To Revise or Not to Revise? Investigating the Behavior of X-13ARIMA-SEATS Seasonal Adjustment Revisions as New Series Values Are Added

Nicole Czaplicki, Kathleen McDonald-Johnson
U.S. Census Bureau, Washington, DC 20233

Abstract
How many historic seasonally adjusted values should we revise with each release of new time series estimates? We can revise the entire span, we can revise a minimum number of recent values, or we can choose an intermediate approach. To answer the question, we investigate how much the seasonal adjustment changes as we add new series values. We seasonally adjust using X-13ARIMA-SEATS with the X-11 seasonal adjustment method and regARIMA (regression plus ARIMA) model forecasts, using real U. S. Census Bureau series and simulated ARIMA model series. We assess the fluctuation of the seasonally adjusted estimates and how quickly they stabilize to a final value, particularly with respect to properties like ARIMA model coefficients and seasonal filter length. These results provide the foundation for determining the preferred number of revisions.

Key Words: ARIMA modeling, Regression, RegARIMA, Time Series

Disclaimer: Any views expressed are those of the authors and not necessarily those of the U. S. Census Bureau.

1. Background

The U. S. Census Bureau uses X-13ARIMA-SEATS for seasonal adjustment (U. S. Census Bureau 2015b). We select regARIMA models for forecasting and estimating regression effects like outliers and calendar effects and X-11 seasonal filters for seasonal adjustment. The software is the most recent in the line of development that has included X-11, X-11-ARIMA and updates (Dagum 1988), and X-12-ARIMA. Once a year, after reviews of the time series and seasonal adjustment settings, subject matter expert reviewers implement new seasonal adjustment specifications that stay in place until the next review. These changes to the specifications generally coincide with historic corrections and benchmarking that revise a year or more of the values of the original time series. With the annual revisions to the original series, the Census Bureau revises multiple years of the seasonal factors as well. For typical ongoing releases, however, the number of seasonal factors to revise has been a recurring question.

In the early days of seasonal adjustment, the Census Bureau's standard approach was to run the software once a year and project seasonal factors for the next year. Here we use the term "seasonal factor" in a broad sense, meaning a combined adjustment factor that might include adjustments for calendar effects like trading day or moving holidays. Following research by McKenzie (1984), most programs in the Census Bureau's economic directorate switched to some form of concurrent seasonal adjustment, incorporating the new series values when calculating seasonal factors each month (or quarter).

For a monthly survey, during a typical month, the Census Bureau collects new series values to add to the end of the series and revises some prior values to incorporate updated or additional response. For instance, for the Value of Construction Put in Place Survey (www.census.gov/construction/c30/c30index.html), we release a preliminary estimate for month *t*, first revised estimate for month *t*-1, and second revised estimate for month *t*-2. Similarly, for the Advance Monthly Sales for Retail and Food Services Survey (MARTS) and its companion the Monthly Sales for Retail Trade and Food Services Survey (MRTS) (www.census.gov/retail/index.html), the first estimate is an *advance* estimate for month *t*, approximately nine business days after the end of that month. In the same release, we provide a preliminary estimate for month *t*-1 and a revised estimate for month *t*-2. Compared to the later preliminary release, the advance values are early estimates from a smaller respondent sample, and the series are more aggregated across industries so there are fewer advance series. By the time we release the preliminary estimates, more detailed industry categories are available. After an additional month, we release revised estimates for these detailed categories. These are revisions to the original (unadjusted) series.

During a production concurrent seasonal adjustment run, the model form and X-11 seasonal filters stay the same, although we calculate new model coefficients. Just the addition of one new month at the end of the original series can change seasonal factors for the entire series, regardless of whether we have revised any prior months.

Ultimately, each statistical area decides how many seasonally adjusted values to update with each new series value. Some revise the entire seasonally adjusted series, and some revise only as far back as the revisions to the original series. Each new or revised month of the original series (as described above) is accompanied by a new or revised seasonally adjusted value. In addition, the retail trade (MARTS and MRTS) program revises the seasonally adjusted values for the current month a year ago (*t*-12) and prior month a year ago *(t*-13). This approach ensures that comparisons of month-to-month change for the current month and for the current month a year ago employ seasonal factors from the same seasonal adjustment vintage using all currently available information. At the Census Bureau, the retail trade program is one of two programs that target the year-ago factors for revision; wholesale trade is the other. Only the retail trade program also revises the prior month a year ago because of the importance of using the same seasonal adjustment settings and the same estimate of the seasonal pattern for these critical comparisons.

To gain perspective on prevailing practices, we contacted colleagues at other official statistics agencies and found that they vary. Some agencies revise no previous values with each release and some allow all of them to change. For example, in practice, the Instituto Nacional de Estadística y Geografía (INEGI) allows the full seasonally adjusted series to change. For most surveys, Statistics Canada revises one more period of the seasonally adjusted series than of the original series (that is, if the original series has revisions to *t*-1, the seasonally adjusted series has revisions back to *t*-2). For the Canadian Labour Force Survey, they do not revise the prior months of the seasonally adjusted series but just add the new concurrent estimate of time *t*.

Choosing a revisions policy can be challenging. It is a balancing act between providing the best estimates using all currently available information and adding uncertainty to the estimates. One compromise could be to revise only some of the seasonal factors at key time points in the estimation cycle, as the MRTS does by revising seasonal factors 12 and 13 months ago.

Some issues related to revisions are of a practical nature. First, when using a shortened model span and estimating calendar effects, we might avoid revising the seasonally adjusted series before the start of the model span. Second, if revising the seasonally adjusted series introduces too much noise, we might avoid revising, although "too much" requires definition. When combining or comparing series from different programs on a regular basis, we might align the number of revisions even if individually, the programs would have chosen different revision policies. Furthermore, when choosing X-11 seasonal filters, we expect changes to be small beyond half the filter length when tabulated data replace the forecasts, and we can take into account the most usual filter choices when determining a revision policy. Table 1 below shows the weights of typical seasonal filters that the Census Bureau uses.

