The effect of ARIMA forecasting on revisions to seasonally adjusted time series

Kaylene Murdoch, Victoria Wilcox, Geoff Kuzmicich, and Frances Krsinich
Statistics New Zealand

Kaylene Murdoch
Analytical Support Division
Statistics New Zealand
kaylene_murdoch@stats.govt.nz
1 Summary

We conducted an investigation into the effectiveness of the ARIMA forecasting option in X-12 ARIMA for reducing revisions to seasonally adjusted series. The default ARIMA models were found to fit and forecast 80 percent of our sample of Statistics New Zealand series acceptably well. However, a reduction in revisions was observed only for series that were already subject to large percentage revisions.

2 Introduction

Statistics New Zealand publishes seasonally adjusted estimates for several hundred economic time series. Revisions to these estimates are an inconvenience to analysts and policy-makers.

Nevertheless revisions are inevitable. Even if the raw data is unchanged, the addition of each new measurement provides a fresh perspective into the most recent observations of the series. This leads to a revised estimate of the seasonally adjusted series. Revisions, therefore, occur regardless of the seasonal adjustment procedure.

However in moving-average methods of seasonal adjustment, such as the Census Method II X-12 variant used at Statistics New Zealand, a characteristic of the seasonal adjustment procedure is itself implicated as a contributor to revisions. In this method an asymmetric filter is used at the end of the series and a symmetric filter is used in the body of the series. Once a data point is no longer in the asymmetric part of the filter it is affected differently by the symmetric filter, requiring its seasonally adjusted estimate to be revised.

Clearly the X-12 contribution to revisions is undesirable and should be minimized. Forecasting future series values (Dagum 1975) can solve the problem. The forecasts are appended to the series of actual data and the seasonal adjustment performed as before. The asymmetric filter now affects mainly just the forecasted values, which are then discarded.

The ability of the forecasting method to reduce revisions depends on how good the forecast is. If the forecast is very poor then the asymmetric filter may have been better. If the forecast is very good then revisions may be very small indeed. Somewhere in between is the plausible case of a forecast that is just good enough to constitute an improvement over the use of the asymmetric filter. The trick to minimizing revisions is in the calculation of forecasts that are good enough.

Some years ago, Statistics Canada implemented the forecasting solution by incorporating a facility for fitting Autoregressive Integrated Moving Average (ARIMA) models to time series in the seasonal adjustment package, X-11 ARIMA (Dagum 1980). Forecasting future series values using ARIMA models, Statistics Canada was able to show (Dagum 1975, 1982, 1992) that where the fits satisfied certain criteria built into the program, they were able to reduce the magnitude of revisions to Canadian seasonally adjusted estimates by around 20–30 percent (Dagum 1992).

A significant aspect of X-11 ARIMA, and the more recent version, X-12 ARIMA, is that it fits ARIMA models in an automatic way suitable for use in an environment where a large number of seasonal adjustments are performed regularly. Identification of an appropriate ARIMA model, which would usually involve a degree of handcrafting, is achieved through the pre-selection of a limited number of default models. The suitability of each model is determined from three fit diagnostics. If more than one model fits, the best is chosen automatically. The five default models chosen by Statistics Canada were those thought to represent economic time series well. A later empirical study into how well these models fitted Canadian time series (Chiu, Higginson, & Huot 1985) largely confirmed the initial default model selection.

Early ad-hoc attempts at Statistics New Zealand to implement the ARIMA forecasting option on New Zealand series were unsuccessful. With the critical values of the diagnostics set at values with which Statistics Canada experienced success limits (forecasting error less than 15 percent, chi-square < 0.05, moving average characteristic equation coefficient < 0.9), the program indicated that no default model fitted. It was not clear whether this problem was widespread amongst New Zealand series, indicating unsuitability of the default models, or whether our early experience just happened to be with particularly difficult series.
What was needed was a comprehensive exploration of the potential of ARIMA forecasting at SNZ. Do the X-12 ARIMA default models fit any SNZ series? If not, are there other defaults that would be better? Would ARIMA forecasting actually reduce revisions in Statistics New Zealand series?

