Empirical Evaluation of X-11 and Model-based Seasonal Adjustment Methods December 2007

Stuart Scott, Richard Tiller, Daniel Chow
U.S. Bureau of Labor Statistics

KEY WORDS: time series decomposition, X-12-ARIMA, TRAMO/SEATS

1. Scope and main findings

Scope

ARIMA-based and X11 (X-11) seasonal adjustments are compared for time series from three major economic areas, employment, consumer prices, and producer prices, all statistics produced by the U.S. Bureau of Labor Statistics (BLS). The U.S. Census Bureau, in collaboration with the Bank of Spain, has produced experimental software, which we call X12/SEATS, to handle both X11 and ARIMA-based adjustments. We shorten the latter to SEATS, which comes from the Bank of Spain's TRAMO/SEATS software, with SEATS an acronym for *signal extraction in ARIMA time series*. X12/SEATS greatly facilitated the work of a BLS team which carried out the evaluation. A third method, structural model-based adjustment, has been studied, but it will not be discussed here, since software limitations make it less suitable for production.

The scope of the evaluation is 82 series from the three program areas, 25 from the Current Employment Statistics program, 35 from the Consumer Price Index, and 22 from the Producer Price Index. "Automatic" adjustments (AUTO) are evaluated for all 82 series and "analyst" adjustments (ANALY) for a subset of 19 series. For the AUTO stage, the software chooses the ARIMA model, the adjustment options, and the outliers. For the ANALY stage, which is more labor-intensive, team members attempt to improve on results from AUTO.

Many other studies of this type have been carried out, including Census studies Hood & Findley (1999) and Hood (2002), to be discussed in Section 5, and unpublished studies from a Eurostat 1998 conference. While the number of series in the present study is not particularly large, this study is perhaps notable for being carried out by a BLS-wide team of seasonal adjusters and for detailed analysis of each series. Furthermore, some connections can be made between our empirical results and theoretical results, for instance, in Depoutot and Planas (1998).

Main findings

1) Both methods give satisfactory results overall.

Both give acceptable adjustments for a majority of series. For series with little evidence of seasonality, both tend to give weak or negative statistics. There is not much evidence of residual seasonality with either method.

2) SEATS is more flexible in estimating a seasonal component.

This is an obvious finding, since its models have a continuum of parameter values, while X11 has a discrete set of filter choices. The difference is appreciable for a number of series in our study, especially in the AUTO stage.

3) SEATS tends to estimate a more stable seasonal component in the AUTO stage.

For several series, this is preferable to X11, which estimates a highly variable seasonal. For a number of series, SEATS estimates a deterministic seasonal, which may or may not be preferable.

4) X11 produces a smoother trend (trend-cycle).

This finding fits with studies, e.g., Dagum & Luati (2002), which show favorable smoothing properties of X11's Henderson filters.

5) With SEATS, sometimes it's difficult to find a good-fitting model with a valid decomposition into components.

The BLS team has recommended that BLS programs consider SEATS adjustments. The question arises: given that neither method entirely dominates and the current X11 method mostly works acceptably well, why go to this extra work? X11 appears to "over-adjust" some series. Also, the seasonality is mild in many price series and can be difficult to adjust, especially for series subject to outliers, such as energy and food prices. Thus, SEATS adjustments may suggest how to improve X11 adjustments or may replace them, when warranted. The next sections attempt to illustrate and explain these findings.

2. Diagnostics

Before making comparisons across methods, it is important to (1) test for presence of seasonality in the observed series and (2) test acceptability of seasonal adjustment with each individual method, using diagnostics pertinent to that method.

X11 diagnostics

We lean heavily on the traditional X11 quality control statistics developed by Statistics Canada and described in Ladiray & Quenneville (2001). We use especially the stable F statistic F_S, the combined statistic for identifiable seasonality M7, and the summary Q measure (the version omitting M2). The individual M statistics and Q are scaled to lie between 0 and 3 with smaller values indicating a better adjustment. We adopt the following viewpoint for interpreting these statistics:

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

We also make use of M10 and M11, which measure the stability of seasonal factors in recent years. When the other statistics lie in the gray area, these help us decide whether an adjustment is acceptable or not. They usually confirm visual impressions from the monthplot (described below).

ARIMA modeling diagnostics

Our principal tool for assessing model adequacy is the widely-used Ljung-Box (LB) statistic, built from autocorrelations of residuals. Of particular interest are autocorrelations at low lags, say 1 to 4, and at seasonal lags 12 and 24. Low LB p-values (below .05) at lags 12 and 24, denoted LBP12 and LBP24, result from one or more high residual autocorrelations and indicate model inadequacy. While model formulation and model adequacy are important for X11 adjustment, they are central to SEATS. X12/SEATS generates forecast error statistics based on the model. It reports AAPE (average absolute prediction error) for the last three years of the input data. Large AAPE values, say 15% or more, indicate shortcomings in the model.

In some cases, the initial model does not yield a suitable decomposition. Then we need to review the replacement model and possibly consider additional models. Here again, the LB statistics are useful.

Of particular interest is the seasonal moving average parameter. When it is close to -1, the seasonal factors are highly stable; when close to 0, the factors tend to change rapidly. Series graphs and knowledge of the series can help assess how much movement in the seasonal is desirable.

Additional SEATS diagnostics

SEATS provides a theoretical framework for developing diagnostics. However, some proposed diagnostics (e.g., Findley, McElroy, & Wills, 2004) are still in the experimental stage and lack established guidelines for use. Other diagnostics assess the quality of the decomposition and the rate of damping of revisions as new data are added. There are a number of statistics for assessing normality of residuals. In general, some departure from normality may not be important, but excessive skewness or kurtosis may suggest additional checking for outliers. Further study of diagnostics from SEATS is recommended.

Cross-method statistics

Presence of seasonality

The first step is to assess presence of seasonality. Our principal test is to look for peaks at the seasonal frequencies in the spectrum of the differenced observed series. The seasonal frequencies correspond to behavior occurring every 12 months, every 6 months, etc. As already mentioned, a key statistic for X11 is the stable F statistic. For SEATS, selection of an ARIMA model with seasonal differencing provides evidence for seasonality. We have found the spectral evidence extremely useful. For only a handful of series in the study does either the X11 or SEATS method give strong evidence for seasonal adjustment when the spectrum does not.

Residual seasonality

After seasonal adjustment, we can check for residual seasonality using the spectra of the differenced seasonally adjusted series, the irregular, and especially important for SEATS, the ARIMA residuals. Significant residual seasonal peaks in the spectrum (RSDPKS) suggest that the seasonal adjustment is inadequate. Here, significance is based on the empirical criterion of the X12 software. Lag 12 and lag 24 autocorrelations in the residuals can also be examined. Research on formal statistical tests for significance of spectral peaks appears in recent work by Evans, Holan, and McElroy (2006).

Stability of seasonal factors

The monthplot displays seasonal factors grouped by month. Each monthly subplot shows the seasonal factors for that month across time, with a horizontal line at the mean value. This graph gives the overall seasonal pattern and shows how much movement is present in each month's factors. We find this graph very useful for comparing methods. Usually, we have less confidence in an adjustment with a rapidly moving seasonal component (unless that is required to capture seasonality). A simple statistic for gauging stability of seasonal factors is the F statistic from a one-way analysis of variance carried out on the final seasonal component. Recall that X11's stable F is computed from an intermediate stage of its adjustment. This "overall F statistic" to some extent quantifies the visual impression from a monthplot. Large values are usually preferred, but not always.

The statistic is essentially infinite for a deterministic seasonal, while we may prefer an adjustment showing some movement in the seasonal.

Reliability or stability of the seasonal adjustment

Assuming that seasonality is removed as mentioned above, another desirable property is reliability in estimation of the seasonally adjusted series. One widely used measure is the *revision in the estimates* as more data become available. For a given month, we compute the absolute difference between the initial estimate of the seasonally adjusted series and the final (or, in practice, a nearly final) estimate. We then compute an average of this difference across a specified test period. Given that our spans are relatively short, we specify a one- or two-year test period and compute a "final" value from a seasonal adjustment extending two years beyond the end of the test period. For most series, the test period is 2001-02, with final estimates from a seasonal adjustment with data through 2004. In our study, we focus on the 75th percentile, the 50th percentile and maximum of revisions.

A second set of diagnostics for stability comes from *sliding spans statistics* developed in Findley, Monsell, Shulman, and Pugh (1990). For most series in the study, we carry out seasonal adjustment on four spans of data, each successive span dropping an initial year and adding a year at the end. We focus on the 60th percentile and maximum of the maximum per cent difference (MPD) of seasonally adjusted monthly change. For both sliding spans and revision statistics, we want most of the values to be acceptably small, and, somewhat arbitrarily, we adopt as "most" 75% for seasonal adjustment revisions and 60% for seasonally adjusted change. There is a less stringent criterion for change, since one month with a large difference can cause two large change differences. We also want to check for large maximum values. We focus on percentiles, rather than the 3% rule sometimes suggested by Census, since the 3% rule is not very discriminating for series in the three programs studied here.

