Labour Force Survey: alternative imputation during the coronavirus (COVID-19) pandemic

During normal operation, Labour Force Survey (LFS) person-weighted datasets use a roll-forward imputation. This article discusses an alternative imputation method that might better cope with the unprecedented labour market circumstances brought about in the pandemic and highlights the impact on main labour market estimates.

Table of contents

1 . Main changes

- The impact of the alternative imputation methodology on the story of headline labour market statuses (employed, unemployed or economically inactive) was minimal, with a maximum of 0.3 percentage points difference, within sampling variability.

- The alternative imputation had a larger impact on total hours worked estimates.

- The impact of the alternative imputation was not equally distributed between industries.

2 . Overview

During normal operation, Labour Force Survey (LFS) person-weighted datasets use a roll-forward imputation. If we have a previous response, followed by a period of non-response, the previous response will be rolled forward to be used in the next period. This roll-forward is only allowed for one period.

The basis for this method is, although things change, responses in a previous quarter are a very good indicator of what that person would be doing in the next quarter. We only roll-forward for one quarter because by a second quarter, six months after the last recorded interview, circumstances become much more likely to have changed. If we fail to get a response for a second consecutive quarter they are dropped from the survey.

Another advantage with the roll-forward method is that it gives a fully cohesive record for the period, with variables such as labour market status, industry, occupation, earnings, hours and so on, all being sensibly aligned with one another, because they all come from the same interview for that specific person.

Issue

Although this method is sensible at most times, there was concern that the rapid rate of change of circumstances during the coronavirus (COVID-19) pandemic might mean that what someone was doing in the previous quarter was no longer a good predictor of what they were doing in the current quarter. Although this could affect many variables, our primary areas of concern were whether rapid job losses would mean that the previous quarter's labour market status became a poor indicator of current labour market status. Also, with the shutdown of certain industries and the introduction of the furlough scheme, whether the number of hours worked in the previous quarter would be a reasonable estimate of the number of hours worked in the following quarter.

We decided to look at alternative imputation methods that might better cope with the unprecedented labour market circumstances brought about in the pandemic.

Method

Since June 2020, we have been trialling an alternative method involving donor imputation. Instead of rolling responses forward from the previous quarter, missing values because of non-response are imputed from another respondent, referred to as the "donor". Where data would have been rolled forward, a suitable donor who returned data in the current period is searched for and their responses used to impute for the missing values. Missing respondents whose responses were to be imputed – referred to as "recipients" – were matched to donors using a "nearest neighbour" approach.

Potential suitable donors are first identified using information on age, sex, geography, labour market status, industry, occupation, hours and ethnicity. Details of the individual variables that were used and how they feed into the identification of a donor can be found in Section 5: Nearest neighbour methodology.

Once a suitable donor is identified we use their responses for suitable variables relating to hours worked and employment status, and some related variables (see Section 5). All other variables continue to be rolled forward.

We limited the method to these main variables because imputing values for all variables would involve much greater complexity. For example, some variables, such as someone's educational status, would be more appropriate to be rolled forward than taken from a donor. Whereas other variables, such as industry, would be more appropriate rolled forward if the labour market status has not changed, but would be more difficult if someone were imputed to change labour market status from unemployed to employed.
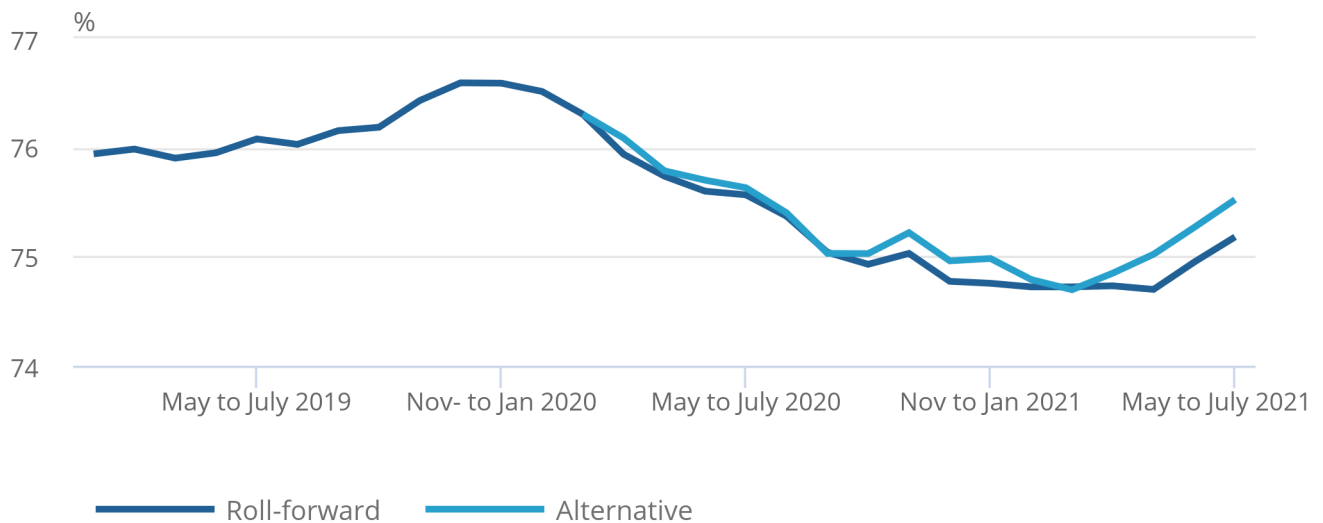
3 . Findings

