Article

Improvements to construction statistics: Addressing the bias in early estimates of construction output , June 2018

An analysis of revisions to monthly construction output estimates and detail on the improvements that will be implemented to address the bias in the early survey estimates.

Table of contents

1 . Main points

- Output in the construction industry statistics have been subject to a bias in early estimates, caused by late responses to the survey returns.

- An improved imputation methodology will be implemented, alongside an adjustment system, which together will considerably enhance the revisions performance of construction.

- These new methods will be used when Blue Book 2018-consistent data are used for the first time in the Quarterly national accounts: January to March 2018 publication on 29 June 2018.

- This is the latest in a series of improvements to Office for National Statistics's construction statistics.

2 . Overview of construction statistics

Office for National Statistics (ONS) is responsible for publishing the following three main datasets on the construction industry:

- [Output in the construction industry](#)

- [New orders in the construction industry](#)

- [Construction Output Price Indices (OPIs)](#)

In December 2014, the Department for Business, Innovation and Skills (BIS) announced the suspension of its publication of Construction Price and Cost Indices (CPCIs). This led to the suspension of the [National Statistics](#) status for construction output, new orders and prices. Responsibility for the OPIs subsequently transferred to ONS in April 2015.

A range of improvements have been implemented to all three datasets since this point and the Office for Statistics Regulation is currently re-assessing the extent to which they now meet the professional standards set out in the statutory [Code of Practice for Statistics](#).

An [impact of improvements to construction statistics](#) article was published in September 2017, detailing improvements that were incorporated to construction output statistics for UK National Accounts, The Blue Book 2017, including nominal data adjustments, a seasonal adjustment review and the output price indices.

The updated output price indices replaced the interim method, which was first published in June 2015, following the incorporation of a mark-up for profit margin, a revised methodology for the labour series, new weights and data sources, and a full review of the methodology used.

Following the inclusion of [Value Added Tax (VAT)](#) as an additional data source for construction output statistics in December 2017, the construction statistics development programme has been focused on two main priority areas:

- improving revisions performance and reducing bias in early estimates of nominal output data

- improving the model used to estimate regional and lower sector level estimates using an enhanced new orders dataset

Output in the construction industry has often been subject to revisions, which can occur for many reasons. It has been identified that the current imputation methodology has been causing a bias to early survey estimates. This article details the existing methodology and explains how it has contributed to previous revisions. It then sets out the improvements that will be made to address this bias in early survey estimates, with a new methodology and an adjustment system to account for the bias that remains.

A separate article has been published today (4 June 2018), detailing the new model for calculating regional and sub-sector estimates of construction output.

3 . Revisions to construction output

Output in the construction industry measures the volume and value of construction work by businesses in the construction industry in Great Britain, using the Monthly Business Survey (MBS) as the primary data source. The Quality and Methodology Information report contains information about how the output is created from this data.

Revisions are a natural occurrence to statistical publications and can occur when changes are made to methodology, or from more data becoming available and being used to recalculate the estimates.

For construction output statistics, "more data becoming available" consists of many different types of revision. For instance:

- late responses to surveys, with actual data replacing imputations

- changes to original returns, for example, revising the figures reported

- HM Revenue and Customs (HMRC) Value Added Tax (VAT) returns supplementing MBS data for small- and medium-sized businesses when VAT estimates become available each quarter

- revisions to seasonal adjustment factors, which are re-estimated every month and reviewed annually; as the seasonal adjustment series used in construction covers a shorter period than many other Office for National Statistics (ONS) economic outputs, the revisions from seasonal adjustment changes have been bigger in construction output than other similar releases

- revisions to the input series for the Output Price Indices

The most consistent cause of revisions, of these types, is the receipt of late responses to the survey. Late response will cause an impact on revisions, where the response differs from the value that has been imputed for that business. There are three imputation methods that are used and these are explained in Table 1.

Table 1: Explanation of imputation methods

Method	Constructed imputation	Forwards imputation	Backwards imputation
Purpose	Estimate for the non-response of a business, for the first period it is on the sample. The survey's sample rotation leads to approximately 600 new businesses joining the sample each month, maintaining the total sample size of 8,000.	Estimate for the non-response of a business, which has been on the sample for more than one period.	Providing an updated estimate for non-response, in cases where a business' first response to the survey is not for its first period on the sample. It is used for all periods that are prior to the period of its first response.
Level	Estimated at strata level (distinct UK SIC 2007 industry and size of business), separately for all 11 questions on the survey.	Estimated at UK SIC 2007 industry level only, separately for all 11 questions on the survey.	Estimated UK SIC 2007 at industry level only, separately for all 11 questions on the survey.
Contributing information	The return (R_p) and annual turnover (T), according to the Inter-Departmental Business Register (IDBR), of businesses at the same level, who have returned for the period.	The return for both the estimation period (R_p) and the previous period (R_{p-1}), for businesses who have returned for both periods.	The return for both the estimation period (R_p) and the subsequent period (R_{p+1}), for businesses who have returned for both periods.
Link Calculation (Ratio of Means)	$(R_p) / (T)$, where the sums include all contributing businesses at the chosen level	$(R_p) / (R_{p-1})$, where the sums include all contributing businesses at the chosen level	$(R_p) / (R_{p+1})$, where the sums include all contributing businesses at the chosen level
Trimming Factor	Individual business ratios are calculated, and ranked: (R_p / T). Businesses amongst the largest 10% of ratios are trimmed, before the link is calculated.	None	None
Link Application	The link is applied to the annual turnover of the non-responding business.	The link is applied to the value held for the non-responding business, for the previous period. This value could be a return, construction or imputation.	The link is applied to the value held for the non-responding business, for the subsequent period. This value could be a return, or imputation.
Average amount of sampled businesses for first iteration (2016 reference periods)	3.6%	31.4%	0.0%
Average amount of sampled businesses for fifth iteration (2016 reference periods)	1.8%	19.9%	1.8%

