Seasonal Adjustment of QGDP: The Implications of Outliers

Victoria Wilcox
Statistics New Zealand

Victoria Wilcox
Statistics New Zealand
Analytical Support Division
Christchurch

Statistics New Zealand
Te Tari Tatau

Published in June 2000 by
Statistics New Zealand
Te Tari Tatau
Wellington, New Zealand

Catalogue Number 01.096.0000
ISBN 0-478-20755-7

Seasonal Adjustment of QGDP:
The Implications of Outliers

Summary

The seasonal adjustment package used by Statistics New Zealand, X12-ARIMA, has an automatic procedure for dealing with outliers, which works well in most cases. However in certain circumstances outliers need to be dealt with explicitly.

In this paper we consider the situation of the Quarterly Gross Domestic Product series, in which there are outliers in two consecutive March quarters. When the second of these outliers was the last point in the series the automatic procedure treated neither quarter as an outlier. A simple, practical, solution was to treat one of the outliers explicitly by prior-adjusting the original series. This enabled the other outlier to be identified automatically by X12-ARIMA in the usual way.

Acknowledgement:

I would like to thank Kaylene Murdoch for her contributions and comments to this report during its preparation.

Victoria Wilcox

Analytical Support

April 2000

1 Introduction

Outliers can cause a variety of problems when seasonally adjusting a time series, the major one being revisions in the time series when further data are obtained.

When the Quarterly Gross Domestic Product series to March 1998 was seasonally adjusted it resulted in major revisions to previously published values. This alerted the National Accounts section of Statistics New Zealand to a problem that was affecting the seasonal adjustment. The March 1998 Gross Domestic Product figure was low, as was the March 1997 figure. These two low March figures in a row caused the seasonal adjustment package to identify a new seasonal pattern, leading to larger than usual revisions.

Statistics New Zealand policy is to minimise the number and magnitude of revisions in seasonally adjusted series. We therefore examined the Gross Domestic Product seasonal adjustment procedure to determine whether these revisions were justified.

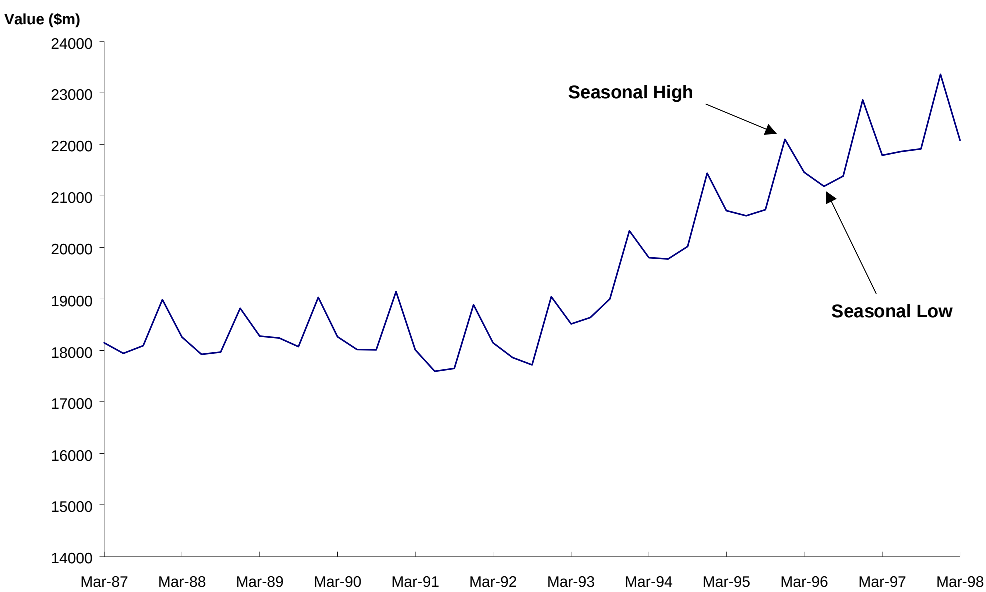
This report outlines the problem in more detail, together with the solution that was implemented.