In this report we discuss a project that was carried out at Statistics New Zealand to establish whether the magnitude of revisions to New Zealand seasonally adjusted estimates could be reduced through the use of the ARIMA forecasting option in X-12 ARIMA. We carried out the project in two stages. Firstly we evaluated how well the X-12 ARIMA default models fitted New Zealand series. We did this by attempting to fit each default model (and two additional models) to a range of different series. Each fit was evaluated via eight fit diagnostics. Secondly, we considered whether ARIMA forecasts were actually capable of reducing revisions to New Zealand seasonally adjusted series. To quantify the revisions we used the X-12 ARIMA history spec to compare the magnitude of revisions with and without ARIMA forecasting.

3 Sample selection
The first step was to select a representative sample of Statistics New Zealand time series on which to test the ARIMA modelling and forecasting procedures. The sample we chose comprised the most important published Statistics New Zealand series — 62 quarterly series and 31 monthly series — 93 in total. For simplicity we only included series for which we knew that a multiplicative decomposition model was appropriate. Table 1 gives a breakdown of the sample by sector and sampling frequency.

Only series that were long enough for ARIMA modeling were included. The series lengths were limited by how much of each series was stored in our time series management software. For example, early portions of series for which there had been a permanent change in seasonality, or parts of series prior to a re-base, would not have been stored in our current system. Series lengths ranged from 6-33 years but were usually about 10 years.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Monthly</th>
<th>Quarterly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail Trade</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Gross Domestic Product</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Building Consents</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Migration</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Balance of Payments</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Overseas Trade</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Manufacturing</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Household Labour Force</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>31</strong></td>
<td><strong>62</strong></td>
</tr>
</tbody>
</table>

4 Fitting the default models: methodology
We chose to use the SAS software package, rather than X-12 ARIMA, as the tool for testing the fitting of ARIMA models to the New Zealand series. This was mainly because SAS enabled us to evaluate a wider range of fit diagnostics. We first passed each series through an X-12 ARIMA adjustment so that outliers did not influence the model fits. We then used the log transform of the resulting E1 table (that is, the original data adjusted for zero-weighted outliers) as input into SAS.

* Thirteen Local Authority series were also initially included in the sample but were later withdrawn due to an undiagnosed problem in fitting more than one ARIMA model to these series when using X-12 ARIMA.
To each test series we attempted to fit not just the five X-12 ARIMA default models but, to enable comparison with the results of Chiu et al (1985), all seven models considered by them. These models are as follows, in standard ARIMA \((pdq)(PDQ)_s\) format:

\[
\begin{align*}
(0,1,1)(0,1,1)_s & \quad \text{(X-12 ARIMA default model)} \\
(0,1,2)(0,1,1)_s & \quad \text{(X-12 ARIMA default model)} \\
(0,2,2)(0,1,1)_s & \quad \text{(X-12 ARIMA default model)} \\
(2,1,2)(0,1,1)_s & \quad \text{(X-12 ARIMA default model)} \\
(2,1,0)(0,1,1)_s & \quad \text{(X-12 ARIMA default model)} \\
(1,1,0)(0,1,1)_s & \\
(2,1,0)(0,1,2)_s
\end{align*}
\]

The fits were evaluated using the eight different criteria that Chiu et al. (1985) used to evaluate the ARIMA models fits to Canadian series, namely:

1) Stationarity (the fitted model should be stationary, as evidenced by the absolute value of the coefficients of the AR characteristic equation being less than 1).

2) Invertibility (the fitted model should be invertible, as evidenced by the absolute value of the coefficients of the MA characteristic equation being less than 1).

3) Underdifferencing (there should be no underdifferencing, as evidenced by the absolute value of the coefficients of the AR characteristic equation being less than 0.9).

4) Overdifferencing (there should be no overdifferencing, as evidenced by the absolute value of the coefficients of the MA characteristic equation being less than 0.9).

5) Randomness of residuals (the residuals should appear to be random, as measured by the \(Q\)-statistic, \(Q \geq 0.05\)).

6) Small parameters (there should be no parameters whose absolute values are 0.1 or less, which suggests a simpler model might be more appropriate).

7) Correlation of parameters (the parameter correlations should be less than or equal to 0.9).

