To Adjust or Not to Adjust? An Empirical Evaluation of Time Series with Unstable Seasonal Patterns

Demetra Lytras[1]
U.S. Census Bureau

Abstract

Often a time series exhibits a fairly stable seasonal pattern, making the decision to seasonally adjust it straightforward. However some series may show indications of seasonality but have a swiftly changing seasonal pattern. One school of thought is that all such series should be seasonally adjusted to ensure there is no residual seasonality in the published composite adjustments, while another says that only series that have a stable adjustment should be seasonally adjusted. This paper examines series with unstable seasonal patterns to determine whether it is better to adjust or not to adjust.

Key Words: time series, seasonal adjustment, X-13ARIMA-SEATS, seasonality

1. Introduction

For most monthly and quarterly economic time series from a survey that regularly seasonally adjusts its data, the decision as to whether to adjust a particular series is straightforward. These series either have a strong indication of a seasonal pattern and a resulting adjustment that is acceptably stable, or are not seasonal.

However, for some series the decision is more difficult. These series may

- have some seasonality diagnostics indicating that the series is seasonal while others do not find seasonality;
- have evident seasonal autocorrelation that is weak, so that the seasonal pattern changes rapidly with time and may be difficult to accurately estimate;
- be seasonal but unstable, so that the adjusted series is subject to overly large revisions;
- have changed over time, so that series that used to have a measurable seasonal pattern no longer do, or vice versa.

This paper examines series which fall in the first two categories to determine the effect of adjusting these mildly seasonal series. In particular, is the seasonal adjustment procedure in X-13ARIMA-SEATS (X-13A-S) effective in removing the seasonal pattern, and does it estimate the effect accurately? This is done using simulated monthly time series. Section 2 describes how the series are simulated, and Section 3 gives the results.

2. Simulation Methodology

The purpose of the simulation study was to determine how accurate the identification of seasonality is in series with a swiftly changing seasonal pattern, and how accurate the seasonal adjustment is. To test this, series were created with a known trend, seasonal pattern, and irregular component.

1. *Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance applied. (Approval ID: CBDRB-FY19-ESMD-B00012)*

The trends were obtained from an X-11 adjustment (using X-11 filters combined with forecasts from regARIMA models) of 10 published seasonal series from the Manufacturers' Shipments, Inventories, and Orders (M3) Survey. Data from this survey can be found at https://www.census.gov/manufacturing/m3/index.html. Estimates from the M3 survey are subject to measurement error and sampling error; because they are not based on a probability sample, the sampling error cannot be measured. The trends varied in level, smoothness, and the presence of sharp changes.

To create the seasonal patterns, 10 20-year monthly series were simulated from a (1 0 0)12 model with Phi = 0.4, and 10 more with Phi = 0.9. See Findley, Lytras, and Maravall (2016) for a discussion on some properties of these seasonal autoregressive series. These 20 series were then seasonally adjusted with X-13A-S, using a multiplicative adjustment, the true ARIMA model (for forecasting purposes), and an X-11 seasonal adjustment. Because the seasonal autocorrelation decays so quickly with the Phi = 0.4 series, a 3x3 filter was used to adjust these series. For the 0.9 series, the program was allowed to select the filter. Table 1 shows the estimated Phi for the series, the seasonal filter used (** indicates the filter was selected by the program), and seasonality diagnostics for these simulated series – the seasonal frequencies with a visually significant (V.S.) peak in the spectrum of the original series, the p-value from the QS calculated on the original series and the last eight years of the original series, and the D8F and M7. These diagnostics are described in Lytras (2007, 2015) and Findley, Lytras, and McElroy (2017). The diagnostics indicate seasonality when there is a V.S. peak at s1, s2, s3, or s4; the QS p-value < 0.01; D8F > 7; and M7 < 1.

When Phi = 0.4, the series does not have a very stable seasonal pattern. The autocorrelation between a value and the value one year ago is 0.4; observations two years apart have an autocorrelation of only 0.16. The only diagnostic that consistently detected seasonality in these series was the QS statistic of the original series. The spectrum of the original series, D8F, and M7 only rarely detected seasonality.