2 Seasonal adjustment

Statistics New Zealand deals with many time series that exhibit seasonal variation. For example, Quarterly Gross Domestic Product (QGDP) is a seasonal series (see Figure 1). There are, typically, peaks at the March quarters (seasonal highs) and dips at the June quarters (seasonal lows).

Figure 1

QGDP – A seasonal series



Seasonal adjustment aims to remove these seasonal highs and lows with smoothing techniques. This results in a smoother series, which reveals better the series' underlying behaviour.

In the seasonal adjustment procedure used at Statistics New Zealand, a time series is assumed to be composed of independent factors or components:

$$OriginalSeries = Trend * SeasonalFactor * Irregular$$
$$O = T * S * I$$

(NB This is a multiplicative model – occasionally an additive model is used)

- The Trend component (T). This component is an estimate of the underlying movement in the series and thus describes its long-term behaviour. This is useful if one wants to examine changes in the level of a series over a number of years.
- The Seasonal factor (S). This component is an estimate of the effect of any seasonal (annual) pattern present in the series. This is the component we aim to remove in seasonal adjustment.
- The Irregular component (I). This is the component we are concerned with in this report. It consists of all elements remaining once the other components are estimated and removed. The irregular component is assumed to be unstructured, and to consist of random, normally distributed events, which sum to zero over time.

Statistics New Zealand uses the X12-ARIMA seasonal adjustment package for seasonally adjusting its time series. X11 was first developed by the U.S. Bureau of the Census (Shiskin, 1967), which was then enhanced by Statistics Canada to produce X11-ARIMA (Dagum, 1992), and then further enhanced by the U.S. Bureau of the Census to produce X12-ARIMA (Bureau of the Census, 1995).

Seasonal adjustment within all three packages is an iterative process. In simple terms the process is as follows:

1. The original series is ‘detrended’. The underlying movement of the series (the trend) is estimated using a weighted moving average across a number of consecutive quarters. This preliminary trend series is then removed from the original series leaving a detrended series. Following this, any outliers are identified, then removed in subsequent iterations.

2. Preliminary seasonal factors are calculated by taking a weighted moving average, across the same quarter of each year, of the preliminary detrended series. Note, the seasonal factors are calculated across the same quarters. For example, to calculate the seasonal factor for March 1997 we mostly use the March 1995, March 1996 and March 1998 figures.

3. The preliminary seasonally adjusted series is calculated by applying the preliminary seasonal factors to the original series.

This iterative process repeats itself three times before producing a final set of seasonally adjusted figures. For a full explanation of how seasonal adjustment is performed within X12-ARIMA see Shiskin (1967).

