Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

SOME PROPERTIES OF THE SEASONAL ADJUSTMENT DIAGNOSTICS IN X-12-ARIMA

Catherine C. Hood and Roxanne M. Feldpausch[1]

ABSTRACT

Progress in seasonal adjustment depends on the development not only of methods that better account for the various components of time series, but also the development of better diagnostics. A successful seasonal adjustment can depend as much on the diagnostics as on the methods. We will present a study to evaluate the performance of the diagnostics as they relate to the performance of X-12-ARIMA and SEATS on a large sample of real and simulated economic time series. Our goal in the study was to see if we could find something in the diagnostics of either X-12-ARIMA or SEATS that would indicate either a preference for X-12-ARIMA or SEATS—or would show that the adjustment would be poor with either program.

KEY WORDS: Seasonal Adjustment Diagnostics; Spectral Diagnostics; Time Series.

1. BACKGROUND

1.1 TRAMO/SEATS, X-12-ARIMA, and X-12-SEATS

TRAMO/SEATS and X-12-ARIMA and are two widely-used seasonal adjustment programs. TRAMO/SEATS and X-12-ARIMA are based on different methods for seasonal adjustment. TRAMO (Time series Regression with ARIMA noise, Missing observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series) are linked programs developed by Agustin Maravall and Victor Gomez to seasonally adjust time series using ARIMA-model-based signal extraction techniques. SEATS uses signal extraction with filters derived from an ARIMA-type time series model that describes the behavior of the series. This method is based on work by Hillmer and Tiao (1982) and Burman (1980), among others. See also Maravall (1993) and Gomez and Maravall (1997).

X-12-ARIMA is the U.S. Census Bureau's latest program in the X-11 line of seasonal adjustment programs. By default, X-12-ARIMA uses signal-to-noise ratios to choose from a fixed set of moving-average filters, often called X-11-type filters. X-12-ARIMA is based on the well-known X-11 program (Shiskin, Young, and Musgrave, 1967) and Statistics Canada's X-11-ARIMA and X-11-ARIMA/88 (Dagum, 1988). Major improvements in X-12-ARIMA over X-11-ARIMA/88, are discussed in Findley, Monsell, Bell, Otto, and Chen (1998).

Not only do SEATS and X-12-ARIMA have very different approaches to seasonal decomposition, the programs also have very different seasonal adjustment diagnostics. SEATS diagnostics consist mainly of model-fit diagnostics. Early versions of X-11 were criticized for being a black box—no diagnostics for the adjustments. Among the improvements introduced by Statistics Canada for X-11-ARIMA was the use of ARIMA models for forecasting and the addition of diagnostics to give users some ideas of the properties of their series and their adjustments. X-12-ARIMA includes regression models with ARIMA errors (regARIMA models) and several types of model and seasonal adjustment diagnostics, including spectral graphs and stability diagnostics, which are discussed below. The

[1]Catherine C. Hood, U.S. Census Bureau, ESMPD Time Series Methods Staff, Washington, DC, USA 20233 (catherine.c.hood@census.gov) and Roxanne M. Feldpausch, U.S. Census Bureau, ESMPD Time Series Methods Staff, Washington, DC, USA 20233 (roxanne.feldpausch@census.gov). This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Census Bureau, with permission and assistance from Agustin Maravall, also has a version of X-12-ARIMA with access to the SEATS method. The new program is temporarily named X-12-SEATS. In this paper, we use the term "X-12" when we want to refer to both X-12-ARIMA and X-12-SEATS. X-12-SEATS adjustments do not always exactly match SEATS adjustments. The main difference between the SEATS distributed by Dr. Maravall and the SEATS module of X-12-SEATS is that SEATS uses conditional likelihood for estimating AR models and X-12-SEATS uses exact maximum likelihood. So for AR models, the adjustments from SEATS and X-12-SEATS will be very similar, but not identical. X-12-SEATS is still in development and is not yet distributed by the Census Bureau. For more information, please see the paper by Monsell, Aston, and Koopmans (2003).

