Office for National Statistics

Article

Construction development: improvements to regional and sub-sector level estimates, June 2018

Improvements to the model used to estimate regional and sub-sector level construction output data, including the impact of the changes.

Table of contents

1 . Main points

- Office for National Statistics (ONS) has developed a new model to more accurately apportion construction output by region and sub-sector, using new orders data from both ONS and Barbour ABI.

- The new model adopts a two-staged approach; spreading the value of new orders worth £50 million and above using project-specific start dates and durations, while spreading lower-value aggregated totals for projects worth between £100,000 and £50 million using weighted average durations.

- The new model will first be implemented in the Construction output in Great Britain: April 2018 and new orders January to March 2018 publication, released on 11 June 2018.

- No revisions will be made to top-level construction output, however, revisions to lower-level sub-sector and regional data estimates, displayed in Tables 5 and 6 of Construction output in Great Britain, will be evident from 2010 onwards.

2 . Introduction

We produce monthly statistics on the value of construction work done in Great Britain. Data by sub-sector and regional splits are published in the first month of each quarter, estimated using new orders data. The modelling of this lower-level data has been necessary because it is not possible to accurately obtain the granular breakdown of work required from either the Monthly Business Survey (MBS) data or Value Added Tax (VAT) turnover data.

As part of the ongoing Office for National Statistics (ONS) construction statistics development programme, we have worked closely with the Construction Statistics Steering Group. This group provides a forum for ONS to engage with main users of construction statistics on the development of ONS-published construction statistics, including other government departments, industry experts and academics, to identify areas for improvement. Following extensive engagement with this group, we have developed a new model to estimate regional and sub-sector construction output.

In addition to this article, which focuses on the model used to create estimates of regional and sub-sector output, further improvements are being made to reduce the level of revisions to ONS construction output figures, detailed in the article, Improvements to construction statistics: addressing the bias in early estimates of construction output, June 2018.
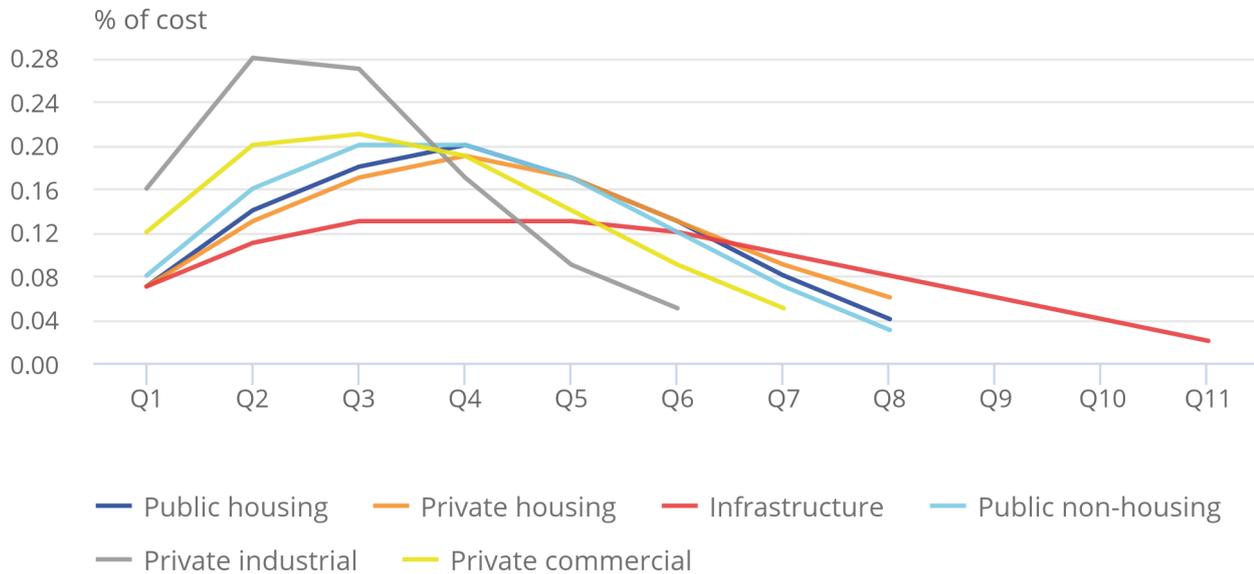
3 . The current model

The value of construction output by region and sub-sector is currently estimated using new orders data supplied by Barbour ABI. The new orders data include information on new contracts awarded by type of work and by region where the work is to be carried out.