When Phi = 0.9, the seasonal pattern is more stable. Strong seasonal autocorrelation between observations persists for years; the autocorrelation is 0.9 for observations one year apart, 0.81 at two years apart, 0.729 at three years, and so on. All diagnostics indicated that each series was strongly seasonal. For these series, the 3x3 filter was selected as the final seasonal filter for seven series, and the 3x5 filter for three series.

The scale of the seasonal factors coming out of the X-11 adjustment was different for the 0.4 and 0.9 series. To make the results comparable and similar to reality, the seasonal factors were scaled so that their mean deviation from 1 was 0.049. This value is the average deviation from 1 of the seasonal factors of the published M3 series which were identified as strongly seasonal and for which X-13A-S automatically selects a multiplicative adjustment. Figure 1 shows one set of Phi = 0.4 seasonal factors and one set from the Phi = 0.9 series. Both sets of factors exhibit a swiftly changing seasonal pattern, but the Phi = 0.4 patterns shift direction and cross the 1.0 axis line more often. Table 2 summarizes these series of monthly seasonal factors.

Table 1: Model information and seasonality diagnostics of the simulated series. "**" indicates the seasonal filter was selected by X-13A-S. A diagnostic is in bold when it indicates the series is seasonal.

Phi	*Series*	*Estimated Phi*	*V.S. Spectrum Peaks*	*QS*	*QS (Last 8 years)*	*D8F*	*M7*	*Seasonal Filter*
0.4	1	0.345	**s4**	**0.000**	**0.000**	1.135	2.062	3x3
	2	0.466		**0.000**	**0.000**	0.905	2.363	3x3
	3	0.376		**0.000**	**0.001**	2.041	1.537	3x3
	4	0.484	**s3 s4**	**0.000**	**0.000**	6.203	**0.899**	3x3
	5	0.376		**0.000**	**0.001**	1.203	1.980	3x3
	6	0.447		**0.000**	**0.000**	2.511	1.384	3x3
	7	0.372		**0.000**	**0.000**	2.503	1.489	3x3
	8	0.327	**s3**	**0.000**	**0.188**	1.266	2.268	3x3
	9	0.461	s5	**0.000**	**0.000**	4.078	1.106	3x3
	10	0.422		**0.000**	**0.000**	2.938	1.259	3x3
0.9	1	0.906	**s1 s4**	**0.000**	**0.000**	**26.659**	**0.411**	3x3 **
	2	0.889	**s1 s2 s3 s4**	**0.000**	**0.000**	**34.080**	**0.398**	3x3 **
	3	0.920	**s2 s4 s5**	**0.000**	**0.000**	**25.741**	**0.488**	3x3 **
	4	0.870	**s3 s4**	**0.000**	**0.000**	**29.124**	**0.403**	3x5 **
	5	0.907	**s2 s3 s4**	**0.000**	**0.000**	**19.078**	**0.460**	3x3 **
	6	0.888	**s1 s3 s4 s5**	**0.000**	**0.000**	**23.214**	**0.462**	3x3 **
	7	0.926	**s1 s2**	**0.000**	**0.000**	**33.378**	**0.504**	3x3 **
	8	0.889	**s1 s2 s3 s4 s5**	**0.000**	**0.000**	**28.038**	**0.394**	3x5 **
	9	0.912	**s2 s4**	**0.000**	**0.000**	**49.316**	**0.330**	3x3 **
	10	0.918	**s2 s3 s4**	**0.000**	**0.000**	**62.020**	**0.313**	3x5 **

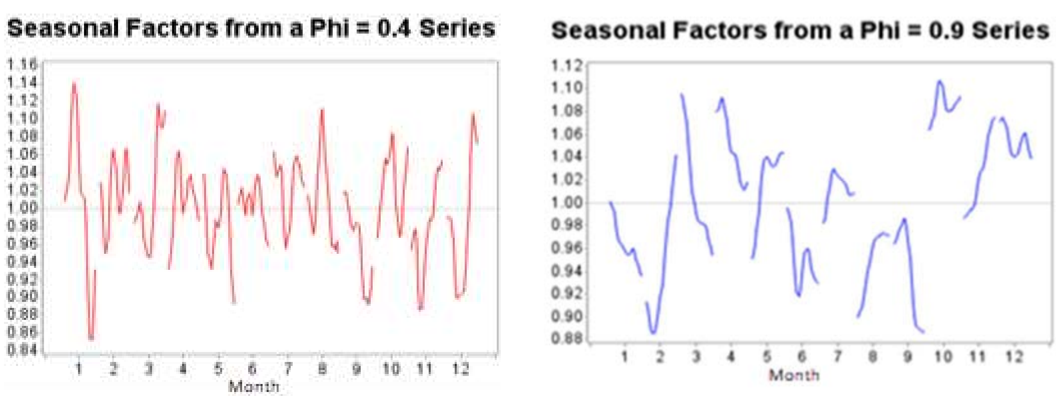

Figure 1: One of the sets of seasonal factors from the Phi=0.4 series and the Phi = 0.9 series

