

BIS Building Materials Statistics:

Work carried out under the Quality Improvement Fund 2011-12

**July 2012**

**Sumit Rahman and Charles Lound**

**Methodology Advisory Service,**

**Office for National Statistics**

# BIS Building Materials Statistics: Work carried out under the Quality Improvement Fund 2011-12

This report covers three pieces of methodological work carried out for the Department for Business Innovation and Skills on Building Materials Statistics. This work was funded by the Quality Improvement Fund (QIF) of the UK Statistics Authority.

The three pieces of work cover:

* Seasonal adjustment of the times series presented in the Monthly Statistics of Building Materials and Components
* Sample design work recommending a new approach for the Sand and Gravel and Blocks Surveys
* Imputation for newly sampled producers where no previous value is available to inform the imputation

To see the detail in the charts in the first section, this report should be printed in colour.

The same section includes embedded Excel workbooks in the Word version of this document. If presented in other formats, this embedding may not be preserved, in which case these files should be made available with the document.

# Seasonal adjustment

***Summary of main points***

* BIS should start seasonally adjusting key series using some version of the X-12-ARIMA software
* We have provided ‘spec files’ for use with X-12-ARIMA for deliveries of sand and gravel, ready-mixed concrete, bricks, blocks and cement
* BIS should consider publishing the seasonally adjusted time series as well as publishing the graphs of the series (and/or the trends), to allow users to calculate growth rates of the seasonally adjusted series
  1. The Monthly Statistics of Building Materials and Components bulletin[[1]](#footnote-1) contains graphs for seven of the published series, using a moving average to smooth out seasonal movements. Three of the series (deliveries of bricks, blocks and cement) are monthly and are graphed in one chart as index numbers with base year 1983. Two quarterly series (deliveries of sand and gravel, and of ready-mixed concrete) are graphed together in a separate chart, again with 1983 as a base year for indices. Finally, total imports and total exports are graphed in current prices.
  2. We have considered the series on the first two graphs for this report. They are reproduced below.
  3. The bulletin does not currently publish the index numbers that result in these graphs (although these are available in the accompanying Excel spreadsheet).
  4. We can think of these graphs as graphs of asymmetric, left-sided, moving averages with equal coefficients. So the value for 2011 quarter 2 is the mean of the values for the four quarters from 2010 quarter 3 to 2011 quarter 2. Moving averages are a simple way of smoothing a series, and if the series is thought to be seasonal then using a moving average of length one year is a simple way of removing most of the seasonality. (If the seasonal effects are additive and fixed from one year to the next, then a simple moving average – one with equal coefficients – will remove the seasonal effects.)



* 1. One problem with symmetric moving averages is that values cannot be calculated for the ends of the series. This means that, for example if a 12 month moving average were being used then there would be a delay of six months before the estimate for a new month could be produced[[2]](#footnote-2). BIS deals with this problem by using the asymmetric, left-sided moving average. But then this means that the graph is no longer appropriately centred: turning points on the graph will be seen half a year ‘too late’.
  2. For example, in the graphs above we see in January 1995 a clear turning point for bricks, blocks and ready-mixed cement. But because of the misleading centring of the moving averages, the turning points are really in the summer of 1994 (refer to later graphs of the seasonally adjusted series to see this).
  3. There are a number of other ways of dealing with this end of series problem, but the recommended method for the Government Statistical Service for seasonal adjustment is to use the X-12-ARIMA software (or other versions of this), and this will produce series that have removed identifiable seasonal effects and generally have good properties at the end points.

## Overview of method

* 1. We assume that the time series of interest, *Y*, is decomposable into a trend series (commonly denoted *C*), a series of seasonal effects (*S*) and a residual irregular series (*I*). We expect C to be relatively smooth and the seasonal effects for any given month to vary smoothly from one year to the next (though they need not vary smoothly from one month to the next within any year). Usually we find with economic time series that a multiplicative decomposition is appropriate[[3]](#footnote-3), so that. In this case the seasonal effects are called seasonal factors and these are dimensionless numbers that we expect to vary about the value one. The seasonally adjusted series is.
  2. ONS publishes a *Guide to Seasonal Adjustment[[4]](#footnote-4)* which gives a good description of the method for estimating *C*, *S* and *I*. The basic idea is to use a moving average on *Y* to estimate the trend *C*, divide *Y* by *C* to obtain a detrended series *SI*, then apply a set of moving averages on the *SI* series (for each month/quarter of the year) to estimate *S*, and therefore a seasonally adjusted series *CI*. But this process is repeated two more times to arrive at progressively better estimates. Depending on the stage of the iterative process, different moving averages can be used which are more appropriate for the sort of series you expect to be faced with at that stage.
  3. The basic method includes steps to identify outliers and reduce their influence on the various moving averages. But for this method to work at its best, the original series may need to be ‘cleaned’. This is typically done using regression techniques, which are combined with so-called ARIMA modelling so that the regressions can be used validly within time series. We can adjust for outliers (e.g. industrial action which depresses output for one month only, which we wish to exclude temporarily to enable estimation of trends unaffected by the strike but will include in the final seasonally adjusted series) and for *calendar effects*.
  4. One calendar effect is the Easter effect: because the date of Easter varies from one year to the next, and it often has a significant impact on the series of interest, this can disrupt the estimation of the seasonal effects in March and April (or in quarters 1 and 2).
  5. Another is the trading day effect. Since more construction work takes place in the week rather than at weekends, we would expect to see higher demand for materials in those periods that happen to contain more weekdays. Similarly, the opportunity for production of materials is higher in periods with more weekdays. For example, there were 21 weekdays in July 2011 and 23 in August 2011 because an extra weekend happened to fall in August, so even if there were no other differences between those months we might expect to see higher production, deliveries and sales in July. Adjusting for the trading day effect allows us to have a time series that is more consistent between months (or quarters), so that we are comparing ‘like with like’ and this also leads to more efficient application of the time series methods
  6. The use of ARIMA models also allows us to use symmetric moving averages, as these models enable the software to extend the end of series with forecasts and backcasts, so that the moving averages can run right to the (true) ends of the series.

## Objectives

* 1. A look at the graphs for the three monthly series shows that the they tend to move in the same direction (although there is some growth in the blocks series during the period 1997-2006 when the other series are broadly flat).
  2. It makes sense to aim for some consistency in the seasonal adjustments that we recommend for the series. There is less need to do this between the monthly and quarterly series as the seasonal adjustment parameters are likely to be different anyway, but consistency between the three monthly series is a sensible aim. This means that although we focussed on the series individually, we would use findings from investigating one series and revisit earlier series. As an example, when looking at the blocks series, after some potential outliers were identified for adjusting, we then revisited the bricks series and investigated the impact of adjusting for these outliers. Nevertheless, the aim of consistency was not at the cost of producing good quality seasonal adjustments for each series.
  3. Because users of the BIS statistical publication are familiar with the current bulletin and the graphs published by the Department, we have made a point of comparing the trend series (*C*) with the published graphs, and on considering the quality of the trend series that the software estimates, as well as the more common quality measures for the seasonally adjusted series.
  4. The business area that produces these statistics does not have much experience in using the X-12-ARIMA software, so we have tried to produce specification files that are straightforward to use and are robust, so that regular updating is less likely to be needed. We have not recommended the use of separate files of prior adjustments, so that all adjustments are performed by the main body of the software itself. The department will need to decide on a revisions policy to deal with the issue of how the seasonal adjustment process is affected by each new data point.[[5]](#footnote-5)

## Results – monthly series

* 1. The recommended seasonal adjustment for the bricks series can be compared with the unadjusted series in the following graph:



* 1. The adjusted series’ graph certainly looks smoother, and there is no evidence reported of residual seasonality in the X-12-ARIMA standard output. The following graph shows the trend series that has been estimated for the bricks series and compares this with the BIS published graph. The dotted line is what is currently published, and the pink line is the same graph centred as it would be had a symmetrical moving average been used instead of the current asymmetric one.