Figure 1: Employment rates have been minimally affected by the alternative imputation method, although it suggests stronger growth in recent periods

UK employment rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

Figure 1: Employment rates have been minimally affected by the alternative imputation method, although it suggests stronger growth in recent periods

UK employment rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021
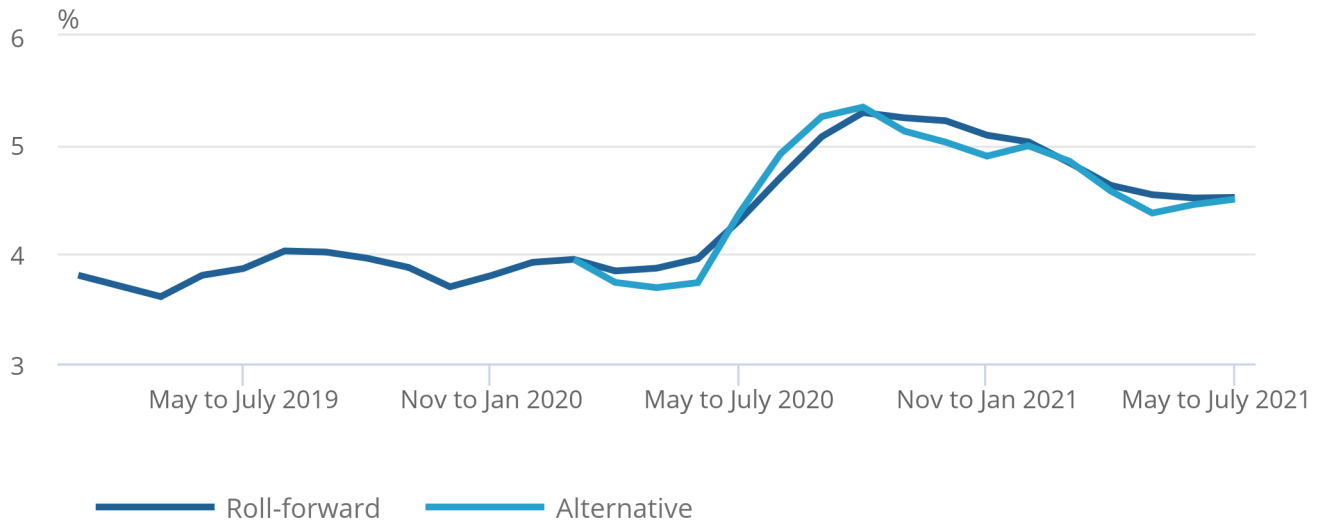
Source: Office for National Statistics - Labour Force Survey

Figure 2: Unemployment rates have been minimally affected by the alternative imputation methodology

UK unemployment rates, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

Figure 2: Unemployment rates have been minimally affected by the alternative imputation methodology

