Empirical Evaluation of X-11 and Model-based Seasonal Adjustment Methods

Richard Tiller, Daniel Chow, and Stuart Scott

Bureau of Labor Statistics.
Massachusetts Avenue, N.E, Washington, D.C. 20212, tiller_r@bls.gov

Introduction

For over three decades X11 has been the standard approach used to seasonally adjust time series at the Bureau of the Labor Statistics (BLS). In recent years the model based approach has gained ground as an alternative approach. This study compares X-11 and model-based seasonal adjustments for 82 series produced by the BLS. We use "X-11" to refer to the seasonal adjustment method developed by Shiskin et. al (1967) as implemented in the enhanced version X-12 ARIMA (Findley et. al., 1998). The model based approach we use is known as SEATS (Signal Extraction of ARIMA Time Series), originally implemented by the Bank of Spain (Gomez and Marvall, 1997). This paper was developed from a larger study conducted at BLS (Scott, Tiller, and Chow, 2007) to evaluate SEATS as a potential supplement to X11.

This study has greatly benefited by the availability of a beta version of the X12-SEATS program under development by the Bureau of the Census in collaboration with Agustin Maravall (Findley, Monsell, et al, 2005). This program provides a unified model-based pre-adjustment framework within which either approach may be invoked and provides a common set of diagnostic and evaluative tools for comparing results from the two methods.

Design of Study

<u>Data</u>

Our series are selected from three BLS programs: Current Employment Statistics (CES), the Consumer Price Index (CPI), and the Producer Price Index (PPI). The aim has been to select a diverse set with respect to classification (industry or item structure), level of aggregation, and difficulty in seasonal adjustment. All are national series. We use 12 years of recent data where practical. For some price series, fewer years are available or important interventions are not available for earlier years, so data spans as short as 8 years are used.

The X12-SEATS software (Findley, Monsell, et al, 2005) provides a controlled environment for comparing the decompositions from X-11 and SEATS. Both methods start from identical "RegARIMA" fits of the original series, which includes identical ARIMA models and pre-adjustment options such as outliers and calendar effects. The difference between the two methods lies in how the filters or moving averages are produced to perform the decomposition of the series into trend, seasonal and irregular components. SEATS filters are derived directly from the ARIMA model and hence are tailored to the specific properties of the series as reflected in the estimated model. In contrast, X11 selects from a set of predetermined filters that have been shown to work well for a wide variety of series.

There is a well known fundamental conceptual issue that must be considered in comparisons of estimated unobserved components from alternative seasonal adjustment methods. Because the components are in general unidentified, all methods must impose specific assumptions, either explicitly or implicitly, in order to decompose a series. When two approaches are based on different but plausible assumptions about the behavior of the unobserved components, no objective criteria exist for judging which empirical decomposition is best (Bell and Hillmer, 1984). Depoutot and Planas (1998) provide conditions under which X-11 and SEATS filters are similar and thus comparisons of the decompositions are meaningful. Filter comparability is data specific in the sense that it depends on the properties of the time series under analysis. We draw on this work in our evaluation of the methods applied to BLS data.

In the first stage of our work, X12/SEATS is run with automatic selection of X11 filters, ARIMA models, outliers, and adjustment mode (multiplicative or additive). Denoted as AUTO, this stage allows the software to adjust automatically each series, in order for us to gain insight into how well X11 and SEATS perform with as little user input as possible. There are some exceptions. Known calendar effects are included in the case of several CES employment series. For difficult price series, such as energy-related series, interventions and outliers used in the official adjustments are our starting point. Experience has shown that automatic detection is inadequate for these effects; for instance, an intervention variable commonly used for price series is not part of automatic detection. Also, CES calendar effects differ from calendar effects options provided by the software.

In the second evaluation stage, called the analyst adjustment stage and denoted ANALY we have selected a subset of series from each program for more in-depth analysis. Comparisons are made across X11 and SEATS results. When X11 quality control (QC) statistics or ARIMA model diagnostics show problems or defects we attempt to address them. Other times, in order to find a feasible decomposition, SEATS discards the model initially identified, in which case we review the model choice. In other cases, we try to improve on the choice of outliers from AUTO.

Diagnostic Testing

When performing seasonal adjustment for any method it is important to test for three basic conditions: (1) the observed series is seasonal, (2) the seasonal effects can be estimated reliably, and (3) no residual seasonality is left in the adjusted series.

A variety of diagnostic tools are available some of which are method specific and others of a generic nature that can be used across methods. For X-11 we have the F-test from the original X11 and the more extensive M and Q tests from X-11 ARIMA. X12 provides a set of generic tests that include sliding span diagnostics, frequency spectrum estimates, and revision history statistics that are suitable for comparing methods.

The individual M statistics and Q2 are scaled to lie between 0 and 3 with smaller values indicating a better adjustment. We adopt the following guidelines for interpreting these statistics:

less than 0.8	diagnostics favorable
between 0.8 and 1.2	gray area
greater than 1.2	diagnostic unfavorable

SEATS provides a theoretical framework for evaluating various diagnostics (Maravall, 2003). However, much of this work is still in the experimental stage and not ready for implementation (Findley et al, 2004 and Evans, Holan, and McElroy, 2006). Other SEATS specific diagnostics assess the quality of the decomposition: for each component in terms of the estimation error, the forecast errors, variances and rate of damping of the revisions as new data are added. For this study the most important SEATS diagnostics are the ones related to the overall adequacy of the ARIMA model fit. While model formulation and model adequacy are important for X11 adjustment, they are central to SEATS.

Table 1 below presents the key diagnostics we use to compare across methods and assess suitability of adjustment by the individual methods.