Table 1. X-11 seasonal filters and their weights

Year-5	*Year-4*	*Year-3*	*Year-2*	*Year-1*	*Center*	*Year+1*	*Year+2*	*Year+3*	*Year+4*	*Year+5*
					3x3					
			1/9	2/9	3/9	2/9	1/9			
					3x5					
		1/15	2/15	3/15	3/15	3/15	2/15	1/15		
					3x9					
1/27	2/27	3/27	3/27	3/27	3/27	3/27	3/27	3/27	2/27	1/27

Elliott, McLaren, and Zhang (2007) listed several causes of revisions to published estimates. Here we concentrate on revisions that occur from added series values with no other changes to the original series.

2. Simulation Study

2.1 Methods

First, we wanted to examine the revisions to the seasonally adjusted estimates under somewhat "sterile" conditions. Namely, we wanted to look at seasonal time series without outliers, calendar effects or other confounding influences, other than white noise, that would interfere with the estimation of the seasonal pattern. We simulated three sets of 500 series from the airline (0 1 1)(0 1 1) model (Box and Jenkins 1970) with seasonal moving average parameter set to 0.3, 0.5 and 0.8, respectively. We set the nonseasonal moving average parameter to 0.3 for all series.

2.1.1 Seasonal Adjustment Options

We used Win X-13 (U. S. Census Bureau 2015a) to create spec files with the following settings:

- Span ending in 2001.1, with a start date of 1991.Jan
- Log transformation
- No regression variables and no automatic outlier identification
- Airline model
- Number of forecasts equal to half the filter length
- A seasonal filter of 3x3 for seasonal theta=0.3, 3x5 for 0.5 and 3x9 for 0.8.

2.1.2 Revisions Calculations

We seasonally adjusted each of these series 96 times with X-13ARIMA-SEATS Version 1.1 Build 17, adding a new month of data each time. We saved the seasonally adjusted series (Table D11 of the output) and the month-to-month changes (Table E6 of the output) for each run. To examine the change in the seasonally adjusted estimates between runs, we calculated the percent change in the estimate between two successive runs as:

$$100 * \left(\frac{\widehat{SA}_{t|t+k} - \widehat{SA}_{t|t+k-1}}{\widehat{SA}_{t|t+k-1}} \right) \qquad (1)$$

Since the month-to-month changes are already in percentages, we simply compute the difference as:

$$100 * \left(\widehat{MM}_{t|t+k} - \widehat{MM}_{t|t+k-1} \right) \qquad (2)$$

We refer to these differences as the lag *k* incremental revisions to the seasonally adjusted series and month-to-month changes, respectively. These revisions differ from those offered by the history spec in X-13ARIMA-SEATS in that they are calculated between successive runs rather than between the lagged period in question and the target period, either concurrent (t=0) or final (t=T). For each series, we calculated the average absolute revision for lags 1-60 and then ranked them from largest to smallest. We then compiled the ranks for all series within a seasonal filter category and examined the frequency of each ranking at each lag.

2.2 Results

For the seasonally adjusted series, we found, as expected, the largest average absolute revisions often occurred at either lag 1 or lag 12 (see Table 2). However, for the series adjusted with a 3x3 or 3x5 filter (seasonal theta equal to 0.3 or 0.5, respectively), a number of series had their largest average absolute revision at lag 6 or lag 7. We did not find this same behavior for the series adjusted with the 3x9 filter. Upon inspection of the spec files, we found that – as expected – the 3x3 series had 30 forecasts and the 3x5 series had 42 forecasts, or half the filter length for the selected filter. However, the 3x9 series had 60 forecasts, not the expected 66 forecasts (half of the length of a 3x9 filter) because previous releases of X-13ARIMA-SEATS allowed a maximum of 60 forecasts, and Win X-13 was programmed accordingly. We changed the number of forecasts for the series adjusted with a 3x9 filter to 66 forecasts and ran everything again. With these new seasonally adjusted series with the 3x9 filter, we found a number of series with the largest average absolute revisions at lag 6 and a few at lag 7, compared to the results with only 60 forecasts where not a single series had a lag 6 or lag 7 average absolute revision among its top 10 largest revisions.

We ran the series that utilized 3x3 and 3x5 filters again, now with 24 and 36 months of forecasts, respectively, to compare revisions from whole-year forecasts (a popular choice) to those from half-filter-length forecasts. With these forecast lengths, the large revisions at lags 6 and 7 were gone.

It is interesting to note that, for the seasonally adjusted estimates, with whole-year forecasts, the largest revisions often occur at lag 1. However, with half-filter-length forecasts, the largest revisions most often occur at lag 12. Also, note that for the series adjusted with 3x5 and 3x9 filters we see some series with lag 24 ranked in the top three (for the 3x9 filter a number of series had lag 36 ranked third, 20.4% with 60 forecasts and 43.6% with 66 forecasts. Lag 36 is not shown to save space). We do not observe many series with large revisions at later seasonal lags for series adjusted with the 3x3 filter. Recall the table of seasonal filters in Section 1. With a 3x9 seasonal filter, the highest weights are applied to the nearest three years before tapering off. The 3x5 weights taper off after one year and the 3x3 weights taper of immediately with the greatest weight only applied to the month of concern.

Since the MRTS revises the prior month a year ago (*t*-13), we included lag 13 in our analysis. With whole-year forecasts, lag 13 had the third largest average absolute revision at least 30% of the time. When forecasts were equal to half the filter length, lag 13 was ranked fourth (not shown) or higher over 30% of the time for series adjusted with a 3x3 filter or 3x5 filter. These results lend support to the MRTS policy to revise seasonally adjusted estimates 13 months in the past.