New orders data are used to calculate proportions for sub-sector and region, which are used to apportion high-level sector output estimates from the monthly construction survey used to calculate the total value of construction work in the industry.

To create these proportions for region and sub-sector, the value of the new orders project is spread across a set duration from the quarter in which the new orders contract was awarded. The duration of the project is taken from the Quarterly Inquiry of Projects in Progress (PROBE) Survey. The PROBE Survey, which was carried out by the Department for Business, Innovation and Skills (now known as the Department for Business, Energy and Industrial Strategy) – and which is now discontinued following data collection ceasing in 2010 – was used to track the output of projects through their life cycles.

Figure 1: Periodic weights and durations by sector

Figure 1: Periodic weights and durations by sector

% of cost

Public housing — Private housing — Infrastructure — Public non-housing
— Private industrial — Private commercial

Source: Quarterly Inquiry of Projects in Progress survey

Figure 1 shows the default average durations and cost weights used in the current model. All projects within the same sector assume the same length. Projects for each sub-sector are modelled using the higher-level sector duration and cost distribution. For example, water projects are modelled using the length and distribution of the higher-level sector of infrastructure. The values for each sub-sector and region for each quarter created using the periodic weights and durations displayed in Figure 1 are then used to apportion construction output.

Limitations of the current model

We have identified three major limitations of the current model:

- all construction projects within the same sector are assumed to have the same duration

- all construction projects are assumed to start in the same quarter they are ordered

- all construction new orders contracts awarded are assumed to then result in construction work taking place
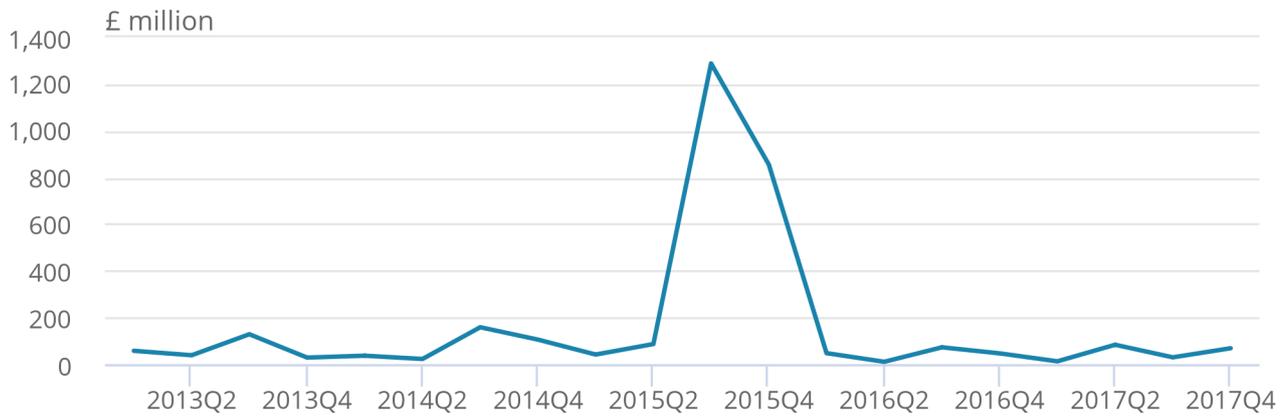
The limitations of the current model were particularly evident in the infrastructure sector, especially the sewerage sub-sector. Figure 2 shows total sewerage new orders between 2013 and 2017. An exceptionally high level of sewerage new orders were placed in Quarter 3 (July to Sept) 2015 and Quarter 4 (Oct to Dec) 2015.

