

Impact of imputation changes in employment statistics from Pay As You Earn Real Time Information methodology

An assessment of changes to the way missing data is modelled in monthly estimates of employees from Pay As You Earn (PAYE) Real Time Information (RTI) data.

Contact:
ONS Labour Market team,
HMRC RTI Statistics
labour.market@ons.gov.uk
rstatistics.enquiries@hmrc.gov.uk
+44 1633 455400

Release date:
4 March 2024

Next release:
To be announced

Table of contents

1. [Main changes](#)
2. [Overview of imputation methods](#)
3. [Rationale for updating the imputation method](#)
4. [Effects of the imputation changes](#)
5. [Further work](#)
6. [Earnings and employment data](#)
7. [Cite this methodology](#)

1 . Main changes

- Statistics on employment from Pay As You Earn (PAYE) Real Time Information (RTI) are modelled estimates in which future payments can affect historic months, leading to a need to impute future or missing payment submissions.
- During periods of high growth, revisions to estimates suggested that the imputation model was not appropriately taking into account recent trends, leading to bias in early estimates.
- Improvements were made to the imputation model to make probability weights both more accurate and more responsive to recent trends.
- Analysis of estimates since these changes suggest that they have improved the accuracy of imputation.
- Although the changes to the method have improved estimates, we continue to advise caution in use of the flash estimate.

2 . Overview of imputation methods

Converting payments data to jobs data

Each month, HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS) jointly release [estimates of payrolled employees and their pay using HMRC's Pay As You Earn \(PAYE\) Real Time Information \(RTI\) bulletins](#).

PAYE RTI data include a record of payments employers make to their employees. The [European System of Accounts, ESA 2010 \(PDF, 6.40MB\)](#) recommends earnings are recorded in the period work is done rather than the period earnings are paid. To make these estimates consistent with those definitions, multiple methods are employed to convert a payments dataset more formally towards a dataset of jobs and pay rates, including imputation.

Main reasons for imputation

To build PAYE RTI estimates of payrolled employees, payments data are spread back across time to build a dataset of when employees worked and earnings for those periods. We refer to this as the calendarisation methodology. This approach can mean that payments yet to be received in the future can affect employment estimates for the recent months. To provide timely estimates and improve accuracy, those future payments need to be imputed.

Another reason for imputation is because of submissions not always being correctly provided to HMRC on time. HMRC guidance states that employers should provide submissions on or before the date of the payment, but a small amount of submissions can be made late or missing completely. There can also be inaccuracies corrected in later submissions, which can revise earlier estimates.

More information on the calendarisation process and how we make imputations can be found in our [Monthly earnings and employment estimates methodology](#).

3 . Rationale for updating the imputation method

Changes to revisions

As the published figures are estimates and include imputation, revisions to the data are not unexpected. Revisions are monitored and regularly reviewed. The most recent month (or the "flash" estimate) is where we expect to see the largest revisions. However, revisions can change data further back in the time series.

In 2021 and early 2022, estimates showed strong growth in payrolled employees. Users of the publication started to raise concerns about revisions to the estimates. Firstly, revisions were consistently larger in scale throughout this period. Secondly, revisions were consistently downward, while revisions should be random if the imputation process was unbiased.

Further investigation on this showed two separate effects taking place. Firstly, revisions between the flash estimate and the second estimate published a month later were increasing. Secondly, cumulative revisions which aggregated over time were downwards biased depending on the time of year, which was exacerbating the effects seen.