UK unemployment rates, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021
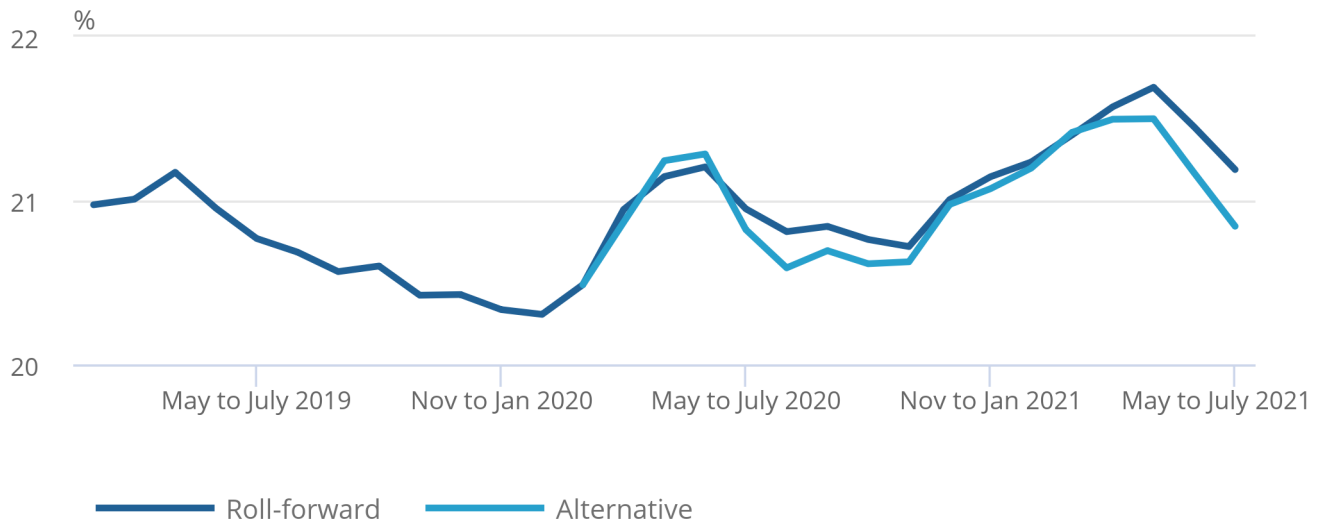

Source: Office for National Statistics - Labour Force Survey

Figure 3: Economic inactivity rates have been minimally affected by the alternative imputation method, although it suggests stronger decreases in recent periods

UK economic inactivity rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

Figure 3: Economic inactivity rates have been minimally affected by the alternative imputation method, although it suggests stronger decreases in recent periods

UK economic inactivity rates, people aged 16 to 64 years, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

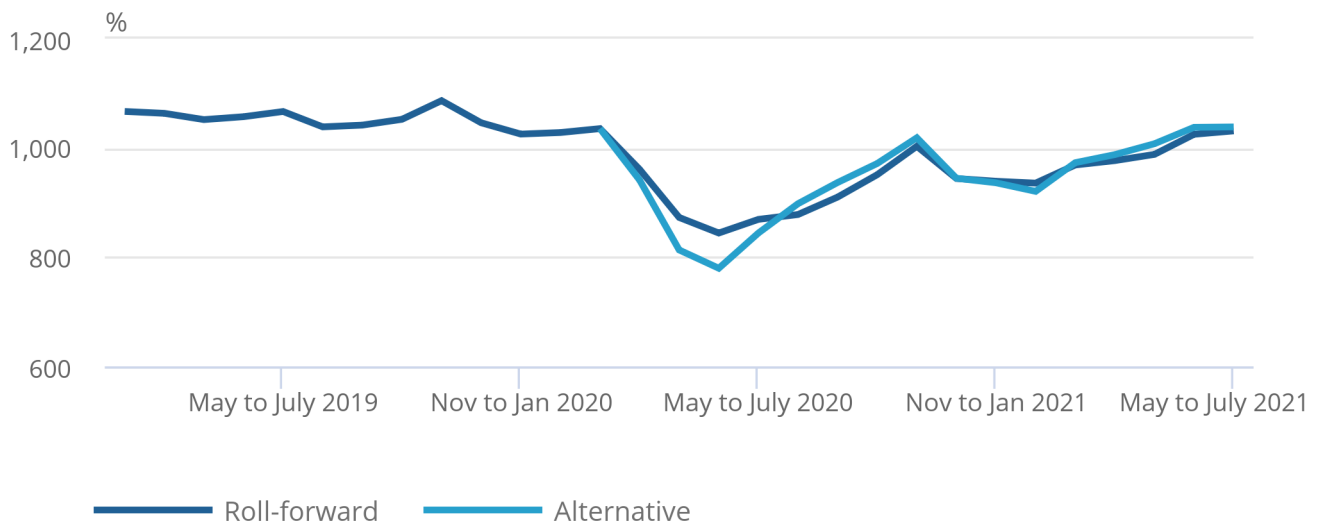**Source: Office for National Statistics - Labour Force Survey**

The impact of the alternative imputation methodology on the story of headline labour market statuses was minimal, with a maximum of 0.3 percentage points difference, shown in Figures 1 to 3. These differences are within sampling variability of the survey estimates.