Table 2: Descriptive summary statistics of the Phi = 0.4 and Phi = 0.9 monthly seasonal factor series

For each monthly seasonal factor series…	*Phi = 0.4*	*Phi = 0.9*
Average number of times they cross one	2.08	0.86
Average changes in direction	4.03	2.65
Average spread	0.16	0.09
Average year-to-year change	0.018	0.008

To simulate the irregular components, five 20-year white noise series were simulated and centered on one. These series were then stretched so that the irregular was small compared to the seasonal factors (group A); the irregular was comparable to the seasonal factors (group B); and the irregular was large compared to the seasonal factors (group C). This is to identify differences in the accuracy of the adjustment depending on the scale of the irregular. Figure 2 shows one of the Phi = 0.4 seasonal factors overlaid with one of the irregular components, showing all three irregular levels.

The simulated series were then obtained by multiplying each trend, seasonal pattern, and irregular component, resulting in 500 series in each of groups A, B, and C for each Phi (3000 total series). These series were adjusted using the X-11 method, with:

- Log transformation
- Automatic regARIMA model identification and one year of forecasts
- Identification of additive outliers and level shifts
- Seasonal filter selected using the global moving seasonality ratio (seasonalma = msr)
- Sigma limits 1.5 to 2.5

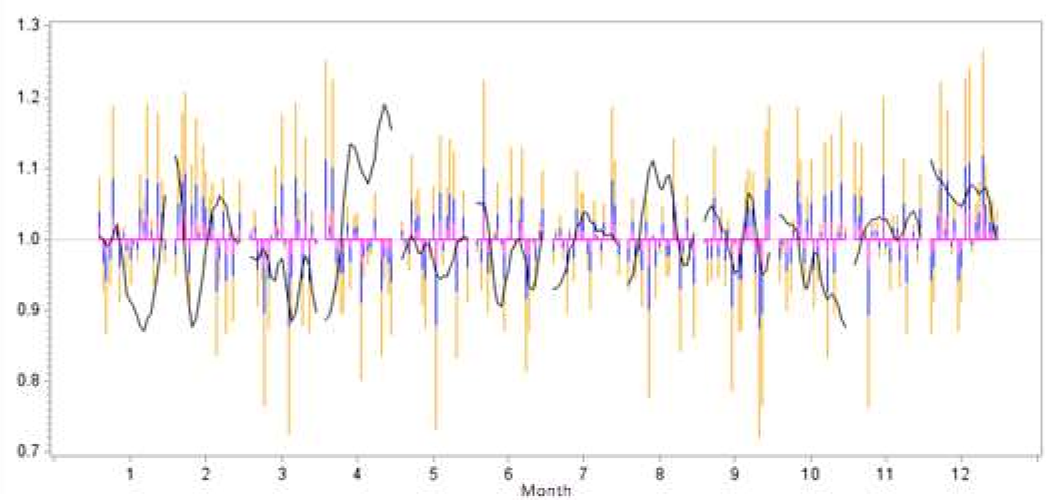

Figure 2: Monthly seasonal factors from a Phi = 0.4 series along with one of the simulated irregular components at all three levels.

An additional 50 group A, B, and C series were created and adjusted using only the trends and irregular components. These nonseasonal series provide a base against which to measure the results of the seasonal series.

3. Results

3.1 Seasonality of the Simulated Series

Table 3 shows the proportion of series from each group exhibiting seasonality according to each seasonality diagnostic. (M7 < 1; D8F > 7; visually significant (V.S.) seasonal peak (s1 – s4) in the spectrum of the original series; QS with p<0.01 for the full span and the last eight years of the series, for the original series and the prior adjusted series; and significance of seasonal regressors added to the model (run with model (0 1 1)), where the F-test has p<0.05.)

