

# BLS WORKING PAPERS



U.S. Department of Labor  
U.S. Bureau of Labor Statistics  
Office of Employment and Unemployment Statistics

## Weekly Seasonal Adjustment – A Locally-weighted Regression Approach

**William P. Cleveland**, Federal Reserve Board  
**Thomas Evans**, U.S. Bureau of Labor Statistics  
**Stuart Scott**, U.S. Bureau of Labor Statistics

Working Paper 473  
March 2014

# Weekly Seasonal Adjustment – A Locally-weighted Regression Approach

William P. Cleveland, Thomas Evans, and Stuart Scott

## 1. Introduction

The *Economist* (Feb 25, 2010) and others have spoken of a “data deluge” in these times. Statistical agencies extend their coverage to smaller and smaller geographical areas. Financial data are available at very high frequencies. In like fashion, weekly time series offer timely updates for government and private observers of the economy while they wait for the release of monthly or quarterly economic indicators. The unemployment insurance (UI) claims series produced by the U.S. Department of Labor and money supply series from the U.S. Federal Reserve are prominent examples.

Weekly series have characteristics which make seasonal adjustment inaccessible or unsatisfactory with most available software. These series are typically compiled for weeks ending on a given day of the week, Saturday in the application presented here. The number of Saturday’s within a year can be either 52 or 53 and their position varies from year to year. These features violate the basic periodic time series structure assumed by X-12-ARIMA (soon to be X-13ARIMA-SEATS), TRAMO/SEATS, and STAMP. The SABL method of Cleveland, Dunn, and Terpenning (1978) transforms weekly data to create a period 52 series and applies robust versions of the seasonal and trend smoothers of X-11. The Kalman filter methods of Gersch and Kitagawa (1983) also assume a fixed number of periods, but could be extended to add multiple regressions and their corresponding hyperparameters. Additionally, holidays require more attention; in the U.S., most switch weeks across years, which must be accounted for.

Periodic splines are used in work by Poirier (1973), Harvey and Koopman (1993), and Harvey, Koopman, and Riani (1997). The last paper develops structural models with stochastic trend, seasonal, and holiday effects, allowing for estimation using Kalman filter techniques. The seasonal component is modeled as a periodic spline with time-varying parameters. The spacing of the spline knots and values of the variance ratios have significant implications for the results, adding complexity to the method.

A locally-weighted least squares procedure is suggested here, which can be used with a weekly design matrix having 52 or 53 observations in a year. The procedure starts from the regression method in Pierce, Grupe, and Cleveland (1984), which assumes a deterministic seasonal component. The method is currently being applied at the Bureau of Labor Statistics, the Federal Reserve, and the Bank of Canada. Much of this material has appeared in a *JOS* article, Cleveland and Scott (2007).

The method is developed in the next section, followed by sections on holidays/outliers and diagnostics. Section 5 contains a detailed application. Section 6 describes a software implementation MoveReg based on a FORTRAN executable and a SAS interface. The interface makes execution more convenient and provides valuable displays, but the FORTRAN program can be run independently.

## 2. The Method

The basic model is a regression on sine and cosine terms evaluated over the days of a year. Using a sufficient number of terms of increasing frequency in cycles per year allows representation of any regular seasonal pattern. Since a given day of the month has a corresponding day of the year (leap years will be dealt with later), within month patterns can also be represented in this manner. The design matrix with columns of sine and cosine terms is augmented by columns specifying holiday and outlier effects.

While the basic seasonal effects are tied to the day of the year, weeks ending on a particular day of the week are not. For a series with weeks ending on Saturdays, the program determines on which day of the year the first Saturday falls. With this information, weekly data values can be associated with the correct days of the year with corresponding sine and cosine values. If Saturday fell on the first day of the year, there would be 53 weeks that year. Weeks ending on Saturday will be used as illustration, but the algorithm and computer program implementing it are general and can be used for weeks ending on any day of the week.

When approaching a new series, a fixed weight regression with only 12 or 24 sine and cosine terms is run. This can be done by running the software described below with simple settings or even with an external program. The fixed weight regression captures enough of the seasonality to reveal much about what outlier or holiday specifications will be needed. After adding appropriate additional terms, the locally weighted regression is carried out to estimate regression coefficients so that seasonal effects can evolve over time. In this context one can determine better how many trigonometric coefficients will ultimately be needed and proceed with further specification of outliers and holidays.

The analysis begins with a regression model for a series  $y$ , which is the observed series after suitable transformation and detrending. Series used in this paper were logged and differenced. Thus, the model for  $y$  consists of a seasonal component and error.

$$y = X\beta + e \quad (1)$$

For the seasonal component of year  $s$ , we employ trigonometric variables with fundamental frequency  $1/365$ ,

$$X_s(t, 2j-1) = \sin(2\pi ij / 365), \quad X_s(t, 2j) = \cos(2\pi ij / 365), \quad (2)$$

where  $i=i(s,t)$  is the day of the year  $s$  on which week  $t$  ends and  $j = 1, 2, \dots, p/2$ . We choose  $p$  sufficiently large to capture the dynamics of the seasonal pattern. The index  $t$  runs from 1 to  $n_s$ , which is 52 or 53. We stack the yearly matrices into an overall design matrix in  $X$  in levels. For leap years, the index  $i$  is set to  $i-1$  for days after February 29 so that these days have the same day-of-the-year index in computing the trigonometric terms for leap years as for other years. Now let  $X$  represent the first difference of this matrix. Assuming  $X$  is defined for  $K$  complete years, it has dimension  $n \times p$ , where  $n = \sum_{s=1}^K n_s$ . To achieve a weighted regression, we employ an  $n \times n$  diagonal weight matrix  $W$  and apply the standard solutions.

$$\hat{\beta} = (X'WX)^{-1} X'Wy, \quad \hat{y} = X\hat{\beta} \quad (3)$$

The regression parameter estimates  $\hat{\beta}$  minimize  $(y - X\beta)'W(y - X\beta)$ . The term of (3) requiring an inverse corresponding to a three-year series may be expanded as

$$\begin{bmatrix} X'_1 & X'_2 & X'_3 \end{bmatrix} \begin{bmatrix} w_1 I_{n_1} & & \\ & w_2 I_{n_2} & \\ & & w_3 I_{n_3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \sum_s w_s X'_s X_s \quad (4)$$

For  $\sum w_s = 1$  and identical  $X_s$  matrices (e.g. for monthly data),  $X'WX$  simplifies to  $X'_s X_s$ . Even though the  $X_s$  vary from year to year for weekly series, the  $X'_s X_s$  matrices are identical for years with 52 weeks. Years with 53 weeks are infrequent enough so we can evaluate (4) simply as  $X'X$ . Also, we have

$$X'Wy = w_1 X'_1 y_1 + w_2 X'_2 y_2 + w_3 X'_3 y_3$$

Looking back at (3), we see  $\hat{\beta}$  is a weighted sum of regression coefficients for individual years. Our estimated seasonal component for year  $s$  becomes

$$\hat{y}_s = X_s (X'X)^{-1} \sum_i w_i X'_i y_i.$$

Use of identical weights  $w_s$  would correspond to the results in Pierce, Grupe and Cleveland (1984). To allow for moving seasonality, we can apply the above method once for each year, choosing a weight matrix  $W$  geared to that particular year and using the results only for that year's seasonal component. Let  $I_p$  represent a  $p \times p$  identity matrix, and let  $w_{ij}$  be the weight applied to year  $j$  to estimate year  $i$  factors.

With three years, we may write

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} X_1(X'X)^{-1} & & \\ & X_2(X'X)^{-1} & \\ & & X_3(X'X)^{-1} \end{bmatrix} \begin{bmatrix} w_{11}I_p & w_{12}I_p & w_{13}I_p \\ w_{21}I_p & w_{22}I_p & w_{23}I_p \\ w_{31}I_p & w_{32}I_p & w_{33}I_p \end{bmatrix} \begin{bmatrix} X_1'y_1 \\ X_2'y_2 \\ X_3'y_3 \end{bmatrix} \quad (5)$$

The choice of the  $w_{ij}$  for each year is to be determined. Years close to the year being estimated should get most weight. The X-11 procedure uses variations of the  $3 \times 5$  seasonal filter for this purpose, applying asymmetric versions near the ends of series. We use a formula from signal extraction theory (see Cleveland and Tiao (1976) and references). The seasonal factors for a given week of the year (or month for monthly data) are assumed to follow an autocorrelated random walk. The detrended data are this seasonal part plus white noise, which means no autocorrelations in the detrended, seasonally adjusted series of lag one year. Given the model

$$\begin{aligned} y_t &= u_t + e_t \\ (1-B)(1-\phi B)u_t &= a_t \end{aligned}$$

with white noise terms  $e_t$  and  $a_t$ , the weights to estimate  $u_t$  given  $y$  form the desired  $W^*$  matrix. These are obtained from

$$E(u|y) = (I + \nu \Sigma_u^{-1})^{-1} y = W^* y$$

where  $\nu = \sigma_e^2 / \sigma_a^2$  and  $\Sigma_u$  is the autocorrelation matrix of  $u$ . The values  $w_{ij}^*$  are the weights  $w_{ij}$  in (5). The weights are more concentrated (distant years have less impact) for smaller values of  $\phi$  and  $\nu$ , but the pattern is much more sensitive to changes in  $\nu$ . Two examples of  $W^*$  for a series of length 9 years are given in Table 2.1. The rows are labeled for the year being estimated and contain the weights for that year (bold) and adjacent years. Note that the last rows show symmetric weight patterns. For seasonal series with more noise, it makes sense to use more data to extract an estimate of the seasonal signal. With  $\nu=10$ , the first three years provide more than 80 percent of the weight for Year 1; with  $\nu=24$ , they provide about two-thirds of the weight.

Table 2.2 shows the weights for X-11 seasonal adjustment with  $3 \times 5$  and  $3 \times 9$  seasonal filters (cf. Shiskin, Young, and Musgrave, 1967). While there is a rough correspondence with the filters from the signal extraction formula, the signal extraction weights exhibit exponential decay and tend to concentrate more on the year being estimated.

The results of these operations might be termed a seasonal kernel regression, with the shape of the kernel and smoothing parameter determined by  $\nu$  (or  $\phi$  and  $\nu$  if  $\phi$  is also allowed to vary). Use of the signal extraction formula automatically supplies correct kernel shapes at the ends of the series. It is up to the analyst to choose the number of trigonometric terms and the value of  $\nu$ . It is best to set some reasonable values for  $\phi$  and  $\nu$  (e.g. 0.5 and 15) and adjust the number of frequencies first, as likelihood ratio tests for adding new frequencies based on the increased degrees of freedom used are fairly straightforward and the results are fairly sensitive to the number of frequencies used. It is reasonable to start with 6 frequencies (12 terms) and add new frequencies 6 at a time rather than one at a time.

