STATISTICAL ARTICLE
ERTHYGL YSTADEGOL

Seasonal Adjustment and Road Casualty Data

Introduction

This Statistical Article reports the results of carrying out an experimental seasonal adjustment of Welsh road accident and casualty data. It uses the standard Government Statistical Service approach for carrying out seasonal adjustment of time series data. The statistical analysis underlying this article was first carried out by the Welsh Government and the results were then repeated and reviewed by the Methodology Directorate of the Office for National Statistics (ONS). This is the first time that road casualty data have been seasonally adjusted in this way within the United Kingdom.

As a result of this work, the Transport Statistics section within the Welsh Government are considering whether or not to publish seasonally adjusted quarterly road casualty data alongside the current unadjusted, that is as reported, quarterly data; and using these seasonally adjusted data to replace the current practise of estimating current trends in road safety in Wales by showing simple moving averages of the most recent twelve months of road casualties data.

We seek your views about this decision; and we welcome any other comments you have on the analysis presented in this article. If you have any comments or require further information please contact us on the details below:

Tel: 029 2082 5413
E-mail: stats.transport@wales.gsi.gov.uk

Table of Contents

Date of Publication: 10 April 2012
Next update: Not a regular publication

Authors: Scott Dewis and Mark Hogan, Knowledge and Analytical Services
E-mail: stats.transport@wales.gsi.gov.uk
Telephone: 029 2082 5413
Twitter: www.twitter.com/statisticswales | www.twitter.com/ystadegaucymru

What is seasonal adjustment?

Time series data often display 'seasonality'; that is they have an observable component consisting of a fairly common shape repeated every twelve months.

This pattern is called the seasonal affect; in order to understand the patterns in a time series we sometimes want to remove this seasonal component in order to get a better understanding of period to period movements of the time series data. For this reason, statistical offices across the world use a variety of methods called 'seasonal adjustment' to remove this seasonal affect. Seasonal adjustment is usually, though not exclusively, applied to economic data.

The standard Government Statistical Service approach for carrying out seasonal adjustment of time series data is through a computer program called X-12-ARIMA, originally developed by Statistics Canada. This approach is based on either of two alternative underlying models of the data. These models are that the time series is either:

- Additive model : $Y = C + S + I$
- Multiplicative model: $Y = C \times S \times I$

Where Y is the original series, C is the trend-cycle, S is the seasonal component, and I is the irregular component. The box below gives the rationale behind this approach and it is taken from the ONS's "Guide to Seasonal Adjustment with X-12-ARIMA". (ONS 2007)

Components of a time series

Time series can be thought of as combinations of three broad and distinctly different types of behaviour, each representing the impact of certain types of real world events on the data. These three components are: systematic calendar-related effects, irregular fluctuations, and trend behaviour.

Systematic calendar related effects comprise seasonal effects and calendar effects. Seasonal effects are cyclical patterns that may evolve as the result of changes associated with the seasons. They may be caused by various factors, such as:

- Weather patterns: for example, the increase in energy consumption with the onset of winter;
- Administrative measures: for example, the start and end dates of the school year;
- Social / cultural / religious events: for example, retail sales increasing in the run up to Christmas;
- Variation in the length of months and quarters due to the nature of the calendar.

Other calendar effects relate to factors which do not necessarily occur in the same month (or quarter) each year. They include:

- Trading day effects which are caused by months having differing numbers of each day of the week from year to year: for example, spending in hardware stores is likely to be higher in a month with five, rather than four, weekends;
- Moving holidays, which may fall in different months from year to year: for example Easter, which can occur in either March or April.

Taken together these effects make up the seasonal component.

Irregular fluctuations may occur due to a combination of unpredictable or unexpected factors, such as: sampling error, non-sampling error, unseasonable weather, natural disasters, or strikes. While every member of the population is affected by general economic or social conditions, each is affected somewhat differently, so there will always be some degree of random variation in a time series. The contribution of the irregular fluctuations will generally change in direction and/or magnitude from period to period. This is in marked contrast with the regular behaviour of the seasonal effects.

The trend (or trend cycle) represents the underlying behaviour and direction of the series. It captures the long-term behaviour of the series as well as the various medium-term business cycles.

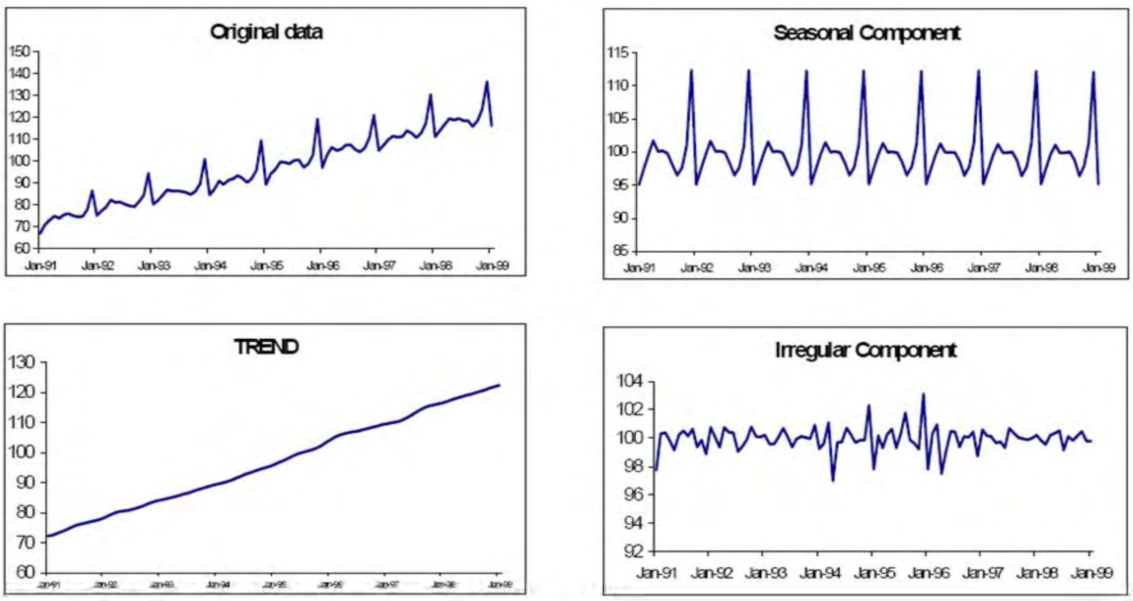
The seasonally adjusted series is calculated by a process of decomposition of the original time series into these three components, and then removing the seasonal component. So that the seasonally adjusted series is made up from both the trend-cycle and the irregular components of the original time series:

- For the additive model : Seasonally adjusted series =$Y - S = C + I$
- For the multiplicative model: Seasonally adjusted series = $Y / S = C \times I$.

In a multiplicative decomposition, the seasonal effects change proportionately with the trend. If the trend rises, so do the seasonal effects, while if the trend moves downward the seasonal effects diminish too. In an additive decomposition the seasonal effects remain broadly constant, no matter which direction the trend is moving in. In practice, most time series exhibit a multiplicative relationship and the multiplicative decomposition usually provides the best fit (ONS 2007).

A seasonally adjusted series is the result of 'removing information' from the original time series by removing the seasonal component. The argument for doing this is that it will improve our understanding of the data because the seasonal component usually contains information that is easily explained, and its dominance in the period-to-period movements in the original time series may mask important non-seasonal information. For example December retail sales are usually higher then November, but is this because of Christmas (seasonal effect) or a growing economy (trend effect)? The seasonal adjustment process removes the seasonal component to leave the trend and the irregular components and these are usually of more interest. For example the trend could show a turning point in GDP (recession or boom), while the irregular could show the affect of a natural disaster or social unrest on GDP.

To help visualise the time series and its various components, the charts below show a hypothetical original time series together with the process of decomposition into its component series of the trend-cycle, the seasonal and the irregular components. The purpose of seasonal adjustment is to estimate and remove the seasonal component from the original data to give a seasonally adjusted estimate. By removing the seasonal component, it is easier to focus on other components.



Source: Office for National Statistics

How is seasonal adjustment performed?

Seasonal adjustment is carried out through a computer program called X-12-ARIMA. It was developed by United States Bureau of the Census (1998) as an extended and improved version of the Statistics Canada X-11-ARIMA method. The program broadly runs through the following steps:

- The series is modified by any user defined *prior adjustments*;
- The program fits a *regression-ARIMA model* to the series in order to: detect and adjust for outliers and other distorting effects, to improve forecasts and seasonal adjustment; detect and estimates additional components such as calendar effects; and extrapolate forwards (forecast) and backwards (backcast) an extra one to three years of data;
- The program then uses a series of *moving averages* to decompose a time series into the three components. It does this in three iterations, getting successively better estimates of the three components. During these iterations *extreme values* are identified and replaced;
- A wide range of *diagnostic statistics* are produced, describing the final seasonal adjustment, and giving pointers to possible improvements which could be made.

The method is also used by many of the leading national statistical institutes across the world including the ONS who use it for economic time series data such as the unemployment count, the Retail Prices Index (RPI), Balance of Payments, and Gross Domestic Product (GDP).