Table 1: Summary of Major Diagnostics

Type	Description	Abbreviation
Cross-method		
Presence of Seasonality	peaks at seasonal frequencies in the spectrum of the differenced observed series	Ori Pks
	stable F statistic from a one-way analysis of variance carried out on the SI ratios	Stable F
Residual Seasonality	Significant residual seasonal peaks in the spectra of the differenced seasonally adjusted series, the irregular, and the ARIMA residuals	rsdpks
Stability of Seasonal Factors	Stable F applied to the final seasonal component	Overall F
Reliability of Seasonal adjustment	Distribution of absolute percent revisions of the monthly seasonal adjustments, concurrent to near final	

Table 1: Summary of Major Diagnostics		
Type	Description	Abbreviation
	Median	Rmed
	75th percentile	R75p
	maximum of revisions	Rmax
	Sliding spans –distribution of maximum per cent difference (MPD) of seasonally adjusted monthly change across 2-4 span values	
	Median	Cmed
	60th percentile	C60
	maximum of revisions	Cmax
Smoothness	Root mean square of the first difference of seasonally adjusted series or trend series	RMSD
X11 Diagnostics	Stable F, described above	Stable F
	M7 measures amount of stable seasonality present relative to amount of moving seasonality	M7
	Q_2 weighted average of M1 and M3-M11 from the original set of 11 statistics	Q_2
ARIMA Model Diagnostics	Ljung-Box statistic for autocorrelations in model residuals for first 12 and 24 lags provides an overall goodness of fit test	LB
	average absolute in sample percent forecasting error for the last 3 years	AAPE
	The version of the AUTOMODEL option based on TRAMO provides a test for the presence of seasonality (seasonal differencing) and the stability of seasonality depends on the magnitude of the seasonal MA parameter	
SEATS Diagnostics	The most important diagnostics specific to SEATS are the ARIMA model diagnostics listed above.	

Results

Table 2 presents summary information on the adequacy of the ARIMA models automatically selected for each of the series in this study. Recall that both X11 and SEATS begin with identical REGARIMA models. Thirty-four percent of the series (28 of 82) fail at least one of the 4 principal diagnostic tests shown below. In the analyst (ANALY) stage acceptable models are found in most of the cases studied. For 4 of these series SEATS automatically rejects the initial model as unsuitable for seasonal adjustment and actually finds an alternative model with an adequate fit.

Table 2
Goodness of Fit Tests of Automatically Selected ARIMA Models

	LB12	LB24	AAPE	RSDPKS	One or more
# Models Failing Test	13	14	3	8	28

Key:

LB12[24] Ljung-Box test over first 12 [24] lags
AAPE Average absolute percent forecast error test
RSDPKS test for seasonality in the residuals

The overall quality of the initial seasonal adjustments by X11 is shown in Table 3. A total of 22 series (27%) are marginal or failing based on one or more of the 4 quality control tests shown below. There is a large gray area for these statistics, and some adjustments of these series are found to be adequate after further analysis. Also, in the analyst stage of this study we were able to make a number of improvements.

Table 3
X11 Quality Control Statistics

	Stable F (< 10)	M7 ($\geq$ 0.8)	Q_2	RSDPKS	One or more
# Series Marginal or Failing	20	13	3	1	22

Key:

Stable F- test for stable seasonality
M7 - amount of stable seasonality relative to moving seasonality
Q2 - weighted average of 10 quality control statistics, exclude cases where M7 $\geq$ 0.8
RSDPKS- test for seasonality in the seasonally adjusted series

The characteristics of the ARIMA models used by SEATS are shown in Table 4. In all but 6 cases these models are identical to the ones used by X11 for outlier detection and forecast extension. Note that 7 series are non-seasonal (no seasonal differencing or seasonal parameters). SEATS does not seasonally adjust these series but does estimate a trend and irregular component. X11 automatically seasonally adjusts all series, whether seasonal or not, and it is left to the analyst to determine the appropriateness of the adjustment based on an extensive set of diagnostic tests.

Interestingly, 80 percent of the series have airline models (011)(011), which take the following form,

$$\nabla\nabla^{12} y_t = (1+ \theta_1 B)(1+ \theta_{12} B^{12}) a_t$$

where ∇^k is the difference operator, $\nabla^k y_t = y_t - y_{t-k}$, B is the backshift operator, $B^k y_t = y^t - y_{t-k}$ and a_t is the white noise disturbance.

With only an ordinary θ_1 and seasonal θ_{12} moving average coefficient, this class of models is simple and relatively easy to interpret. The MA coefficients are related to the stability of the trend and seasonal components, respectively. Specifically, the closer these coefficients are to -1.0 the more stable are their related components. Of particular interest for this study is that for many airline models there is a close correspondence between the model-based seasonal adjustment and the X11 adjustment. Planas and Depoutot (1998) demonstrated that a large set of X11 trend/seasonal filter combinations can be closely approximated with the Wiener-Kolmogorov filter of SEATS derived from the airline model with appropriately selected regular and seasonal moving average parameters. There are also major differences between the two approaches particularly when seasonality is deterministic or rapidly changing which corresponds to extreme values of the seasonal MA parameter, θ_{12} near -1.0 and at the other extreme close to zero or positive. We will examine this correspondence in more detail shortly.

Table 4 also shows the distribution of the MA coefficients. The bin values represent the interval greater than the adjoining lower bin and less than or equal to the current bin number. Focusing on the seasonal coefficient, we see that 22 percent of the models have values between -0.75 and -0.99, which implies stable stochastic to fixed seasonal patterns. In fact, 10 series have seasonal coefficients between -0.99 and -1.0, which implies a deterministic pattern where the seasonal factor for any given month is fixed across all years. Thirty-six percent have rapidly changing seasonal patterns (-0.50 to 0.00). The trends are much less stable than the seasonal patterns since no more than 6 series have MA coefficients smaller than -0.50.