Figure 2: Total new orders for sewerage, current price, non-seasonally adjusted

Great Britain, 2013 to 2017

Figure 2: Total new orders for sewerage, current price, non-seasonally adjusted

Great Britain, 2013 to 2017

£ million

[Line chart showing total new orders for sewerage from 2013Q2 to 2017Q4. Y-axis ranges from 0 to 1,400 in increments of 200. Values remain low (around 50-150) for most periods, with a sharp peak reaching approximately 1,300 around 2015Q3, dropping back to low levels by 2016.]

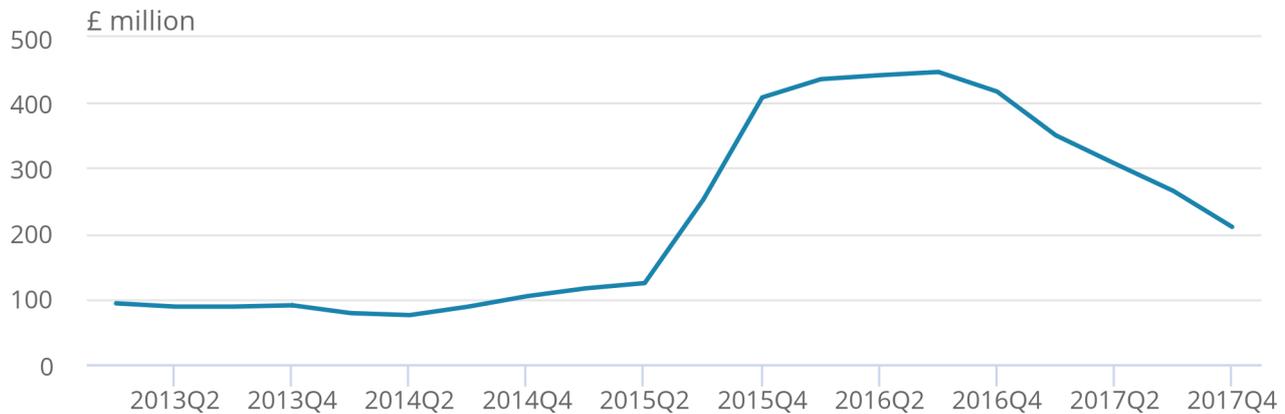**Source: Office for National Statistics, New Orders in the Construction Industry**

These new orders were subsequently fed into the model immediately, leading to the visible increase in sewerage output, particularly from Quarter 4 2015 to Quarter 1 (Jan to Mar) 2017, as shown in Figure 3.

Figure 3: Construction output for sewerage, current price, non-seasonally adjusted

Great Britain, 2013 to 2017

Figure 3: Construction output for sewerage, current price, non-seasonally adjusted

Great Britain, 2013 to 2017

£ million

Source: Office for National Statistics, Output in the Construction Industry

This spike in sewerage output has been identified as unrepresentative of activity within the industry, as actual construction work did not start on these sewerage new orders immediately and will take a much longer period to complete than the assumed length of 11 quarters (as shown in Table 1). In addition to increasing the sewerage output level, the levels associated with other sub-sectors within infrastructure will subsequently also be reduced as a result.

Similarly, other new orders have been identified as being included in the model too early – representing work that was ordered well in advance of the intended start date, or work that was ordered but subsequently delayed.

Following the availability of additional data, we have identified the opportunity to improve this model and have worked closely with the ONS Construction Statistics Steering Group to address each of the three stated limitations.

4 . New model

A new model has been developed that refines the current approach of using new orders data to model construction output by sub-sector and region. The new model has been developed using more detailed data from our current new orders data supplier; Barbour ABI.

In September 2017, the contract for the supply of Office for National Statistics (ONS) new orders data was re-tendered. Following the completion of this process, the contract was re-awarded to Barbour ABI. The tender "Statement of Requirements" included a request for additional data, including the start date, duration and end dates of the higher-value projects, to be used to improve the model used to estimate regional and sub-sector construction output. Barbour ABI data were the subject of a Quality assurance of administrative data used in construction statistics report, published in March 2018.