Table 2 Frequency of average absolute revisions rankings by lag for seasonally adjusted estimates for simulated data (in percentages)

3x3 Filter

	Lag 1		*Lag 2*		*Lag 6*		*Lag 7*		*Lag 12*		*Lag 13*		*Lag 24*	
Forecasts	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*
Rank 1	91.6	28.6	0.0	0.0	0.0	11.8	0.0	7.0	8.4	52.6	0.0	0.0	0.0	0.0
Rank 2	8.4	50.0	35.2	0.6	0.0	11.8	0.0	13.4	50.6	23.2	5.6	0.0	0.0	0.0
Rank 3	0.0	13.6	37.4	54.0	0.0	13.0	0.0	8.6	28.4	9.0	30.8	0.8	0.6	1.0

3x5 Filter

	Lag 1		*Lag 2*		*Lag 6*		*Lag 7*		*Lag 12*		*Lag 13*		*Lag24*	
Forecasts	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*
Rank 1	67.4	2.4	0.0	0.0	0.0	1.8	0.0	1.4	32.6	94.4	0.0	0.0	0.0	0.0
Rank 2	29.0	62.8	11.8	0.0	0.0	7.4	0.0	3.2	54.0	2.8	1.2	0.0	3.6	23.8
Rank 3	3.4	24.8	36.8	20.0	0.0	5.8	0.0	6.8	4.0	2.4	34.4	4.2	15.0	36.0

3x9 Filter

	Lag 1		*Lag 2*		*Lag 6*		*Lag 7*		*Lag 12*		*Lag 13*		*Lag 24*	
Forecasts	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*
Rank 1	62.2	0.0	0.0	0.0	0.0	29.8	0.0	8.0	37.8	62.2	0.0	0.0	0.0	0.0
Rank 2	7.4	0.6	0.0	0.0	0.0	11.2	0.0	29.6	61.0	1.4	1.2	0.0	30.4	57.2
Rank 3	10.0	7.8	0.0	0.0	0.0	6.6	0.0	4.4	1.2	20.8	43.6	0.0	24.8	1.4

For the month-to-month changes, the lag 1 revisions were not as prominent, especially with forecasts equal to half the filter length. For the 3x3 filter with 30 forecasts and the 3x9 filter with 66 forecasts, the largest revisions were split between lag 6 and lag 12. Otherwise, the largest revisions most frequently occurred at lag 12, with some at lag 1 for the 3x3 filter with 24 forecasts. The next largest revisions were spread over many lags, including lag 5, lag 11, and lag 24. We saw some of the larger revisions to the seasonally adjusted estimates at lag 24, so naturally this followed in the month-to-month changes. That lags 5 and 11 should also have some of the larger revisions is not surprising. The largest revisions to the seasonally adjusted series often occur at lag 12. Recall that the month-to-month change for a given month is the percent change between that month and the month prior. So, the lag 11 estimate of the month-to-month change is the percent change between the lag 11 seasonally adjusted estimate of one month and the lag 12 seasonally adjusted estimate of the prior month. Likewise, the lag 5 estimate of the month-to-month change is the percent change between the lag 5 seasonally adjusted estimate of one month and the lag 6 seasonally adjusted estimate of the prior month, and we already saw that there were large lag 6 revisions to the seasonally adjusted estimates with certain forecast lengths.

Table 3 Frequency of average absolute revisions rankings by lag for month-to-month change estimates for simulated data (in percentages)

3x3 Filter

	Lag 1		*Lag 5*		*Lag 6*		*Lag 11*		*Lag 12*		*Lag 18*		*Lag 24*	
Forecasts	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*
Rank 1	14.2	0.0	0.0	0.0	0.0	31.0	0.0	0.0	85.8	69.0	0.0	0.0	0.0	0.0
Rank 2	63.2	0.0	0.0	2.2	0.0	33.2	2.0	2.2	14.2	28.0	0.0	0.0	20.6	33.6
Rank 3	7.6	0.6	0.0	30.8	0.0	6.6	56.2	28.4	0.0	0.6	0.0	0.6	19.8	25.6

3x5 Filter

	Lag 1		*Lag 5*		*Lag 6*		*Lag 11*		*Lag 12*		*Lag 18*		*Lag 24*	
Forecasts	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*
Rank 1	0.6	0.0	0.0	0.0	0.0	8.6	0.0	0.2	99.4	91.4	0.0	0.0	0.0	0.0
Rank 2	31.0	0.0	0.0	0.0	0.0	19.6	0.2	38.6	0.6	8.6	0.0	0.0	68.2	71.6
Rank 3	14.6	0.0	0.0	1.4	0.0	15.4	42.6	35.0	0.0	0.0	0.0	0.0	19.8	26.8

3x9 Filter

	Lag 1		*Lag 5*		*Lag 6*		*Lag 11*		*Lag 12*		*Lag 18*		*Lag 24*	
Forecasts	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*	*60*	*66*
Rank 1	0.0	0.0	0.0	0.0	0.0	46.4	0.0	0.0	100.0	53.6	0.0	0.0	0.0	0.0
Rank 2	3.8	0.0	0.0	0.0	0.0	5.8	0.0	0.0	0.0	21.6	0.0	24.0	96.2	47.8
Rank 3	35.8	0.0	0.0	14.4	0.0	6.2	0.8	0.0	0.0	4.8	0.0	6.0	3.8	24.1

Win X-13’s coding only 60 forecasts instead of 66 for a 3x9 filter turned out to be a happy accident that helped demonstrate the effect of the number of forecasts on the timing of the largest revisions, to both the seasonally adjusted estimates and the estimated month-to-month changes.