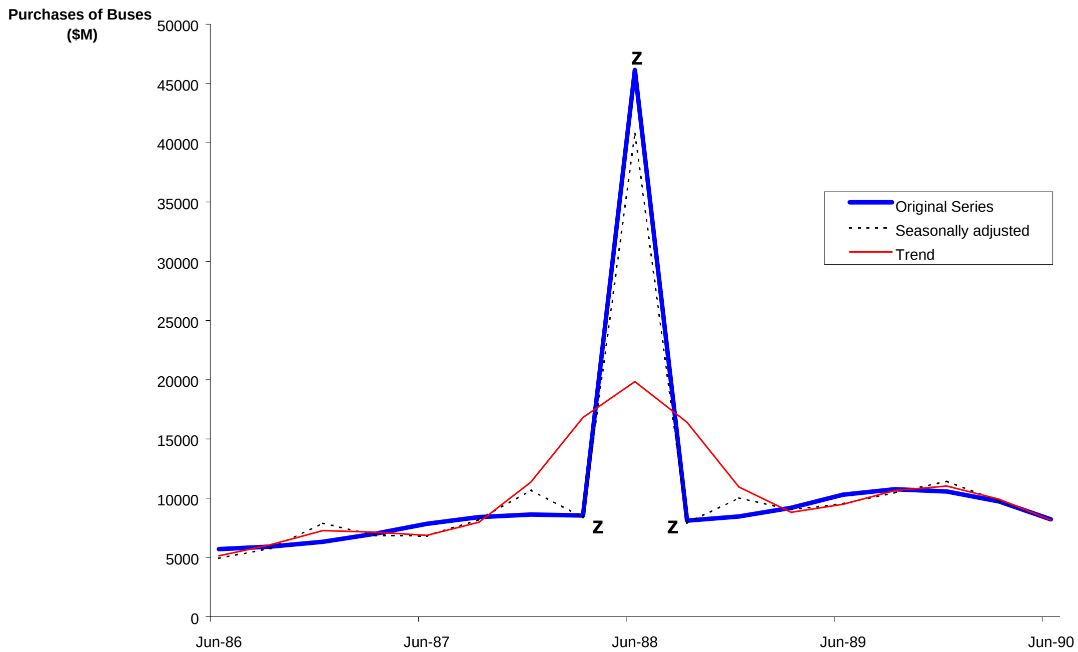
Treatment of outliers

Occasionally the irregular component of a series may contain some identifiable, non-random event. Where this is a major shift in the amplitude of the time series, the event is called a shock. Shocks can be divided into two categories: additive outliers and level shifts. Additive outliers are one-off events that cause a spike in the series. Level shifts are shocks that have a permanent effect. Additive outliers and level shifts can be specified and modelled in a seasonal adjustment.

Within X12-ARIMA, outliers can be dealt with *implicitly* or *explicitly*. In Step 1 of the seasonal adjustment process above, X12-ARIMA automatically identifies outliers using predetermined limits. If a value exceeds this limit it is classified as an outlier and the value is down-weighted in the subsequent iterations. This method of dealing with outliers is the *implicit* method and has the advantage that it is automatic and requires no manual intervention. The downside of this method, however, is that one iteration is completed before the outlier is down-weighted and, hence, the outlier influences the preliminary estimate of the trend and seasonal factor. If the outlier is extreme enough it will affect the final estimates of the trend and seasonal factor. This is illustrated in Figure 2[1]. The outlier still affects the final estimate of the trend, even though it has been zero-weighted (as indicated by a 'z').

Figure 2.

Effect of an Additive Outlier



If subject-matter specialists have good a priori reason to suspect that a data point is an outlier, then another option is to override X12-ARIMA's implicit outlier treatment. We do this by *prior adjusting* the series. This is the *explicit* method of treating outliers. The prior adjustment is done before any preliminary estimates of trend and seasonal factors are made, therefore explicitly treating an outlier will avoid the final estimates being affected by the outlier at all. Note, however, that at Statistics New Zealand we generally do not explicitly treat outliers unless further data confirms the outlier is an additive outlier, as opposed to a level shift. The exception to this is with Overseas Trade Imports data. New Zealand often imports large items, such as the frigates; all items over $100M are treated explicitly to avoid preliminary estimates of trend and seasonal factors being affected.

