On Seasonality: Comparing X-13ARIMA-SEATS Diagnostics for Quarterly Series

Demetra Lytras
U.S. Census Bureau

Abstract
X-13ARIMA-SEATS has several diagnostics for determining whether an original time series is seasonal and for detecting residual seasonality in an adjusted series. However, seasonality diagnostics are more limited for quarterly series, as the spectrum diagnostic is available only for monthly series. Familiar diagnostics that are available for quarterly series either apply only to X-11 adjustments and not to SEATS adjustments or require an additional run. X-13ARIMA-SEATS's newest seasonality diagnostics are the QS statistics, which were first available in the Bank of Spain's SEATS software. They measure positive seasonal autocorrelation in the original and adjusted series and are available for both monthly and quarterly, X-11 and SEATS adjustments. This paper compares their efficacy in finding seasonality to that of the spectrum, the F-test of seasonal regressors, and various X-11 diagnostics using both real U.S. Census Bureau series and simulated seasonal and nonseasonal series.

Keywords: seasonal adjustment, X-11, SEATS, time series

1. Introduction

X-13ARIMA-SEATS has diagnostics both for determining whether a series is seasonal and whether an adjustment has residual seasonality. However, as the spectrum diagnostic is no longer available for quarterly series in X-13ARIMA-SEATS, seasonality diagnostics for quarterly series are more limited. The QS statistics, originally created for use in the TRAMO-SEATS program (Gómez and Maravall 1996), were added to the program partly to address this limitation.

In Lytras, Feldpausch, and Bell (2007), the efficacy of various seasonality diagnostics available in X-13ARIMA-SEATS was studied using simulated monthly time series. In this paper, the same diagnostics (the F-test for seasonal regressors, the X-11 diagnostics D8F and M7, and the spectrum) are studied on simulated and real quarterly series. This study also extends this investigation into residual seasonality diagnostics and the new QS statistics.

2. Diagnostics

2.1 QS Statistics

The QS diagnostics were developed by Agustin Maravall, then at the Bank of Spain, for the TRAMO-SEATS program (Gómez and Maravall 1996). They were added to X-13ARIMA-SEATS with the 2013 release.

The QS statistic looks for positive seasonal autocorrelation in a series to test the hypothesis that there is no seasonality in the series. To calculate the QS, first an appropriate order of differencing is chosen for the series. For the irregular component, the differencing order is set to zero. For the original or seasonally adjusted series, the differencing is first set to max(1, min(d + D, 2)), where d is the order of nonseasonal differencing and D the order of seasonal differencing in the ARIMA model used for the series. The differencing is set to 1 if there is no ARIMA model. If the differencing is 1, a test is performed to check whether the autocorrelations decay too slowly. For quarterly series, first the autocorrelations of the first-differenced series are calculated. If the autocorrelations at lags 1 through 4 are all greater than 0.2, then the differencing order is reset to 2.

With the differencing order decided, the quarterly QS is derived by first calculating the autocorrelation at lags 4 and 8 from the appropriately differenced series. If the autocorrelation at lag 4 is less than or equal to zero, then QS = 0. Otherwise,

$$QS = n(n+2)\left\{\frac{r_4^2}{n-4} + \frac{\max(0, r_8)^2}{n-8}\right\}$$

where r_i is the autocorrelation at lag i and n is the series length minus the order of differencing. QS is approximated by a chi-squared distribution with two degrees of freedom, according to Maravall (2012).

In X-13ARIMA-SEATS, seven QS diagnostics are given. These are listed in Table 1, along with the series on which these are calculated for an X-11 adjustment. X-13ARIMA-SEATS calculates these statistics using the full span of data and again on the span of data on which the spectrum is calculated, which by default is 96 observations. So for monthly series these statistics are given on the last eight years of data as well as on the full span, but for quarterly series they are given on the full span and the last 24 years. In this paper, the QS are calculated also on the last eight years, to see how the quarterly series behave on a shortened span. X-13ARIMA-SEATS calls these QSS in some of its output; in this paper they are referred to as QS8, to distinguish them from the Quarterly Services Survey discussed in Section 4.

Table 1: The QS statistics and the series on which they are calculated

QS statistic	**Series on which it is calculated**
Original Series	A1: original series
Original Adjusted for Extremes	B1: prior-adjusted series (original adjusted for outliers, trading day, holidays)
Seasonally Adjusted Series	D11: seasonally adjusted series
Seas. Adj. Adjusted for Extremes	E2: seasonally adjusted series, but with the trend estimate replacing values identified as extreme
Irregular	D13 with outlier adjustment: irregular series with additive outliers and temporary changes removed
Irregular Adjusted for Extremes	E3: irregular component, but with values identified as extreme replaced with 0 (additive) or 1 (multiplicative)
Residuals	Residual component

2.2 X-11 Seasonality Tests

The M7, D8F, and D11F statistics are outputs of the X-11 seasonal adjustment method; they have no corollary in a SEATS adjustment. The D8F test for stable seasonality tests whether all quarterly means are equal using a detrended version of the prior adjusted series. M7 brings together the D8F tests for stable and moving seasonality to test for what the program refers to as "identifiable" seasonality. For more on these statistics, see Lytras et al. (2007) or Ladiray and Quenneville (2001). In this paper we use the traditional cut-off values of these statistics; a series is thought to be seasonal if the D8F statistic is greater than 7 or M7 is less than 1.