* 1. This graph illustrates the point made earlier in paragraph 1.05 that the asymmetric moving average results in an average that is lagged by six months. For example, there is a turning point in the trend estimated by the seasonal adjustment software to have taken place in mid-1994, but the published graph turns in January 1995. Of course the pink line does not extend as far as the dotted line (or the blue trend line), which is the drawback of the simple symmetric moving average (or moving total).
  2. The trend series and the simple moving average are similar to each other, although the trend series appears to have more pronounced peaks. The paths in 1988 look rather different, but over the course of nearly 30 years there is a good degree of agreement. We investigated the effect of using a different moving average for the estimation of the trend from that automatically chosen by the X-12 algorithm. A shorter ‘Henderson filter’[[6]](#footnote-6) actually resulted in a seasonal adjustment that was better according to some of the quality measures, but the trend estimate becomes even less smooth and we judge that there is no compelling reason to change the trend in this way.
  3. The recommended seasonal adjustment for the blocks leads to the following graphs:



* 1. We see again that the trend series has more pronounced peaks than the graph that is currently published by BIS, but the general agreement is close once more, once we have corrected the lag in the BIS graph. For this series we found that a shorter Henderson filter improved most of the key quality measures for seasonal adjustment, but the resulting trend is probably dominated too much by relatively short-term fluctuations. The following graph shows the trend term in these cases:



* 1. The fact that there was no compelling reason to change the Henderson filter for bricks (or for cement) is itself another reason to keep the longer filter for blocks as well.
  2. We found with the blocks series that there were rather more outliers detected automatically by the algorithm than there were for bricks or cement. Many of these extra ones came from very early in the series, before March 1986. The outliers automatically arising from the cement series were different again from the bricks’, but we have arrived at a set of six outliers that result in satisfactory seasonal adjustments for all three monthly series. These six outliers usually are common to at least two of the automatically identified outlier sets and generally we have been able to determine something extreme in the weather conditions in the month for that outlier.
  3. We have included only so-called additive outliers to be adjusted for by X-12-ARIMA. These are temporary adjustments so should not affect the trend series but still appear in the ‘irregular’ series and therefore also the seasonally adjusted series. Sometimes the algorithm identified level shifts as well but these have not been included. These were seen in early 1989 for bricks and blocks and in 2009 for cement but keeping these level shifts does not seem to help anything much, as the steep declines appear to take place over several months and are picked up adequately by the trend. Including the 1989 level shift in particular seems to distort the trend badly, making the peak in January 1989 even more pronounced. If there is indeed a peak at this time, which the simple moving average is not picking up here because of the steep decline throughout 1989, then the trend series is still picking this up without needing the level shift adjustment.



* 1. Next we can look at the results for the cement series’ seasonal adjustment:



* 1. The basic pattern of results is the same, though here the agreement between the simple (symmetric) moving average and the trend appears to be especially close. For this series the algorithm preferred the additive decomposition for the seasonal adjustment model. In fact, under the default settings the algorithm preferred this for all three series. However for the bricks series we had a number of compelling reasons to prefer the multiplicative model: there was evidence that the size of the irregulars was reducing as the trend reduced; the SI series were much better (in that they were more stable and amenable to having the seasonal factors estimated by moving averages); the quality of the seasonal adjustment was better; the ARIMA model that was used to describe the series was more parsimonious. These reasons were less evident with the other two series, but as there were few ways in which these were better with the additive model we were content to use the multiplicative decomposition in all three cases.
  2. Ultimately this is a judgement call and the next time BIS reviews these statistics and their methodology the department could reconsider which decomposition to use. There is a large knock-on effect on the other parameters (for example the set of outliers automatically identified by the algorithm changes, the ARIMA model changes, the evidence for trading day and Easter effects change, *etc*) and we did not have time to explore this fully for all three monthly series.[[7]](#footnote-7)
  3. The ARIMA model that was chosen for all three was the same, the well-known ‘airline’ model[[8]](#footnote-8), described as (0, 1, 1)(0, 1, 1)12 in the standard ARIMA notation. There was a case for a more complicated model for the cement series, with the software selecting a (2, 1, 1)(0, 1, 1)12 model (two more parameters than with the airline model), but there was no meaningful improvement to the seasonal adjustment of the cement series with this model compared with the airline model.
  4. For each of the three monthly series we found evidence for an Easter effect. The best fitting variable to add to the data cleaning procedure (the ‘regARIMA’ module) was the ‘Easter[1]’ variable. This means that the Easter effect is to change the level of the series for just one day – in other words, the Easter effect is short term and does not last for, say, a full week[[9]](#footnote-9). We also found for each series a trading day effect. (These calendar effects were not always significant with the use of an additive seasonal decomposition, but were clearly identifiable with the multiplicative model, which gives us another reason to prefer the multiplicative model.) The trading day effects were adequately modelled using the single trading day regression variable which models a single effect for weekends and a single effect for weekdays. Slightly better fits were found with the model that fits a separate effect for each day of the week, but the cost in extra parameters to be estimated outweighs the improved fit, and there was certainly no discernible difference in the seasonal adjustment.
  5. One difference in our treatment of the cement series is that we found that there were more problems with identifying trading day effects and removing residual seasonality. There were fewer problems when we allowed the software to select outliers automatically, but a simple way to remain consistent with the outliers chosen for the other series and remove the problems with the trading days and residual seasonality was to eliminate the first year of the series (1983). This led to a change in the ‘seasonal filter’ for the series.
  6. The seasonal filter is the moving average that is used on the *SI* series to estimate the seasonal factor series *S*. The length of the filter used is determined by considering the relative volatility of the irregular series to the seasonal series. We found that the value of this relative volatility was close to the cut-off value used by the algorithm to determine the seasonal filter, but this became more stable if 1983 was ignored.
  7. So for the purposes of fitting the regARIMA model we ignored 1983. We tested that the seasonal adjustment was stable once 1983 was ignored for model fitting, and it was. We also investigated the effect of ignoring 1983 for the other series, but there was no improvement to these seasonal adjustments, including to their stability.
  8. We did find that there were convincing reasons to have different sets of seasonal filters for each series. For cement there was no need to deviate from the algorithm’s automatic choice. For bricks the *SI* graph for January indicated that the seasonal filter was not working well, although it was satisfactory for the other months. A shorter filter for January improved the seasonal factors for January and improved the overall seasonal adjustment. (This included using the ‘history’ feature of the X-12-ARIMA software to see the effect on revisions histories: there were smaller revisions, on average, to the end of the series when new data points were added – this was considering both the seasonally adjusted series and the trend series.)
  9. For blocks, a shorter seasonal filter was considered for all months. This led to seasonal factors that tracked the *SI* ratios better, but for April we had the seasonal factors changing frequently between being greater than and then less than one. The *SI* ratios in fact suggest that perhaps there is no real seasonal effect in April. By changing the filter to a longer one (a 3x9 seasonal filter), the April seasonals are stable and very close to one, which seems to be the most appropriate result.
  10. So to summarise the recommended settings for the three series:

|  |  |  |  |
| --- | --- | --- | --- |
| ***Setting*** | **Bricks** | **Blocks** | **Cement** |
| *Decomposition* | Multiplicative | Multiplicative | Multiplicative |
| *ARIMA model* | Airline | Airline | Airline |
| *Seasonal filters* | 3x3 for January, 3x5 for the rest | 3x9 for April, 3x3 for the rest | 3x5 for all months |
| *Trend filter* | Henderson 13 | Henderson 13 | Henderson 13 |
| *Calendar effects* | Easter[1], weekend/weekday trading day | Easter[1], weekend/weekday trading day | Easter[1], weekend/weekday trading day |
| *Additive outliers* | January 1987  February 1991  February 1996  February 2009  January 2010  December 2010 | January 1987  February 1991  February 1996  February 2009  January 2010  December 2010 | January 1987  February 1991  February 1996  February 2009  January 2010  December 2010 |

* 1. The graph below shows the trend series in a similar format to the graphs that BIS published. We can see that the turning points are now six months earlier than in the published graph, and the trends are smoother in appearance as well.