Why use seasonal adjustment with respect to road casualties?

Transport is intrinsically related to government economic policy (and associated strategies) as it is to the daily economic experiences of the Welsh public. Therefore (based on the strength of this relationship) a discourse could be formulated that if economic series are to be seasonally adjusted due to factors such as the calendar, timing decisions, weather and expectation so too could aspects of transport statistics such as road casualties. Furthermore the effects of various holidays and periods would be estimated by seasonal adjustment and would be of interest to policy experts and stakeholders.

It is not a difficult concept to understand how for example weather might impact on road casualties, especially those associated with certain vehicle types.

How was seasonal adjustment carried out with respect to road casualties?

Road casualties time series data could be seasonally adjusted as whole; and/or the killed and seriously injured (KSI) component could be adjusted separately from the slight casualties component. In addition, the time series could be adjusted as a whole (called <u>direct</u> seasonal adjustment), or it could be split into a number of component series and each of these adjusted separately before adding the seasonally adjusted components back together to get an adjusted total (called <u>indirect</u> seasonal adjustment). For the road casualties data there is a further choice in that the total could be split into components in a variety of ways, for example by road user type (pedestrian, motorcyclist, etc), by age and sex, by region of Wales, by type of accident and so on. For road casualty data we could, also, in principle adjust either monthly or quarterly data (as we have the exact date for every accident and the resulting casualties).

Inspection of the road casualties data suggested that the criteria for deciding on the parameters for this preliminary seasonal adjustment should be based on:

- Getting reasonably high numbers of casualties for the data series that were to be adjusted. This is so that the seasonal element would not be dominated by the irregular component of the time series.
- Take an indirect approach to seasonal adjustment, as prior inspection of the data series showed that some of the components exhibit very different seasonal patterns from other component series.

This, in turn, determined the initial choices about the series that were going to be adjusted. The choices were:

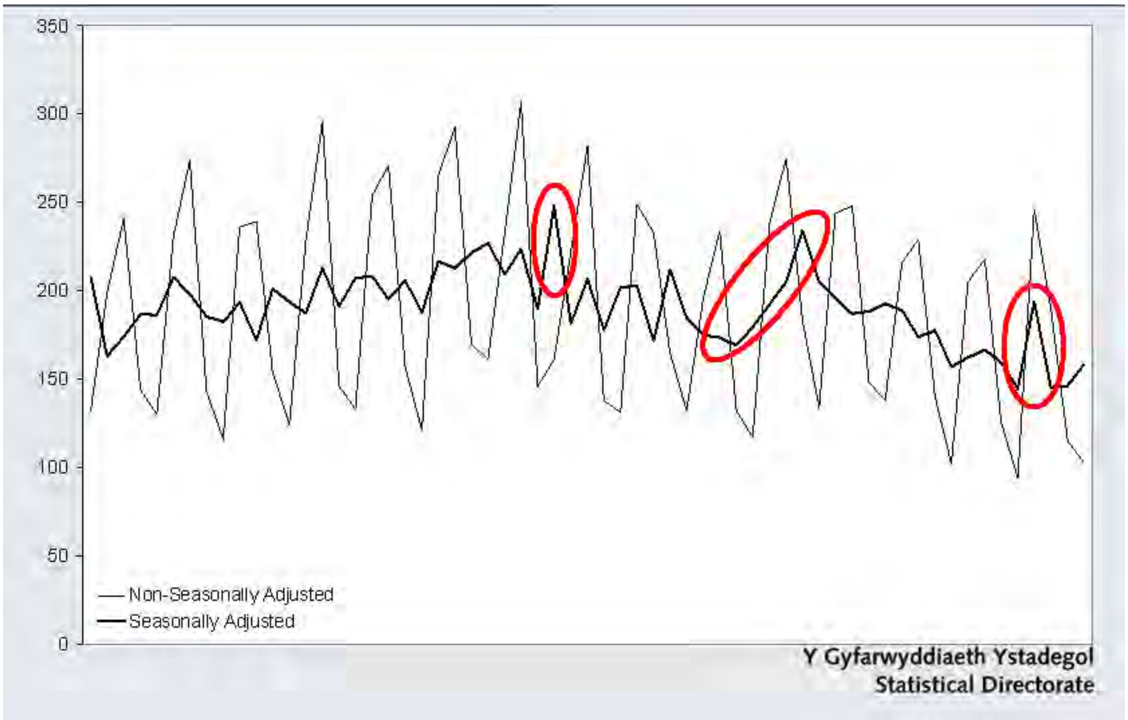
- To adjust quarterly data, not monthly data.
- To adjust total casualties, and not split this between the KSI and the slight casualties components. As the slight casualties component is much larger than the KSI component, then this dominates movement in the total casualties series.
- To adjust by component series, where the component is for 'casualties by road user type'.

The component series for road user type are:

- Pedestrian casualties: Here casualties tend to peak in Q4 (the September to December quarter) and to a lesser extent in Q2; with troughs in casualties tending to occur in Q1 and Q3. Working age adult casualties are linked with evening entertainment (they peak at 11pm and 2am at weekends) and this more prevalent in the run up to Christmas. The trough at the beginning of the year is linked to poor weather keeping people at home, and the trough in Q3 is linked with the school summer holiday (child pedestrian casualties tend to peak after school).
- Pedal cyclist and motorcyclist casualties: In both cases casualties peak in Q2 and Q3 (with Q3 tending to be a little higher). This is clearly a weather effect as people tend to cycle more in spring and summer when the weather is warmer.
- Car, Taxi and Minibus Users Road Casualties: Here the pattern tends to be for a trough in casualties in Q1, with casualties then rising through the year to peak in Q4. It is difficult to identify the factors behind this pattern (and this understanding is not, in fact, needed for the seasonal adjustment process).
- Other vehicles: The seasonal pattern is harder to identify, perhaps because the number of casualties is fewer and the range of vehicles included, so the irregular element tends to dominate the time series. However a pattern can be identified that shows that casualties tend to be lower in Q1 and higher in Q3 and Q4 each year.

Appendix 1 on page 7 below shows more precisely how the adjustment was carried out. The strength of this approach is, however, demonstrated by the chart below. It shows the non-seasonally adjusted motorcycle casualty data compared with the seasonally adjusted motorcycle casualty data. The areas highlighted are interesting. The first one shows that even though the actual figure reduced, the seasonally adjusted series shows that after considering the seasonal affect, the number of casualties was actually higher than would have been expected for the time of year. The second highlighted area shows an increasing trend over 4 quarters which was hidden by the seasonal variation. The third highlighted area shows that during the last peak, even allowing for the seasonal affect the figure still increased. Also from this highlighted area you can see the rise due to the trend or irregular (thick black line) and the rise due to the seasonal (gap between thick black line and thin black line).

Motorcycle Casualty Casualties



Quality issues

The quarterly figures can be revised throughout the year as police and local authorities in Wales provide late amendments. So the quarterly breakdown data for the current year would be provisional at the point of publication and be subject to validation change alongside an expectation of a significant increase in the number of casualties as the police conduct and conclude any investigations. Similarly, the figures for earlier years may differ from those previously published. The figures cover only road accidents reported to the police and involving personal injury.

The (x12arima) specification used for each seasonally adjusted series would be subject to an annual review. This would entail an in depth investigative process examining the spec's parameters. An expert review carried out by ONS (appendix 1) has been carried out for the road casualty time series. Quarterly sense checks would be put in place to support this annual review. These would be more basic in nature but would highlight any anomalies in the data and any problems that may arise from them.

References

Clive W. J. Granger, Collected Papers of Clive W. J. Granger, Volume 1, Spectral Analysis, Seasonality, Nonlinearity, Methodology, and Forecasting, 2001.

ONS Methodology and Statistical Development: Guide to Seasonal Adjustment with X-12-ARIMA, 2007. http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/time-series-analysis/index.html

Clive W. J. Granger Seasonality: Causation, Interpretation, and implications, 1979.

Appendix 1: Seasonal Adjustment Review of the Welsh Government Road Casualties Time Series

Prepared for the Transport Statistics section of the Welsh Government by the Time Series Analysis Branch, Methodology Directorate, Office for National Statistics

Date: January 2012

Authors:

Dr Gary Brown
Head of Time Series Analysis Branch
Email: gary.brown@ons.gsi.gov.uk
Telephone: 01633 455649

Daniel Ayoubkhani
Assistant Methodologist, Time Series Analysis Branch
Email: daniel.ayoubkhani@ons.gsi.gov.uk
Telephone: 01633 456391

1. Introduction

In December 2011, Gary Brown and Daniel Ayoubkhani (Time Series Analysis Branch, Office for National Statistics) were contracted by Mark Hogan (Labour Market Statistics, Welsh Government) to seasonally adjust six road casualty time series produced by the Welsh Government. This report details the findings of that seasonal adjustment review.