Average amount of sampled businesses for thirteenth iteration (2016 reference periods)	1.5%	18.7%	2.5%
Factors that lead to revisions after first iteration	Replaced by late return. Replaced by backwards imputation. Recalculation of construction link, following the inclusion of additional responders.	Replaced by late return. Replaced by backwards imputation. Recalculation of imputation link, following the inclusion of additional responders.	None, as there are no backwards imputations in the first iteration.

Only a return or constructed imputation value can determine the amount of work conducted by a business; while forward and backward imputations can then be applied to estimate how that value may have changed – using information calculated from the relative performance of businesses who have responded.

Revisions are documented in the revisions triangles datasets, which are available for one-month growth and three-month growth. These highlight the history of revisions to headline construction output statistics, at the seasonally adjusted and volume level. However, the impact of changes to survey data on revisions can be isolated by examining revisions to the unconstrained survey totals for all construction work. Figure 1 compares three iterations of this data for the 2016 reference period and highlights the upward revisions that have been made to these values.

Although only data for 2016 are displayed, similar revisions are found throughout the time series, dating back to the previous imputation methodology change during 2011. The first iteration represents the survey totals at the time of first iteration for each reference period and similarly for the fifth and thirteenth iterations.

Figure 1: Difference between the first, fifth and thirteenth iterations of unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016

Figure 1: Difference between the first, fifth and thirteenth iterations of unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016
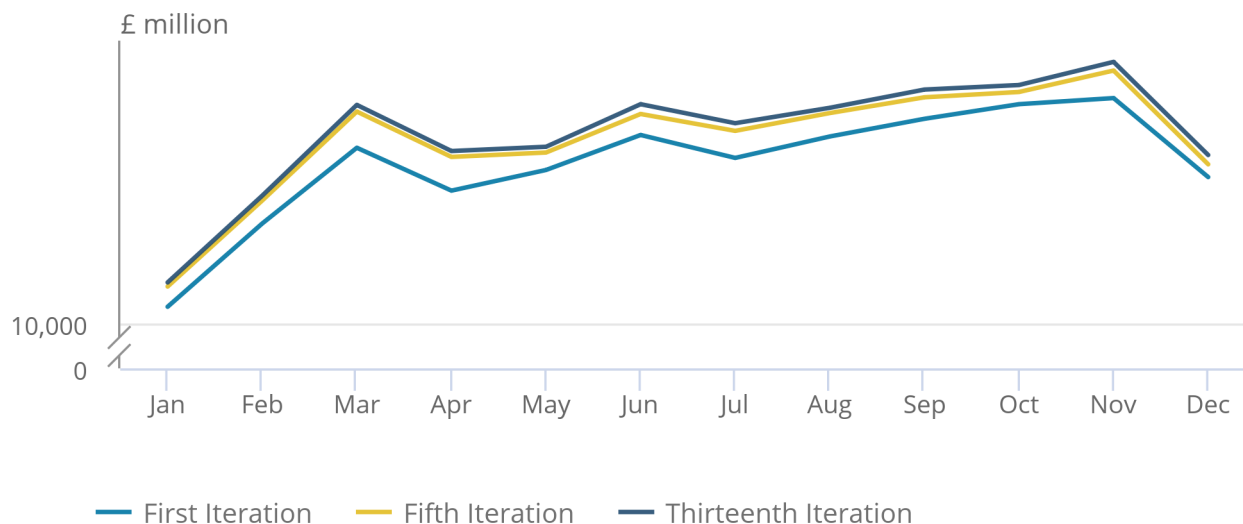
Source: Office for National Statistics

Notes:

1. These data differ from the current price non-seasonally adjusted data published in Table 4 of Output in the construction industry, as these data are constrained, in accordance with the standard revisions policy for national accounts.