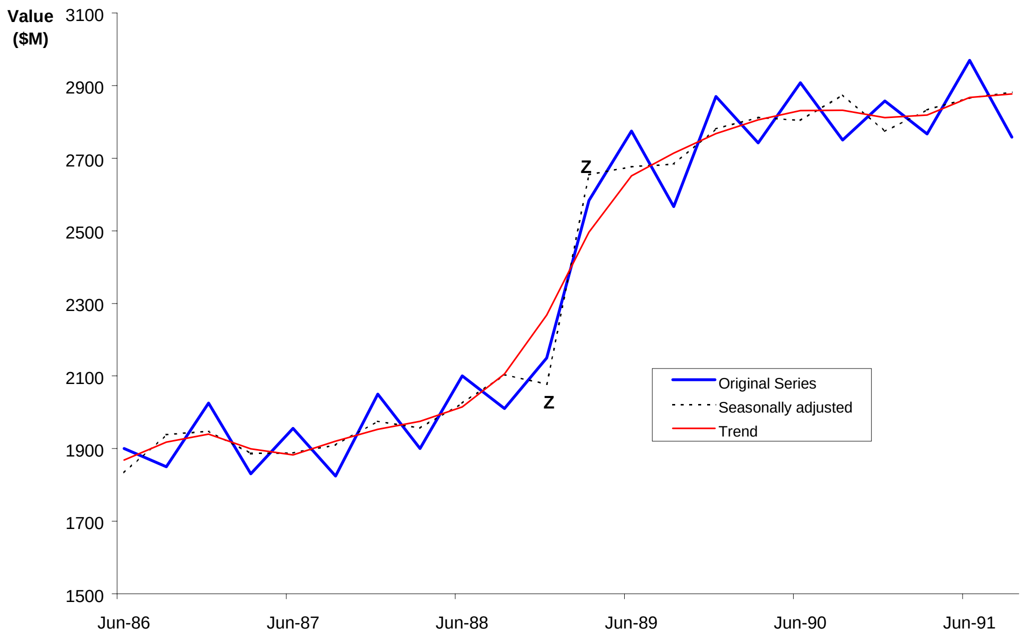
1 Data from Purchases of Buses time series, Gross Domestic Expenditure.

Prior adjustment is performed using the regression option in X12-ARIMA. Here the structure in the irregular component, due to the additive outlier, is modelled using a regression technique with the additive outlier as the regressor. The calculated parameter provides an estimate of the size of the additive outlier and this is then removed before seasonal adjustment takes place (Findley et al, 1998).

For completeness, Figure 3 shows an example of a level shift. Here the time series has experienced a lasting interruption[2]. Again, the figures around the interruption are influenced when the series is seasonally adjusted – these have been given zero weight.

Figure 3.

Effect of a level shift



3 Gross Domestic Product

Gross Domestic Product (GDP) represents 'the total market value of goods and services produced in New Zealand after deducting the cost of goods and services utilised in the process of production, but before deducting allowances for the consumption of fixed capital' (New Zealand System of National Accounts (NZSNA), 1998). The Quarterly GDP (QGDP) figure is a widely used economic indicator whose change is generally interpreted in terms of the scale of economic growth. Within Statistics New Zealand, it is the National Accounts section that is responsible for producing the QGDP series.

In order to calculate QGDP, National Accounts collates a large amount of data from various parts of the economy and condenses this into a single figure. The main sources of information include the Quarterly Manufacturing Survey, Wholesale Trade Survey, Producers Price Index, Retail Trade Survey, and the Annual Enterprise Survey. The figure released is the percentage change from the last quarter. For example, when first released, the March 1998 seasonally adjusted QGDP figure was –0.9 percent. This means there was a 0.9 percent decrease in final goods and services produced between December 1997 and March 1998.

