the base period, OD19, and the quarter JS20 is given by:

$$\gamma_{JS20,EU} - \gamma_{OD19,EU} = \frac{N_{JS20}}{N_{JS19}} - \frac{N_{OD19}}{N_{OD18}} \qquad (3)$$

Let a_k denote the ratio of the population size to the number of RTI employees in quarter k; that is:

$$a_k = \frac{N_k}{E_k}$$

Equation (3) can then be expressed in terms of the a-ratios and RTI employee totals as follows:

$$\gamma_{JS20,EU} - \gamma_{OD19,EU} = \frac{a_{JS20}E_{JS20}}{a_{JS19}E_{JS19}} - \frac{a_{OD19}E_{OD19}}{a_{OD18}E_{OD18}} \qquad (4)$$

Let:

$$b_{JS20} = \frac{a_{JS20}}{a_{JS19}}$$

and

$$b_{OD19} = \frac{a_{OD19}}{a_{OD18}}.$$

These ratios can also be expressed as:

$$b_{JS20} = \frac{N_{JS20}}{N_{JS19}\left(1 + \theta_{RTI,JS20,EU}\right)},$$

and

$$b_{OD19} = \frac{N_{OD19}}{N_{OD18}\left(1 + \theta_{RTI,OD19,EU}\right)}$$

Each of the ratios can be seen as the ratio of the population size to its naïve estimate, which is based on the assumption that the population grows at the same rate as RTI employee total and hence likely to be biased. We now consider the difference between the two ratios, which we denote by delta; that is:

$$\delta = b_{JS20} - b_{OD19} \qquad (5)$$

Equation (4) then becomes

$$\gamma_{JS20,EU} = \gamma_{OD19,EU} + \left(b_{OD19} + \delta\right)\frac{E_{JS20}}{E_{JS19}} - b_{OD19}\frac{E_{OD19}}{E_{OD18}} = \gamma_{OD19,EU} + b_{OD19}\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) + \delta\left(1 + \theta_{\ldots}\right)$$

In [Appendix 3](#), we show that when assumptions A1 and A2, defined above, hold, we have:

$$\delta \approx \tau\frac{\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right)}{\left(1 + \theta_{RTI,JS20,EU}\right)}$$

where tau is an unknown parameter that takes values between -1 and 0.

Then, Equation (6) becomes

$$\gamma_{JS20,EU} - \gamma_{OD19,EU} \approx \left(b_{OD19} + \tau\right)\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) \qquad (7)$$

We show that to minimize the average prediction error and mean square error with respect to the unknown distribution of tau, population growth should be estimated by setting the value of tau in Equation (7) to -1/2. Its estimator; is given by:

$$\hat{\gamma}_{JS20,EU} = \gamma_{OD19,EU} + \left(b_{OD19} - \frac{1}{2}\right)(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}) \qquad (8)$$

Using past data, we can estimate the value of the population growth of the base period. The estimate of the ratio b_{OD19} is 0.99, which can be approximated by 1. Hence, the approximately minimum prediction estimator is given by:

$$\hat{\gamma}_{JS20,EU} = \hat{\gamma}_{OD19,EU} + \frac{1}{2}(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}) \qquad (9)$$

If the b-ratios do not vary over time, then, given they are close to 1, it can be shown that the change in population growth rates is approximately equal to the change in RTI growth rates. If the b-ratios vary, which we saw in previous data, then, in the absence of information on the magnitude of the variation, the average prediction error is minimised by adjusting the RTI growth rates by half. So, the parameter tau accounts for the sensitivity of the method to the assumption on the equality of the b-ratios. We hence refer to the parameter tau as the model-sensitivity parameter.

