# X12-ARIMA Supplement

*Contents*

## X11 — X11 Seasonal Adjustment

**X11** performs a seasonal adjustment on a series using the Census X11 method as revised for the Census Bureau's X12-ARIMA program. Most of the changes to this have been to the precision of the calculations. Many of the filters used in the original Census-X11 came from lookup tables, which in many cases rounded coefficients to just three or four significant digits. The new **X11** engine uses filters that are generated as needed to the full precision available. This also adds two new adjustment modes (log additive and pseudo-additive) and more flexible allowance for holiday effects.

The **X11** instruction is only available in the *Professional* version of RATS—the command will have no effect in the standard versions. Please contact Estima if you are interested in upgrading to the *Professional* version.

```
X11( options ) series start end
```

### Wizard

Use the *X11* wizard on the *Data* menu.

### Overview

**X11** can only be applied to monthly or quarterly series containing at least three years of data. Note that the X11 method will not necessarily work well with some types of series:

- It assumes that the seasonal component changes, if at all, in a fairly smooth fashion.
- It assumes that the seasonality is primarily a function of the calendar.

If, instead, the series shows a fairly sharp break in its seasonal pattern, or if the seasonal is more a function of some other variable (e.g. weather-driven electric demand), X11 is likely to leave some obvious seasonality in all, or at least part, of the data. One simple way to check how well X11 is doing on a new series is to run it first on the first 3/4 of the data, then on the last 3/4. Compare the overlapping range.

### Parameters

| | |
|---|---|
| *series* | Series to be seasonally adjusted. |
| *start end* | Range to use in **X11** seasonal adjustment. If you have not set a **SMPL**, this defaults to the largest range of *series*. |

Two additional parameters, *adjusted* and *factors*, used in earlier versions have been replaced by options of the same names. RATS will still recognize these parameters, but we recommend that you use the option in any new programs.

## Options

**mode=[multiplicative]/additive/pseudoadditive/logadditive**
**multiplicative/[nomult]** for compatibility with older versions of RATS)

The MODE option chooses which of the four adjustment modes to use. These are described later in the supplement. The older MULTIPLICATIVE option still works, but we would recommend correcting any of your programs to use the new option.

**print=[none]/short/standard/long/full**
**decimals=***number of decimals to show in output* **[depends upon data]**

PRINT controls the amount of output displayed by the **X11** instruction:

| | |
|---|---|
| NONE | This produces no output—used for production runs where you don't need to examine the output. |
| SHORT | This choice produces the minimum amount of printed output—only the initial series and final component estimates. |
| STANDARD | This is the standard level of X11 output—still primarily the final estimates. |
| LONG | This setting includes most of the important intermediate steps in addition to the standard output. |
| FULL | This reports on every step in the main adjustment process. |

DECIMALS controls the number of digits right of the decimal to display in any of the output. This has no effect on the level of precision of the calculations or the output series, just the printed output. By default, this depends upon the data.

**adjusted=***series for the seasonal adjusted data*
**factors=***series for the seasonal factors*

Use these if you want to save the seasonally adjusted series and/or the series of seasonal factors computed by **X11**. The seasonal factors include forecasts for three years beyond the adjustment period.

**[graduate]/nograduate**
**lower=***Lower limit standard errors from the mean* **[1.5]**
**upper=***Upper limit standard errors from the mean* **[2.5]**

Graduating extremes reduces the effect of outliers on the estimates of the seasonal factors. This only makes a great deal of difference if a series has a good deal of non-seasonal variability. The actual process used in graduating extreme values is fairly complicated, but basically, it drops data points which are more than "Upper Limit" standard errors from the mean (within a certain calculation), leaves unchanged those closer to the mean than "Lower Limit" standard errors and makes a smooth transition from inclusion to exclusion for those in between. The default values for the limits are 1.5 and 2.5.

**prefactors=***preliminary adjustment factors* **[not used]**

X12-ARIMA emphasizes the use of preliminary adjustment factors to take care of various types of outliers, rather than using the internal outlier detection engine (which is still present). These are usually estimated using **BOXJENK**, and created using the `ADJUST` option.

For multiplicative and log-additive adjustments, these should be in the form of factors, that is, 1.0 means no adjustment. If you are doing a log-additive adjustment starting with a **BOXJENK** model applied to the log of the series, you will have to transform the factors generated by **BOXJENK** from their additive form to the multiplicative form.

