Effects of Transformation Choice on Seasonal Adjustment Diagnostics and Forecast Errors

Golam M. Farooque
Economic Statistical Methods and Programming Division
U.S. Census Bureau, Washington, D.C. 20233

Abstract: Usually the variations in economic time series increase as the level of the series increases. Traditionally, the U.S. Census Bureau adjusts time series data for seasonal variations using logarithmic transformations for forecast extension and a multiplicative model for the forecast-extended series because the seasonal variations in most seasonal economic time series increase and decrease proportionally with increases and decreases in the level of the series. If a series, however, does not show an increasing variance, then there is no need for a log transformation; the series can be adjusted additively.

In this study we examined whether we could get better seasonal adjustment diagnostics and regARIMA model-selection diagnostics with an additive model for those series that might not need a log transformation. Using the Census Bureau's seasonal adjustment program X-12-ARIMA Version 0.3, we computed various seasonal adjustment diagnostics for 263 U.S. Import/Export series and 271 Manufacturers' Shipments, Inventories and Orders series. Our results show that out of 120 series that do not need log transformations but can be seasonally adjusted based on the seasonal adjustment diagnostic M7, 29 of them have better seasonal adjustment and modeling diagnostics with an additive adjustment.

1. Introduction

With economic time series, it is very common that the variations in the series increase as the level of the series increases. The logarithmic transformation usually adequately converts nonconstant variation into constant variation, hence, it stabilizes the variance and the series can then be modeled with the Box-Jenkins methodology (Box, Jenkins, and Riensel 1994). Log transformations also affect the forecast error substantially if the variation in the series occurs at the end of the series because most forecast methods place more weight on the most recent data. On the other hand, if a series does not show increasing variance, there is no need for log transformations to stabilize the variance. Taking logs when the seasonal pattern does not change with the level has the effect of producing a series whose seasonal variability does change with the level.

In adjusting the data for seasonal variations, a fundamental question is whether to decompose the series with additive decomposition or multiplicative decomposition, i.e., whether to use no transformation or a log transformation of the series when running the X-12-ARIMA seasonal adjustment program (Findley, Monsell, Bell, Otto, and Chen 1998). Traditionally, the U.S. Census Bureau has used a log transformation with a multiplicative model for seasonal adjustment.

The U.S. Census Bureau publishes the seasonal factors that are obtained from X-12-ARIMA. With multiplicative decomposition, the seasonal factors are centered on one, always positive, and divided into the original series. With additive decomposition, the seasonal factors are centered on zero and subtracted from the original series. For production or publication, it is advantageous to have one kind of seasonal factor: two kinds of seasonal factors could be a problem. At the Census Bureau, we usually assume that the series need log transformations and apply the multiplicative decomposition.

In this study, we examined whether we could get better seasonal adjustment diagnostics and regARIMA model-selection diagnostics by not using logs and by using an additive adjustment for those series that do not need a log transformation. We used the transformation choice to suggest the mode of seasonal adjustment, not the other way around. We used the mean square error (MSE) of out-of-sample forecasts from X-12-ARIMA to compare the forecast performance of the regARIMA models of the two transformation options. The modeling diagnostics, especially the out-of-sample forecast error diagnostic are important because X-12-ARIMA uses forecasts of the series to produce the seasonal adjustment. And if the forecasts are affected by the transformation, then the transformation does have the potential to affect the adjustment and its diagnostics.

---

2. Methodology

In Section 2.1, we describe the data and in Section 2.2, we discuss the steps used to identify the transformation choices, compute the seasonal adjustment and model-selection diagnostics for the transformation choices and compare the quality of adjustments of both transformation choices using the computed diagnostics.

2.1 Data

In this study, we ran X-12-ARIMA for 263 U.S. Import/Export series and 271 Manufacturers' Shipments, Inventories, and Orders (M3) series. The Import/Export series start in January 1989 and end in November 2001. The M3 series start in January 1992 and end in October 2001.

2.2 Steps to Identify the Transformation Choices and for Computing the Diagnostics for Transformation Choices

We applied the following steps. We used the U.S. Census Bureau's software, X-12-Write Version 1.1 to create the input files (Hood 2003).