2. All data included in this article is weighted data, as explained in the Construction Output QMI.

The fifth iteration has been chosen, as the majority of revisions to survey data have taken place by this point in time. There was an average monthly revision of £324 million to the unconstrained survey totals, between the first and fifth iteration in 2016 – which is an average increase of 2.6%. This compares with an average monthly revision of £421 million, between the first and thirteenth iteration.

Additionally, the fifth iteration is now the final iteration in which survey data will be the sole data source for all months, following the implementation of VAT turnover data into national accounts. Annex A provides an illustrative example of when current price data sources are combined for monthly construction output. Subsequent iterations for future periods will therefore feature revisions where survey data is replaced by VAT data and the pattern of these revisions will be regularly monitored.

The thirteenth iteration marks one year after the initial iteration and is the final iteration in which revisions to the survey data are processed. Survey data will only be re-processed after 13 months if there is a need to retrospectively apply a methodological change, such as with the size-band adjustments in September 2017.

Of the businesses that are imputed for in the first iteration, the 2016 data show that 95% do see a change in value by the thirteenth iteration, in accordance with the "Factors that lead to revisions after first iteration" section. In 55% of cases, these were positive changes, while 40% had a negative change.

Table 2 displays the average absolute sum of revisions to unconstrained survey totals, separated by whether the value of revision was positive or negative. Across 2016, the average absolute sum of positive revisions was £820 million, while the corresponding figure for negative revisions was £391 million. The sum of positive revisions was always greater than the sum of negative revisions for every month in 2016. There was therefore an average monthly revision of £430 million in 2016, following the revision of imputed values.

This consistent upward revision indicates that there is a bias in the early survey estimates. To identify where this revision is largest, the data can be separated by revisions from constructed imputation values; forward imputation values, which are ultimately sourced from a constructed value; and forward imputed values, which are ultimately sourced from a returned value.

Table 2: Average absolute sum of revisions from imputation methods, from first iteration to thirteenth iteration, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain 2016, £ million

Imputation method	Average absolute sum of revision		Net average absolute sum of revision
	Positive Change	**Negative Change**	
Constructed imputation	159	14	145
Forward imputation from constructed value	241	19	223
Forward imputation from returned value	420	358	62
Total	820	391	430

Source: Office for National Statistics

Table 2 shows that revisions from constructed imputation values on average contribute £145 million to the total revision; while values that have been forward imputed from a constructed imputation value in a previous month are contributing an average of £223 million to the total revision. In each case, the current constructed imputation method is the cause of this initial under-estimate, with both showing notably more positive revisions than negative.

It is therefore evident that revisions from constructed imputation are the main cause of the bias in early survey estimates.

Figure 2 demonstrates the timing of these revisions, focusing only on revisions directly from constructed imputation values. For the businesses who had a constructed imputation value in the first iteration, the total sum has been calculated, and Figure 2 displays the revisions that occurred in subsequent iterations (at which point, many of the constructions will have been replaced by either a return or backwards imputation).

Figure 2: Revisions from first iteration, to total value of businesses who had constructed imputation values for the first iteration, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016

Figure 2: Revisions from first iteration, to total value of businesses who had constructed imputation values for the first iteration, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016
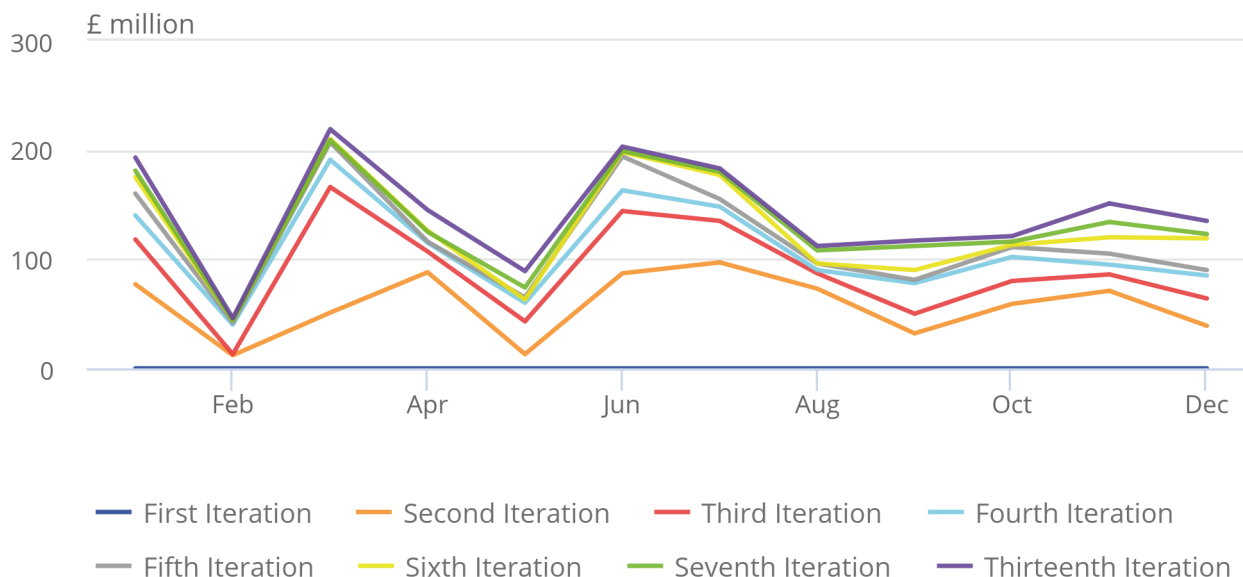
Source: Office for National Statistics