**extension=***series with out-of-sample forecasts of the input series*
**leads=***number of periods of extension*

The `EXTENSION` series is used in computing some of centered moving averages within the **X11** engine to reduce the bias in the end effects on the filters.

## Internal Regression Effects

These options control an internal regression of the irregular component on various dummy variables.

**tradeday=[none]/apply**

This allows you to apply a "trading day adjustment" for variation due to the number of Mondays, Tuesdays, etc., in a month. A typical series which would benefit from trading day adjustment is a total retail sales series, where there would be considerable predictable variation among the days of the week.

| | |
|---|---|
| `NONE` | The trading day option is not applied. |
| `APPLY` | Computes the trading day factors and applies them to the final adjustment. |

In previous versions of RATS, the following holidays were switch options. These are still supported, though it's recommended that any of these be done as preliminary factors instead. All of the holiday shifts are adjusted for long-run mean values, which prevents them from picking up a spurious trend effect for particular ranges of data.

**easter=***number of days before Easter at which effect is felt*

It's assumed that the level of activity is different for this number of days before Easter. This generates a dummy which splits this among February, March and April based upon the number of days falling in each month. The analogue to the old switch option is `EASTER=21`, though the calculation is now done differently.

**laborday=***number of days before Labor Day for the effect*

It's assumed that the level of activity is different for this number of days prior to Labor Day. This generates a dummy which splits this between August and September based upon the number of days falling in each month. The value for this that would be equivalent to the old `LABORDAY` switch is `LABORDAY=8`.

**thanksgiving=***number of days before Thanksgiving for the effect*

The value which gives the old adjustment is –1, that is the day after Thanksgiving. The level of activity is assumed to be different from this point on to December 24.

**critical=***critical value (t-stat) for outlier detection* **[default depends upon number of data points, typically around 3.8]**

The internal regression includes automatic detection and removal of additive outliers. This is to prevent contamination of the estimates of the calendar effects by very large outliers. Data points for which an additive dummy has a *t*-statistic bigger than the indicated limit are removed from the regression. These are added sequentially using a forward pass and possibly deleted using a backwards one.

## Examples

This seasonally adjusts DEUIP (German Industrial Production) by applying the multiplicative decomposition. The adjusted series and the seasonal factors are saved into DEUADJIP and DEUFACT, respectively.

```
calendar(m) 1961:12
allocate 1988:1
open data examx11.rat
data(format=rats) / deuip
x11(mode=mult,adjusted=deuadjip,factors=deufact) deuip
```

This uses the GMAUTOFIT procedure to automatically select a model for the log of the data, and then estimates the selected model using **BOXJENK**, with automatic detection of outliers. The adjustment series needs to be anti-logged before it can be used as a preliminary set of factors.

The value of FINAL at the end of the data is subtracted before the *exp* is taken so the adjustment will leave the end of data value at its observed level. This is done because the level shift and temporary change dummies are defined from t0 on, and so will give non-zero shift values to the end of the data, rather than the beginning. (The adjusted data will be the same either way; any printed output looks more natural with this correction). The series is adjusted using the log-additive model, with a full set of printed output.

```
@gmautofit(report,diffs=1) ldata
boxjenk(ar=%%autop,diffs=1,ma=%%autoq,sar=%%autops,$
   sdiffs=1,sma=%%autoqs,method=bfgs,outliers=standard,$
   adjustments=final) ldata
set prior = exp(final-final(2008:7))
x11(mode=logadd,prefactors=prior,print=full) u36cvs
```

## Seasonal Adjustment Methods

### X11 Seasonal Adjustment

The `X11` instruction in RATS uses the U.S. Census Bureau's X11 seasonal adjustment procedure to estimate and remove seasonal variations from a set of data, producing a "seasonally adjusted" series.

The primary purpose of seasonal adjustment is to allow you to examine non-seasonal trends in a set of data, which may be partially or completely masked by the seasonal variations. X11 is a signal extraction procedure, extracting the non-seasonal "signal" from the obscuring seasonal "noise."

It should be noted that X11 is not based upon an underlying statistical model for the seasonal process; instead it has evolved through many years of development and practice. It is designed to work best on series in which the primary source of seasonality is the calendar (as opposed to weather or something similar) and in which the seasonal changes smoothly, if at all, over the data.