8) Forecasting error — the mean absolute percentage error of the one-year-ahead out-of-sample forecasts should be less than or equal to 15 per cent.

The reader is referred to Chiu et al (1985) for fuller description of these criteria.

Note that the X-12 ARIMA default criteria are a subset of the above, namely 4, (overdifferencing), 5 (randomness of residuals), and 8 (forecasting error). The default critical values are the same as those used by Chiu et al (1985). The forecasting error criterion is estimated slightly differently in X-12-ARIMA in that within-sample rather than out-of-sample forecasts are used (ie the last year of data used to calculate forecast errors is used in the estimation of the model). However this difference is unlikely to be significant (see for example Lothian and Morry 1978).

The Chiu et al (1985) criteria above were chosen for evaluating ARIMA model fits to New Zealand series in preference to the X-12 ARIMA criteria because the former were able to give us greater confidence that a model that passed was truly a good representation of the series. We therefore expected that any model fit which passed the Chiu criteria would in X-12 ARIMA produce forecasts that were good enough to reduce the magnitude of revisions to seasonally adjusted estimates.
5 Fitting the default models: results

We attempted to fit each of the seven ARIMA models above to all of the 93 New Zealand series we tested. The main results were as summarized in Table 2.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Monthly</th>
<th>Quarterly</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of series where no acceptable ARIMA model was found</td>
<td>10</td>
<td>5</td>
<td>15 (16%)</td>
</tr>
<tr>
<td>Number of series where at least one acceptable ARIMA model was found</td>
<td>21</td>
<td>57</td>
<td>78 (84%)</td>
</tr>
<tr>
<td>Total number of series</td>
<td>31</td>
<td>62</td>
<td>93 (100%)</td>
</tr>
</tbody>
</table>

Note that:

- An ‘acceptable’ model is one that passed the eight criteria of Chiu et al (1985) listed in Section 4. Since acceptable models were found for 84 percent of the series, we expected that ARIMA forecasting had the potential to reduce revisions in about this proportion of Statistics New Zealand series.

- Only five of the seven models were required for the results in Table 2, therefore the remaining two models (including the existing X-12 ARIMA default model (2,1,2)(0,1,1)) were not of importance in the New Zealand context. Of the five useful models remaining, all but one was an existing X-12 ARIMA default. Because this model, (1,1,0)(0,1,1), fitted many series better than any other model, we proposed that it should replace (2,1,2)(0,1,1) as a default for New Zealand seasonal adjustments. Note however that there was no series which was fitted solely by (1,1,0)(0,1,1) so, if only the five Canadian default X-12-ARIMA models were applied to our test series, we would still found that 84 percent of the series fitted adequately.

- The quarterly series were fitted better than the monthly series and those quarterly series unable to be fitted all came from Balance of Payments. However we could not draw conclusions on the relationship between whether a series fitted well and other variables such as the length of the series and the sector of the economy from which it arose. This is because the sample of 93 series was small and came from a small number of sectors within which variables such as frequency and length were identical.

6 Quantifying reduction in revisions: methodology

We had established that, with one change, the existing default ARIMA models in X-12 ARIMA were applicable to New Zealand series. We wished now to use these default models in X-12 ARIMA to seasonally adjust the New Zealand series. The aim was to see if using the ARIMA forecasting option did actually reduce revisions.

6.1 Quantifying revisions

For a user of Statistics New Zealand seasonally adjusted data, the desired properties of revisions are:

- The concurrent seasonally adjusted value (ie the seasonally adjusted value when the point was the last in the series) is as close as possible to the final value that is reached when many data points have been added to the series (this may be some years after the data point was first added to the series).

- Seasonally adjusted values subsequent to the first converge as smoothly as possible to the final value. Overshooting of the final value, for example, is undesirable.

Therefore to assess the merits of ARIMA forecasting we needed to compare the revision between the concurrent seasonally adjusted value and the final value for the cases of ARIMA forecasting and no ARIMA forecasting. The idea was to examine whether ARIMA forecasting provides concurrent seasonally adjusted estimates that are closer to the final seasonally adjusted estimates than the concurrent estimates derived with no ARIMA forecasting.
In order to examine whether the convergence from current to final seasonally adjusted estimate was smooth, we also needed to look at seasonally adjusted estimates intermediate between these two periods, for the cases of ARIMA forecasting and no ARIMA forecasting.