The quarterly time series to be reviewed consist of flow data from 1996Q1 to 2011Q1:

- pedestrian road casualties
- pedal cyclist road casualties
- motorcycle users road casualties
- car, taxi and minibus users road casualties
- other vehicles road casualties
- total road casualties (derived as the sum of the preceding five series)

2. Methods

The latest version of the X-12-ARIMA software released by the US Census Bureau is build 192. However, the software used in the review was X-12-ARIMA build 188 for consistency with Welsh Government production practices.

For each of the six series, a “default” specification (spec) file was run and automatic selections were then validated, modified (where required) and finally fixed. The contents of the default spec file can be found in Annex A. The following specifications were reviewed for each series as part of this SAR:

- the decomposition of the series
- the fitted ARIMA model
- prior adjustments for unusual features in the data and Easter effects
- the seasonal moving average length(s)
- the trend moving average length
- other specifications as necessary

3. Results

Table 1 below summarises final seasonal adjustment specifications for each of the six road casualty time series, as determined during the SAR.

Table 1

Table 1: Summary of seasonal adjustment specification for Welsh road casualty data

Time Series	Decomposition	ARIMA Model	Regression Model	Seasonal Filter(s)	Trend Filter
Pedestrian	Multiplicative	(0,1,1)(0,1,1)	Easter[1], SB1999.1, AO1999.4, LS2004.3, AO2006.1, AO2009.4, AO2010.1	3x3	7-term
Pedal Cyclist	Multiplicative	(0,1,1)(0,1,1)	Easter[15], LS2006.1, AO2007.3, AO2008.3	3x9	7-term
Motorcycle Users	Multiplicative	(0,1,1)(0,1,1)	Easter[1], AO2003.1, AO2006.4, AO2010.2	3x9	7-term
Car, Taxi and Minibus Users	Multiplicative	(0,1,1)(0,1,1)	None	3x9	7-term
Other Vehicles	Multiplicative	(0,1,1)(0,1,1)	AO1998.2, LS1999.1, AO2002.4, SB2004.1, SB2006.3	3x9	7-term
Total	Multiplicative	(0,1,1)(0,1,1)	None	3x9	7-term

4. Detailed Analysis

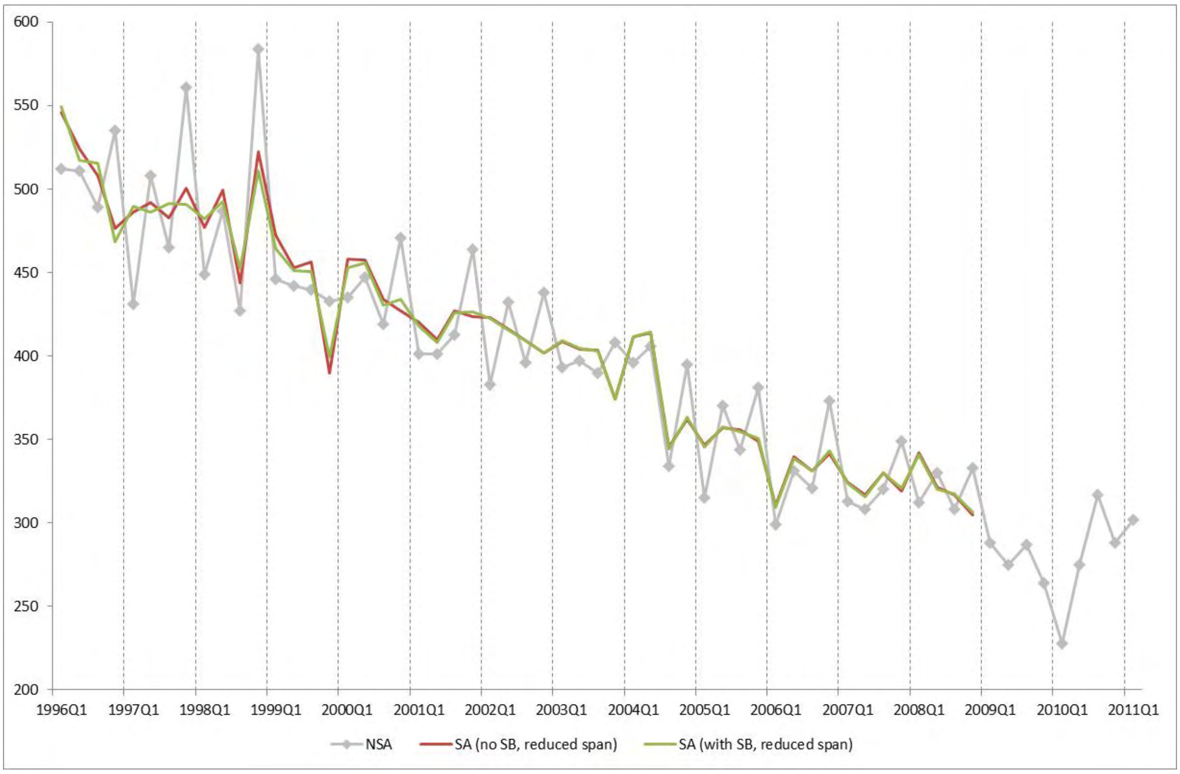
4.1 Pedestrian Road Casualties

Visual inspection of the unadjusted time series showed a possible seasonal break in 1999Q1 and erratic behaviour at the current end of the series. Hence, as a first step the default spec file was only run on the stable part of the series - 1999Q1 to 2008Q4.

An Easter effect with a build-up period of one day ("Easter[1]") and a level shift in 2003Q4 ("LS2003.4") were automatically identified by the default spec file. The pickmdl procedure selected the (0,1,1)(0,1,1) ("Airline[1]") ARIMA model, but only as a default model after all five candidate models failed the selection tests. The more sophisticated automdl procedure was subsequently run and a (0,0,0)(1,0,0) ARIMA model with a constant term was identified. Comparison of results from the two alternative ARIMA models showed no tangible difference.

A number of additive outliers (AOs) were identified via visual inspection of the plotted unadjusted (normally denoted "non-seasonally adjusted" - NSA) and seasonally adjusted (SA) data, and hence the regression model was changed to (Easter[1], AO1999.4, AO2003.4, LS2004.3, AO2006.1). The pickmdl procedure was then re-run and the Airline model was again selected, but this time as the preferred model, rather than as the default model. Comparison with (0,0,0)(1,0,0) again showed no differences. As the Airline model is very versatile, and fits the majority of time series well, it is preferred here.

Figure 1: Initial adjustment of pedestrian road casualties

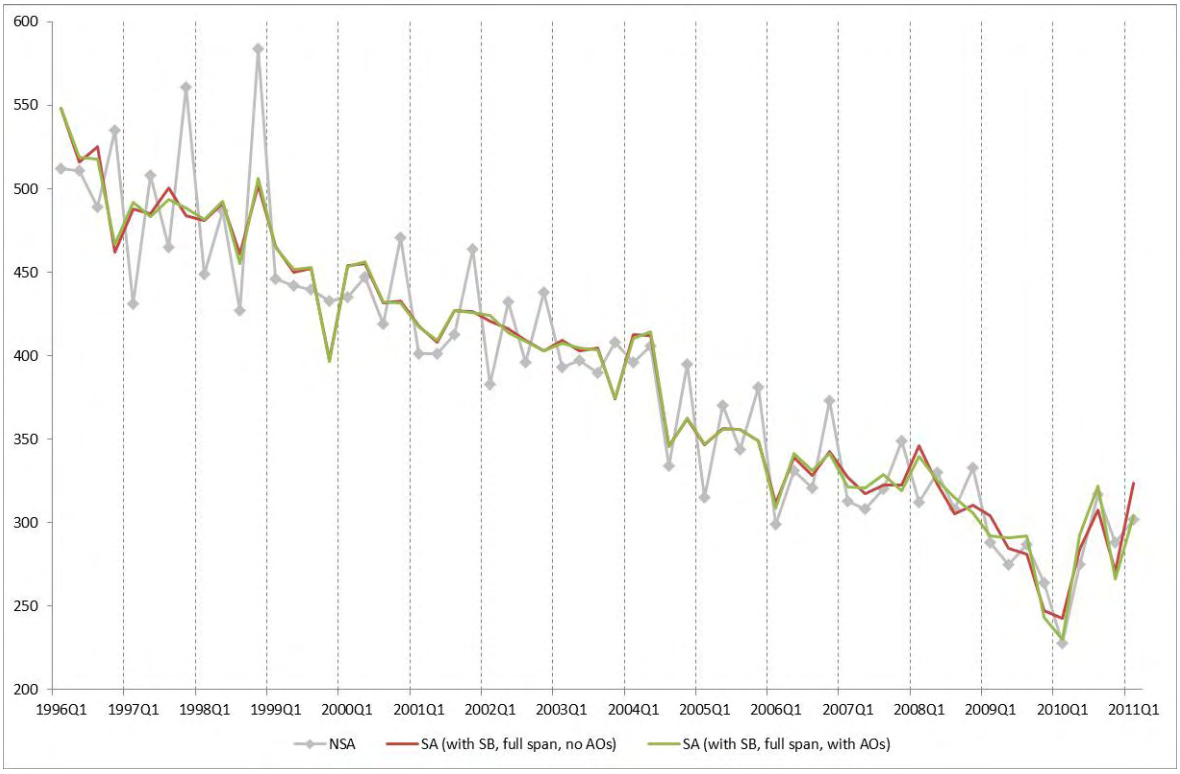


[1] Known as the airline model after being chosen to fit airline data in the book "Time Series Analysis: Forecasting and Control" by Box and Jenkins (1976)

The span was now increased to include the beginning of the series, and the seasonal break in 1999Q1 was tested. The regressor “SB1999.1” was found to be statistically significant (p=0.01), especially for Q3 and Q4 (AO2003.4 was removed from the model due to insignificance once SB1999.1 was added). The SA series with and without prior adjustments for the SB, covering the reduced span, are illustrated in figure 1 Above. Prior adjusting for the SB makes a significant difference to the quality of the SA series, especially over the span 1997 to 1999.