Table 2.1. Year Weights for Two Choices of  $\phi$  and  $\nu$

a.  $\phi=0.5, \nu=10$

Year									
1	<b>0.384</b>	0.271	0.169	0.096	0.049	0.023	0.008	0.002	-0.002
2	0.271	<b>0.264</b>	0.197	0.127	0.074	0.039	0.019	0.007	0.002
3	0.169	0.197	<b>0.218</b>	0.170	0.114	0.068	0.037	0.019	0.008
4	0.096	0.127	0.170	<b>0.203</b>	0.163	0.111	0.068	0.039	0.023
5	0.049	0.074	0.114	0.163	<b>0.200</b>	0.163	0.111	0.074	0.049
Phase shift	1.254	0.706	0.358	0.146	0.000	-0.146	-0.358	-0.706	-1.254

b.  $\phi=0.5, \nu=24$

Year									
1	<b>0.285</b>	0.226	0.165	0.116	0.078	0.052	0.035	0.024	0.018
2	0.226	<b>0.215</b>	0.174	0.130	0.092	0.064	0.044	0.031	0.024
3	0.165	0.174	<b>0.177</b>	0.148	0.113	0.083	0.060	0.044	0.035
4	0.116	0.130	0.148	<b>0.160</b>	0.138	0.109	0.083	0.064	0.052
5	0.078	0.092	0.113	0.138	<b>0.155</b>	0.138	0.113	0.092	0.078
Phase shift	2.001	1.319	0.788	0.366	0.000	-0.366	-0.788	-1.319	-2.001

Table 2.2. Year Weights for X-11 Seasonal Filters

a.  $3 \times 5$

Year									
1	<b>0.283</b>	0.283	0.283	0.150	0.000	0.000	0.000	0.000	0.000
2	0.250	<b>0.250</b>	0.250	0.183	0.067	0.000	0.000	0.000	0.000
3	0.150	0.217	<b>0.217</b>	0.217	0.133	0.067	0.000	0.000	0.000
4	0.067	0.133	0.200	<b>0.200</b>	0.200	0.133	0.067	0.000	0.000
5	0.000	0.067	0.133	0.200	<b>0.200</b>	0.200	0.133	0.067	0.000

b.  $3 \times 9$

Year									
1	<b>0.246</b>	0.221	0.197	0.173	0.112	0.051	0.000	0.000	0.000
2	0.208	<b>0.192</b>	0.176	0.160	0.144	0.092	0.028	0.000	0.000
3	0.173	0.163	<b>0.154</b>	0.143	0.133	0.123	0.079	0.032	0.000
4	0.141	0.137	0.132	<b>0.128</b>	0.123	0.117	0.113	0.075	0.034
5	0.111	0.111	0.111	0.111	<b>0.111</b>	0.111	0.111	0.111	0.111

The problem of choosing  $\nu$  remains. In kernel or smoothing spline regressions, the smoothing parameter is optimized by minimizing some sort of penalized residual sum-of-squares, or a cross-validation technique. As suggested in Härdle et al. (1988), convergence is slow and the surface rather flat. Given that there is no simple objective criterion for a seasonally adjusted series, it may be helpful to create seasonal factors for a set of  $\nu$  values. One can compute the smoothness of the resulting seasonally adjusted series using concurrent and projected factors and the size of revisions when new data are added. This is similar to the approach in Grillenzoni (1994) of minimizing one-step-ahead projection errors.

In a standard regression the number of degrees of freedom used in the regression is given by  $tr(X(X'X)^{-1}X')$ . In the case of weighted regression this generalizes to  $tr(X(X'\Omega X)^{-1}X'\Omega)$ , where  $\Omega$  is a diagonal matrix containing the weights. For our application we used the trace of the matrix in (5),

which equals the trace of the annual weight matrix partially illustrated for particular choices of  $\phi$  and  $\nu$  in Table 2.1. This is in agreement with calculations described in Zhang (2003).

The program begins by running an unweighted or global regression on the specified model. The observed series is adjusted for all variables in the model other than the trigonometric variables, so that what remains is, at least in principle, weekly seasonality plus noise. Then, the weighted regressions described above are carried out to obtain estimates of trigonometric coefficients for successive years, one year at a time. As mentioned, holiday coefficients are estimated with a fixed-coefficient regression. Estimating an “average” holiday effect across calendar variations helps stabilize holiday adjustment. Also, some holiday events do not occur every year, so a long span provides more occurrences for estimation. Section 5 has some examples. The trigonometric and built-in holiday sections of the design matrix are automatically extended two years beyond the input data span to compute projected seasonal factors. These projected factors are needed when seasonal adjustment is not being done contemporaneously.

### 3. Holidays and Outliers

The program contains calendar features to assist in handling holidays and outliers. A flexible treatment of U.S. holidays is built into the program, and holidays of other countries or other adjustments can be entered as user-supplied variables. A list of the precoded holidays is in Table 3.1. The application in Section 5 will illustrate these features. It is important to estimate holiday effects, because many change weeks of the year from one year to the next. The program asks for a holiday code (which of the precoded holidays is at issue or 0 for a user-supplied variable), and for the parameters of a daily weight pattern around the holiday. The first parameter is the number of weights to be used (often 1), then the index in the pattern of the day of the holiday, and finally the weights. If the effect of the holiday is only to cancel activity on that day, a single weight is fine. The value of the weight is not important, as its effect is scaled by the regression coefficient estimated. A weight value of 1.0 or 0.1 usually gives a regression coefficient with a good printable size. In the case of Christmas (or Easter) sales in previous days might be boosted. If so, one could specify several positive weights with Christmas falling on the last one or on a final zero weight. The program determines in which week of the year and where in the week the holiday falls, and assigns the weights to the correct week or weeks. While statistical tests could be tried comparing various weight patterns, the patterns are often dictated by external sources of information.

In addition to the holidays which are coded into the program, the user may read in a column of the design matrix as a user-specified holiday effect. This would be a column of data with non-zero entries in rows corresponding to dates when the effect was present. This column should include dates two years beyond the latest date being used in estimation to allow for projected seasonal factors.

Table 3.1. Holiday Codes

Code	Holiday	Code	Holiday
0	User-supplied effect	6	Christmas
1	Easter	7	July 4
2	Labor Day	8	Presidents Day
3	New Year’s Day	9	Thanksgiving
4	Memorial Day	10	M.L. King Day
5	April 15 taxes	11	Veterans Day
		12	Columbus Day

The program also includes a way to compensate for additive outliers; it does not do automatic identification. There are plans to expand the program to handle a broader range of outlier/intervention effects, including level shifts. Identification of additive outliers may come from known events or analysis of graphs. The program takes as input parameters the number of outliers to be treated and a corresponding sequence of dates on which they occur in terms of year and week of the year. The week of the year can be easily identified from the data listings generated by the program (see Section 6 for more details).

#### 4. Diagnostics

As described in Section 2, the first step in the method is to fit a global regression model. A standard ANOVA table exhibits the overall contribution from seasonal, holiday, outlier, and linear trend variables. In addition, t-statistics and p-values are available for individual trig coefficients, holidays, additive outliers, intercept and slope variables, and, when included, user-supplied variables. This global regression has the following functions:

Table 4.1. MoveReg Graphs

Plot	Description
Time series plot	Shows the unadjusted and seasonally adjusted series for the full time span. Additive outliers are marked by a circle, and official U.S. recessions are shaded in gray.
Time series plot	Same as the first, but only for the last 5 years.
Time series plot	Same as the first, but only for the last 2 years.
Plot of initial and revised seasonally adjusted series, along with the unadjusted series, for the last year	The initial or preliminary adjusted values use trig coefficients based on data through the next-to-last year (so that values for the last year come from projected factors); the revised values use trig coefficients from data through the last year.
Average seasonal factors by week; plot of seasonal factors	The plot of averages highlights the overall seasonal pattern across the year. The time series plot can be helpful in assessing whether strength of seasonality is increasing or not.
Projected seasonal factors	Time series plot of projected seasonal factors (typically 2 years), with 3 previous years of factors.
Seasonal sub-plots by week	Assesses movement of factors for individual weeks. The straight lines are the weekly means for the factors.
Trig coefficient sub-plots by frequency	Indicates which frequency terms are significant and how much volatility they exhibit. The straight lines are the means by frequency.
Autoregressive spectrum	The SAS interface utilizes X-12-ARIMA to calculate the spectrum for the differenced seasonally adjusted and unadjusted data. The horizontal lines are the medians.

- (1) allows the user to assess the overall model and the value of each individual variable in the model;
- (2) adjusts the series for all submitted variables except the seasonal variables.

Then, the weighted regressions are carried out. With this procedure, all but the seasonal variables are treated as deterministic, with all years contributing equally to estimation of their effects. Typically, for holidays or other special variables, it is helpful to base estimation on as much data as possible. In addition,  $R^2$  and Ljung-Box statistics help assess overall suitability of the model.

The SAS interface part of the MoveReg program produces 10 graphs by default and outputs them to a pdf. Some of the graphs are displayed for the Unemployment Insurance application in Section 5. Table 4.1 has an annotated list of the graphs in the order that they are produced.

## 5. Example: Initial Claims, U.S. Unemployment Insurance Program

Among important U.S. economic time series is Initial Claims from the Unemployment Insurance program. Individuals apply for unemployment benefits at government offices across the country. Claims data from these local offices are forwarded first to the state office, and then to the Department of Labor (DOL), Washington, D.C. Every Thursday ([http://www.doleta.gov/ETA\\_News\\_Releases/](http://www.doleta.gov/ETA_News_Releases/)), Initial Claims (IC) are reported for the previous week ending on Saturday and Continuing Claims for the week before that. With less than a week's time lag, the Initial Claims series is clearly one of the most timely indicators of the state of the economy. Press releases and many media accounts feature weekly figures and 4-week moving averages, both seasonally adjusted. During the economic recession beginning in 2007 and continuing into the recovery period, these numbers have been scrutinized closely.