2 The data is artificial.

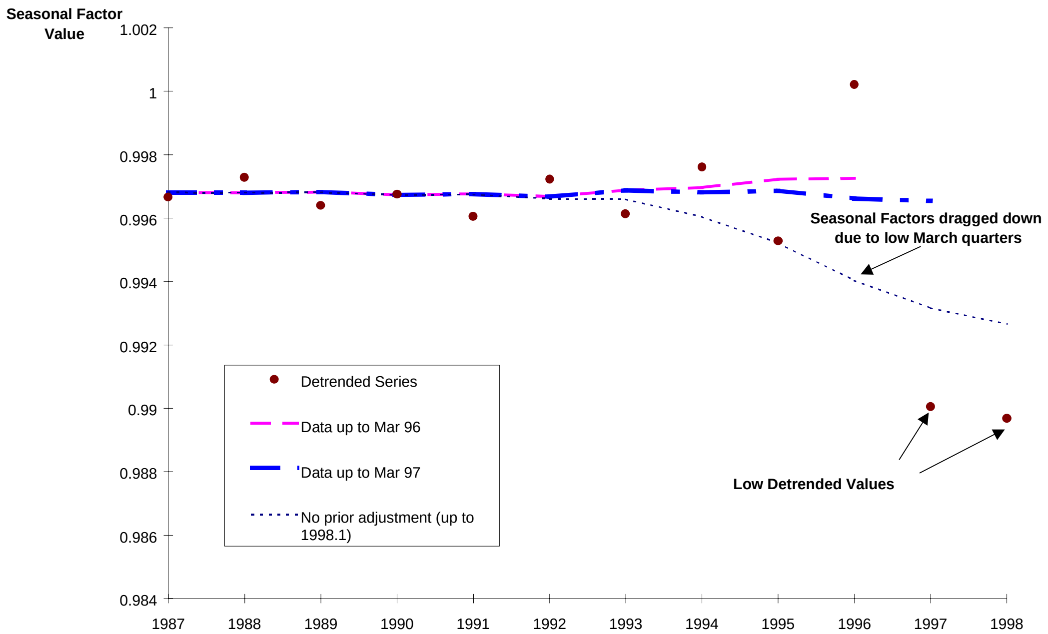
4 An outlier problem in the QGDP series

The March 1998 QGDP figure was lower than usual. Expert opinion from the National Accounts section suggested that this was due to a combination of factors, including a New Zealand drought, and the ongoing Asian economic crisis. It was expected that the effect was a ‘one-off’ and that there would not be a lasting effect on the time series. The March 1998 quarter was therefore likely to be an additive outlier.

However, the 1997 March quarter had also been unusually low, for different reasons. The opinion of subject-matter experts, and the data values around March 1997, confirmed that this was also a temporary event and should be treated as an additive outlier. The detrended March quarter values are presented graphically in Figure 4.

Figure 4.

Effect of outliers on March seasonal factors



Two years in a row with additive outliers in the same quarter presents a problem in X12-ARIMA seasonal adjustment. Before the March 1998 QGDP figure was obtained, X12-ARIMA had correctly recognised the March 1997 figure as an outlier, without any explicit intervention. This was because the March 1997 detrended value is somewhat lower than earlier March detrended values. Figure 4 shows that the March seasonal factors for data up to 1997 are similar to the seasonal factors for data up to 1996 because March 1997 has been treated as an outlier.

However, when seasonal adjustment is performed on the series including the March 1998 figure, there are now two low detrended March values in a row. The 1997 value is now not recognised as an outlier because its low value is seemingly validated by the March 1998 low value. Similarly, the March 1998 value is not recognised as an outlier. X12-ARIMA now gives both the 1997 and 1998 March quarter values full weight in the calculation of seasonal factors, effectively treating the phenomenon as a level shift. Figure 4 shows that the March seasonal factors for data up to 1998 are strongly affected by the March 1997 and 1998 values because these have not been treated as outliers.

The treatment of these data in X12-ARIMA would be entirely appropriate were it felt that the unusual situation of a level shift affecting only the March quarter had been observed. However, National Accounts experts felt that this pattern of low March quarter values was very unlikely to continue. Accepting the default X12-ARIMA treatment would lead to large upward revisions of the 1997 and 1998 March seasonally adjusted values as more usual March quarter values were obtained. This would happen because X12-ARIMA would give increasingly lower weight to the two points, as their outlier status became more apparent. It is Statistics New Zealand policy that the number and magnitude of revisions must be minimised. It was, therefore, clear that the March 1997 and 1998 values needed to be identified as outliers explicitly.

5 The solution

Since X12-ARIMA was unable to identify both the March 1997 and 1998 QGDP figures as additive outliers[3], we needed to remove the effect of these outliers explicitly. This was done using the additive outlier adjustment technique in X12-ARIMA.