The new model will be implemented back to Quarter 1 (Jan to Mar) 2010. To do this, data from the now discontinued ONS New Orders Construction Survey have also been used between 2010 and 2013. These survey data have been quality assured using Barbour ABI's administrative data from this period. In addition, the new model for "spreading" new orders data will be applied to some aggregated totals from pre-2010, to "spread" these values into 2010 and beyond. Users must be aware that all published regional and sub-sector estimates before 2010 will still be modelled using the current method.

The new method for estimating regional and sub-sector output will be implemented back to Quarter 1 2010 because this coincides with a current methodological break, whereby ONS construction output data transitioned from a quarterly to a monthly release, and where ONS new orders data transitioned from a monthly to a quarterly release. Revisions will therefore be seen in the back series of Tables 5 and 6 of construction output from 2010 onwards.

The new model adopts a two-staged approach, with different approaches taken for new orders worth over £50 million in value and those below £50 million in value. This value threshold has been selected because of data availability restrictions on projects with a value less than £50 million, and to give extra consideration to the largest value new orders, which have a significant bearing on the estimations of regional and sub-sector construction output.

Approach to be taken for new orders worth over £50 million

Projects above the chosen threshold value of £50 million will be modelled on a project-by-project basis. The value of the new order will be used to model output by spreading its value from its start date across the expected duration of the project using cost distribution curves, which are modelled from the now discontinued Projects in Progress Survey (PROBE), examples of which are shown in Figure 4.

Figure 4: Cost distribution curves

Figure 4: Cost distribution curves
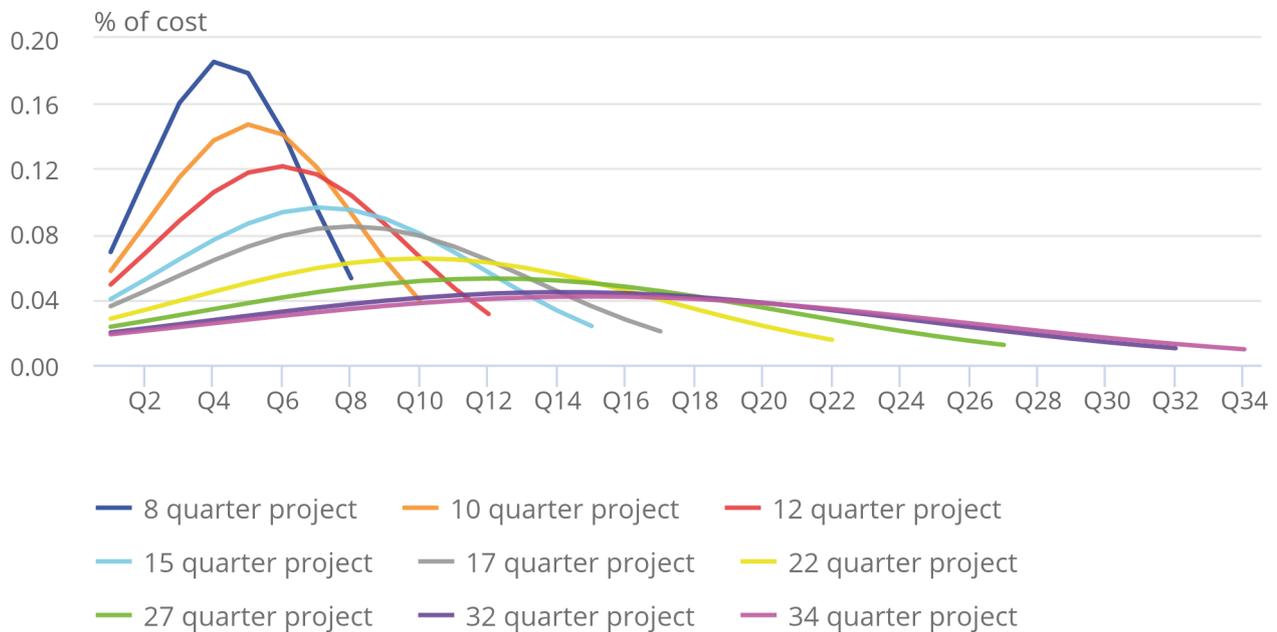
Source: Office for National Statistics, Quarterly Inquiry of Projects in Progress Survey

Whereas in the previous model the periodic weights and durations were selected on the basis of sector, for example, eight quarters for private housing, the new model provides a more flexible approach. The distribution of the project cost is determined by the duration of the project rather than the sector. This provides the capability to model a project over any length of time.