The D11F test checks the seasonally adjusted series for residual seasonality. Like the D8F, it is a one-way ANOVA test with a null hypothesis of equality of all quarterly means. The test is calculated on the full series and again on the last three years, with a warning that this last test may have invalid results if the series level changes rapidly over the last three years.

2.3 Seasonal Regressors

When X-13ARIMA-SEATS fits a regARIMA model with fixed seasonal effects, it outputs an F-test for whether the parameters are all collectively zero. Lytras et al. (2007) shows the calculation of this statistic and looks at how it performs in identifying seasonality in nonseasonal and seasonal simulated monthly series. It was shown to have a consistent significance level when the correct ARIMA model was specified and high power, but when the model was incorrectly specified the significance level became somewhat distorted. This paper investigates the properties of the seasonal regressors on quarterly series when the model is either specified correctly or automatically chosen.

Findley and Lytras (2014) suggested that seasonal regressors could be used to check for residual seasonality by running the seasonally adjusted series through X-13ARIMA-SEATS with seasonal regressors specified. As seasonal adjustment should have removed a mean effect from the overall series, this test is best performed on a subspan of the series.

2.4 Spectrum

Lytras et al. (2007) describes the spectrum diagnostic and its application to monthly series. For a quarterly series, there is only one seasonal frequency, 0.25, to check for a peak. With the release of X-13ARIMA-SEATS the spectrum diagnostic became restricted to monthly series.

The program calculated the spectrum on the differenced, transformed prior-adjusted series to look for seasonality, and on the seasonally adjusted series and irregular to look for residual seasonality. For this study we ran all series in X-12-ARIMA, the predecessor to X-13ARIMA-SEATS, to find spectrum information for these quarterly series. The traditional "visually significant" (v.s.) cut-off is used to determine if the spectrum indicates seasonality: a peak is v.s. if it is greater than the median of all 60 measured frequencies and higher than both its nearest neighbors (those frequencies 1/120th away) by at least 6 "stars," or 6/52nds of the range.

3. Results on Simulated Series

3.1 Nonseasonal Series

Ten and twenty year series were simulated from (1 1 0) and (0 1 1) models, with the parameter, φ or θ, having values of 0.3, 0.6, or 0.9. For each combination, 1,000 series were simulated, for a total of 12,000 series. The series were run in X-13ARIMA-SEATS four times: with the correct model specified, with the automodel procedure selecting the model, with seasonal regressors added to the correct model, and with seasonal regressors and automodel. These last two runs were used only for the results of the F-test for seasonal regressors; X-11 seasonality test and QS statistic results come from the first two runs. Also, since X-13ARIMA-SEATS does not include spectrum plots for quarterly series, X-12-ARIMA was used to get spectrum results, with the correct model specified. Results from the seasonality tests are given in Table 2 for (1 1 0) series and Table 3 for (0 1 1) series.

Lytras et al. (2007) showed significance levels of M7, D8F, the spectrum, and the F-test of seasonal regressors for monthly 20-year series when the correct model was specified. There are some large differences in the results for these quarterly series. For monthly series, M7 and D8F identified series as seasonal less than 1% of the time for (0 1 1) series; for (1 1 0) series between 8% and 11% of series were identified as seasonal with M7 and 6% to 17% with D8F, depending on the parameter. For quarterly series, M7 finds between 4% and 10% of (0 1 1) and 13% to 19% of (1 1 0) series seasonal. The 40-quarter series are more likely to have M7<1 than the 80-quarter series. For the (1 1 0) series, M7 is more likely to find a series with a small parameter to be seasonal when the series has 40 observations, but is more likely to find a series with a large parameter to be seasonal when the series has 80 observations. For the (0 1 1) series, larger parameter estimates lead to a smaller chance of identification as seasonal for both series lengths. For all models and series lengths, M7 is more likely to identify a series as seasonal when the ARIMA model is automatically chosen than when the correct model is specified; the number of additional identified series is larger for the 10-year series than for the 20-year series. (The ARIMA model selection affects the prior adjustment, which for these series is solely outlier adjustment, as they have no calendar effects, and the forecasts, which are used in the moving averages used to detrend the series.)

D8F, a component of the M7 calculation, follows these same patterns for quarterly series but identifies fewer series as seasonal – roughly about half of what M7 does for all series. This is not consistent with the previous results from the monthly series; in particular for the (1 1 0) series with a larger parameter estimate, D8F identified more series as seasonal than M7.

Although the M7 and D8F are more likely to identify a quarterly series as seasonal than a monthly series, the spectrum diagnostic does the opposite. However, monthly results look for a spectral peak over four seasonal frequencies and a 96-observation series, whereas these quarterly results look at only one seasonal frequency over a 40- or 80-observation series, so for this diagnostic, largely different results are less surprising. Between 1% and 10% of quarterly series are identified as seasonal, depending on model and series length. For both the (0 1 1) and (1 1 0) series, a larger parameter is less likely to lead to a series with a v.s. peak. The 80-observation series are also less likely to be found seasonal than the 40-observation series.