In [Appendix 3](), we show that we can reduce the prediction error of the estimator given in equation (8) by using adjusted RTI growth rates. These rates account for the background growth rate among UK employee nationals, which is affected only minimally the growth of the UK born population. For the quarter OD19, the adjusted RTI growth rate is given by:

$$\theta_{RTI,OD19,EU,adj} = \theta_{RTI,OD19,EU} - \theta_{RTI,OD19,UK} \qquad (10)$$

The adjusted RTI growth rate is defined in the same way for other quarters.

Hence, the alternative estimator of growth, which uses adjusted RTI employee growth rates, is given by

$$\hat{\gamma}_{JS20,EU,adj} = \hat{\gamma}_{OD19,EU} + \frac{1}{2}(\theta_{RTI,JS20,EU,adj} - \theta_{RTI,OD19,EU,adj}) \qquad (11)$$

The prediction error of this estimator is given by

$$e_{JS20,adj} \approx \left(\hat{\gamma}_{OD19,EU} - \gamma_{OD19,EU}\right) + \left(\frac{1}{2} + \tau\right)(\theta_{RTI,JS20,EU,adj} - \theta_{RTI,OD19,EU,adj}) \qquad (12)$$

It can be seen that the prediction error will be small if the actual unknown value of tau is close to -1/2, which is equivalent to the actual change in population growth rate being close to half that of the change in RTI employee growth rate.

Sensitivity analysis

To carry out sensitivity analysis, we considered the estimator

$$\hat{\gamma}_{JS20,EU,b} = \hat{\gamma}_{OD19,EU} + b\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) \qquad (13)$$

where the coefficient b can take different values but applied to periods between 2016 and 2019 for the October-December quarter, with the quarter October-December 2015 as the base period. This estimator is based on the UK-unadjusted RTI growth rates.

Table A1 shows the estimates we obtained when using Equation (13) with values of the model coefficient between 0.3 and 0.6. It can be seen that the estimates aren't very sensitive to small changes of the value of model coefficient but, overall, it performed much better when the model coefficient was set to 0.4 or 0.5, producing estimates fairly close to LTIM estimates overall. Similar results were obtained using adjusted RTI growth rates.

Table A1: RTI-based estimates of year-on-year change in population size by country of birth for different values of the model coefficient (figures are in 1,000s)

| Period | Country of birth | | Model coefficient | | | |
		LTIM	0.3	0.4	0.5	0.6
October - December 2016	EU	133	185	168	151	134
	Non-EU	179	138	132	126	120
October - December 2017	EU	99	148	114	79	44
	Non-EU	199	158	155	153	151
October - December 2018	EU	75	129	85	42	-2
	Non-EU	205	168	170	172	174
October - December 2019	EU	50	117	74	31	-11
	Non-EU	282	238	263	287	312

Source: Labour Force Survey (LFS), Real-Time Information (RTI), Long-Term International Migration (LTIM)

Equation (13) can also be written as

$$\hat{\gamma}_{Q,EU} - \hat{\gamma}_{Q_0,EU} = b \left(\theta_{RTI,Q,EU} - \theta_{Q_0,EU} \right) \qquad (14)$$

where Q_0 denotes the base period and Q denotes a period later than the base period. If past data were available to compute the change in RTI employee growth rates and population growth rates from a base period Q_0 for several time points Q, that is:

$$\theta_{RTI,Q,EU} - \theta_{RTI,Q_0,EU}$$

and

$$\hat{\gamma}_{Q,EU} - \hat{\gamma}_{Q_0,EU}$$

We could estimate b by fitting a regression model with no intercept.

We fitted a ratio model to the changes between population growth rates of the same quarter in four years from 2016, with the changes in RTI employee growth rates being the model covariate. We obtained estimates of the coefficient b equal to 0.45 and 0.42 for the EU and non-EU sub-populations, respectively. Both estimates are close to 1/2, the value of the coefficient of the minimum average prediction error estimators we derived.