Step 1: In this step, we identified the series that do not need a log transformation. We ran X-12-ARIMA for 534 Import/Export and M3 series using its **automatic transformation** option. When **function = auto** is specified for a series, the program chooses between no transformation and a log transformation by fitting a regARIMA model to the untransformed and transformed series and choosing the transformation based on Akaike's Information Criterion Corrected for sample size (AICC). The AICC of the log transformation needs to be bigger by 2 than the AICC of no transformation for X-12-ARIMA to choose no transformation over the log transformation (for details, see the X-12-ARIMA Reference Manual (U.S. Census Bureau 2002)).

An example of an input file with the automatic transformation option is:

```
series { ...}
transform{function = auto}
automdl {savelog=amd}
regression {aictest = (td easter) savelog = aictest}
outlier {types = all}
x11{}
...
```

Based on the AICC test, this run identified that out of 534 series, 136 series do not need log transformations.

Step 2: In this step, we obtained the regARIMA model for the 136 series that do not need log transformations. We created separate input files for each of the 136 series: one input file with the log transformation and another with no transformation. We used the automatic model selection option and typical initial regression and outlier options. We ran X-12-ARIMA using the two input files as shown below.

```
series{...}
transform{function = log}
automdl{savelog = amd }
regression{ aictest=( td easter ) savelog=aictest }
outlier{types=all}
x11{}
...
```

```
series{...}
transform{function = none}
automdl{savelog = amd}
regression{ aictest=( td easter ) savelog=aictest }
outlier{types=all}
x11{mode = add}
...
```

Step 3: In this step, we computed the seasonal adjustment diagnostics: M7, sliding spans month-to-month percent change (MM) and maximum percent differences (MPD), average absolute revisions of concurrent adjustments (ARCSA), and average absolute revisions of the month-to-month differences of the adjustments (ARMMDA), and model-selection diagnostics: Ljung-Box Qs (LBQ) for correlated residuals, residual spectral peaks, and mean squared error (MSE) of out-of-sample forecasts. For details on seasonal adjustment diagnostics, see the X-12-ARIMA Reference Manual (U.S. Census Bureau 2002), Ladiray and Quenneville (2001), and Findley, Monsell, Shulman, Pugh (1990); and for model-selection diagnostics, see Findley, et al. (1998).

We hard coded the ARIMA models and regression variables obtained in Step 2 for both transformation choices and reran X-12-ARIMA with the hard coded input files. We applied a fixed 3X5 filter in all series to compare the sliding spans diagnostics of both transformation choices. If no filter is specified, the program may choose different final filters for the two transformation choices for the same series, and if it does, we cannot easily compare the sliding spans diagnostics. Examples of the two input files:

```
series{...}
transform{function = log}
arima {model = (x x x)(y y y)}
regression{ variables = (m n)}
outlier{types=all}
x11{seasonalma = s3x5 savelog = (m7, q)}
estimate {}
check{ print = all   savelog = lbq }
forecast{ maxlead = 24  }
slidingspans{savelog=percent additivesa = percent }
history{estimates= (fcst sadj sadjchng)}
```

```
series{...}
transform{function = none}
arima{model = (w w w)(z z z)}
regression{variables = (p r)}
outlier{types=all}
x11{seasonalma = s3x5   savelog = (m7 q )}
estimate{mode = add}
check{ print = all   savelog = lbq }
forecast{ maxlead = 24  }
slidingspans{savelog=percent additivesa = percent}
history{estimates= (fcst sadj sadjchng)}
```

In this step, we also identified the nonseasonal series to exclude from the study based on the seasonal adjustment diagnostic M7 and the spectral graphs of the original series. We found that 120 series out of 136 are seasonal.

Step 4: In this step, we compared the quality of the seasonal adjustment of the run with no transformation to the run with the log transformation using diagnostics from Step 3: MM, MPD, ARCSA, ARMMDA, LBQ, residual spectral peaks, and MSE of out-of-sample forecasts.