Figure 4: The alternative imputation method impacted estimates of total hours worked to a greater extent than labour market statuses, particularly in initial stages of the pandemic

UK total hours worked, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

Figure 4: The alternative imputation method impacted estimates of total hours worked to a greater extent than labour market statuses, particularly in initial stages of the pandemic

UK total hours worked, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, January to March 2019 to May to July 2021

Source: Office for National Statistics - Labour Force Survey

The alternative imputation did have a larger impact on total hours worked estimates, particularly at the start of the coronavirus (COVID-19) pandemic, as shown in Figure 4. In the initial stages of the pandemic, the alternative methodology suggested a decrease in hours approximately 7.6% more than the original methodology. While these differences have decreased in later stages of the pandemic, the alternative methodology suggests both faster increases and decreases in total hours than the roll-forward method.

Figure 5: Average weekly hours worked in accommodation and food services was most affected by the alternative imputation methodology

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2019 and April to June 2020
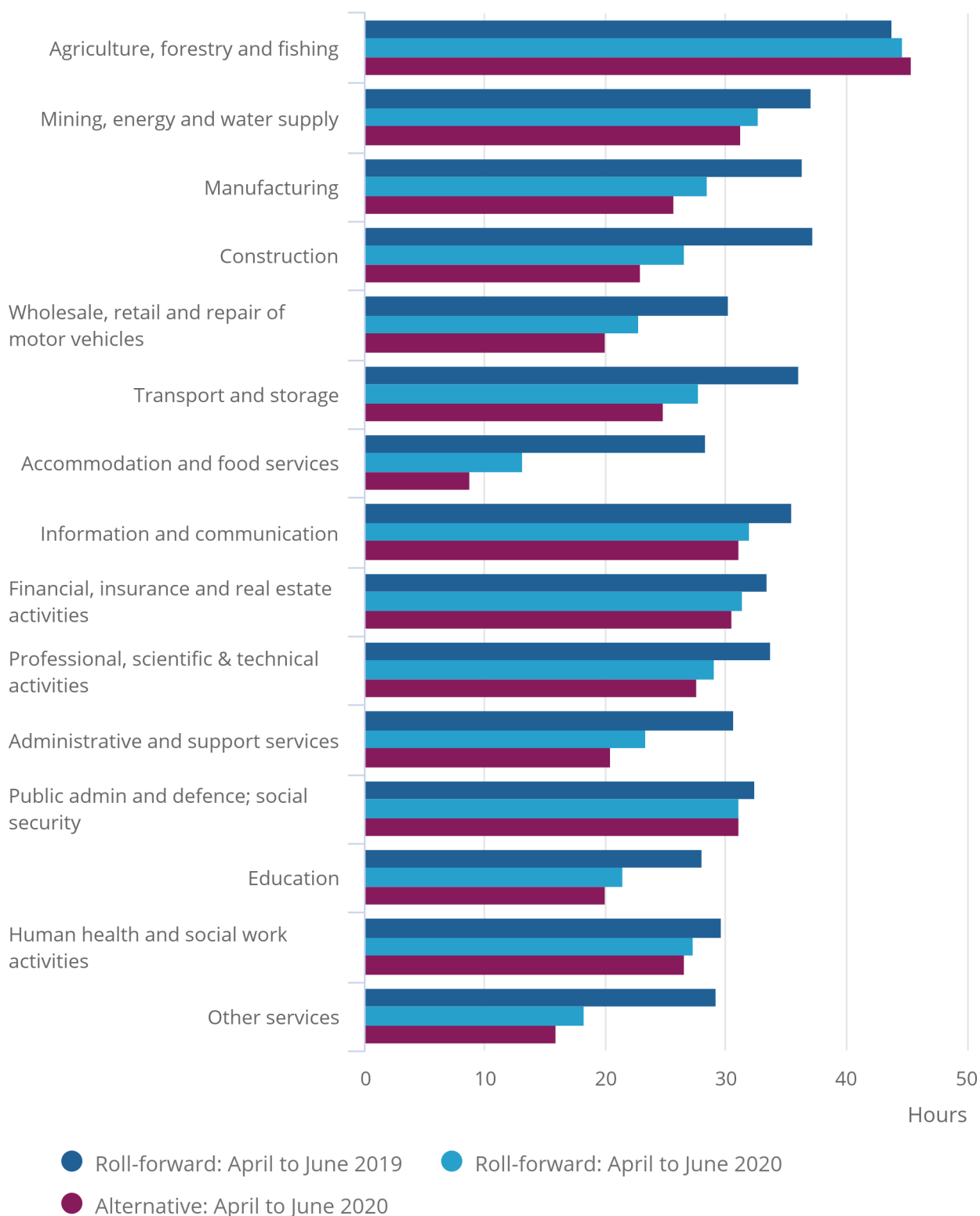
Figure 5: Average weekly hours worked in accommodation and food services was most affected by the alternative imputation methodology