9 . Appendix 2 - Comparing LFS-based and RTI-based estimates

As can be seen in Table A2, according to Real Time Information (RTI) data, the year-on-year employee total growth rate for EU nationals decreased from 2.6% in Oct-Dec 2019 to -4.6% in Jul-Sep 2020, whereas using Labour Force Survey (LFS) data the year-on-year total employment growth rate for the EU born decreased from 1.6% to -16.2% over the same period. For the non-EU, RTI data show a decrease from 8.4% to 2.4% whereas LFS data show a decrease from 3.8% to -4.8%. The width of the range of growth rates is larger in the LFS for both the EU and non-EU sub-populations.

Table A2: 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)

10 . Appendix 3 - Derivation of the optimal value of the model-sensitivity parameter tau

We present the derivation for the model with unadjusted Real Time Information (RTI) growth rates based on the b-ratios.

The b-ratios b_{OD19} and b_{JS20} can be written as:

$$b_{OD19} = \frac{1 + \gamma_{OD19,EU}}{1 + \theta_{RTI,OD19,EU}}$$

and

$$b_{JS20} = \frac{1 + \gamma_{JS20,EU}}{1 + \theta_{RTI,JS20,EU}}.$$

The expression of the difference between them, delta, can then be written as:

$$\delta = \frac{1 + \gamma_{JS20,EU}}{1 + \theta_{RTI,JS20,EU}} - \frac{1 + \gamma_{OD19,EU}}{1 + \theta_{RTI,OD19,EU}}$$

which yields, after some algebra, the expression

$$\delta = \frac{\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}}{(1 + \theta_{RTI,JS20,EU})} \frac{1}{(1 + \theta_{RTI,OD19,EU})} \left(-1 + \frac{\gamma_{JS20,EU}(1 + \theta_{RTI,OD19,EU}) - \gamma_{OD19,EU}(1 + \theta_{RTI,JS20,EU})}{\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}} \right)$$

As:

$$1 + \theta_{RTI,OD19,EU} \approx 1$$

and:

$$1 + \theta_{RTI,JS20,EU} \approx 1 ,$$

the expression of delta can be approximated as:

$$\delta \approx \frac{\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}}{(1 + \theta_{RTI,JS20,EU})}\tau$$

where tau is defined as:

$$\tau = -1 + \frac{\gamma_{JS20,EU} - \gamma_{OD19,EU}}{\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}}$$

If assumptions A1 and A2, defined above, hold, then we have:

$$-1 \leq \tau \leq 0$$

These assumptions should hold in most cases when change is not very small to be swamped by sampling and non-sampling error because employees form a large part of the population and the labour market is more changeable than the overall population. Table A3.1 shows the change in growth rates and the ratio between the changes. It can be seen that when the change in RTI growth rates is relatively large, the change in population growth rates is lower by a factor close to 0.5. On the other hand, when the change in RTI growth rates is relatively small, the change in population growth rates is much higher.

Table A3.1: Comparing changes in population and RTI employee growth rates

	EU			Non-EU		
	Population (%)	RTI (%)	Ratio	Population (%)	RTI (%)	Ratio
October - December 2016	-3.3	-5.4	0.61	0.5	-1.1	-0.41
October - December 2017	-4.6	-10.4	0.44	0.6	-0.4	-1.51
October - December 2018	-5.4	-12.7	0.42	0.8	0.4	2.03
October - December 2019	-6.0	-13.2	0.46	2.2	4.5	0.48

Source: Labour Force Survey (LFS), Real-Time Information (RTI), Long-Term International Migration (LTIM)

We can write:

$$\gamma_{JS20,EU} - \gamma_{OD19,EU} = b_{OD18,OD19}\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) + \delta\left(1 + \theta_{RTI,JS20,EU}\right)$$

So, substituting delta by its approximation given above, we obtain:

$$\gamma_{JS20,EU} \approx \gamma_{OD19,EU} + \left(b_{OD18,OD19} + \tau\right)\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right).$$