Figure 2 highlights that the revisions occur in stages, with the largest revision occurring between the first and second iterations, an average of £58 million across 2016. The next largest occurs between the second and third iterations, and the between-iteration revision continues to decrease over time, up to the thirteenth iteration. This explains why there will often be upward revision to month-on-month growth rates.

Using December as an illustrative example, the first month-on-month growth rate for a period of December will be a comparison between the first iteration of December data and the second iteration of November data. The second month-on-month growth rate will then be a comparison between the second iteration of December data and the third iteration of November data. On average, the difference between versions one and two of December is larger than the difference between versions two and three of November – therefore it is to be expected that the month-on-month growth rate for December will be revised upwards between its first and second iteration.

A similar pattern to Figure 2 is found, when analysing the same data for businesses that were forward imputed for at the first iteration.

This evidence has highlighted that there are a larger amount of positive revisions from imputed values than negative revisions, which is resulting in a bias to early survey estimates. In particular, there are a larger amount of positive revisions to values calculated using the constructed imputation method. This therefore highlighted a need to review the imputation methodology.

4 . Reviewing the imputation methodology

To identify whether the bias in early estimates can be reduced, alternative imputation methodologies have been investigated for all three of the approaches detailed in Table 1.

The imputation methodology already uses the ratio of means approach, which is consistent with other short-term indicators (retail sales, Index of Services and Index of Production) and is recognised in the Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys EDIMBUS manual (PDF, 799KB) (see C.4.2) as international best practice for imputation when the contributor has a valid value in the previous period.

Alternatives for the forward imputation method therefore looked at the level at which it is performed, such as incorporating a distinction by size of business, or targeting the total construction level and apportioning down to the lower question level. None of the approaches considered were found to cause a significant improvement to the revisions that occurred.

For the constructed imputation methodology, the current inclusion of trimming the largest 10% of ratios was identified as a main concern for why under-estimation has occurred under the current methodology. As the trimming was not balanced, with no trimming of the smallest 10%, a reduction in this one-sided trimming can only result in larger construction links and therefore high constructed imputation values.

Additionally, while the use of trimming is appropriate for mean-of-ratio imputation, it is not necessary for the ratio-of-means approach, which is used here.

Through the analysis of historical data, it was possible to calculate constructed imputation values for alternative methodologies and compare these values to the returns that were received as late responses. This allowed to identify the methodology that would have minimised the revisions to total construction output and this identified that the best approach would be to not include any trimming.

This analysis also provided evidence to support a change to the level of construction imputation, from strata-based to the UK Standard Industrial Classification: SIC 2007 industry level only, in line with the other imputation methods.

The new methodology for constructed values will not use any trimming and will be calculated at the industry level.

This methodology change has been approved by ONS's Methodological Assurance for Statistical Transformation (MAST) group, who agreed that the existing methodology was not fit for the purpose and that the chosen methodology is the best available new method.

5 . Impact of improved imputation methodology

This new methodology has been processed through a test system, to calculate an indication of what the revisions performance would have been for the unconstrained survey totals for all construction work, had this method been used instead. This has produced an indication of the reduction in revisions that would have occurred, resulting from updated values for both constructed imputation values and construction-based forward imputations.

Table 3 provides an updated version of Table 2, with the revision associated with constructed values now being both smaller and more balanced.

Table 3: Average absolute sum of revisions from imputation methods, from first iteration to thirteenth iteration, following implementation of improved methodology, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain 2016, £ million

| Imputation method	Average absolute sum of revision		Net average absolute sum of revision
	Positive Change	**Negative Change**	
Constructed imputation	42	33	9
Forward imputation from constructed value	133	121	12
Forward imputation from returned value	420	358	62
Total	595	512	83

Source: Office for National Statistics

The net average absolute sum of revision for both methodologies is compared in Figure 3, highlighting that the total amount of revision has reduced from £430 million to £83 million, a reduction of approximately 80%.