UK average weekly hours by industry, people aged 16 years and over, not seasonally adjusted, roll-forward imputation and alternative imputation, April to June 2019 and April to June 2020

Legend:
- Roll-forward: April to June 2019
- Roll-forward: April to June 2020
- Alternative: April to June 2020

Source: Office for National Statistics – Labour Force Survey

The impact of this alternative imputation was not equally distributed between industries. As Figure 5 shows, the industries that were most impacted by a reduction in hours in the initial stages of the pandemic were also those where the alternative imputation made the biggest difference. For April to June 2020, accommodation and food service activities was the most affected, with the alternative imputation suggesting average hours 34% lower than using the roll-forward imputation method, followed by construction at 14% below. During the pandemic the industry hours estimates from the alternative imputation have been published in extra worksheets in dataset HOUR03.

4 . Future development

This alternative imputation has offered a different picture to help users understand the impact of the coronavirus (COVID-19) pandemic on estimates of headline labour market statuses and hours worked. However, because of the complexities involved in incorporating this new methodology into the existing systems, it is not intended to extend its use on the Labour Force Survey (LFS) person-weighted datasets beyond the impact of the main pandemic-related labour market policies. Further research into this methodology will continue, particularly regarding its potential application in the Labour Market Survey (LMS), currently under development.

5 . Nearest neighbour methodology

The alternative imputation method involves donor imputation. Instead of rolling responses forward from the previous quarter, missing values because of non-response are imputed from another respondent who returned data in the current period, referred to as the "donor". Missing respondents whose responses were to be imputed – referred to as "recipients" – were matched to donors using a nearest neighbour approach. Potential suitable donors were first identified using the following variables:

- current AGE - variable derived separating respondents into groups of under 16 years, 16 to 64 years, 65 to 70 years and 70 years and over; this ensures a child aged under 16 years is not matched with a respondent aged over 16 years

- current SEX - men and women

- current COUNTRY - resident country within the UK

- current GOVTOF - UK government office region of residence

- previous ILODEFR - economic activity

- previous INDS07M - industry section in main job (Standard Industrial Classification: SIC07)

- previous SC2010MMJ - major occupation group in main job (mapped Standard Occupational Classification: SOC10)

- previous STAT - employment status

- previous FTPT - full-time or part-time employment

- previous INECAC05 - detailed economic activity

Secondly, for each potential donor, "distance" measures were calculated for current age, ethnicity, and previous SUMHRS, as follows:

- current AGE: (donor AGE minus recipient AGE) divided by 14

- current ethnicity: if donor ETHUKEUL = recipient ETHUKEUL (if the ethnicities match), then ethnicity distance = 0, else ethnicity distance = 1

- previous SUMHRS: abs(donor previous SUMHRS minus recipient previous SUMHRS) divided by 97

These distance measures using age, ethnicity and hours were then combined to provide an overall combined distance measure. The potential donor with the minimum distance measure, so the one that was most like the recipient, was selected as the donor. Once a suitable donor had been identified, their responses to the following variables were imputed to replace the recipients' missing data:

- BACTHR – total actual hours, main job, excluding overtime

- SUMHRS – total actual hours, main and second job, including overtime

- TOTHRS – total actual hours, main and second job, including overtime

- TTACHR – total actual hours, main job, including overtime

- TTUSHR – total usual hours, main job, including overtime

- YLESS20 – reason for working fewer hours than usual in reference week

- WRKING – whether did paid work in reference week

- YTETJB – whether has paid job in addition to government training scheme

- TYPSCH12 – type of work scheme

- FTPT – full-time or part-time employment

- ILODEFR – economic activity

- INECAC05 – detailed economic activity

- JBAWAY –- whether temporarily away from paid work

- STAT – employment status

These variables were selected as the impact of this alternative imputation method on labour market status and hours worked was the priority at this time. Imputing values for all variables would involve greater complexity that, because of the urgency of the project, was not considered warranted at this time.

6 . Related links

Employment in the UK: October 2021
Bulletin | Released monthly
Estimates of employment, unemployment and economic inactivity for the UK.