The Effect of Forecast Quality on Seasonal Adjustment Revisions

Nicole Czaplicki, U.S. Census Bureau

Abstract

When data are available, the X-11 method uses symmetric moving average filters, utilizing the same amount of data before and after the point of interest. For the most relevant points at the end of the series, we have two options: use asymmetric filters or extend the series with forecasts and use "symmetric" filters where possible. At the U.S. Census Bureau, most economic series are extended with RegARIMA models that incorporate holiday and trading day effects. We apply symmetric filters (with forecasts) and asymmetric filters (without forecasts) to empirical data from series with large forecast errors to assess the revision size-effects.

1. Introduction

Forecasts extension is a key part of any seasonal adjustment procedure. With the addition of the popular ARIMA models, X-11-ARIMA (Dagum, 1975) applied the X-11 method to time series extended with forecasts allowing points near the end of the time series to be seasonally adjusted using the same symmetric filters as those near the center of the series. Seasonally adjusted series published by the U.S. Census Bureau utilize the X-11 method currently deployed in X-13ARIMA-SEATS.

There has been much work done on the effect of forecasts on seasonal adjustment revisions with concurrent seasonal adjustment. Bobbitt and Otto (1990) found significantly smaller revisions between concurrent and final estimates with forecasts than without. Moreover, they showed that the differences in revisions were smaller for estimates calculated with a full set of forecasts compared to estimates calculated with only one year of forecasts.

Thus, it is clear from past research that when forecasts are reasonable, extending a series with forecasts for seasonal adjustment is preferred to the alternative method, asymmetric filters. For example, Figure 1 (see appendix) shows construction spending on pavement lighting with 36 months of forecasts. This series has a very obvious, stable seasonal pattern and the confidence bounds on the forecasts are tight. We can expect that the estimates we obtain from performing a moving average calculation with these forecasts will be reasonably close to the estimates we will obtain when the forecasts are replaced with real data.

On the other hand, some series are more difficult to forecast and thus, subject to large forecast errors. One example is the construction expenditures on rubber and plastics plants series shown in Figure 2 (see appendix). This seasonal series has large forecast errors and has very wide confidence bounds. Naturally, we have questioned whether using forecasts with such large errors can still minimize seasonal adjustment revisions. In this paper, we seek to answer this question by comparing revisions to seasonally adjusted estimates with and without forecasts for series with large forecast errors.

2. X-11 Background and History

The X-11 method decomposes time series into trend-cycle (C), seasonal (S), and irregular (I) components. The seasonally adjusted series is calculated by removing the seasonal component from the calendar adjusted original series. The decomposition can be either additive (1) or multiplicative (2). Multiplicative seasonal adjustment is the more common of the two for economic series at the Census Bureau. Only multiplicative seasonal adjustment was used for this study.

$$X_t = C_t + S_t + I_t \quad (1)$$

$$X_t = C_t \cdot S_t \cdot I_t \quad (2)$$

The X-11 method decomposes time series into these components by iterating between estimating the trend-cycle and seasonal components using a series of moving average filters. In general, these calculations take the form:

$$Y_t = \sum_{i=-F}^{F} w_i X_{t+i}$$

where $t_0 \leq t \leq T$, and the filter is length is 2F+1.

Near the center of the series, ($t_0 < t - F,\ t + F < T$) the filter is symmetric ($w_{-i} = w_i$). However, near the ends of the series we do not have all of the X_t's needed to calculate Y_t using the typical symmetric filters. Prior to X-11-ARIMA, the only option for these points was to use asymmetric filters applied to the data that we do have. X-11-ARIMA extended the series with forecasts from ARIMA models. When enough forecasts were generated, symmetric filters could be used to calculate seasonally adjusted values for all points in the time series. This approach generally resulted in smaller revisions to seasonal factors, thereby giving a more stable seasonal adjustment. Likewise, backcasts can be used to apply symmetric filters to values at the beginning of the series, though we will not consider backcasts for this paper.

2.1 Example: 3x3 Filter

To illustrate the difference between the two methods, we will look at 3x3 seasonal filter weights that would be used in one X-11 iteration.