6.2 The X-12 ARIMA .spc file

To prepare for seasonally adjusting our sample series using X-12 ARIMA we first modified the default ARIMA models in the file referred to by the X-12 ARIMA automdl spec so that it contained models identified earlier as suitable ARIMA defaults for Statistics New Zealand. The resulting file therefore contained:

- \((0,1,1)(0,1,1)_s\) (existing X-12 ARIMA default model)
- \((0,1,2)(0,1,1)_s\) (existing X-12 ARIMA default model)
- \((0,2,2)(0,1,1)_s\) (existing X-12 ARIMA default model)
- \((2,1,0)(0,1,1)_s\) (existing X-12 ARIMA default model)
- \((1,1,0)(0,1,1)_s\) (new X-12 ARIMA default model)

We then constructed two .spc files for each series — one to seasonally adjust the series using the ARIMA forecasting option and the other to seasonally adjust it without ARIMA forecasting. The Appendix lists an example .spc file where ARIMA forecasting was used. Important aspects of the configuration of the .spc file are:

- We configured the automdl spec to direct the program to select the best model of our defaults (the alternative is to select the first model that fits).
- The critical values for the diagnostics used by X-12 ARIMA to assess the fits are the same as the critical values we used for those same diagnostics to assess the ARIMA model fits in Section 4.
- To quantify revisions, we employed the history spec which requests a sequence of runs from a sequence of truncated versions of the time series. This creates historical records of revisions from initial seasonal adjustments.

6.3 The history spec

We employed the history spec both for when ARIMA forecasting was used and for when it was not used. For each series which X-12 ARIMA could fit adequately with an ARIMA model this spec was able to provide:

- Concurrent seasonally adjusted values for all points in the series beyond a lead time (6, 8, or 12 years) determined automatically according to the length of the longest seasonal filter used for that series.
- Final (ie the latest) seasonally adjusted values for these points.

To examine the smoothness of the path of the revisions, we treated two series as case studies. We tracked the path that seasonally adjusted values take from their first appearance as concurrent estimates through to their most recent appearance as final estimates. For these series we used the history spec to calculate:

- \(n\)-period-later seasonally adjusted values, that is, the seasonally adjusted values for points after \(n\) additional data points had been obtained, where \(n\) was such that the estimate was intermediate between concurrent and final.

In all cases we allowed the parameters of the ARIMA model to vary from one history fit to the next, but the history spec keeps the model itself fixed for a given series.
Comparison of ARIMA model selection in SAS and in X-12 ARIMA

Before we discuss the results of our analysis of revisions using the history spec, we discuss in this section any differences between the ARIMA models selected by X-12 ARIMA in the procedure described in Section 6 and the ARIMA models selected by our SAS investigation described in Section 4.

7.1 Expectations

We expected some differences between the SAS and X-12 ARIMA procedures in best ARIMA models found because these procedures differed slightly in the way fits were evaluated and the way outliers were treated.

There were eight fitting criteria used in SAS and just three in X-12 ARIMA. We expected then that models identified using our SAS procedure would not necessarily include the best model chosen using the X-12 ARIMA procedure. For example, there was no screening in X-12 ARIMA for models with small parameters, so we expected that where there was a difference between a model chosen by X-12 ARIMA and the ones identified by SAS, the SAS model would be the more parsimonious one.

We also expected that since SAS had more criteria for a model fit to pass, there would be cases where an acceptable model could be found by X-12 ARIMA but not by SAS (because the model found by X-12 ARIMA had failed one of the additional SAS criteria).

Also, in our SAS analysis, the input data set was the E1 output table from X-12 ARIMA, that is, the raw data adjusted for extreme outliers. However in the X-12 ARIMA analysis the input data set was simply the raw data. We chose to use the raw data as input into X-12 ARIMA because we wanted to simulate the actual process that would be adopted if the ARIMA forecasting option were to be taken up for day-to-day seasonal adjustment. However, since outliers are not removed in X-12 ARIMA before the fitting of ARIMA models, this meant that we could expect some differences between SAS and X-12-ARIMA in whether an appropriate ARIMA model could be found. We expected that there would be cases where a model could be found in SAS but no model could be found using X-12 ARIMA because of the presence of extreme outliers.