The 20-year monthly series simulated in Lytras et al. (2007) had significance levels close to 0.05 for the seasonal regressors when they were run with the correct model specified. For the quarterly series simulated here, this is not always the case. For the 20-year quarterly series with a parameter of 0.9, 5.1% of (1 1 0) and 5.9% of (0 1 1) series had a p-value of 0.05 or less, so these were fairly consistent. However, as the parameter got smaller, generally more series had a small p-value, with up to 7.6% seasonal for the 20-year (1 1 0) $\varphi = 0.3$ series. With one exception ((1 1 0) series with $\varphi = 0.3$), the seasonal regressors were more likely to identify seasonality for the 10-year series than for the 20-year series. Series length has the largest effect for the (0 1 1) series with θ = 0.3, where 6.8% of 20-year and 9.5% of 10-year series have a p-value less than 0.05.

The choice of ARIMA model does not affect the M7 or D8F results excessively. However, the seasonal regressors are more likely to indicate a seasonal series when the ARIMA model is automatically chosen than when the correct model is specified. This is particularly true for the 10-year series, for which up to 18% more series may be identified as seasonal when the model is automatically selected. The effect of the model selection has a larger impact when the true model is (1 1 0); for (0 1 1) models, up to 6% more series have seasonal regressors with a p-value less than 0.05, compared to 18% for (1 1 0).

Table 2: Percentage of diagnostics indicating seasonality for quarterly (1 1 0) series; unless stated otherwise, at an α =0.05 level.

φ	Len	Model	M7 <1	D8F > 7	V.S. Peak	Seas Reg	QS (A1)	QS (B1)	QS8 (A1)	QS8 (B1)
0.3	40	auto	15.4	8.3	--	18.4	2.2	2.6	2.2	2.7
0.3	40	(1 1 0)	14.3	7.4	10.4	6.9	1.0	1.1	0.6	0.7
0.3	80	auto	15.4	8.2	--	9.6	0.9	0.9	1.0	1.0
0.3	80	(1 1 0)	14.6	7.9	7.2	7.6	1.1	1.2	0.3	0.3
0.6	40	auto	18.8	10.9	--	25.5	1.6	1.7	1.6	1.9
0.6	40	(1 1 0)	17.4	10.0	6.2	7.6	0.8	0.6	0.5	0.5
0.6	80	auto	14.3	7.1	--	9.6	1.7	1.7	0.6	0.6
0.6	80	(1 1 0)	14.2	6.3	2.6	6.4	1.6	1.6	0.4	0.4
0.9	40	auto	19.2	11.0	--	12.7	1.6	1.6	0.9	0.9
0.9	40	(1 1 0)	18.2	10.7	1.6	6.1	1.2	1.2	0.7	0.7
0.9	80	auto	13.4	6.5	--	6.1	0.9	1.0	0.4	0.4
0.9	80	(1 1 0)	13.2	6.2	0.9	5.1	0.9	0.9	0.5	0.5

Table 3: Percentage of diagnostics indicating seasonality for quarterly (0 1 1) series; unless stated otherwise, at an α=0.05 level.

θ	Len	Model	M7 <1	D8F > 7	V.S. Peak	Seas Reg	QS (A1)	QS (B1)	QS8 (A1)	QS8 (B1)
0.3	40	auto	9.7	5.4	--	14.5	3.3	4.1	3.0	3.5
0.3	40	(0 1 1)	8.6	4.9	8.5	9.5	2.3	2.5	2.2	2.2
0.3	80	auto	7.1	3.2	--	7.2	3.1	3.2	2.6	2.5
0.3	80	(0 1 1)	6.9	3.0	6.6	6.8	2.1	2.1	2.2	2.2
0.6	40	auto	8.3	3.7	--	12.6	4.1	4.4	4.4	4.8
0.6	40	(0 1 1)	7.1	2.9	5.8	6.9	3.6	3.6	4.0	4.0
0.6	80	auto	5.1	1.8	--	8.0	3.6	3.8	3.8	4.1
0.6	80	(0 1 1)	5.2	1.5	3.7	5.9	3.3	3.3	3.2	3.2
0.9	40	auto	6.9	3.3	--	12.0	3.9	4.5	4.8	5.1
0.9	40	(0 1 1)	6.3	3.0	4.2	7.2	3.7	3.7	4.4	4.3
0.9	80	auto	4.1	1.9	--	8.9	5.3	5.6	3.8	4.0
0.9	80	(0 1 1)	4.0	1.9	2.2	5.9	5.0	5.4	3.7	3.8

The QS diagnostics seem to do the best job in not finding seasonality in these nonseasonal series. Depending on the model, series length, and series used for QS calculation, between 0.9% and 5.6% of series have a QS with a p-value less than 0.05, and between 0.3% and 5.1% have an eight-year QS with a p-value less than 0.05. Overall, the proportion of QS diagnostics indicating seasonality are smaller for the (1 1 0) series than the (0 1 1) series. The QS of the original series (A1) performs similarly to that of the prior adjusted (B1) series, with only up to a 0.4% difference when the correct model is specified and an 0.8% difference when the model is automatically selected. Series length does not seem to play a consistent role in the significance of these regressors. The prior-adjusted series is slightly more likely to be found seasonal, particularly for the (0 1 1) series. With some exceptions, the eight-year version of the diagnostic tends to find fewer series seasonal.

3.2 Simulated Airline Series

Ten and twenty year monthly and quarterly series were simulated from an airline model with the nonseasonal parameter 0.3 or 0.9 and the seasonal parameter 0.3, 0.6, or 0.9. One thousand series were simulated for each combination. The series were run in both X-13ARIMA-SEATS and X-12-ARIMA (for spectrum results only) with the airline model specified and either the automatically chosen seasonal filter or the stable seasonal filter. The series were also run in X-13ARIMA-SEATS with (0 1 1) + seasonal to test the significance of the seasonal regressors. To test for residual seasonality, the seasonally adjusted series were run through X-13ARIMA-SEATS with (0 1 1) + seasonal; for ten year series the span was limited to the last five years, and for twenty year series it was limited to the last eight years.

