Dealing with High Frequency Time Series: Seasonal Adjustment of Road Traffic Data

Atanaska Nikolova and Duncan Elliott;
Office for National Statistics, UK

Abstract

Time series in official statistics are typically of monthly, quarterly, or annual frequency. In recent years there is an increasing interest in higher frequency data, which is timelier and can be used to complement traditional lower frequencies and serve as a faster indicator for changes in the economic climate. This research compares different methods for seasonal adjustment of road traffic data from Highways England, which is available at hourly frequency, but is currently aggregated to monthly for the purpose of regular seasonal adjustment. Some of the methods used include fractional airline decomposition and state space modelling, using newly developed capabilities of the software JDemetra+. The estimated seasonal and calendar effects are compared between different aggregations of the data, investigating how information from higher frequencies (e.g. daily and weekly effects) can be used to inform movements in the monthly equivalent.

Key Words: high frequency; mixed frequency; seasonal adjustment; time series;

Background

Time series can come at many different frequencies, but in the context of official statistics in the UK, the main frequencies are annual, quarterly, and monthly. With data science techniques popularity on the rise, as well as methods such as web scraping, administrative data sources have become more accessible. This provides many opportunities to enhance and inform current statistics, for example with the use of leading indicators. Fast economic indicator series can be valuable in helping to anticipate and predict unexpected changes in the economy, particularly in the face of uncertainty.

One issue with novel data sources is that they may come in a format or frequency that is not typically dealt with in official statistics, meaning that the established methods for analysis and processing may not be fit for purpose. The purpose of this research was to explore a high frequency time series data set which has recently been adopted by the ONS as a faster indicator of economic activity. The data source is from Highways England, and represents traffic flow (counts) from major roads and English motorways measured on a 15-minute interval for different types of vehicles based on length. The different traffic variables available are listed in Table 1. Average speed is also available, but the current focus is on flow as it is considered more informative of economic activity. For example, decreasing numbers of heavy goods vehicles may signify decrease in trade and thus be used as a contributor to the indication of economic health.
Table 1: Variables in original data set

<table>
<thead>
<tr>
<th>Speed: all vehicles</th>
<th>Flow: all vehicles</th>
<th>Flow: &lt;5.2m</th>
<th>Flow: 5.2m-6.6m</th>
<th>Flow: 6.6m-11.66m</th>
<th>Flow: &gt;11.66m</th>
<th>NoB date/time</th>
</tr>
</thead>
</table>

In the current publication of the Office for National Statistics (2019), the data is aggregated up to monthly frequency and seasonally adjusted to feature in the “Road traffic commentary” section. It is important to note that this output is not classed as official statistics and is still under development. The main motivation behind this research is to explore methods for seasonal adjustment of the road traffic flow data at a higher than monthly frequency, which can potentially render the output timelier and with higher granularity. Current official guidelines on seasonal adjustment methods and software are not tailored to deal with higher than monthly frequency (ESS, 2015), so this paper focused on experimental techniques which are still under development. These are early release functions that will make part of the officially recommended by Eurostat software for seasonal adjustment, JDemetra+ version 3. The functions are open source and available from GitHub from NBB Research and Development branch (nbbrd/jd3-rtests).

Data

For the purpose of this research, the data set was aggregated up to daily frequency and the main variable used was traffic flow for all vehicle sizes. Due to many missing values after 2014 as well as large outliers due to sensor failures, the span for the analysis was restricted between 01/01/2006 to 31/12/2014 (Figure 1). Following aggregation into daily, there was one missing data point for 13/03/2011 (due to all hourly observations missing for that day), leading to a missing data point for March 2011 in the monthly data.

Figure 1 Traffic flow section for all vehicles, England. Daily (top) and monthly (bottom) frequency.
There were other single instances of missing values in the hourly data (e.g. a single 15-minute interval missing within a day), but these cases were still averaged into a daily aggregation in order to preserve a continuous data span. The data at both monthly and daily frequency clearly displays cyclical variations indicating seasonality. The autoregressive spectrum of the daily flow suggests strong intra-weekly variations, as peaks can be observed at frequencies of 52, 104, 156 (indicated by green vertical lines on Figure 2), signifying periodic phenomena occurring once, twice and three times per week.

![Figure 2: Autoregressive spectrum of daily flow. Green lines indicate activity peaks once, twice and three times a week.](image)

Software and methods

R was used to run all analyses. Monthly aggregates were seasonally adjusted using RJDemetra package, while daily aggregates were seasonally adjusted using functions from jd3-rtests. The three main functions used were fractional airline decomposition, X11, and STL. Testing for residual seasonality was performed by also using jd3-rtests functionalities, namely the F-test and QS test functions.

Analyses

- Fractional Airline Decomposition

Model-based approach similar to SEATS decomposition (Signal Extraction for ARIMA Time Series) where a \((0 1 1)(0 1 1)\) ARIMA model is adapted to fractional frequencies using Taylor expansion, rather than integer frequencies. The benefit of this approach is that missing data is allowed. The seasonal adjustment is done is
stages for the different periodicities within the daily data, starting with 7, then 30.44, and finally 365.25. With each step, the seasonally adjusted output from the first frequency is fed as an input to the seasonal adjustment for the next one. The resulting seasonal adjustment after the full process can be seen on Figure 3.