7.2 Results

Table 3 illustrates that of the 93 series, 80 percent could be fitted by an ARIMA model in X-12 ARIMA. At an overall level, this was consistent with the 84 percent for which an acceptable model was found using SAS (see Table 2).

<table>
<thead>
<tr>
<th>Fitting of ARIMA models using X-12 ARIMA</th>
<th>Monthly</th>
<th>Quarterly</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of series where no acceptable ARIMA model was found</td>
<td>12</td>
<td>7</td>
<td>19 (20%)</td>
</tr>
<tr>
<td>Number of series where at least one acceptable ARIMA model was found</td>
<td>19</td>
<td>55</td>
<td>74 (80%)</td>
</tr>
<tr>
<td>Total number of series</td>
<td>31</td>
<td>62</td>
<td>93 (100%)</td>
</tr>
</tbody>
</table>

As expected, however, there were some differences between the models chosen by X-12 ARIMA and the models that were chosen for the same series using SAS.

There were 15 series for which no acceptable model was found in SAS. For 10 of these series there was no model found using X-12 ARIMA either. This left five series for which a model was found in X-12 ARIMA even though it wasn’t in SAS. In these five cases the Q-statistic was acceptable in the X12-ARIMA analysis but not in the SAS analysis, suggesting a sensitivity of this statistic to the slight differences between the SAS and X-12 ARIMA datasets.

For nine series no fit was found in X-12 ARIMA, although a fit was found in SAS. These appear to be series which have extreme outliers which were removed before the SAS analysis but prevented a model fitting successfully in the X-12 ARIMA analysis.
It therefore appeared that any difference between models found between the SAS and X-12 ARIMA methods was explainable by differences between the methodologies. The only critical issue that we noted that compromised the X-12 ARIMA forecasting technique was that outliers apparently prevented an acceptable ARIMA model being found for nine series.

8 Quantifying reduction in revisions: results

As can be seen in Table 4, 64 percent of the series for which an acceptable ARIMA model was found in X-12 ARIMA were too short for the history analysis, which requires a lead time of 6–12 years before the period used for the analysis. This was disappointing, nevertheless there were still 27 series that could be analyzed. These results were used to assess how well seasonal adjustment revisions could be reduced for SNZ series through the use of ARIMA forecasting.

Table 4

<table>
<thead>
<tr>
<th>Suitability of series for history analysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of series too short for history analysis but having an acceptable ARIMA model</td>
<td>47 (64%)</td>
</tr>
<tr>
<td>Number of series with history analysis &amp; acceptable ARIMA model</td>
<td>27 (26%)</td>
</tr>
<tr>
<td>Total number of series with acceptable ARIMA model</td>
<td>74 (100%)</td>
</tr>
</tbody>
</table>

8.1 Magnitude of concurrent-to-final revisions

For the 27 series for which we could use the history analysis, we compared the concurrent-to-final revisions for the ARIMA forecasting and no ARIMA forecasting cases by looking at the mean absolute percentage revision $R_t$ for each case. This was calculated as

$$R_t = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_{t_{f}} - A_{t_{a}}}{A_{t_{a}}} \right|,$$

where

$A_{t_{a}}$ is the concurrent estimate of the seasonally adjusted series at time $t$, ie the value of the seasonally adjusted series at time $t$ calculated from the series up to time $t$;

$A_{t_{f}}$ is the final estimate of the seasonally adjusted series at time $t$, ie the value of the seasonally adjusted series at time $t$ calculated from the series up to the most recent time point $T$;

$N$ is the number of points used in the history analysis.

For comparison of revisions with and without ARIMA forecasting, ideally the final seasonally adjusted estimate (rather than the concurrent estimate) should have been the denominator in Equation (1), as its value is approximately the same with or without ARIMA forecasting. However we expected that this would not affect the comparison very much since the percentage revisions were always small.