After SB1999.1 was fixed in the model, the current end of the series was included in the span and the automatic outlier identification procedure was turned back on, resulting in AO2009.4 and AO2010.1 being included in the model. Ideally, minimal intervention should be favoured at the current end of the series due to noise at the tails and uncertainty over the future direction of the series. However, not prior adjusting for AO2009.4 and AO2010.1 affects quality before 1999Q1 (ie pre-SB) and results in unrealistic growth at the current end of the series, as illustrated in figure 2.

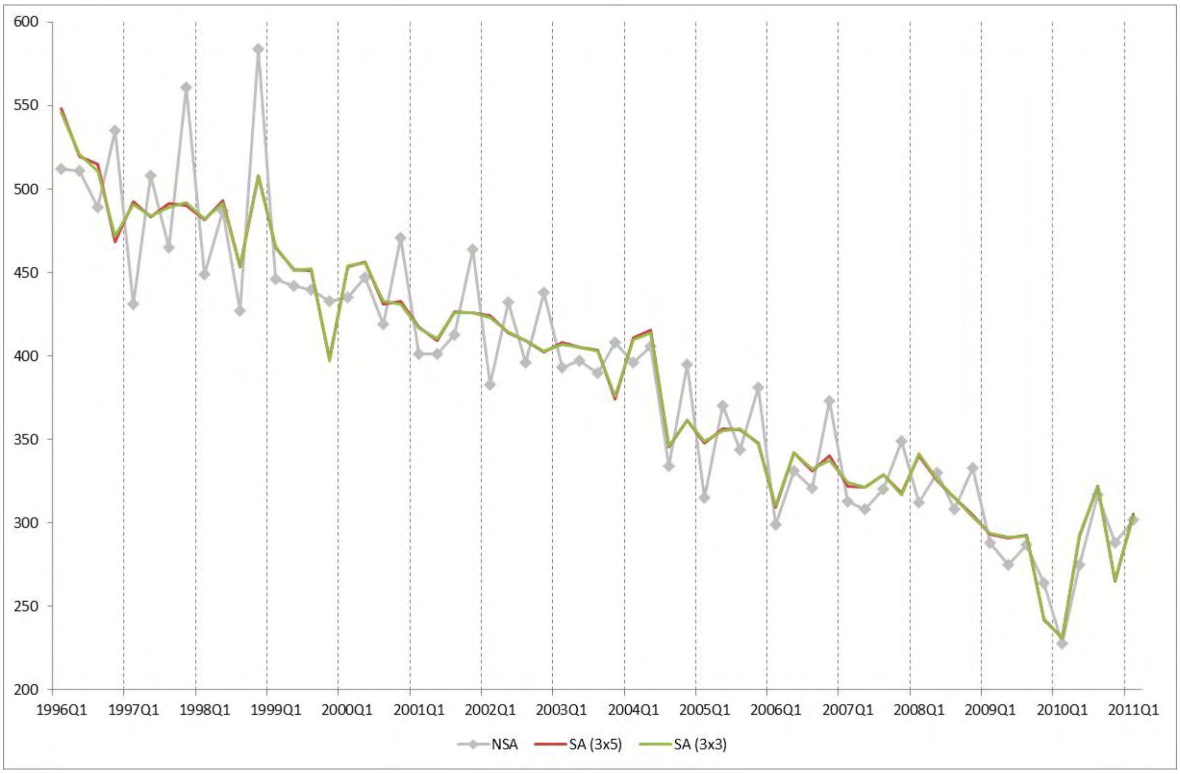
Figure 2: Second stage adjustment of pedestrian road casualties



In order to assess the robustness of these two late AO adjustments, the end date of the analysed span was reduced sequentially, one quarter at a time, back to 2010Q1 with the automatic outlier identification procedure turned on. Both AO2009.4 and AO2010.1 were consistently identified (albeit sometimes alongside other AOs), supporting the inclusion of these regressors in the final model. Interventions in place at the current end of the series should be reviewed in the near future, as evidence to support fitting different or additional regressors may emerge.

Finally, a 3x3 seasonal moving average (SMA) was used for all four quarters in order to produce the final SA series, rather than using the 3x5 SMA selected by the default spec file on the basis of the global irregular-to-seasonal (I/S) ratio. This choice of a relatively short SMA better captures the uncertainty at the current end of the series. The practical impact on the SA series is minimal as is illustrated in figure 3 below, where 'SA (3x3)' represents the final SA series for 'pedestrian road casualties'. For consistency with the other road casualty series, a 3x9 SMA was also tested – but this performed significantly worse than either the 3x3 or 3x5 and so could not be used.

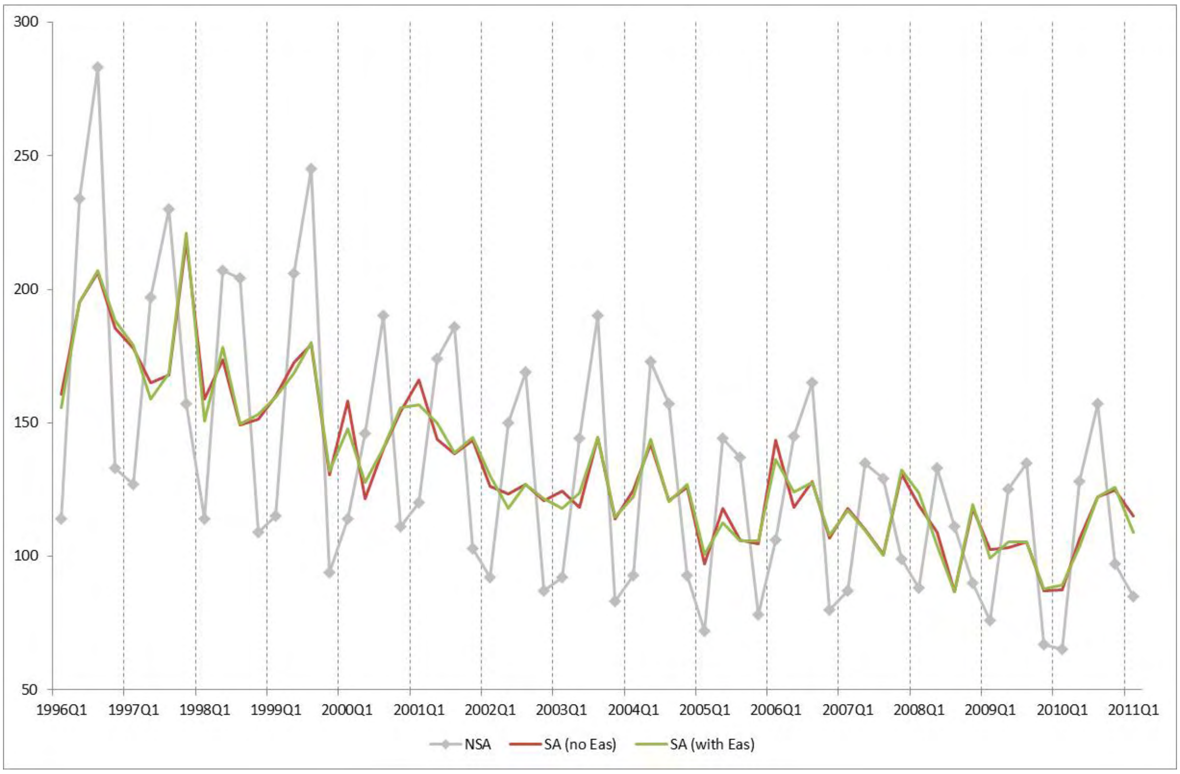
Figure 3: Final adjustment of pedestrian road casualties



4.2 Pedal Cyclist Road Casualties

An Easter[15] effect was identified by the default spec file but the regressor was found to be insignificant (t=-1.73). However, prior adjusting for the Easter effect does make a difference to the SA series, as illustrated in figure 4 below. Two other road casualty series also included an Easter adjustment – but of immediate effect (Easter[1]). Hence, Easter[1] was also tested for this series – for consistency – but performed significantly worse than Easter[15]. An Easter[15] regressor has therefore been included in the model.