## Results – quarterly series

* 1. The two quarterly series that were considered are deliveries of sand and gravel and ready-mixed concrete (RMC).
  2. For sand we found the multiplicative decomposition worked best. The seasonal adjustment was very successful under the default settings, with stable seasonal factors and seasonal filters that fit the SI graphs well. The resulting irregular factors are consequently small and the seasonally adjusted series pretty smooth.



* 1. The simplest ARIMA model which helped the seasonal adjustment was the (0, 1, 0)(0, 1, 1)4 model. A single additive outlier was found for 1986 quarter 1 – this was highly significant (with a *t* value of 4.9) and might be due to a cold February in that year. No Easter effect was found, but a trading day effect was found. This was the model with different effects for each day of the week and there is an improvement to the seasonal adjustment (in particular to the M1 statistic, which means the irregular factors are smaller when we adjust for trading days), so we have kept the trading day effect. This is unusual for quarterly series as the distribution of the days of the week do not change much at all between quarters (unlike for calendar months). Our decision to keep the trading day factors was helped by the fact that a similar effect was found for the RMC series.[[10]](#footnote-10)
  2. The small irregular means that the 5 term Henderson filter is appropriate for the trend filter. For the seasonal filter, the algorithm goes for a 3x5 filter but it is marginal because quarter 1 has a rather different ‘*I*/*S* ratio’ to the other three quarters. An examination of the *SI* graph for quarter 1 below does not give us any cause for concern over the use of the 3x5 filter however, so we have fixed the seasonal filter as 3x5 for all four quarters.



* 1. Finally let us look at the recommended seasonal adjustment for the RMC series:



* 1. The seasonal adjustment is very successful again. The algorithm preferred the additive decomposition, but forcing the multiplicative decomposition improved the M1 statistic considerably, suggesting that the multiplicative decomposition explains more of the variation in the series. As mentioned earlier, a trading day effect was found to improve the seasonal adjustment so we kept it in our final specification. And there was an Easter effect found, too, which improved the M1 statistic considerably.
  2. The automatic selection procedure found a (0, 2, 1)(0, 1, 1)4 ARIMA model to perform better than the airline model, with a better likelihood statistic. But we found that using the airline model actually meant that the series was more stable, in the sense that smaller revisions were made to the end of the series when points were added to it. So we chose the airline model for this series.
  3. The only outlier we found was for the final quarter of 2010 (which was extremely cold). So there was no overlap in the outliers sets for the two series. We looked at including both outliers (1986 quarter 1 and 2010 quarter 4) for both series, but there was no improvement to the sand and gravel seasonal adjustment, so there was no compelling reason to include both outliers for both series.
  4. Finally there was no reason to change the filters, both trend and seasonal, from those the algorithm selected. So the 5 term Henderson trend filter and the 3x5 seasonal filters were fixed for this series. To summarise the settings for each of the quarterly series:

|  |  |  |
| --- | --- | --- |
| ***Setting*** | **Sand and gravel** | **RMC** |
| *Decomposition* | Multiplicative | Multiplicative |
| *ARIMA model* | (0, 1, 0)(0, 1, 1)4 | Airline |
| *Seasonal filters* | 3x5 | 3x5 |
| *Trend filter* | Henderson 5 | Henderson 5 |
| *Calendar effects* | 7 day trading day | 7 day trading day, Easter[1] |
| *Additive outliers* | 1986 quarter 1 | 2010 quarter 4 |

* 1. The graph of the trend series in the same style of the published graphs (see below paragraph 1.02) is as follows.



* 1. There is not much difference between this graph and the published one (apart from the centring issue), although it does show that the trend series is not as smooth as the simple moving averages. We could smooth the trends further by using longer Henderson filters, but we should recall that the *I*/*C* ratios for both series were very low, about 0.5, and this is very far from the 3.5 value that is the usual figure for the algorithm to select a 7 term Henderson filter.

## Quality measures for the seasonal adjustments

* 1. The Quality Centre at ONS has collected and developed a large set of quality measures for statistics[[11]](#footnote-11), and there are several that relating to seasonal adjustment. Chapter 27 of the *Guide to Seasonal Adjustment[[12]](#footnote-12)* describes these measures and gives some examples of them. It groups the measures into four groups: quality measures for the seasonal adjustment, for the forecasts, for the trend estimates and for constraining. BIS could consider using some of these measures if it produces quality reports for these statistics.
  2. To illustrate the use of these measures, we have used the measures relating to the seasonal adjustment and to the trend estimation. Measures have been taken either from the standard X-12-ARIMA output or calculated in the spreadsheet that is attached to this report.
  3. For the quality of the seasonal adjustment, a simple first step is to compare the graph of the original series with that of the seasonally adjusted series.
  4. The M7 statistic is an output of the X-12-ARIMA software and indicates whether there is sufficient stable seasonality in the original series, compared to moving seasonality, for the X-12 method to work effectively. It can be thought of as a continuous statistic for the combined test for identifiable seasonality that the software produces in Table D.8A of the standard output. Values lower than 1 indicate the series is suitable for seasonal adjustment using this algorithm, and lower values indicate more successful adjustments.
  5. The ANOVA statistic for the seasonally adjusted series tells us how much of the month-to-month[[13]](#footnote-13) variation is explained by the trend component. This gives an idea of how much short term growth is due to the irregular component. A low value suggests that monthly changes in the seasonally adjusted series are not a reliable indicator of changes to the trend. The STAR statistic (stability of trend and adjusted series rating) is the mean absolute monthly change in the irregular, so it provides a measure of the volatility of the irregular component. A useful interpretation comes from dividing it by two: this is the expected revision to the most recent estimate when a new point is added to the series.
  6. For the quality of the trend estimates, the visual check could be of the trend and the seasonally adjusted series, or of the trend and the original series. (We have not produced graphs for either of these cases in this report, but they can be easily produced from the spreadsheet data we have included with the report.)
  7. An ANOVA statistic for the trend compared to the original series can be calculated. This gives the proportion of variation in the original series that is due to the variation in the trend. A high figure would indicate that the series is dominated by the trend.
  8. The MCD (months for cyclical dominance) is calculated by X-12-ARIMA and indicates how many months it typically takes for the growth in the trend to become greater than the growth in the irregular component. It is another measure of the volatility of the seasonally adjusted series. An MCD of three, say, means that looking at growth *in the seasonally adjusted series* from month-to-month will not give a reliable indication of how the trend is moving, but a growth rate based on the last three months compared to the three months prior to this will. (Of course such a growth rate would be centred away from the end of the series, so a high MCD indicates that the seasonally adjusted series has poorer quality information about the behaviour of the trend at the end of the series.)
  9. Finally the CTQ (the contingency table Q) shows how frequently the gradients of the trend and seasonally adjusted series have the same sign. If there is no association between the two the CTQ will be 0.5 (or so). This would happen if the trend is very flat or if the series is very volatile. A CTQ close to one means the trend and seasonally adjusted series both tend to move in the same direction month after month.
  10. We can comment on the quality of the seasonal adjustments we have recommended in this report using these measures as follows.
  11. For the bricks series, the M7 statistic is 0.152, which is low, indicating a relatively large ratio of stable seasonality to moving seasonality. This means the series exhibits clear seasonality in a way for which the X-12-ARIMA software can adjust effectively. We can see from the graph that the seasonally adjusted series is smoother than the original series. The ANOVA statistic is low, at 4.4%. This indicates that monthly movements in the seasonally adjusted series are dominated by the irregular component - just 4.4% of the monthly change is explained by the trend. The STAR statistic is 5.3%, which means that the expected revision to the latest point in the seasonally adjusted series, after a new point is added, is about 2.6%. The ANOVA for the trend component compared to the original series is just 0.9%, showing that the monthly variation in the original series is barely affected by the trend component. The MCD is 5, which means that on average it takes 5 months for the trend component to grow more than the irregular component. This suggests that you cannot draw firm conclusions from the monthly growth rate for the seasonally adjusted series, but a 5 month on 5 month comparison is more suitable for this purpose. The CTQ is 61%, which is only a little above 50%. This indicates that whether the trend goes up or down from month to month is only very weakly associated with whether the seasonally adjusted series goes up or down. This is another reason to be careful about reading too much into monthly changes in the seasonally adjusted series.
  12. For the blocks series the M7 statistic is 0.189, which means the series can be adjusted effectively. We can see from the graph that the seasonally adjusted series is smoother than the original series. The ANOVA statistic is low, at 4.0%, so monthly movements in the seasonally adjusted series are dominated by the irregular component. The STAR statistic is 4.6%, which means that the expected revision to the latest point in the seasonally adjusted series, after a new point is added, is about 2.3%. The ANOVA for the trend component compared to the original series is just 0.4%, showing that the monthly variation in the original series is barely affected by the trend component. The MCD is 4. The CTQ is 61%, which is only a little above 50%. So, just as with the bricks, we should be careful about reading too much into monthly changes in the seasonally adjusted series.
  13. For the cement series, the M7 statistic is 0.219, which is still low, so the series is suitable for seasonal adjustment. Once again we see from the graph that the seasonally adjusted series is smoother than the original series. The ANOVA statistic is low, at 2.1%, and the STAR statistic is 4.6%, which means that all three monthly series have similar degrees of volatility in their irregular factors. The ANOVA for the trend component compared to the original series is again very low at 0.4%. The MCD is 5, just as it is for the bricks, and the CTQ is 61%, the same as for the blocks. For all three monthly series we have to be careful about reading too much into monthly changes in the seasonally adjusted series.
  14. For the sand and gravel series, the M7 statistic is 0.172 so this series exhibits clear seasonality. We can see from the graph that the seasonally adjusted series is smoother than the original series. The same is the case for the RMC series, where the M7 statistic is 0.170. The ANOVA statistic for sand and gravel is 34.1%. This indicates that quarterly movements in the seasonally adjusted series are affected more by the irregular component than by the trend, but about one third of the variation in quarterly change is explained by the trend. For RMC the ANOVA is 48.8%, so about half the variation is explained by the trend. The STAR statistic for sand and gravel is just 2.5%, which means that the expected revision to the latest point in the seasonally adjusted series, after a new point is added, is about 1.2%. For RMC it is even lower, with a STAR statistic of 2.2%. This shows that the seasonally adjusted series are not very volatile. The ANOVA for the trend component compared to the original series is just 4.9% for sand and gravel and 8.2% for RMC, showing that the quarterly variation in the original series is barely affected by the trend component. (In fact, given the ANOVA for the trends compared to the seasonally adjusted series, which was non-negligible, this shows that the seasonal component drives quarterly changes.) The QCD is 1 for both series, which suggests that you can reasonably consider quarterly changes in the seasonally adjusted series to indicate changes in the trend. The CTQ is 81% for sand and gravel and 80% for RMC, which is fairly high. This indicates that there is some association between quarterly direction of changes in the trend and seasonally adjusted series.