We propose to use the following estimator:

$$\hat{\gamma}_{JS20,EU} = \gamma_{OD19,EU} + \left(b_{OD18,OD19} + \tau_0\right)\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right)$$

where τ_0 is a specified value. Then, the prediction error is given by:

$$\hat{\gamma}_{JS20,EU} - \gamma_{JS20,EU} \approx \left(\tau_0 - \tau\right)\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right)$$

We want to find the value of τ_0 that minimises the mean prediction error, that is:

$$\min\left(\hat{\gamma}_{JS20,EU} - \gamma_{JS20,EU}\right) \approx \left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right)\min E\left(\tau_0 - \tau\right)$$

Because the value of tau lies mostly between -1 and 0 but its expected value is unknown, the mean prediction error is minimised by setting $\tau_0 = -1/2$, that is in the middle of the interval of plausible values of tau.

Formally, the minimum of the expected prediction error is achieved when

$$E\left(\tau_0 - \tau\right) = 0$$

which yields:

$$\tau_0 = E\left(\tau\right)$$

As E(tau) is unknown, we assume that it's a random variable with a uniform distribution over the interval (-1,0). The estimate of τ_0 is then given by the expectation of E(tau), which is equal to -1/2.

Minimising the mean square prediction error yields the same result.

The few years of RTI data we have available to estimate tau showed that many of the estimates were close to -1/2 (see Table A3.2), which means that the prediction error should be small when setting the factor tau to -1/2. There were also a few estimates of tau outside the interval (-1,0). These occurred when the change in RTI growth rates was relatively small; this is not a surprise as estimates of change become swamped by measurement error and sampling error. However, because the prediction error is proportional to the change in RTI growth rates, it should still be relatively small.

Table A3.2: Estimates of model-sensitivity parameter using past data

	Deviation from ratio of base period		RTI growth rate change		Model-sensitivity parameter	
	EU (%)	Non-EU (%)	EU (%)	Non-EU (%)	EU	Non-EU
October - December 2016	1.6	1.5	-5.4	-1.1	-0.30	-1.30
October - December 2017	4.9	1.0	-10.4	-0.4	-0.47	-2.43
October - December 2018	6.3	0.2	-12.7	0.4	-0.50	0.53
October - December 2019	6.0	-2.4	-13.2	4.5	-0.46	-0.54

Source: Labour Force Survey (LFS), Real-Time Information (RTI), Long-Term International Migration (LTIM)

11 . Appendix 4 - Comparing the original and adjusted RTI-based estimators

We now attempt to show that using adjusted Real Time Information (RTI) rates in the estimator leads to a lower magnitude expected prediction error than when using the original, unadjusted, RTI growth rates.

Let B and C denote the prediction errors when using the predictors/estimators with the original and adjusted RTI growth rates, respectively – in both estimators τ_0 is set equal to -1/2. We then have:

$$B = -\frac{1}{2}\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) - \delta_b\left(1 + \theta_{RTI,JS20,EU}\right)$$

and:

$$C = -\frac{1}{2}\left(\theta_{RTI,JS20,EU,adj} - \theta_{RTI,OD19,EU,adj}\right) - \delta_c\left(1 + \theta_{RTI,JS20,EU,adj}\right) = -\frac{1}{2}[\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}\right) - \left(\theta_{RTI,JS}$$

We compare the magnitude of the errors B and C when growth rates are decreasing, but the results will hold when they are increasing.

Let:

$$B_1 = -\frac{1}{2}(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU})$$

and:

$$B_2 = -\delta_b\left(1 + \theta_{RTI,JS20,EU}\right)$$

Similarly, let:

$$C_1 = -\frac{1}{2}[(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}) - (\theta_{RTI,JS20,UK} - \theta_{RTI,OD19,UK})]$$

and:

$$C_2 = -\delta_c\left[(1 + \theta_{RTI,JS20,EU}) - \theta_{RTI,JS20,UK}\right]$$

Case 1: B_1, B_2, C_1 and C_2 are of the same sign

Without loss of generality, we assume that all the terms are positive.