Initial revisions to the flash estimate

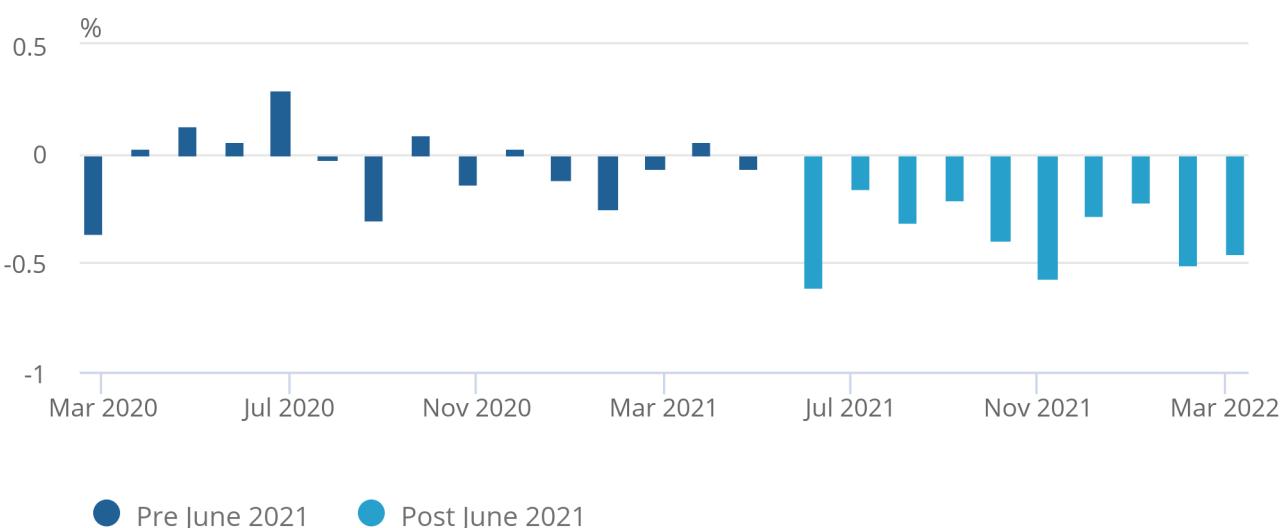
Figure 1 shows initial revisions to employee estimates over two years. The chart is split into two time periods marked by the change in pattern in June 2021. Before this point the revisions to employee estimates varied around the neutral point, which would be expected in a process without bias. However, after June 2021 there was a consistent downward bias and the average size of revisions increased.

Figure 1: Initial revisions for payrolled employee estimates

Payrolled employee revisions, flash vs non-flash month, not seasonally adjusted

Figure 1: Initial revisions for payrolled employee estimates

Payrolled employee revisions, flash vs non-flash month, not seasonally adjusted



Source: Pay As You Earn Real Time Information from HM Revenue and Customs

Cumulative revisions

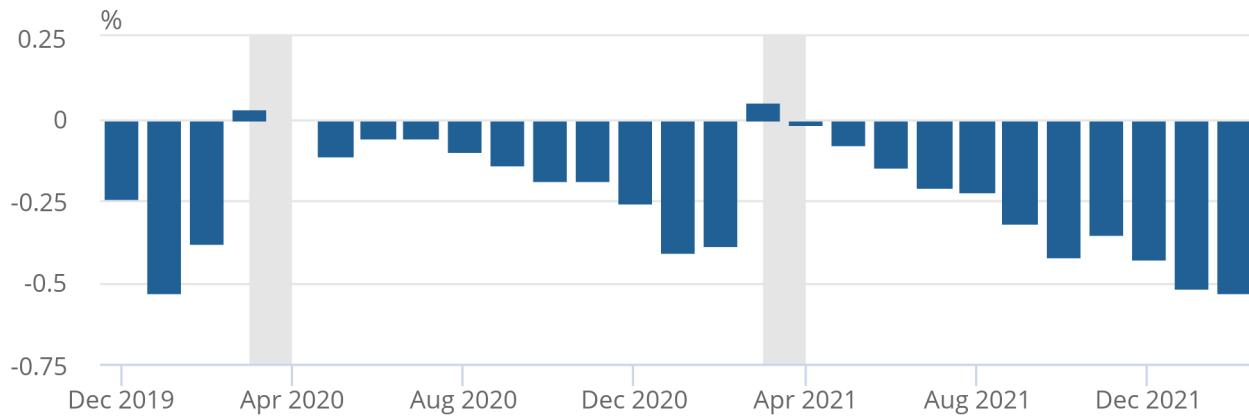
Figure 2 shows cumulative revisions over the course of the publication, with divided sections for tax years. There is a clear downward bias in the revisions for all months other than March and April. The difference grows over the course of the tax year, peaking in January or February, and then repeating.