Figure 5.1 shows the observed and seasonally adjusted series across the years 1988-2009, plus the additive outliers. Three features are apparent: seasonality, volatility, and upturns during the three official recessions, shaded in gray. Figure 5.2 graphs average weekly seasonal factors across the span 1988-2008. An extreme peak occurs at the beginning of each year. Seasonality is low during the second quarter. In the third quarter, a short peak occurs near the beginning, followed by a trough around the end. Figure 5.2 also contains the time series of the seasonal component. It appears that the early peak and the third quarter trough have moderated in recent years. Evident in both series, volatility is due to differences in both reporting practices and economic behavior across the individual states. Naturally enough, the number of claims tends to increase in difficult economic times. Adopted in 2002 at BLS, the locally-weighted regression approach accommodating moving seasonality has significantly improved the seasonal adjustment, in particular in terms of (1) smoothness of the seasonally adjusted series and (2) less evidence of residual seasonality (cf. Cleveland and Scott, 2007). Figure 5.3 has sub-plots of the seasonal factors by week, showing how the factors move across years. Recall that even with deterministic trig coefficients these plots would show variability, due to changing positions of the weeks. Estimating a stochastic seasonal component continues to be pertinent, since the recent recession and changes in economic activity appear to be affecting seasonality. Since it is not entirely clear what is trend and what is seasonality for 2008 and 2009 at the time this is being written, we focus on the period 1988, Week 4 – 2008, Week 4, a series of length 1045 which avoids most of the recession which ended in June 2009.

As described in Section 2, after differencing, the data are modeled as a seasonal time series, along with calendar effects and outliers. The model can produce a seasonal factor for each day of the year. For Week  $i$ , we use the factor corresponding to  $d(i)$ , where  $d(i)$  is the day of the year on which the Saturday of Week  $i$  falls. For example, in 2011, Weeks 1, 2, and 6 have their Saturday's on January 1, January 8, and February 5, so the seasonal factors for those weeks are computed for days 1, 8, and 36. We first address selection of the seasonal variables. A conservative approach is to add trig terms in sets of 6 frequencies,



beginning with the fundamental frequency  $\lambda_1 = 1/365$ . A more selective approach would be more parsimonious, but with over a thousand observations, parsimony is not a serious issue. Figure 5.4 graphs spectra for differenced unadjusted and seasonally adjusted series. The gray lines denote the seasonal frequencies for weekly data, computed as  $7k/365.24, k = 1, \dots, 26$ . The spectrum for the unadjusted series contains peaks at or near several of the seasonal frequencies, even some corresponding to periods nearly as short as “every two weeks.” The adjustment yielding this spectrum is based on setting  $p=60$ , that is, including the first 30 seasonal frequencies. Figure 5.5 shows the coefficients for each sine and cosine variable. (All but the first plot have the same scale.) Most of the visually significant values come from the first 12 frequencies, but others from each set of six seem different from 0. For 26 of 30 frequencies, at least one of the trig terms has a significant t-statistic; these statistics exceed 3 for 21 frequencies. These statistics come from the global regression model (1) with uniform weights.

Table 5.1 provides further confirmation for inclusion of all these seasonal variables. It contains a set of residual F-tests from adding successive sets of 6 frequencies to the regression model. The change in residual sum of squares is attributable to the 12 added variables and compared to the mean residual sum of squares for the expanded model. We see that each F-test supports the inclusion of additional variables.

Every government holiday is potentially significant for this series because it represents loss of a day for filing claims. Table 5.2 shows that holiday variables have t-statistics ranging from 2.6 to 14.2 in magnitude. All require special treatment, since they affect different weeks in different years. July 4 can occur in either Week 27 or 28. While not a Federal holiday, Easter affects claims activity and moves between Weeks 12-17. The largest effect comes from Thanksgiving, since in addition to the Thursday holiday, many state offices are closed on Friday. Even if an office is open, applicants may be less likely to appear that Friday. Special user variables are introduced for occurrence of July 4 on Wednesday and for particular patterns related to Christmas.

A surge in claims activity occurs when the Christmas season comes to a close. This is modeled as a New Year’s Day effect. From our definition of weeks, New Year’s Day is always in Week 1. The New Year’s effect is modeled as an increase generally assigned to Week 2. When January 1 falls on Saturday, the rest of Week 1 is in December, so clearly the surge comes in Week 2. The surge seems to fall in week 2 for other positions of January 1 as well, except when it is on Sunday. In this case, the effect is split between weeks 1 and 2. The holiday specification which gives this result is 8 1 0. 0. 0. 0. 0. 0. 1. 1., where 8 is the number of weights, 1 is the position of January 1 in the weight sequence, and the following 0’s and 1’s are the weights. When January 1 is Sunday, one nonzero weight falls in week 1 and the other in week 2. Otherwise, both nonzero weights correspond to days in Week 2. From this example, we see that the program can handle a variety of holiday effects. A limitation is that holiday effects are treated as time-invariant.

Table 5.1. Sequential F-tests for Inclusion of Seasonal Frequencies  $\lambda$

# $\lambda$ 's	Sum of Squares			Mean Squares		Seq'l F
	Model (df)	Error (df)	Added $\lambda$ 's	Error	Added $\lambda$ 's	
<b>6</b>	10.73 (48)	7.02 (996)	-	.0071	-	-
<b>12</b>	13.78 (60)	3.98 (984)	3.05	.0040	.2540	63.5
<b>18</b>	14.73 (72)	3.03 (972)	0.95	.0031	.0789	25.4
<b>24</b>	15.40 (84)	2.36 (960)	0.67	.0025	.0561	22.4
<b>30</b>	15.70 (96)	2.06 (948)	0.29	.0022	.0248	11.3

Table 5.2. Holiday Effects and Associated t-statistics

<b>Holiday</b>	<b>Factor</b>	<b>t-statistic</b>	<b>p-value</b>
New Year's Day	1.08	8.1	0.0000
Martin Luther King Day	0.83	-11.8	0.0000
President's Day	0.94	-4.0	0.0000
Easter	0.96	-5.7	0.0000
Memorial Day	0.89	-7.2	0.0000
July 4	0.96	-2.6	0.0043
July 4 on Wednesday (user)	1.11	5.1	0.0000
Labor Day	0.89	-7.1	0.0000
Columbus Day	0.96	-2.7	0.0035
Veterans Day	0.87	-9.5	0.0000
Thanksgiving	0.80	-14.2	0.0000
Christmas in Week 53 (user)	1.05	2.4	0.0078
Christmas on Friday (user)	0.91	-3.5	0.0003

Table 5.3. Outlier Effects and Associated t-statistics

<b>Year</b>	<b>Week</b>	<b>Factor</b>	<b>t-statistic</b>	<b>p-value</b>
1989	40	1.19	5.0	0.0000
1992	30	1.41	10.0	0.0000
1993	30	1.34	8.6	0.0000
1993	52	0.87	-3.9	0.0001
1994	5	1.13	3.6	0.0002
1996	3	1.17	4.4	0.0000
2001	38	1.13	2.8	0.0030
2001	39	1.28	4.3	0.0000
2001	40	1.17	2.5	0.0063
2001	41	1.16	2.4	0.0093
2001	42	1.13	2.1	0.0188
2001	43	1.11	2.3	0.0108
2001	47	1.16	3.8	0.0001
2001	48	1.16	3.6	0.0002
2005	37	1.29	5.8	0.0000
2005	38	1.25	4.0	0.0000
2005	39	1.07	1.2	0.1117
2005	40	1.13	2.2	0.0141
2005	41	1.11	2.4	0.0090
2006	1	1.10	2.8	0.0022
2007	2	0.87	-4.1	0.0000

As usual in seasonal adjustment, outlier and intervention specification is an iterative process. We take advantage of the automatic outlier detection in X-12-ARIMA (<http://www.census.gov/srd/www/x12a/>) by submitting residuals from the overall regression model. Candidate outliers are then tested within an expanded overall regression model. The process is repeated until a final selection is made. Dates and t-statistics for the 21 selected outliers are given in Table 5.3. All are simple additive outliers.

Hurricane Katrina struck the U.S. Gulf Coast on August 29, 2005 with devastating effects in Louisiana, Mississippi, and Alabama. This resulted in a spike in Initial Claims, modeled as five outliers. Two runs have been carried out with data through 2006, Week 4, one with and one without outlier treatment. Results are compared to a run with data through the week ending August 27, 2005. We do not want the hurricane, an extraordinary event, to influence estimation of the seasonal component. Figure 5.6 compares seasonal factors from all three runs for the last half of 2005. Differences from the run ending August 27 appear minor, except for three weeks in September and two weeks in October, where clearly the “With treatment” factors agree better.

For the weights matrix, we follow the suggestion in Section 2 of setting the AR parameter  $\phi = .5$  (Box-Jenkins, 1970, sign convention). For  $\nu$ , a comparison of smoothness with  $\nu = 10, 16,$  and  $24$  leads us to select  $\nu = 16$ . Based on the discussion in Section 2, this yields a weighting scheme intermediate between X-11’s 3x5 and 3x9 seasonal filters.

To analyze stability, we compute revisions to seasonal adjustment from two successive runs. Figure 5.7 overlays observed and seasonally adjusted data from 2007, Week 5 through 2008, Week 4. The series in green is the initial or preliminary adjusted series, based on projected factors derived from data through 2007, Week 4. The series in red comes from the adjustment we have been examining, with data through 2008, Week 4. The largest revisions occur in February and April 2007 and January 2008. For all three months the revised series is smoother. The median absolute revision is 4,118 for the year 2007, referring to the period 2007, Week 5 – 2008, Week 4. The maximum absolute revision is 21,873 (6.46%), occurring in 2007, Week 6. Table 5.4 shows revision statistics from these two runs for 2007 and earlier years. We see the revisions decline going back in time as we approach use of the central filter for weighting. The drop in these statistics is especially strong moving back to 2004. For 2002, the median value is below 1000. These numbers are large, reflecting considerable volatility in the data. Still, the seasonally adjusted values themselves are quite large, mostly in the 300,000 – 400,000 range, and Figure 5.7 shows relatively good agreement for most weeks.

Returning to Figure 5.4, there is little evidence of residual seasonality in the seasonally adjusted spectrum. There is a minor peak remaining near the 6<sup>th</sup> seasonal frequency, but even that peak is much reduced from that for the unadjusted series.

Table 5.4. Absolute Revisions by Year in Seasonally Adjusted Series, Based on 2007 Run

<b>Year</b>	<b>Median</b>	<b>Root Mean Square</b>	<b>Maximum</b>
2002	654	1345	4304
2003	1007	1976	4391
2004	1279	2726	6797
2005	2080	3697	11085
2006	2931	4973	15790
2007	4118	6951	21873

Note. The statistics are based on “offset” years. For instance, Year 2007 refers to the 52 week period 2007, Week 5 – 2008, Week 4. Input series for official annual adjustments end with January weeks, since data tend to stabilize somewhat following the December holiday period.