3. Results

We found that out of 534 series, 136 do not need log transformations. Out of the 136 series, 120 were seasonal. Table 1 provides the number of series, out of the 120, that have better MM, MPD, ARCSA or ARMMDA for the run with no transformation than for the run with the log transformation. Note that more series have better MM, MPD, ARCSA, and ARMMDA with no transformation than with the log transformation. In other words, the seasonal adjustment diagnostics for the run with no transformation are better for a larger number of series than for the run with a log transformation of the same series. For example, in Table 1, 55 series have better sliding spans diagnostics with a log transformation and 65 series have better sliding spans diagnostics with no transformation. Similarly, ARCSA and ARMMDA are better for more series with the no-transformation choice than for the log transformation.

Table 1: Diagnostic preferences for the 120 seasonal series where the AICC preferred no transformation

Diagnostic	Diagnostics better for	
	log transformation	no transformation
MM/MPD	55	65
ARCSA	47	73
ARMMDA	50	70
MM/MPD & ARCSA	29	49
MM/MPD & ARMMDA	28	46
MM/MPD, ARCSA, ARMMDA	22	41

Table 1 also shows the number of series with agreement between seasonal adjustment diagnostics when we combined the results of the diagnostics. By combining diagnostics results, we see that the no-transformation choice has more series that have better seasonal adjustment diagnostics than the log transformation. The last row of this table shows that a total of 41 series have better MM (or MPD), ARCSA, and ARMMDA for the no-transformation choice compared to 22 series for the log transformation.

To understand further the effects of transformation choice on the seasonal adjustment quality, we examined the graphs of initial and final estimates of the month-to-month percent changes and initial and final estimates of the seasonal adjustment values. Figures 1, 2, 3, and 4 are graphs of month-to-month percent changes and seasonal adjustment values for the series Imports of Alcoholic Beverages (M01010). In figures 1 and 2, a vertical line shows the change between the two estimates. A diamond marks the final estimate and a circle marks the initial estimate. From these graphs we see that the differences between the initial and final estimates are larger for the log transformation than for no transformation, even close to the end of the series. In figures 3 and 4, the final estimates are graphed as a line and the initial estimates as dots. We see that toward the end of the series, the discrepancies between the final estimates and initial estimates are larger for the log transformation than for no transformation. We used the U.S. Census Bureau's software, X-12-Graph Version 1.2 to create the graphs (Hood 2001).