1. Firstly, the level of the one-off effect in March 1997 was quantified using the regARIMA function within X12-ARIMA. The following line, `regression {variables =(ao1997.1)}`, along with some other model building specifications were added into the specification file. The full specification file is listed in the appendix. The regression command forces X12-ARIMA to treat March 1997 as an additive outlier. X12-ARIMA then estimates the effect of the outlier and prior adjusts the data – up-weighting the March 1997 quarter by an appropriate amount. This process increased the March 1997 figure by $160 million.

2. This quantity ($160 million) was then added to the March 1997 figure in the original series. This effectively eliminated the outlier at March 1997. Such a series is said to be prior adjusted.

3. The prior adjusted data was then seasonally adjusted as usual using X12-ARIMA. Now the March 1998 outlier could be identified implicitly because it was the only outlier.

4. The resulting seasonally adjusted figure for March 1997 was then adjusted back to allow for the $160 million that was added in Step 2.

The additive outlier is part of the irregular component of the series. In the above process we first corrected for the March 1997 value so that the neighbouring data points were not affected when the series was seasonally adjusted. A seasonally adjusted series consists of the Trend component multiplied by the Irregular component, so the March 1997 outlier should still appear in the irregular part of the final seasonally adjusted series. It is for this reason that the last step is a reversal of the March 1997 correction to the seasonally adjusted series.

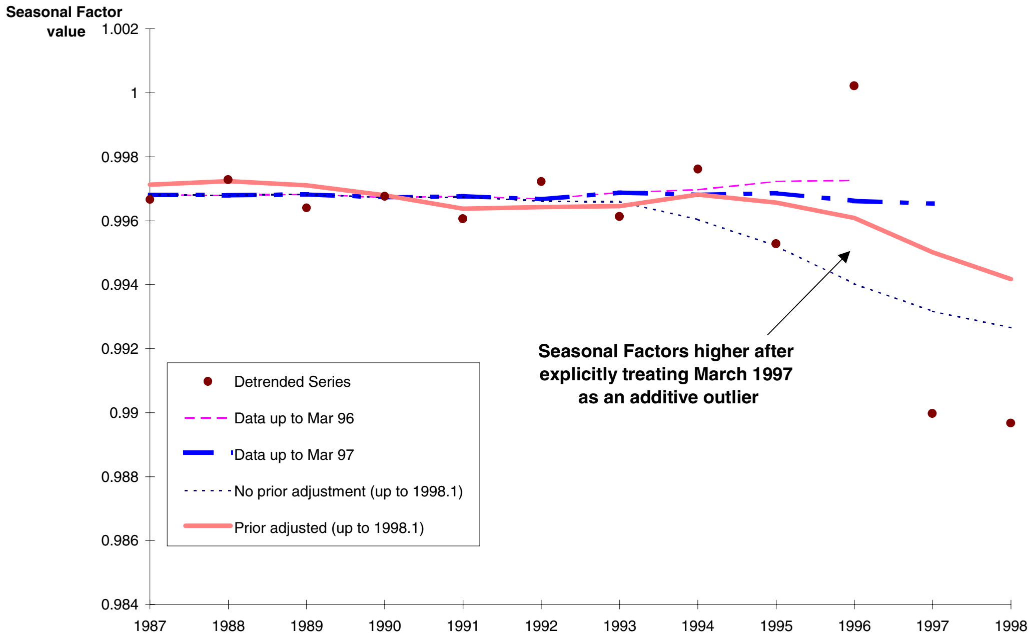
6 Results

Figure 5 shows the March seasonal factors for the QGDP series when the March 1997 outlier was dealt with using the prior adjustment procedure and, subsequently, the March 1998 outlier was dealt with automatically within X12-ARIMA, as described in Section 5. Figure 5 is otherwise identical to Figure 4.