Most of the time the key statistics are acceptably low and the results from X11 and SEATS are close. Also, there are limitations in the statistics we have computed. It is known that time to convergence to final values varies in X11 with the length of the seasonal moving average or filter and in SEATS with the magnitude of the seasonal moving average parameter. Given the length of the input spans in the study, for a large number of series in the study "final" seasonally adjusted values are not achieved in computing either revision statistics or sliding spans statistics. This means that the statistics computed give an incomplete picture of stability.

Smoothness

A simple traditional smoothness measure is root mean square difference (RMSD) in the seasonally adjusted series,

$$RMSD = \left[\frac{1}{N-1} \sum_{t=1}^{N} (A_t - A_{t-1})^2 \right]^{1/2},$$

where A_t denotes the seasonally adjusted series at time t. While a small square difference at a particular time point is not always best, seasonal adjustment is a smoothing procedure and across an entire series a smaller $RMSD$ is likely to be advantageous provided there is no evidence of inadequacies in the seasonal adjustment.

Another set of statistics that can be applied to any method has been borrowed from traditional X11 diagnostics. We compute the *relative contribution of components to change* in the observed series. For a component *X* and a lag *d*, we compute

$$\bar{X}_d = \frac{1}{n-d} \sum_{t=d+1}^{n} \left| \frac{X_t}{X_{t-d}} - 1 \right|$$

(multiplicative case)

for $X = O, T, S, \text{and } I$ and for *d*=1,3,12, and 24. The relative contribution of the irregular I at lag d is $100\% \times \bar{I}_d^2 / \bar{O}_d'^2$, where $\bar{O}_d'^2 = \bar{T}_d^2 + \bar{S}_d^2 + \bar{I}_d^2$ and similarly for the other components. Simple change is used in the additive case. While these statistics don't directly assess quality of seasonal adjustment, they help quantify what is accomplished by a given method. In this study, we have focused on one-month change. We check whether SEATS' more stable seasonal still captures as much variation by examining $\bar{S}_1^2$. In contrast, evidence for X11 usually having a smoother trend comes from seeing smaller values of $\bar{T}_1^2$. Given these differing tendencies, it is of interest to see which puts more into the irregular.

All the above statistics are considered in the study (Scott, Tiller, and Chow, 2007), but this paper will emphasize key diagnostics and graphs. The specialized graphs come from SAS programs developed by Tom Evans, a member of the team. Similar graphs are available from a new version of X12/GRAPH (Lytras, 2007).

3. Examples

PMEAT (PPI for Meats)

Figure 1a. shows the observed and seasonally adjusted series from SEATS, AUTO stage, across the span 1993-2004. Some smoothing is accomplished with seasonal adjustment, but seasonality is not dominant. As seen in Figure 1b. the spectrum of the differenced unadjusted series has strong peaks at the first and 6th seasonal frequencies, which are essentially eliminated with the SEATS adjustment. Figures 1c. and 1d. contain the monthplots for the AUTO X11 and SEATS multiplicative adjustments, respectively. The overall patterns are similar, but SEATS factors are much more stable. For half the months, X11's factors vary between positive and negative seasonality (where 1 or 100% represents neutral seasonality). Barring strong economic evidence for large movement in seasonality, the SEATS adjustment is preferable, especially since, as already noted, its spectrum in Figure 1b. raises no concerns about residual seasonality. The AUTO adjustment by X11 employs the 3x5 seasonal filter. Given the preference for SEATS' more stable seasonal, the ANALY adjustment for X11 uses the 3x9 seasonal filter. This adjustment is still less stable than SEATS', but does represent an improvement, with the stable F improving from 5.8 to 9.2 and Q declining from .96 to .87. This series illustrates Finding 3 in Section 1, which, as seen later, occurs for a number of series.

PDIE2 (PPI for #2 Diesel Fuel)

Figures 2a. and 2b. show the monthplot and spectral plots for the X11 AUTO adjustment. As for PMEAT, the monthplot is highly variable; the stable F, at 7.6, is barely in the acceptable range. Still, from Figure 1b., the observed spectrum has a peak at the second seasonal frequency which is eliminated in the adjusted spectrum. Having a peak at the 2nd seasonal frequency fits with the overall pattern in the monthplot, since the latter's peaks and troughs are roughly six months apart. In the ANALY stage, for X11, the 3x5 seasonal filter is replaced by the 3x9, a new set of 3 outliers is selected, and a different ARIMA model is used for extrapolation. Fig 2c. shows that the resulting monthplot is much more satisfactory. The stable F improves slightly to 8.5; M7 and Q also lie in the acceptable range. SEATS selects a nonseasonal model in the AUTO stage, that is, it does not seasonally adjust the series. For the ANALY stage, a number of seasonal models are imposed. The (010) (011) model is selected and yields the deterministic seasonal in Fig 2d. The spectral plot shows no significant residual seasonality with this adjustment. However, there is some reservation with this adjustment, since, as seen in Table 1 below, the model has a failing Ljung-Box p-value at lag 24. The model used with the X11 ANALY adjustment is a constrained (01[6]) (011) model, with MA parameters at lags 6 and 12 only; the table below shows that this model gives an acceptable fit. Such models can yield acceptable decompositions, but at present the software cannot handle them. For this series, adjustment is accomplished more easily with X11.

Table 1. Model fit for two models for PDIE2

Model	LB24 p-value	AICC
(010) (011)	.003	626.0
(01[6]) (011)	.13	621.3

CBOOKS (CPI for Educational Books & Supplies)

AUTO monthplots in Figure 3a. and 3b. are quite similar and show considerable variability, with X11 somewhat more variable. An (011) (110) model is identified, which SEATS changes to an (011) (011) model for decomposition. This model has $\theta_{12} = -.67$, suggesting a 3x9 or 3x5 seasonal MA for X11, as will be discussed in Section 5. Figure 3c. shows the observed and SEATS seasonally adjusted series. We see the observed series is smooth and dominated by trend. The smoothness translates into a small irregular, leading X11 to select the 3x3 MA. Its overall statistics are satisfactory, but the M8-M11 statistics indicate an excess of movement in the seasonal component. Figure 3d. shows that with the more stable SEATS seasonal component, adjustment eliminates the seasonal peaks. Employing the 3x9 seasonal MA improves the X11 diagnostics displayed in Table 2 below.

Table 2. Quality control statistics for two X11 seasonal MA choices for CBOOKS

Seas. MA	3x3	3x9
Stable F	14.8	20.9
M7	.55	.46
M10	2.00	.71
M11	1.82	.71
Q	.63	.32

The SEATS AUTO adjustment is also flawed since the airline model used for decomposition has a failing Ljung-Box statistic LB(24) with p-value .002. Further review suggests that an (013) (011) model provides an acceptable fit and a valid decomposition. This model also suggests the longer seasonal MA for X11.

SEATS' seasonal parameter again provides helpful guidance for the choice of X11's seasonal MA. Overall, AUTO adjustments from both methods are similar and reasonable, but both need and benefit from analyst review.

SCENTRAN (Employment in Scenic & Sightseeing Transportation)

Figures 4a. and 4b. show the AUTO monthplots for the two methods. They are both highly variable, with the SEATS factors even more variable. In this case, high variability is definitely an advantage since the unadjusted series in Figure 4c. exhibits much stronger seasonality starting in 2001. X11 adjusts rapidly to the stronger pattern by using the 3x3 seasonal filter. SEATS accomplishes this by choosing its MA(12) parameter to be 0. Both methods suppress the seasonal peaks in the spectrum; Figure 4d. exhibits the results with SEATS. This series illustrates Finding 2, SEATS' greater flexibility in estimating a seasonal pattern. (The abrupt change in seasonality for this series can be attributed to difficulty in constructing historical data under a new industry coding system, rather than actual economic changes.)

4. Summary results and X11 filter selection

Neither method provides acceptable results for all series in the AUTO stage. Table 3 shows that one or more major diagnostic has a marginal or failing value for a sizable number of series, about one-fourth with X11 and one-third with SEATS. While some of these adjustments may be found to be adequate after further analysis, Table 3 demonstrates that statistical agencies need to review automatic adjustment of their series.

The methods agree well in their assessment of series with weak seasonality. Spectra of 13 series are weak or lacking in seasonal peaks. Table 4 shows the level of acceptability from the methods for these series. They agree on a characterization as "acceptable", "marginal", or "failing" for 8 of 13 series and for another 3, combining marginal and failing categories. There is one series acceptable to X11 but failing according to SEATS, and one acceptable to SEATS and marginal to X11.