The automatically chosen filter is either a short (3x3), medium (3x5), or long (3x9) filter, with the choice based on the ratio of the movement in the seasonal component to the movement in the irregular component; when the seasonal component is changing rapidly, the short filter is chosen to capture this movement, whereas if the series has a slowly evolving seasonal pattern a long seasonal filter is chosen. For more information on the automatic filter choice, see U.S. Census Bureau (2015). The stable seasonal filter removes the quarterly mean from each observation; for a series with changing seasonality, this is a poor filter selection, as it does not capture and remove the changing effects. One might expect the automatically selected filter to do an adequate job in removing seasonality (although a poorer job than a trained analyst selecting model and adjustment options would do), so few series should show residual seasonality. However, the stable filter will likely do a poor job removing the effects, particularly for series with a small Θ and so an evolving seasonal pattern.

3.2.1 Seasonality

Table 4 shows the seasonality diagnostics for the airline series. D8F, M7, seasonal regressors, and the full-span QS for the original series and the prior adjusted series have at least 99% of series identified as seasonal for all parameters and series lengths. The spectrum does a poorer job, particularly for the 10-year series and the larger Θ values.

The eight-year QS is less likely to identify the series as seasonal than the full-span QS. Interestingly, the QS calculated on the last eight years of a 10-year series is slightly less likely to identify seasonality than the QS calculated on the last eight years of a 20-year series.

Table 4: Percentage of diagnostics indicating seasonality of quarterly airline series; unless stated otherwise, at an α =0.05 level.

Θ	θ	Len	V.S. Ori Peak	D8F >7	M7 <1	Seas Reg	QS (A1)	QS8 (A1)	QS (B1)	QS8 (B1)
0.3	0.3	40	96.8	99.9	99.8	99.9	100.0	100.0	100.0	100.0
0.3	0.3	80	99.1	99.8	99.7	99.8	100.0	99.8	100.0	99.8
0.3	0.9	40	94.9	99.7	99.7	99.8	100.0	99.8	100.0	99.8
0.3	0.9	80	99.4	99.8	99.8	100.0	100.0	100.0	100.0	100.0
0.6	0.3	40	96.7	99.9	100.0	100.0	99.7	99.6	99.8	99.6
0.6	0.3	80	98.0	99.9	100.0	100.0	99.9	99.6	99.9	99.6
0.6	0.9	40	93.4	99.6	99.6	99.7	99.5	99.1	99.5	99.1
0.6	0.9	80	96.9	99.7	99.7	99.8	100.0	99.6	100.0	99.6
0.9	0.3	40	95.1	99.7	99.8	99.7	99.0	98.4	99.0	98.4
0.9	0.3	80	96.7	99.9	99.9	99.9	99.6	99.2	99.6	99.2
0.9	0.9	40	92.8	99.4	99.6	99.6	99.1	98.2	99.1	98.3
0.9	0.9	80	94.3	99.9	99.9	99.9	99.6	97.9	99.6	97.9

3.2.2 Residual Seasonality

Tables 5 and 6 show residual seasonality diagnostics of the seasonally adjusted series with an automatic and stable filter, respectively; Table 7 shows the diagnostics for the irregular component. We expect to find few series with residual seasonality when the automatic filter is used, and this is what we find. For the spectrum, the 10-year series were slightly more likely to have a v.s. seasonal peak, but only up to 0.6% of the series had a v.s. seasonally adjusted series peak and 1.0% had a v.s. irregular peak. Most series groups had 0.1% or fewer series with a v.s. seasonally adjusted series peak and 0.3% or fewer with a v.s. irregular peak. The D11F test over the entire series identified almost no residual seasonality, while the D11F over the last three years found up to 1.3%.

Table 5: Percentage of diagnostics indicating residual seasonality of quarterly airline series run with the automatic filter choice; unless stated otherwise, at an α =0.05 level.

Θ	θ	Len	V.S. SA Peak	Seas Regs	D11F	D11F Last 3yr	QS (D11)	QS8 (D11)	QS (E2)	QS8 (E2)
0.3	0.3	40	0.6	4.4	0.0	0.3	0.0	0.2	0.0	0.0
0.3	0.3	80	0.2	1.4	0.0	0.2	0.1	0.1	0.0	0.0
0.3	0.9	40	0.1	3.7	0.1	0.9	0.4	0.4	0.1	0.1
0.3	0.9	80	0.0	0.7	0.0	0.7	0.0	0.3	0.0	0.0
0.6	0.3	40	0.0	3.2	0.0	0.2	0.0	0.0	0.0	0.0
0.6	0.3	80	0.0	1.0	0.0	0.3	0.0	0.0	0.0	0.0
0.6	0.9	40	0.1	2.6	0.2	0.4	0.0	0.1	0.0	0.0
0.6	0.9	80	0.0	0.4	0.0	1.3	0.0	0.0	0.0	0.0
0.9	0.3	40	0.1	4.3	0.1	0.2	0.0	0.0	0.0	0.0
0.9	0.3	80	0.0	0.7	0.0	0.6	0.0	0.0	0.0	0.0
0.9	0.9	40	0.0	1.9	0.0	0.7	0.0	0.1	0.0	0.0
0.9	0.9	80	0.0	0.4	0.1	1.3	0.0	0.0	0.0	0.0

Table 6: Percentage of diagnostics indicating residual seasonality of quarterly airline series run with a stable seasonal filter; unless stated otherwise, at an α =0.05 level.