Because growth rates are assumed to be decreasing, B_1 and B_2 are positive. Also, growth rates should be much lower than 1, making the terms:

$$(1 + \theta_{RTI,JS20,EU})$$

and:

$$(1 + \theta_{RTI,JS20,EU}) - \theta_{RTI,JS20,UK}$$

almost always positive. Therefore, delta$_b$ and delta$_c$ are almost always negative. The difference in magnitude between the prediction errors B and C is given by:

$$\Delta_{b,c} = B - C = -\frac{1}{2}(\theta_{RTI,JS20,UK} - \theta_{RTI,OD19,UK}) - (\delta_b - \delta_c)\left(1 + \theta_{RTI,JS20,EU}\right) - \delta_c\theta_{RTI,JS20,UK}$$

The last term is very small but positive.

Given that:

$$\delta_b < 0$$

and:

$$\delta_c < 0$$

the inequality:

$$E\left(\delta_b\right) \leq E\left(\delta_c\right)$$

should hold as in the adjusted RTI rates we account for background change in employment.

So, we obtain:

$$E(\Delta_{b,c}) = E\left(B - C\right) > -\left(\theta_{RTI,JS20,UK} - \theta_{RTI,OD19,UK}\right) > 0$$

which proves that on average the magnitude of prediction error of the estimator based on the adjusted RTI growth rates is lower.

So, when B_1, B_2, C_1 and C_2 are of the same sign, it is very likely that on average the magnitude of the prediction error of the estimator based on the adjusted RTI growth rates is lower than that based on the original, unadjusted, RTI growth rates.

Case 2: B_1 and C_1 are positive but B_2 and C_2 are negative

Let:

$$\Delta_{b,c,1}$$

denote the difference between B_1 and C_1 and let:

$$\Delta_{b,c,2}$$

denote the difference between B_2 and C_2.

then, we have:

$$E(\Delta_{b,c,1}) = E(B_1 - C_1) = -\frac{1}{2}(\theta_{RTI,JS20,UK} - \theta_{RTI,OD19,UK}),$$

and:

$$E(\Delta_{b,c,2}) = E(B_2 - C_2) = -E(\delta_b - \delta_c)(1 + \theta_{RTI,JS20,EU}) - E(\delta_c)\delta_{RTI,JS20,UK} \approx -E(\theta_b - \theta_c)(1 + \theta_{RTI,JS20,EU}).$$

Substituting delta$_b$ and delta$_c$ with their expressions yields:

$$E(\Delta_{b,c,2}) \approx -(1 + \theta_{RTI,JS20,EU}) E\left(\tau_b \frac{\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU}}{1 + \theta_{RTI,JS20,EU}} - \tau_c \frac{\theta_{RTI,JS20,EU,adj} - \theta_{RTI,OD19,EU,adj}}{1 + \theta_{RTI,JS20,EU,adj}}\right)$$

Assume that:

$$E(\tau_b) = E(\tau_b) = -\frac{1}{2}$$

we then have:

$$E\delta_{b,c,2} \approx \frac{1}{2}\left(\theta_{RTI,JS20,EU} - \theta_{RTI,OD19,EU} - \frac{1 + \theta_{RTI,JS20,EU}}{1 + \theta_{RTI,JS20,EU,adj}}(\theta_{RTI,JS20,EU,adj} - \theta_{RTI,OD19,EU,adj})\right) \approx \frac{1}{2}(\theta_{RTI,JS20,UK}$$

hence, the difference in expected prediction error is given by:

$$E(\Delta_{b,c}) = E(\Delta_{b,c,1}) + E(\Delta_{b,c,2}) \approx 0$$

We can thus conclude that the two estimators should have approximately an expected error of the same magnitude.

So, overall, the adjusted RTI-based estimator has an expected prediction error that is equal or lower than that of the original RTI-based estimator.