Figure 3: Net average absolute sum of revisions from imputation estimation methods, from first iteration to thirteenth iteration, for current and improved methodology, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016

Figure 3: Net average absolute sum of revisions from imputation estimation methods, from first iteration to thirteenth iteration, for current and improved methodology, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016
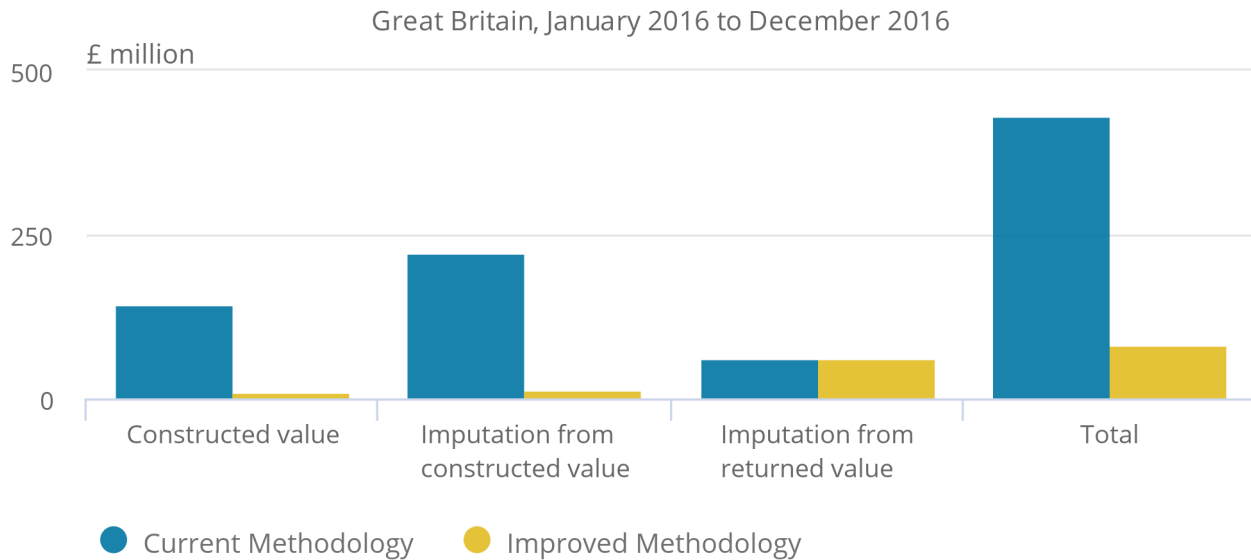
Source: Office for National Statistics

6 . Remaining bias under improved imputation methodology

The new methodology for constructed imputations has accounted for the majority of the existing bias in early estimates of survey data, but the remaining bias is still statistically significant for revisions at the current price, non-seasonally adjusted level.

Figure 4 documents the revisions that occur from imputation methods, between the first iteration and the next four subsequent iterations – for both the previous and improved methodologies. As Table 3 shows us, the revisions for the improved methodology are now caused primarily by imputations that are ultimately sourced from a returned value, under the improved methodology.

Figure 4: Average revisions from imputation, from first iterations to subsequent iterations, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016

Figure 4: Average revisions from imputation, from first iterations to subsequent iterations, unconstrained, all construction work, non-seasonally adjusted, current price

Great Britain, January 2016 to December 2016
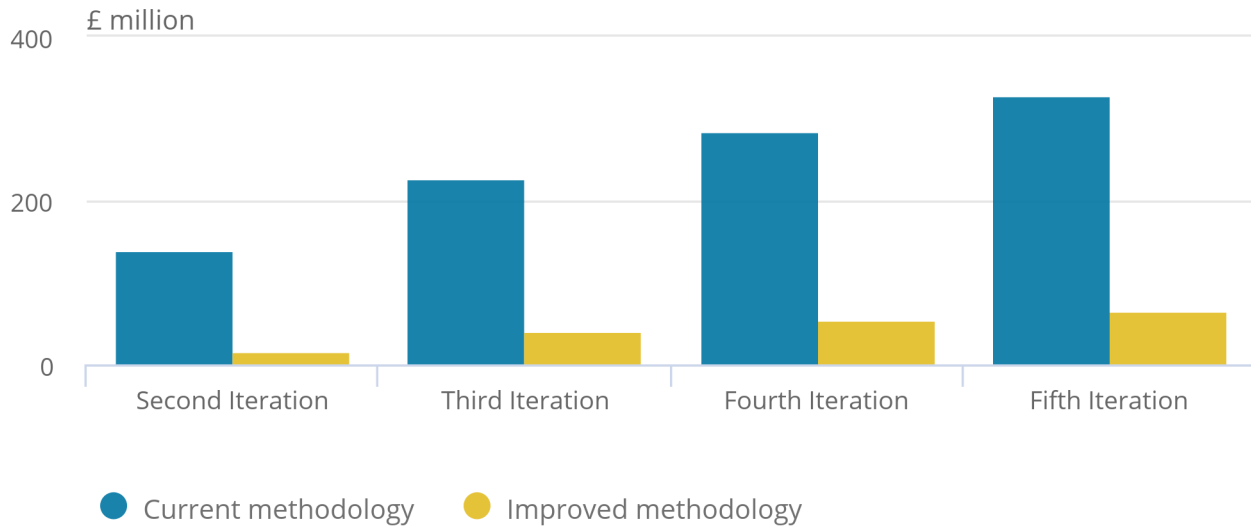
Source: Office for National Statistics

Under the previous methodology, all months of 2016 were upwardly revised between the first and fifth iterations. Now under the improved methodology, 3 of the 12 months have received a downward revision. The average revision by the fifth iteration remains positive, but is considerably reduced, from £328 million to £66 million.