3 At the time this adjustment was first performed, March 1998 was the last data point in the series. We therefore had no further data with which to confirm that March 1998 was, in fact, an additive outlier rather than a level shift. We therefore assumed that it was an additive outlier purely on the basis of the opinion of National Accounts subject-matter experts. Subsequent data has proven the assumption to be valid.

Figure 5

Effect of outliers on March seasonal factors after the additive outlier technique was performed



It can be seen that when we use the explicit method to prior adjust the data, the seasonal factors for the latter part of the series are not as low as when prior adjustment is not performed. There is still, nevertheless, a slight decrease in the seasonal factor line. This is again due to the iterative nature of X12-ARIMA. March 1998 is not explicitly treated and consequently the outlier influences the final seasonal factor estimate. At the time this adjustment was first performed, March 1998 was the last data point in the series. We, therefore, had no further data with which to confirm that March 1998 was in fact an additive outlier rather than a level shift. For this reason, we could not explicitly treat March 1998 until further data (for June, September etc) confirmed its outlier status. The treatment of March 1998 by X12-ARIMA implicitly was sufficient.

Since this procedure was performed, we have obtained further data and have been able to see how the series has actually reacted with a further March data value. Figure 6 shows the March seasonal factors for the QGDP series *to March 1999*, when the March 1997 outlier was dealt with using the prior adjustment procedure. The seasonal factor clearly returns to its normal March level quicker where prior adjustment has been performed.

Figure 6

Seasonal Factors with data up to March 1999

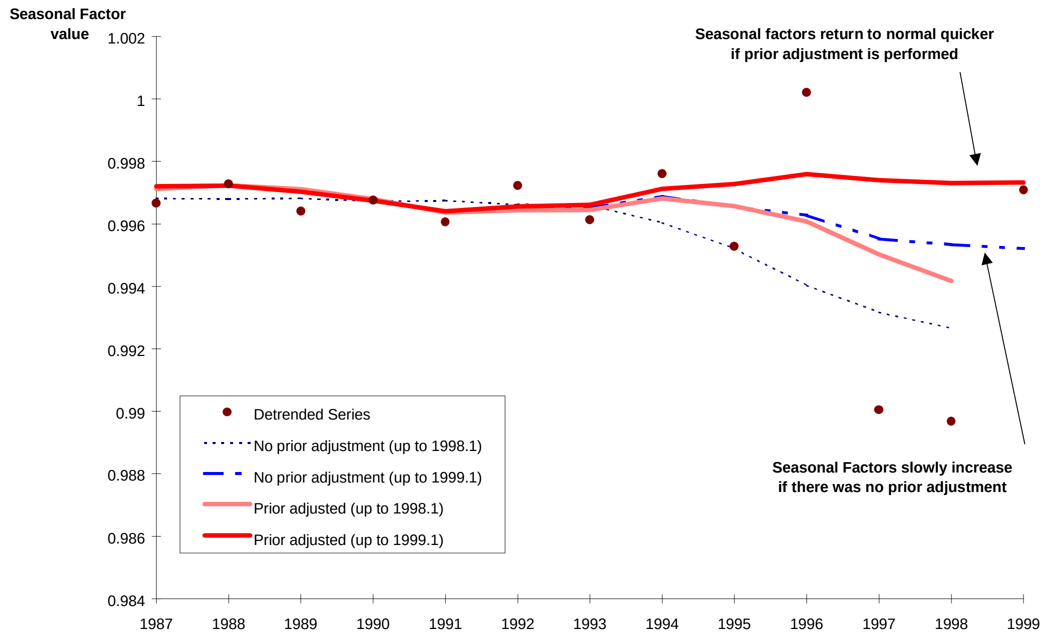


Figure 6 indicates the revision pattern that would have been observed had we not dealt explicitly with the outlier in March 1997. If we didn't prior adjust the series, the seasonal factor would take a lot longer to return to its normal level.