In addition to the PROBE data suggesting that the distribution of project costs is not spread evenly over the time, the Construction Statistics Steering Group also felt that a curved approach would be more appropriate than a constant spread of value. New cost curves have therefore been calculated, which are broadly normal distribution approximations of the previously used weights, displayed in Figure 1. Project duration information will be regularly updated as projects progress, providing more accurate details, which means that revisions to estimates will be inevitable.

Approach taken for £100,000 to £50 million projects

The approach taken for new orders with a value of between £100,000 and £50 million is similar in theory to the approach used in the current model. The new orders data for this value cohort are provided at an aggregated level, with a total value provided for each sub-sector, in each region, for each quarter.

To spread these aggregated totals, some assumptions are necessary. As these values are only provided with information on the quarter in which the new order was placed, each aggregated total must have a start date and duration applied to it. To do this, an additional average duration dataset from Barbour ABI has been used to created two weighted averages for projects worth between £100,000 and £50 million. These are:

- weighted gap equals the difference between new orders date and start date

- weighted duration equals the difference between the start date and end date

These two variables are created as an average by sub-sector, split by public and private work, by firstly multiplying the value of individual projects by the given gap or duration and then grouping these values by sub-sector and dividing them by the total sub-sector value.

Table 1: Weighted gaps and weighted duration, by sub-sector, quarters

Sub-sector	Gap		Duration	
	Public	Private	Public	Private
Agriculture	2	2	4	4
Air	2	2	5	5
Coal	2	2	6	4
Communications	2	2	5	5
Electricity	3	3	4	4
Entertainment	3	3	5	6
Factories	2	2	4	4
Garages	2	2	4	4
Gas	2	2	5	5
Harbours	2	2	7	7
Health	3	3	6	6
Housing	2	3	7	10
Miscellaneous	3	2	7	6
Offices	2	2	6	6
Oil	2	2	6	4
Railways	3	3	6	6
Roads	2	2	7	7
Schools	2	2	5	5
Sewerage	2	2	5	5
Shops	2	2	4	4
Steelworks	2	2	6	4
Universities	2	2	6	6
Warehouses	2	2	4	4
Water	2	2	5	5

Source: Office for National Statistics and Barbour ABI

Table 1 shows the weighted gaps and durations by sub-sector used to allow the new orders worth between £100,000 and £50 million to be used to model output. The weighted gaps and durations have been selected using a pre-determined selection criteria, ensuring a sufficient sample size for each sub-sector is used. If the individual sub-sector has not met the criteria, a sector-level weighted gap and duration are used.

The aggregated totals for projects worth between £100,000 and £50 million will be used to model output using the cost curves shown in Figure 4 from the start date to end date, where the start date is equivalent to the new orders date plus the weighted gap, and the end date is equal to the start date plus weighted duration.

For example, once the aggregated total for public railways projects worth between £100,000 and £50 million is provided by Barbour ABI in, for example, Quarter 1 2016, the weighted gap of three quarters will be applied. This will essentially delay the impact of these totals on the output apportioned to railways until Quarter 4 (Oct to Dec) 2016, when the total will then be spread over the set duration for public railways projects – six quarters – using the cost distribution curves.

Once projects from both value cohorts have been spread across their duration, the values for each sub-sector and region in each quarter are summed and constrained to output. These values are then used to model both regional and sub-sector quarterly construction output.

The new model for estimating regional and sub-sector construction output has been discussed at, and agreed by, ONS's Methodological Assurance for Statistical Transformation group.