Table 4 **Characteristics of SEATS ARIMA Models** **(6 models changed from initial identification used by X11)**							
Model Form			Distribution of MA Coefficients					
			θ_1			θ_{12}		
Type	#	%	Bin	Freq	%	Bin	Freq	%
(011)(011)	66	80.5	-0.99	1	1.25	-0.99	10	13.33
(111)(011)	4	4.9	-0.74	1	1.25	-0.87	11	14.67
(110)(011)	1	1.2	-0.50	4	5.00	-0.74	14	18.67
(012)(011)	1	1.2	-0.25	9	11.25	-0.62	13	17.33
(021)(011)	1	1.2	0.00	16	20.00	-0.50	14	18.67
(211)(011)	1	1.2	0.25	27	33.75	-0.37	5	6.67
(311)(011)	1	1.2	0.50	9	11.25	-0.25	4	5.33
Non-seasonal	7	8.5	0.74	5	6.25	-0.12	3	4.00
Total	82	100.0	1.00	8	%	0.00	1	%

How similar the SEATS and X11 seasonal adjustments are in practice depends not only on how comparable their filters are but also on how each method automatically selects a filter. SEATS filter selection is automatically determined by the values of the estimated MA parameters. Since these parameters vary continuously, SEATS provides an infinite range of filters from which to choose.

X12 ARIMA optionally implements automatic seasonal filter selection based on the so-called moving seasonality ratio (MSR) procedure in ARIMA/88 (Dagum, 1988). This procedure selects from a 3×3 (5 year length), 3×5 (7 year length) or 3×9 (11 year length) seasonal moving average. The choice of which moving averages to use depends on the global MSR ratio which is the average absolute year-to-year percentage change in the irregular ($\bar{I}$) to that in the seasonal component ($\bar{S}$).

$$MSR = \bar{I} \big/ \bar{S} = \sum_j n_j \bar{I}_j \Big/ \sum_j n_j \bar{S}_j$$

where, $\bar{I}_j = \left(\sum_{i=2}^{n_j} I_{i,j} / I_{i-1,j} \right) \Big/ n_j - 1,\ \bar{S}_j = \left(\sum_{i=2}^{n_j} S_{i,j} / S_{i-1,j} \right) \Big/ n_j - 1, \quad j = 1 \ldots 12$ n_j = number of years with data for month j.

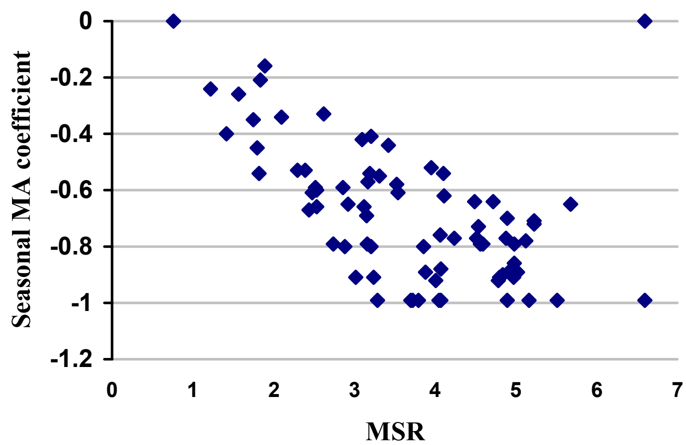
The larger is MSR the longer the seasonal filter tends to be and the smoother is the evolution of the estimated seasonal factors. This makes intuitive sense because a stable seasonal will have a small over-the-year change and thus tend to have a large MSR. Conversely, a small MSR implies the seasonal is changing rapidly requiring a shorter filter. The actual selection process is implemented by dividing the potential MSR values into broad zones as shown in Table 5. Two buffer zones, B and D, have been added to prevent frequent switches in filter length for MSR values near the boundaries. If a value falls into either of these zones, up to 5 years of the latest data may be dropped, the MSR recomputed and if it continues to be in either B or D the 3x5 filter will be selected.

Table 5 Criteria for selection of the X11 seasonal moving average based on MSR				
A 3×3	B	C 3×5	D	E 3×9
0	2.5	3.5	5.5	6.5

MSR

To see how this process relates to SEATS selection of seasonal filters we show in Figure 1 a scatter plot of the estimated seasonal MA coefficients with the empirical MSR values for our series. As expected, large values of the MSR tend to be associated with small values of θ_{12}. The correlation coefficient is -0.71. This negative relationship, however, is weakest for θ_{12} values near -1.0 where the associated MSR values range from about 3.3 to 6.6.

Figure 1: Scatter Plot of Seasonal MA coefficients and MSR values



Matching the selection criteria in Table 5 with the range of MSR values in Figure 1, we see that the X12 automatic selection procedure is strongly oriented to short filters. In Table 6 we make a direct comparison of the X11 and SEATS seasonal filters selected for each of our series. Specifically we compare the frequency of the length of the X11 seasonal filter closest to the SEATS filter with the frequency of the actual seasonal filter lengths selected by the MSR procedure.

Of the 75 series identified by SEATS as seasonal 28 percent have highly stable seasonality implying the use of the long 3×15 seasonal filter. An additional 18.7 percent have fairly stable seasonality implying the use of a 3×9 seasonal filter.

X11 relies on the 3×5 filter for 65 series in contrast to SEATS use of this filter for 27 series. The two methods do closely agree on the use of the 3×3 filter. The longest seasonal filter selected automatically by the MSR procedure is the 3×9 but for only two series. One of these series appears to be non-seasonal from the SEATS perspective since it has an ARIMA model with no seasonal part. Nevertheless, it generated a very high MSR value. The 3×15 filter cannot be selected by X11 since it requires 20 years of observations which exceeds the length of our longest series. The stable seasonal option, a simple average of the detrended values for each month separately, is available but is excluded from the MSR selection procedure. X11 will default to this filter only for short series.