## 6. Software

### *Description*

The MoveReg software for carrying out weekly seasonal adjustment consists of a FORTRAN executable and a SAS interface. The original code for weekly seasonal adjustment with moving seasonality was written by William P. Cleveland in the FORTRAN programming language. This code is available for both SOLARIS and Windows XP, but can be compiled for other platforms. Since BLS began using this program, modifications have been made to the FORTRAN code to add diagnostics and improve the machine readability of the output files. The FORTRAN program requires two input files and prints seven output files in text format. The SAS interface simplifies execution and aids in the preparation of the input files.

The rest of this section presents in some detail features and instructions for the software. More complete documentation is available at the BLS website [to be supplied]. This site contains (1) a MoveReg software manual, (2) FORTRAN executables, (3) sample SAS programs, and (4) a full example with data, input and output files, and analysis of the results.

### *Input and output files*

The two input files are called CONTROL.TXT and DATA.TXT. CONTROL.TXT specifies program options and parameters. A complete example for the Initial Claims series of Section 5 is shown below:

#### Example 6.1. Sample CONTROL.TXT file

```
52 0
21 13 2 2 nout hol nfilt mxtype
0.4 16 phi sigratio
60
1988 01 16 2008 01 19 7
40 1989 30 1992 30 1993 52 1993 5 1994 3 1996 38 2001 39 2001
40 2001 41 2001 42 2001 43 2001 47 2001 48 2001 37 2005 38 2005
39 2005 40 2005 41 2005 1 2006 2 2007
0 0 0 3 10 8 1 4 7 2 12 11 9
8 1 0 0 0 0 0 0 1. 1. New Year
1 1 1. M. L. King Day
1 1 1. President's Day
8 8 1. 0 0 0 0 0 0 0 0 Easter
1 1 1. Memorial Day
1 1 1. July 4
2 2 0. 1. Labor Day
1 1 1. Columbus Day
1 1 1. Veterans Day
1 1 1. Thanksgiving
```

Key:

Line	Description
1	NPER 52 for weekly , 12 for monthly
	LOC 1 for fame database, <sup>1</sup> 0 for ascii file (data.txt)
2	NOUT no. of AO's (70 maximum)
	HOL no. of built-in holidays

<sup>1</sup> The BLS version of the program does not support FAME.

- NFILT           width of detrending filter (2 for the factor  $I-B$ )
- MXTYPE        1 for trapezoidal weighting (not documented), 2 for ARIMA weighting
- 3 PHI           AR parameter  $\phi$  (only when MXTYPE=2)
- SIGR        variance ratio  $v$  (only when MXTYPE=2)
- 4 NFS          no. of terms in trig seasonal = 2 x no. of frequencies
- 5 KBYR,KBMO,KBDAY,LYR,LMO,LDAY,IWKD  
               starting & ending dates for input data (4-digit year, month, day of month),  
               days in week
- 6               dates of outliers (week & year for each)
- 7               holiday codes preceded by '0' for each user variable
- 8               information on built-in holidays, 1 line per holiday:  
               length of holiday effect, position of holiday in weight  
               pattern, weight pattern (the most common weight pattern is 1 1 1, which simply  
               generates a 1 in the weeks where the holiday occurs and zeroes elsewhere).

DATA.TXT contains the input data plus any user-defined events. A shortened example follows:

Example 6.2. Sample DATA.TXT file (extract)

```
ic
      1149  52  1988   01   30  2010   01   30   7   0
(46x,f8.0)
30JAN1988   ic           1 30  1988   5       1   395000
06FEB1988   ic           2  6  1988   6       2   381000
13FEB1988   ic           2 13  1988   7       3   335000
20FEB1988   ic           2 20  1988   8       4   316000
27FEB1988   ic           2 27  1988   9       5   324000
05MAR1988   ic           3  5  1988  10      6   312000
12MAR1988   ic           3 12  1988  11      7   294000
19MAR1988   ic           3 19  1988  12      8   276000
26MAR1988   ic           3 26  1988  13      9   269000
```

Key:

Line 1:       title, identifying info

Line 2:       NOBS, NPER, KKBYSR, KKBMO, etc., IWKD, KBB

              number of observations, number of periods, start year, start month, start day, end year,  
               end month, end day, number of days in a week, '0'

Line 3:       FORTRAN input format (change as needed); note that the data column is read following  
               the format line

Line 4:       date, variable name, month, day, year, week of year, cumulative week, data value (@46, 8.0)

User-defined variables (if any) follow the input data series. An example of the coding is given below. The variable value is zero except when Christmas falls in week 53. This series must be extended to cover the forecast period. Here, the length is 1149, 1045 for the input data plus 104 for the forecast period.

### Example 6.3. Sample specification of a user variable

```
ic USER FACTORS FOR CHRISTMAS IN WEEK 53
 1149 52 1988 01 16          2010          1          16 7 0
(46x,f6.0)
XMAS IN WEEK 53  January  16 1988  3  1  0
XMAS IN WEEK 53  January  23 1988  4  2  0
XMAS IN WEEK 53  January  30 1988  5  3  0
XMAS IN WEEK 53  February  6 1988  6  4  0
XMAS IN WEEK 53  February 13 1988  7  5  0
:
XMAS IN WEEK 53  December 17 1988 51 49  0
XMAS IN WEEK 53  December 24 1988 52 50  0
XMAS IN WEEK 53  December 31 1988 53 51  1
XMAS IN WEEK 53  January  7 1989  1 52  0
XMAS IN WEEK 53  January 14 1989  2 53  0
```

#### Key:

Line 1: title of user variable

Line 2: number of observations, number of periods, start year, start month, start day, end year (include forecasts), end month, end day, '7', '0'

Line 3: FORTRAN input format

Line 4: variable title, month in letters, day, year, week in year, cumulative week, data value (only the last column is actually read)

ANOVA.OUT is a text file with the test statistics discussed in Section 4. There are seven other output files, some in table form and some more amenable to further calculations.

#### *Run instructions with the SAS interface program*

This section describes how to carry out seasonal adjustment with the interface. The input control and data files are automatically created by the SAS code which can significantly reduce the possibility of errors. Note that the interface is available for both UNIX and Windows. If one chooses to execute the program from the command line without the interface, control and data files need to be created with a programming language like SAS or with a text editor. Requirements and examples for setting up these files are in the detailed MoveReg documentation.

Before starting the interface, one or more data files must be created. The time series to be adjusted can be supplied as a SAS data set or a text file. Example 6.2 exhibits the format for the latter option. If "Num of Additive Outliers" is positive, these are to be specified in a file OUTLIERS.TXT. The format should be that of lines 6-8 in Example 6.1. If treatment for U.S. holidays is desired, a text file HOLIDAYS.TXT needs to be supplied. This file should have one line for each holiday, as in Example 6.1, with the same logic (but not necessarily the same weight structure). User-defined events can be added to the data input file by providing SAS code to create dummy variables; a supplied program USER\_EVENTS.SAS can be modified for this purpose.

To begin MoveReg, execute the supplied SAS program, and the window in Figure 6.1 appears. Next, as shown, select "Start" from the Program menu. This will bring up a window to specify program options seen in Figure 6.2. This "Start" window allows the user to specify required program parameters. All available settings can be specified with this window except for holiday weights and user-defined events as explained above.

Figure 6.1. Entry/Exit Window for MoveReg Interface

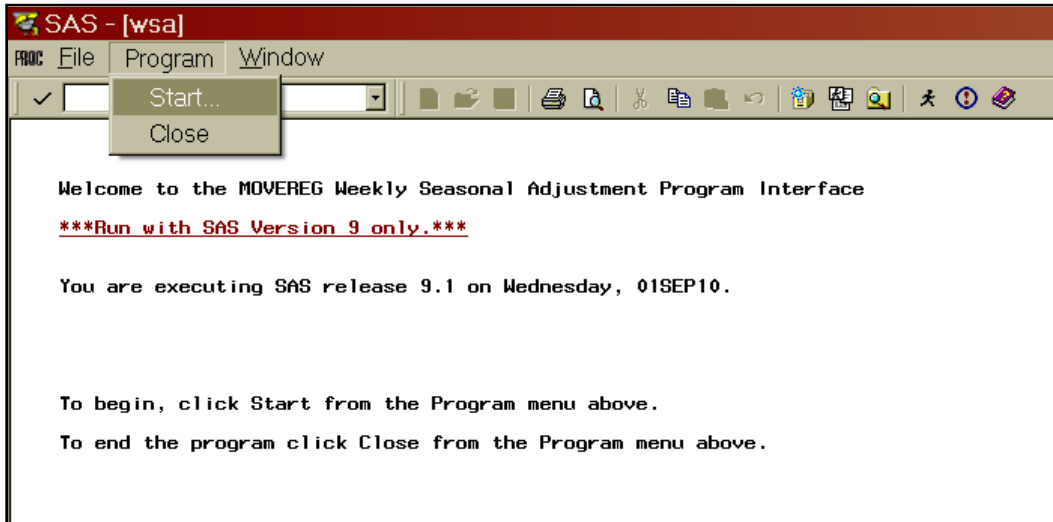
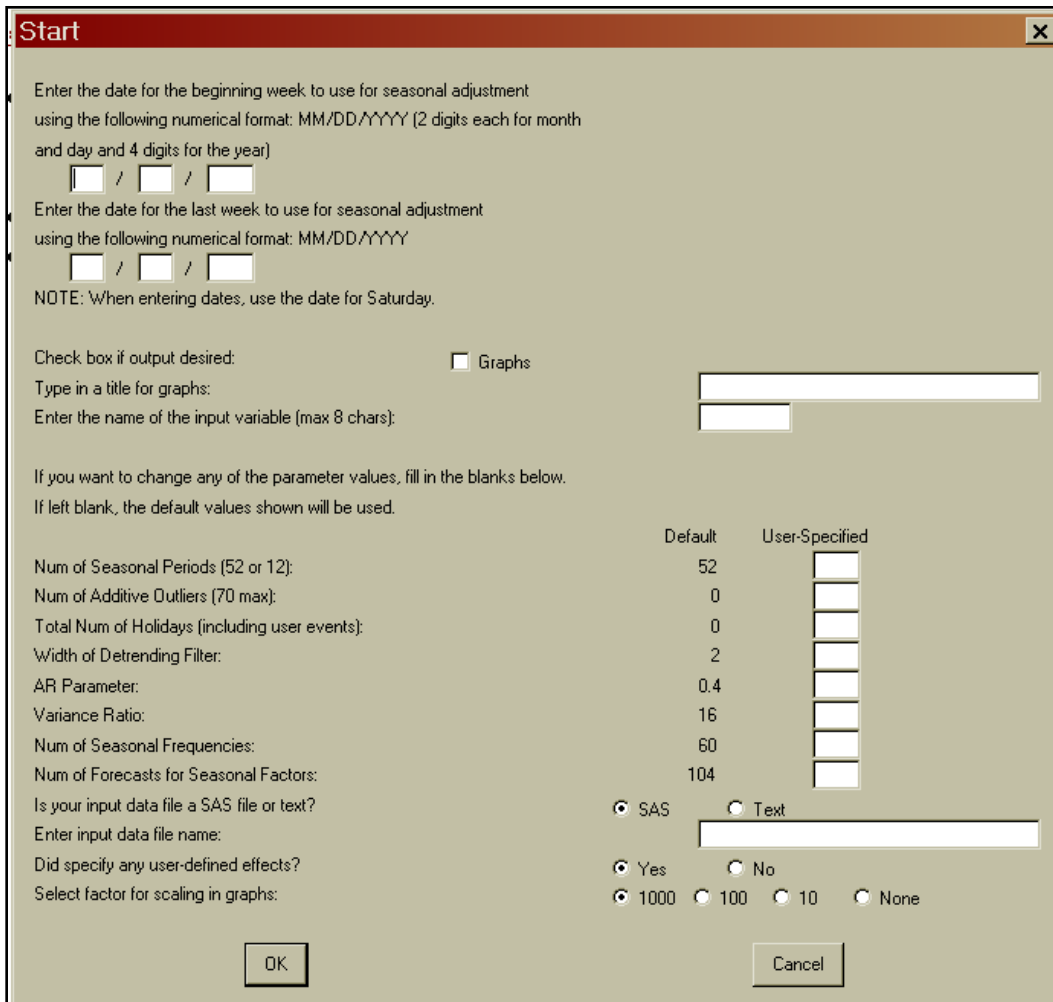