5 . Impact of new model

Once the new model is implemented, revisions back to Quarter 1 (Jan to Mar) 2010 will be evident. However, it is worth noting that no revisions will be made to top-level, chained volume measure or current price totals, only the lower-level sub-sector and regional estimates will be revised. These revisions will be available in Construction output in Great Britain: April 2018 and new orders January to March 2018 publication, released on 11 June 2018. However, for the purpose of this article, some provisional revisions to specific sub-sectors and regions, up to Quarter 4 (Oct to Dec) 2016, can be seen in this section.

Sub-sector impact: example

As already eluded to in Section 3 of this article, the sewerage sector had been identified by both Office for National Statistics (ONS) and the Construction Statistics Steering Group as a particular sub-sector of concern. Figure 5 shows the impact of applying the new model on the value of output apportioned to the sewerage industry in comparison with the previous model.

Figure 5: Sewerage construction output – current and new model, current price, non-seasonally adjusted

Great Britain, 2010 to 2016

Figure 5: Sewerage construction output – current and new model, current price, non-seasonally adjusted

Great Britain, 2010 to 2016
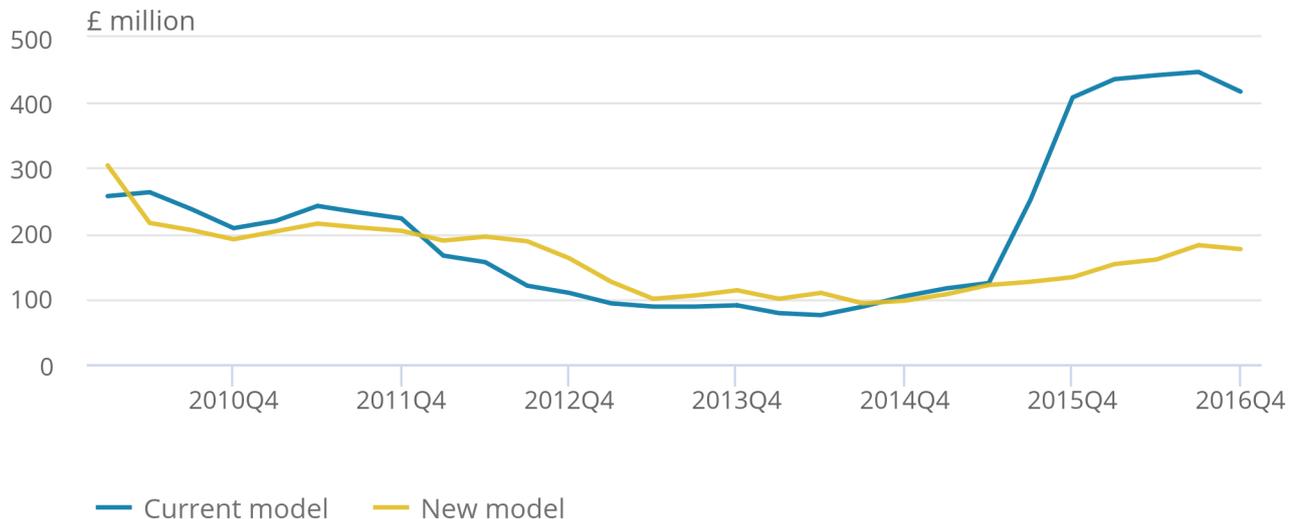
Source: Office for National Statistics and Barbour ABI

As explained in Section 2 of this article, the sewerage sector previously experienced an unrepresentative spike in output beginning in Quarter 3 (July to Sept) 2015, because multiple high-value new orders were being used to model sewerage output before their start date. However, using the new model, which delays the projects being used to model output until their construction start date, and then spreads the values across the genuine expected duration of the project, the sewerage totals are far more representative of actual construction activity within the industry.

Despite the new model producing considerably lower estimates for the sewerage sub-sector in 2015 to 2016 compared with the current model, the amount of construction work apportioned to sewerage does begin to rise in 2016 as a result of one of the high value projects beginning construction work. As the other high value sewerage new orders also begin construction work in the future, the amount of construction output apportioned to the sewerage sub-sector will continue to rise.