Θ	**θ**	**Len**	**V.S. SA Peak**	**Seas Regs**	**D11F**	**D11F Last 3yr**	**QS (D11)**	**QS8 (D11)**	**QS (E2)**	**QS8 (E2)**
0.3	0.3	40	17.5	21.3	0.2	28.4	62.6	36.8	61.5	38.0
0.3	0.3	80	68.0	72.8	0.3	76.5	98.6	83.1	98.4	82.7
0.3	0.9	40	12.4	20.1	1.9	31.4	61.7	35.2	61.6	37.8
0.3	0.9	80	57.4	75.9	0.9	79.1	98.0	82.6	98.0	82.9
0.6	0.3	40	8.9	10.0	0.0	7.6	16.3	5.9	17.3	6.9
0.6	0.3	80	41.0	39.6	0.0	37.5	69.2	38.8	68.1	41.4
0.6	0.9	40	4.3	7.9	0.3	9.9	16.5	5.6	16.4	6.4
0.6	0.9	80	29.4	44.2	0.0	45.1	70.6	38.9	71.5	40.7
0.9	0.3	40	3.1	7.4	0.1	2.9	1.8	0.5	2.0	1.4
0.9	0.3	80	7.4	4.3	0.0	6.3	4.4	2.3	5.8	3.4
0.9	0.9	40	0.7	3.6	0.0	3.1	2.1	1.5	2.4	2.1
0.9	0.9	80	1.4	3.9	0.0	10.3	5.1	3.4	6.2	4.2

Table 7: Percentage of diagnostics applied to the irregular component indicating residual seasonality of quarterly airline series run with the automatic and the stable filters; unless stated otherwise, at an α =0.05 level.

			Seasonal filter = automatic					*Seasonal filter = stable*				
Θ	**θ**	**Len**	**V.S. Irr Peak**	**QS (D13)**	**QS8 (D13)**	**QS (E3)**	**QS8 (E3)**	**V.S. Irr Peak**	**QS (D13)**	**QS8 (D13)**	**QS (E3)**	**QS8 (E3)**
0.3	0.3	40	1.0	0.5	0.9	0.3	0.3	9.4	54.2	35.8	50.5	34.5
0.3	0.3	80	0.1	0.5	0.3	0.0	0.1	49.0	96.7	76.8	95.7	75.9
0.3	0.9	40	0.5	0.4	0.4	0.2	0.2	8.1	53.2	31.7	52.8	33.4
0.3	0.9	80	0.2	0.0	0.3	0.1	0.0	46.9	96.1	76.0	95.6	75.2
0.6	0.3	40	0.0	0.1	0.0	0.0	0.0	2.7	15.0	7.1	17.0	8.5
0.6	0.3	80	0.0	0.0	0.0	0.0	0.1	17.6	60.7	35.2	61.3	37.8
0.6	0.9	40	0.3	0.0	0.0	0.0	0.0	3.1	12.9	5.4	13.8	5.4
0.6	0.9	80	0.1	0.0	0.1	0.0	0.1	19.9	61.7	34.0	60.9	35.5
0.9	0.3	40	0.3	0.0	0.0	0.0	0.0	0.6	2.1	1.4	3.1	2.3
0.9	0.3	80	0.1	0.0	0.0	0.0	0.1	1.3	5.1	2.6	6.9	4.0
0.9	0.9	40	0.3	0.0	0.1	0.0	0.0	0.3	1.9	1.7	2.6	2.4
0.9	0.9	80	0.0	0.0	0.0	0.0	0.0	0.6	4.6	2.8	6.0	4.0

The QS statistics found almost no residual seasonality for adjustments using the automatic seasonal filter choice. For the seasonally adjusted series (D11), the most residual seasonality was found for 10-year series with θ = 0.9 and Θ = 0.3; 0.4% of series had a QS with a p-value less than 0.05. The QS of the seasonally adjusted series adjusted for extremes (E2) identified only one series as seasonal. The QS of the irregular (D13) and the irregular adjusted for extremes (E3) were slightly more likely to find residual seasonality, up to 0.5%.

The seasonal regressors fit to the seasonally adjusted series with automatically selected seasonal filters were the most likely to find residual seasonality. For up to 4.4% of the 10-

year and 1.4% of the 20-year series, the F-test for the regressors had p-values less than 0.05. The shorter series were more likely to have residual seasonality found, perhaps due to the adjustment or due to the shorter data span on which the seasonal regressors were run (five years vs. eight years). The parameter estimates have some effect on the residual seasonality. For all parameter/length groups, fewer series were seasonal when the θ parameter was 0.9 than when it was 0.3. With one exception – 10-year series with $\theta = 0.3$ – residual seasonality was found more often when the Θ was smaller.

Overall, for all diagnostics, seasonality was much less likely to be detected in these seasonally adjusted airline series than in the simulated nonseasonal series in Section 3.1.

Results are more interesting for the series run with a stable filter. First, more residual seasonality is consistently found with a 20-year series than a 10-year series. As the seasonal pattern has more time to evolve and therefore would be adjusted worse with a stable filter, this would make sense. Each diagnostic consistently finds fewer series with residual seasonality as the Θ parameter gets larger; as the stable filter is a better fit for series with a larger parameter value (in fact, a SEATS filter is practically stable as the Θ approaches 1), this result makes sense.