1.2 Judging the Accuracy of the Adjustments

How do we judge between the two adjustments? This has been a much debated question in the statistical literature; see for example Bell and Hillmer (1984).

One problem in evaluating seasonal adjustment software is that we don't know the truth—we don't know if we have the correct seasonally adjusted series. However, if we decompose a group of series into the seasonal, trend, and irregular, and recombine them to form new series, we will know what our components should be, and we can judge the accuracy of the adjustments from the created series as compared to the true seasonally adjusted series.

For the simulated series only, we judged the accuracy of the adjustments based on how close we can come to the actual seasonally adjusted series. We were able to calculate the actual seasonally adjusted series, x_t because we created the series by combining components from existing decompositions. See Section 2 for more details on how we simulated the series. For each adjustment we calculated an average absolute percent difference

$$AAPD = N^{-1} \sum_{i=1}^{N} \frac{\left| x_t - \hat{x}_t \right|}{x_t}$$

where N is the number of data points in the series, x_t is the actual seasonally adjusted series, and $\hat{X}_t$ is the estimated seasonally adjusted series.

1.3 Previous Studies

For our previous study, we chose three series with very different trends and different seasonal factors. Then we computed the components with both X-12-ARIMA and SEATS to generate six trends and six sets of seasonal factors and then recombined the irregulars to form three sets of irregulars. We didn't want to use trends and seasonal factors from the same program when we recombined them, so this gave us 54 simulated series. Please see Hood, Ashley, and Findley (2000) for more information on how we created the simulated series.

In the earlier study, we found that for many series, the adjustments from X-12 and SEATS were almost identical except that X-12-ARIMA did better with short series (monthly series with four to seven years of data). We also found that the SEATS diagnostics were inadequate to find problems with the adjustments and that SEATS needed more diagnostics before we could recommend using SEATS for production work at the Census Bureau. This led to our current study using more series.

2. METHODS

2.1 Series

We used two sets of series. The first set was 267 published series from the Census Bureau's Import/Export series. The series are ideal to use for research purposes because they show a wide range of characteristics. Some of the series are very stable, but many have a large irregular component.

We constructed a second set of series similar to the simulated series we used in the previous study (Hood, Ashley, Findley, 2000). For the simulated series, we started with trend/cycle, seasonal, and irregular components from the 267 import/export series and recombined them. We used both X-12-ARIMA and SEATS to generate the components we used. We multiplied the trend/cycle from a given series by the seasonal factors from a different series and then multiplied again by the irregular from a different series. From our original 267 series, we generated 1400 series.

2.2 Running the Programs

For consistency, we used the ARIMA model chosen automatically by TRAMO for both types of adjustments. We hardcoded the transformation choice, the ARIMA model, and trading day and Easter effects that were chosen automatically. Then for the same regARIMA model, we used X-12-SEATS to get seasonal adjustments and diagnostics from both an X-11/X-12-type adjustment and a SEATS-type adjustment, using either an x11 or a seats spec in the input file to X-12-SEATS.

In the example input file for X-12-SEATS shown below, we commented out the x11 spec (with the #), and we used a seats spec for producing the seasonal adjustment.

Example Input File for a SEATS Run

```
series{  file = 'm00020.dat'
  format = 'datevalue'
  savelog = peaks   }
transform{ function = log }
arima{ model = (0,1,1)(0,1,1) }
outlier{ types=all }
forecast{ maxlead = 24 }
check{ print = all   savelog = lbq }
#x11 {  }
seats{  }
slidingspans{ savelog=percent }
history{  estimates= (fcst aic sadj sadjchng) }
```

2.3 Diagnostics Used

For an adjustment to be acceptable, there should be no residual seasonal or calendar effect present in the seasonally adjusted series or in the irregular component. For series that are long enough, the best diagnostic to find residual seasonality or calendar effects is the spectrum, discussed below in Section 2.3.1.