Table 6
Seasonal Filter Comparisons

MA12 parameter intervals	Frequency of values	% of seasonal series	X11 filter closest to SEATS	Frequency of MSR selection of X11 filters	% of all series (82)
-0.88 to -1.00	21	28.0	3×15	0	0
-0.75 to -0.87	14	18.7	3×9	2	2.44
-0.51 to -0.74	27	36.0	3×5	65	79.27
0.00 to -0.50	13	17.3	3×3	15	18.29
Seasonal	75				
Non-seasonal	7				

Total	82	100.00		82	100.00

We investigate the appropriateness of SEATS selection of longer seasonal filters by looking at the goodness of fit of those airline models with seasonal coefficients of -0.9 or lower and more generally by examining their spectral test for residual seasonality in the corresponding seasonally adjusted series. Only 4 out of the 17 airline models with small seasonal MA coefficients showed clear evidence of lack of fit. There is also the possibility that the use of a long filter could lead to under adjustment resulting in residual seasonality in the adjusted series. Table 7 reports no residual seasonality in almost all of the X11 and SEATS adjustments (this is also true for the irregular series). A minor exception is the three SEATS adjusted PPI series. These are special series having ARIMA models with no seasonal part. In this case SEATS performs no seasonal adjustment while X11 does but with poor quality control statistics.

Table 7: Number of Seasonally Adjusted Series with Residual Seasonality in the Spectrum of their First Differences

	CES	CPI	PPI
X11	0	1	0
SEATS	0	1	3

The CPI for eggs illustrates how the two methods differ in over-the-year variability in seasonal factors when θ_{12} is near -1.0. The SEATS adjustment is deterministic while X-11 selects a 3×5 moving average for the seasonal filter (MSR = 4.9). The seasonal sub-plots are shown in Figures 2 and 3. Even though the X11 factors are changing rapidly its average seasonal pattern is very similar to SEATS. Looking at the spectrum of the first differenced original and adjusted series in Figure 4, SEATS compares favorably with X11. Both methods successfully remove the seasonal peaks in the original series but X11 removes more variation around the seasonal frequencies as a consequence of using the shorter 3x5 filter. The seasonally adjusted series are compared in Figure 5. Seasonality is not particularly strong in the original series and the two adjusted series appear fairly close even with very different seasonal stability properties.

For a few series in the analysts stage of our study we experimented with changing the automatic MSR choice of a 3x5 seasonal filter to a 3x9 when θ_{12} was close to -1.0. Basing the X11 seasonal filter choice on θ_{12} tended to result in improvements over the MSR procedure as shown in Table 8 below.

Table 8 : X11 Quality Control Statistics for 3×5 and 3×9 Seasonal Filters

	#2 Diesel Fuel		Mixed Fertilizers		Publication & Printed Matter		Switching Equipment	
Filter length	3×5	3×9	3×5	3×9	3×5	3×9	3×5	3×9
stable F	7.60	8.5	22.7	35	57.3	94.3	6.9	10.1
M7	0.83	0.78	0.69	0.44	0.31	0.22	0.96	0.7
Q2	0.78	0.55	0.47	0.31	0.25	0.13	0.66	0.59

Table 9 presents smoothness comparisons for the seasonal factors and the trend. Our focus here is on seasonality; we discuss the trend shortly. Our measure of smoothness is the stable F-statistic applied to the seasonal factors (Table 1). This table shows that overall SEATS seasonal factors are more stable for about two-thirds of our series. This follows, of course, from X11's use of shorter seasonal filters. Other studies have found similar results. Depoutot & Planas found that out of 7,372 Airline series, the closest approximating X11 filter was either the 3×15 or 3×9 for 56 percent of the series and only 26% for the 3×5.

Table 9: Smoothness Comparisons for 75 Seasonal Series

	Number	% of total
X11 seasonal factors smoother than SEATS	22	29.3

Table 9: Smoothness Comparisons for 75 Seasonal Series		
	Number	% of total
X11 trend smoother than SEATS	75	100.0

Having illustrated how the MSR procedure of X11 is prone to selecting shorter filters than SEATS with series where the seasonal MA coefficients were close to -1.0 we now look at the opposite extreme where seasonality is rapidly changing. Even here, SEATS sometimes shows an advantage with its ability to select from a much wider range of filters than X11.

Scenic and Sightseeing Transportation series presents an extreme case of model failure resulting from the introduction of NAICS in 2001 and the reconstruction of the original SIC based series to reflect NAICS for previous years. In 2001 there is a clearly visible break in seasonality (see Figure 6). The AUTOMDL spec of X12/SEATS initially fits a (3,0,1) (0,1,1) model with parameter values that yield an invalid decomposition. SEATS first tries to approximate this model with an airline model and estimate its parameters from the data. This fails, so SEATS goes to default values of 0.99 for the regular MA coefficient and 0 for the seasonal MA. The first value implies an unstable trend and the second rapidly changing seasonality. The seasonal factors derived from this model are much less stable than the X11 factors (Figures 7 and 8), but this is an advantage. It helps SEATS adjust to the abrupt change in seasonal pattern, which results in a better adjustment than that provided by X11 (Figure 6). Of course, the proper way to seasonal adjust this series is to split it into two parts and adjust each part separately, but this exercise illustrates the flexibility of SEATS filters to adapt to unstable patterns.

Trend

We now consider comparisons of the trends for both methods. We use as our smoothness measure the root mean square of the first difference of the trend series. Based on this measure, Table 8 reports that X11 produces smoother trends than SEATS for all series. Neither approach, however, produces very smooth trends. This is well known from studies of the spectral properties of the frequency response functions for the respective trend filters. Both methods produce trend filters that have poor cut off properties in the high frequency range. For long run analysis further smoothing of the trend may be necessary for both approaches (see Dagum, 1996 and Kaiser and Maravall, 2001).

Reliability

We present sliding span statistics for the initial X11 adjustments in Table 10 as an additional evaluative tool. Our focus is on the 60th percentile of the distribution of the month-to-month change in the seasonally adjusted estimates across overlapping spans. The rule of thumb is that three percent is too high for C60. Based on this critical value we could not identify a single inadequate adjustment. This seems due to the fact that most of our series do not have very large seasonal swings over the year. Because of difficulties in calibrating the critical value for the C60 statistic, we have not found this test useful for distinguishing between adequate and inadequate seasonal adjustments. It is, however, useful for comparing methods.

