

Assessment of Diagnostics for the Presence of Seasonality

Catherine C. Harvill Hood
Catherine Hood Consulting, USA
Email: cath@catherinechhood.net

Abstract

An important criterion when assessing the quality of a seasonally adjusted series is the presence of residual seasonality. X-12-ARIMA automatically checks for residual seasonality using spectral diagnostics which are generally considered reliable if there are at least 60 points in the series. However, this is longer than most quarterly series. This paper uses a large sample of both real and simulated quarterly and monthly series to look at various options available in X-12 for spectral estimation and other diagnostics that are sometimes used to detect seasonality. The spectral diagnostics worked well for short quarterly series, especially with some of the newer options available in X-12.

Keywords: seasonal adjustment, spectral graphs

1. Introduction

One of the most important criteria when assessing the quality of a seasonally adjusted series, and more generally of a seasonal adjustment method, is the presence of residual seasonality. Residual seasonality can result from inadequacies in the adjustment procedures or ARIMA models chosen, and also possibly from seasonal effects that are difficult to estimate, such as series with highly variable effects. With indirectly adjusted aggregate series, residual seasonality can result when some of the component series are inadequately adjusted. Testing for the presence of seasonality is also useful when looking at raw data to see if seasonal adjustment is even necessary.

Not only is it an important issue, residual seasonality in the seasonally adjusted series or the irregular component can almost always be eliminated by using different seasonal adjustment options. In most cases, residual seasonality can be removed by shortening the span of the data used for seasonal adjustment or by changing the seasonal filters. For very long series, the same seasonal filter or model may not be able to adequately estimate the seasonality through the entire length of the series.

So far, X-12-ARIMA (U.S. Census Bureau, 2005) is the only software that automatically checks for residual seasonality. The test is done using spectral diagnostics, but spectral diagnostics could be used for any seasonal adjustment method. Without X-12 or X-13, a basic, but potentially costly, test to check for residual seasonality is to simply rerun a seasonal adjustment. If the software detects some seasonality, then there is a problem in the adjusted series.

Though the spectral diagnostics in X-12-ARIMA are very useful, they are generally considered reliable if there are at least 60 points in the series. For quarterly series, this would mean at least 15 years of data. Many economic time series do not have 15 years of data. Therefore, estimating residual seasonality for quarterly series presents a challenge. The default estimator for the spectrum in the currently released version of X-12-ARIMA is the 30th-order autoregressive, AR(30), model. This paper looks at using the 10th-order autoregressive, AR(10), model instead. Generally the AR(10) model would give too many false peaks for most monthly series, but could be practical for use in quarterly series.

Because it is a common, but not entirely recommended, practice to use the M & Q diagnostics available in X-11-ARIMA/X-12 for the detection of residual seasonality, I also looked at these diagnostics. Another possible diagnostic for the detection of residual seasonality is a chi-square test on a set of fixed seasonal effects (seasonal dummy regressors). This is similar to the test in SEATS (Gomez and Maravall, 1997) where SEATS looks at the significance of individual months using individual t-test values (Maravall, 2004). These diagnostics are dependent on the ARIMA model used, so I investigated these diagnostics under many different ARIMA models.

2. Descriptions of the Diagnostics

2.1 Spectral Graphs

Spectral graphs of the detrended seasonally adjusted series and the detrended irregular component are included automatically with X-12-ARIMA output as the primary method to detect residual seasonality. For series that are not too short, the spectral graph is the most sensitive of all the tests for residual seasonality.

Any time series can be analyzed from two different points of view: time and frequency. The two approaches are complementary, giving us two different views of the same information. To analyze a series in the frequency domain, we measure the strength of the different frequencies in decibels. The graph of the frequencies versus the decibels is called the periodogram or the spectrum. Spectral analysis allows us to see the relationships between the frequencies. We can quantify the importance of certain frequencies of interest relative to frequencies of other components.