If the series is not extended with forecasts, only values up to time *t* can be used in the calculations for points near the end of the series. Thus, for a monthly series, the estimate of the seasonal component, $\hat{S}_{t|t}$, at time *t* given all of the information up to time *t* (commonly called the concurrent estimate) is given by:

$$\hat{S}_{t|t} = \frac{5}{27} X_{t-24} + \frac{11}{27} X_{t-12} + \frac{11}{27} X_t$$

When 12 months of data are added, the $t + 12$ value is incorporated into the calculation for $\hat{S}_t$.

$$\hat{S}_{t|t+12} = \frac{3}{27} X_{t-24} + \frac{7}{27} X_{t-12} + \frac{10}{27} X_t + \frac{7}{27} X_{t+12}$$

After another 12 months of data have been added, the symmetric filter, centered at time *t*, is used.

$$\hat{S}_{t|t+24} = \frac{1}{9} X_{t-24} + \frac{2}{9} X_{t-12} + \frac{3}{9} X_t + \frac{2}{9} X_{t+12} + \frac{1}{9} X_{t+24}$$

Notice that since the weights must sum to one, the weights given to the series at time t, $t - 12$, and $t - 24$ change as new values are added.

When the series is extended with at least 24 forecasts, symmetric filters can be used for the concurrent estimate, treating the forecasted values as if they were true data points. Let $\widehat{X_t}$ denote the forecasted values.

$$\hat{S}_{t|t} = \frac{1}{9} X_{t-24} + \frac{2}{9} X_{t-12} + \frac{3}{9} X_t + \frac{2}{9} \hat{X}_{t+12|t} + \frac{1}{9} \hat{X}_{t+24|t}$$

Note that both the weights applied and the data used in the calculation for the concurrent estimate, $\hat{S}_{t|t}$, differ between the case with forecast extension and the case without. As a result, the concurrent estimates could be quite different between the two treatments.

2.2 Forecasting in Practice

Further releases in the X-11 family of programs incorporated additional regression variables such as trading day, Easter (Bell and Hillmer, 1983), and outliers (Bell, 1984), which can improve forecast quality.

Dagum (1978) gave some guidelines as to forecast quality, stating that ARIMA models used with the X-11 method should fit the data well and produce "reasonable" forecasts. She defines "reasonable" as forecasts with a mean absolute error less than 5% for "well behaved" series and less than 10% for highly irregular series.

At the Census Bureau, forecast extension has been the rule, regardless of the size of forecast errors. Bobbitt and Otto (1990) found that revisions were minimized by extending the series to the full forecast horizon (F months), but that extending the series with even a single year of forecasts could significantly reduce the magnitude of revisions. Bobbitt and Otto also noted that extending a series with forecasts from the airline model (rather than the selected model) still offered improvements in revisions over not forecasting at all. Moreover, they found no significant difference in revisions between series extended with forecasts from the airline model and those using forecasts from a carefully selected model, suggesting that model selection may not be of great importance when it comes to reducing revisions.

Since forecast extension has been shown to reduce revisions in seasonally adjusted series, X-13ARIMA-SEATS, the most recent release in the X-11 family of programs, extends the series with a year of forecasts by default (as did previous releases of the program). Therefore, to obtain an adjustment that does not utilize forecast extension, zero forecasts must be explicitly specified.

3. Forecast Quality and Revisions

Although Dagum (1978) offered guidelines for maximum forecast errors that should be permitted when extending a series with forecasts for seasonal adjustment, these are rarely checked in practice. Forecast errors are frequently used in comparing competing models during model selection but have not been used to determine how to treat data at the end of the series for seasonal adjustment.

In this study, we investigated the effect of forecast quality on revisions to seasonally adjusted estimates. Revisions to seasonally adjusted estimates come from different sources depending on how we treat the points near the end of the series. Here we will consider the total revision, R_t, or the absolute difference between the concurrent estimate, $\widehat{Y}_{t|t}$, and the final estimate, $\hat{Y}_{t|t+F}$. This is somewhat of a simplification of the final estimate. The value of the weights beyond F months are very small but still nonzero. Therefore, $\hat{Y}_t$ is revised beyond F months, but by a very small amount. The concurrent and final estimates without forecast extension are given by (3) and (4), respectively. The corresponding calculations with forecast extension are provided in (7) and (8), respectively. Symmetric weights are denoted by w_i and the asymmetric weights used for the concurrent estimate are denoted $\widetilde{w}_i$.

$$\hat{Y}_{t|t} = \sum_{i=-F}^{0} \widetilde{w}_i X_{t+i} \quad (3)$$

$$\hat{Y}_{t|t+F} = \sum_{i=-F}^{F} w_i X_{t+i} \quad (4)$$

$$R_t = \left|\hat{Y}_{t|t+F} - \hat{Y}_{t|t}\right|$$

$$= \left|\sum_{i=-F}^{F} w_i X_{t+i} - \sum_{i=-F}^{0} \widetilde{w}_i X_{t+i}\right| \quad (5)$$

$$= \left|\sum_{i=1}^{F} w_i X_{t+i} + \sum_{i=-F}^{0} (w_i - \widetilde{w}_i) X_{t+i}\right| \quad (6)$$

If the series is not extended with forecasts, revisions come from changes to filter weights on existing data points and the addition of new data points into the moving average (6).