We also wanted to be able to identify the series that were nonseasonal. For this we used 1) the F-test for stable seasonality (associated with the D8 Table in X-12-ARIMA) based on an ANOVA to measure the between-months variation and 2) spectral graphs of the original series to look at the series in the frequency domain.

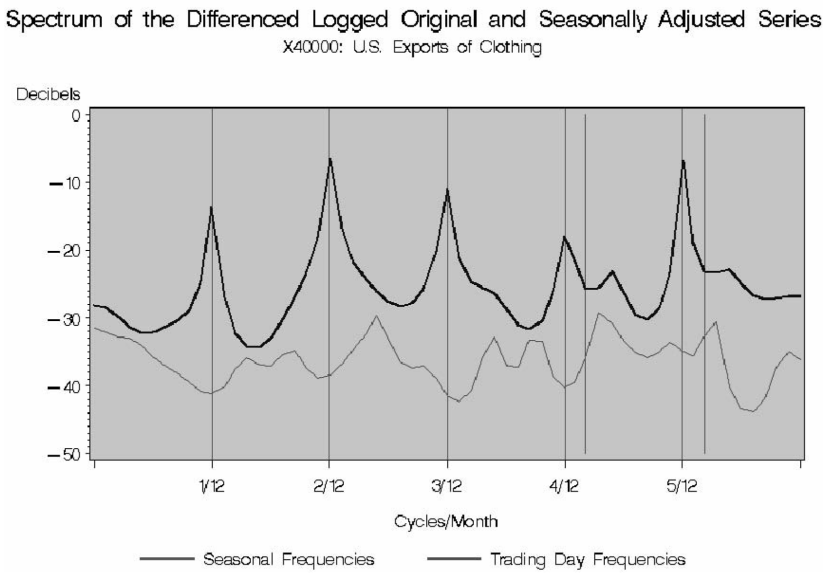
Once we knew an adjustment was acceptable, we wanted to evaluate the diagnostics in both programs as they relate to the accuracy measures. Because revisions in the seasonally adjusted series are important at the Census Bureau, we looked at the stability diagnostics in X-12: the revisions for the seasonally-adjusted series and the sliding spans diagnostics. In Section 2.3.2, we discuss the stability diagnostics in X-12. We also looked at the monitoring and quality diagnostics (Ms and Qs) available in X-11-ARIMA and X-12-ARIMA; we describe them in Section 2.3.3.

2.3.1 Spectral Graphs

X-12 automatically estimates three spectra whenever seasonal adjustment is requested: the spectrum of the differenced original series, the spectrum of the differenced seasonally adjusted series, and the spectrum of the final irregular component. Seasonal frequencies are marked by vertical lines at k/12 cycles/month for $1 \leq k \leq 5$. Trading day frequencies are marked by vertical lines at 0.348 and 0.432 cycles/month (Cleveland and Devlin, 1980). A visually significant peak at any of the seasonal or trading day frequencies for either the seasonally adjusted series or the irregular is a signal of a possible residual seasonal or trading day effect. The spectral graphs as described above are available only in X-12.

Figure 1 is an example of a spectral graph for a seasonal series. The graph shows seasonal peaks in the original series (the dark line), but the seasonal peaks are suppressed in the spectrum of the seasonally adjusted series (the lighter line below).