For a typical monthly economic time series with a significant seasonal component, there can be several different kinds of effects, for example, there could be a quarterly effect, a biannual effect, and an annual effect in the same series. However, for quarterly series, the possibilities are much smaller because of the timing of the data collection. In a quarterly time series with a significant seasonal component, the amplitudes that dominate the spectral graph are amplitudes associated with components that repeat every year (i.e., four quarters) or every two quarters. In X-12-ARIMA, the seasonal frequencies at $\frac{1}{4}$ cycles per quarter (an annual effect) and $\frac{1}{2}$ cycles per quarter (a biannual effect) are marked in the graphs. If the series is seasonal, there would be peaks at one or both of the two seasonal frequencies.

In a spectral graph of a time series, the low frequencies at the left of the graph correspond to slowly changing components, like the trend, while the higher frequencies correspond to rapidly changing components, like the irregular. If a series has strong long-term trend movements, the low frequencies associated with the long-term trend movements can have amplitudes that dominate the spectrum. Therefore, some kind of prior detrending, such as differencing, is required before the spectrum calculations so we can see clearly the amplitudes at the seasonal frequencies.

To have a spectral estimator with enough resolution so that peaks can be sharply defined, X-12-ARIMA, by default, uses an AR(30) model fit to each output series and evaluated at 61 frequencies between 0 and $\frac{1}{2}$ cycles/quarter. In the latest version of X-12-ARIMA (Version 0.3), the order of the AR model for the spectral estimate can be changed from the default.

The rule for a seasonal peak being flagged as “visually significant” by X-12-ARIMA is that the value at the seasonal frequencies must be six “stars” or asterisks higher than either neighboring frequency in the plot found in the output file of X-12-ARIMA. (For the plot, the range of the frequencies is divided in 52 parts, so each asterisk in the plot equals $\frac{1}{52}$ nd of the range.) The higher the peak is above its neighbors, the more important the peak. In addition to the “six-star” rule, a seasonal peak also must be higher than the median of the frequencies in the graphs for X-12-ARIMA to flag the peak. For more information on how spectrum peaks are flagged, see Soukup and Findley (1999).

For more information on how the spectral diagnostics are used in both SEATS and X-12-ARIMA, see the papers by Maravall (2004) and Findley and Hood (2000).

2.2 D8 F-test and M7 of X-12-ARIMA

Included after Table D8 in the X-12-ARIMA output are some ANOVA tables to help measure stable and moving seasonality. Generally for F-tests, one would use a cut-off value of 4.0, but because we have generally correlated data in the time series, a cut-off value of 7.0 is recommended by the authors of X-11-ARIMA (Lothian and Morry, 1978).

The first ANOVA table given tests for the “presence of seasonality assuming stability” using the variation between months. The F-value from this table is referred to in Table F2.I as the “F-test for stable seasonality from Table D 8” and so has come to be known as the “D8 F-test”. It is also sometimes referenced as FS, the F-value to test for seasonality. A second ANOVA table after Table D 8 labeled as the “moving seasonality test.” It is sometimes referenced as FM, the F-value to test for moving seasonality. M7 is a descriptive statistic based on the FS (the D8 F-test) and FM. M7 was designed by the authors of X-11-ARIMA to determine whether seasonality can or cannot be

identified by X-11 (Lothian and Morry, 1978). Its interpretation is straightforward: a value greater than 1.0 indicates no identifiable seasonality.

It is a common misconception that the M7 and D8 F-test available in X-12 are for the detection of residual seasonality. Both diagnostics were meant to help the user decide if the original series is seasonal, and some users run the seasonally adjusted series back through X-12-ARIMA to look at the values for M7 and the D8 F, though the M7 diagnostic was not designed for this purpose. Also, the M7 is susceptible to the model used and to the type of adjustment used by X-12-ARIMA, so to be used for checking residual seasonality, it would require careful modeling of the seasonally adjusted series. The D8 F-test is somewhat more stable than the M7 diagnostic, but is still dependent on the extreme values chosen in the X11 procedure.