Figure 6.2. Start Window.



The beginning and ending dates are entered at the top of the window. Next, one can indicate if graphs are desired, type in a title, and specify a name for the time series. The eight small boxes that follow are for settings in the control file. The “Width of Detrending Filter” is to specify differencing; the “AR parameter”  $\phi$  controls how fast the seasonal factors move; the “Variance Ratio”  $\nu$  is a noise-to-signal ratio. It works in a similar way to the seasonal moving averages in X-11; the lower the ratio, the more concentrated the weights. The “Num of Seasonal Frequencies” indicates how many sine and cosine pairs are needed for the trig seasonal component; and the “Num of Forecasts for Seasonal Factors” is simply the number of desired projected seasonal factors. The input data can be in either a SAS database or a text file (check the appropriate radio box), and the name of the data file is entered in the next box. Then, if user-defined events are needed for the input data file, one selects the “Yes” radio button in the next line. Finally, scaling choices are provided through radio buttons in the last line for the graphs that are output to a pdf.

A SAS macro is provided to change the default options for the “Start” window. This macro also allows the user to select graphs for the pdf. The choices include time series plots for 1) the unadjusted series, the seasonally adjusted series, and outliers; 2) revisions for the last year; 3) average seasonal factors by week; 4) seasonal factors over time; 5) projected seasonal factors; 6) seasonal sub-plots by week, 7) trig coefficient sub-plots; and 8) a spectral plot.

## References

- Box, G.E.P., and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control*, Holden Day: San Francisco.
- Cleveland, W.P. and Tiao, G.C. (1976), “Decomposition of seasonal time series: a model for the X-11 program,” *Journal of the American Statistical Association*, 71, 581-87.
- Cleveland, W.P. and Scott, S. (2007), “Seasonal Adjustment of Weekly Time Series with Application to Unemployment Insurance Claims and Steel Production,” *Journal of Official Statistics* 23, 209-221.
- Cleveland, W.S., Dunn, D.M., and Terpenning, I.J. (1978), “SABL: A Resistant Seasonal Adjustment Procedure with Graphical Methods for Interpretation and Diagnosis,” with comments, in *Seasonal Analysis of Economic Time Series*, Arnold Zellner (ed.). U.S. Department of Commerce, Washington, D.C. 201-241.
- Gersch, W. and Kitagawa, G. (1983), “The prediction of time series with trends and seasonalities,” *Journal of Business and Economic Statistics*, 1, 253-264.
- Grillenzoni, C. (1994), “Optimal recursive estimation of dynamic models,” *Journal of the American Statistical Association*, 89, 777-787.
- Härdle, W., Hall, P. and Marron, J.S. (1988), “How far are automatically chosen regression smoothing parameters from their optimum?,” *Journal of the American Statistical Association*, 83, 86-95.
- Harvey, A. and Koopman, S.J. (1993), “Forecasting hourly electricity demand using time-varying splines,” *Journal of the American Statistical Association*, 88, 1228-36.
- Harvey, A., Koopman, S.J., and Riana, M. (1997), “The Modeling and Seasonal Adjustment of Weekly Observations,” *Journal of Business and Economic Statistics*, 15, 354-368.
- Pierce, D.A., Grupe, M.R. and Cleveland, W.P. (1984), “Seasonal adjustment of the weekly monetary aggregates: a model-based approach,” *Journal of Business and Economic Statistics*, 2, 260-270.
- Poirier, D.J. (1973), “Piecewise regression using cubic splines,” *Journal of the American Statistical Association*, 68, 515-524 (Corr: vol. 69, 288).
- Shiskin, J., Young, Alan H., and Musgrave, J.C. (1967), *The X-11 variant of the census method II seasonal adjustment program, Technical Paper 15*, Bureau of the Census, Washington, D.C.
- Zhang, Chunming (2003), “Calibrating the degrees of freedom for automatic data smoothing and effective curve checking,” *Journal of the American Statistical Association*, 98, 609-628.