3. Revisions With Retail Sales Data

3.1 Data and Methods

For our real-world investigation, we looked at 63 monthly retail sales series from the Monthly Retail Trade Survey as published on the U. S. Census Bureau’s website. Some of these series are composites of published and unpublished series. We used data from January 1992 to December 2008 and examined revisions from January 2001 to December 2008. The series estimates are subject to sampling error; more information on the survey design and estimation methods are available at the website www.census.gov/retail/how_surveys_are_collected.html. We followed a similar procedure as with the simulation study. First, we created spec files with Win X-13, and then we seasonally adjusted the series 96 times, adding one month of data with each run. We created the spec files with Win X-13 using the following settings:

- Span ending in 2001.1
- Log transformation
- Tested for outliers, trading day (6 coefficient or 1 coefficient), Easter (1, 8 or 15), constant
- Automatic model selection, mixed=no
- Forced seasonal (0 1 1) component
- Number of forecasts equal to half the filter length
- Filter length automatically selected
- Saved d11 and e6 tables

We calculated the average absolute incremental revisions defined by (1) and (2), ranked the average revisions, and computed the frequencies of each ranking at each lag. The typical patterns of incremental revisions to the seasonally adjusted estimates and the month-to-month change estimates are illustrated in Figure 1 by the revisions to women's clothing stores sales seasonally adjusted and month-to-month change estimates.

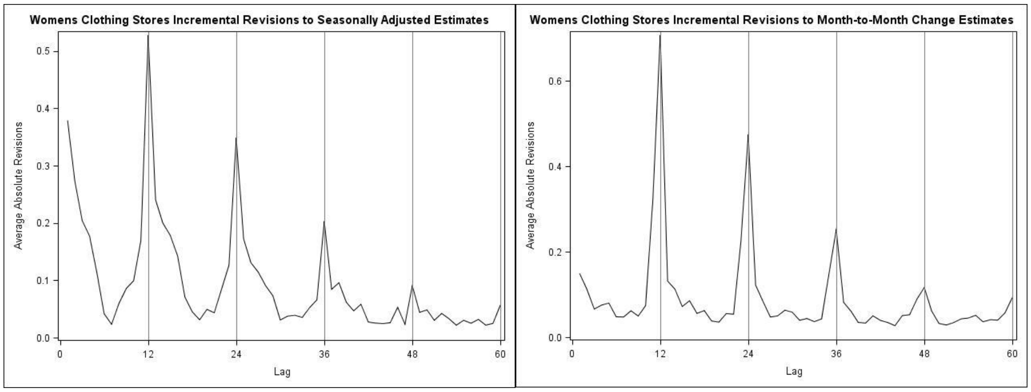

Figure 1 Incremental revisions to the seasonally adjusted and month-to-month change estimates for women's clothing retail sales. Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

We repeated this with 24 months of forecasts for series adjusted with a 3x3 filter and 36 months of forecasts for series adjusted with a 3x5 filter. Only one series was adjusted with a 3x9 filter, so we omit that series from this comparison.

Table 4 Frequency of average absolute revisions rankings by lag for seasonally adjusted estimates for retail data (in percentages)

3x3 Filter

	Lag 1		*Lag 2*		*Lag 6*		*Lag 7*		*Lag 12*		*Lag 13*		*Lag 24*	
Forecasts	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*
Rank 1	22.7	22.7	0.0	0.0	0.0	0.0	0.0	0.0	77.3	77.3	0.0	0.0	0.0	0.0
Rank 2	68.2	68.2	13.6	18.2	0.0	0.0	0.0	0.0	9.1	4.5	0.0	0.0	9.1	9.1
Rank 3	9.1	9.1	72.7	59.1	0.0	0.0	0.0	0.0	13.6	18.2	0.0	0.0	4.5	13.6

3x5 Filter

	Lag 1		*Lag 2*		*Lag 6*		*Lag 7*		*Lag 12*		*Lag 13*		*Lag 24*	
Forecasts	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*
Rank 1	30.0	35.0	0.0	0.0	0.0	0.0	0.0	0.0	70.0	65.0	0.0	0.0	0.0	0.0
Rank 2	27.5	25.0	12.5	12.5	0.0	0.0	0.0	0.0	12.5	17.5	5.0	5.0	35.0	32.5
Rank 3	32.5	30.0	12.5	15.0	0.0	0.0	0.0	0.0	12.5	12.5	15.0	12.5	22.5	25.0

Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

Table 5 Frequency of average absolute revisions rankings by lag for month-to-month change estimates for retail data (in percentages)

3x3 Filter

	Lag 1		*Lag 5*		*Lag 6*		*Lag 11*		*Lag 12*		*Lag 18*		*Lag 24*	
Forecasts	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*	*24*	*30*
Rank 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
Rank 2	4.5	4.5	0.0	0.0	0.0	0.0	27.3	27.3	0.0	0.0	0.0	0.0	68.2	68.2
Rank 3	13.6	9.1	0.0	0.0	0.0	0.0	54.5	59.1	0.0	0.0	0.0	0.0	27.3	27.3

3x5 Filter

	Lag 1		*Lag 5*		*Lag 6*		*Lag 11*		*Lag 12*		*Lag 18*		*Lag 24*	
Forecasts	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*	*36*	*42*
Rank 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
Rank 2	0.0	0.0	0.0	0.0	0.0	0.0	10.0	10.0	0.0	0.0	0.0	0.0	90.0	90.0
Rank 3	0.0	0.0	0.0	0.0	0.0	0.0	40.0	40.0	0.0	0.0	0.0	0.0	10.0	10.0

Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

As we can see in Table 4 and Table 5, the large revisions at lag 6 observed in the simulated series are not present in the retail series. Moreover, there is very little if any difference in the frequency of revisions rankings between the different forecast lengths, as we observed in the simulated data.