Table 3. Key Diagnostics – Counts of Marginal or Failing Series

SEATS	LB12	LB24	AAPE	RSDPKS	One or more
# Models failing test	13	14	3	8	28
X11	Stable F (< 10)	M7 (> 0.8)	Q	RSDPKS	One or more
# Series marginal or failing test	20	13	3	1	22

Key for SEATS:

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

Key for X11:

- Stable F – test for stable seasonality
- M7 – combined test for identifiable seasonality
- Q – weighted average of 10 "quality control" statistics (omitting M2)
- RSDPEAKS – test for seasonality in the seasonally adjusted series

Table 4. Series Weak or Lacking in Seasonal Spectral Peaks

		SEATS			
		Acceptable	Marginal	Failing	Total
X11	Acceptable	2	0	1	3
	Marginal	1	2	1	4
	Failing	0	2	4	6
	Total	3	4	6	13

Table 5. 75th Quantile of Revision Statistics for Seasonally Adjusted Series

AUTO Stage	$\theta_{12} \le -.70$ (more stable)	$\theta_{12} > -.70$ (less stable)
X11 smaller	4	10
About equal	9	11
SEATS smaller	27	14
Total	40	35

Interestingly, the two series with acceptable diagnostics from both probably can be legitimately seasonally adjusted, showing that the spectral diagnostic is not foolproof.

Summary sliding span and revision statistics are small with both methods for most series, with SEATS tending to have slightly smaller values. Table 5 shows frequency counts for series with a smaller 75th quantile for revisions for each method, with a breakdown according to the value of the seasonal MA parameter θ_{12}. The table shows that in the cases with a less stable seasonal component, according to the model, each method is preferred roughly an equal number of times. However, for the cases with a more stable seasonal, SEATS dominates, with X11 having smaller revisions for only 4 or 10% of the cases.

We examine the empirical results in Table 5 in the light of theoretical results linking overall X11 filters (the so-called linear approximation filters) to ARIMA models. Building from the initial paper by Cleveland & Tiao (1976), Depoutot and Planas (1998) make connections between X11 filter choices and airline models. Table 6 contains a simplified (perhaps over-simplified) version of Table 1 of Depoutot and Planas. Based on a rather natural distance function between implied seasonal filters, it shows a range of

Table 6. Correspondence between X11 Seasonal MA Options and SEATS Seasonal Parameter θ_{12} (based on Depoutot & Planas, 1998)

θ_{12}	X11 Seasonal MA Option
(-.45, 0)	3x3
(-.65, -.45)	3x5
(-.75, -.65)	3x9
(-1.0, -.75)	3x15 or deterministic

Table 7. Choice of X11 Seasonal MA for 75 Series

Seasonal MA	AUTO (I/S Ratio)	D&P match
3x3	15	13
3x5	59	27
3x9	1	14
3x15 or deterministic	0	21

θ_{12} values corresponding to four possible X11 seasonal filter (MA) choices. Table 7 shows counts for X11's actual selection in the AUTO stage and counts based on Table 6. The agreement is quite close for the 3x3 filter, but X11 selects the 3x9 filter only once, while the Depoutot and Planas criterion suggests a longer filter than the 3x5 or a deterministic seasonal for 35 series. This confirms Finding 3 and shows that the PMEAT and CBOOKS examples are not isolated cases. Table 8 depicts X11's automatic selection of the seasonal filter. Based on a global "I/S" ratio, it makes one of three filter choices. The criterion (cf. Ladiray & Quenneville, 2001) works in the right direction: more movement in the seasonal, relative to the irregular, (i.e., small values of I/S) suggests a shorter seasonal filter and vice versa. However, the 3x5 filter appears to be chosen too often, in part due to the existence of the gray areas. While this is probably imposed to promote stability across years (to avoid frequently changing the filter), our study gives concrete evidence that basing the X11 seasonal filter on the value of θ_{12} can often improve seasonal adjustment.

Table 8. Selection of X11 Seasonal MA Based on Global "I/S" Ratio

Ratio	MA
<2.5	3x3
(2.5, 3.5)	gray area
(3.5, 5.5)	3x5
(5.5, 6.5)	gray area
>6.5	3x9

5. Comparison with two Census studies

Hood & Findley (1999)

Hood and Findley evaluate seasonal adjustment from X11 (using X-12-ARIMA) and SEATS (using TRAMO/SEATS) based on simulated series. The 54 series are formed from combinations of estimated components from X11 and SEATS adjustments of real series. Their creative method in some sense avoids bias in favor of either method, but may or may not yield realistic series. They evaluate the methods according to agreement between true and estimated seasonally adjusted series. A principal measure of agreement is relative root mean square deviation (RRMSQD) from true values. Overall, they find that seaonal adjustments tend to be quite similar and agreement statistics close. They make the following points:

(1) SEATS' adjusted series come closer to the true values for series with large irregulars.
Using 12-year spans and based on the RRMSQD statistic, each method is closer for 25 series and there is a virtual tie for the remaining 4. One-third (18 of 54) are designed to have relatively large irregulars. SEATS' values are closer than X11's for 16 of these series.

(2) SEATS induces seasonality into some series.
The spectrum of the seasonally adjusted series has seasonal peaks for 5 series where there are not peaks in the original series.

(3) X11 tends to be closer for very short series.
Based on adjustment with 4-year spans, X11 is closer for 39 series, SEATS for 11, and the methods tie for 4.

Point (1) seems consistent with BLS results. X11 may be choosing the 3x5 filter, allowing some of the large irregular to enter the seasonal component. Point (2) does not occur for the BLS series, but certainly suggests paying attention to diagnostics. A positive aspect of X-12 software is its automatically warning of residual seasonal peaks. With respect to Point (3), such short series should in most cases be adjusted with a deterministic seasonal component, where the differences should be small, so this finding does not seem compelling.

Hood (1992)

Hood considers 260 import/export series from Census' Foreign Trade program. As for the BLS study, X12/SEATS is used for both X11 and SEATS adjustments. Revision statistics are similar for the most part: average absolute revision in the seasonally adjusted series from the methods differs by more than 0.5 for only 18 series and relative mean absolute deviation differs by more than 0.05 for only 25 series. Hood's main findings are that (1) series with larger X11 revisions mostly occur when SEATS' θ_{12} parameter is large in magnitude (near -1) and X11's I/S ratio is less than 5 and (2) series with larger SEATS revisions mostly occur when θ_{12} is moderate (-.6 to -.4) and the I/S ratio is large. That is, smaller revisions are associated with longer filters. No preference between methods is made, since longer filters tend to converge to final values more slowly, not fully captured by the revision statistics.

As with the earlier Census study, finding (1) is consistent with BLS' finding. The BLS study goes further by preferring SEATS in such cases, based on more satisfactory monthplots and (even with the more stable seasonal components) a lack of residual seasonality. Furthermore, the BLS revision statistics are more nearly "final," since there are at least 24 months beyond each month in its test period, while 40% of Census' test months have only 1-23 months beyond.

Hood reports that SEATS' model selection can cause unstable adjustments. A (100) (100) model is selected for a grain exports series. For some spans, SEATS decomposes the series with this model and for others it switches to an airline model. Large revisions result. Revisions are substantially reduced with an "analyst" adjustment specifying an airline model.

In the BLS study, when SEATS switches models, the results are usually acceptable. However, experience with other series tells us that automatic model identification can sometimes select models which may be all right for forecasting but are not suitable for decomposition. Consistent with Hood, we feel that analyst review of model selection is needed.

6. Main findings and recommendations

Both theoretical and empirical work point to shortcomings in X11's automatic selection of seasonal moving averages. Depoutot and Planas provide guidelines for choosing both trend and seasonal MA's for X11 through comparison with ARIMA model parameters. The present BLS study and the two Census studies all find that SEATS tends to have smaller revisions than X11 when the ARIMA seasonal parameter is large in magnitude. In the BLS study, there are definite benefits from the SEATS adjustments. For a number of series, monthplots exhibit a more acceptable level of variability in the seasonal component and the seasonally adjusted spectrum shows less departure from the original spectrum, except for eliminating the seasonal peaks.

Given this body of evidence, we recommend that Census consider basing its automatic seasonal MA selection on the seasonal parameter in the ARIMA model for the series, when available. As mentioned earlier, in part X11's selection is biased toward the 3x5 seasonal MA due to the gray areas in its selection procedure. Giving that up could sacrifice stability in adjustments across time. This points to an intrinsic advantage with SEATS adjustment: a modest change in the seasonal parameter can be expected to cause a modest change in the overall seasonal adjustment filter length, while any time X11 switches seasonal MA's, the effect on overall filter length will be appreciable.