Among the advantages of X11 are:

- It is widely used and the most generally accepted method of seasonal adjustment.
- The component "model" is a fairly natural one if we are approaching seasonal adjustment as the elimination of systematic "noise" components.
- Some potentially important and predictable effects, such as trading day variation in retail sales numbers, are hard to handle in other methods.

### Other Adjustment Procedures

There are, of course, other ways of dealing with seasonal variations. Unquestionably the simplest is a comparison with the same period in the previous year: "January sales are up 3% compared with January of last year." The great disadvantage, of course, is that it lacks immediacy—we can't see the underlying trend over the past few months.

Regression methods, such as dummy variables, can also be used, but they require that the seasonal variations be constant from year to year, or at least that they change in a fashion that can be explained with a relatively small number of parameters.

Time series methods, such as Box-Jenkins seasonal models, provide an explicit statistical model for the seasonal. They treat the seasonal as an integral part of the process, and are a good choice when you need forecasts of the unadjusted data. If, however, you are approaching your data with the "seasonal as noise" philosophy, they are not as appropriate.

See Bell and Hillmer (1984) for a survey of seasonal adjustment techniques.

## Two-Sided Averages

As a signal-noise extraction problem, X11 is “entitled” to bring the entire data series to bear when estimating the seasonally adjusted value at any point. It does this by using several types of centered moving averages to estimate the trend-cycle and seasonal components. Centered moving averages have the desirable statistical property that they do not induce any “phase shift” in a series—there is no tendency for the seasonally adjusted series to lead or lag the original data.

Adjustment (or smoothing methods, more generally) based upon two-sided averages have two significant disadvantages:

- They cannot easily be coaxed into providing forecasts of the input series, where a one-sided technique such as Box-Jenkins or exponential smoothing directly produces a forecasting model.
- The data near the beginning and end of a series have to be treated as special cases.

In some very simple ratio-to-moving average seasonal adjustment procedures, the second problem is treated by simply dropping data from each end. (Ratio-to-moving average estimates the seasonal factors by dividing the series by a centered moving average—the multiplicative variant of X11 is a big brother of these techniques).

Note, by the way, that no procedure (other than regression on dummies) is immune from problems such as these: a one-sided technique has a problem with the beginning of the data set. However, losing the *last* few data points usually defeats the whole purpose of seasonally adjusting the data, since we want to know what is happening *now*. Instead of dropping the points, X11 instead uses moving averages which become increasingly one-sided as the ends of the data are approached. This allows computation of seasonally adjusted data for the entire range of the data, but it means that as more data are obtained, not only will the last few seasonally adjusted values change due to the presence of new information, but also they will change because the actual calculations have a different form. The X12-ARIMA procedure, described below, is one proposed method of reducing the effect of this problem.

## X12-ARIMA

As you obtain more data, your estimates of the seasonally adjusted data will change (one hopes for the better), as there is additional information with which to both estimate the recent trend, and also to refine the estimate of the seasonal component (which will affect primarily the same months in previous years). However, since the seasonally adjusted values are the “data” on which you may be basing decisions, you would not like to see really large changes in your past data.

As noted above, the standard X11 procedure has one additional source of revision: the change in the form of the trend-cycle moving average for the most recent data. Some theoretical work has shown that this leads to larger than “optimal” revisions in the seasonally adjusted data. The X11-ARIMA procedure (see Dagum (1978)) was designed to reduce the effect of this source of revision error. Basically, it uses forecasts to fill out the data set so that the normal centered moving average can be used on all of the actual data points.

The current U.S. Census Bureau software is dubbed X12-ARIMA. This includes an enhanced variant on the original X11 algorithm, combined with an ARIMA modelling procedure for producing forecasts of the seasonal data and/or isolating systematic components of the "irregular" part of the process. Beginning with version 7.2, RATS has included most of the new capabilities of this in the combination of the **X11** and **BOXJENK** instructions.

**BOXJENK** handles any ARIMA modelling that is required. It now has new `GLS` and `OUTLIER` options for estimating "RegARIMA" models and locating outliers within the raw series. You can use the `ADJUSTMENTS` option on **BOXJENK** to produce any preliminary regression factors, and you can use a **UFORECAST** instruction to generate forecasts for extending the data. By separating analysis into the two instructions, you can use some other type of model (such as a state-space model estimated with **DLM**) to generate the forecasts.

If you create an adjustment series with **BOXJENK**, you input that into **X11** using the `PREFACTORS` option. If you would like to use forecasts to extend the series, use the `EXTENSIONS` and `LEADS` options on **X11**.