If the series is extended with enough forecasts so that symmetric filters can be used, the filters' weights are the same for both the concurrent and final estimates. Therefore, revisions come only from the forecast errors (10).

$$\hat{Y}_{t|t} = \sum_{i=-F}^{0} w_i X_{t+i} + \sum_{i=1}^{F} w_i \hat{X}_{t+i|t} \quad (7)$$

$$\hat{Y}_{t|t+F} = \sum_{i=-F}^{F} w_i X_{t+i} \quad (8)$$

$$R_t = \left|\hat{Y}_{t|t+F} - \hat{Y}_{t|t}\right|$$

$$= \left| \sum_{i=-F}^{F} w_i X_{t+i} - \sum_{i=-F}^{0} w_i X_{t+i} - \sum_{i=1}^{F} w_i \hat{X}_{t+i|t} \right| \quad (9)$$

$$= \left|\sum_{i=1}^{F} w_i \left(X_{t+i} - \hat{X}_{t+i|t}\right)\right| \quad (10)$$

4. Data and Methods

This study follows the outline of the analysis done by Bobbitt and Otto (1990) and utilizes many of the same methods.

We focus on the following two research questions: 1) When forecast errors are large, are revisions to seasonally adjusted estimates smaller without forecasts than with forecasts? 2) If revisions are smaller with forecasts, does the number of forecasts used affect the size of revisions?

We selected monthly series from the Value of Construction Put in Place Survey (www.census.gov/construction/c30/methodology.html) over the span of 1993 to 2009. This survey provides monthly estimates of the total dollar value of construction work done in the United States. The estimates are subject to error, including sampling error, nonsampling error, and other measurement errors for the indirect estimates. Information about the estimation methods is available on the survey website. The selected study period was January 2000 to December 2004 in an attempt to minimize the impact of the Great Recession. Since the goal was to focus on series with large forecast errors, we selected series with average absolute percent error for one-year ahead out of sample forecasts of 20% or more, twice the limit recommended by Dagum for highly irregular series.

Current production models and filters were used and remained fixed throughout the study. Automatic identification of additive outliers, level shifts, and temporary change outliers was permitted throughout the entire span of the series. Model parameters were re-estimated with each run of X-13ARIMA-SEATS.

Three different forecast treatments were compared for each series, zero forecasts, one year of forecasts, and a full set of forecasts, dependent on the seasonal filter length. These three treatments are denoted as 0, 12, and *F*, respectively. We define a full set of forecasts, *F*, as the number of forecasts needed for the last point in the time series to be adjusted using symmetric seasonal filters applied to observed and forecasted data in a single X-11 iteration. This translates to 24 forecasts for a 3x3 filter, 36 for a 3x5 filter, and 60 for a 3x9 filter.

Data users generally prefer the difference between the concurrent and final estimates to be as small as possible. Due to the iterative nature of the X-11 method, seasonally adjusted estimates do not achieve their final value for many years after the concurrent estimate, if they achieve a final value at all. However, revisions are generally trivial after a value is adjusted with symmetric filters using only real data, in other words after another *F* months of data have been added. For this study, the final seasonally adjusted estimate, $\hat{Y}_{t|t+F}$, is the seasonally adjusted estimate for the value at time *t* calculated after an additional *F* months of data have been added to the end of the series. We use the final estimate for the treatment without forecast extension as the final estimate for all revisions calculations, thus comparing all concurrent estimates to the same target. Recall from (4) and (8) that the final seasonally adjusted estimates *should* be the same with or without forecast extension. The revisions for month *t* between the final estimate and the concurrent estimates under the three forecast treatments are defined as:

$$R_{t,0} = \left|Y_{t|t+F,0} - Y_{t|t,0}\right|$$

$$R_{t,12} = \left|Y_{t|t+F,0} - Y_{t|t,12}\right|$$

$$R_{t,F} = \left|Y_{t|t+F,0} - Y_{t|t,F}\right|$$

We compared mean squared revisions, mean absolute revisions, and maximum absolute revisions for levels, log levels, and month-to-month changes. We calculated an analysis of variance (ANOVA) for each of these revisions measures to check the significance of the forecast method effect; in other words, to test whether revisions differed by forecast method. Since the data are heteroskedastic, the ANOVA was conducted on the ranks of the revisions measures rather than the revisions measures themselves. The magnitude of revisions varied greatly by series so we used the series mean as a blocking factor. Thus, the formula for the ANOVA is given by:

$$rank(R_{i,j}) = \mu + s_i + f_j + e_{i,j}$$

where $rank(R_{i,j})$ is the rank of revisions measure R for series i and forecast treatment j, μ is the overall mean, s_i is the series mean for series i, and f_j is the method effect for forecast treatment *j*. We also examined pairwise comparisons to test for differences in revisions between each pair of forecast treatments. Where the revisions measures were normally distributed, we compared pairwise with a paired t test. Where they were not normally distributed, we used a Wilcoxon signed rank test ("Wilcoxon signed-rank test", 2015)

5. Results

The results of the ANOVAs, paired t tests, and Wilcoxon signed rank tests are presented in Tables 1 to 3 for mean squared revision, mean absolute revision, and maximum absolute revision, respectively. In each table, the ANOVA column gives the F statistic of the forecast method effect from the ANOVA and columns 0-F, 0-12, and 12-F contain the t statistic from the paired t test or W statistic from the Wilcoxon signed rank test. Each statistic is accompanied by its corresponding p value.

In each of our ANOVAs, the forecast method effect is significant. As in previous studies, the magnitude of revisions differs across forecast treatments.

It is with the paired comparisons that we truly get at the two research questions. For the first, we looked for evidence of smaller revisions for estimates calculated without forecast extension than with either of the two forecast treatments. In particular, we looked for negative t statistics from the paired t test and negative W statistics from the Wilcoxon signed rank test for the 0-F and 0-12 comparisons. In fact, we found no such evidence. Every statistic for the 0-F and 0-12 comparisons was positive, with all but one (maximum absolute revision of the level) significant at the 0.05 level. Thus, even with large forecast errors, smaller revisions are obtained with forecast extension than without.

Another interesting result lies in the paired comparisons between revisions with one year of forecasts and with a full set of forecasts. There was a significant difference in revisions for only two of nine measures (maximum absolute revisions for log-levels and month-to-month changes). This suggests that with large forecast errors, there is little if any improvement in revisions achieved by forecasting a full *F* months rather than only 12 months. This differs from the result from Bobbitt and Otto who found a significant reduction in revisions using a full set of forecasts compared to a single year of forecasts.

Additionally, we looked at the correlations between forecast errors and the log revisions measures (Table 4). Intriguingly, the correlations between revisions measures and forecast errors sharply decline from the lead-1 error to the lead-12 error. This is somewhat puzzling because the value one year away receives a much higher weight in the moving average calculation than the value one month away. Thus, we would expect the lead-12 error to be more highly correlated with revisions than the lead-1 error. Figure 3 (see appendix) shows the weights associated with a 3x5 seasonal filter and a 13-term Henderson filter; the pattern is similar for other filter combinations. It is also interesting to note that the correlations for revisions measures without forecast extension are almost as high as, or at times even higher than the corresponding correlations for revisions measures with forecast extension.

Table 1. Mean Squared Revisions

Revision Measure	ANOVA (p)	0-F (p)	0-12 (p)	12-F (p)
*Level	12.54 (0.000)	578.5 (0.000)	754.5 (0.000)	106.5 (0.491)
Log-Level	21.73 (0.000)	4.11 (0.000)	5.14 (0.000)	-0.72 (0.474)
*Month-to-Month	23.49 (0.000)	543.5 (0.000)	761.5 (0.000)	109.5 (0.479)

Table 2. Mean Absolute Revisions

Revision Measure	ANOVA (p)	0-F (p)	0-12 (p)	12-F (p)
*Level	14.70 (0.000)	604.5 (0.000)	751.5 (0.000)	135.5 (0.380)
Log-Level	18.90 (0.000)	4.67 (0.000)	6.74 (0.000)	-0.34 (0.735)
Month-to-Month	15.95 (0.000)	2.53 (0.014)	5.23 (0.000)	-1.16 (0.251)