We then looked to see if $R_t$ tended to be smaller for the case of ARIMA forecasting. The results were as follows:

Table 5

<table>
<thead>
<tr>
<th>Effect of ARIMA forecasting on mean percentage revisions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of series better (smaller $R_t$) with ARIMA forecasting</td>
<td>14</td>
</tr>
<tr>
<td>Number of series same (same $R_t$) with ARIMA forecasting</td>
<td>4</td>
</tr>
<tr>
<td>Number of series worse (greater $R_t$) with ARIMA forecasting</td>
<td>9</td>
</tr>
<tr>
<td>Total number of series</td>
<td>27</td>
</tr>
</tbody>
</table>
Table 5 illustrates that not all series showed a reduction in revisions from concurrent to final estimates. Just 14 of the 27 series showed smaller revisions where ARIMA forecasting was used. This left 13 series where the use of the ARIMA forecasting option gave the same or increased concurrent-to-final revisions.

Of the series that showed a reduction in mean percentage revisions where ARIMA forecasting was used, the mean reduction in revisions was 13 percent and the maximum reduction in revisions was 30 percent. These tended to be series for which the percentage revisions were relatively high, with or without the ARIMA forecasting option. The magnitude of these reductions in revisions was similar to that experienced by Statistics Canada.

Figure 1

**Relationship between mean percentage revision (without the ARIMA forecasting option) and the decrease in mean percentage revision experienced with the ARIMA forecasting option**

![Graph showing relationship between mean percentage revision and reduction in revisions.]

Of the series that showed an increase in mean percentage revisions where ARIMA forecasting was used, the mean increase in revisions was 24 percent and the maximum increase in revisions was 60 percent. However these series were generally those for which revisions were already small.

Figure 1 illustrates the relationship between mean percentage revision (without the ARIMA option) and the reduction in mean percentage revision that occurred with the ARIMA option. It suggests that for the more volatile series, where the mean percentage revision is above about 1.5 percent, the ARIMA forecasting option indeed causes a reduction in revisions. As an example of the scale of the reduction, a series with a mean percentage revision of 2.0 percent would be expected to have a mean percentage revision of 1.7 percent (ie a reduction of 15 percent) with the ARIMA forecasting option. For the more stable series, where the mean percentage revision is less than about 1.5 percent, ARIMA forecasting may actually increase revisions.

8.2 Convergence from concurrent estimate to final estimate

For two series, Total Quarterly Gross Domestic Product (QGDP) and Balance of Payments (BOP) Services Credits, we ran the history analysis not just for the concurrent and final estimates but also for some intermediate estimates. These intermediate estimates were those after 1, 2, 3, 4, and 5 additional data points had been obtained.

Figures 2 and 3 are plots for points from Total QGDP and BOP Services Credits. The plots illustrate what happens to a data point as it progresses from being a concurrent estimate through to a final stable value once many further data points have been obtained.
The revisions of a seasonally adjusted value from the Total QGDP series are represented in Figure 2. This was a series that had mean percentage revisions of only 0.26 percent without ARIMA forecasting and showed no change with ARIMA forecasting. It can be seen that for this particular point the concurrent seasonally adjusted estimate is closer to the final estimate where an ARIMA forecast has been used. However the path from the concurrent estimate to the final estimate is not direct but involves an overshoot of the final estimate. This behaviour was observed in other points in this series.

The revisions of a point from the BOP Services Credits series are illustrated in Figure 3. This was a series that had mean percentage revisions of 1.46 percent without ARIMA forecasting and 1.39 percent with ARIMA forecasting. Here we can see improved revision behaviour for the case of ARIMA forecasting, which was fairly typical of points in this series. The concurrent seasonally adjusted estimate is better where ARIMA forecasts are used and both methods converge in a well-behaved manner to the final estimate.