The D11F over the full span is rarely identified as seasonal, which makes sense as it measures for equality of seasonal means, and all seasonal means should be made equal by use of a stable filter. When the span is limited to the last three years, the D11F is more likely to find a series seasonal. Up to almost 80% of series are identified as seasonal with the three-year D11F with $\Theta = 0.3$; with $\Theta = 0.9$, only 10.3% are.

Overall, the spectrum is the diagnostic least likely to identify residual seasonality. For the 10-year $\Theta = 0.3$ series, only 12.4% to 17.5% of series have a v.s. peak in the spectrum of the seasonally adjusted series, while 57.4% to 68.0% have one for the 20-year series. For one group – $\Theta = 0.9$ and $\theta = 0.3$ – which the QS and the D11F considered to be the group with the least residual seasonality, the spectrum diagnostic found more residual seasonality than all the other diagnostics. The spectrum of the seasonally adjusted series was more likely to find seasonality than the spectrum of the irregular for all parameter/length groups.

Seasonal regressors are less likely than all diagnostics except the spectrum to find seasonality in the $\Theta = 0.3$ series, and are among the mostly likely to identify it in the $\Theta = 0.9$ series.

The QS results depended largely on the Θ parameter and on the series length, and on the θ parameter to a much smaller extent. When the $\Theta = 0.3$, up to 98.6% of series are identified as having residual seasonality, versus up to 71.5% of $\Theta = 0.6$ and 6.2% of $\Theta = 0.9$ series. The groups most likely to test positive for residual seasonality are all 20-year series; for the 10-year series, the most seasonal group finds 62.6%, 17.3%, and 3.1% of series to have residual seasonality for $\Theta = 0.3$, 0.6, and 0.9, respectively. This could be because the stable filter works better when the series has a shorter time to evolve, or because of the fewer observations upon which the QS is calculated. The QS8 are always calculated on the last eight years of the data. For every parameter/length group, fewer series are found seasonal when the QS is calculated on the last eight years, though the amounts by which they differ depend on parameter and length. For the $\Theta = 0.3$ group, about 25% fewer 10-year series and 15% fewer 20-year series are found seasonal with the QS8 than with the full span QS. However, for $\Theta = 0.6$, around 10% fewer are found for the 10-year series and 30% fewer for the 20-year series.

When $\Theta = 0.3$ or 0.6, the QS is more likely to identify seasonality in the seasonally adjusted series or the extreme-adjusted seasonally adjusted series than in the irregular or extreme-adjusted irregular. When $\Theta = 0.9$, the QS on the irregular adjusted for extreme values is most likely to find seasonality. For small Θ values, the extreme-adjusted series is slightly less likely to have a significant QS than its unadjusted version; the opposite is true for larger Θ.

4. Results for Real Economic Time Series

4.1 Data

Seasonality diagnostics were examined for 175 quarterly Census Bureau series. These include:

- 5 e-commerce sales series, from the Monthly Retail Trade Survey (MRTS). All 5 series are currently seasonally adjusted, although only the total retail series is published. Series span from fourth quarter 1999 to first quarter 2014.
- 147 series from the Quarterly Services Survey (QSS). Of these, 12 are currently being seasonally adjusted, although series not being adjusted may be seasonal. Series have different start dates, but all series end in first quarter 2014.
- 23 series from the Quarterly Financial Report (QFR), of which 7 are currently being seasonally adjusted, and several others are being considered for adjustment. The adjusted series start in fourth quarter 2000 and end in first quarter 2014; the other series have various start dates but all end in first quarter 2014.

All three surveys are sample surveys, and are subject to both sampling and nonsampling error. Descriptions of the survey methods of the surveys are available at www.census.gov/econ/other-econ.html.

The 24 series currently being seasonally adjusted are expected to be seasonal; the remaining series, while not seasonally adjusted, may still be seasonal.

Series were run with the specification file used in production in 2014, if one existed; for the unadjusted QFR series, they did not. They were also run in X-13ARIMA-SEATS with all automatic modeling options and either the automatic seasonal filter or a stable seasonal filter, and again in X-12-ARIMA to find the spectrum results. The whole series was run with (0 1 1) + seasonal regressors to find seasonality results. For residual seasonality, the seasonally adjusted series were run with (0 1 1) + seasonal regressors with either the last eight years or the last half of the series, whichever was smaller. QS8 is calculated only for series at least eight years long; this makes up 7 adjusted and 12 unadjusted QFR series, 10 adjusted and 37 unadjusted QSS series, and all 5 of the e-commerce series.

4.2 Seasonality

Table 8 shows the seasonality diagnostics for the series run using the production specification for all but the QFR-no adjustment series, which use specifications with automatic options.

Table 8: Percentage of real quarterly Census Bureau series with a seasonal diagnostic; unless stated otherwise, at an α =0.05 level. Series are separated by survey, and whether they are currently being seasonally adjusted (adj) or not (no adj).