7 Conclusions

The QGDP series to March 1998 presents a case where X12-ARIMA's automatic method for identifying outliers does not work well – two outlying March quarters in a row. Were the automatic procedure relied on here, the result would be unacceptably high revisions to the seasonally adjusted series in subsequent quarters.

Instead, we have shown that if one outlier is treated explicitly using X12-ARIMA's regression command, the second outlier is now within the scope of the automatic procedure. The resulting seasonal factors are now less strongly affected by the outliers and the subsequent revisions are therefore smaller.

The technique outlined in this paper is simple, but provides a quick practical solution in a production environment with tight deadlines. More sophisticated models might alternatively be employed (Kendall and Ord, 1990), but it is unlikely that they would minimise revisions sufficiently further to justify the extra effort.

8 Bibliography

Barnett V and Lewis T 1994. *Outliers in Statistical Data*, 3rd Edition. Chichester: Wiley.

X-12-ARIMA Reference Manual Beta Version 1.0. Bureau of the Census 1995 http://www.census.gov/pub/ts/x12a/final

Dagum EB 1988 X-11-ARIMA/88 *Seasonal Adjustment Method - Foundations and Users' Manual,* Ottawa: Statistics Canada.

Findley DF, Monsell BC, Bell WR, Otto MC and Chen BC 1998. *New capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program, Journal of Business & Economic Statistics,* 16, 127–177.

Harvey AC 1989 *Forecasting, structural time series models and the Kalman filter.* Chapter 3, Cambridge: Cambridge University Press.

Kendall M and Ord JK 1990. *Time Series.* 3rd Edition, New York: Oxford University Press.

Shiskin J, Young AH and Musgrave JC 1967. *The X-11 Variant of the Census Method II Seasonal Adjustment Program.* Technical Paper No. 15. Washington DC: US Department of Commerce.

Statistics New Zealand 1998. *New Zealand System of National Accounts.*

Appendix 1: Specification file

This specification file is from the version of X12-ARIMA in use at the time of this work. A new version of X-12-ARIMA has been released subsequently and some of the options have been changed. This file will need to be converted for use in the new version.

```
series    {title = "QGDP Total - Direct"
          comptype = add
          format = edit
          file = "s:\s_surmth\x12a\consolid\consol.dat"
          name = "consol"
          start = 1986.2
          period = 4
          }

transform { function=log}

regression {variables =(ao1997.1)}

arima { model=(0 1 1)(0 1 1)}

estimate {}

x11{      mode = mult
          sigmalim = (1.8 , 2.8 )
          title = "QGDP Total - Direct"
          print = ( ftestb1  c17  d10  d11  d12  d8  ftestd8 replacsi
          movseasrat  residualseasf  x11diag qstat  specsa  specirr
          seasonalb10)
          save = ( b1  c17  d8  d10  d11  d12  d13 )
          }

regadjust { prior=(ao)
             final=(ao)
             save=(ao)}
```

Appendix 2:
Definitions

Partial weighted: For Statistics New Zealand series this means that the irregular component, of that data point, minus the mean (0.0 for additive series and 1.0 for multiplicative series) multiplied by the estimated standard error of the irregular component of the data point, is between 1.8 and 2.8. If so, the data point is down-weighted using a linear scale between 0 and 1 depending on how high the standard error is. If the standard error is close to 1.8 the weight will be close to 1, and if it is close to 2.8 the weight will be close to 0.

Zero weighted: For Statistics New Zealand series this means that the irregular component, of that data point, minus the mean (0.0 for additive series and 1.0 for multiplicative series) multiplied by the estimated standard error of the irregular component of the data point is above 2.8. This means that the irregular component is replaced by the mean irregular component, this is 1.0 for a multiplicative series, and 0.0 for an additive series.

Weighted moving average: The calculation of a mean value using the points immediately surrounding the data point. The values closest to the middle point have the greatest emphasis placed on them, this weight decreases as one moves further away from the middle point.