Table 3: Proportion of series in each group indicating seasonality

Irregular Group	*M7*	*D8F*	*Spectrum Peak*	*QS Ori*	*QS Prior Adj*	*QS Ori (Last 8 year)*	*QS Prior Adj (Last 8 year)*	*Seasonal Regs*
Phi = 0.4								
A	0.60	0.53	0.88	1.00	1.00	1.00	1.00	0.99
B	0.46	0.22	0.77	1.00	1.00	1.00	0.99	0.81
C	0.02	0.00	0.24	0.81	0.81	0.37	0.36	0.53
Phi = 0.9								
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	0.97	1.00	1.00	0.98	0.98	1.00
C	0.67	0.29	0.53	0.70	0.70	0.38	0.38	0.98
Nonseas								
A	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

There were almost no indications of seasonality in the nonseasonal series, except for one series with a visually significant seasonal peak. Interestingly, most of the nonseasonal series had a seasonal peak in the spectrum of the original series (31 Group A, 27 Group B, and 28 Group C series) but the peak was visually significant for only one series.

Unsurprisingly, the Phi = 0.9 series were identified as seasonal more than the Phi = 0.4 group, and seasonality was identified less as the irregular became larger. The Groups A and B Phi = 0.9 series were consistently identified as seasonal. The Group C series had a large range in seasonality detections amongst the diagnostics, with D8F identifying the least (29%) and the seasonal regressors finding the most (98%).

For the Phi = 0.4 series, seasonality detection was more varied among all groups. The QS identified seasonality at about the same levels as the Phi = 0.9 series: almost all Group A and B series had a significant QS, as did most Group C series when QS was calculated over the full span and almost 40% when it was calculated over the last eight years. Interestingly, this is the one diagnostic for which seasonality was detected more often in the Phi = 0.4 series than the Phi = 0.9 series. M7 and D8F were least likely to find seasonality in the Phi = 0.4 series in all groups.

One would expect the Phi = 0.9 series to be detected as seasonal, at least for Groups A and B; in Group C the large irregular could mask the rather consistent seasonal pattern. There is less certainty about the Phi = 0.4 series. While these series have a modest seasonal relationship, the seasonality is not persistent across multiple years. Many analysts would not consider the seasonality to be stable enough to adjust.

3.2 Residual Seasonality of the Simulated Series

A successful seasonal adjustment will remove the seasonality from the series, so the first step in assessing a seasonal adjustment is to check whether there is residual seasonality. We look for visually significant peaks in the spectrum of the seasonally adjusted series (adjusted for extremes) and the irregular (adjusted for extremes); QS with p-value<0.01 in the seasonally adjusted series, the seasonally adjusted series modified for extremes, the irregular, and the irregular modified for extremes; and the seasonal regressor F-test to have p-value<0.05 when the last eight years of the seasonally adjusted series are modeled with seasonal regressors and model (0 1 1). Extreme value adjustments remove the effects of both identified outliers and extreme values detected as part of the X-11 procedure. Results are in Table 4.

None of the nonseasonal series exhibited any residual seasonality after being seasonally adjusted. Like with the spectrum of the original series, there were a few non-visually significant seasonal peaks in the spectrum of the seasonally adjusted series and the irregular, but not as many as were found in the original (fewer than 5 in each group).

Table 4: Proportion of series exhibiting residual seasonality

Irregular Group	*Seas Adj Spectrum*	*Irregular Spectrum*	*QS Seas Adj*	*QS Extreme Adj Seas Adj*	*QS Irregular*	*QS Extreme Adj Irregular*	*Seasonal Regs (Last 8 years)*
Phi = 0.4							
A	0.012	0.030	0.594	0.006	0.632	0.000	0.024
B	0.000	0.006	0.000	0.000	0.000	0.000	0.000
C	0.004	0.004	0.000	0.000	0.000	0.000	0.000
Phi = 0.9							
A	0.000	0.010	0.002	0.000	0.000	0.000	0.004
B	0.000	0.002	0.000	0.000	0.000	0.000	0.000
C	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Nonseas							
A	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C	0.012	0.030	0.594	0.006	0.632	0.000	0.024

The adjustments of the Phi=0.4 and 0.9 series are almost all successful in removing seasonality. All the 0.9 series and the Group B and C 0.4 series have almost no indications of residual seasonality. However, for the 0.4 Group A series most diagnostics detect residual seasonality in some series. The seasonal regressors and the spectrum of the seasonally adjusted series and the irregular each flag between 1%-3% of the series. Surprisingly, 59% of the QS of the seasonally adjusted series have a significant p-value and 63% of the QS of the irregular are significant. When these series are adjusted for extreme values, fewer than 1% of the series have p-value<0.01. The reason that so many series have a significant QS for the component itself but not for the component modified for extreme values has not been identified.