Survey	**N**	**M7 < 1**	**D8F >7**	**Seas Regs**	**V.S. Ori Peak**	**Ori Peak (not v.s.)**	**QS (A1)**	**QS (B1)**	**QS8 (A1)**	**QS8 (B1)**
QFR (adj)	7	100.0	100.0	100.0	57.1	71.4	42.9	100.0	42.9	85.7
QSS (adj)	12	100.0	100.0	100.0	91.7	91.7	100.0	100.0	100.0	100.0
e-comm (adj)	5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
QFR (no adj)	16	62.5	62.5	75.0	25.0	43.8	43.8	56.3	16.7	41.7
QSS (no adj)	135	84.4	80.7	89.6	51.1	66.7	76.3	75.6	56.8	56.8
All	175	85.1	81.1	89.7	53.1	67.4	74.3	77.1	57.7	66.2

All currently adjusted series have M7 less than one and D8F greater than seven; as these two diagnostics have long been among the chief diagnostics used to determine whether to adjust a series, this is not a surprising result. These diagnostics are among the most likely to say a series is seasonal, with 85% of series indicating seasonality with M7 and 81.1% with D8F. Only the seasonal regressors, which find almost 90% of series as seasonal, identify more seasonality.

The spectrum was the worst performer overall, indicating far fewer series were seasonal than the X-11 diagnostics or the full span QS. Of the adjusted series, 83% had a v.s. peak, but this varied a great deal depending on the survey, with 100% of e-commerce and 57.1% of adjusted QFR series having one. Results were more favorable if the visual significance criteria were relaxed and a peak below the median or less than six stars was allowed.

The QS showed some mixed results. All adjusted QSS and e-commerce series had QS and QS8 of both the original and the prior adjusted series indicating seasonality at a 0.05 level. However, for the adjusted QFR series only three of seven series had a significant QS of the original series, while all seven had a significant QS of the prior adjusted series. These series have some very large outliers, which may provide an explanation as to why seasonality is not found until the series is outlier-adjusted.

For the unadjusted series, the QS found more seasonal series than the spectrum but fewer than the X-11 diagnostics. The QS for the prior-adjusted series was slightly more likely to indicate a seasonal series than for the original series, while the QS for the last eight years was less likely than the full span version to find seasonality.

4.3 Residual Seasonality

Tables 9 and 10 show the residual seasonality diagnostics of the series run with the production specification, the automatic MSR filter, and the stable filter. There are few indications of residual seasonality in the series run with the production spec; the seasonal regressors are significant for two series. There are also few indications of seasonality when the automatic filter selection is used, with all of them occurring in series that are not currently adjusted. When the stable filter is used, a few series indicate residual seasonality, again mostly in the currently unadjusted series.

Table 9: Percentage of real quarterly series showing residual seasonality with various diagnostics; unless stated otherwise, at an α =0.05 level. Series are run either with the automatic filter (msr), the stable filter (stab), or the 2014 U.S. Census Bureau specification file (prod).

Survey	N (Num >8yr)	SA Peak	V.S. SA Peak	Seas Regs	D11F	D11F Last 3yr	QS (D11)	QS8 (D11)	QS (E2)	QS8 (E2)
QFR (prod)	7 (7)	0.0	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0
QFR (msr)	7 (7)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
QFR (stable)	7 (7)	14.3	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0
QSS (prod)	12 (10)	0.0	0.0	8.3	0.0	0.0	0.0	0.0	0.0	0.0
QSS (msr)	12 (10)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
QSS (stable)	12 (10)	33.3	8.3	25.0	0.0	8.3	8.3	10.0	0.0	10.0
e-comm (prod)	5 (5)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
e-comm (msr)	5 (5)	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
e-comm (stable)	5 (5)	80.0	60.0	60.0	20.0	80.0	80.0	80.0	80.0	80.0
QFR no adj (msr)	16 (12)	0.0	0.0	12.5	6.3	0.0	0.0	0.0	0.0	0.0
QFR no adj (stab)	16 (12)	0.0	0.0	18.8	6.3	6.3	6.3	8.3	25.0	25.0
QSS no adj(prod)	135(37)	0.0	0.0	8.1	5.2	3.0	3.7	10.8	3.0	8.1
QSS no adj (msr)	135(37)	2.2	1.5	16.3	0.7	0.7	0.7	0.0	0.0	0.0
QSS no adj (stable)	135(37)	10.4	3.0	16.3	3.0	3.7	8.9	2.7	5.2	2.7
Total (prod)	159(59)	0.0	0.0	8.1	4.4	2.6	3.1	6.8	2.6	5.1
Total (msr)	175(71)	2.9	1.1	13.7	1.1	0.5	0.5	0.0	0.0	0.0
Total (stab)	175(71)	13.1	4.6	18.3	3.5	6.3	10.3	9.9	8.6	12.7

Table 10: Real quarterly series showing residual seasonality with various diagnostics; unless stated otherwise, at an α = 0.05 significance level.