Figure 5.1. Initial Claims for Unemployment Insurance

1988, Week 4 - 2010, Week 4

NBER Recessions in Gray

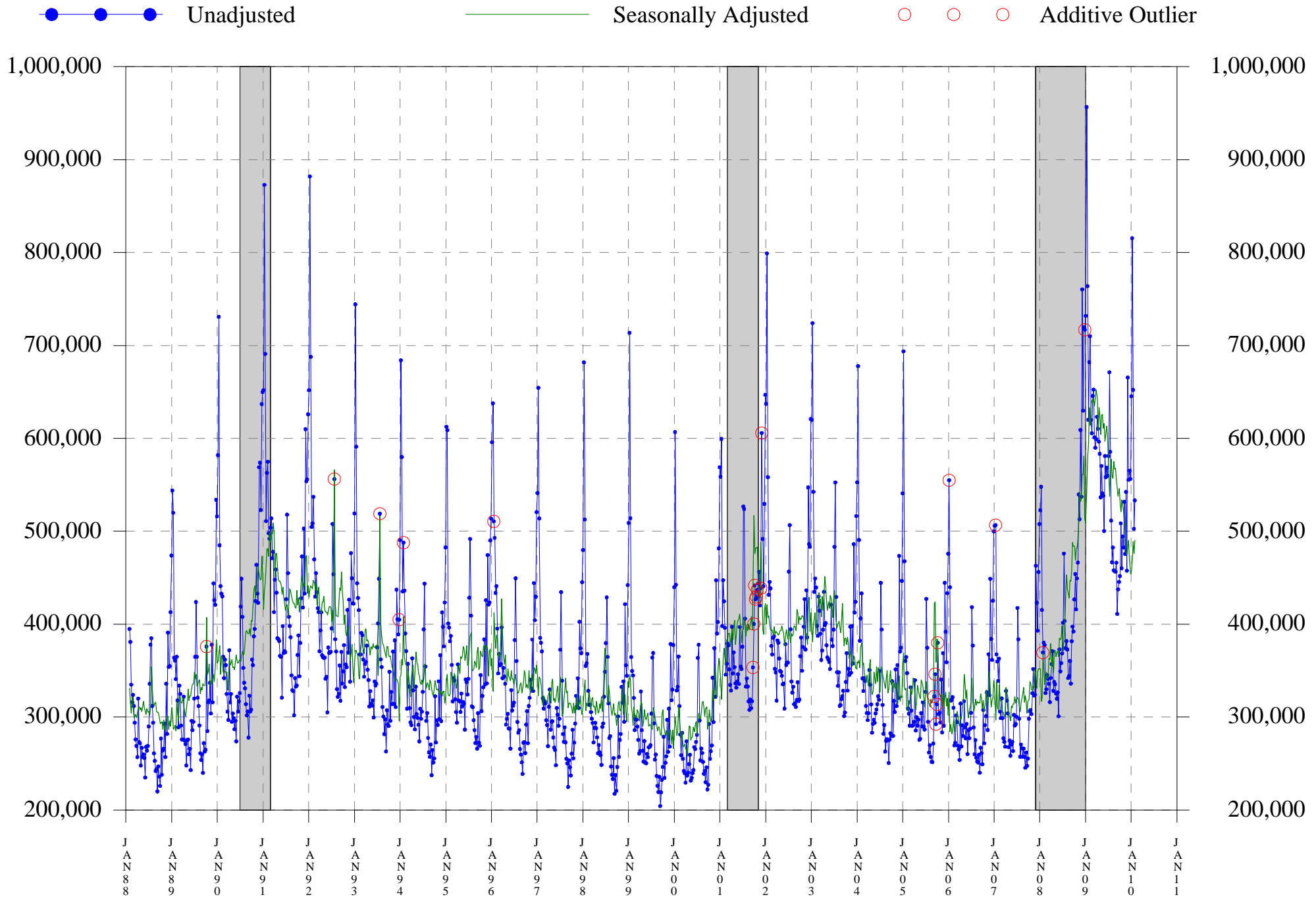
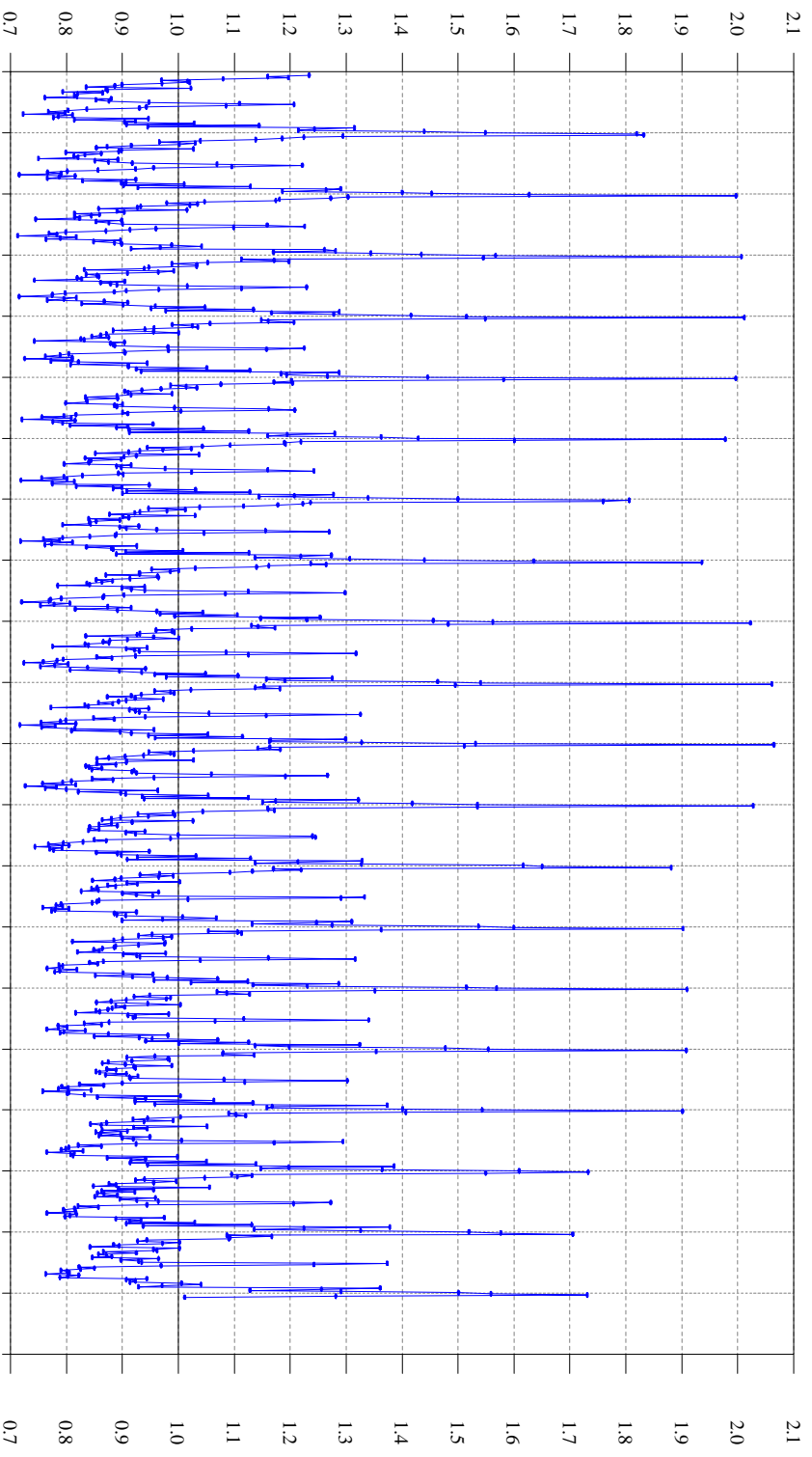
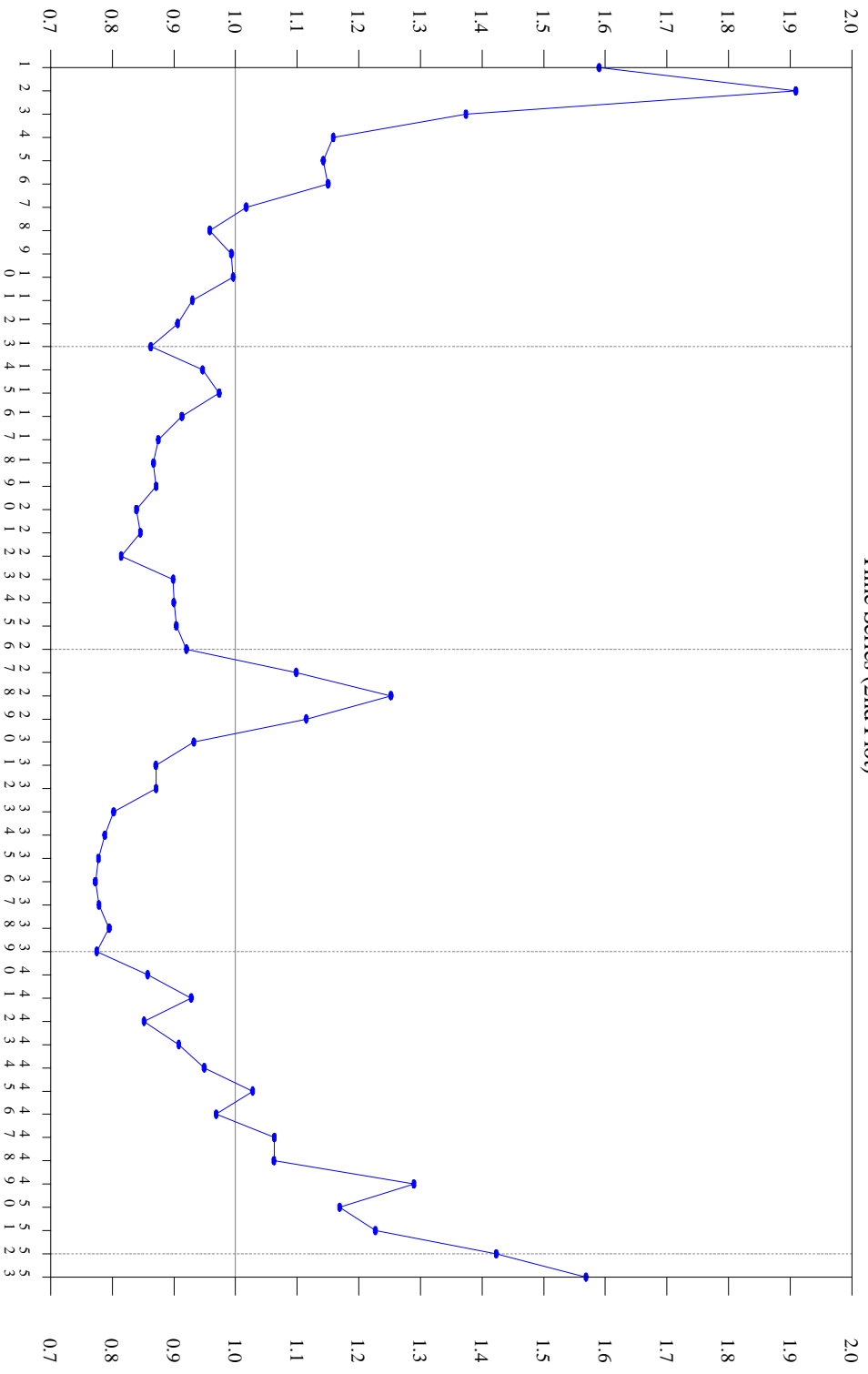


Figure 5.2. Initial Claims Seasonal Factors, 1988, Week 4 - 2008, Week 4  
 Weekly Averages (1st Plot)  
 Time Series (2nd Plot)