Figure 4: Adjustment of pedal cyclist casualties



According to X-12-ARIMA's 'sliding spans' and 'revisions history' diagnostics, the SA from the default run is unstable, meaning that movements in the SA series will change in magnitude and direction whenever new observations are added to the current end of the series. This is indicated by:

- 16% of seasonal factors being flagged as unstable
- 48% of quarter-to-quarter changes in the SA series being flagged as unstable
- an average absolute percentage revision, from first to final SA estimates, over the last five years of the series of 2.7%

The most unstable part of the SA series was determined as being the span 2006 to 2009. After inspection of quarter-on-quarter growth rates in the NSA series over this span and the plotted NSA and SA series, LS2006.1, AO2007.3 and AO2008.3 were added to the regression model. The SA is more stable as a result of these changes, indicated by:

- 2% of seasonal factors being flagged as unstable
- 19% of quarter-to-quarter changes in the SA series being flagged as unstable
- an average absolute percentage revision, from first to final SA estimates, over the last five years of the series of 1.6%

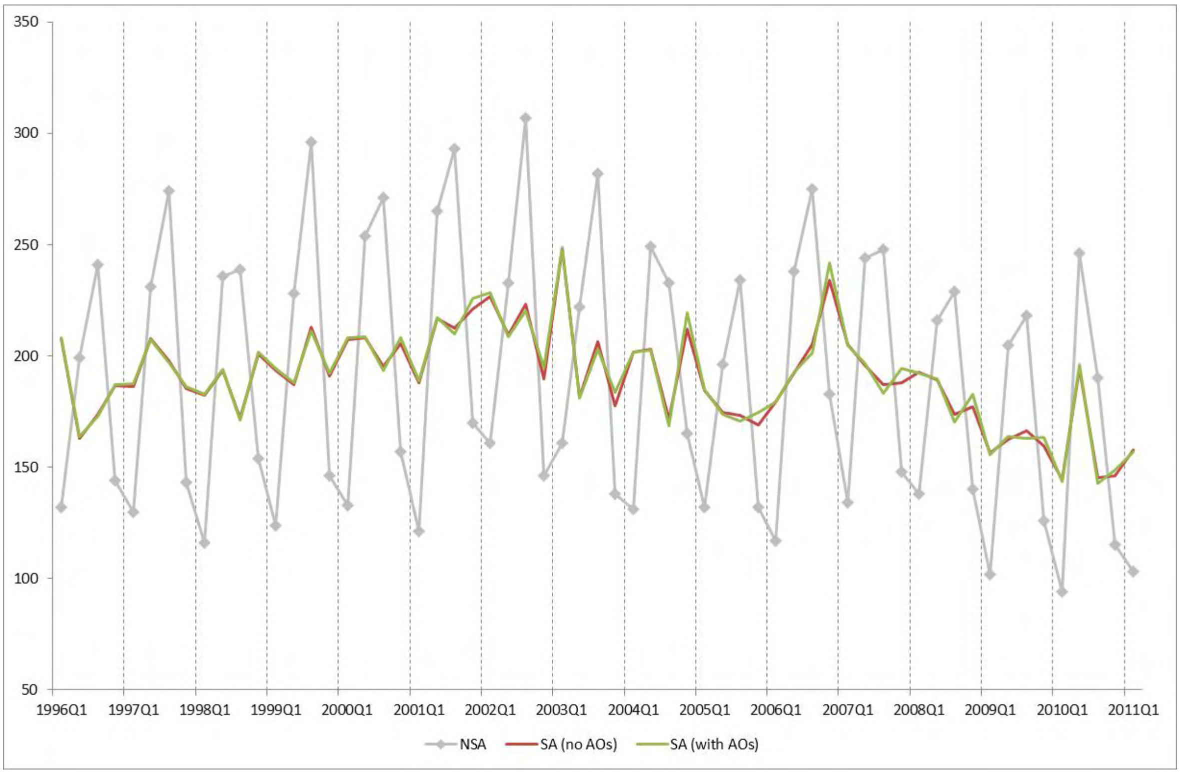
In addition, the “sigma limits” used to down-weight extreme values in the irregular component were investigated – increasing them from their default values of 1.5 (for the lower limit) and 2.5 (for the upper limit) to 1.8 and 2.8. However, this change had minimal impact, so was not applied.

4.3 Motorcycle Users Road Casualties

No regression effects were identified during the default run. However, after inspection of quarter-on-quarter growth rates in the NSA series and the plotted NSA and SA series, AO2003.1, AO2006.4 and AO2010.2 were added to the regression model. The decision to prior adjust for these effects was made because:

- all three regressors were found to be statistically significant
- prior adjustment makes a difference to the SA series, confirmed by visual inspection of the plotted data (see figure 5 below, where 'SA (with AOs)' represents the final SA series for 'motorcycle users road casualties')
- in addition, an overall view of the diagnostics available in X-12-ARIMA (none of which should be trusted in isolation, but which as a group give a guide to quality) shows improvement:
    - o the average absolute percentage revision, from first to final SA estimates, over the last five years of the series falls from 0.9% to 0.8% (indicating that prior adjustment results in a more stable SA series)
    - o the percentage of unstable quarter-to-quarter changes in the SA series falls from 25% to 23% (there are no unstable seasonal factors for either adjustment)
    - o the average absolute percentage error of within-sample forecasts over the last three years of the series falls from 10% to 7% (indicating that forecasts appended to the fitted model are likely to be more accurate when AO regressors are included)
    - o the value of Akaike's Information Criterion (the F-adjusted variant, AICC) associated with the fitted model falls from 502 to 486 (indicating that the increase in model complexity is more than offset by the improvement in model fit)

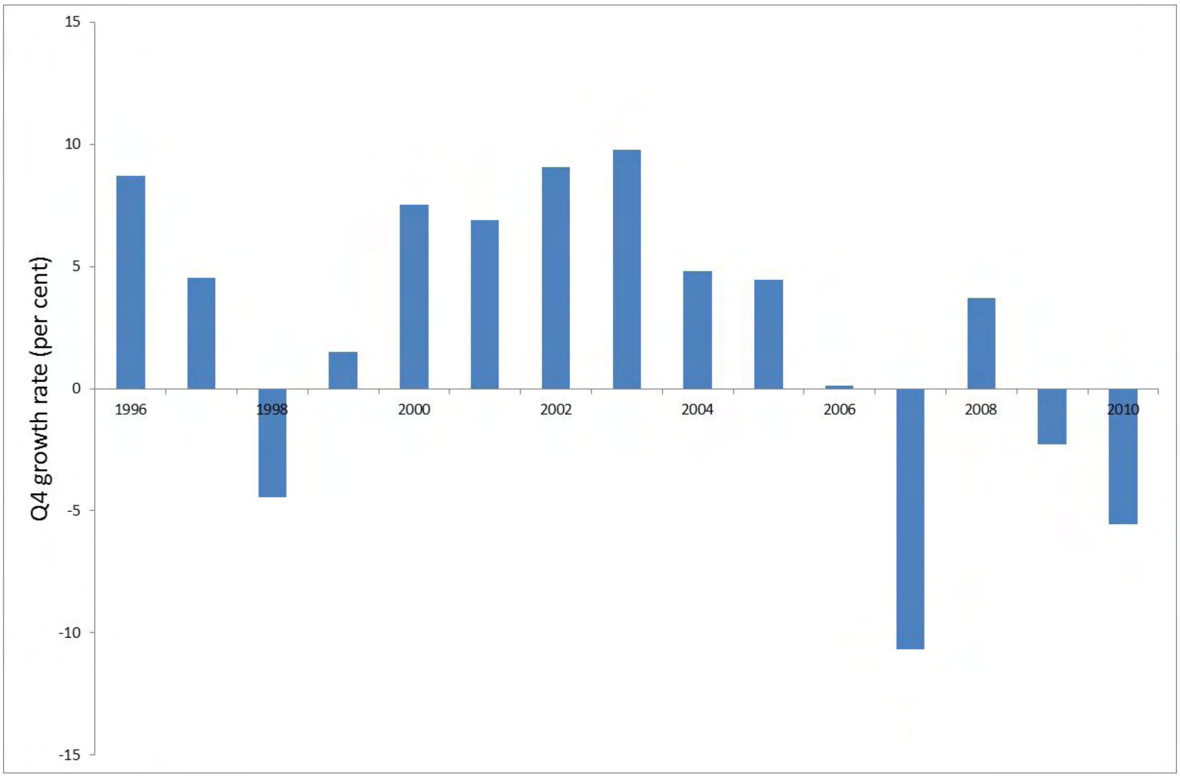
Figure 5: Adjustment of motorcycle user casualties



4.4 Car, Taxi and Minibus Users Road Casualties

No regression effects were identified during the default run. However, visual inspection of the plotted NSA series reveals a possible SB in 2006Q4, from which point onwards the seasonal peaks in the series move from Q4 to Q3. The possible SB can also be seen in the plotted quarter-on-quarter growth rates in the NSA series for Q4, which are illustrated in figure 6.