Survey (adjustment)	N (Num>8yrs)	Irr Peak	V. S. Irr Peak	QS (D13)	QS8 (D13)	QS (E3)	QS8 (E3)
QFR (prod)	7 (7)	14.3	0.0	0.0	0.0	0.0	0.0
QFR (msr)	7 (7)	0.0	0.0	0.0	0.0	0.0	0.0
QFR(stable)	7 (7)	14.3	0.0	0.0	0.0	0.0	0.0
QSS (prod)	12 (10)	0.0	0.0	0.0	0.0	0.0	0.0
QSS (msr)	12 (10)	0.0	0.0	0.0	0.0	0.0	0.0
QSS (stable)	12 (10)	16.7	0.0	16.7	20.0	16.7	30.0
e-comm (prod)	5 (5)	0.0	0.0	0.0	0.0	0.0	0.0
e-comm (msr)	5 (5)	0.0	0.0	0.0	0.0	0.0	0.0
e-comm (stable)	5 (5)	60.0	0.0	80.0	80.0	80.0	80.0
QFR no adj (msr)	16 (12)	12.5	0.0	0.0	8.3	0.0	0.0
QFR no adj (stable)	16 (12)	0.0	0.0	25.0	25.0	25.0	25.0
QSS no adj (prod)	135 (37)	0.7	0.0	1.5	8.0	1.5	5.4
QSS no adj (msr)	135 (37)	3.0	0.7	1.5	2.7	0.0	0.0
QSS no adj (stable)	135 (37)	9.6	1.5	13.3	8.1	9.6	8.1
Total (prod)	159 (59)	1.2	0.0	1.3	5.1	1.3	3.4
Total (msr)	175 (71)	3.5	0.5	1.1	2.8	0.0	0.0
Total (stab)	175 (71)	10.9	1.1	16.0	16.9	13.1	18.3

The QS applied to the irregular component finds residual seasonality in 16.0% of the stable adjustments; the other three QS diagnostics trail behind, with 13.1% for the irregular adjusted for extremes, 10.3% for the seasonally adjusted series, and 8.6% for the seasonally adjusted series adjusted for extremes.

Unlike what was found with the simulated data, the eight-year versions of the QS diagnostics find a higher proportion of series with residual seasonality. However, this result is misleading, as the QS8 results are found on 71 eight-year series, whereas the full-span QS results are found on 175 series ranging from 4.5 to 14.5 years, with 104 of them shorter than eight years. For the stable adjustments, residual seasonality is found in 16.9% of the irregular components, 18.3% of the irregular adjusted for extremes, 9.9% of seasonal adjustments, and 12.7% of seasonal adjustments adjusted for extremes. While the full-span versions found more seasonality in the unadjusted components, the eight-year QS found more seasonality in the extreme-adjusted component.

There is a seasonal peak in the irregular for 10.9% of stable-adjusted series and one in the seasonally adjusted series for 13.1%; however, when the peaks are limited to "visually significant" peaks, only 1.1% of series have a v.s. irregular peak and 4.6% have a v.s. seasonally adjusted series peak. Compared to the QS, the spectrum seems to underperform.

The seasonal regressors run on either the last eight years or the last half of the seasonally adjusted series, whichever is smaller (except for the extremely short 4.5-year series, where the span used was the last three years, the program series length minimum), were the most likely to find residual seasonality in the stable adjustments. Of the series, 18.3% had seasonal regressors with a p-value of less than 0.05. The automatic filter adjustments did not fare much better; 13.7% had significant seasonal regressors; all the series with residual seasonality are not currently adjusted. The seasonal regressors were the only diagnostic finding much residual seasonality in the series adjusted with the automatic filter; the other diagnostics found residual seasonality in 0% to 3% of those series.

5. Conclusions

The new QS diagnostics perform well in identifying seasonality in seasonal series and in not identifying it in nonseasonal series. Overall, they tend to perform better or as well as the other diagnostics, and have the advantage of being built into the program. The program's default settings do not output the QS on the last eight years of the series for quarterly series as they do for monthly; as the eight-year version performs worse overall than the full span version, this is not a concern.

The spectrum does a poor job in identifying seasonality in quarterly series, particularly when using the traditional visual significance criteria. It identifies fewer nonseasonal series as seasonal than the M7 and D8F, but more than the QS. It is the worst diagnostic at identifying seasonality in airline series and also at identifying residual seasonality in series where it is expected to be found.

The seasonal regressors also perform poorly on these quarterly series. They are the most likely to find seasonality where seasonality is not expected but find less residual seasonality than the QS diagnostics for the poorly adjusted simulated series.

The results in the real series are perhaps negatively impacted by the varying series lengths; series are between 4.5 and 14.5 years, but are not separated by length for this study. Future research will look at results with the real data series separated by length, and also at how the QS diagnostics perform on simulated and real monthly series.

Acknowledgments

Thank you to my reviewers Kathleen McDonald-Johnson, Jenny Thompson, James Livsey, Grant Degler, and Gregory Cepluch, for their comments and insights.

References

Findley, D. F. and Lytras, D. (2014). Detecting Seasonality in Seasonally Adjusted Time Series. Unpublished paper, U.S. Census Bureau.

Gómez, V. and Maravall, A. (1996). *Programs TRAMO and SEATS: Instructions for the user,* (beta version: June 1997). Banco de Espana, Servicio de Estudios, DT 9628. (Updates and additional documentation found at www.bde.es/webbde/es/secciones/servicio/software/econom.html).

Ladiray, D. and Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method*, New York: Springer. Lecture Notes in Statistics, Vol. 158.

Lytras, D. P., Feldpausch, R.M., and Bell, W.R. (2007). Determining seasonality: A comparison of diagnostics from X-12-ARIMA. *Proceedings of the International Conference on Establishment Surveys III.* (CD-ROM).

Maravall, A. (2012). Seasonality tests and automatic model identification in TRAMO-SEATS. Banco de Espana, Servicio de Estudios.

U.S. Census Bureau (2015). *X-13ARIMA-SEATS Reference Manual, Version 1.1*, Time Series Staff, Center for Statistical Research and Methodology, Washington, DC, available at www.census.gov/srd/www/x13as/.