Figure 1. Example of Spectral Graph of the Original and Seasonally Adjusted Series



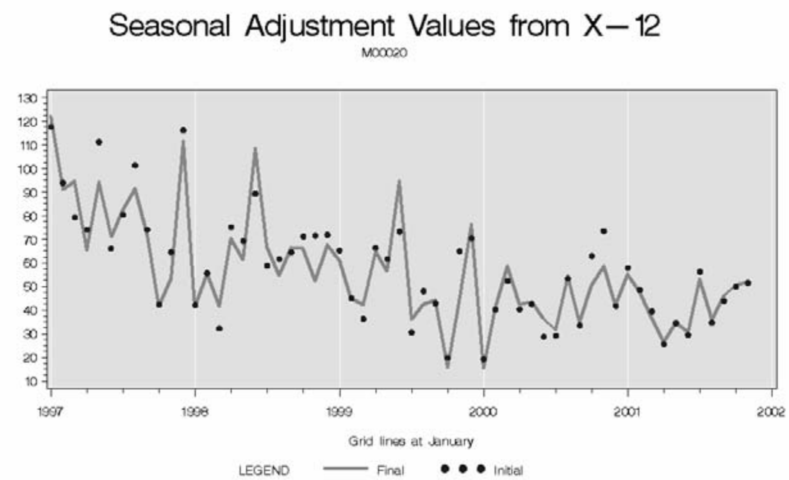
2.3.2 Stability Diagnostics

The stability diagnostics available in X-12 are available for both X-11-type and SEATS-type adjustments. For more information on the stability diagnostics in X-12, please see Findley, et al.(1998). The sliding spans diagnostics were developed at the Census Bureau (Findley, Monsell, Shulman, and Pugh, 1990).

The revision history procedure computes a sequence of runs from truncated sets of data, allowing the user to compare revisions from the initial estimate to the most recent estimate. Let X_t be a time series defined for t=1,2,...,N. Let $A_{n|T}$ be the seasonal adjustment of X for observation *n* calculated using $X_1, X_2, \ldots, X_T$, where $n \leq T \leq N$. Define $A_{n|n}$ to be the *initial* or *concurrent* seasonal adjustment — the first seasonal adjustment for observation *n*. Define $A_{n|N}$ to be the *final* or full-series seasonal adjustment — the seasonal adjustment for observation *n* including all the data up to observation N.

Revisions can be quantified by the mean and maximum absolute percent difference between the initial and final estimate for the seasonal adjustment and between the initial and final estimate of the percent change for the seasonal adjustment. We often look at graphs of the initial and final adjustments. Figure 2 below shows the final seasonally adjusted series and the initial estimates for every month from January 1997 to November 2001.

Figure 2. Example of Seasonal Adjustment Revision Graph (Final and Initial) from X-12-ARIMA



For this series, there exist several differences between the initial estimates of the seasonally adjusted series and the final estimate of the seasonally adjusted series. You will notice that the points at the end of the series agree more closely, which is to be expected, because at the end of the series, there are not as many new data points between the initial estimate and the final estimate of the seasonally adjusted series.

The purpose of the sliding spans diagnostics is to compare adjustments from overlapping subspans of the series. X-12-ARIMA looks at four spans (if there are enough data) beginning every new span one year earlier. Let X_t be a time series defined for t=1,2,...,T. The last span will end at point T, the last point in the series. For the second-to-last span, X-12-ARIMA takes one year of data off the end of the series. The length of the spans stays the same, so the starting point for this span is a year earlier than the date for the last span. If there are not enough data for four spans, the program will calculate three spans or two spans. X-12-ARIMA calculates seasonal adjustments for all spans of data separately resulting in four (or three or two) different estimates of a large number of points. X-12 looks at the maximum difference between the estimates.

The X-12-ARIMA Users Guide (U.S. Census Bureau, 2002) and the X-12-ARIMA output file contain some guidelines for the sliding spans diagnostics. It is suggested that the 75^{th}-percentile of the Maximum Percent Differences (MPD) across the spans for seasonal factors or the seasonally adjusted series should be less than 3 percent and that the 60^{th}-percentile of the MPD across the spans for month-to-month (or quarter-to-quarter) changes should be less than 3 percent. There are no guidelines for the revision histories.