Figure 6: Growth rates for Q4 compared with Q3 for not seasonally adjusted car users casualties



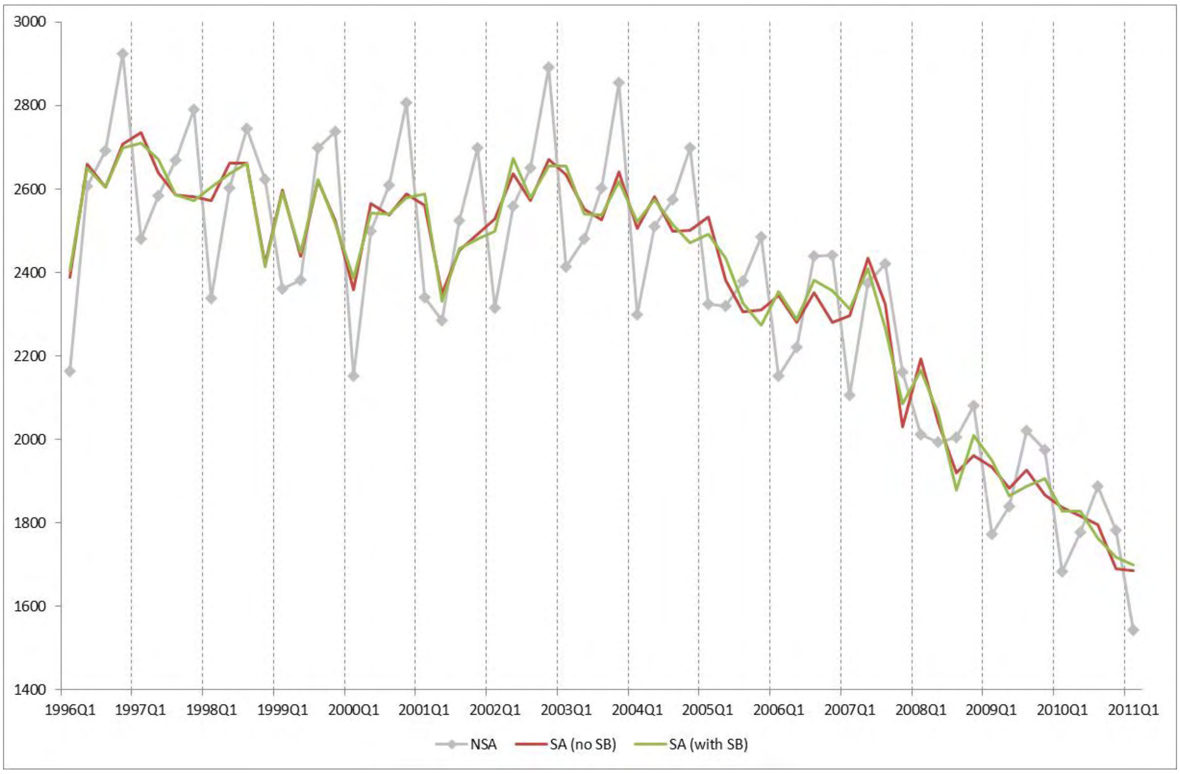
When SB2006.4 was tested for, the overall effect was found to be insignificant (although the individual regressors for Q3 and Q4 were found to be marginally significant). Although there are clear differences between the SA series with and without prior adjustments for SB2006.4 (see Figure 7), it is not obvious which adjustment is better. The SA series with prior adjustments is likely to be more stable according to the 'sliding spans' and 'revisions history' diagnostics, but the SA series without prior adjustments is likely to be more interpretable according to X-12-ARIMA's monitoring and quality assessment statistics. There is no evidence of residual seasonality in either of the SA series.

The final spec file does not specify prior adjustments for SB2006.4. This decision was taken because:

- there is a lack of evidence of the presence of the SB (for example, the effect cannot be easily seen in the SI (de-trended) series split by quarter)
- the SB is not statistically significant
- there is a lack of evidence that prior adjusting for the SB will result in a better quality SA
- not including SB regressors in the model results in a more parsimonious specification

The series 'SA (no SB)' in figure 7 represents the final SA series for 'car, taxi and minibus users road casualties'. The decision not to prior adjust for SB2006.4 should be reviewed in the near future, when more data at the current end of the series become available.

Figure 7: Adjustment of car, taxi and minibus users casualties

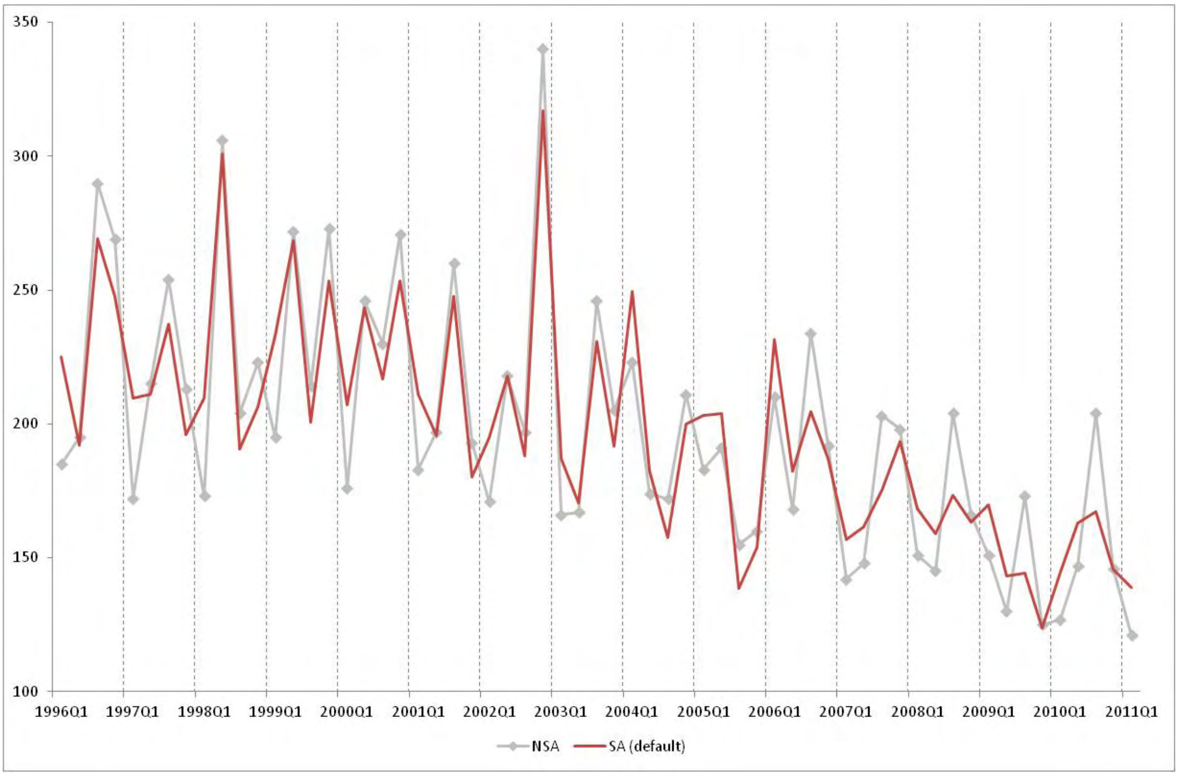


4.5 Other Vehicles Road Casualties

The default run chose a log transformation, the Airline model (selected as the default model) and no prior adjustments. However, this simplistic specification results in a poor SA because:

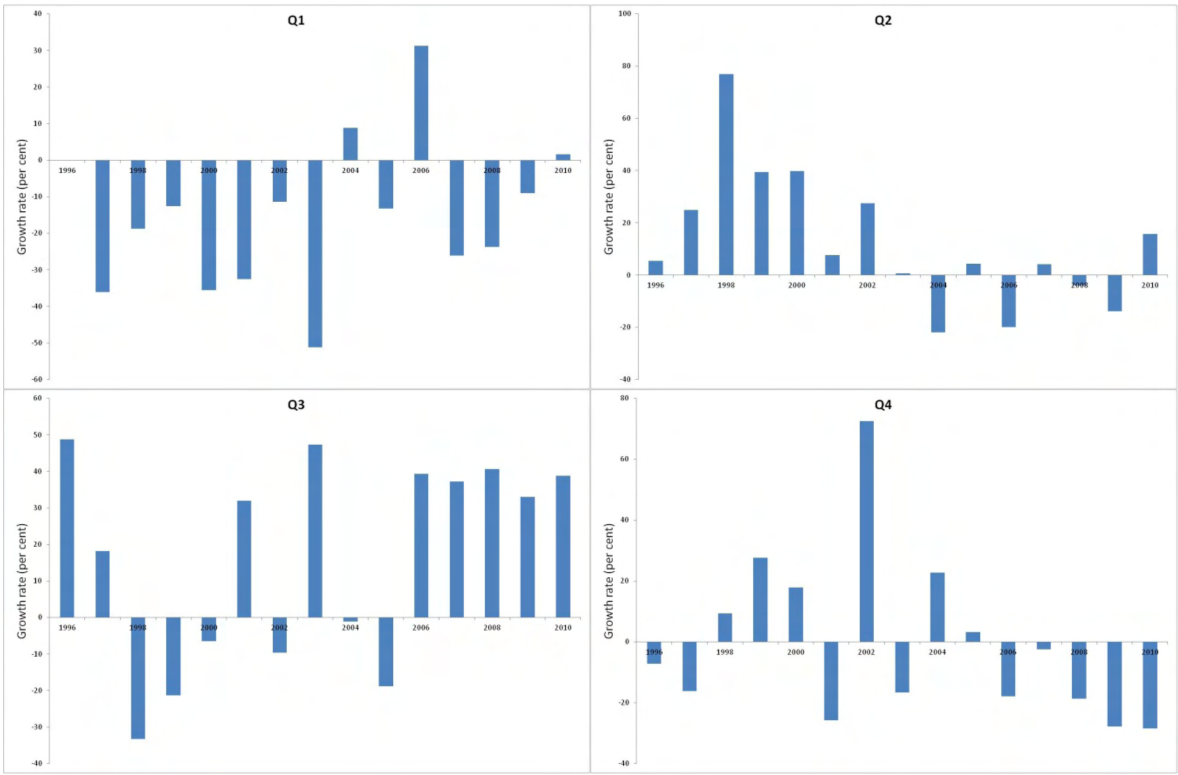
- it is difficult to interpret due to irregular movements, indicated by the plotted NSA and SA series (see figure 8) and failure against the M1-M3 statistics
- it is difficult to interpret due to fluctuations in the seasonal component, indicated by the plotted NSA and SA series and failure against the M8-M11 statistics
- it is unstable, indicated by 73% unstable seasonal factors and 83% unstable quarter-to-quarter changes in the SA series
- it is potentially subject to large revisions, indicated by an average absolute percentage revision, from first to final SA estimates, over the last five years of the series of 3.9%

Figure 8: Initial analysis of other vehicles casualties



After visual inspection of the plotted NSA and SA series, AO2002.4 was added to the regression model and two potential SBs were identified; one in 2004Q1 (consistent decline in Q1 and growth in Q2 before this date) and the other in 2006Q3 (consistent growth in Q3 and decline in Q4 after this date). The possible SBs can also be seen in the plotted quarter-on-quarter growth rates in the NSA series, which are illustrated in figure 9 below.

Figure 9: Growth rates for each quarter compared with the previous quarter for not seasonally adjusted other vehicles casualties



Visual inspection of the plotted NSA and SA series confirms that prior adjusting for both SB2004.1 and SB2006.3 improves SA quality, especially from 2004 onwards (see figure 10). However, it is still difficult to interpret the SA series prior to 2004; a number of additional modifications were investigated to overcome this problem (see “2 SB plus mods” in figure 10 for their combined effect):

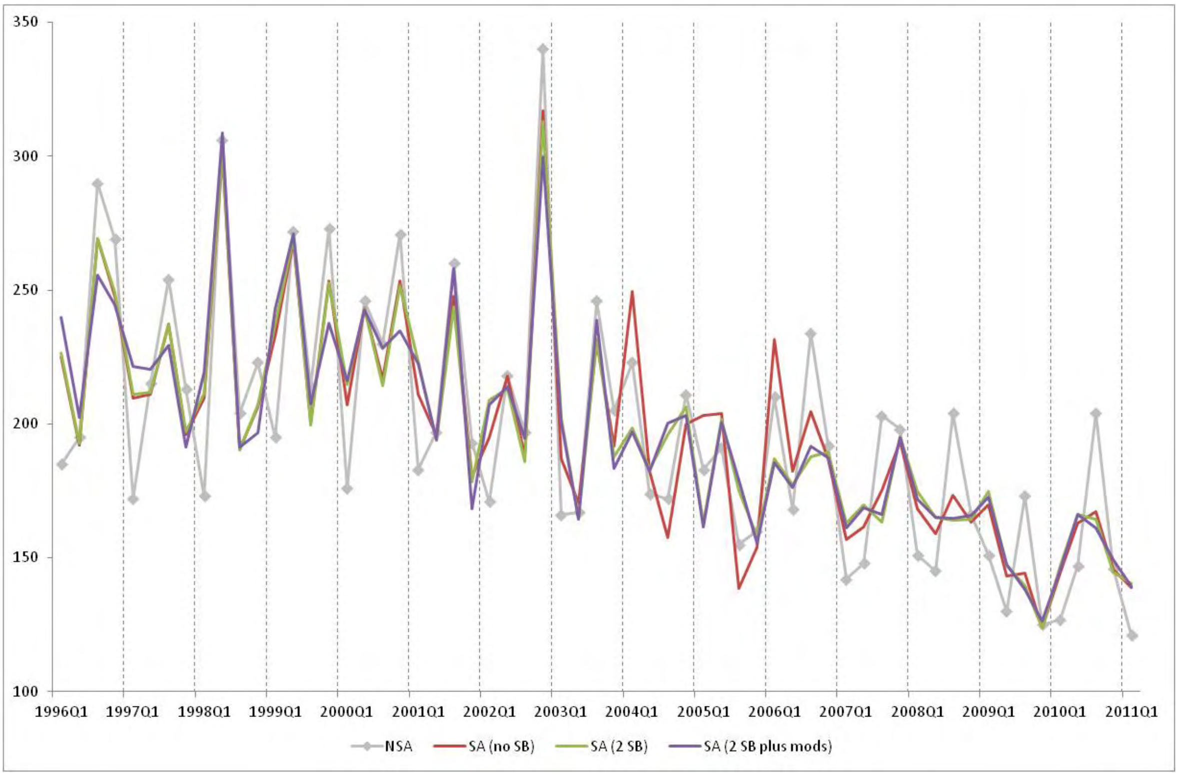
- automatic outlier identification procedure was re-run, but this time just over the span of the series up until the end of 2003 and with the critical t-value reduced to 3.00 – this identified AO1998.2 and LS1999.1: these improved the fit so were added to the regression model
- different SMAs were tried for different quarters (rather than using the same SMA length for all four quarters), chosen on the basis of the I/S ratio split by quarter – 3x9 SMA for Q1 and Q2, 3x3 SMA for Q3 and 3x5 SMA for Q4: these made little difference so were not adopted
- the ‘calendarsigma=all’ argument was included in the ‘x11’ spec so that quarterly standard deviations (rather than the global 5-term moving standard deviation) in the irregular component were used to down-weight extreme values (as the standard deviation in the irregular component is notably higher for Q3 than for the other three quarters): this made little difference so was not adopted

The additional regressors resulted in an overall improvement:

- the M1-M3 statistics (strength of the irregular) still failed, but the M8-M11 statistics (fluctuations in the seasonal component) passed and the overall Q statistic reduced

- the percentage of unstable seasonal factors, the percentage of unstable quarter-to-quarter changes, and the average absolute percentage revision (from first to final SA estimates) over the last five years of the series, all reduced

Figure 10: Adjustment of other vehicles casualties



A notable quality issue for the SA series is potential instability. Although prior adjusting for SB2004.1 and SB2006.3 and including additional regressors limited the instability (as described by X-12-ARIMA's 'sliding spans' diagnostic), it is still relatively high. The series clearly exhibits some seasonal behaviour, as indicated by:

- a combined test result of "identifiable seasonality present"
- an F-value associated with stable seasonality of 70.7
- an M7 statistic value of 0.265
- peaks in the spectrum of the differenced, prior adjusted NSA series at both seasonal frequencies

The series should therefore be seasonally adjusted, but data producers and users should be made aware of the potential instability in the SA series. This instability may manifest itself as changes in magnitude and/or direction of movements in the SA series whenever new observations are added to the current end of the series. The Welsh Government may wish to conduct a more detailed revisions analysis for this (and other) SA series in order to obtain further insight into potential instability.