Figure 3: Seasonal adjustment of daily traffic flow with Fractional airline model. Green line represents the seasonally adjusted estimates (SA), red line is the trend.

Although the seasonal adjustment process has removed some of the periodic variations, it can be clearly seen that there still are consistent peaks towards the end of each annual period.

- **X11**

This is a non-parametric approach based on iterative moving average filters to smooth the seasonal fluctuations, and Henderson filters to extract the trend. The technique is the same as the traditional X-11 algorithm applied for monthly and quarterly time series, but adapted for non-integer periodicities. The principle is the same as with the fractional airline decomposition, whereby the seasonal adjustment is done in stages with sequential removal of the three periodicities within the daily data, starting with 7 and trend filter length 9, then 30.44 and trend length 65, and finally 365.25 and trend length 367. The final result is illustrated on Figure 4. One problem with this approach is that it does not support missing values, so the one
missing data point was imputed by averaging the two neighbouring values on either side.

![Figure 4: Seasonal adjustment of daily traffic flow using X11. Green line represents the seasonally adjusted estimates (SA), red line is the trend.](image)

As with the fractional airline model approach, the seasonally adjusted values appear volatile. The residual peaks in December are even larger, while the values for the rest of the year appear smaller in amplitude than for the fractional airline model (apart from various extreme values throughout).

- **STL**

The Seasonal and Trend Decomposition using Loess (STL) method is similar to X11, but is based on local weighted regression, rather than moving average and it also allows for missing values. Same iterative process was followed as with the other two methods. For periodicity 7, the trend window was set to 75 and the seasonal window to 15, all other parameters were left as default. The resulting seasonal adjustment is displayed on Figure 5.

Again, there are consistent large peaks in December present in the seasonally adjusted values, while other major variations appear smoothed out, and the trend component is also very smooth.
Figure 5: Seasonal adjustment of daily traffic flow with STL.

Residual seasonality tests

Visual inspection of the three versions of seasonally adjusted data suggests that there is residual seasonality that has not been cleared out. QS and F-test were performed (functions available from GitHub jd3-rtests) on the final seasonally adjusted data. The QS test checks for autocorrelations at seasonal lags, while F-test is computed on the coefficients of dummy seasonal variables. Results are summarised in Table 2.

Table 2: Residual seasonality tests on daily traffic flow (test value and corresponding p-value). Significant figures suggest residual seasonality and are marked in red.

<table>
<thead>
<tr>
<th></th>
<th>Fractional airline</th>
<th>X11</th>
<th>STL</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTest</td>
<td>0.01</td>
<td>0.52</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>p = 1</td>
<td>p = 1</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>QSTest</td>
<td>1.41</td>
<td>6.18</td>
<td>143.75</td>
</tr>
<tr>
<td></td>
<td>p = 0.5</td>
<td>p = 0.05</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>
It appears that even though on visual inspection all methods seem to have residual seasonality left, the fractional airline decomposition displays the least statistical evidence, although it is marginal for the QS statistic. One of the most obvious issues is the large drops of traffic flow around end of December and beginning of January, which appear problematic for the seasonal adjustment estimation. It is expected that traffic flow would be low at holidays such as Christmas and New Year’s Eve, and perhaps higher than usual just before and after. Figure 6 marks the dates of Christmas and Easter Sunday in the daily data, so the associated activity around these days can be clearly seen. It should be noted that these data were not treated for any calendar effects prior to seasonal adjustment, which can explain the problematic residual seasonality.

Figure 6: Christmas (X, green line) and Easter Sunday (E, red line) position on non-seasonally adjusted daily flow data.

Seasonal adjustment of monthly aggregates

Seasonal adjustment of daily data can be a complex process where careful modelling of the different frequencies is required, in order to minimise volatility and seasonal leakage in the seasonally adjusted estimates. Given that there are no established and official methods for this purpose, it can be a challenging task for national statistical institutes. Official outputs need to be based on well established and tested techniques. Therefore, an alternative way of using information from high frequency data is to extract useful information in order to perform prior adjustments for monthly (or quarterly) seasonal adjustment, using well established techniques.

The following analyses were performed on the monthly aggregate of the average daily traffic flow values, using RJDemetra. Firstly, seasonal adjustment was performed using automatic detection of parameters, after that some parameters were altered based on characteristics of the daily equivalent of the data. Finally,
seasonal adjustment with Easter correction was performed using estimated effects from the daily data. All analyses were done with the X13 method.

Automatic detection of parameters suggested an Airline model \((0 \ 1 \ 1)(0 \ 1 \ 1)\) and no transformation, trading day effect, Easter[1] correction, as well as a range of outliers (see Table 3 for details of RegARIMA effects).