Why do the retail series exhibit such a different pattern of revision ranking frequencies from the simulated series? There are several differences between the two types of series. For one, the retail series are subject to outliers and calendar effects, whereas our simulated series were not. We did adjust for outliers and calendar effects in the retail series when they were identified, but other effects could have gone undetected.

Another difference lies in the ARIMA model identification. We generated the simulated series with the airline model, so there is no possibility of series model misspecification. In contrast, we do not know the best ARIMA model for each retail series; those models were estimated from the data at hand. We utilized the automatic model selection procedure to choose a model for each series. Furthermore, there may not be an ARIMA model that describes each of the retail series well. Thus, the quality of the model fit (or even the appropriateness of the utilized model) is one obvious difference between the retail and simulated series. X-13ARIMA-SEATS uses the regARIMA model to extend the series with forecasts, which it then treats as “real” data while decomposing the series into the trend, seasonal, and irregular components. This allows X-13 to use symmetric filters for more data points and to minimize the use of X-11 asymmetric filters. In reality, even the “symmetric” filters are asymmetric when forecasted values are used because the forecasts are functions of past values. Large revisions to seasonally adjusted estimates can occur when the difference between a previously forecasted value and the actual value that replaces it are large. A poor model fit can result in large forecast errors if the model continues to be poor going forward. But even with a great model fit, large forecast errors can still occur when there are unanticipated changes to the series, such as a change in the seasonal pattern.

Another difference is the composition of the respective sets of time series. We know the simulated time series are composed of a moving average component (in the form of the airline model) and a white noise component. Because this is a simple process, it is easier for X-13ARIMA-SEATS to decompose this into its seasonal, trend and irregular components. This means that the decomposition is more consistent when new months are added, resulting in the largest revisions where we would expect them. In contrast, the retail series are affected by unpredictable events and multiple sources of noise (for instance, sampling and nonsampling errors) that can interfere with the estimation of the seasonal pattern. If we have a slightly different estimated seasonal pattern when adding new months of data, we could get unexpected large revisions because the forecast models do not anticipate changes in seasonal patterns. Moreover, the nature of the error processes in the retail series is unknown. The errors in the retail data could follow different patterns than the errors in the simulated data, resulting in a different pattern of revisions.

3.2 Revision Magnitude

We now have a good understanding of the timing of the largest revisions to the seasonally adjusted estimates and month-to-month change estimates. But what about the *size* of the revisions? Is the largest average absolute revision *that* much greater than the second largest? How do the sizes of the revisions compare across series?

We know that some of the retail series have larger revisions than others just due to the nature of the series. For example, the average absolute revisions to the seasonally adjusted estimates from the shoe stores series range from 0.039% to 0.816%, whereas the range for grocery stores is 0.004% to 0.120%. A revision of 0.1% would seem trivial in the shoe stores series but would be quite sizeable in the grocery stores series. Even though the average absolute revisions for each series that we have computed are all in percentage terms, it does little to aid us in forming a unified revisions policy. Consequently, it is difficult to compare across series the size of revisions at a given lag for one series relative to other revisions for that series by looking at revisions in percentage terms alone. Our goal in comparing revisions across series was to look for patterns in the relative magnitude of revisions.

Therefore, we normalized our revisions calculations by dividing the average absolute revision at each lag by the overall average absolute revision for the series. Normalized revisions allowed us to compare the relative magnitude of revisions for one series to those of other series at a given lag.

3.2.1 Relative Revisions by Seasonal Filter

In Figure 2, normalized lag 1 revisions of the seasonally adjusted estimates for each of the 63 retail sales series are stratified by filter length. There is a clear cluster of large average absolute revisions adjusted with a 3x3 filter. An analysis of variance indicates that there is a significant difference in the mean revision among filters at lag 1 and lag 36 at the 0.05 level. With lags 12 and 24, there is little difference in the size of relative revisions between the 3x3 and 3x5 series. By lag 36, the 3x9 series has the largest average relative revision followed by a cluster of 3x5 series. This is not unexpected when we consider the filter weights provided in Table 1. For the 3x3 filter, the weight for the value three years removed is zero. Whereas with the 3x9 filter, the value three years removed receives the same weight as the current month. The 3x5 series has a small but nonzero weight here. The greater weight corresponds to a greater revision.

Also, note the scale on these graphs. Recall that we divided the average absolute revision at each lag by the overall average for the series. With most of the lag 1 and lag 12 revisions exceeding two times the average revision for the series, and some as large as four or six times the average, it is clear that not only do these lags most often have the largest revisions, but also the magnitude of revisions at these lags are quite large relative to other revisions in the series. In Figures 2 and 3, it is apparent that many of the lag 12 revisions are larger than the lag 1 revisions. Controlling for the series itself, the difference between the lag 1 and lag 12 revision is significant. Even though the models might be adequate, the information added at t+12 affects the estimation of the seasonal pattern at time t. The new value replaces the forecasted value in the seasonal filter, causing a larger re-estimation than occurs for lags 1 through 11.

3.2.2 Relative Revisions by Seasonal Theta Estimate

We also examined revisions grouped by the seasonal theta estimate. Large estimates for the seasonal theta parameter typically indicate a more stable seasonal pattern. We hypothesized that a large seasonal theta would correspond to smaller relative revisions. This is most evident in the lag 2 revisions shown in Figure 4. The series with small seasonal theta parameter estimates (shown in blue) had the largest relative revisions and the series with large seasonal theta estimates (green) had the smallest relative revisions.

The difference in mean revision across the three theta estimate categories is statistically significant at the 0.05 level.