## Further considerations

### Revisions policy

* 1. BIS needs to consider a revisions policy to hold if it decides to move to this method of seasonal adjustment. A new data point can result in revisions to the whole of the seasonally adjusted series, but this might not be welcomed by users. One solution is to state in advance when the back series will be revised, and how far back. A possible policy is during the year to revise back to the start of the last completed year, but when publishing for December to open up the whole series for revision. This might be a suitable time to revisit the parameters of the seasonal adjustment (e.g. decide if the filters need to be changed). When part of the series is closed to revision, the X-12-ARIMA software will still of course produce revised estimates for the closed period, but BIS would simply not publish these until they become open for revision again. This is just a suggestion: it would be sensible to develop a revisions policy in consultation with users.

### Presentation

* 1. The department needs to consider how to present the results of the seasonal adjustment process. We have shown how the current graphs look if the moving averages/totals are replaced by their trend series. ONS recommends that seasonally adjusted series should be graphed and their values published, while if the trend is to be shown it should be graphed with the seasonally adjusted series, with a dotted line used to indicate the trend for the last part of the series (according to the MCD or QCD – so if the MCD is 5 then the last 5 months of the trend should be dotted), and the values should be made available on request. (See chapter 26 of the *Guide to Seasonal Adjustment[[14]](#footnote-14)* for more details.) Once again it would be sensible for BIS to consult with users of these statistics to determine what sort of presentation would be best.

### Constraining to annual totals

* 1. There is no mathematical reason why the annual totals of a seasonally adjusted series should equal the annual totals of the unadjusted series. If the seasonal factors for a year average out close to one, and the trend is reasonably stable over the course of the year then there should be close agreement. In practice users of official statistics often prefer the annual totals to agree in this way, and X-12-ARIMA has options to do this. We can constrain the seasonally adjusted series so that calendar years agree, or financial years – but not both simultaneously.
  2. Such constraining does distort the seasonal adjustment though and the quality measures for constraining give some indication of the distortion that occurs. The measure comparing annual totals is included in the attached spreadsheet on quality measures and the mean difference between the annual totals for the original series and their seasonally adjusted versions is typically about 0.3% (0.5% for sand and gravel).
  3. We tried constraining the sand and gravel series, since this had the worst average discrepancy. The mean absolute percentage difference for individual points is just 0.09%, with a median of 0.05% and maximum of 0.45%, so there is little apparent distortion caused by the constraining.
  4. Constraining can cause more distortion in monthly series where there are trading day effects identified and corrected for. In the blocks series, which had the smallest difference from paragraph 1.66, the mean absolute percentage difference for points is 0.29%, with a median of 0.21% and a maximum of 2.1%, so there is more distortion here.[[15]](#footnote-15) Nevertheless the graph shows that there is practically no difference by eye between the constrained and unconstrained series.



* 1. Once again BIS should consult with users to determine whether constraining should be used for the seasonally adjusted series.

### Sub-series

* 1. We have not looked at the sub-components of these series in this report. There are a number of ways to breakdown the series. For example the bricks series can be broken down by brick types, brick materials or by region of production. It is possible to seasonally adjust these sub-series and then sum them to arrive at an indirectly adjusted total series for bricks. The ‘composite’ option in the X-12-ARIMA software allows us to compare the indirectly and directly adjusted series to help decide which approach is better to adopt.
  2. However there are some issues that BIS needs to be aware of if it wishes to explore this. For the regional breakdown there is a lot of suppression of values following disclosure control, but the X-12-ARIMA algorithm needs complete time series. The breakdown of bricks into brick types and brick materials are different and separate breakdowns, so would lead to two different seasonal adjustments of the total, if this were to be indirectly adjusted. If BIS preferred to stick with a direct adjustment of the total, then the seasonally adjusted sub-series would no longer sum to the total series. The other problem with adjusting the sub-series is that these series would usually be subject to relatively greater sampling error, potentially leading to less effective seasonal adjustment (as the sub-series are probably subject to more noise).
  3. We recommend that BIS spends some time getting used to the production of these higher level seasonally adjusted series before considering the seasonal adjustment of the sub-series, in consultation with users.

**Specification files for the seasonal adjustments**

Bricks

series{

title="Deliveries of bricks"

start=1983.jan

period=12

decimals=0

name="Bricks"

file="H:\My Documents\BIS\second project\seasonal adjustment\bricks.txt"

save=(a1)

}

transform{

function = log

}

arima{

model = (0 1 1)(0 1 1)

}

regression{

variables=(Easter[1] td1coef ao1987.jan ao1991.feb ao1996.feb ao2009.feb ao2010.jan ao2010.dec)

}

x11{

appendfcst=yes

trendma=13

seasonalma=(s3x3 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5 s3x5)

save=(d8 d9 d10 d11 d12 d13 d16)

}

Blocks

series{

title="Deliveries of blocks"

start=1983.jan

period=12

decimals=0

name="Blocks"

file="H:\My Documents\BIS\second project\seasonal adjustment\blocks.txt"

save=(a1)

}

transform{

function = log

}

arima{

model=(0 1 1)(0 1 1)

}

regression{

variables=(Easter[1] td1coef ao1987.jan ao1991.feb ao1996.feb ao2009.feb ao2010.jan ao2010.dec)

}

x11{

appendfcst=yes

trendma=13

seasonalma=(s3x3 s3x3 s3x3 s3x9 s3x3 s3x3 s3x3 s3x3 s3x3 s3x3 s3x3 s3x3 )

save=(d8 d9 d10 d11 d12 d13 d16)