2.3 Other Tests

In SEATS, Agustín Maravall has used a significance test on the individual months or quarters to determine if there is significant seasonality in the original series, to warrant seasonal adjustment. A test of the individual months could also be used to test for residual seasonality. X-12-ARIMA also contains built-in seasonal dummy variables. Produced as a result of the regression are t-tests on the individual months or quarters and a chi-square test to test for the significant of the set of seasonal regressors. Though not used commonly by X-12-ARIMA users, both the t-tests and chi-square values could be used. These tests also have the potential to be susceptible to changes in the model, and would require modeling of the already adjusted series. Therefore, research is required to test the stability of the diagnostic against changes in the model.

Agustín Maravall in his 2004 paper suggests the use of the Kendall-Ord nonparametric test (Kendall and Ord, 1990) and mentions some success with the diagnostic. However, nonparametric tests of this sort usually are not feasible for short or quarterly series.

3. Methods

3.1 Running the Series

I began the research with 314 seasonal quarterly economic time series that were long enough (more than four years) to run in X-12-ARIMA and SEATS. I ran all 314 series through X-12-ARIMA using seasonal dummy regressors along with the following nonseasonal ARIMA models: (2 1 0), (0 1 2), (0 1 0), (0 0 0), and the results from the Version 0.3 automatic modeling selection including the regressors. For comparison, I also looked at the results using the model from the automatic modeling procedure using no regressors. For all 6 models above, I used both the AR(30) spectral estimator and the AR(10) spectral estimator. I looked at spectral results using both the outlier-adjusted original series and the prior-adjusted original series.

I collected the spectrum peak “star” values, M7, Q2, D8 F-value, and the Chi-square p-value for the seasonal regressors. Though I did look at spectral graphs for the series, I depended more on the number of stars listed in the diagnostic output file. I used Excel for most of the analysis, including X-12-Data (Feldpausch, 2003) and the Excel version of X-12-Rvw.

3.2 Simulated Series

To test the sensitivity of the diagnostics on various levels of residual seasonality, I constructed simulated series from a set of known trends, seasonal components and irregular components (Hood, Ashley, and Findley, 2000). This allows control of the various levels of each component, so I constructed some series with strong seasonality, some with no seasonality, and some with very little seasonality similar to the level of seasonality that we’d like to detect when looking for residual seasonality. The impact of changes in the models and options for the diagnostics can then be compared to what we know is the correct answer for the given series.

The focus for this part of the study was on the spectral diagnostics along with the tests for the seasonal regressors because these will be the diagnostics most useful generally. For example, applying the M7 diagnostic designed for the raw, original series in X-12-ARIMA, to a SEATS adjustment seemed a bit dubious.

4. Results

4.1 Results for the Real Series

For seasonal series, we would expect to see peaks at the seasonal frequencies. Using the AR(30) estimator with the series, because there are less than 60 points, the peaks were often very short and flat and difficult to distinguish. The peaks are much easier to distinguish using the AR(10) estimator. Assuming that the AR order is held constant, the results for the outlier-adjusted series and the prior-adjusted series were almost identical, so I focused on the prior-adjusted spectrum.

The results in the table below show there is also an effect from the model used. For the (0 0 0) model, which would generally be inappropriate for these series, the outlier sets chosen automatically by the program were not ideal, and the spectrum results generally suffered. The peaks are labelled as either S1 (meaning $\frac{1}{4}$ cycles per quarter) or S2 (meaning $\frac{1}{2}$ cycles per quarter). The ARIMA models given in Table 1 are the nonseasonal part of the model. In this table and in all subsequent tables, the models include fixed seasonal regressors for the seasonal part of the model.

Table 1. Results for Spectral Peaks

ARIMA model *	Spectral AR order	Peak	# >6 stars	# 5-6 stars	# 0-5 stars	Average Stars (total # of peaks)
(2 1 0)	10	S1	71	21	95	5.3 (187)
		S2	98	12	133	5.4 (243)
(2 1 0)	30	S1	9	9	81	2.6 (99)
		S2	29	6	107	3.9 (142)
(0 0 0)	10	S1	55	9	112	4.1 (176)
		S2	69	14	148	4.7 (231)
(0 0 0)	30	S1	3	10	94	2.2 (107)
		S2	28	3	111	3.6 (142)

* Nonseasonal part of the ARIMA model. Fixed seasonal regressors used for all runs.