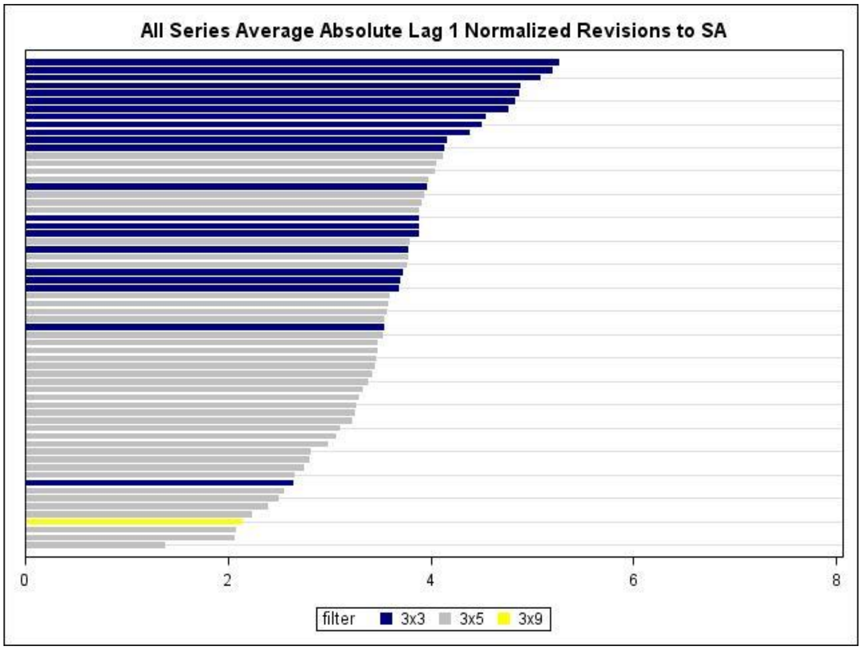

Figure 2 Average absolute relative lag 1 revision to the seasonally adjusted estimates for retail sales series. Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

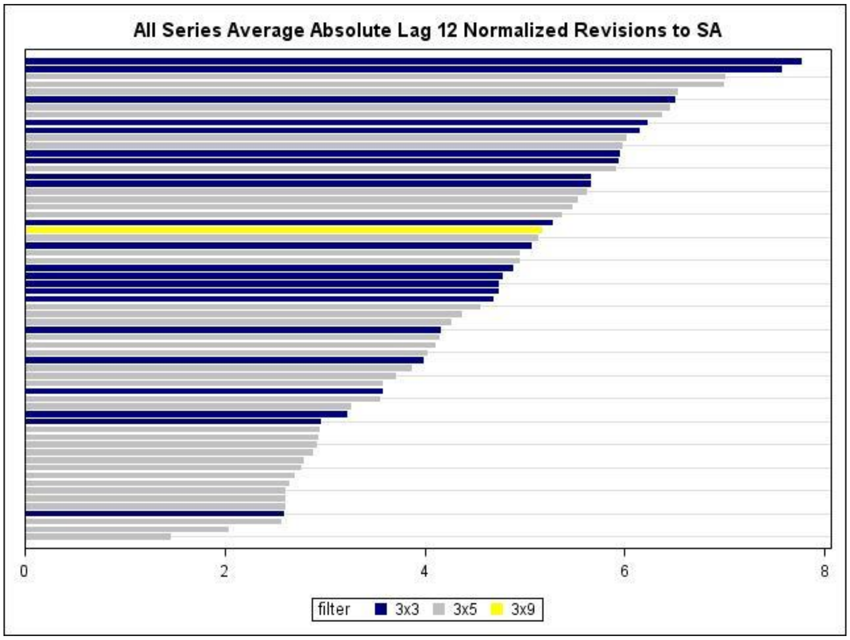

Figure 3 Average absolute relative lag 12 revision to the seasonally adjusted estimates for retail sales series stratified by filter. Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

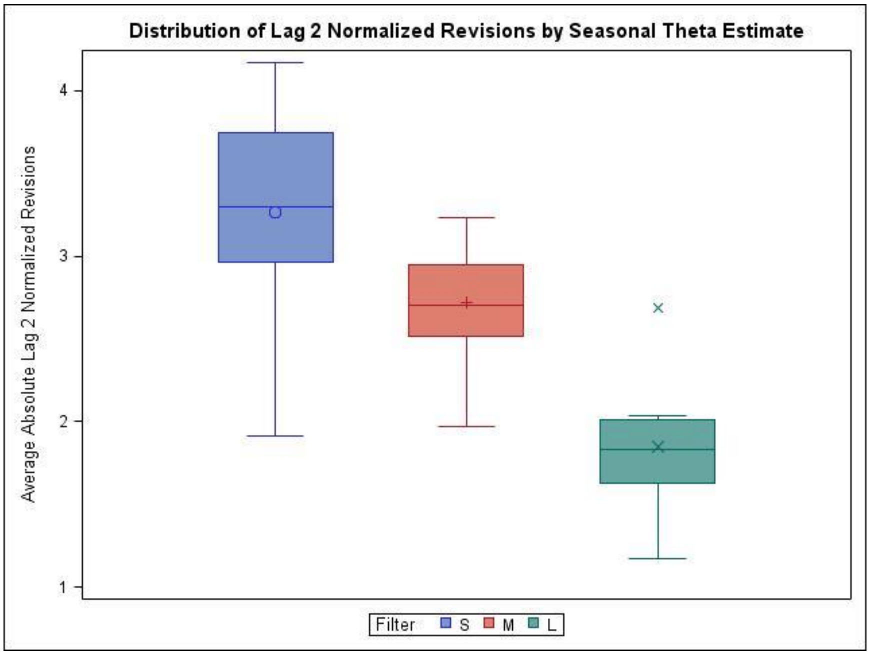

Figure 4 Average absolute relative lag 2 revision to the seasonally adjusted estimates for retail sales series by seasonal theta estimate:
S=small, Θ < 0.4; M=medium, 0.4 ≤ Θ ≤ 0.7; L=large, Θ > 0.7. Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

The normalized revisions of the month-to-month change estimates exhibited similar patterns to the seasonally adjusted estimates and we will not discuss them here.