3.3 Accuracy of the Adjustments

The average absolute percent difference (aapd) between the true unadjusted series and each estimated seasonally adjusted series was calculated. Table 5 shows the mean aapd of the series with no seasonal pattern, the Phi = 0.4 pattern, and the Phi = 0.9 pattern.

Table 5: Means of average absolute percent differences between the true unadjusted series and the estimated seasonally adjusted series, along with the standard error

Seasonal Pattern	*Group A aapd (s.e.)*	*Group B aapd (s.e.)*	*Group C aapd (s.e.)*
Phi = 0.4	1.4812 (0.0057)	2.3443 (0.0066)	4.0857 (0.0128)
Phi = 0.9	0.9083 (0.0041)	1.8292 (0.0074)	3.7121 (0.0206)
None	0.6245 (0.0109)	1.5666 (0.0329)	3.4466 (0.0743)

The errors in the estimates are smallest when the series being adjusted have no seasonal component, and largest for the series with the Phi = 0.4 seasonal pattern. As expected, the errors are also largest when the irregular is large compared to the seasonal component. Differences are greater between irregular groups than they are between the three seasonal groups. For the Group A series, the average error ranged from 0.6% for the nonseasonal series to 1.5% for the Phi = 0.4 series. For Group C, the average error ranged from 3.4% to 4.1%.

Figure 3 shows the true seasonal factors and the estimated Group A, B, and C seasonal factors of a Phi = 0. 4 and Phi = 0.9 series. These two series have the largest aapd among all series in their group that were identified as seasonal by all diagnostics. The Phi = 0.4 series is especially interesting in August. The true seasonal starts below 1, rises to about 1.1, and then falls back below 1. The Groups A and B series mimic this behavior, but when the series has the large Group C irregular, the estimated seasonal factor never rises above 1.

Table 6 shows the average absolute percent differences between the true trend and the estimated trends. For both the Phi = 0.4 and 0.9 series, there was less error in the estimate of the trend than that of the seasonally adjusted series. However, in Group A and B the nonseasonal series had larger errors when estimating the trend than when estimating the seasonal adjustment. (This difference is not significant for Group B.) The Phi = 0.4 series had the largest differences between their seasonal adjustment errors and their trend errors, with the seasonal adjustment error about 0.5% to 0.8% larger than the trend error.