Table 10
60 th Percentile (C60) of Change in X11 Adjusted Series Across Overlapping Spans

interval	Frequency of C60
0 to 0.81	67
0.82 to1.08	5
1.09 to 1.62	1
1.63 to 2.20	2
2.21 to 3.00	0
Total	75

In Table 11 we compare the sliding spans and revision statistics for X11 and SEATS. In order to have a fair comparison, we restrict our analysis to those series with seasonal coefficients no smaller than -0.695 since the corresponding model filters are very close to the X11 filters in terms of their effective filter length (Findley et al, 2003). For the sliding spans percentiles X11 has smaller values than SEATS for less than one third of the series. For the revision statistics, the 75th percentile for X11 is smaller in slightly over half of the series but for the median and the maximum percent revision SEATS produces smaller values in more than half the series. The actual numerical percentage point differences between X11 and SEATS (Table 12) are not very large, however.

Table 11
Stability Comparisons, X11 Smaller than SEATS
(33 series with airline ARIMA models with parameter values allowing sliding span comparisons)

	cmed	c60p	cmax	rmed	r75p	rmax
Number	10	8	8	13	17	16
%	30.3	24.2	24.2	39.4	51.5	48.5

Table 12
Distribution of Sliding Span Differences (X11 minus SEATS)

	cmed	c60p	cmax	rmed	r75p	rmax
median	0.01	0.01	0.09	0.00	-0.01	0.00
mean	0.02	0.03	0.18	0.07	0.03	0.03
max	0.42	0.50	2.85	1.90	1.06	1.22
min	-0.08	-0.09	-1.16	-0.18	-0.07	-0.64

Robustness features

The dependence of SEATS on the overall ARIMA model is its defining characteristic. This raises two potential problems. First, what happens if an adequately fitting model can not be found for a series? Inadequate models may have an adverse effect on X11 but the potential damage is much greater for SEATS. Second, the best-fitting ARIMA model is not necessarily decomposable into trend, seasonal and irregular components. Certain restrictions have to be imposed and not all models will satisfy these restrictions. What happens if the ARIMA model identified is not suitable for deriving a seasonal decomposition?

These two problems, poorly fitting models and models inappropriate for seasonal adjustment, would seem to limit the applicability of SEATS. In practice, these two problems are often related: a bad fitting model is often non-decomposable. When this happens, SEATS seeks to find a decomposable model that adequately fits the data. If this search fails, SEATS will default to an airline model (with judicious selection of parameter values) which can produce filters that rival the X11 filters in terms of good empirical properties.

Most of the time X12/SEATS' automatic selection of model and outliers works acceptably well, but by no means always. Close to 20% of the seasonal series exhibit lack of model fit. Out of the 5 series where an adequate but non-decomposable model was selected, 4 had failing or marginal diagnostics. SEATS found better fitting decomposable models for 3 of the series. Surprisingly, in a few examples the default filter chosen by SEATS, when no adequate model fit could be found, actually outperformed what AUTOMODEL selected. From these examples as well as others we did not find consistent evidence, in the presence of model failure or the lack of a decomposable initial model selection, that SEATS performed worse than X11 in terms of quality of the seasonal adjustment. An issue that remains, however, is gauging confidence in the results when model fit criteria are not met.

Simple moving averages like those used by X11 are not robust against outliers. Prior to the introduction of intervention models and automated outlier detection, the original developers of X11 provided a partial solution to this problem which is still routinely implemented. The software computes a moving standard deviation with which it checks a preliminary estimate of the irregular or noise term for outliers. With default settings, X11 invariably identifies a number of points as extreme and down weights them to avoid their affecting trend and seasonal estimates. While the technique used is old-fashioned, it still provides X11 with a considerable level of robustness against the occurrence of one-time outliers in real time.

Summary

The major strength of SEATS is that it selects from an infinite number of filters based on the estimated characteristics of the individual series. In contrast, X11 has a much smaller set of filters to choose from. The X11 automated filter selection procedure is further limited to only three seasonal filters (and three Henderson filters). While X11 is capable of handling a wide range of patterns, SEATS wider flexibility and its continuous filter selection process does provide important additional capabilities but it does not come without costs. Each method has its own strengths and weaknesses.

Identification of deterministic seasonality by SEATS in some of the series is a major difference with X11. Except for series shorter than 5 years, the X11 selection rule always assumes seasonality is stochastic. In terms of our generic tests this does not result in unsatisfactory decompositions. If the user is not convinced that a deterministic result from SEATS is appropriate, perhaps due to a suspect model, X11 may be the better choice. The use of shorter filters does protect against under adjustment.

The danger with shorter filters is that the seasonal component will absorb additional irregular variation and remove more variability from the original series than necessary for an adequate adjustment. For many of the series where SEATS estimates fixed seasonality, X11's automatic choice of the 3×5 filter appears to over-adjust the series as evidenced by both QC and spectral diagnostics. In the X11 ANALY adjustments, use of the 3×9 filter allows for moving seasonality, but gives a more stable seasonal and improved diagnostics.

Most often SEATS views seasonality as stochastic but still tends to produce more stable seasonal factors than X11. SEATS identifies 28 series as highly stable but still stochastic as compared to only two series identified by X11 with filters longer than the 3×5. In these cases X11 filters tend to be in close agreement with SEATS filters but the X11 automated procedure selects shorter filters. Using SEATS model parameters to suggest a longer seasonal filter for X11 may be a helpful solution. In particular, for some difficult-to-adjust price series, this has potential for reducing the amount of intervention treatment required. For many series, using the 3×9 filter reduces variability in the X11 seasonal component to an acceptable level, as evidenced by diagnostics and graphs.