Figure 1

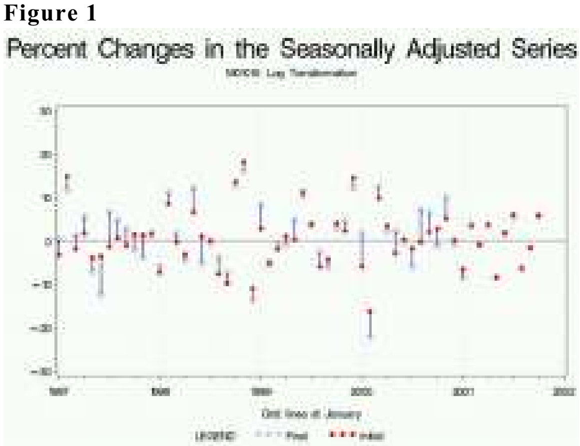

Figure 2

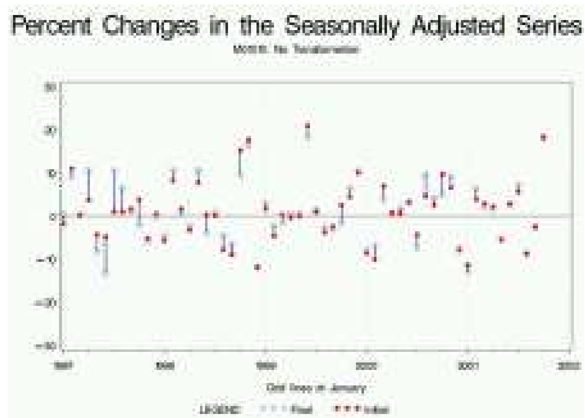

Figure 3

Figure 4

In Table 2, we compare the MSE of out-of-sample forecasts of the run with the log transformation to the run with no transformation. From this table we see that the no-transformation choice has more series with better leads 1 and 12 MSE of out-of-sample forecasts than the log transformation. For example, of the 120 series, 53 series have better leads 1 and 12 MSE of out-of-sample forecasts for the no-transformation choice compared to 31 series for the log transformation. This table also shows how many of the 120 series have smaller seasonal adjustment diagnostics (MM/MPD, ARCSA, and ARMMDA) and model- selection diagnostics (MSE of out-of-sample forecasts). From this table we see that with all diagnostics together, the no-transformation choice has more series than the log transformation where these diagnostics are smaller. We see that at a minimum, out of the 534 series, 29 series should not be log transformed, but should be adjusted additively with no transformation.

Table 2: Series with better model selection diagnostics: MSE of out-of-sample forecast errors and seasonal adjustment diagnostics: MM/MPD, ARCSA, and ARMMDA

Model Selection Diagnostics	Diagnostics better for	
	log transformation	no transformation
MSE of FCE at lead 1	42	78
MSE of FCE at lead 12	47	73
MSE of FCE at leads 1 & 12	31	53
MM/MPD, ARCSA, ARMMDA, and MSE at lead 1	12	34
MM/MPD, ARCSA, ARMMDA, and MSE at lead 12	15	35
MM/MPD, ARCSA, ARMMDA, and MSE at leads 1 & 12	8	29

Table 3 shows the model-fit diagnostics, LBQs at lags 12 and 24 and residual trading day spectral peaks and residual seasonal spectral peaks, for the 29 series that do not need log transformations. We see that there is no significant difference in trading day or residual seasonal spectral peaks between the two transformations. Similarly, there is no significant difference in LBQs at lag 12 or at lag 24 between the two transformations.

Table 3: Model fit diagnostics: LBQs and residual spectral peaks for 29 series

Diagnostic	log transformation	no transformation
Significant LBQ at lag 12	0	1
Significant LBQ at lag 24	1	2
Residual trading day spectral peak	6	4
Residual seasonal spectral peak	8	9

4. Conclusions

In this study, we examined whether we can get better seasonal adjustment and regARIMA model-selection diagnostics using the additive adjustment rather than the multiplicative adjustment. Based on seasonal adjustment diagnostics MM/MPD, ARCSA, and ARMMDA, and the model-selection diagnostic MSE of out-of-sample forecast errors, we find that there are at least 29 series that do not need log transformations. An additive adjustment of these series would produce better modeling and seasonal adjustment diagnostics.

It is important that the organizations that produce seasonal adjustments perform the appropriate transformation when adjusting the series for seasonal variations instead of using the transformation dictated by production and publication. The U.S. Census Bureau has already added a new table (E16) in X-12-ARIMA providing seasonal factors that are centered on one and can be divided into original series irrespective of the type of transformation applied to the series as long as the series contains only positive values. This new table will allow the users to change the transformation choice without making large changes to production programs already in place.

Acknowledgments

The author would like to thank David F. Findley, Catherine C. Hood, and Kathleen M. McDonald-Johnson for their valuable comments on several drafts of the paper.

References

Box, G.E.P., G.M. Jenkins, and G. Riensel (1994). *Time Series Analysis: Forecasting and Control, 3rd Edition.* Prentice-Hall: New York, NY.

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto and B.-C. Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 16: 127-176.

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

Hood, C.C. (2001), “X-12-Graph: A SAS/GRAPH program for X-12–ARIMA output, User’s Guide for X-12-Graph Interactive for PC/Windows, Version 1.2,” U.S. Census Bureau, U.S. Department of Commerce.

_______(2003), “X-12-Write: A SAS program to write X-12-ARIMA input files, Version 1.1 in SAS 8.0,” U.S. Census Bureau, U.S. Department of Commerce.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method.* Springer-Verlag: New York, NY.

U.S. Census Bureau (2002), *X-12-ARIMA Reference Manual, Final Version 0.2*, Washington, DC: U.S. Census Bureau.