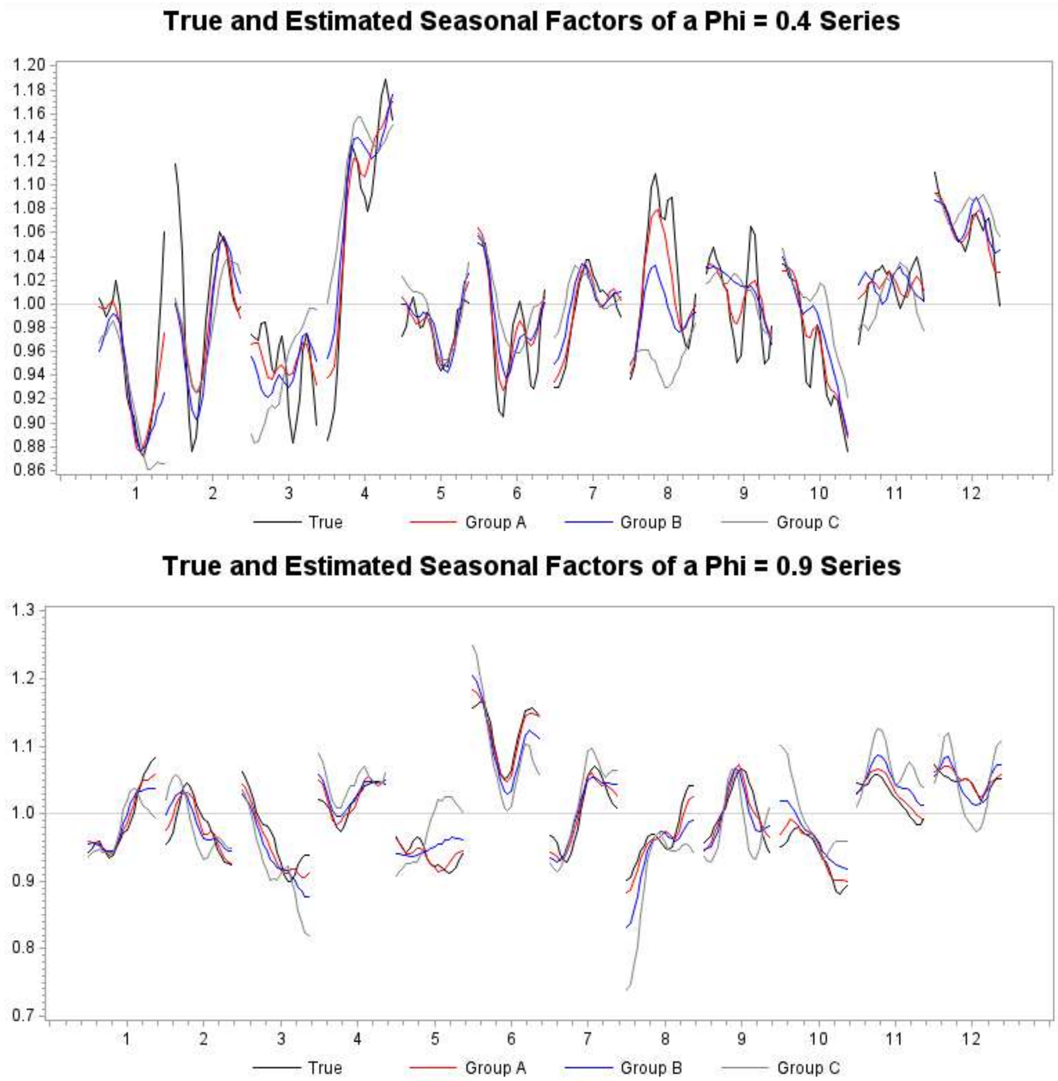

Figure 3: Monthly seasonal factors from one Phi = 0.4 and one Phi = 0.9 series, showing the differences between the true seasonal factors and the Group A, B, and C estimates

Table 6: Means of average absolute percent differences between the true trend and the estimated trends, along with the standard error

Seasonal Pattern	*Group A aapd (s.e.)*	*Group B aapd (s.e.)*	*Group C aapd (s.e.)*
Phi = 0.4	0.9511 (0.0071)	1.7394 (0.0131)	3.3289 (0.0295)
Phi = 0.9	0.801 (0.0071)	1.6664 (0.0137)	3.2869 (0.0297)
None	0.7373 (0.019)	1.6444 (0.0415)	3.2672 (0.0911)

3.4 Stability of the Adjustments

Findley and Monsell (1984) introduce the sliding spans and history diagnostics for helping to determine whether a series which shows evidence of seasonality can be reliably estimated. The sliding spans diagnostic divides the series into four overlapping subspans, adjusts each span, and calculates the maximum percent difference (mpd) in the estimates of the seasonal factors and month-to-month changes in the seasonally adjusted series for each point. The proportion of months with an mpd > 3% is found for each series. Findley and Monsell recommend that a series with greater than 25% of seasonal factors or 40% of month-to-month changes flagged is too unstable to be reliably adjusted.

Table 7 shows the mean, standard deviation, minimum, and maximum percentage of months with flagged seasonal factors and flagged month-to-month changes in the seasonally adjusted series for each group. It also shows the percentage of series in each group that would be considered too unstable to adjust. All Phi = 0.9 and most Phi = 0.4 Group A series have a stable adjustment. With the Group B irregular, most Phi = 0.9 and half of the Phi = 0.4 series are considered stable enough to adjust. Almost no Group C series are stable.