Overall, all three empirical studies find results of similar quality for a majority of series. X11 continues to apply somewhat more broadly than SEATS. Sometimes it can be difficult to find a good-fitting, decomposable model for a series, which reduces confidence in SEATS adjustments. Acceptable X11 adjustments may often be found more easily for such series, given greater experience with its diagnostics and its options. Furthermore, Aston et al. (2007) suggest that limitations of the (011) seasonal part of ARIMA models may lead to seasonal components which are too stable.

Based on its findings, the BLS team carrying out the study has recommended that BLS programs consider using model-based adjustment with SEATS (1) as a supplement to assist in carrying out X11 adjustment and (2) in place of X11 adjustment where it offers improvement. Further study of SEATS diagnostics is also recommended for effective implementation of these uses of SEATS.

The views expressed are those of the authors and do not represent official positions of BLS.

Acknowledgments

The study has been carried out by a BLS-wide team consisting of Nicole Brooks, Dave Byun, Daniel Chow, Tom Evans, Mike Giandrea, Chris Manning, Jeff Medlar, Randall Powers, Stuart Scott (team leader), Eric Simants, Michail Sverchkov, Richard Tiller, Daniell Toth, and Jeff Wilson.

We are especially grateful to David Findley, Census, for detailed comments on the study. We also benefitted from comments on drafts of the report from Jim Buszuwski, John Eltinge, Jurgen Kropf, Jim Thomas, and Peter Zadrozny, all from BLS. We appreciate very much support for the project from John Eltinge and other members of the BLS Innovation Board and support for staff participation from other BLS managers.

References

Aston, J. A., Findley, D. F., McElroy, T. S., Wills, K. C. and Martin, D. E. K. (2007), "New ARIMA Models for Seasonal Time Series and their Application to Seasonal Adjustment and Forecasting." U. S. Census Bureau SRD Research Report

Cleveland, W. P. and Tiao, G. C. (1976), "Decomposition of Seasonal Time Series: A Model for the Census X11 Program," *Journal of the American Statistical Association* 71, 581-587

Dagum, E. B. and A. Luati (2002), "Smoothing Seasonally Adjusted Time Series," *ASA Proceedings of the Joint Statistical Meetings*, 665-670

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

Findley, D., T. S. McElroy, and K. C. Wills (2004), "Modifications of SEATS' Diagnostic for Detecting Over- or Underestimation of Seasonal Adjustment Decomposition Components," *ASA Proceedings of the Joint Statistical Meetings*

Findley, David F., Monsell, Brian C., Shulman, Holly B., and Pugh, M. G. (1990), "Sliding-Spans Diagnostics for Seasonal and Related Adjustments," *Journal of the American Statistical Association* 85, 345-355

Hood, Catherine C. and Findley, David F. (1999), "An evaluation of TRAMO/SEATS and comparison with X-12-ARIMA," *ASA Proceedings of the Business and Economic Statistics Section*, 150-155

Hood, Catherine C. (2002), "Comparison of Time Series Characteristics for Seasonal Adjustments from SEATS and X-12-ARIMA," *ASA Proceedings of the Joint Statistical Meetings*

Ladiray, Dominique and Benoit Quenneville (2001), *Seasonal Adjustment with the X-11 Method*, Springer, New York

Lytras, Demetra (2007), "X-12-Graph: A SAS/GRAPH® Program for X-12-ARIMA Output," U.S. Census Bureau Users Guide

Scott, Stuart, Richard Tiller, and Daniel Chow (2007), "Empirical Evaluation of X11 and Model-based Methods for Three BLS Programs," internal report to BLS Innovation Board

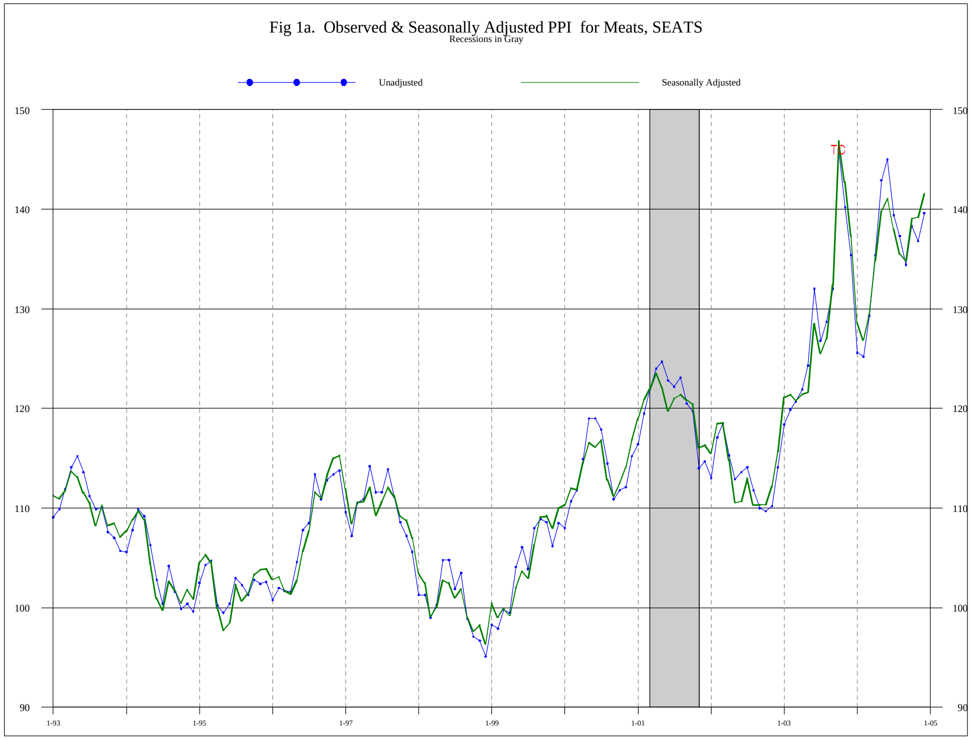

Fig 1a. Observed & Seasonally Adjusted PPI for Meats, SEATS
Recessions in Gray
Unadjusted
Seasonally Adjusted
TC

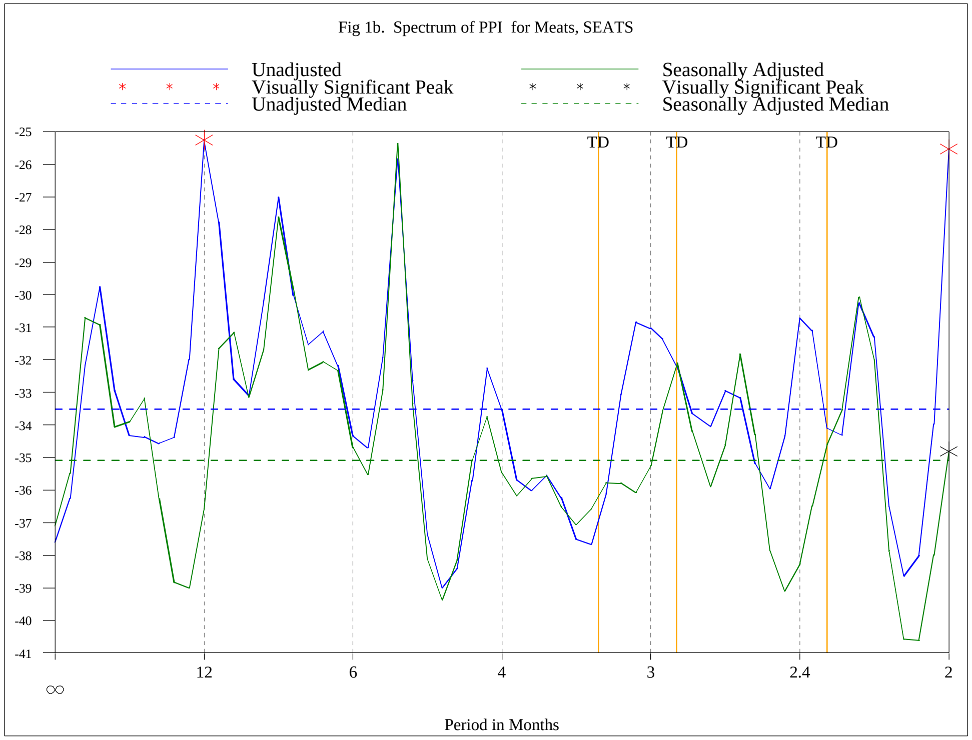

Fig 1b. Spectrum of PPI for Meats, SEATS
Unadjusted
Visually Significant Peak
Unadjusted Median
Seasonally Adjusted
Visually Significant Peak
Seasonally Adjusted Median
TD
Period in Months