Full flexibility of the SEATS filter selection may not always be desirable. This may be the case when observed data is frequently revised or for short series where the ARIMA parameter estimates may be unreliable. Even though the airline model is robust a careful check of the adequacy of the model is important particularly when the estimated coefficients take extreme values. Special problems can arise with non airline models. For example, a model with one or more ordinary AR parameters may lead to undesirable decompositions particularly where complex roots are assigned to the seasonal component. This can lead to unstable seasonal factors which in some cases may be avoided by allocating the roots to the transitory component.

Taking into account the strengths and limitation of both X11 and SEATS, our study concludes that relying solely on one method is undesirable. The combined use of both methods provides the user with more tools than either separately to tackle the difficult problems encountered in seasonal adjustment. This point has been long recognized, if not often practiced, as evidenced by a quote from early developers of the model based approach (Box, Hillmer, and Tiao, 1978),

> "The empirical method [X11] and the model-based method … are sometimes thought of as rivals. But they are only rivals in the same sense that the two sexes are rivals. In both cases, isolation is necessarily sterile, while interaction can be fruitful."

With the availability of X11 and SEATS in a truly integrated package we are now capable of putting this into practice at BLS.

References

Bell, William R. and Hillmer, Steven C. (1984), "Issues Involved with the Seasonal Adjustment of Economic Time Series" with comments and reply, Journal of Business & Economic Statistics 2, 291-349

Box, G.E.P., HILLMER, S.C. and TIAO, G.C. (1978). "Analysis and Modeling of Seasonal Time Series," in Seasonal Analvsis of - Economic Time Series, ed. A. Zellner. Washington. D.C.: U.S. Dept. of Commerce. Bureau of the Census. 309-334.

Dagum, E. B. (1988). X-11-ARIMA/88 seasonal adjustment method - foundations and users manual. Statistics Canada.

Dagum, E. B. (1996)"A New Method to Reduce Unwanted Ripples and Revisions in Trend-Cycle Estimates from X11 ARIMA", Survey Methodology, 22, n.1, 77-83..

Depoutot, R. and C. Planas (1998), "Comparing Seasonal Adjustment and Trend Extraction Filters with Application to a Model-based Selection of X11 Linear Filters," Tech. Paper TP361, Joint Research Centre, Ispra, Italy

Evans, T. D., S. H. Holan, and T. S. McElroy (2006), "Evaluating Measures for Assessing Spectral Peaks," 2006 Proceedings of the American Statistical Association, Business and Economics Section [CD-ROM], Alexandria, VA, forthcoming.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," with discussion, Journal of Business and Economic Statistics 16, 127-177

Findley, David F., Brian C. Monsell, et al (2005), X-12-SEATS Reference Manual, Version 0.3, Bureau of the Census, Washington, DC.

Gomez, V. and A. Maravall (1996), Programs TRAMO (Time series Regression with ARIMA noise, Missing values, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series): Instructions for the User, Working Paper 9628, Servicio de Estudios, Banco de Espana

Kaiser, R. Maravall A., 2001. Measuring Business Cycles in Economic Time Series, Lecture Notes in Statistics, 154, Springer-Verlag, New York.

Maravall, A. (2003), "A Class of Diagnostics in the ARIMA-Model-Based Decomposition of a Time Series," memorandum, Bank of Spain

Scott, Stuart, Richard Tiller, and Daniel Chow (2007), " Empirical Evaluation of X-11 and Model-based Seasonal Adjustment Methods," Joint Statistical Meetings 2007, Business & Economic Statistics Section

Shiskin, J., Young, A. H., and Musgrave, J. C.(1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Bureau of the Census Technical Paper 15, U. S. Department of Commerce, Washington, DC

Figure 2: Eggs , X11 Seasonal subplots)

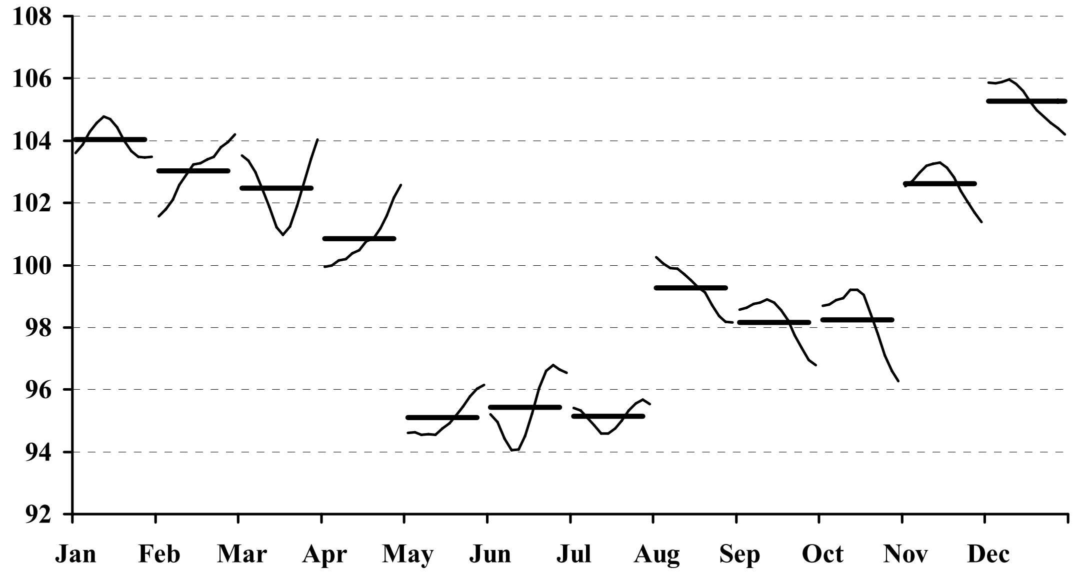


Figure 3: Eggs , SEATS Seasonal subplots)