Table 7: Average percentage of months flagged as unstable in the seasonal factors (SF) and month-to-month changes in the seasonal adjustment (MM)

		Phi = 0.4					Phi = 0.9				
Group		*Mean*	*Std Dev*	*Min*	*Max*	*Percent Unstable Series*	*Mean*	*Std Dev*	*Min*	*Max*	*Percent Unstable Series*
A	SF	17.4	5.9	2.1	34.4	9.6	4.9	4.2	0.0	20.8	0.0
	MM	31.9	7.4	5.3	49.5	9.8	11.1	6.8	0.0	32.6	0.0
B	SF	23.6	7.3	5.6	50.9	39.6	13.7	6.4	0.0	33.3	4.2
	MM	39.7	8.5	15.9	58.9	54.8	27.0	8.5	6.5	55.1	7.4
C	SF	44.9	8.4	23.6	67.6	99.4	40.3	7.2	20.4	66.7	99.6
	MM	62.5	8.9	33.6	79.4	98.2	59.0	9.9	34.3	83.2	99.0

The history diagnostic measures the percent change in the seasonal adjustment and the month-to-month change in the seasonal adjustment between the initial estimate of a time point (when the series ends with that observation) and the final estimate of the time point (when the series extends to the final observation). Each series was run with the history diagnostic run over the last eight years. Table 8 shows the mean, standard deviation, minimum, and maximum average revisions for the 500 series in each group. The history diagnostic has no recommended level of adequate stability. Users develop cutoffs from knowledge of the series or use the diagnostic for comparison purposes.

Table 8: Average absolute percent difference between the initial and the final seasonal adjustment (SA) and month-to-month change in the seasonal adjustment (MM)

		Phi = 0.4				Phi = 0.9			
Group		*Mean*	*Std Dev*	*Min*	*Max*	*Mean*	*Std Dev*	*Min*	*Max*
A	SA	1.71	0.30	1.11	2.58	1.16	0.29	0.56	2.21
	MM	1.90	0.27	1.18	2.64	1.21	0.20	0.62	2.04
B	SA	2.00	0.29	1.29	2.82	1.58	0.28	0.89	2.46
	MM	2.60	0.33	1.74	3.41	2.04	0.36	1.14	3.12
C	SA	2.86	0.36	2.04	3.91	2.60	0.31	1.80	3.60
	MM	3.97	0.62	2.62	5.78	3.58	0.56	2.32	5.53

Within each group, the average revision and the average error in estimating the adjustment are only weakly correlated. Pearson' correlation statistic for Group A is 0.3 for both the Phi = 0.4 and Phi = 0.9 series; the Group C series are not correlated; and the Group B Phi = 0.9 series has correlation -0.25 while the Phi = 0.4 series is not correlated.

4. Conclusion

Deciding whether to seasonally adjust a series with a quickly changing seasonal pattern can be difficult. First one must determine whether a series is seasonal, and as Section 3.1 shows the seasonality diagnostics can give conflicting information for these mildly seasonal series. The X-11 method is largely effective in removing the seasonal patterns from these series, but these adjustments are less accurate and less stable than adjustments from series with a more stable seasonal pattern. The analyst must then weigh which is more important: removing all indications of seasonal autocorrelation in the series or producing seasonally adjusted series which are stable and not subject to large revisions.

Acknowledgements

Thank you to Kathleen McDonald-Johnson and Colt Viehdorfer for their comments on this paper.

5. References

Findley, D.F, Lytras, D, and Maravall, A (2016) "Illuminating Model-Based Seasonal Adjustment with the First Order Seasonal Autoregressive and Airline Models." *SERIEs* 7: 11-52.

Findley, D.F., Lytras, D., and McElroy, T. (2017) “Detecting Seasonality in Seasonally Adjusted Monthly Time Series.” Research Report Series *Statistics #2017-03*. Washington, DC: U.S. Census Bureau.

Findley, D.F. and Monsell, F. (1984) “New Techniques for Determining If a Time Series Can Be Seasonally Adjusted Reliably, and Their Application to U.S. Foreign Trade Series.” Statistical Research Division Report Series #RR-84/30. Washington, DC: U.S. Census Bureau.

Lytras, D. (2015) “On Seasonality: Comparing X-13ARIMA-SEATS Diagnostics for Quarterly Series.” *Proceedings of the Business and Economic Statistics Section: American Statistical Association*. Alexandria, VA: American Statistical Association.

Lytras, D., 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.*