# Details of the X11 Seasonal Adjustment

## The Basics of X11

This section of the manual presents more detailed information on the X11 seasonal adjustment procedure. While Census X12-ARIMA has many new features, at its heart is something very similar to the original X11 algorithm as described in Technical Paper Number 15 of the Bureau of the Census.

The procedure can only be used with series which have monthly or quarterly frequencies, and requires a minimum amount of data to function correctly. You must have at least three years of data (36 months or 12 quarters), or the adjustment will not be done at all. Although the **X11** instruction will work as long as this requirement is met, using data that contains fewer than ten years of observations will produce results that are unreliable at best. (Consider that with six years of data, there are only six Januaries). For the same reasons that any pattern becomes more apparent the more times it is repeated, the X11 procedure's ability to distinguish seasonal variations from other fluctuations in the data becomes more accurate with more data.

To determine the seasonal pattern in a set of data, X11 requires that the other factors which affect the data be separated from the seasonal influences. The nonseasonal components which influence economic time series include the trend-cycle ($C$) and the irregular ($I$).

- The Trend-Cycle component is the primary "signal" revealed in the seasonally adjusted series, from which the seasonal "noise" has been removed. The trend-cycle includes the effects of any long term overall trends and the business cycle.
- The Irregular component contains the residual variations in a series, which cannot be attributed to Seasonal or Trend-Cycle variations. The Irregular will include the influences of major non-repeating events which may affect the data (such as droughts, political events, etc.), as well as measurement errors.

The basic assumption behind the seasonal adjustment process is that the series being adjusted ($X$) is the result of some combination of these non-seasonal factors and the seasonal factor ($S$). By repeatedly estimating and extracting (by division or subtraction) these components in various combinations, **X11** eventually produces a final seasonally adjusted series (*SA*) which consists only of the trend-cycle and irregular components of the original series. The following is the decomposition for the four adjustment mode:

| Mode | Model for X | Model for SA |
|---|---|---|
| Multiplicative | $X = C \times S \times I$ | $SA = C \times I$ |
| Additive | $X = C + S + I$ | $SA = C + I$ |
| Pseudo-Additive | $X = C \times (S + I - 1)$ | $SA = C \times I$ |
| Log-Additive | $\log X = C + S + I$ | $SA = \exp(C + I)$ |

Multiplicative and log-additive are the two most common types of adjustment. Additive is the only one which can be used on series which have non-positive values. Pseudo-additive will generally be similar to multiplicative, since $S \times I \approx (S-1)+I$ if $S$ is near 1. However, if some months have very small seasonal factors (say, less than .7), adjustment with the pseudo-additive mode can be superior. This can occur when a few months have predictably very small values (due to vacations or weather).

In additive mode, all the components are in the same units as the data. For multiplicative and pseudo-additive, $C$ is in the natural units of the data, while $S$ and $I$ are ratios centered about 1. In any output, they are scaled by 100.

In log-additive, the calculations are done additively on the log scale as shown in the table. However, the output shows $C$ in the natural units of the data, and $S$ and $I$ as ratios scaled by 100. (All components are anti-logged before being shown). As a result, its output will look very similar to the multiplicative mode. The final estimate of $C$ is "bias-corrected" to take into account the effects of anti-logging a series subject to estimation error.

## The X11 Options

The **X11** instruction offers you several options when doing seasonal adjustments, as shown on pages 2–5. You specify which mode (multiplicative, additive, pseudo-additive or log-additive ) will be used for the adjustment. You can choose the cutoffs for handling extreme values, and how much detail will be included in the printed output. There are also options (described later) for dealing with systematic "calendar" effects (holidays and trading day effects).

More information on using these operations and on how they are handled by the procedure can be found following the description of the X11 procedure.

### Output Retrieval Options (FACTORS, ADJUSTED)

The `FACTORS` and `ADJUSTED` options allow you to retrieve the final estimates of the seasonal factors ($S$) and the final estimates of the seasonally adjusted data ($SA$). The seasonal factors are in ratio form (centered about 1) for all modes except additive.

### MODE

A multiplicative or log-additive decomposition is usually in order for any data which includes an overall trend as the seasonal variations in the data are likely to be proportional to the level of the data. Additive decomposition is for situations in which seasonal variations are independent of the level of the data.