2.3.3 M and Q Quality Diagnostics

The M and Q Monitoring and Quality Diagnostics were developed at Statistics Canada and included in X-11-ARIMA. They were designed to point out characteristics of the series that may cause problems with the adjustment. The 11 M statistics range from 0.0 to 3.0. The Q is a weighted average of the 11 M statistics. Any M or Q greater than 1.0 indicates a source of potential problems for the adjustment procedure. For more information on the individual statistics, please see Lothian and Morry (1978) or Ladiray and Quenneville (2001).

2.4 Steps

After we ran the program for both types of adjustments, we separated out the nonseasonal series using spectral diagnostics and F-tests. Next we looked for residual seasonality using spectral diagnostics. For the simulated series with no residual seasonality, we next checked the accuracy of the adjustments and compared our accuracy measures for the series with the stability diagnostics from X-12.

3. RESULTS

3.1 Running the Programs

We had no problems running the 267 published series through X-12. However, out of the 1400 simulated series, 175 (12.5%) series had fatal errors when run in with the SEATS method in X-12-SEATS. Eight out of the 175 series also had fatal errors because of problems with the ARIMA model chosen by TRAMO. We also found 83 (5.9%) series that needed an additive adjustment. For ease of computations, we also eliminated the additive series. This left us with 1142 series for analysis.

3.2 Nonseasonal Identification

Of the 1142 series, 357 series were not seasonal based on X-12's D8 F-test, spectral graphs of the original series, and the fact that TRAMO chose a $(0\ 0\ 0)_{12}$ seasonal model. A nonseasonal model from TRAMO is not necessarily the best way to identify nonseasonal series. TRAMO chose a $(0\ 0\ 0)_{12}$ seasonal model for 180 (50.4%) series out of the 357 that we decided were not seasonal. TRAMO did not choose a $(0\ 0\ 0)_{12}$ seasonal model for any series that were seasonal. If we had a nonseasonal model from TRAMO, we could not get a SEATS-type seasonal adjustment.

3.3 Residual Seasonality

For both the seasonal and nonseasonal simulated series, SEATS-type adjustments had more series with residual seasonality than the X11-type adjustments. We had noticed in past studies that SEATS can induce residual seasonality into the seasonally adjusted series of a nonseasonal series. So we looked for residual seasonality in the seasonally adjusted series and the irregular for all the series. As Table 1 shows, both programs can leave some series with residual seasonality, but for nonseasonal series SEATS is more likely than X-12-ARIMA to induce residual seasonality into the seasonally adjusted series. See Section 4 for more information.

Table 1. Number of Series (and Percentage of Total for the Row) with Residual Seasonal Peaks in the Seasonally Adjusted Series and/or the Irregular Component

	Total	SEATS-type	X11-type	Both
Series identified as nonseasonal	357	60 (16.8%)	7 (2.0%)	2 (0.6%)
Series identified as seasonal	785	38 (4.8%)	14 (1.8%)	1 (0.1%)
All Series	1142	98 (8.6%)	21 (1.8%)	3 (0.3%)

3.4 Accuracy

There is correlation between sliding spans and revisions diagnostics and our accuracy measure for both the X11-type and SEATS-type seasonal adjustments. Figures 3 and 4 below show the Average Absolute Percent Difference (AAPD) between the truth and the X-12-ARIMA adjustment versus either 1) the Maximum Percent Difference across the spans from the sliding span diagnostics, or 2) the Average Absolute Revision in the seasonally adjusted series from the history diagnostic. We chose a cut-off of 3% for the AAPD somewhat arbitrarily in the graphs below. Very similar results hold for SEATS-type adjustments.

Relationships between other diagnostics and the accuracy measures are not so clear. There was very little correlation between the M and Q diagnostics and our accuracy measures. Figure 5 below shows the Q Statistic. We see quite a few series with failing Q diagnostics that show very good performance according to our accuracy measure.