Figure 2: Cumulative revisions for payrolled employee estimates

Difference between non-flash estimates and latest estimate, not seasonally adjusted

Figure 2: Cumulative revisions for payrolled employee estimates

Difference between non-flash estimates and latest estimate, not seasonally adjusted



Source: Pay As You Earn Real Time Information from HM Revenue and Customs

Improvements to imputation to address revisions

Where employees have historic data, but recent payments are not yet received or missing, their last payment is carried forward. For each of these imputed payments, a probability weight is created based on historical data. This is an estimate of the probability that the job has continued. The update to this imputation process which was implemented in July 2022 is comprised of two elements:

- monthly splits to deal with annual cycles in cumulative revisions
- a scaling factor to deal with recent developments in initial revisions

The monthly split change meant that a separate probability was calculated for each month of the year, and then applied as a weight depending on which month the last real payment was received. This accounts for the differences in revisions patterns being seen through the year, as in Figure 2.

Using a scaling factor made the calculation of these weights more responsive to recent trends. A limitation of the original approach is that, when creating the probabilities for the factors using historical data, sufficient time must be left to identify when a missing payment will be received. This creates a lag between the historical data used to calculate the probabilities, and the period to which they can be applied. If the composition or behaviour of the population changes, the accuracy of the probabilities will reduce.

The scaling factor corrects for this by calculating the probability weights based on historic data, but also equivalent weights are calculated on the latest three months and compared with the weight from historic data. This creates a ratio of how much recent months diverge from historic trends. The weight based on historic data is still the main probability applied but is scaled upwards or downwards based on an average of this divergence ratio from the latest three months of data. This better aligns the probability of still being in employment with current patterns, but still allowing weights to be built on a long-term annual cycle.

4 . Effects of the imputation changes

After the introduction of the imputation update in our July 2022 publication, we have continued to monitor revisions and compare them against the previous trends. Figure 3 shows the scale and patterns of initial revisions in the data since the introduction of the flash estimate in April 2020, split into three time periods:

- March 2020 to May 2021 – this is the early period of our revisions, before the consistent negative bias emerged
- June 2021 to June 2022 – this is the period which was affected by the negative bias, before our implementation of the imputation update
- July 2022 to September 2023 – this is the latest period, following the imputation update

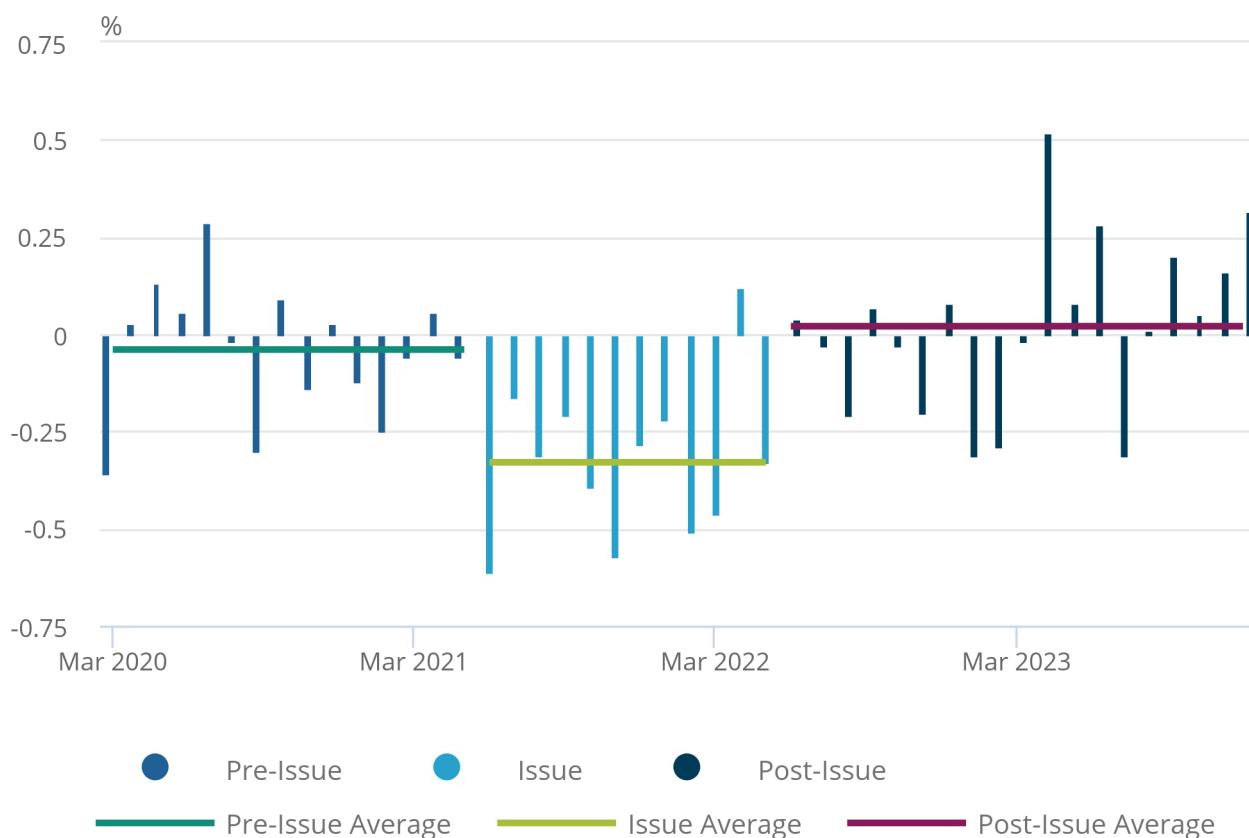
The more recent revisions to our flash estimates show a removal of the consistent negative bias that we had previously observed.