Table 3. Maximum Absolute Revisions

Revision Measure	ANOVA (p)	0-F (p)	0-12 (p)	12-F (p)
*Level	6.21 (0.003)	271.5 (0.076)	440.5 (0.003)	-23.5 (0.879)
Log-Level	10.53 (0.000)	4.10 (0.000)	2.99 (0.004)	2.74 (0.008)
*Month-to-Month	10.85 (0.000)	551.5 (0.000)	378.5 (0.012)	524.5 (0.000)

Tables 1-3. The first column contains the F statistic of the forecasting method effect from the ANOVA accompanied by the corresponding p value. The last three columns contain the t statistic of the paired t test with the corresponding p value when the data are normally distributed. For data that are not normally distributed, denoted by *, the Wilcoxon signed rank test was used and the W statistic from this test is provided with the corresponding p value.

Table 4. Spearman Correlation Coefficients Between Revisions Measures and Forecast Errors

	Forecast Error		
Revision Measure	Lead 1	Lead 12	Lead 24
Mean Squared Log Revision			
No Forecast	.811	.464	.200
12 Forecasts	.859	.521	.251
F Forecasts	.824	.469	.226
Mean Absolute Log Revision			
No Forecast	.789	.449	.204
12 Forecasts	.854	.520	.255
F Forecasts	.827	.486	.239
Maximum Absolute Log Revision			
No Forecast	.809	.409	.148
12 Forecasts	.776	.403	.173
F Forecasts	.735	.358	.144

6. Conclusions

For this study, we compared revisions to seasonally adjusted estimates under three different forecast lengths, focusing on series with large forecast errors. In summary, we did not find any results to suggest that revisions could be improved by not utilizing forecasts, even when those forecasts are subject to large errors. Nearly all of the paired t tests and Wilcoxon signed ranks tests indicated revisions were significantly smaller when forecasts were used. However, extending series with a single year of forecasts may be sufficient to reduce revisions as there seems to be little improvement when further forecasts are added. Furthermore, these results support the default setting in X-13ARIMA-SEATS of using one year of forecasts automatically and the current Census Bureau practice of using forecasts for all seasonal adjustments. However, forecast errors should not be discarded from the seasonal adjustment dialog. Forecast errors remain an effective tool for model comparisons and give insight into the predictability, or lack thereof, of a given series.

References

Bell, W. R. (1984), "A Computer Program (TEST) for Detecting Outliers in Time Series," Bureau of the Census, Washington, D.C.

Bell, W.R. and Hillmer, S.C. (1983), "Modeling Time Series with Calendar Variation," *Journal of the American Statistical Association*, 78, 526.

Bobbitt, L. and Otto, M. C. (1990), "Effects of Forecasts on the Revisions of Seasonally Adjusted Values Using the X-11 Seasonal Adjustment Procedure," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 449-453 www.census.gov/srd/papers/pdf/rr90-09.pdf

Dagum, E.B. (1975), "Seasonal Factor Forecasts from ARIMA Models," Proceedings of the International Statistical Institute, 40th Session, 3 Warsaw, 206-219.

Dagum, E. B. (1978), "Modeling, Forecasting and Seasonally Adjusting Economic Time Series with the X-11 ARIMA Method," *Journal of the Royal Statistical Society, Series D (The Statistician)* Vol. 27, No. 3/4, 203-216.

Wilcoxon signed-rank test. (2015, August 11). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:10, October 5, 2015, from en.wikipedia.org/w/index.php?title=Wilcoxon_signed-rank_test&oldid=675600016

Appendix: Graphs

Figure 1: Construction expenditures on pavement lighting. Source: Value of Construction Put in Place Survey (1993-2009) http://www.census.gov/construction/c30/historical_data.html