Figure 3. Sliding Spans Diagnostics Plotted Against Average Absolute Percent Differences (AAPD)

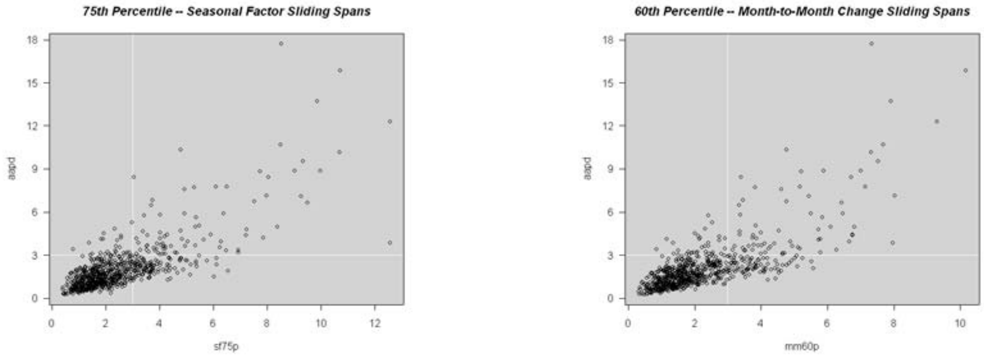


Figure 4. Revisions Diagnostics Plotted Against Average Absolute Percent Differences (AAPD)

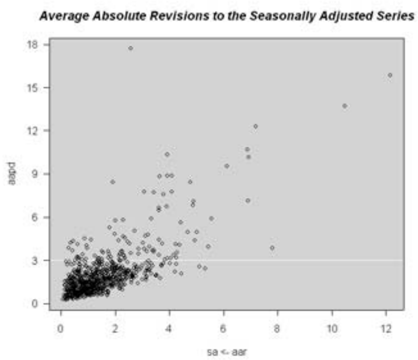
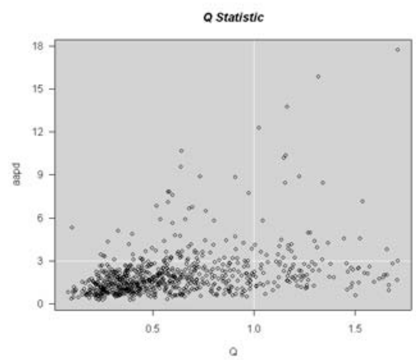


Figure 5. Q Statistics Plotted Against Average Absolute Percent Differences (AAPD)



4. EXAMPLES

4.1 Residual Seasonality

As we saw in Table 1, SEATS can induce residual seasonality into the seasonally adjusted series when it is adjusting a series that should not be adjusted. Spectral diagnostics for the original series and the seasonally adjusted series would help to avoid this problem. In Figure 6 note the seasonal peak at 1/12 in the seasonally adjusted series from SEATS where there is no seasonal peak in the original series. Figure 7 shows a similar spectral graph from an X-11-type of adjustment for comparison.

In many cases, this problem occurred when TRAMO chose a seasonal component in the ARIMA model for a nonseasonal series, and then SEATS changed the model. In most of the examples we have found, TRAMO chose a seasonal AR(1) model for a series with very little or no seasonality and SEATS changed the model to a seasonal IMA(1,1), using a seasonal filter for a nonseasonal series.

Figure 6. Spectral Graph of the Original and Seasonally Adjusted Series from a SEATS Adjustment

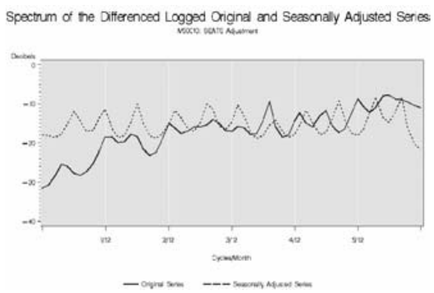
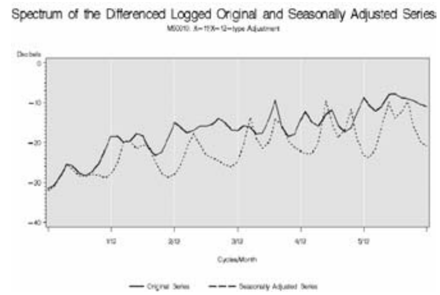