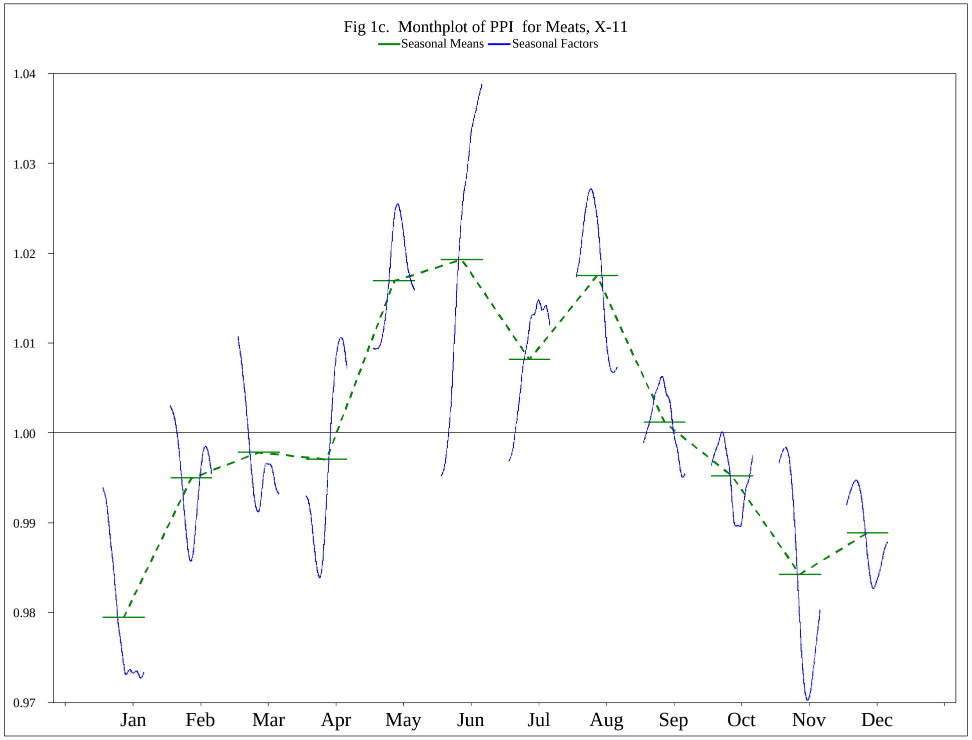

Fig 1c. Monthplot of PPI for Meats, X-11
Seasonal Means
Seasonal Factors
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

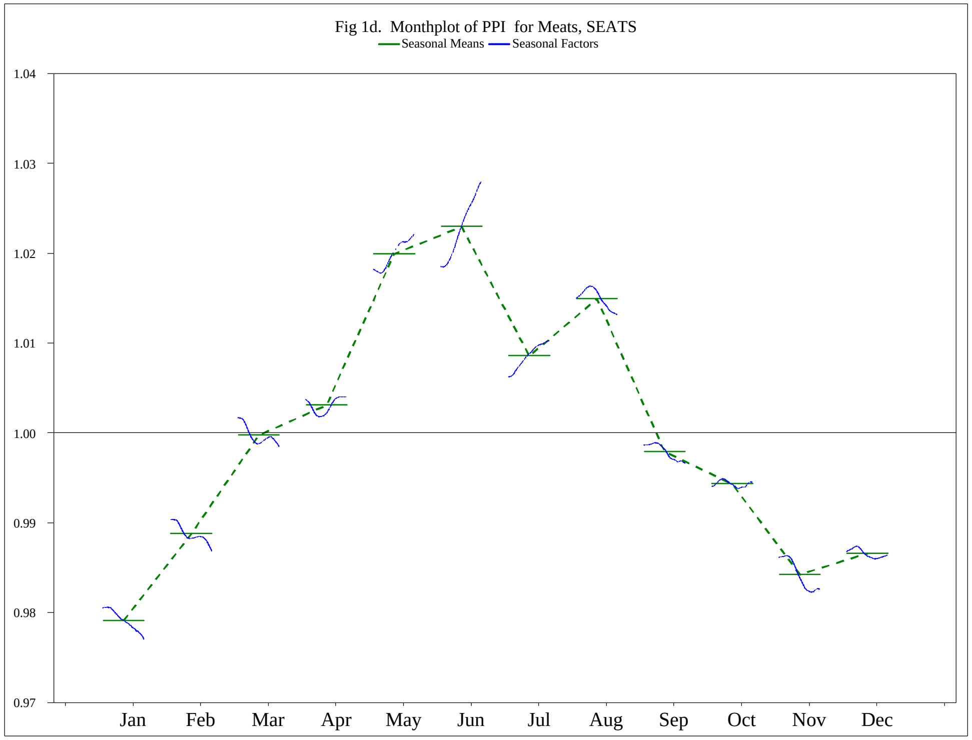

Fig 1d. Monthplot of PPI for Meats, SEATS
Seasonal Means
Seasonal Factors
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