The peak information given in Table 1 is not mutually exclusive, meaning that some series would have had a peak for both S1 and S2 and some had only 1 peak. Results from the other models, i.e., (0 1 2) and (0 1 0), were similar to the results given above using (2 1 0).

All of the regression- and ANOVA-based diagnostics (M7, D8 F-test, and Chi-square test on the seasonal regressors), as with the spectrum, are dependent on the model chosen. The results for any of the diagnostics were inadequate using seasonal regressors and the (0 0 0) model, which I knew would be a bad model for seasonal series.

For these series, I expected the diagnostics would say the series were seasonal. I looked first at the number of the 314 seasonal series that would be rejected as being nonseasonal with the various models. It can be seen clearly in Table 2 below that a bad model, (0 0 0), affects all the diagnostics. There were fewer false rejections with the Chi-square test and the D8 F-test using a cut-off value of 4.0.

Table 2. Results for M7, D8 F-test, and Chi-square Test

ARIMA Models	# of Series Chosen as Nonseasonal			Chi-Sq p>5%
	M7>1	D8F<7	D8F<4	
(2 1 0)	28	41	16	13
(2 1 2)	37	45	18	21
(0 1 2)	33	43	17	13
(0 1 0)	36	43	23	31
(0 0 0)	210	200	154	236

I also counted the number of series that were in disagreement between the diagnostics were using different models, excluding the (0 0 0) model. For most series and most models, there was very little disagreement as to whether or not a series was seasonal. An example of one such comparison is given in Table 3.

Table 3. Results for M7, D8 F-test, and Chi-square Test

Models for Comparisons	# of Series in Disagreement			
	M7	D8F>7	D8F>4	Chi-Sq
(2 1 0) & (2 1 2)	14	16	3	16
(2 1 0) & (0 1 2)	10	6	3	16
(2 1 0) & (0 1 0)	12	14	6	24

Though I ran all models for all series, in the tables that follow I will only list the (2 1 0) model as an example of the performance of the diagnostics under nonseasonal model that contains a first difference.

4.2 Results for the Simulated Series

4.2.1 Results for AR(10) and AR(30) Spectral Estimators

For the series with strong seasonality, the AR(10) estimate in the spectrum gave us the best results, finding that all 144 series were in fact seasonal. Even with seasonal factors on the order of 50% to 150%, the AR(30) estimate did not always identify the seasonality present.

Table 4. Results for AR(10) and AR(30) estimators in series with strong seasonality

Models	AR	# of seasonal series	# of series with visually significant peaks
(2 1 0)	10	144	144 (100%)
(2 1 0)	30	144	134 (93%)
(0 0 0)	10	144	144 (100%)
(0 0 0)	30	144	122 (85%)

For the series with no seasonality, it was somewhat expected that there would be more false positives with the AR(30) estimator. In fact, the opposite was true. Therefore, for the rest of the study I focused on the AR(10) estimator.

Table 5. Results for AR(10) and AR(30) estimators in series with no seasonality

Models	AR	# of seasonal series	# of series with visually significant peaks
(2 1 0)	10	0	57 (40%)
(2 1 0)	30	0	65 (45%)
(0 0 0)	10	0	51 (35%)
(0 0 0)	30	0	72 (50%)

4.2.2 Comparison of the Diagnostics

Comparing the spectral diagnostics to the D8 F-test and the Chi-square test on the seasonal regressors, we can see that the model does have some effect on the results. For the strongly seasonal series, with a (0 0 0) model plus seasonal regressors model, the D8 F-test and the Chi-square test both have trouble identifying the seasonality that is present, as shown in Table 6 below.

Table 6. Results for series with strong seasonality

Diagnostic	Model used	
	(2 1 0)	(0 0 0)
D8 F-test	144 (100%)	49 (34%)
Chi-square	144 (100%)	0 (0%)
AR10 Spectrum	144 (100%)	144 (100%)

For the series with no seasonality, using a model with AR or MA terms, such as the (2 1 0) model plus the seasonal regressors, causes some of the series to be identified as seasonal. As seen in both Tables 6 and 7, the spectrum is less susceptible to the model than the other diagnostics.