### Graduation Limits (GRADUATE,UPPER,LOWER)

Because X11 makes extensive use of averages of a relatively small number of data points (to make it possible for the series to drift over the data set), extreme values in the data can throw off estimates unless they are dealt with in some way. Rather than simply discarding outliers above a single point, RATS uses a graduated method in which values slightly below the exclusion cutoff are included but not given full weight. There are thus two limits: above the upper limit, the observation is dropped completely, below the lower limit it is given full weight, in between it is given partial weight. The default values for these options will generally be fine.

### PREFACTORS

The `PREFACTORS` option allows you to pull known parts of the irregular factors out of the data right from the start. These are typically generated out of a **`BOXJENK`** instruction with the `ADJUSTMENTS` option. Note that these might need to be transformed to be compatible with the form of the irregular component for the adjustment mode that you've selected. For instance, if you run a **`BOXJENK`** on the log of the data, you'll have to transform the adjustment series by `EXP`'ing in order to use the factors with either the multiplicative or the log-additive adjustment mode. Note, by the way, that you can't use this with the pseudo-additive mode.

### EXTENSION and LEADS

The *`series`* parameter on **`X11`** gives the actual data to be adjusted. If you want to extend that with forecasts to permit better estimates of the trend-cycle near the end of the data, you input those forecasts with the `EXTENSIONS` option. Use `LEADS` to indicate how many forecasted data points are to be used.

### PRINT and DECIMALS

The `PRINT` option allows you to select how much output will be produced by the seasonal adjustment operation. There are five levels (including `NONE`, which produces no output). At the end of this chapter is a list of all the data tables that can be produced by the **`X11`** instruction. This list will tell you which tables will be printed for a selected level of output.

The `DECIMALS` option allows you to control how many decimals are shown in the output. By default, RATS decides this based upon the range of values in the series, showing a deciding how many digits to show.

## The X11 Algorithm

To produce the seasonally adjusted series, X11 must be able to estimate the trend-cycle component. However, to estimate from the original data a trend-cycle estimate which is not contaminated with the seasonal requires a rather wide moving average. X11 begins with this rather rough estimate of the $C$ component, using it to generate preliminary estimates of the other components. This allows computation of an approximate seasonally adjusted series, and a more sophisticated method of determining the trend-cycle can be applied to that. This in turn allows more accurate estimates of the other factors, and of the seasonally adjusted series.

This is repeated three times. The complexity of the process is due to concern with contamination of the estimates of the trend-cycle and seasonal by extreme irregulars (outliers). The seasonal factors, in particular, can be strongly influenced by these since they are done with relatively narrow filters. What ends up happening is that (for a multiplicative decomposition), the data can be represented by

$$X = C \times S \times (I_N \times I_O)$$

where the irregular component is broken into a part consisting of identifiable outliers $I_O$ and a part that are just normal irregular ($I_N$). $I_O$ will be 1.0 for most data points. The calculation of the $C$ and $S$ components is done with the outlier-adjusted series

$$X_N = X / I_O$$

It's estimation and refinement of the $I_O$ part that requires multiple passes.

We will go through this procedure step by step, using a multiplicative decomposition for a monthly series as an example. The basic mathematics of the process will be covered fairly explicitly, but without going too deeply into topics such as the theory and detailed calculations of moving averages.

The four steps of the process are labeled parts A, B, C, and D. Part B is described in a fairly detailed manner, while parts C and D are covered more briefly, except where they differ from part B. The designation of each table produced by **`X11`** is indicated in parentheses at the point it is generated.

## X11 Part A

Call the original series $X_A$. If there are preliminary factors (`PREFACTORS` option), those are extracted now. The preliminary factors are the first attempt to eliminate outliers, or any other identifiable part of the irregular component. The input to the next part is now $X_B = X_A / PF$. Otherwise, $X_B = X_A$.