The fact that the remaining revisions are not constant also means that revisions to month-on-month growth rates can still be expected in future. This is illustrated after the implementation of the improved methodology as, whilst the early bias in the survey estimates has reduced, there is a non-constant sign of revision where some periods are revised upwards and some downwards.

7 . Adjustment for remaining bias

While the results in Sections 5 and 6 display a significant improvement following the new constructed imputation methodology, they do also show that a positive average revision remains in the early estimates of the survey data. We therefore will be incorporating an additional adjustment system, to account for the remaining bias in the early estimates of survey data.

The use of quality adjustments is common across national accounts and short-term output indicators to address for various conceptual and data quality issues. For example, the Index of Services apply these quality adjustments (PDF, 128KB). It is also not a new concept for construction output estimates to have a quality adjustment applied. An adjustment is used for the preliminary gross domestic product (GDP) estimate to address for the lower data content and the bias introduced from this earlier response. This is stated in Section 6 of the GDP preliminary publication.

The new quality adjustment facility will allow a decaying adjustment to be applied to the data in the early estimates of monthly construction output to account for the remaining bias. This adjustment will be applied at the aggregate level and seeks to address the remaining problems caused by late responders differing from early responders.

The objective of this quality adjustment facility will target a position of the survey data at its fifth iteration, which is its final iteration before Value Added Tax (VAT) turnover can be used as a data source in the estimates of monthly construction output. This is explained in Section 3 and within an illustrative quarter in Annex A. This differs from the current quality adjustment that is currently applied for the preliminary GDP estimate, as this has the target position of the first iteration of construction output.

Using historic data, analysis has been carried out to assess the most appropriate quality adjustment to apply. The analysis highlighted that there is a relationship between the number of imputations and the size of revision. We will apply a multiplicative quality adjustment, to account for the remaining bias from imputation methods in the early survey estimates. This will use historical data and will consider:

- the mean average adjustment for each of the first four iterations, against the targeted fifth iteration position

- the average rates of imputation at each stage

The quality adjustment facility will be regularly reviewed, in collaboration with colleagues from Methodological Assurance for Statistical Transformation (MAST), and updated with the latest data to ensure its suitability within the publication. This will include consideration of whether there is a need to extend the adjustment beyond the fifth iteration, where VAT data becomes an additional contributing factor to revisions.

Also, the new quality adjustment will factor in the change to the new GDP publication model. As a result of this new publication model, the first construction figure used in GDP estimates will have a higher data content than the current figures used in the preliminary estimate of GDP. However, there will be a reduced response for the third month in the quarter for monthly construction estimates, due to the earlier finalisation date (See Figure 3 in the GDP publication model hyperlink). Therefore, in the future, the third month in the quarter will receive a larger initial quality adjustment for this reduced response, when compared with the first quality adjustment for other months.

A review of both the imputation methodology along with the use of the quality adjustment facility will be undertaken when the survey is fully transformed as part of transformation of economic statistics. This will be towards the end of 2019.

8 . Impact analysis on construction output

To analyse the impact of the change to the imputation methodology and the incorporation of the quality adjustment for the remaining bias, the previously published datasets have been recalculated. This has provided us with indicative results for what construction output would have looked like.

As stated in Section 3, the current bias saw an average monthly increase in value of 2.6% to the unconstrained survey totals for 2016, between the first and fifth iteration. The change to the imputation methodology reduces the average increase to 0.5%, with a further reduction of the bias in revisions to 0.06% upon application of the quality adjustment.

Figure 5 highlights the impact this has had on the 2016 monthly values. It portrays the revision to the level of unconstrained surveys totals, by comparing the growth in levels from the first to fifth iteration, for the improved methodology against the current.

Figure 5: Percentage growth to unconstrained, all construction work, current price, non-seasonally adjusted totals, from first iteration to fifth iteration

Great Britain, January 2016 to December 2016

Figure 5: Percentage growth to unconstrained, all construction work, current price, non-seasonally adjusted totals, from first iteration to fifth iteration