Figure 3: Initial revisions to flash estimates for payrolled employees since implementation of the imputation methodology update

Payrolled employee initial flash estimate revisions, flash estimate versus non flash estimate, not seasonally adjusted

Figure 3: Initial revisions to flash estimates for payrolled employees since implementation of the imputation methodology update

Payrolled employee initial flash estimate revisions, flash estimate versus non flash estimate, not seasonally adjusted



Source: Pay As You Earn Real Time Information from HM Revenue and Customs

Effect on cumulative revisions

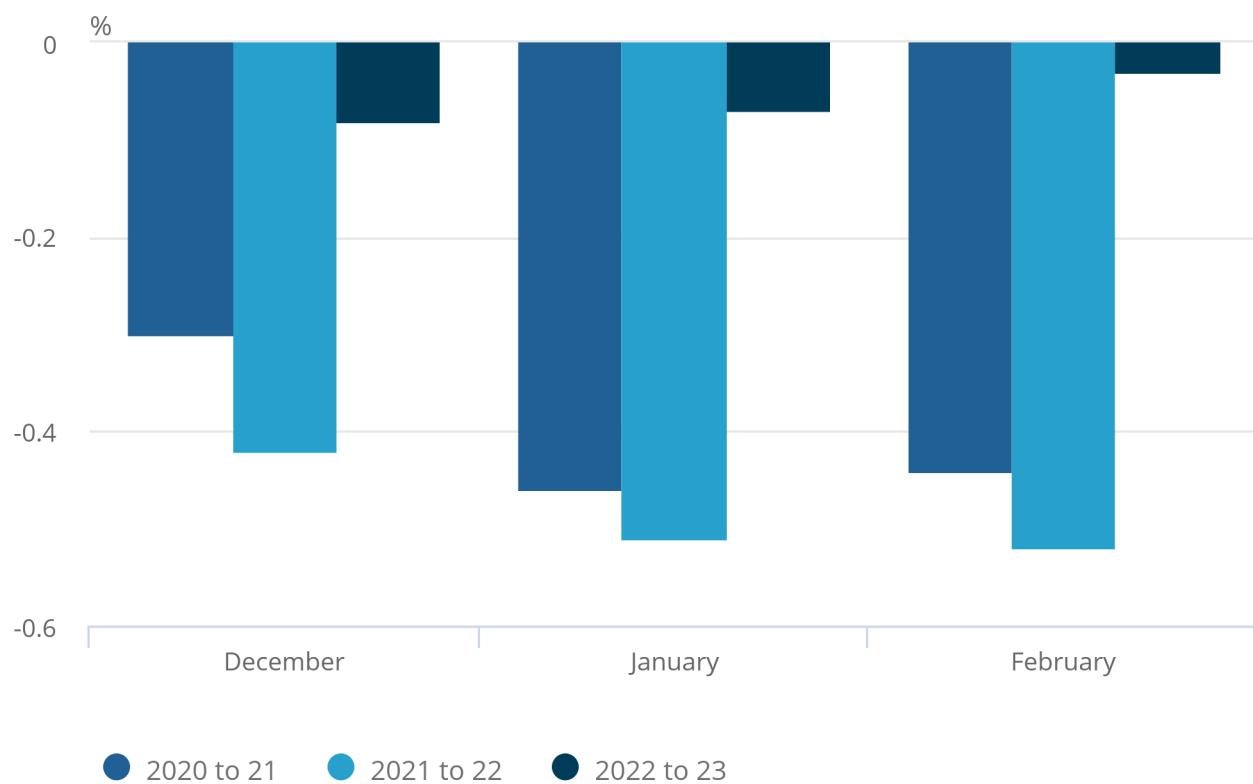
Evaluating cumulative revisions for the new post-imputation update data is slightly more difficult than for the initial revisions, because the most recent months do not have full data yet. However, measuring revisions accumulated until November of the following year, it does appear that the most substantial elements of concern have been addressed. In particular, the large downwards revisions seen post-flash estimate in the December, January, and February months have been substantially reduced in the 2022 to 2023 period (Figure 4).