## X11 Part B

### Compute SI Ratios

With a multiplicative decomposition, the series $X_B$ is assumed to be represented by

(B1) $X_B = C \times S \times I$

The first step is to compute an estimate of the trend-cycle component using a 12 term moving average $X_B$ for (table B2). This estimate for the trend-cycle (we'll label it $C_1$) is divided into $X_B$ to eliminate the $C$ term from the decomposition. This results in an unmodified *SI* (Seasonal–Irregular) ratio:

(B3) $X_B / C_1 = \left(C \times S \times I\right) / C_1 = \left(S \times I\right)_1$

### Replace Extremes

The next step is to treat the extreme values of these *SI* ratios. Preliminary seasonal factors ($S_1$) are computed by using a five-term moving averages of like months. This estimate of the seasonal component is then divided into the *SI* ratios to give a preliminary estimate of the irregular component:

$$(S \times I)_1 \,/\, S_1 = I_1$$

Because outliers in the irregular component may result in inflated values of the final seasonal factors, extreme values are removed or weighted using the graduated extremes method (see page 16). *SI* ratios corresponding to weighted irregular values (those which did not receive full weight) are replaced using an average of the ratio multiplied by its assigned weight and the nearest four ratios in the data that received full weights. (These replacement values appear in table B4).

### Preliminary Seasonal Adjustment

Another weighted five-term average is applied to the series of now modified *SI* ratios each month separately to obtain a new estimate of the seasonal factors, $S_2$ (table B5). These factors are then divided into $X_B$ to produce a preliminary seasonally adjusted series:

(B6) $X_B / S_2 = \left(C \times S \times I\right) / S_2 = \left(C \times I\right)_2$

### Trend-cycle and Unmodified SI Ratios

At this point, the variable trend-cycle routine (described later) is applies to $\left(C \times I\right)_2$ to estimate the trend-cycle. This is table B7, what we'll call $C_2$.

Once again, the working series is divided by trend-cycle values to produce unmodified *SI* ratios; however, the updated $C_2$ values are used this time:

(B8) $X_B / C_2 = \left(C \times S \times I\right) / C_2 = \left(S \times I\right)_2$

### Estimating the Irregular Component

A new set of seasonal factors, which we will label $S_3$, is produced (by a five-term moving average) from these *SI* ratios. These factors are divided into the *SI* ratios to produce new estimates of the irregular component:

$$(S \times I)_2 / S_3 = I_3$$

The extreme values of the irregular component are treated using graduated extremes, and corresponding values of the *SI* ratios are replaced, just as before (B9).

### Second Preliminary Adjusted Series

A five-term moving average is again applied, this time to the modified *SI* ratios, to estimate a new set of seasonal factors, $S_4$ (B10). These factors are then divided into the working series to produce a new preliminary seasonally adjusted series:

(11) $$X_B / S_4 = (C \times S \times I) / S_4 = (C \times I)_4$$

A preliminary irregular series is then obtained by dividing the seasonally adjusted series by $C_2$:

(B3) $$(C \times I)_4 / C_2 = I_4$$

The method of graduated extremes is once again employed to handle extreme values of $I_4$. Keep in mind that $I_4$ has been computed from *SI* ratios that had already been modified once for extreme values, making the treatment of extremes an iterative process. Table B17 reproduces the standard deviations and the preliminary weights for the irregular component that is determined in this process.

### Calculation of Outlier Weights

The only thing that will be retained from Part B is its identification of extreme values. The series for adjustment for identifiable outliers is defined by

(20) $$I_{OB} = \frac{I_4}{(1-w) + wI_4}$$

where the $w$ are the final weights from the iteration. This is (as expected) 1.0 for data points given full weight, and is $I_4$ for any receiving zero weight and somewhere in between for partially down-weighted data.

## X11 Part C

### Modifying the Original Series

The working series for part C is $X_B$ modified for the outliers identified in part B.

(C1) $X_C = X_B / I_{OB}$

### Repeating the Procedure

The same steps used in part B are then applied, but they are applied to $X_C$. However, seven-term (rather than five-term) moving averages are used in computing the seasonal factors and the early step for identifying outliers is skipped (since we're already using a series modified for outlies).

$X_C$ is used up until the seasonal factors $S_4$ (table C10) have been computed. At that point, we switch back to using $X_B$, that is, the new estimate of the seasonally adjusted series is

(11) $X_B / S_4 = (C \times S \times I) / S_4 = (C \times I)_4$

This includes the previously identified extreme and near-extreme values. However, we have improved estimates of the other components, which we use to recompute the outlier weights, which we'll call $I_{OC}$ (table C20). As with part B, this is the only information retained from this pass.