Figure 7. Spectral Graph of the Original and Seasonally Adjusted Series from an X11-type Adjustment



SEATS diagnostics for "under-adjustment" are not adequate to find residual seasonality. One problem is that the diagnostics are biased, so it is difficult to see problems. For more information, see Findley, Wills, Aston, Feldpausch, and Hood (2003).

4.2 Revisions

There were several series where the AAPD for the SEATS adjustment was much larger than for the X-12 adjustment. For all of these series, SEATS would change the model from the one hardcoded into the program—the model chosen by TRAMO. This result holds true in both X-12-SEATS and SEATS (2000 and 2002 versions). Below is an example that demonstrates the problem.

SEATS prefers balanced models, meaning models where the total AR order, including differencing, equals the total MA order. For a grain exports series, X00GRN, TRAMO chose an ARIMA(1 0 0)(1 0 0)$_{12}$ model with a constant, trading day, and Easter regressors. Even though TRAMO's regARIMA model was hardcoded into SEATS, SEATS would sometimes change the model, with very unstable results. With a model like the AR model for X00GRN, depending only on the span of data we used, SEATS would sometimes use the given model, and sometimes SEATS used different models, including an airline model. Figure 8 below shows the revisions for the default SEATS adjustment. In Figure 9, for all the dates circled, SEATS changed the model to an Airline model – (0 1 1)(0 1 1).

Figure 8. Seasonal Adjustment Revision Graph (Final and Initial) from A Default SEATS Adjustment

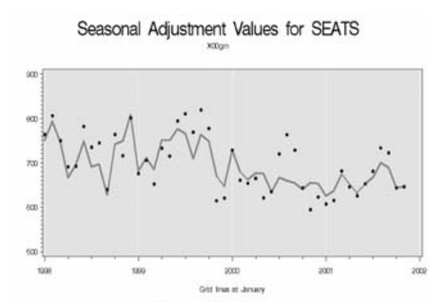
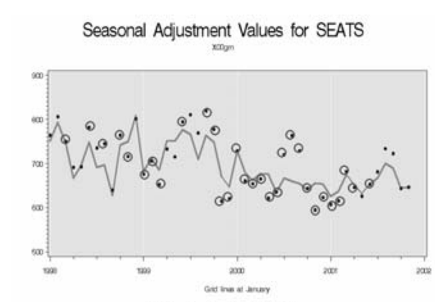


Figure 9. Seasonal Adjustment Revision Graph with Airline Model Marked with Circles



If we set the model for the SEATS adjustment to the airline model, the revisions improve dramatically as shown in Figure 10. With an airline model, instead of the default model from TRAMO, SEATS can find an admissible decomposition and the revisions are smaller. The revisions are still larger than the revisions from the default X-11-type adjustment, shown in Figure 11.

Figure 10. Seasonal Adjustment Revision Graph from a SEATS-type Adjustment Using the Airline Model (Nondefault)

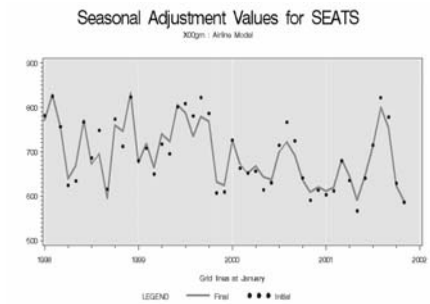
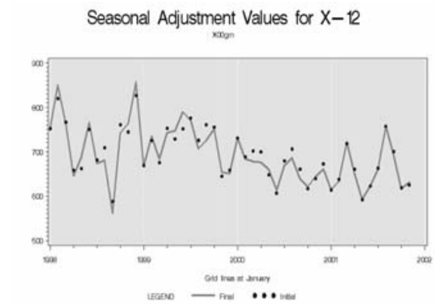


Figure 11. Seasonal Adjustment Revision Graph from a Default X-11-type Adjustment



5. CONCLUSION

Diagnostics in X-12 are essential to see problems in seasonal adjustments. The sliding spans and revisions diagnostics available in X-12 can point to series that possibly should not be adjusted because both X-12-ARIMA and SEATS cannot separate the components.