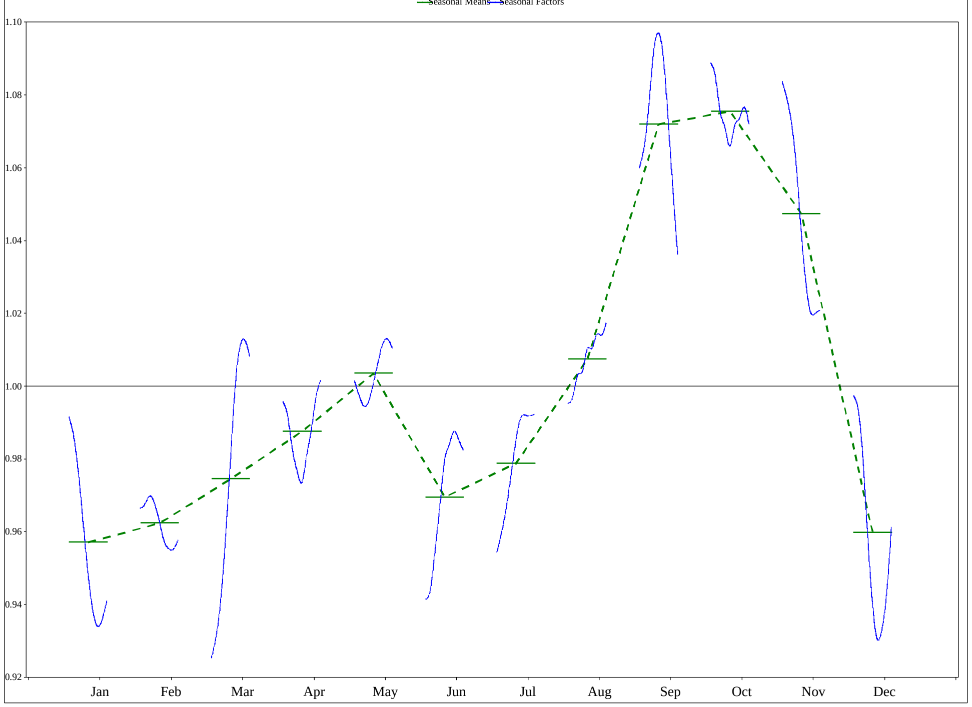

Fig 2a. Monthplot of PPI for #2 Diesel Fuel, X-11 AUTO

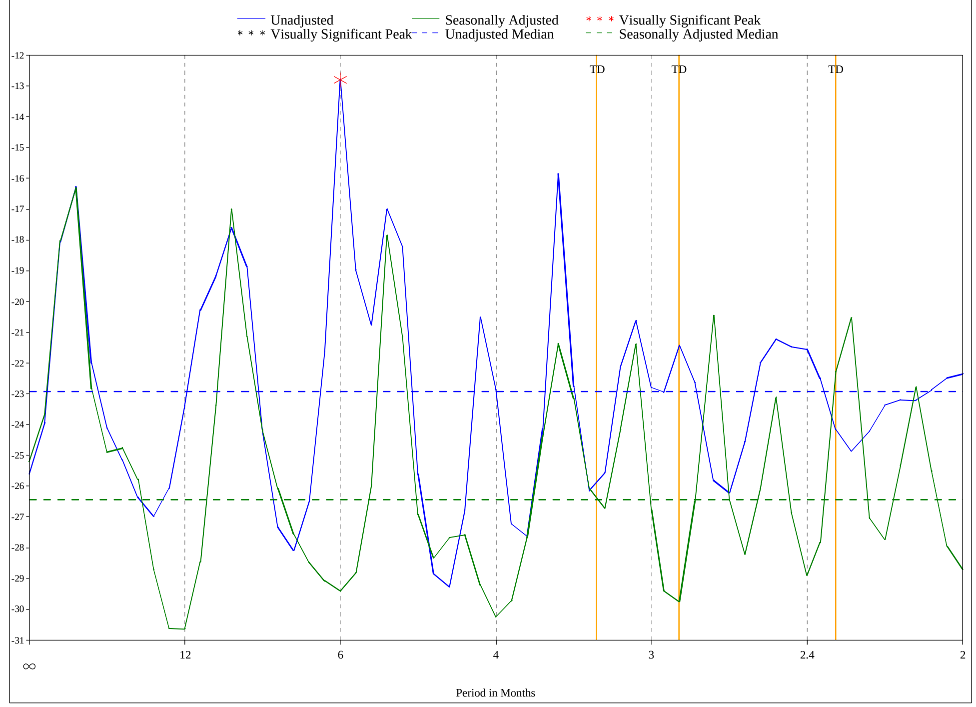

Fig 2b. Spectrum of PPI for #2 Diesel Fuel, X-11 AUTO

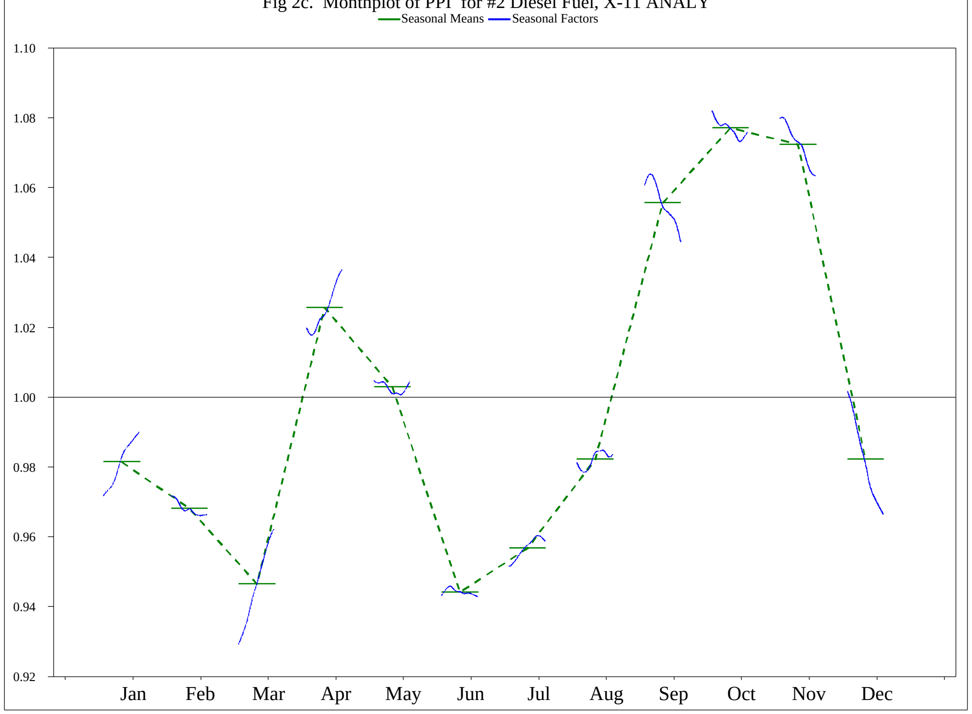

Fig 2c. Monthplot of PPI for #2 Diesel Fuel, X-11 ANALY

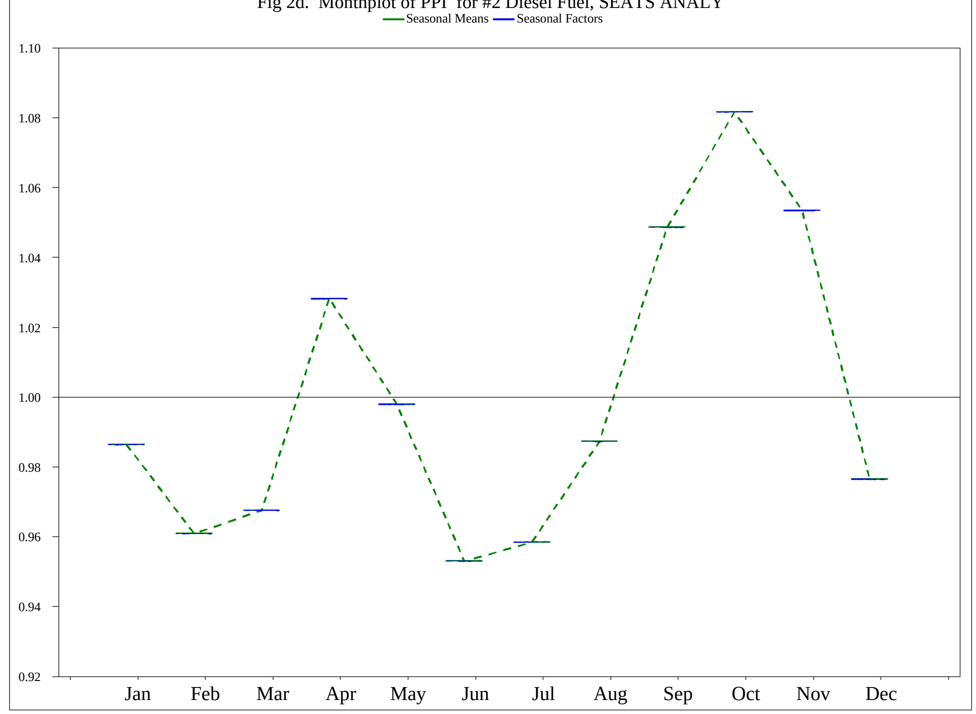

Fig 2d. Monthplot of PPI for #2 Diesel Fuel, SEATS ANALY

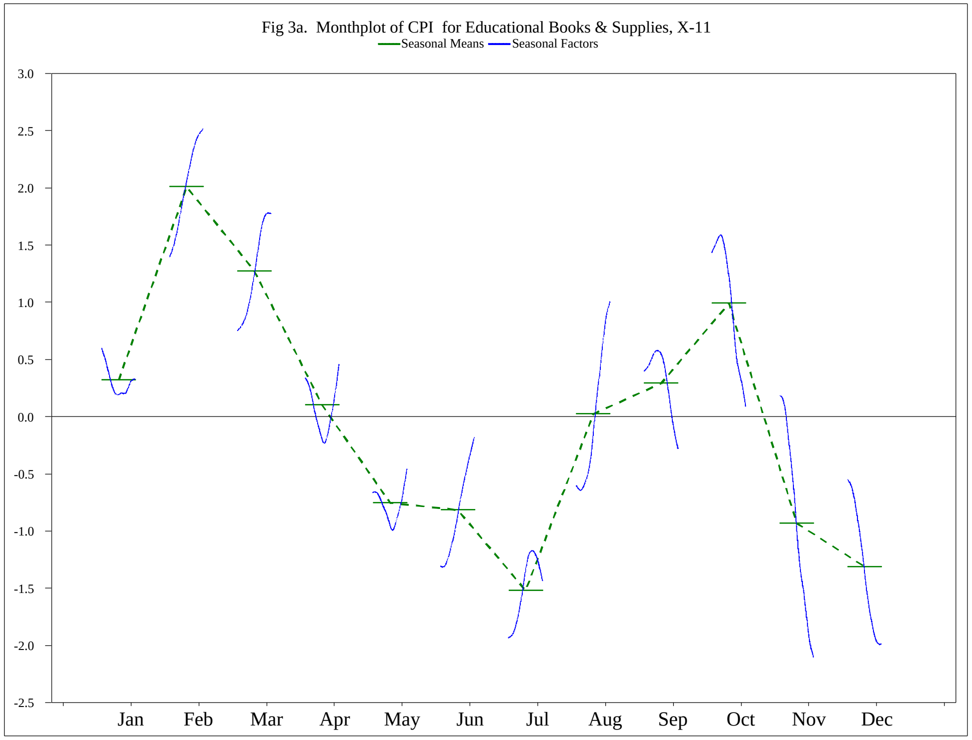

Fig 3a. Monthplot of CPI for Educational Books & Supplies, X-11
Seasonal Means
Seasonal Factors
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