## X11 Part D

As before, the original series is modified by removing the irregular outlier component identified in Part C. The working series is thus:

(D1) $X_D = X_B / I_{OC}$

The steps of part B (as modified in part C) are repeated using the new series through the estimate of the final set of seasonal factors (series $S_f$, table D10). These factors are then divided into the *original* series (before any prior adjustments) to produce the final seasonally adjusted series:

(11) $SA \equiv X_A / S_f = (C \times S \times I) / S_f = (C \times I)_f$

The final trend-cycle component ($C_f$, table D12) is computed by applying the variable trend-cycle routine to the final *modified* seasonally adjusted data:

$$X_D / S_f = (C \times I)_4$$

A final irregular component is computed by dividing the final seasonally adjusted series by the final trend-cycle:

(13) $(C \times I)_f / C_f = I_f$

## Variable Trend-Cycle Estimation

This first computes an estimate of $C$ by applying a 13-term Henderson moving average to the $CI$ series. This is divided into the series to get a preliminary estimate of the irregular component. Next, the ratio of the average month-to-month percent changes for both the $I$ and $C$ components is computed.

This $I/C$ ratio gives an indication of how important the irregular variations are compared to the overall trend-cycle. RATS recomputes the $C$ estimate, this time using a 9, 13, or 23 term average depending on the values of the $I/C$ ratio. Small values indicate that the irregular is not very significant compared to the trend-cycle, and so a shorter average can be used. Larger values of the ratio indicate a more significant irregular component, which means a longer average has to be used to estimate the trend-cycle accurately.

## Graduated Treatment of Extremes

In each pass of the adjustment procedure, RATS computes an $SI$ (Seasonal–Irregular) ratio (or difference, if using an additive decomposition). The $SI$ ratio is an estimate of the combination of the seasonal component with the irregular. Since extreme outlying values can throw off the seasonal calculations, these extreme values must be treated in some manner to remove their effects from the calculations. Rather than simply removing values which deviate from the mean beyond an arbitrary cut-off value, the X11 procedure uses a graduated treatment of extremes. This procedure also allows the user some flexibility in determining the limits that will be used to determine which values are "extreme."

The graduated treatment of extremes is a two step process. Moving five-year standard deviations ($SD$) are computed for each value of the irregular component being checked for extremes. Values of the series are compared to the standard deviation using specified upper and lower limits ($UL$ and $LL$).

First, any value greater than the product $\left(UL \times SD\right)$ is removed. The standard deviation is recomputed. Second, the values are compared to the (new) standard deviation, and weights are assigned to the data based on this comparison. Values which are greater than $\left(UL \times SD\right)$ are assigned a weight of zero. Values between $\left(LL \times SD\right)$ and $\left(UL \times SD\right)$ are given partial weights, ranging from zero at the upper limit to full weight at the lower limit. Values below the lower limit also receive full weight. These weights are then used in computing replacement values for the $SI$ ratios from which the irregular component was derived.

The default values are 2.5 for the upper limit and 1.5 for the lower limit. You can enter different values using the `LOWER` and `UPPER` options. You may want to use higher limits if you are working with a very smooth series, or lower limits for a highly irregular series.

## Tests for Stable Seasonality

These are done during Part D. The final *unmodified SI* ratios are computed. The statistical definition of stable seasonality is based on:

$$SI_{it} = SI_i^* + u_{it}$$

where $SI_{it}$ is the estimated *SI* ratio for month $i$, year $t$, $SI_i^*$ is the mean value of *SI*'s for month $i$, and $u_{it}$ is a residual. Two tests are done. The first is a standard analysis-of-variance test for equality among the $SI_i^*$. The second is a Kruskal-Wallis non-parametric test for the same thing. Rejection of the null hypothesis, (i.e., a high $F$-statistic or low significance level) would be the comforting result.

There are several things to note about this test:

- It is *not* a test of whether or not X11 has worked adequately. If you seasonally adjust a non-seasonal series, you should get a very low $F$-statistic. You will only get a high $F$-statistic if there is seasonality and X11 estimates systematic differences among the months.
- If moving seasonality is present in the series (which is permitted under X11), the test may quite properly reject stable seasonality, even though a distinct seasonal pattern is apparent.
- If the variance in the irregular is very large compared to the variance of the seasonal, the test may have extremely low power. In such situations, X11 (or any other seasonal adjustment procedure) will not be particularly reliable.

## Moving Seasonality Test

This is also done during Part D, and also applied to the final unmodified *SI* ratios. In

$$\left|SI_{it} - 1\right| = \alpha + \mu_t + \lambda_i + u_{it}$$

where $\mu_t$ is the year $t$ effect and $\lambda_i$ is the month $i$ effect, the hypothesis of interest is that the $\mu_t$ are equal. (This is a two-way analysis of variance test). If there is moving seasonality, the $\lambda_i$ *aren't* constant (which is a maintained hypothesis); instead, some months tend to have seasonal factors which increase over time and other months (necessarily) decrease. This should induce a significant difference among the $\mu_t$, and thus a significant test statistic.