}

Cement

series{

title="Deliveries of cement"

start=1983.jan

period=12

decimals=0

name="Cement"

file="H:\My Documents\BIS\second project\seasonal adjustment\cement.txt"

save=(a1)

modelspan=(1984.jan,)

}

transform{

function = log

}

arima{

model = (0 1 1)(0 1 1)

}

regression{

variables=(Easter[1] td1coef ao1987.jan ao1991.feb ao1996.feb ao2009.feb ao2010.jan ao2010.dec)

}

x11{

appendfcst=yes

trendma=13

seasonalma=(s3x5)

save=(d8 d9 d10 d11 d12 d13 d16)

}

Sand

series{

title="Deliveries of sand and gravel"

start=1983.1

period=4

decimals=0

name="Sand"

file="H:\My Documents\BIS\second project\seasonal adjustment\sand.txt"

save=(a1)

}

transform{

function = log

}

arima{

model=(0 1 0)(0 1 1)

}

regression{

variables = (td ao1986.1)

}

x11{

appendfcst=yes

trendma=5

seasonalma=(s3x5)

save=(d8 d9 d10 d11 d12 d13 d16)

}

Ready-mixed cement

series{

title="Deliveries of ready mixed concrete"

start=1983.1

period=4

decimals=0

name="RMC"

file="H:\My Documents\BIS\second project\seasonal adjustment\RMC.txt"

save=(a1)

}

transform{

function = log

}

arima{

model=(0 1 1)(0 1 1)

}

regression{

variables=(Easter[1] td ao2010.4)

}

x11{

appendfcst=yes

trendma=5

seasonalma=(s3x5)

save=(d8 d9 d10 d11 d12 d13 d16)

}

## Spreadsheets containing results from seasonal adjustment

In the spreadsheets provided, you can see each series and its estimated components from the seasonal adjustment process. The spreadsheets named ‘bricks’, ‘blocks’ etc contain two sheets called ‘master’ and ‘graphs’.

In column E of the master sheet is the original series. The most important series after this to look at are: column N (d11, the seasonally adjusted series), column O (d12, the trend series), column P (d13, the irregular series) and column Q (d16, the seasonal factors). You will see that column E *equals* column O *times* column P *times* column Q (this is the decomposition of the original series) and that column N *equals* column O *times* column P (this is). The difference between columns M and Q is that M is the ‘pure’ seasonal effect of each month (or quarter), while Q contains this and also the calendar effects (trading days and Easter in this case).

There are graphs of the original series, the seasonally adjusted series, the trend series and the BIS moving averages in various permutations on the graphs sheet. But, there is another graph in the master sheet (scroll to cell T70 or so) which allows you to see the ‘*SI* ratios’ (you can use the spinner buttons to see the *SI* ratios for different months/quarters, or alternatively alter the cell value in V72). These show how well the seasonal factors follow the ‘detrended’ series.

    

The quality spreadsheet contains much the same data but additionally contains calculations for the various quality measures discussed above. Column B contains the original series. Cells P4 and Q4 are not calculated in the spreadsheet: these results are taken from the X-12-ARIMA output. Column Y is taken from the X-12-ARIMA output as well and is used to provide one of the quality measures for constraining, the average percentage difference between the annual totals of the constrained and unconstrained series, in cell X4.



# Sample design work

***Summary of main points***

* Future rounds of the Sand and Gravel survey should adopt a probability proportional to size sample design for the land-won survey, using the AMRI *total excluding fill* as the size measure
* Rather than attempting to re-balance the monthly samples for the Blocks survey, the whole survey should be moved to a monthly design
  1. The proposal for the QIF project includes the requirement to address the concerns that the samples for the Sand and Gravel and Blocks Surveys have departed from the intended optimal designs over time and that this may have a bearing on both the accuracy of estimates and, possibly, the coherence of short and long-term estimates. These concerns were described in our previous methodology project [report](http://www.bis.gov.uk/assets/biscore/statistics/docs/b/bisreportfinalrev711.doc)[[16]](#footnote-16), in the section on weighting and standard errors for these surveys.

## Sand and Gravel Survey

### Current design

* 1. The Annual Minerals Raised Inquiry (AMRI) is intended to be a census of sites for producers of a wide range of non-energy minerals. The Sand and Gravel (S&G) Survey is based on a sample taken from the AMRI from two years before. So, for example, the 2010 S&G sample was drawn from the 2008 AMRI survey. The sand and gravel producers are divided into land-won, for which a sample of AMRI producers is used, and marine-won, for which all producers are included.
  2. The sample of land-won producers is described in the survey documentation as being a *cut-off* sample in which sites exceeding a regional threshold for total sales of sand and gravel (excluding *fill*) are included in the sample. In spite of this description, the 2010 S&G sample clearly included producers with low level or even zero sales recorded in AMRI. It appears that, in practice, the sample is retained over time and only topped up with any producers now exceeding the sales threshold.

Figure : S&G sales (2010) versus AMRI sales (2008)

r=0.22

r=0.62

r=0.09

r=0.75

r=0.21

r=0.66



* 1. The inclusion of low sales sites in the sample is shown in the charts in which use a linked file to plot sales of different types of product, as measured in the S&G in quarter one of 2010 against the 2008 AMRI total sales of sand and gravel (excluding fill). These products are those shown in monthly report.
  2. As well as illustrating that cases from the lower end of the AMRI sales distribution are in fact included, the charts also show the relationship between the outcomes from the two surveys. For *total sales*, for *sand for concreting* and for *gravel for concreting and other*, the relationship appears to fit a standard ratio model with points scattered about a line through the origin and increasing variance further from the origin. For the other three products the same relationship is not apparent (and the pattern is not revealed by the removal of the distinct outlier at the top of two of the charts.)
  3. To confirm that the sampled cases are not exclusively from the top end of the AMRI sales distribution, shows summary values for the AMRI measure and confirms that the distribution for the sampled and non-sampled cases has a considerable overlap. This does also show that the sampled cases have a much higher standard deviation so that the sample does, at least to some extent, oversample the higher variance part of the population which will lead to a lower sample variance compared with an equal probability sample. Note that while the S&G sample is not exclusively the largest producers, it does not include any producers with a zero AMRI total.

Table : Summary of total AMRI sales (excluding fill) for S&G sampled and non-sampled cases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N | Minimum | Mean | Maximum | Std. Deviation |
| Whether sampled for S&G | | | | | |
| No | 308 | 0 | 34,279 | 336,611 | 49,451 |
| Yes | 213 | 11,062 | 178,891 | 1,058,740 | 155,728 |

### Objectives for design

* 1. Having investigated the existing sample design, we now consider how we could re-design the sample to best meet the needs of the survey. Estimates from the Sand and Gravel Survey are shown in tables 4, 5 and 6 of the Monthly Statistics of Building Materials and Components[[17]](#footnote-17).
  2. Table 4 in the monthly bulletin shows sales of different types of sand and gravel, including breakdowns by product type, an overall total and a separate column showing how much was marine-dredged. Quarterly results are summed to give survey estimates for completed years, but since the AMRI results are reproduced in the same table, there is an obvious inconsistency between the two sets of figures[[18]](#footnote-18).
  3. One clear objective is to minimise the variance of these quarterly estimates which should also help reduce the difference between the summed quarterly survey estimates and the annual AMRI estimates.
  4. It may be desirable to replace the summed quarterly returns with the AMRI totals so that only one set of annual totals is shown. The quarterly results for previous years could be scaled to sum to the AMRI total, although this would result in a revision after each year is completed. Such a revision is already done for the monthly Blocks estimates once each quarter is complete, although the lag until the full AMRI results are available would be longer.
  5. Table 5 in the monthly bulletin shows the sales figures for the most recent quarter and the same quarter one year before, broken down by region. The same regional breakdown is given for each of the previous two annual totals, based on summed quarterly results (rather than the AMRI results.)
  6. Having a regional breakdown suggests that the sample should be stratified by region. Although this may conflict with the requirement for optimal precision on national estimates there, so some consideration could be given to this balance. Assuming the main focus is on the national estimates, we have not pursued this further.
  7. For coherence with other annual estimates and to produce more precise estimates, the regional breakdown could be based on the full AMRI data once available, although this would require a revision due to the lag in obtaining the AMRI results.
  8. Finally, table 6 breaks the most recent six quarterly estimates down by English and Welsh counties and Scottish Regions. This breakdown is into 60 areas, so the sample of 213 companies (in Q3 2010) is very sparsely spread.