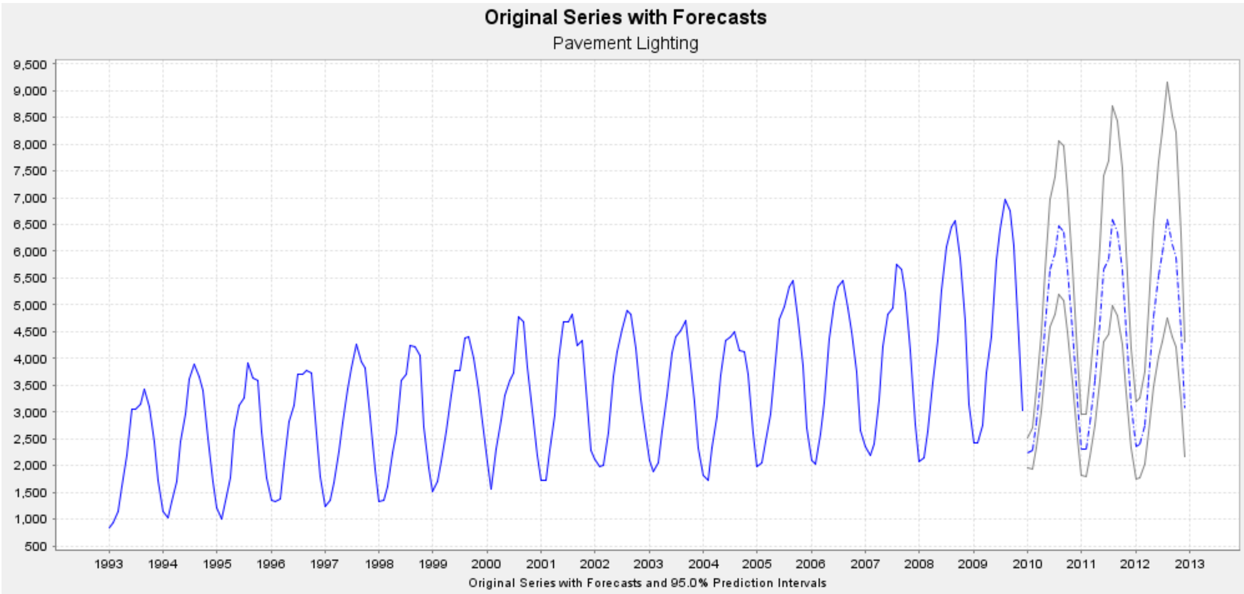

Figure 2: Construction expenditures on rubber/plastics plants. Source: Value of Construction Put in Place Survey (1993-2009) http://www.census.gov/construction/c30/historical_data.html

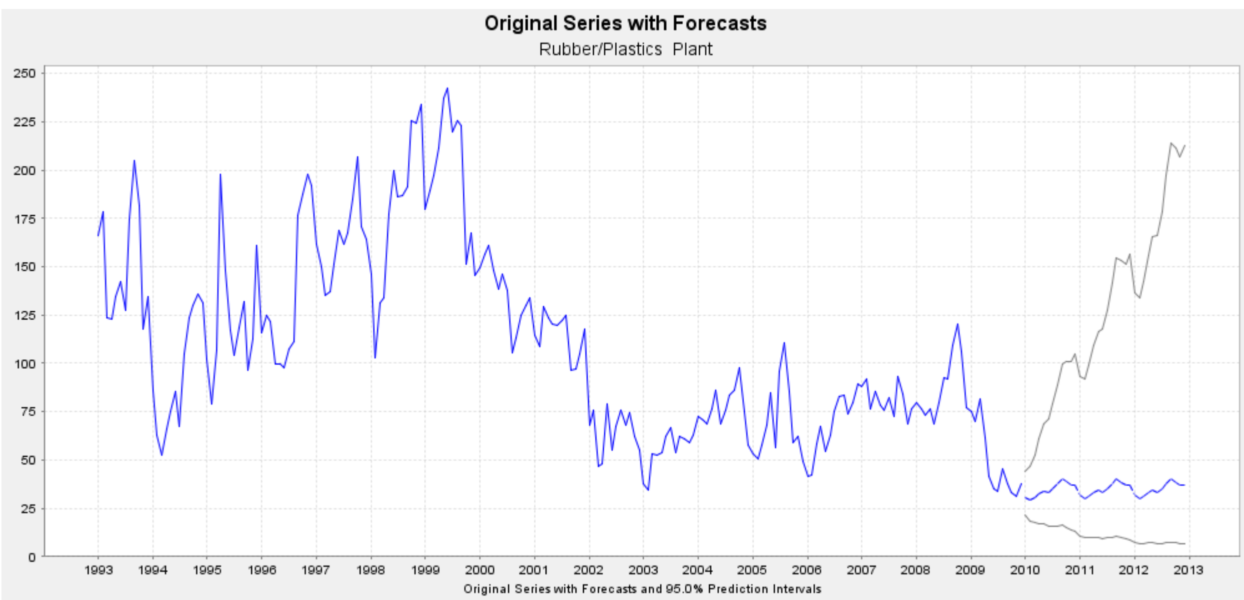

Figure 3. Seasonal Adjustment Weights with 3x5 Seasonal Filter and 13-term Henderson Trend Filter