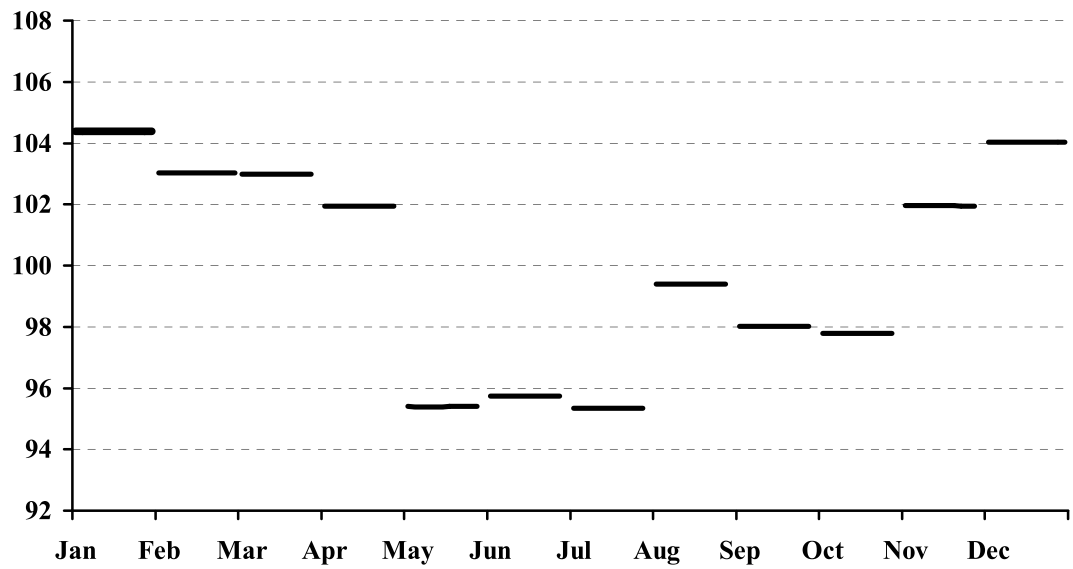


Figure 4: Eggs, Spectrum of Original and Seasonally Adjusted Series

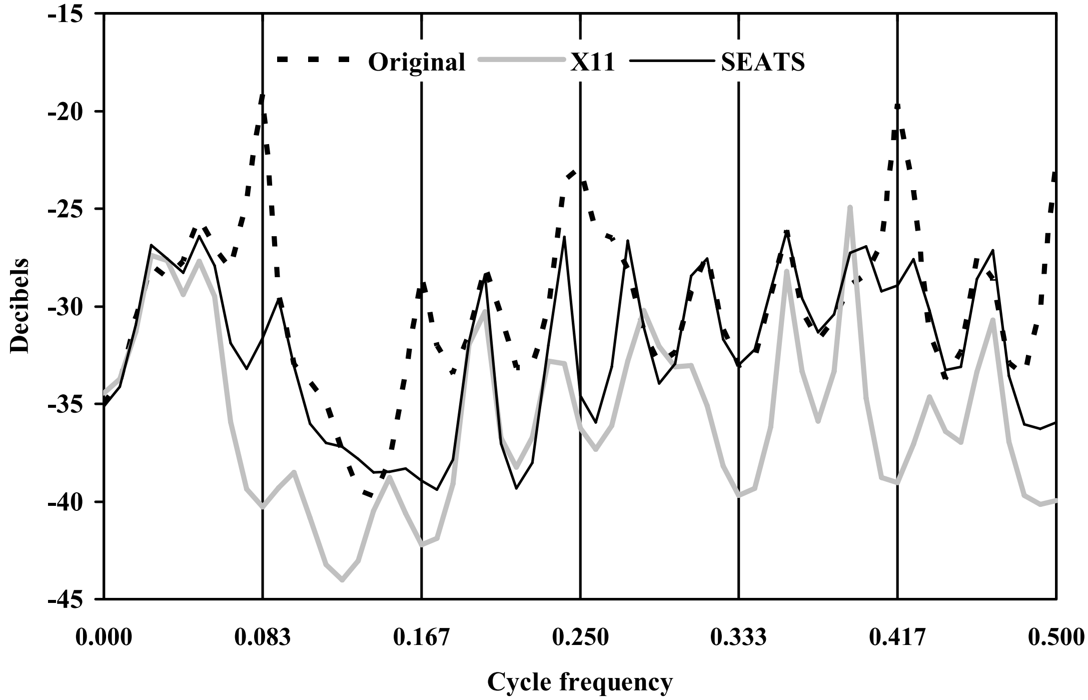


Figure 5: Eggs, Original and Seasonally Adjusted Series

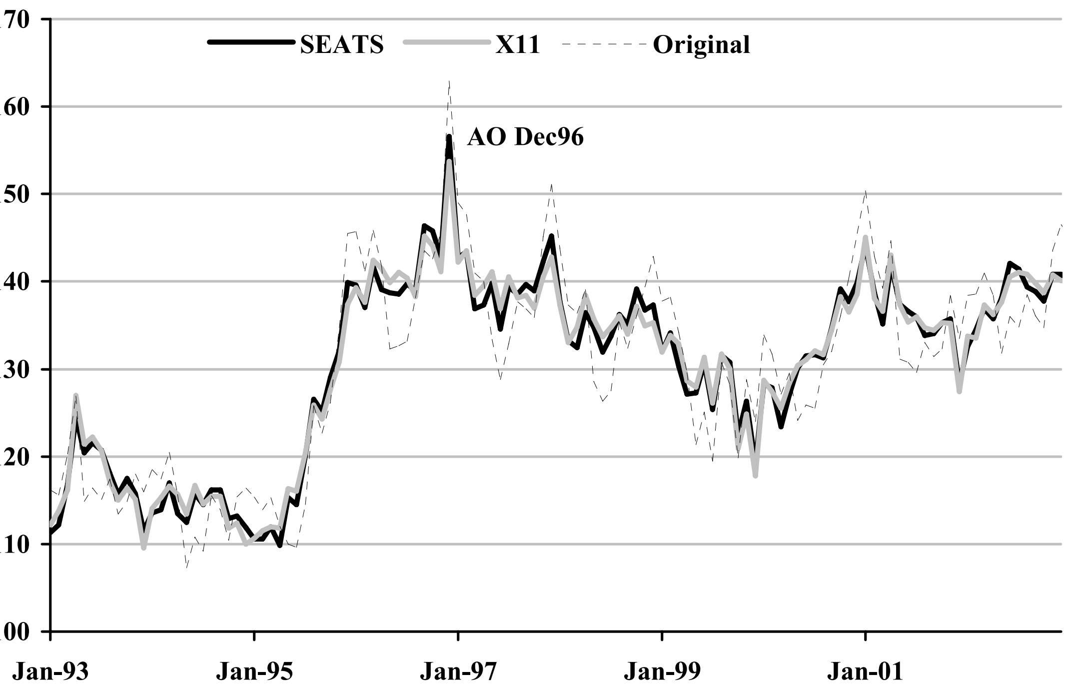