### Proposal for re-design

* 1. The existing S&G data are available only for the existing design and the uncertainty around that design and its application make it difficult to use that data to evaluate potential designs. Also any such evaluation would be a prediction based only on the actual realised sample and therefore be subject to sampling error[[19]](#footnote-19).
  2. Under the ratio model, if the model is true, the optimal design is to simply select the cases with the largest AMRI values. The risk is that some groups of cases may depart systematically from the model. We note from the early scatterplots that the ratio model fits well for most products, but apparently not for two out of the six shown. Giving the whole population a chance of selection is a safety against this departure from the model and is the more standard approach for official statistics.
  3. The obvious choice is probability proportional to size (PPS) sampling – which is close to the (model-based) optimum but allocates a probability of selection throughout the AMRI frame. In choosing the best measure of size, *Cochran[[20]](#footnote-20)* (in §9A.4) makes the point that the theoretical optimum would be the current values of the survey variable, but as these are not available, then ‘if the [values of the survey variable] are reasonably stable through time, the most recently available previous values of the [survey variable] may be the best measures of size’. He also points out that in the sampling process we need to use a single size variable for all estimates, rather than a different one for different estimates. Given the strong relationship for three out of five measures in Figure 1, including the overall total, the variable *AMRI total (excluding fill)* appears to be a good candidate.
  4. To apply the method Särndal *et al* (p93, under remark 3.6.4) describe a simple method by Sunter. This achieves exact PPS except for, possibly, the smallest units. We have successfully tested this method with the AMRI in a spreadsheet.
  5. Some extra attention is required for any cases that report zero in the AMRI, as these get a zero chance of selection under PPS. We can consider whether it is better to omit these cases from the survey, which is similar to current practice) or assign a small probability of selection across all the smallest cases. In making this decision, we also need to take into account the practical issue of sending a survey form and chasing up a producer who is known to have recently reported zero production. On balance, given the potential for only a small bias in the weighted estimate, we consider it reasonable to allow these cases to have zero chance of selection.
  6. Noting that the correlation between the S&G outcome for fill and the above AMRI measure was the lowest, we also investigated using total sales including fill as the size measure. Qualitatively, the shapes of the scatter plots with this alternative size measure appear to be the same as those shown earlier in . In we see that the correlations between the S&G measures and the AMRI total are similar, once an extreme outlier on the AMRI measure has been excluded. The one relatively large improvement is, as expected for the fill measure, but even there the correlation is not raised to a level where the estimates are likely to benefit markedly. In view of this, we recommend that the size measure remains total sales, excluding fill.

Table : Correlations between S&G outcomes and potential size measures for sampling

|  |  |  |  |
| --- | --- | --- | --- |
|  | AMRI measure used for size | | |
|  | Total excluding fill | Total including fill | Total including fill, excluding outlier |
| *S&G measure* |  |  |  |
| Sand for Building | 0.22 | 0.13 | 0.21 |
| Sand for Concreting | 0.66 | 0.41 | 0.63 |
| Gravel for Concreting | 0.62 | 0.41 | 0.20 |
| Sand and Gravel for Coating | 0.21 | 0.13 | 0.20 |
| Fill | 0.09 | 0.07 | 0.12 |
| Total Sand and Gravel | 0.75 | 0.48 | 0.75 |

* 1. We have calculated the sampling probabilities associated with a PPS design, using the AMRI measure *total sales excluding fill* as the size measure. To get an idea of the impact of this proposed change in design, we can compare these probabilities of selection with those observed for the current sample. To do this we have modelled the observed selection probabilities using a logistic regression model. We used *total sales excluding fill* and its logarithm as the candidate auxiliary variables, with only the logarithm coming out as significant.
  2. The charts in Figure 2 plot the selection probabilities for sites in the AMRI sample calculated for a PPS sample and the fitted probabilities of selection from the above logistic regression model, with the PPS design plotted as the bolder line. (Note that the x-axis is incomplete on the left.) These are shown for both the 2007 and 2008 AMRI sampling frames. We also show the observed average probabilities of selection for cases grouped by the size variable to illustrate the appropriateness of the form of the model for the inclusion probabilities.
  3. The charts show that the overall pattern of inclusion probabilities is similar under the existing and proposed sampling designs, with low probabilities of selection for the smaller AMRI producers, rising to inclusion with probability one for the larger producers. In fact, the PPS design appears to include more of the large producers with probability one.
  4. Because the probabilities of selection are similar, we would expect that the sampling variability in the estimates under the existing and proposed PPS schemes would be similar, with perhaps some benefit for the PPS scheme from the definite inclusion of more of the large producers. In addition, the proposed PPS scheme uses a form of systematic random sampling from an ordered list, which should lead to some benefits in terms of a reduced variance. We cannot produce a direct estimate of the likely variance benefits because of some uncertainty over the exact derivation of the existing design, which although described here in terms of inclusion probabilities was not in fact a random design[[21]](#footnote-21).
  5. The point about the inclusion of somewhat more large producers under the PPS proposed design than under the current scheme is relevant to the practical application of the sampling scheme. The survey managers were initially wary of a random scheme with substantial change from year-to-year since they had invested substantial effort in building relationships with these producers to attempt to gain a response. Since the PPS scheme ensures more are definitely included, this ongoing relationship will be ensured, as long as these remain large producers.

Figure : Comparison of selection probabilities under current and proposed design



### Estimation issues on Sand and Gravel

* 1. As noted earlier, there doesn’t seem to be any benefit to replicating two sets of figures for annual sales of sand and gravel in Table 4 of the monthly statistics release. Annual figures for the regional breakdown in table 5 and 6 are based on the S&G survey. We might expect the AMRI estimates to be better, as they are based on a census of sites, although as we noted earlier, the AMRI data will be available later than the collated S&G results.
  2. At present, the estimation method for S&G involves using recorded sales of the same product from the AMRI as the auxiliary variable in the ratio estimator. (See our earlier report for details[[22]](#footnote-22)). This effectively means that the weights are different for each product, but since there is no multivariate analysis of these data, this does not present a problem. Also, as shows, the relationship between the S&G variables and the equivalent AMRI variable is generally stronger than with the overall AMRI total. Therefore, we recommend that this current estimation method for S&G is retained.

Table : Correlations between S&G variables and equivalent AMRI variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | AMRI variable used for size | | |
|  | Total including fill | Product specific | Product-specific excluding outliers (# outliers)\* |
| *S&G variable* |  |  |  |
| Sand for Building | 0.13 | 0.56\* | 0.80 (1) |
| Sand for Concreting | 0.41 | 0.70 | 0.70 (-) |
| Gravel for Concreting | 0.41 | 0.76 | 0.76 (-) |
| Sand and Gravel for Coating | 0.13 | 0.11 | 0.11 (-) |
| Fill | 0.07 | 0.13\* | 0.36 (2) |
| Total Sand and Gravel | 0.48 | 0.48\* | 0.75 (1) |

\* Three correlations are particularly affected by outliers and increase markedly when these are omitted

## Blocks Survey

### Current Design

* 1. Producers of blocks are divided into fixed monthly and quarterly samples with, currently, 48 and 45 producers respectively. The quarterly estimates are based on the merged monthly and quarterly sample returns and so are based on a census of producers reporting production for all three months in the quarter. The monthly estimates are weighted using the ratio of the total for the whole sample to the monthly sample, aggregated for the previous rolling year, as described in our previous report. The monthly estimates are revised once the quarterly estimates are available.
  2. Evidence from our earlier work showed that the allocation to the monthly and quarterly samples is not balanced with respect to production of each type of block or with respect to region. In particular, producers of aerated blocks are all in the monthly sample. This lack of balance is partly countered by the weighting, but will add unnecessarily to the month-on-month variation and to the extent of the revision on completion of the quarter.
  3. A further, although perhaps less significant point is that the quarterly results are collated from data collected in two different ways: monthly and quarterly. Also the quarterly returns cannot be split into monthly data to retrospectively look at more stable monthly patterns.