Great Britain, January 2016 to December 2016
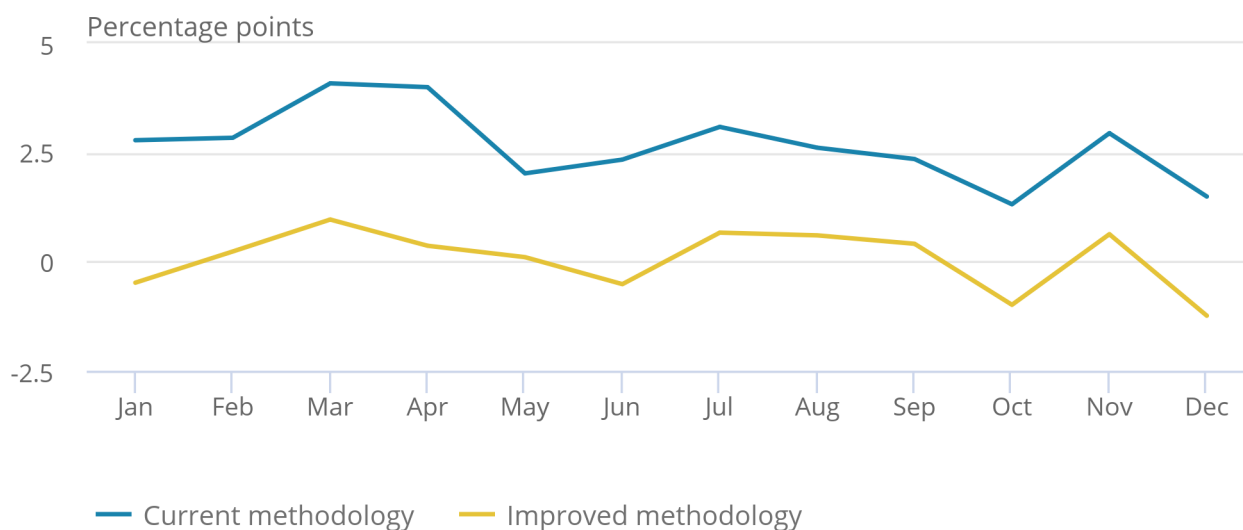
Source: Office for National Statistics

By the fifth iteration, the extent to which the monthly level had increased varied from 1.3% to 4.1% under the existing methodology, a positive bias that is not seen under the new methodology, with growths varying from -1.2% to 1.0%.

At the headline, seasonally adjusted chained volume measure level, a point of comparison can be made at the publication for the October 2016 reference period, where construction output was open to revisions back to January 2015.
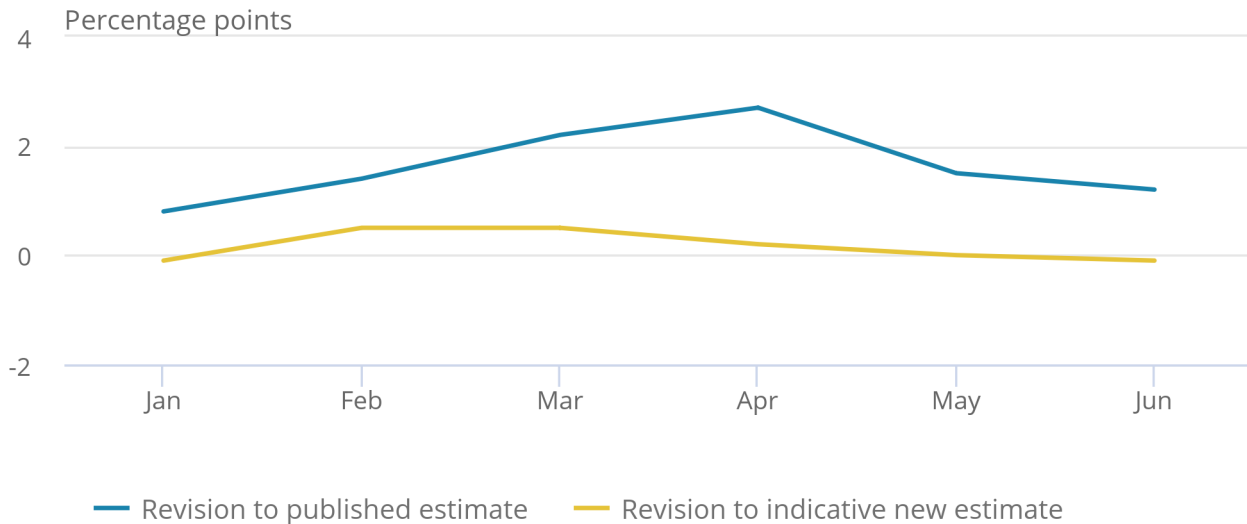
For the months of January 2016 to June 2016, Figure 6 displays the revision to three-month on three-month growth rate from first publication to the October 2016 publication period, for both the published estimate and indicative new estimate. Figure 7 displays the revision to month-on-month growth rate, for the same timeframes.