3.3 Path to a Final Value

The goal when revising any published estimate is to provide a better estimate of the true unobserved value than the previously published estimate, given the most recent information available. When revising any estimate multiple times, we run the risk of revising in either direction, with each revision potentially oscillating around what will ultimately be the final estimate. This type of random up and down, yo-yo like revision pattern could undermine the confidence data users have in the estimates. Therefore, it would be desirable if revisions to seasonally adjusted estimates and month-to-month changes were not completely random. The ideal case would be if revisions were monotonically decreasing (or increasing) as the adjustments approach the final estimate, but we know this is not the case.

To test the randomness of the revisions path, we conducted a Wald-Wolfowitz test (Wald and Wolfowitz 1940), also called a runs test, on each of the seasonally adjusted and month-to-month change estimates. We centered the estimates about a “final” seasonal adjustment or month-to-month change estimate. These estimates are subject to small revisions after real series estimates replace forecasts in the calculation of the seasonal factors. This occurs after two years for series adjusted with a 3x3 filter, three years for a 3x5 filter and five years for a 3x9 filter. For simplicity, we will take the “final” value to be the estimate three years after the concurrent estimate. This centering resulted in 60 tests per estimate type (seasonally adjusted or month-to-month change) for each series, one for each estimate that we had at least three years of data afterwards. For all 63 series, every test indicated that the revisions were not random. However, the path to a final value is not a monotonic one. We repeated the Wald-Wolfowitz test, this time centering about the concurrent estimate. Again, every test indicated the revisions were not random and not monotonic.

This result was an important one to confirm. It validates the practice of revising seasonally adjusted estimates at all and gives us confidence that revisions to our estimates are beneficial.

3.4 Mean Squared Revision of the Estimates

Our goal when studying revisions to seasonally adjusted data is to understand how we get from the concurrent estimate to the final estimate by studying the timing and magnitude of revisions to the estimates. We are also interested in identifying factors that can help us predict both the timing and the magnitude of those revisions. So far, we have focused on factors that the practitioner controls, namely the choice of seasonal filter and forecast extension length. However, another factor that could aid in the prediction of revisions lies outside of the control of the practitioner: the series itself.

We focused on just a few of many series characteristics that could affect revisions. For example, we considered the distribution of the differences between intermediate estimates and the “final” estimate. In particular, we wanted to look at how far away each seasonally adjusted estimate for each month was from its “final” estimate and how those differences between intermediate estimate and “final” estimate compared across months within a series. Again, we defined the “final” seasonal adjustment estimate to be the estimate three years after the concurrent estimate.

Since we wanted to compare the deviations of the intermediate estimates across each month of the series, we constructed a time series of mean squared revisions, defining each month as in (3).

$$\widetilde{x_t} = \frac{1}{36}\sum_{i=0}^{35}\left(x_{t|t+i} - x_{t|t+36}\right)^2 \quad (3)$$

This gave us a 60-month long time series of mean squared revisions for each of the 63 retail sales series. After graphing a couple of these series, as in Figure 5, it became clear that for some series there was a pattern to the mean squared revisions that seemed seasonal in nature.

Thus, we read the mean squared revision series into X-13ARIMA-SEATS to check seasonality diagnostics and autocorrelations. We checked the QS statistic, the autocorrelation function (ACF), and the spectrum of the series of mean squared errors, as shown in Table 6. However, the spectrum and QS diagnostics may be of limited effectiveness for five-year-long series.

Table 6: Number of mean squared error series with diagnostics indicating seasonality

Diagnostic	*Number of series*
S1 peaks	14
S2 peak	5
S3 peak	7
S4 peak	6
Significant QS	28
Lag 12 ACF	36

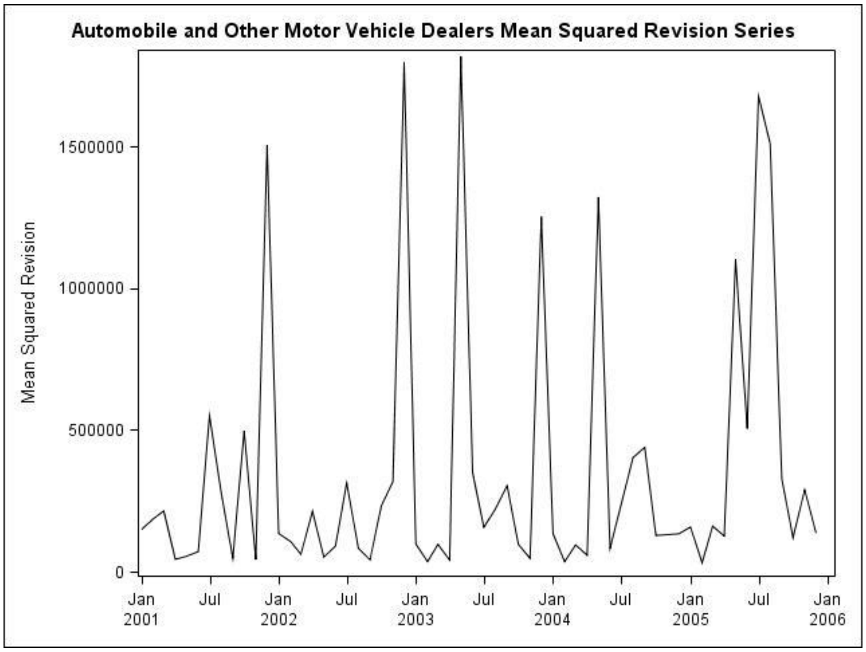

Figure 5: Time series of mean squared revisions for sales at automobile and other motor vehicle dealers. Source: Monthly Retail Trade and Food Service Survey, 2001-2008.