Regional impact: example

Revisions will also be made to the regional breakdown of construction output. For example, Figure 6 shows the value of output apportioned to Wales using both the current and new models. Wales experienced a spike in 2015, which occurred because of a high-value infrastructure new order in Quarter 1 2015, which was immediately used to model output. This spike was identified as unrepresentative of genuine construction activity following consultation with users, like that seen in the sewerage sub-sector.

Great Britain, 2010 to 2016

Figure 6: Construction output in Wales – current and new model, current price, non-seasonally adjusted

Great Britain, 2010 to 2016
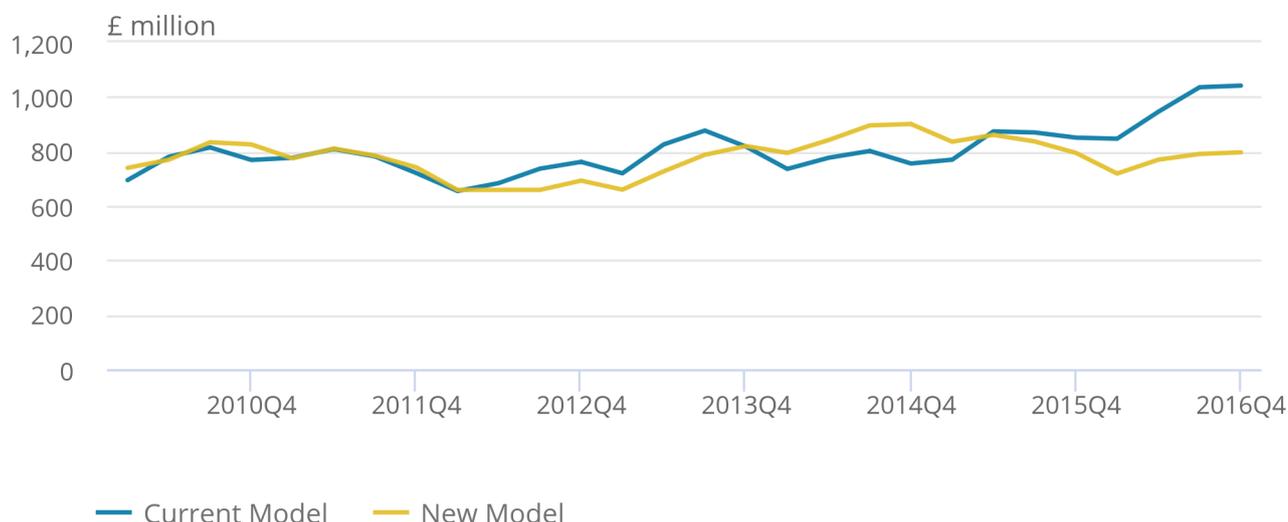
Source: Office for National Statistics and Barbour ABI

In reality, actual construction work for the high-value new order causing the spike in construction output in Wales has, at the time of writing, still not actually begun, even though the project had fully fed through the current model. Using the new model, the new order will not be used to estimate output until its given start date, and then spread across its actual construction duration.

In addition to the two examples of sub-sector and regional revisions given in this section, full revisions back to 2010 will be made to the value estimates for each respective sub-sector and region.

6 . Implementation

Following consultation with the Office for National Statistics (ONS) Construction Statistics Steering Group, it is clear that once implemented, the new model will lead to an improved accuracy in the regional and sub-sector breakdown of construction output, by addressing the three major limitations of the existing model.

Users should be aware that pre-2010 data will still be modelled using the previous model used to apportion regional and sub-sector output, and will therefore be unrevised. Data from 2010 onwards will be estimated using the new method outlined in this article. As a result of the constraining of data used in both the current and new model, no level shift will be evident between pre- and post-2010 region and sub-sector estimates.

The full results of the new model will first be seen in Tables 5 and 6 of [Construction output in Great Britain: April 2018 and new orders January to March 2018](#), released on 11 June 2018.

7 . Further work

We believe that the new model outlined in this article represents a large improvement in comparison with the current model. However, as part of the new model, some assumptions still need to be made. As part of the ongoing construction statistics development programme, we will continue to investigate potential areas for development, including the access and use of new data sources, which may result in further improvements to our estimates of regional and sub-sector construction output estimations.