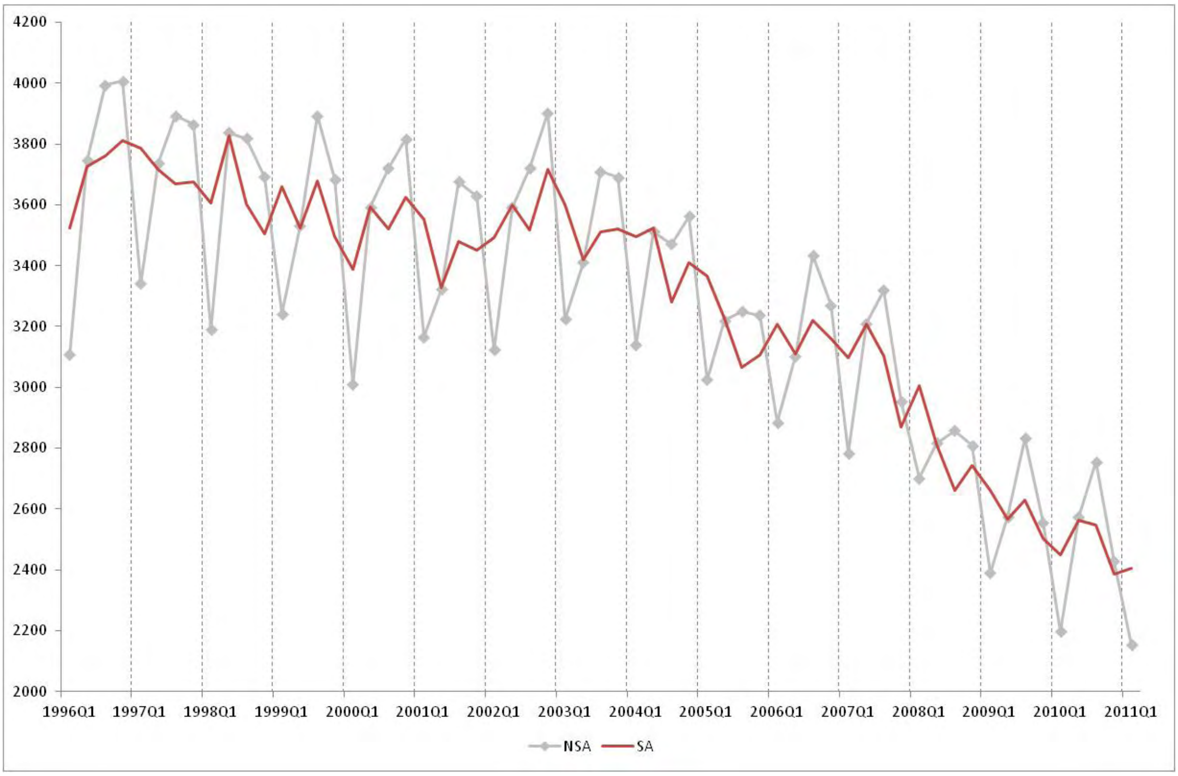
4.6 Total Road Casualties

The NSA series for ‘total road casualties’ exhibits a very similar pattern to that for ‘car, taxi and minibus users road casualties’. This is not surprising, given that the latter is the biggest single component of the former. The SA specifications for ‘total road casualties’ are the same as those for ‘car, taxi and minibus users road casualties’ and are based entirely on the selections from the default run. These specifications result in an adequate SA series in terms of it being:

- free of residual seasonality
- interpretable
- stable

The series ‘SA’ in figure 11 shows a well-behaved SA series for ‘total road casualties’.

Figure 11: Direct adjustment of total road casualties



4.7 Direct versus Indirect Seasonal Adjustment of Total Road Casualties

Two options exist for seasonally adjusting the aggregate 'total road casualties' series:

- direct estimation, whereby the aggregate SA series is itself seasonally adjusted (see section 3.6 above)
- indirect estimation, whereby the aggregate SA series is derived via aggregation of its SA sub-components

Indirect estimation preserves additivity between the aggregate SA series and its SA sub-components. This additivity will not hold under direct estimation, as seasonal adjustment is a non-linear process. However, SA aggregates that have been estimated indirectly are never of better quality than those that have been estimated directly. Figure 12 illustrates both directly and indirectly estimated SA series for 'total road casualties'. The two SA series are broadly similar, but there are some notable differences, particularly over the span 2003 to 2007.

Figure 12: Comparison of direct and indirect adjustment of total road casualties

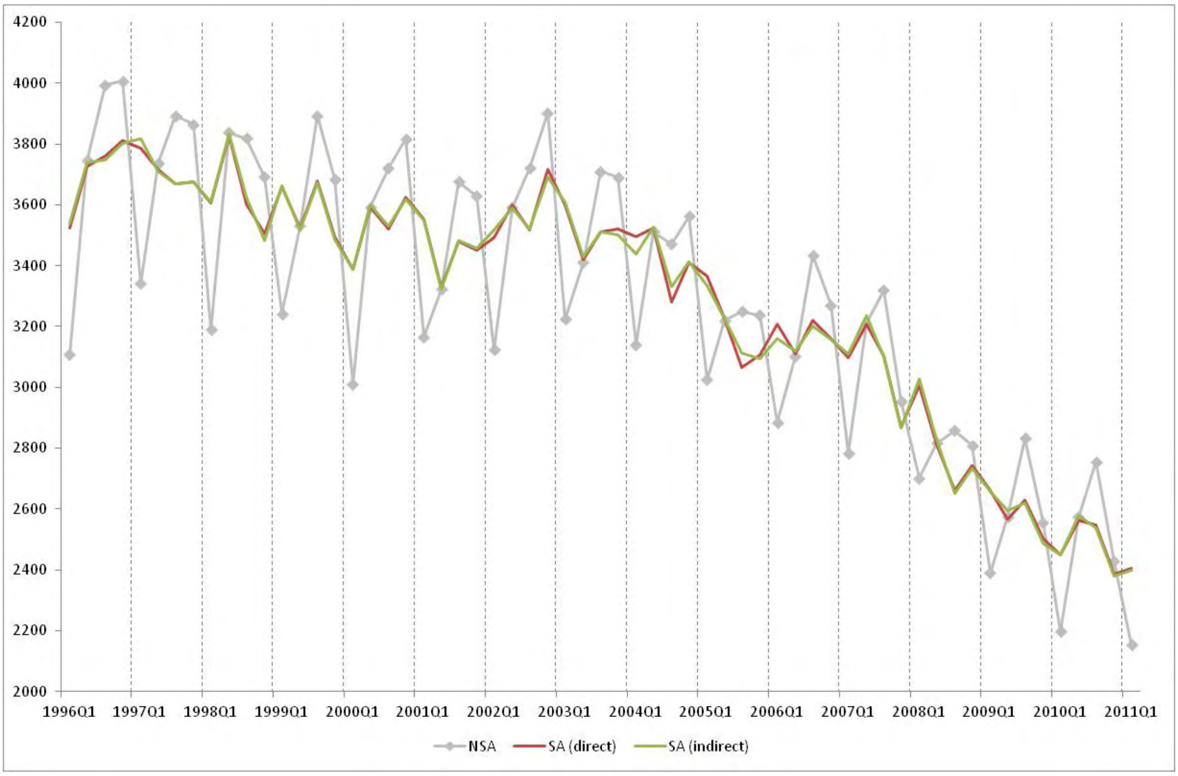


Table 2 below reports some quality measures for each of the two SA series (a lower value is preferred for each measure). The directly estimated SA series is slightly smoother, particularly at the current end of the series, whilst the indirectly estimated SA series might be easier to interpret and is prone to slightly smaller revisions: unsurprising as it is an aggregate of five components, whose revisions tend to cancel, rather than the revisions of a single series.

In principle, direct estimation is preferred. In practice, however, the choice of estimation method appears to make little difference to the quality of the SA series in this case. The Welsh Government should consider user requirements and preferences when deciding which estimation method (direct or indirect) to use for seasonally adjusting the 'total road casualties' time series.

Table 2: Comparison of quality of direct and indirect		
Measure	**Direct**	**Indirect**
Smoothness Measure R1 *RMSE, Whole Series*	124.660	122.050
Smoothness Measure R1 *RMSE, Last 3 Years*	97.952	99.895
Smoothness Measure R2 *RMSE, Whole Series*	0.016	0.018
Smoothness Measure R2 *RMSE, Last 3 Years*	0.012	0.017
Q Statistic Value	0.64	0.54
Number of Failed M Statistics	4	1
% Unstable Seasonal Factors	0.0	0.0
% Unstable Quarter-to-Quarter Changes in the SA Series	2.3	0.0
Average Absolute % Revision *First to Final SA, Last 5 Years*	0.94	0.51

5. Recommendations

The ONS's Time Series Analysis Branch recommends that the Welsh Government:

- Upgrades its seasonal adjustment software to X-12-ARIMA build 192, and in time to X-13ARIMA-SEATS.
- Implements the seasonal adjustment specifications detailed in this report using the spec ('.spc'), regression matrix ('.rmx') and permanent prior adjustment ('.ppp') files provided in conjunction with this report.
- Reviews the seasonal adjustment conducts of road casualty time series (or asks TSAB to do so) every year, or at least every two years, and fixes its seasonal adjustment specifications in between reviews for quarterly production running (this approach is known as partial concurrent adjustment – see the ESS Guidelines on Seasonal Adjustment[2] for more details).
- Reviews the interventions in place at the current end of the 'pedestrian road casualties' time series in the near future.
- Reviews the decision not to prior adjust the 'car, taxi and minibus users road casualties' time series for SB2006.4 in the near future.
- Makes data producers and users aware of potential instability in the 'other vehicles road casualties' SA series: the Welsh Government may wish to conduct a more detailed revisions analysis for this (and other) SA series in order to obtain further insight into potential instability.
- Considers user requirements and preferences when deciding which estimation method (direct or indirect) to use for seasonally adjusting the 'total road casualties' time series.

[2] http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-09-006

Default Specification File

```
series{
file="xxxx.dat"
start=1996.1
period=4
decimals=2
save=a1
}

transform{
function=auto
}

pickmdl{
}

regression{
aictest=easter
}

outlier{
}

x11{
appendfcst=yes
save=(d8 d9 d10 d11 d12 d13)
}

slidingspans{
}

history{
}
```