Figure 6: Revision to three-month on three-month growth rate, from first publication to October 2016 publication, all construction work, chained volume measure, seasonally adjusted

Great Britain, January 2016 to June 2016

Figure 6: Revision to three-month on three-month growth rate, from first publication to October 2016 publication, all construction work, chained volume measure, seasonally adjusted

Great Britain, January 2016 to June 2016

Percentage points

Legend:
— Revision to published estimate
— Revision to indicative new estimate

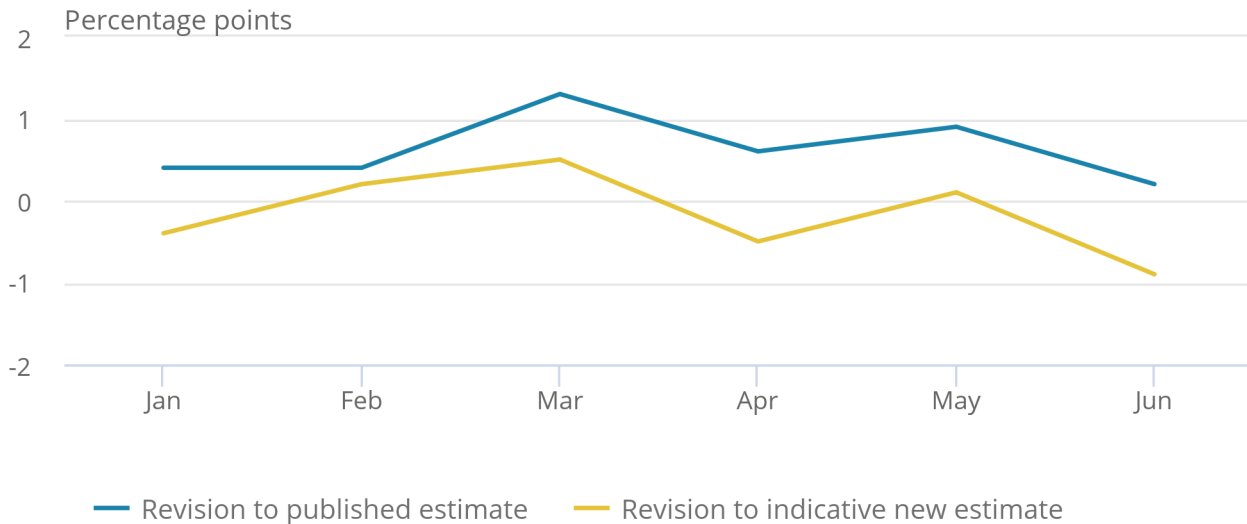**Source: Office for National Statistics**

Figure 6 shows a clear reduction to the percentage point revision to three-month on three-month growth rate with the indicative new estimates, when compared with published estimates, with a maximum revision of 0.5 percentage points – compared with a minimum revision of 0.8 percentage points.

Figure 7: Revision to month-on-month growth rate, from first publication to October 2016 publication, all construction work, chained volume measure, seasonally adjusted

Great Britain, January 2016 to June 2016

Figure 7: Revision to month-on-month growth rate, from first publication to October 2016 publication, all construction work, chained volume measure, seasonally adjusted

Great Britain, January 2016 to June 2016

Source: Office for National Statistics

Whilst the new imputation methodology and implementation of a quality adjustment facility have removed a statistically significant bias in the early iterations of the survey returns, revisions are still prevalent. However, these revisions are consistently smaller for the indicative new estimates in the month-on-month growth rate as shown in Figure 7. This also shows a combination of both positive and negative revisions, rather than all revisions being positive.

9 . Implementation

The new methodology and quality adjustment system will be adopted for the first time in the Quarterly national accounts: January to March 2018 release on 29 June 2018, which is consistent with the 2018 Blue Book and the subsequent Construction output in Great Britain: May 2018 publication on 10 July 2018.

Within the processing system, the new imputation methodology will be fully implemented, to ensure that all imputations are updated to reflect the new methodology, but the published data will only be affected from 2017 onwards. The updated series will be constrained to the previous series by growth, to prevent any step-changes in the data.

The target of the improvements is to remove the bias in revisions for future reference periods. The main impact is likely to be seen in the most recent five monthly estimates. There will be little impact for older periods, as they already include almost full Monthly Business Survey and Value Added Tax content.

10 . Annex A: Combining current price data sources for monthly construction output

An illustrative example as to when current price data sources are combined for monthly construction output is shown in Table 4. This is the latest quarter within the quarterly national accounts to currently incorporate Value Added Tax (VAT) turnover data and illustrates how the individual months change data sources as they go through iterations of the data. The fifth iteration of a release month is the last position when survey data are the sole indicator for all reference months within a quarter. This was the position as at the Quarterly national accounts: October to December 2017 published on 29 March 2018.

Table 4: Source of current price data for monthly construction output by month of publication and release month for Quarter 3 (July to September) 2017

Release Month	July 2017 reference month	August 2017 reference month	September 2017 reference
First Iteration	Survey only	Survey only	Survey only
Second Iteration	Survey only	Survey only	Survey only
Third Iteration	Survey only	Survey only	Survey only
Fourth Iteration	Survey only	Survey only	Survey only
Fifth Iteration	Survey only	Survey only	Survey only
Sixth Iteration	Survey only	Survey only	Survey and VAT
Seventh Iteration	Survey only	Survey and VAT	Survey and VAT
Eighth Iteration	Survey and VAT	Survey and VAT	Survey and VAT

Source: Office for National Statistics