### Figure 5.3. Initial Claims Seasonal Factors

Weekly Sub-Plots Across Years

— Factors — Means

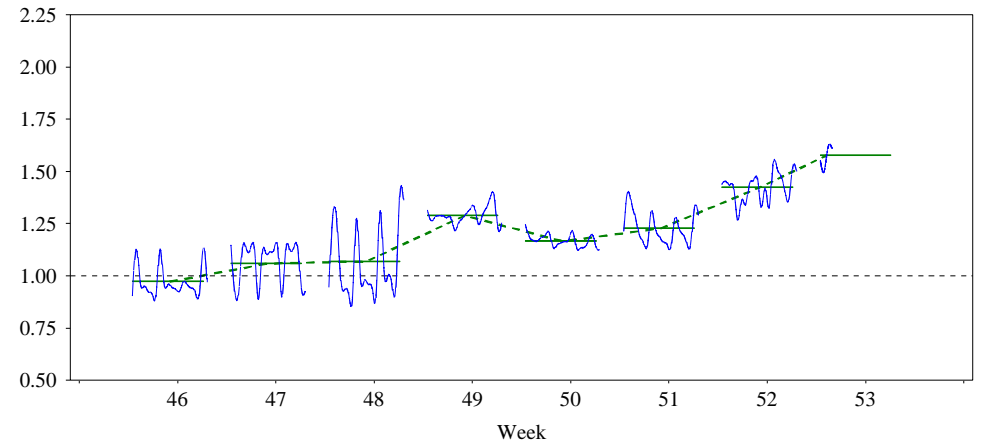
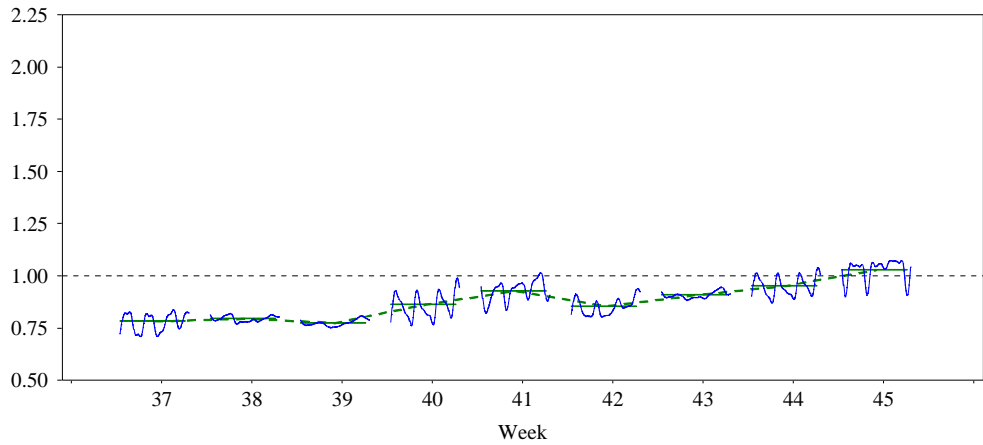
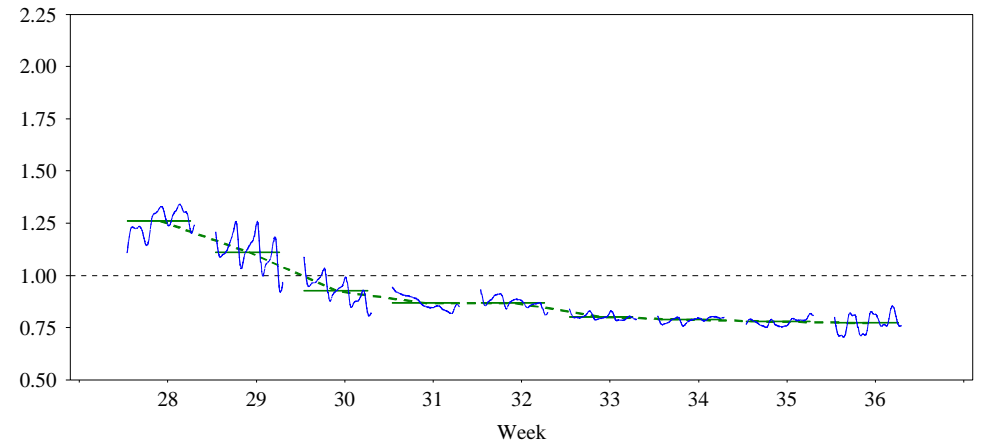
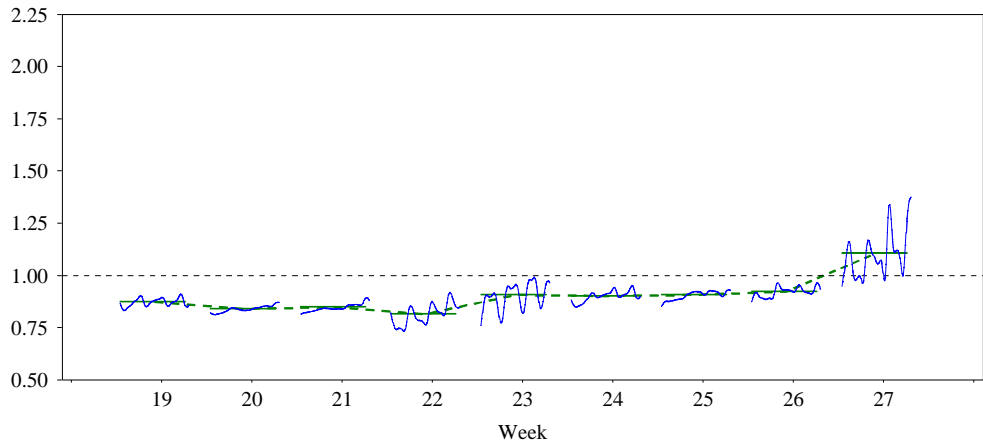
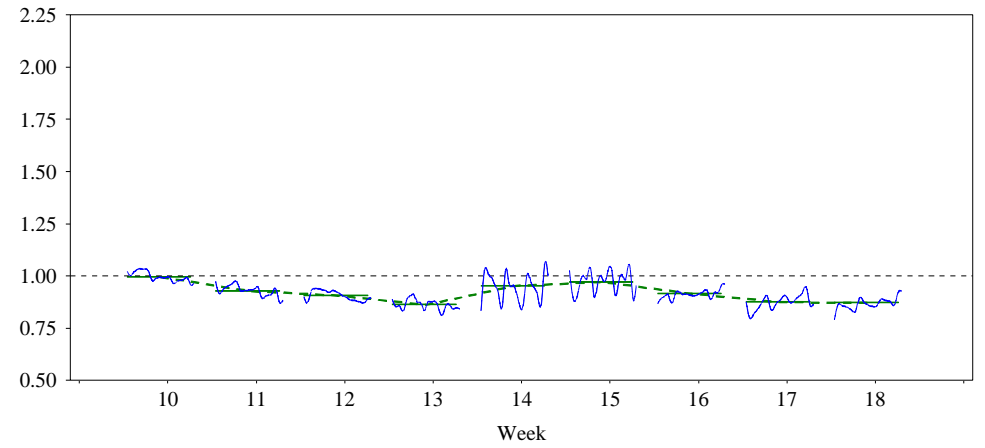
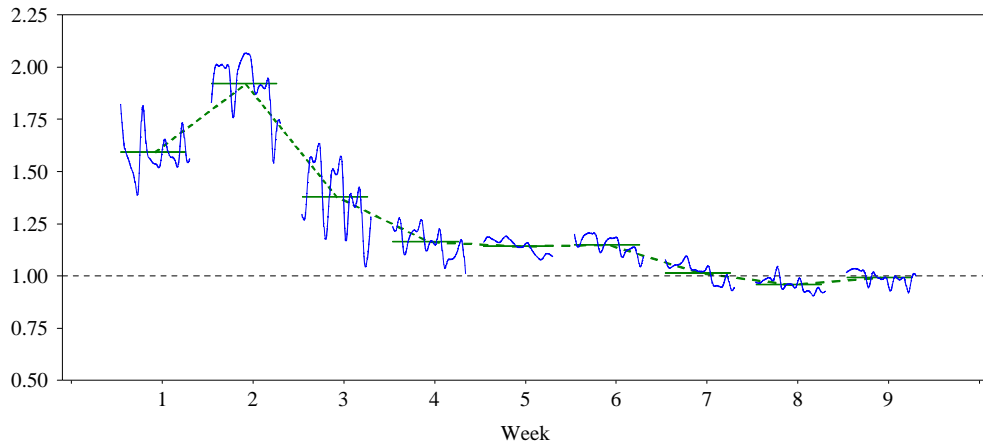


Figure 5.4. Initial Claims Spectra for Differenced Series  
Autoregressive Estimates