**Table 3: RegARIMA effects from automatic parameter identification with RJDemetra X13**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>-0.45</td>
<td>-0.74</td>
</tr>
<tr>
<td>Tuesday</td>
<td>0.91</td>
<td>1.46</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>Thursday</td>
<td>1.49</td>
<td>2.30</td>
</tr>
<tr>
<td>Friday</td>
<td>1.39</td>
<td>2.22</td>
</tr>
<tr>
<td>Saturday</td>
<td>-1.43</td>
<td>-2.32</td>
</tr>
<tr>
<td>Easter[1]</td>
<td>-5.67</td>
<td>-4.41</td>
</tr>
<tr>
<td>AO (12-2010)</td>
<td>-18.35</td>
<td>-6.63</td>
</tr>
<tr>
<td>AO (1-2010)</td>
<td>-18.96</td>
<td>-6.77</td>
</tr>
<tr>
<td>AO (1-2013)</td>
<td>-16.70</td>
<td>-5.98</td>
</tr>
<tr>
<td>AO (2-2009)</td>
<td>-14.64</td>
<td>-5.20</td>
</tr>
<tr>
<td>AO (1-2009)</td>
<td>-11.15</td>
<td>-3.79</td>
</tr>
</tbody>
</table>

Although the automatic identification procedure suggested a trading day effect, a possibility was entertained that a working day contrast may be more appropriate. Estimating the average week-day flow from the daily data shows a sharp drop of traffic counts on Saturdays and Sundays relative to other days, even though there is also an upward curve from Monday through to Friday (Figure 7). In addition, inspection of the Easter build-up from the daily data suggests that there is an Easter-related activity more than one day before Easter Sunday. Figure 8 plots together the mean values in March/April for Thursdays to Saturdays prior to Easter Sunday, Easter Sunday and Easter Mondays (indicated with letter E) next to the mean values for these days when not within an Easter week. Based on the effects in the daily
data, it may be suggested that there is an Easter effect up to Thursday before Easter continuing to Easter Monday.

Seasonal adjustment on monthly data was performed again, this time with a working day contrast and an Easter regression variable extending 3 days before and 1 day after Easter. All other parameters were kept constant. The overlay between the automatic seasonal adjustment and the edited spec can be seen on figure 9.

\[ \text{Figure 7: Mean daily flow per day of week (1 = Monday, 7 = Sunday)} \]
Figure 8: Easter influence on day-of-week, averaged values for March/April in daily data

Figure 9: Seasonal adjustment overlay with original automatic spec (green) and new edited spec (red) with working day and longer Easter influence

The difference between the two seasonal adjustment is not too drastic. Although the effects from the edited specs were significant, the diagnostics in terms of ACF and PACF of the residuals were slightly better with the original automatic spec. Figure 10 suggests that there is some auto-correlation at lag 6. It should be noted that even if a longer Easter effect may be suited for the data, in the analysed span (2006-2014)
there is only one instance of Easter influence potentially crossing between March and April, in 2013.

Finally, an attempt was made to estimate the effect of Easter by fitting an ARIMA model to the daily data, creating a regression variable for Thursday before to Monday after Easter. Other variables were also included, to account for Christmas day and Christmas eve, New Year’s Day and New years’ Eve. Dummy variables were also coded for the months and days of week, in order to be able to estimate the Easter effect without any confounding factors. The resulting coefficients of the Easter influence days were aggregated up to monthly and used to create a bespoke regressor for the monthly flow aggregate. In addition, trading day correction was introduced again, rather than working day. The resulting seasonal adjustment is illustrated on Figure 11 in red, against the original automatic seasonal adjustment (in green).
Seasonal adjustment of traffic flow monthly data with automatic parameter selection (green line) and edited specs based on estimated Easter effects from daily data (red line)

The difference with the automatic specification is not too obvious. The diagnostics are comparable, so one seasonal adjustment is not necessarily preferable over the other. Although information from the daily data can be useful to gain insight, the automatic selection procedures appear to do a good job, at least in the context of monthly seasonal adjustment.

Conclusions and Future directions

Dealing with high frequency time series can be challenging and needs careful consideration if the outputs are to be used for official publications. There are no established methods or official guidance for seasonal adjustment of higher than monthly frequency, which means that an average user may struggle without the appropriate expertise in the area. Daily data can be very useful, informative and attractive due to its timely property, making it suitable for a fast indicator into the state of the economy. However, it should be interpreted with caution as it can also be very volatile and prone to anomalies.

In terms of the present analysis, it can be concluded that to extract useful meaning from the daily data the series need to be prior adjusted accordingly for calendar effects and outliers. Traffic data can be very sensitive to weather conditions and special one-off events. Therefore, the modelling requires good knowledge of the data and its characteristics. Once pre-cleaning is achieved, other useful methods for testing the quality of the seasonal adjustment can include stability and revisions analysis. Forecast errors can also be useful to assess the quality of the models. Future methods that can be tested on these data include State Space modelling, Wavelets, and SigEx R package (McElroy and Livsey, 2019).
References


McElroy, T., Livsey, J., 2019, Sigex R package, GitHub: https://github.com/jlivsey/sigex

NBB, 2019, jd3-rtests, GitHub: https://github.com/nbbrd/jd3-rtests