Figure 6: **Scenic Transportation, Original Seasonally Adjusted Series**

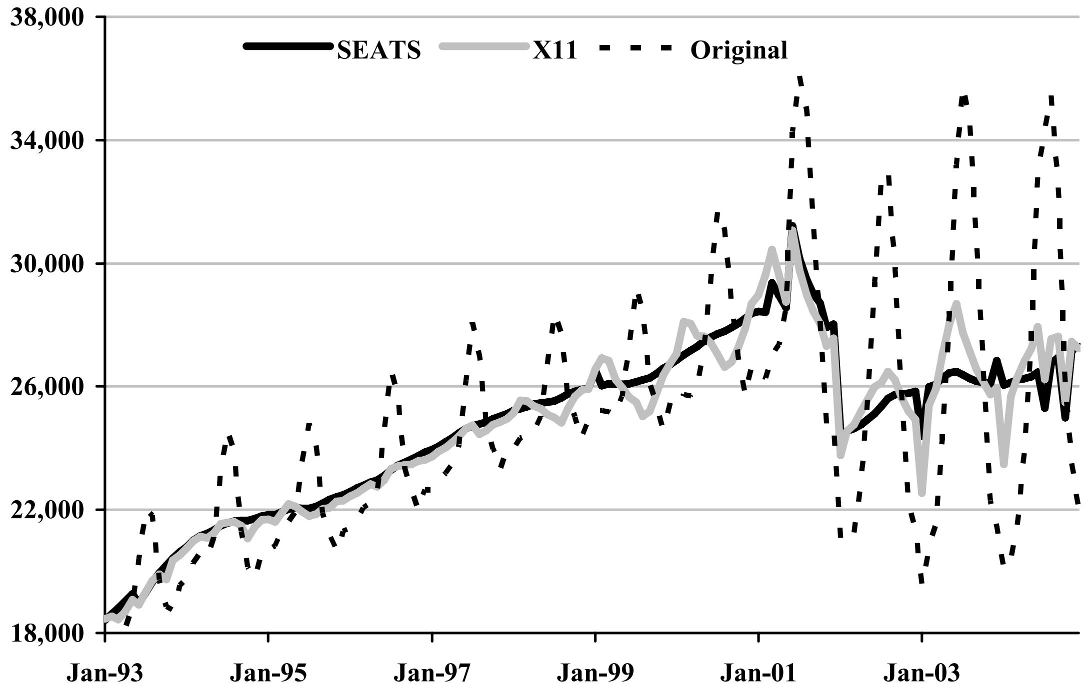


Figure 7: **Scenic Transportation , X11 Seasonal subplots**

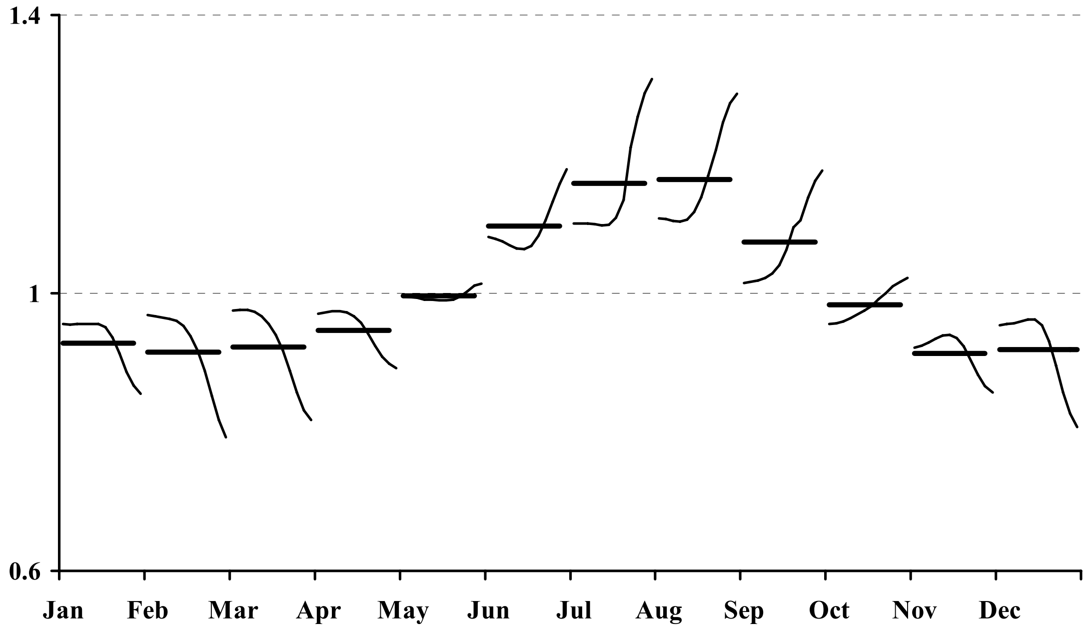


Figure 8: **Scenic Transportation , SEATS Seasonal subplots)**