With 51 out of 63 mean squared revision series displaying some evidence of seasonality, it appears that, at least for some series, revisions are seasonal. In other words, the magnitudes of revisions for some months (Januaries, Februaries, etc.) are consistently greater than other months. Moreover, for some of these series, the magnitude of revisions of one month is correlated with the magnitude of revisions 12 months in the past.

What does this mean for predicting revisions for these series? Well, for example, if we observe large mean squared errors for Januaries in the past, we can expect that next January's seasonally adjusted estimate will also be subject to large revisions before it reaches its final estimate. Understanding the seasonal pattern of revisions to the seasonally adjusted series can help agencies prepare analysts and data users for larger expected revisions to the seasonally adjusted estimates in months where large revisions were observed in the past.

4. Summary

In this study, we examined the size and timing of revisions to estimates of seasonally adjusted series and seasonally adjusted month-to-month changes. The results indicate that the largest revisions to these estimates most often occur 12 months after the concurrent estimate of both the seasonally adjusted series and month-to-month changes. Thus, revisions to the year-ago estimate are a high priority, especially for year-to-year comparisons but also to correct for the large change to the adjustment that occurs with an additional year of data. For the retail series, the filter choice determined the largest revisions at other lags. For example, the 3x3 series saw their next largest revisions at lags 1 and 2 for the seasonally adjusted series and lags 24 and 11 for the month-to-month changes. Series adjusted with a 3x5 filter had their next largest revisions at lags 1 and 24 for the seasonally adjusted series and lags 24, 36, and 11 for the month-to-month changes. Although we only had one series adjusted with a 3x9 filter for our retail study, the relative magnitude of its lag 36 revision suggests that it may be large enough to warrant a revision as well. Further study of real-world series adjusted with a 3x9 filter will help shed more light on later lag revisions and can help inform a revisions policy for these series.

Our results validate the rationale behind the current MRTS policy of revising seasonally adjusted estimates a year after their initial estimate. The overwhelming majority of large revisions at lag 12 in both the retail series and the simulated series suggest that this result is not unique to retail sales series alone. Thus, programs using the X-11 method of seasonal adjustment that are not revising every month should consider revising their seasonally adjusted estimates a year after their initial estimate and perhaps a year prior to any other revised unadjusted estimate.

One takeaway from our simulation study was that under certain conditions, the number of forecasts used to extend the series could affect the timing of the largest revisions. In our study, we saw some large revisions around lag 6, lag 7 and lag 18 when the forecast length was set to half the seasonal filter length (which is a multiple of 6 but not 12). However, this finding did not extend to our retail sales series.

For future research, we would like to investigate the effect of certain regressors, like outliers, trading day effects and moving holiday effects, on revisions. This could also include analysis of how far back to revise when adding new regressors to the model. Although we used these types of regressors in our retail sales series study, we did not study their individual impact on revisions here.

We are also interested in extending this research to other types of series, including series from other subject areas, stock series, and quarterly series. While it seems likely that the pattern and timing of revisions will be similar across these different types of series, it is important to verify. It would also be interesting to see if the seasonal pattern in the mean squared revision series is present in other types of series.

Acknowledgements

We are grateful for comments on earlier drafts from our reviewers Katherine Jenny Thompson, Demetra Lytras and Deanna Weidenhamer. We benefitted from discussions of revisions with William Davie, Jr., David Kinyon, Brian Monsell, and David Findley. We thank colleagues Thomas Evans of the U. S. Bureau of Labor Statistics, Susie Fortier of Statistics Canada, Alex Stuckey of the Australian Bureau of Statistics and Blanca Rosa Sainz López of the Instituto Nacional de Estadística y Geografía for providing background on their agencies' approach to revisions.

References

Box, G. E. P. and Jenkins G. M. (1970), *Time Series Analysis, Forecasting and Control,* San Francisco, CA: Holden-Day.

Dagum, E. B. (1988), "X-11-ARIMA/88 Seasonal Adjustment Method – Foundations and Users' Manual," Time Series Research and Analysis Division, Ottawa: Statistics Canada.

Elliott, D., McLaren, C.H., and Zhang, X. (2007), "Quantifying and Measuring Revisions in Time-Series Estimates," *Proceedings of the Survey Research Section*, American Statistical Association, Joint Statistics Meeting, Salt Lake City, July 29 to August 2, 2007. www.uow.edu.au/~craigmc/index.html

McKenzie, S. (1984), "Concurrent Seasonal Adjustment With Census X-11," *Journal of Business and Economic Statistics*, Vol. 2, No.3, July 1984. www.census.gov/ts/papers/jbes84skm.pdf

SAS Institute, Inc. (2008), "Sample 33092: Wald-Wolfowitz (or Runs) Test for Randomness." support.sas.com/kb/33/092.html

U. S. Census Bureau (2014), "Estimates of Monthly Retail and Food Services Sales by Kind of Business: 2014." www.census.gov/retail/mrts/www/mrtssales92-present.xls

U. S. Census Bureau (2015a), *Win X-13 Version 2.3: A Windows Interface for X-13ARIMA-SEATS*, Time Series and Related Methods Staff, Economic Statistical Methods Division, U. S. Census Bureau, U. S. Department of Commerce. www.census.gov/srd/www/winx13/WinX13Doc.html

U. S. Census Bureau (2015b), *X-13ARIMA-SEATS Reference Manual*, Version 1.1, Time Series Research Staff, Center for Statistical Research and Methodology, U. S. Census Bureau, U. S. Department of Commerce. www.census.gov/ts/x13as/docX13AS.pdf

Wald, A. and Wolfowitz, J. (1940), "On a Test Whether Two Samples Are From the Same Population," *Ann. Math Statist.* 11, 147-162.