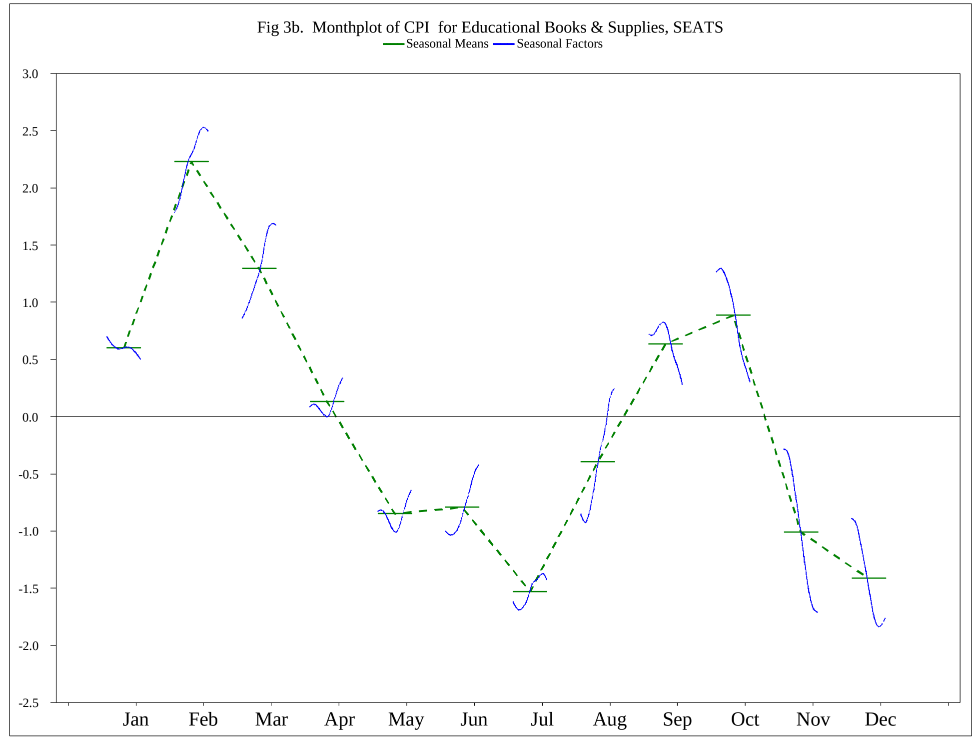

Fig 3b. Monthplot of CPI for Educational Books & Supplies, SEATS
Seasonal Means
Seasonal Factors
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

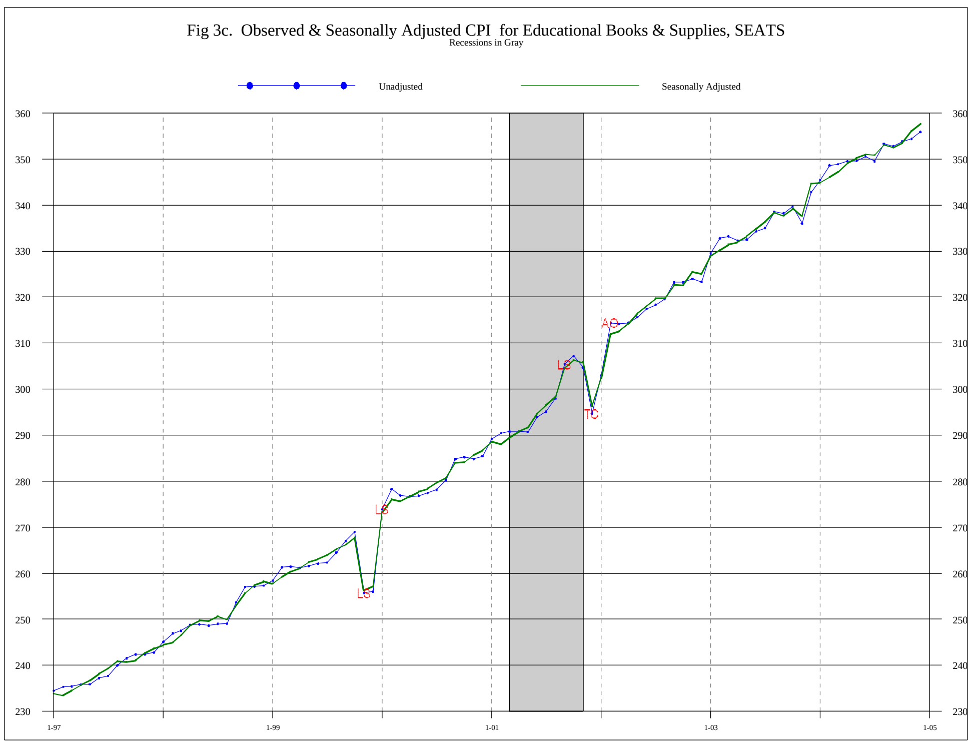

Fig 3c. Observed & Seasonally Adjusted CPI for Educational Books & Supplies, SEATS
Recessions in Gray
Unadjusted
Seasonally Adjusted
360
350
340
330
320
310
300
290
280
270
260
250
240
230
LS
LS
LS
TC
AO
1-97
1-99
1-01
1-03
1-05

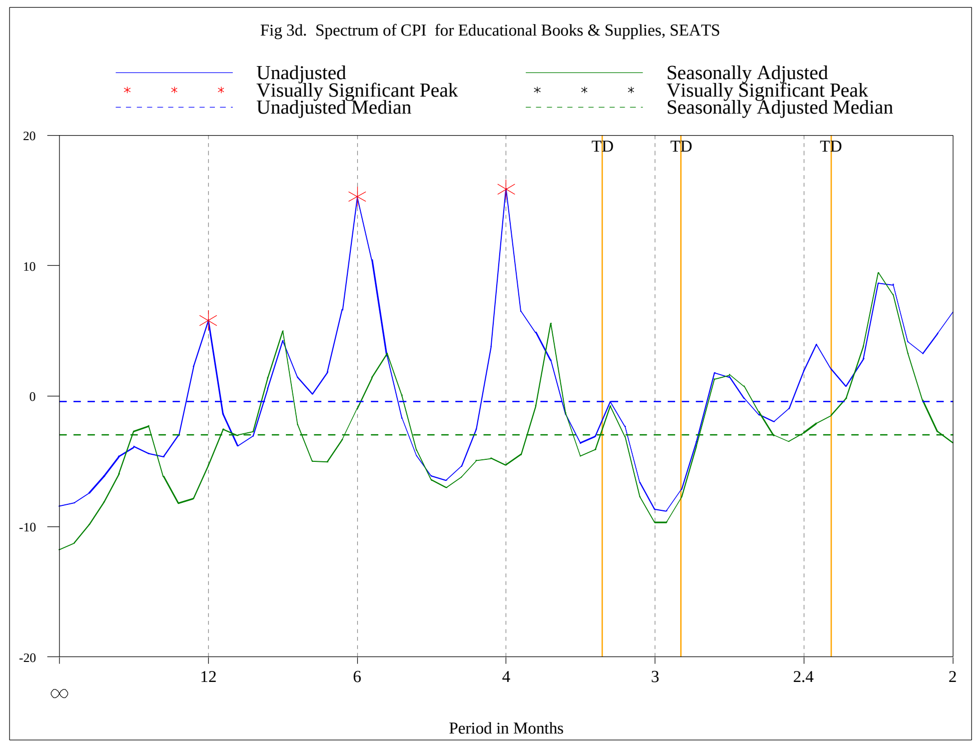

Fig 3d. Spectrum of CPI for Educational Books & Supplies, SEATS
Unadjusted
Visually Significant Peak
Unadjusted Median
Seasonally Adjusted
Visually Significant Peak
Seasonally Adjusted Median
20
10
0
-10
-20
TD
TD
TD
∞
12
6
4
3
2.4
2
Period in Months

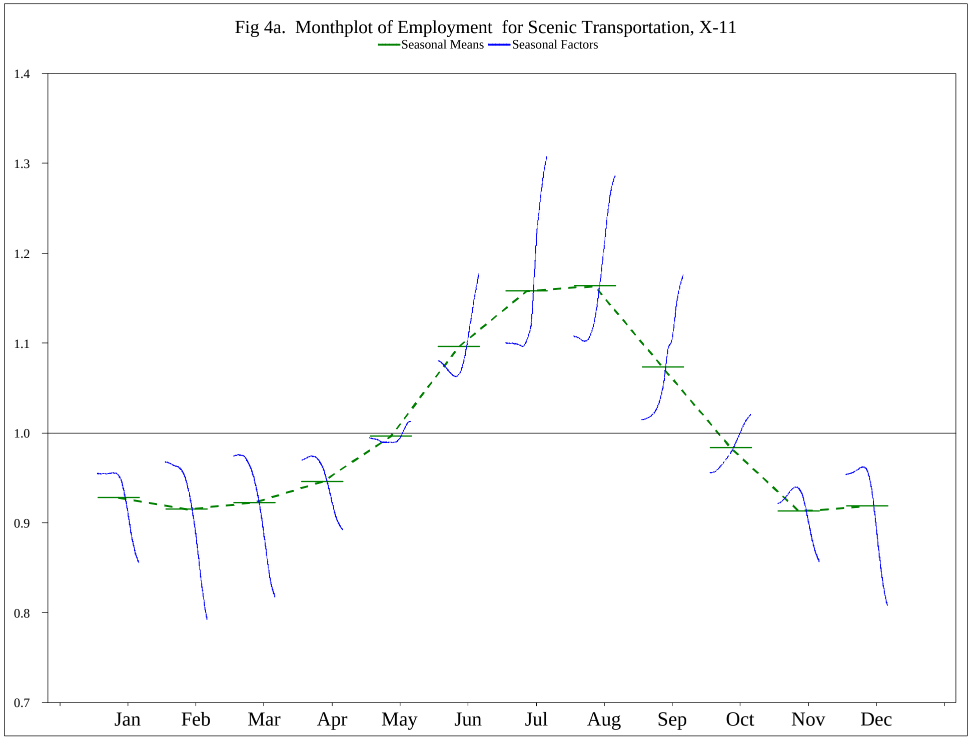
Fig 4a. Monthplot of Employment for Scenic Transportation, X-11
Seasonal Means
Seasonal Factors
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