### Proposals for re-design

* 1. Our initial intention was to propose a move to a design in which the monthly sample was a better-balanced sample from the population, using stratified sampling. However a better solution appears to be to move the whole survey to a monthly design, so that all producers get the same monthly request for data.
  2. We have obtained a cost for this option, which relates to the dispatch and chasing of extra forms and the entry and checking of returned data. The total extra cost for this (small) sample of approximately £1.4k is small enough not to be a determinant of the design decision.
  3. Although this proposal perhaps appears to be a more radical change, any rebalancing of the monthly sample would require swapping producers between the quarterly and monthly sample, so the actual gross change may be quite similar.
  4. This option also appears to be available under a wholly compulsory survey, although it is difficult to get a definitive answer on the legal aspects as the Statistics of Trade Act is not explicit about which surveys are covered.
  5. We must carefully consider the possibility of any extra burden placed on survey responders for those moving from a quarterly to monthly response. If the existing process of quarterly reporting consists of aggregating existing monthly figures, then the change to monthly reporting should lead to little extra effort. However, we recommend checking this in advance with responders to ensure that there is no unforeseen extra burden imposed on businesses from this change.

# Imputing for new sampled elements

***Summary of main points***

* For the land-won sand and gravel inquiry BIS needs to deal with the problem of newly sampled businesses which do not respond to their first request for information
* We have considered a number of mean imputation methods, but all are coherent with the imputation methods we have recommended in previous work for BIS
* We recommend imputing a trimmed class mean, where the trimming is symmetric (5% of responders from each end) and classes employ a regional breakdown as long as there are at least ten responders
  1. In the previous work[[23]](#footnote-23) the ONS Methodology Advisory Service has done on imputation in the building material statistics, the focus has been on imputing for non-responders who have responded in the past. The recommended method that has been accepted and will be adopted is to obtain a (trimmed) mean growth rate for each imputation class and apply it to the response in the previous period.
  2. This method will not work for units that are being sampled for the first time since there is no ‘previous’ figure to apply any growth rate to. In practice this issue is most pertinent for the land-won sand and gravel inquiry. The marine-dredged sand and gravel inquiry is a statutory survey of all producers with high response rates. For the bricks and blocks inquiries, ONS currently sends dummy forms to new panel members for a period or two to establish the firms are valid and producing useable responses. Only once they have sent in one or two responses will ONS actually add them to the panel, so there will be actual previous responses to use for imputation. In any case these will be a very small number of new firms in any one reporting period that will need to have responses imputed.
  3. The land-won sand and gravel is based on a sample of producers and there are likely to be several ‘new’ firms every quarter[[24]](#footnote-24), and with a significant level of non-response the need is clear for a separate imputation method for non-responding new firms. So our investigation focuses on the data for this particular inquiry. The method can be used for the marine inquiry as well, although we have not used data from this inquiry to assess the various methods.

## Methods considered

* 1. We considered using auxiliary information from the IDBR frame to set up linear regression models, but these were not successful. The auxiliary variables considered were turnover, employment and full-time equivalent employment, with the response variable being the reported values for sales of the various types of sand and gravel. The R-squared value for each model was never more than 10% (indicating poor fits) and usually more like 1%. We had similarly poor fits when briefly investigating the slate, bricks and blocks data.
  2. The other obvious strategy to consider was setting up imputation classes for the firms that did respond in a period and imputing some sort of mean class value for the missing values, using the reported levels rather than growth rates. Imputation classes are already used for the current imputation methodology and it is sensible to use the same classes for this method. The classes are a cross-classification of firm size and region. Where a class has more than five reported values in a period that class is used to impute a value; if the class is too small then a coarser classification is used which ignores the regional breakdown.
  3. The main issue with imputing mean values (whether or not trimmed) is that they artificially reduce the amount of variation in the final dataset. Since the microdata are not released by BIS and no estimation is made of variation (whether of parameters or of estimates), there is more reason to accept this imputation strategy, which is closely aligned to the method being introduced following our previous work for BIS. Imputation methods such as random hot-deck imputation or random regression imputation would be fundamentally different from the imputation about to be implemented for ‘old’ firms.
  4. In the imputation for ‘old’ firms, the set of growth rates in a class is trimmed before a simple mean is taken. In the imputation for ‘new’ firms we do not have growth rates so the set of values to take an average from is less highly skewed. This suggests that there is no need to perform an asymmetric trimming method. We focussed on symmetric trimming (e.g. removing 5% of the values from the top and bottom of the distribution) rather than asymmetric (e.g. remove 5% from the bottom and 10% from the top). We checked briefly that this was sensible by looking at one variable and confirming that symmetric trimming worked better than asymmetric trimming and no trimming. We report results from looking at three different symmetric trims: 5% from each end of the distribution, 10% and 20%.
  5. There is also a possibility of adjusting the rule for when to move from the fine imputation classes to the coarse ones. Instead of using the wider classes when n is less than or equal to 5, we can try other values. We investigated using 10 and 20 instead of 5. The effect of having a higher value is to make less use of the regional breakdown.
  6. So we report the results of nine different methods: trimming 5, 10 and 20 per cent and regional classes of at least 5, 10 and 20 respondents[[25]](#footnote-25).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method numbers | | *Condition to use regional imputation classes* | | |
| *n*>5 | *n*>10 | *n*>20 |
| *Amount to trim off each end of the distribution* | 5% | 1 | 2 | 3 |
| 10% | 4 | 5 | 6 |
| 20% | 7 | 8 | 9 |

## Results

* 1. We applied these methods to each of the seven variables for the land-won sand and gravel inquiry (sand for coating, gravel for aggregates etc). The total sales figure for each firm was then calculated by summing the seven values after imputation. This ensures that the totals are consistent, but this could lead to sub-optimal results for the total. So for each method we also considered the effect of imputing the totals directly, and then adjusting the seven originally imputed values so that they sum to the directly imputed total. (We call this the constrained method[[26]](#footnote-26).) These constrained methods are numbered 1a, 2a *etc*.
  2. In order to assess these methods we have used the microdata between 1997 and 2010 for the land-won sand and gravel inquiry. We have found each instance of a ‘new’ firm, treated it as a non-responder and imputed values for the variables according to each of the 18 methods. We compared these imputed values with the actual values in the data and calculate the effect of the imputation by summing across firms for each quarter and arriving at a relative imputation bias for each variable and each quarter. There were 300 instances of new firms in the 51 quarters (1997 quarter 1 to 2010 quarter 2); although for 28 quarters there were no new firms[[27]](#footnote-27).
  3. So, if the sampled units are indexed by *i*, *yi* is the value of the variable (whether production or delivery) and is the imputed value of *yi* if *i* is a non-responder (and equal to *yi* if *i* is a responder) then. The RIB measures the effect of the imputation on the final estimates and we present it as a percentage. To summarise the effect over several periods we report the mean of the *absolute* values of the RIB (so that RIBs of opposite sign do not offset each other). We call this quantity mRIB. (See Fallows and Brown (2007)[[28]](#footnote-28) for more details).
  4. The table below gives the mean relative imputation bias (mRIB) for four variables (sand for building, sand for concreting, gravel for aggregates and fill), the (indirect) total of all seven variables and a weighted mRIB which was used to rank the 18 methods. (The reason that not all seven variables have mRIBs given is that for three of these variables, there was one quarter with sales of zero reported, making the RIB undefined. There are ways we could adjust these so they can still contribute to the overall mRIB, but these three variables are all small contributors and would have very small weight anyway.) The weights are assigned according to the totals observed for each variable, as well as the total, for all the 23 quarters in which new firms were observed. So the indirect total’s mRIB has weight 0.5.
  5. The best performing method is 5 (minimum 10 respondents, trim 10%), but it is very close to 2 (minimum 10 respondents, trim 5%). Since trimming smaller amounts makes use of more of the reported data we think that method 2 is the one to use.
  6. Comparing method *x* with method *x*a lets us see the impact of imputing the total indirectly instead of directly. We see that the impact is low, but perhaps surprisingly usually the unconstrained method works better than the constrained method.
  7. We did a quick check against the current method used by ONS when imputing for new firms in this inquiry (take the latest reported figure from the AMRI and divide it by 4). Because we could only get a few years of AMRI data we could not make a like-for-like comparison but we tried applying this to one quarter of data, using all firms and not just new firms, and found much larger mRIBs for the variables. This gives us confidence that trimming and imputing a class mean, according to either method 2 or 5, will be better than the current method and we recommend that BIS chooses to implement one of these methods. BIS may choose to weight the various mRIBs differently to arrive at a different choice of the 18 methods, but it is unlikely that any sensible weighting will result in a choice beyond methods 1-6.
  8. Therefore, we recommend imputing a trimmed class mean, where the trimming is symmetric (5% of responders from each end) and classes employ a regional breakdown as long as there are at least ten responders.