Figure 4: Cumulative revisions seen in payrolled employee estimates after the second published estimate for three key months of prior concern

Cumulative revisions in payrolled employees after the non-flash estimate, not seasonally adjusted, selected months

Figure 4: Cumulative revisions seen in payrolled employee estimates after the second published estimate for three key months of prior concern

Cumulative revisions in payrolled employees after the non-flash estimate, not seasonally adjusted, selected months



Source: Pay As You Earn Real Time Information from HM Revenue and Customs

5 . Further work

End of year effects

Continued monitoring of revisions and general data patterns after the introduction of the methodological update has revealed some remaining issues. For example, the initial revision for the April 2023 data (the first flash estimate April data since the introduction of the new process) was substantially larger than average. An investigation suggested that imputation patterns for March (which feed into the model for April figures) were less affected by recent trends, therefore adjusting the weighting according to these trends was overcorrecting the data. These issues are being investigated further to ensure there are no hidden factors related to the change in tax year, which are also exacerbating this.

Revisions to pay

Most of this report has been focused on the revisions seen in payrolled employees. However, there are also routine monthly revisions to earnings estimates. Upwards bias seen in pay revisions could suggest a compositional impact of the imputation methodology. The size of the revisions reflects general patterns of greater variability in earnings than in employment, partially because of the skewed nature of earnings towards high-paid individuals.

The differences in processing for earnings imputation mean that detailed analysis is more difficult than for employment. Recent higher than usual pay settlements agreed across industries has increased the size of revisions in the past 12 to 18 months, so this has added some noise to the data. We are investigating whether there is longer term bias that could be addressed through the imputation model. We will update users if further issues with bias or consistency are found, and will inform users through the main bulletin if there are any further planned changes to the imputation model.

6 . Earnings and employment data

[Earnings and employment from Pay As You Earn Real Time Information, non-seasonally adjusted](#)

Dataset | Released 13 February 2024

Earnings and employment statistics from Pay As You Earn (PAYE) Real Time Information (RTI), non-seasonally adjusted. These are official statistics in development.

[Earnings and employment from Pay As You Earn Real Time Information, revision triangle](#)

Dataset | Released 13 February 2024

Revisions of earnings and employment statistics from Pay As You Earn (PAYE) Real Time Information (RTI). These are official statistics in development.

[Earnings and employment from Pay As You Earn Real Time Information, seasonally adjusted](#)

Dataset | Released 13 February 2024

Earnings and employment statistics from Pay As You Earn (PAYE) Real Time Information (RTI), seasonally adjusted. These are official statistics in development.

7 . Cite this methodology

Office for National Statistics (ONS), released 4 March 2024, ONS website, article, [Impact of imputation changes in employment statistics from Pay As You Earn Real Time Information methodology](#)