Thus we have found one case of a series where the concurrent-to-final revisions were on average not improved by ARIMA forecasting and with intermediate revisions often worse; and another case where the concurrent-to-final revisions were on average improved by ARIMA forecasting with intermediate revisions improved. While it cannot be claimed that these two series are necessarily representative, it seems likely that for series where there is an improvement in concurrent-to-final revisions there is also an improvement in the intermediate revisions. Conversely, it seems likely that for series where there is no improvement in concurrent-to-final revisions there is no improvement in intermediate revisions either.
9 Discussion

A reduction in revisions is most desirable for the most volatile series. It is therefore satisfying to find that it is for these series that a reduction in revisions is possible using the ARIMA forecasting option in X-12 ARIMA. The reduction for these volatile series is around 15 percent, in agreement with the claims of Statistics Canada regarding the magnitude of the effect of ARIMA forecasting technique.

For ARIMA forecasting to have benefit only for the more volatile series, the revision problem introduced by the asymmetric filter must be significant only for these series. For the more stable series, the ARIMA forecasts with the fit criteria critical values set at the default levels are just not good enough to reduce the revisions.

The early failure at Statistics New Zealand to implement successfully the ARIMA forecasting option in one or two isolated cases can now probably be explained. Twenty percent of our sample series could not be fitted by one of the default ARIMA models and these initial test series could have fallen into this category. The failure to fit could have related to the nature of the series or to the presence of outliers. One aspect of the nature of New Zealand series which could plausibly cause difficulty in the fitting of ARIMA models is the observation that the variance commonly changes over relatively short time periods (a few years).

The history analysis required longer series lengths than most series in our sample, leaving a sample of just 27 series on which to analyse the effect of ARIMA forecasting on revisions. We therefore recommend that a follow-up study be carried out which treats, as a hypothesis, the findings of this study; that series with mean percentage revisions of 1.5 percent or above show reduced revisions using the ARIMA forecasting option. The series in the new sample should be long enough that the history spec can again be used. About 20 percent of the sample will not be able to be fitted by one of the default models, but the majority of the rest should show improvements in mean percentage revision.

10 Conclusions

We have seen in this study that 80 percent of the Statistics New Zealand series that we tested could be fit by one or more of the ARIMA models in X-12 ARIMA, according to that program’s fit criteria. Not all reasons for failure of the 20 percent in the sample are clear, although it is likely that in some cases outliers prevented a suitable model being found because in X-12 ARIMA these are not removed before the ARIMA modelling is performed.

Of the 80 percent of series in our sample where the ARIMA default models were able to furnish acceptable fits and forecasts, not all series that we could test demonstrated a reduction in revisions through the use of ARIMA forecasts.

Series where the mean percentage revisions without using ARIMA forecasts were already under about 1.5 percent showed no reduction and even a slight increase in mean percentage revision between concurrent and final seasonally adjusted value.

However, series where the mean percentage revisions without using ARIMA forecasts were above about 1.5 percent did show a reduction in mean percentage revision from concurrent to final seasonally adjusted value. A closer look at one such series showed that both individual concurrent estimates and their path to the final estimate tended to be better where ARIMA forecasting was used.

It therefore appears that ARIMA forecasting has the potential to reduce revisions from concurrent to final seasonally adjusted estimates, and to maintain a smooth path between these two estimates, for the majority of the more volatile Statistics New Zealand series. A further study using a different sample of volatile series should test this result. If the findings hold in that study, there is potential for ARIMA forecasting to be included in routine seasonal adjustments for those series where it is likely to be of benefit.
11 References


12 Appendix
Example .spc file used in the methodology described in Section 6.

```plaintext
series{ 
    start=1988.1
    period=4
    title="witharima model"
    file="j:\s_as\arima\quart\revisions\SER22.dat" format=datevalue
}
automdl{ 
    mode=fcst
    method=best
    file ="j:\s_as\arima\x12a.mdl"
    fcstlim = 15.0
    qlim = 5.0
    overdiff = 0.90
    outofsample = yes
}
estimate {} 
transform { 
    function=log
    mode=ratio
}
x11{ 
    mode=mult
    sigmalim=(1.8,2.8)
    print=(e2)
    save=(b1 d8 d10 d11 d12 c17)
}
history{ 
    estimates=(sadj)
    sadjlags=(1,4)
    target=concurrent
    fixmdl=no
    refresh=no
    print=(sarevisions sasummary saestimates )
    save=(sar)
}
```