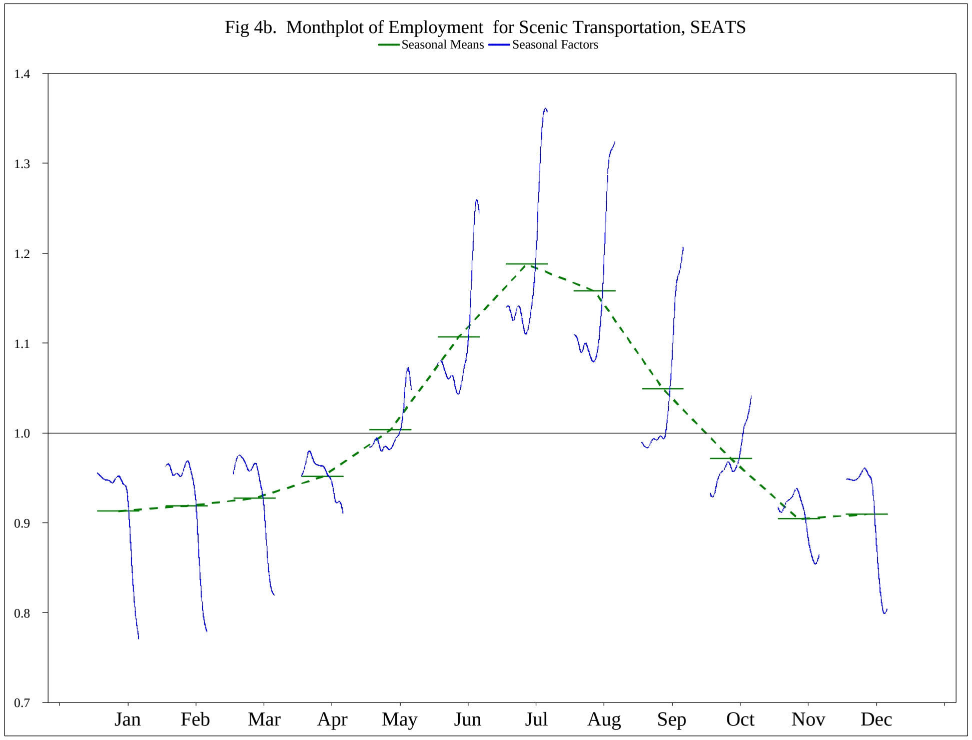
Fig 4b. Monthplot of Employment for Scenic Transportation, SEATS
Seasonal Means
Seasonal Factors
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

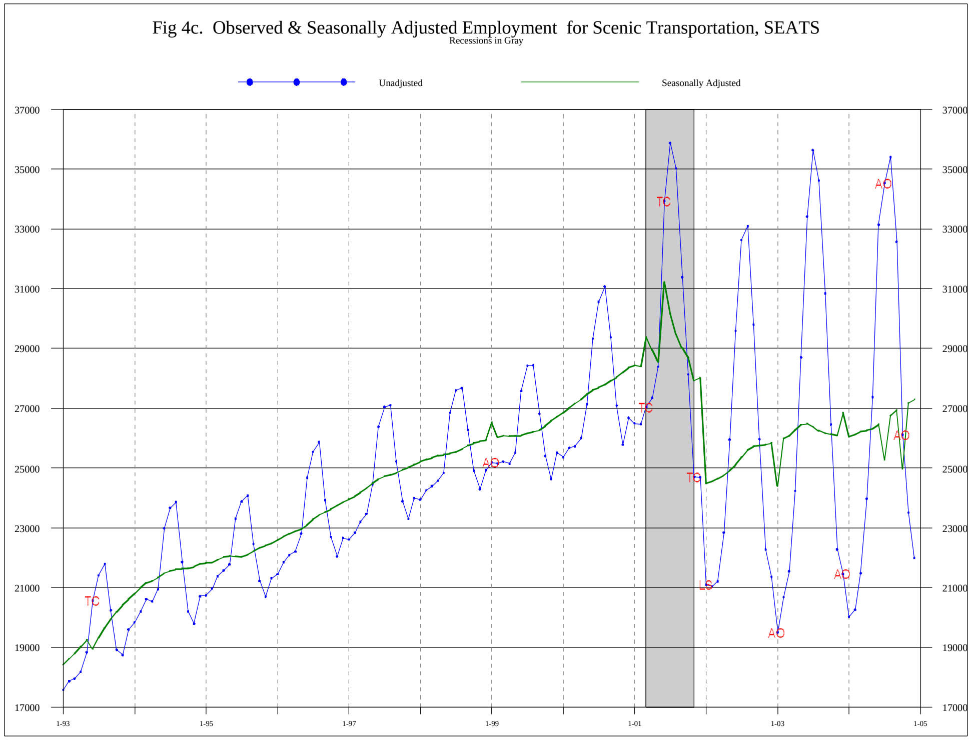
Fig 4c. Observed & Seasonally Adjusted Employment for Scenic Transportation, SEATS
Recessions in Gray
Unadjusted
Seasonally Adjusted
37000
35000
33000
31000
29000
27000
25000
23000
21000
19000
17000
TC
AO
TC
TC
TC
LS
AO
AO
AO
AO
1-93
1-95
1-97
1-99
1-01
1-03
1-05

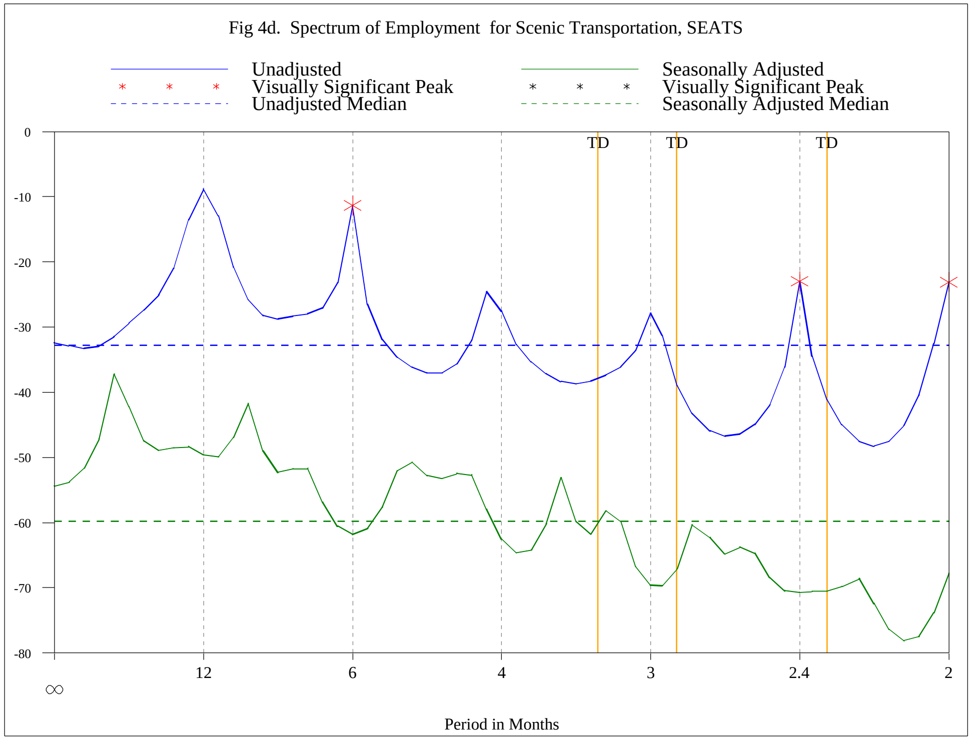
Fig 4d. Spectrum of Employment for Scenic Transportation, SEATS
Unadjusted
Visually Significant Peak
Unadjusted Median
Seasonally Adjusted
Visually Significant Peak
Seasonally Adjusted Median
TD
TD
TD
0
-10
-20
-30
-40
-50
-60
-70
-80
∞
12
6
4
3
2.4
2
Period in Months