Table 7. Results for series with no seasonality

Diagnostic	Model used	
	(2 1 0)	(0 0 0)
D8 F	56 (39%)	0 (0%)
Chi-sq	90 (62%)	0 (0%)
AR10 Spectrum	57 (40%)	51 (35%)

For the series with weak seasonality, I had hoped that all the diagnostics would identify the seasonality. As seen in Table 8, with a (2 1 0) plus seasonal regressors model, the D8 F-test and Chi-square test did a reasonably good job, but failed to find any seasonality using a (0 0 0) plus seasonal regressors model. Therefore, the success or failure of these tests in finding residual seasonal would depend on the model selection. For the spectrum diagnostics, it was not so susceptible to the model, but it did not identify the weak seasonality as often as I had hoped, though still finding seasonality in 60% of the series.

Table 8. Results for series with no seasonality

Diagnostic	Model used	
	(2 1 0)	(0 0 0)
D8 F	111 (77%)	0 (0%)
Chi-sq	124 (86%)	0 (0%)
AR10 Spectrum	86 (60%)	86 (60%)

4.2.2 Monthly Series

There was also a very limited study on some simulated monthly series. Using the AR(10) estimator for the spectrum, I did have quite a few more false positives, as expected, for nonseasonal series.

5. Conclusions

For quarterly series with six to 15 years of data, an AR(10) estimator for the spectrum gives both fewer false positives (peaks for nonseasonal series) and more true positives (peaks for seasonal series) than the default AR(30) estimator. Therefore, use of the AR(10) estimator should be considered for use on quarterly series.

There has been some research on the “six star” rule for monthly series. Similar research might be useful for AR(10) estimators with quarterly series.

I also found the diagnostics were fairly stable between various ARIMA models. However, there can be a lot of disagreement in the diagnostics when comparing runs using an inappropriate model. The practice of using M7 and the D8 F-test to look for residual seasonality in seasonally adjusted series is very much dependent on the model choice, and the practice should be discouraged.

Acknowledgements

This author wishes to thank the following people for their invaluable assistance on this project: Roxanne Feldpausch of the U.S. Census Bureau for her programming assistance and for providing updated copies of X-12-Data, Kathy McDonald-Johnson of the U.S. Census Bureau for her use of the program X-12-Rvw and in programming assistance in working with diagnostic files and Excel, Brian Monsell of the U.S. Census Bureau for providing the necessary modifications to X-12-ARIMA for testing the various spectrum options, David Findley of the U.S. Census Bureau for his advice and support, and Leo Harvill formerly of the Quillen College of Medicine for his support on this research project and other research through the years. Thank you.

References

- Gómez, V. and A. Maravall (1997), *Program TRAMO and SEATS: Instructions for the User, Beta Version*, Banco de España.
- Feldpausch, R. (2003), "X-12-Data: A Program to Convert Excel Spreadsheet Data in a Format X-12 Can Handle," Working paper, U.S. Census Bureau.
- Findley, D.F. and C.C.H. Hood (2000), "X-12-ARIMA and Its Application to Some Italian Indicator Series," *Seasonal Adjustment Procedures – Experiences and Perspectives*, Istituto Nazionale di Statistica, Rome, 231-251.
- Hood, C.C.H., J.D. Ashley, and D.F. Findley (2000), "An Empirical Evaluation of the Performance of TRAMO/SEATS on Simulated Series," in *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 171-176. Alexandria VA: American Statistical Association.
- Kendall, M. and J.K. Ord (1990). *Time Series, 3rd edition*. Oxford University Press: London.
- 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. (2004), "An Application of the TRAMO-SEATS Automatic procedure: Direct versus Indirect Adjustment", Banco de España working paper.
- Soukup, R.J. and D.F. Findley (1999), "On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects after Modeling or Adjustment," *Proceedings of the Business and Economics Section*, Alexandria, VA: American Statistical Association.
- U.S. Census Bureau (2005), *X-12-ARIMA Reference Manual, Version 0.3*. Washington DC: U.S. Census Bureau, U.S. Department of Commerce.