Table : Mean relative imputation bias (mRIB) for each method for each component and for total and weighted mean over all

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Deriving the total as the sum of the components*** | | | |  |  |  |  |  |  |  |
|  |  | Method |  |  |  |  |  |  |  |  |
|  | *weight* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Sand for Building | *0.032* | 11.4% | 11.6% | 12.0% | 11.6% | 11.4% | 11.7% | 11.8% | 12.1% | 12.5% |
| Sand for Concreting | *0.190* | 3.0% | 2.8% | 3.2% | 3.0% | 2.9% | 3.2% | 3.1% | 2.9% | 3.2% |
| Gravel for concreting | *0.270* | 1.8% | 1.8% | 1.8% | 1.9% | 1.8% | 1.8% | 1.9% | 1.9% | 1.9% |
| Fill | *0.008* | 29.2% | 29.5% | 27.0% | 26.7% | 27.2% | 23.2% | 34.0% | 34.5% | 33.7% |
| Total (indirect) | *0.500* | 2.6% | 2.5% | 2.6% | 2.6% | 2.5% | 2.6% | 2.8% | 2.8% | 2.9% |
| Weighted mRIB |  | 2.93% | 2.88% | 2.98% | 2.96% | 2.86% | 2.94% | 3.12% | 3.11% | 3.24% |
|  |  |  |  |  |  |  |  |  |  |  |
| ***Imputing the total directly*** | |  |  |  |  |  |  |  |  |  |
|  |  | Method |  |  |  |  |  |  |  |  |
|  | *weight* | 1a | 2a | 3a | 4a | 5a | 6a | 7a | 8a | 9a |
| Sand for Building | *0.032* | 12.2% | 11.8% | 12.1% | 12.3% | 11.7% | 12.0% | 11.9% | 11.6% | 12.3% |
| Sand for Concreting | *0.190* | 3.5% | 3.5% | 4.0% | 3.4% | 3.3% | 3.7% | 3.8% | 3.9% | 4.4% |
| Gravel for concreting | *0.270* | 2.1% | 2.0% | 1.9% | 2.1% | 2.0% | 1.9% | 2.2% | 2.2% | 2.1% |
| Fill | *0.008* | 25.6% | 25.0% | 21.3% | 26.5% | 24.7% | 23.0% | 25.5% | 26.7% | 23.8% |
| Total (direct) | *0.500* | 2.8% | 2.7% | 2.8% | 2.8% | 2.7% | 2.8% | 2.8% | 2.7% | 2.8% |
| Weighted mRIB |  | 3.23% | 3.14% | 3.22% | 3.21% | 3.07% | 3.17% | 3.32% | 3.28% | 3.37% |

1. See: <http://www.bis.gov.uk/analysis/statistics/construction-statistics/building-materials> [↑](#footnote-ref-1)
2. There is the extra problem that a 12 month symmetric moving average would not be centred on a month – it would be centred between the sixth and seventh months. This is usually dealt with by taking a 2x12 moving average – an average of two 12 month moving averages, which is equivalent to a single weighted 13 month average – but this average, although centred on a month, still could not be calculated for the final six months of the series. [↑](#footnote-ref-2)
3. For some series an additive decomposition is appropriate, where Y=C+S+I; in this case the seasonal effects are in the same units as Y, and we expect them to vary about zero. [↑](#footnote-ref-3)
4. See: <http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/time-series-analysis/index.html> [↑](#footnote-ref-4)
5. The issue of revisions is discussed in paragraph 1.64. [↑](#footnote-ref-5)
6. We compared a 9 term Henderson filter with the 13 term filter. Both are symmetrical moving averages which are effective on series that have had large seasonal effects already removed, so they are applied in the later iterations of the X-12 process. [↑](#footnote-ref-6)
7. It remains clear to us that whether we choose the additive or multiplicative decomposition this represents a clear improvement to the current use of a simple moving average to smooth the series. [↑](#footnote-ref-7)
8. We say the ‘same model’ was used, but the parameters are estimated separately for each series and so are different specific airline models in each case. [↑](#footnote-ref-8)
9. We looked at the Easter[8] and Easter[15] variables but these did not work as well as Easter[1]. [↑](#footnote-ref-9)
10. The use of trading day regressors in these quarterly series should be reviewed the next time BIS reviews the methodology for these statistics. [↑](#footnote-ref-10)
11. <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html> [↑](#footnote-ref-11)
12. <http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/time-series-analysis/guide-to-seasonal-adjustment.pdf> [↑](#footnote-ref-12)
13. For quarterly series, this would be quarter-on-quarter change; similarly for the explanations of the other quality measures in this section. [↑](#footnote-ref-13)
14. <http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/time-series-analysis/guide-to-seasonal-adjustment.pdf> [↑](#footnote-ref-14)
15. Distortion is worse where there are trading day effects because total trading day effects over a year vary from year to year and are not expected to sum to zero, even approximately (unlike with seasonal factors). So constraining in such series is rather unnatural. [↑](#footnote-ref-15)
16. <http://www.bis.gov.uk/assets/biscore/statistics/docs/b/bisreportfinalrev711.doc> [↑](#footnote-ref-16)
17. See: <http://www.bis.gov.uk/analysis/statistics/construction-statistics/building-materials> [↑](#footnote-ref-17)
18. The footnote to the table says that the “two inquiries differ because some respondents are only able to provide estimated information for the quarterly inquiry”. This may be true, but the more substantial difference is that the S&G is based on a sample rather than the whole population. [↑](#footnote-ref-18)
19. It would be possible to model outcomes based on linked AMRI data from two years apart, but given the lack of an existing comparison we did not pursue this. [↑](#footnote-ref-19)
20. Cochran (1977) *Sampling Techniques*, (3rd edition), Wiley [↑](#footnote-ref-20)
21. Predicting the variance for the PPS design is complex, requiring either a simulation study or a calculation of all joint inclusion probabilities. [↑](#footnote-ref-21)
22. <http://www.bis.gov.uk/assets/biscore/statistics/docs/b/bisreportfinalrev711.doc> [↑](#footnote-ref-22)
23. See: <http://www.bis.gov.uk/assets/biscore/statistics/docs/b/bisreportfinalrev711.doc> [↑](#footnote-ref-23)
24. ‘New’ here means sampled in time *t* but not in time *t­*–1. Such a firm may well have been sampled several times over the years. [↑](#footnote-ref-24)
25. In fact later on we introduce a difference between ‘constrained’ and ‘unconstrained’ methods, so in total there are 18 different methods considered eventually. [↑](#footnote-ref-25)
26. The adjustment is done so that the constrained values for each firm are in proportion to the unconstrained values. The ratio of the directly imputed total to the indirectly imputed total is the scaling factor that is applied to the unconstrained values. [↑](#footnote-ref-26)
27. New firms tend to be observed in the first quarter of the year, when ONS refreshes its sample. [↑](#footnote-ref-27)
28. <http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/survey-methodology-bulletin-61.pdf> [↑](#footnote-ref-28)