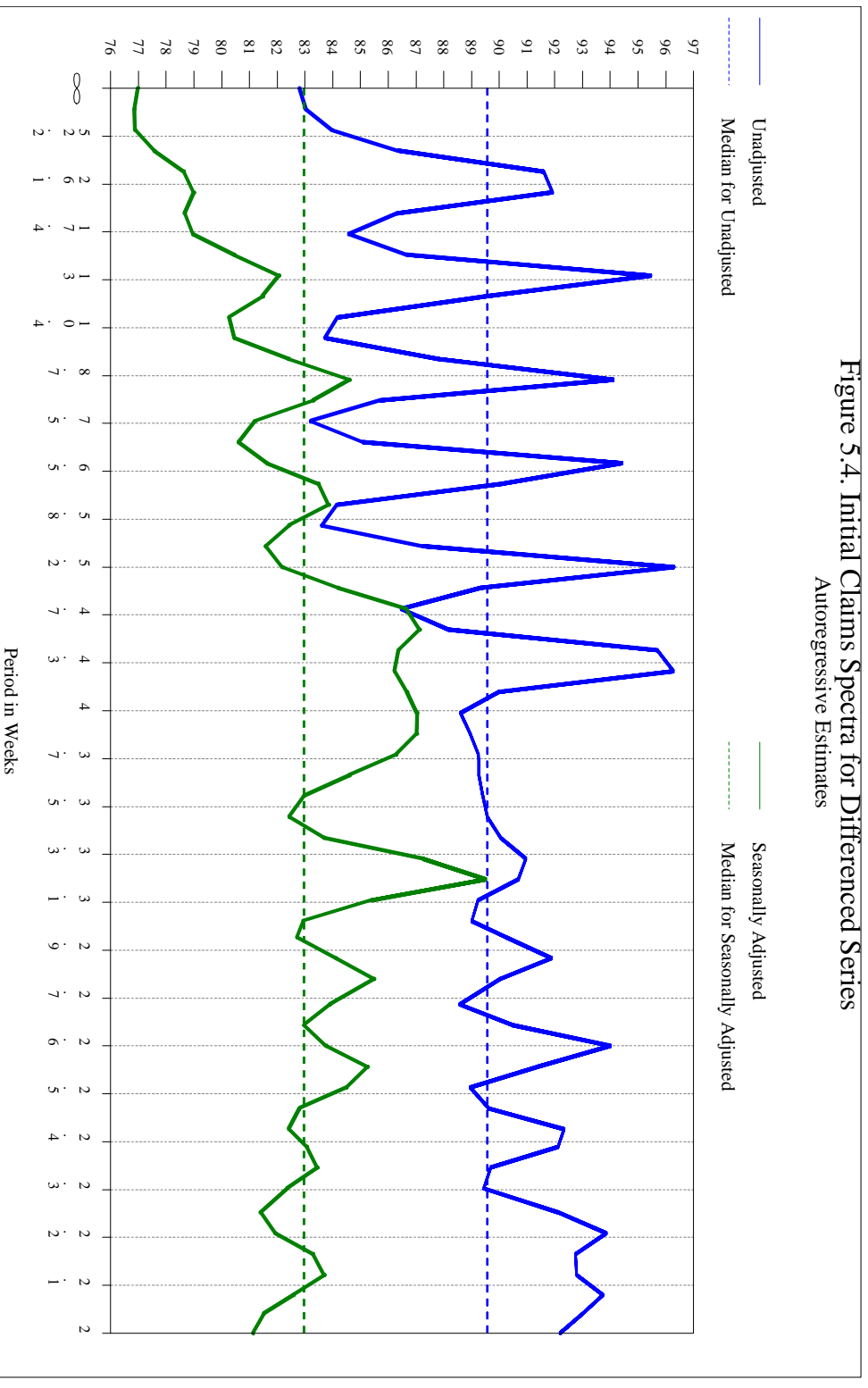


Figure 5.5. Initial Claims Trig Coefficients

Sub-Plots by Sine/Cosine Variable

— Coefficients — Means

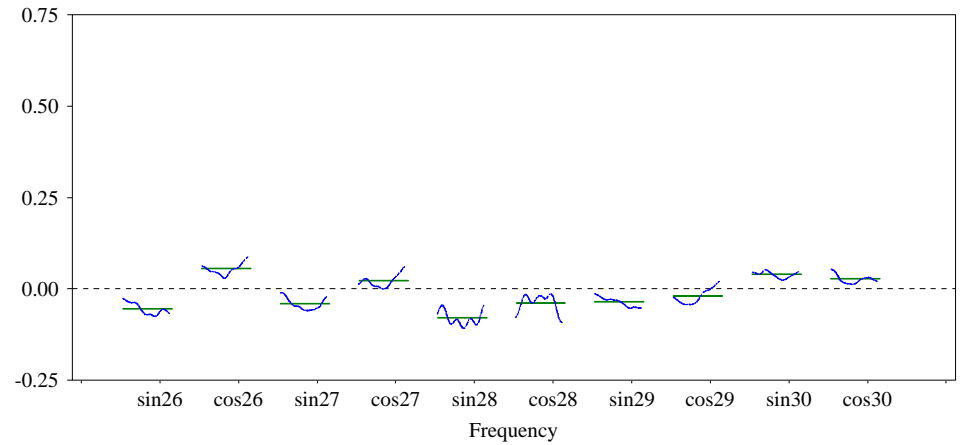
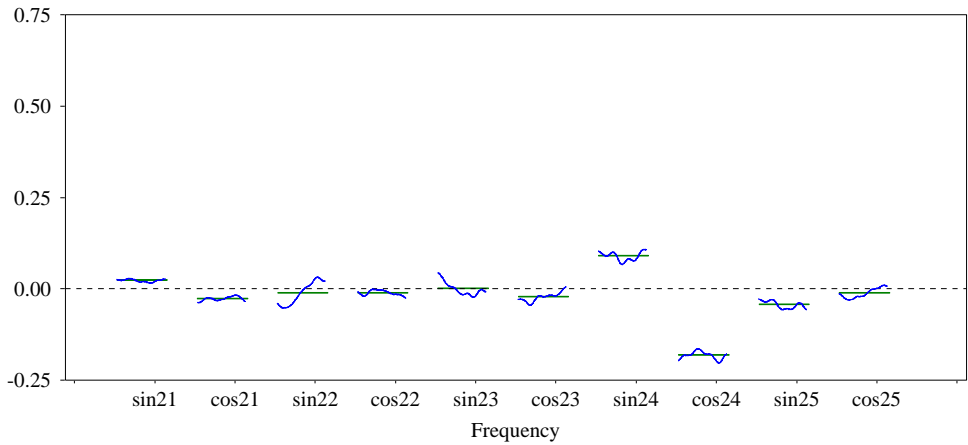
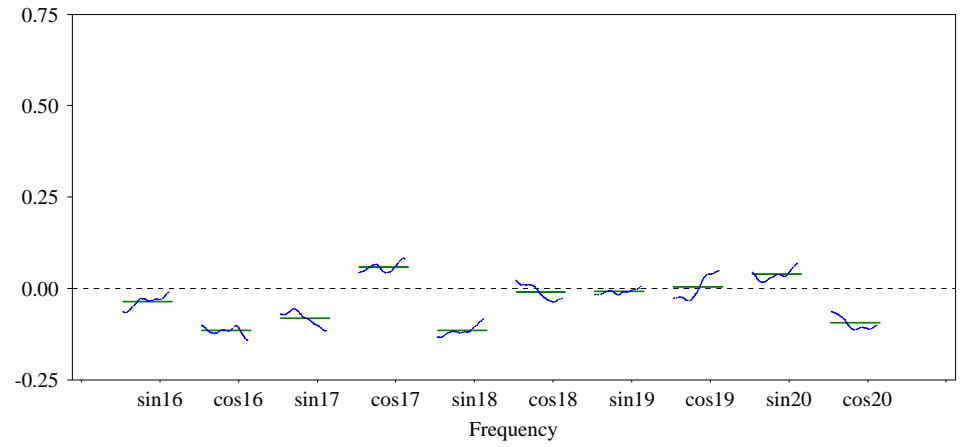
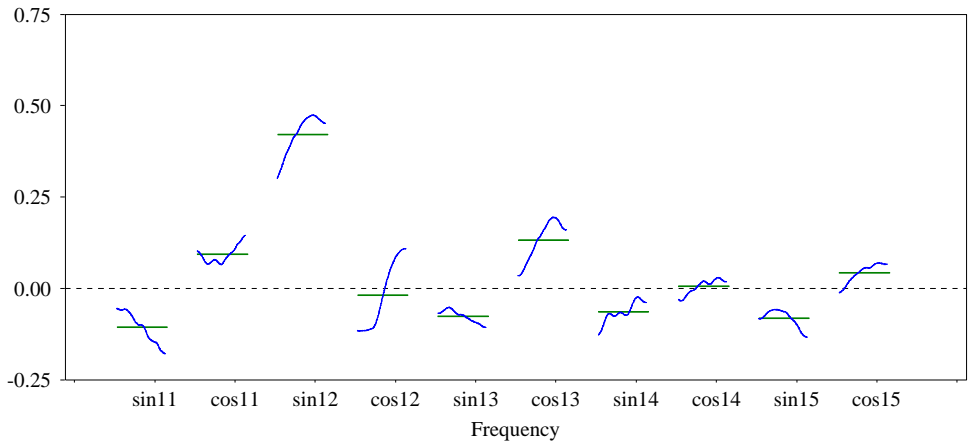
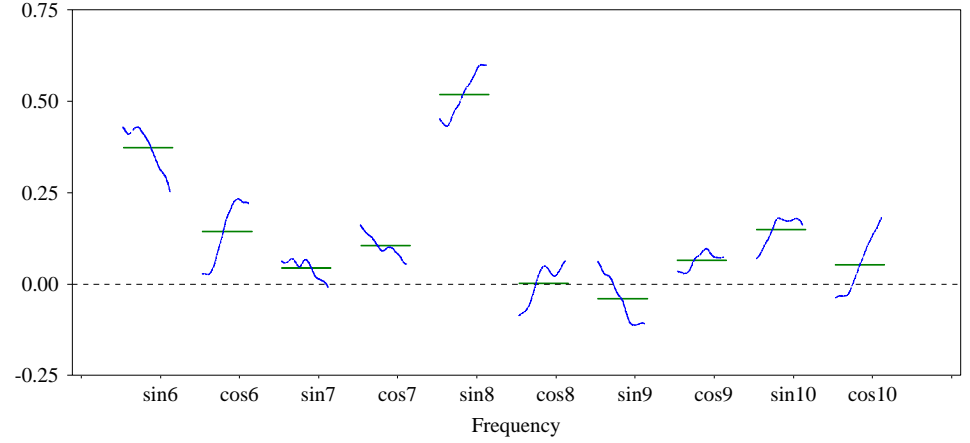
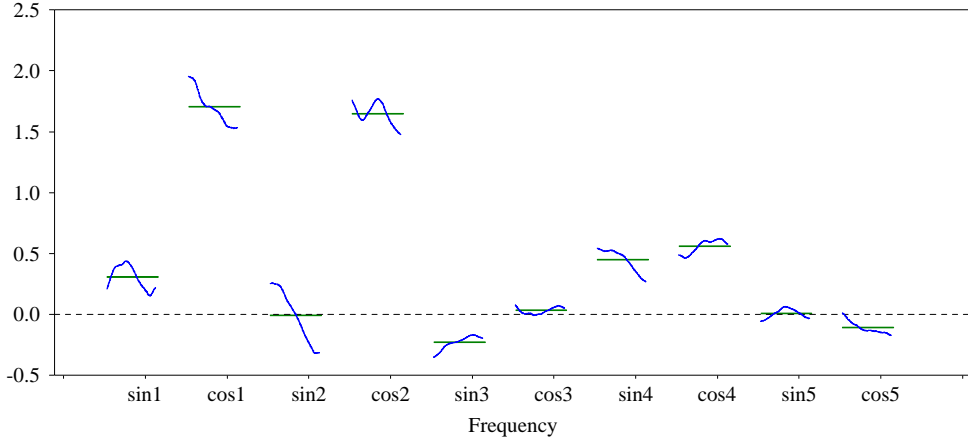


Figure 5.6. Effect on Seasonal Component of Outlier Treatment for Hurricane Katrina  
 Runs Ending in August 2005 (pre-Katrina) and January 2006 (post-Katrina)

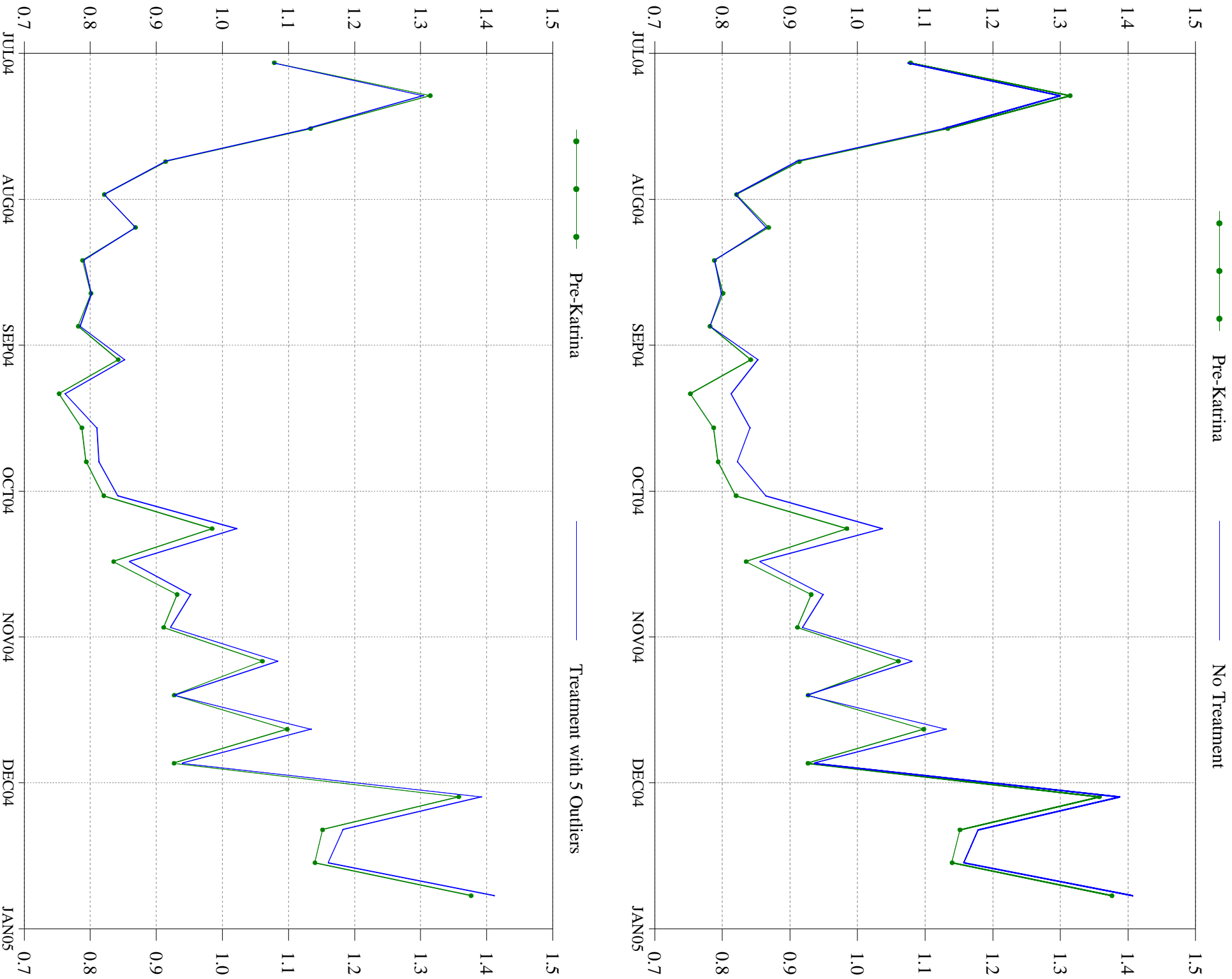


Figure 5.7. Preliminary and Revised Seasonally Adjusted Series

2007, Week 5 - 2008, Week 4

% Revision Statistics:

N = 48, Mean = 0.35, Abs Mean = 1.43

Median = 0.00, Abs Median = 1.26, Abs Max = 6.46