For the next part of our study, we are going to look at more series. We are working on simulating components using ARIMA models. In theory, this should give an advantage to SEATS, but we want to see how both a SEATS-type adjustment and an X11-type adjustment perform. For all of our series, we want to look at more diagnostics and characteristics of the series to try to find patterns that will help us decide between SEATS-type and X11-type adjustments in the future.

ACKNOWLEDGMENTS

We thank David Findley for his advice and encouragement on all the SEATS and X-12-ARIMA studies through the years, and we thank Agustin Maravall for his advice and instruction on how to use TRAMO/SEATS. We are also grateful to James Ashley for help with earlier studies on TRAMO/SEATS, to Brian Monsell and Gianluca Capporella for writing the program X-12-SEATS, and to Brian for software support. Thanks also to David, Brian, Kathleen McDonald-Johnson, and others who gave comments on this paper.

REFERENCES

Bell, W.R. and S.C. Hillmer (1984), "Issues Involved with the Seasonal Adjustment of Economic Time Series" (with discussion), *Journal of Business and Economic Statistics*, 2, pp. 291-349.

Burman, J.P. (1980), "Seasonal Adjustment by Signal Extraction," *Journal of the Royal Statistical Society*, Ser. A, 143, pp. 321-337.

Cleveland, W.S. and S.J. Devlin (1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods," *Journal of the American Statistical Association*, 75, pp. 487-496.

Dagum, E.B. (1988), "X-11-ARIMA/88 Seasonal Adjustment Method - Foundations and Users' Manual," Statistics Canada.

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto and B.-C. Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 16, pp. 127-176.

Findley, D.F., B.C. Monsell, H.B. Shulman, and M.G. Pugh (1990), "Sliding Spans Diagnostics for Seasonal and Related Adjustments," *Journal of the American Statistical Association*, 85, pp.345-355.

Findley, D.F., K.C. Wills, J.A. Aston, R.M. Feldpausch, C.C. Hood (2003), "Diagnostics for ARIMA-Model-Based Seasonal Adjustment," to be published in *Proceedings of the American Statistical Association, American Statistical Association*.

Gomez, V. and A. Maravall (1997), "Program TRAMO and SEATS: Instructions for the User, Beta Version," Banco de Espana.

Hillmer, S.C. and G.C. Tiao (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, 77, pp. 63-70.

Hood, C.C., J.D. Ashley, and D.F. Findley (2000), "An Empirical Evaluation of the Performance of TRAMO/SEATS on Simulated Series," *Proceedings of the American Statistical Association, Business and Economic Statistics Section, American Statistical Association*, pp. 171-176.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method.* Springer-Verlag: New York, NY.

Lothian, J. and M. Morry (1978), "A Test of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Program," Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.

Maravall, A. (1993), "Unobserved Components in Economic Time Series," *Handbook of Applied Econometrics*, ed. M.H. Pesaran, T. Schmidt, M. Wickens, Oxford Basil Blackwell.

Monsell, B. C., J. A. Aston and S. J. Koopmans (2003) ”Towards X-13?”, to be published in *Proceedings of the American Statistical Association, American Statistical Association.*

Shiskin, J., A.H. Young and J.C. Musgrave (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program, " Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce.

U.S. Census Bureau (2002), *X-12-ARIMA Reference Manual, Final Version 0.2,* Washington, DC: U.S. Census Bureau.