## Regression (Trading Day/Holiday Effects)

In addition to the externally specified preliminary factors (`PREFACTORS` option), X11 allows for specialized removal of systematic components from the irregular component using regression. This is done in parts B and C and applied to the $I_4$ estimate of the irregular.

For holiday effects only, a regression of the form

$$I_t - 1 = Z_t \beta + u_t$$

is run, where the $Z_t$ depend upon the options that you choose. The estimated regression factors ($I_R$) are generated by

$$I_{R,t} = 1 + Z_t \hat{\beta}$$

Note that this *could* be negative, though that's highly unlikely with the types of variables being used.

Because the irregular being used isn't modified for outliers already, the regression routine includes automatic outlier detection. From the initial regression, a robust estimate of the standard error of residuals is computed (based upon the median absolute residual). Tests are run on each data point to see whether an additive outlier at that point is statistically significant. The critical value for the *t*-statistic is controlled by the `CRITICAL` option on **`X11`**. If any data point has a significant outlier test, the one with the highest value is added to the regressor set and the regression is re-estimated. This process is repeated until no further outliers are detected.

If any outliers have been detected, a backwards pruning process is started. The standard regression *t*-statistics on the outliers are checked. If any of those are *now* below the critical value, the least significant one is removed. The regression is re-estimated with the smaller set of outliers, and the pruning process is repeated until all remaining outliers are significant.

Note that the outliers are *not* part of the fitted values used in creating $I_R$. They are included only to prevent their contaminating the estimates of the effects of interest.

The irregular adjustment weights are then computed using $I_5 = I_4 / I_R$, and the combined adjustment series for the start of the next part is

$$I_O = \frac{I_5}{(1-w) + wI_5} \times I_R$$

This is shown in tables B20 and C20.

### Holiday Adjustment

The RATS **X11** instruction allows for inclusion of effects for Easter, the U.S. Labor Day and Thanksgiving holidays. Although the fluctuations caused by the variable dates of these holidays are not very significant for most series, they can make a difference with certain types of data. For example, if Thanksgiving occurs relatively early in November, more of the Christmas season for retail sales will be in November than in a year in which Thanksgiving comes late.

Labor Day will affect data for August and September, Thanksgiving will affect November and December, and Easter affects March, April and (possibly) February.

**LABORDAY=***offset*

The offset is the number of days before Labor Day that are assumed to be affected. The $Z_t$ for this is the proportion of the affected days for a given year falling in a month, less the average number over a full 28 year calendar cycle.

Take, for instance, the year 2009. Labor Day is September 7. If *offset*=8, 6 of the covered days will be in September and 2 in August. The proportions will thus be .25 for August and .75 for September. For that value of *offset*, the average proportion of covered days is .6875 for August and .3125 for September. For 2009, the $Z_t$ are 0 for months other than August and September, $0.25-0.6875=-0.4375$ for August and .4375 for September.

**THANKSGIVING=***offset*

The offset is the number of days before Thanksgiving (use negative for days after) at which the Thanksgiving to Christmas effect is felt. The $Z_t$ is the proportion of the days from Thanksgiving–*offset* to December 24 that fall in a month in a particular year, again, less the average proportion over a 28 year calendar cycle.

**EASTER=***offset*

The *offset* is the number of days before Easter at which the Easter effect is felt. The $Z_t$ is the proportion of those days in each month less the long-run average of the proportion.

### Trading Day Adjustment

RATS includes the option of adjusting the series for trading day variations. Trading day variations are due to the fact that a given day of the week may occur more times in one month than in another month. Such variation could be important for a series as retail sales for a business that only operates Monday through Friday. Sales figures would be expected to be lower for a month with five weekends than for the same month with only four. Note that variation among months due simply to different numbers of days (November vs December, for instance) is handled just fine by the standard adjustment process.

For trading day effects, there are six regressors (one for each day except Sunday). The Monday dummy is the difference between the number of Mondays and the number of Sundays in a given month, similarly for the other days of the week. However, we have to make an adjustment to the dependent variable. The seasonal factors that have already been removed have taken with them the "length of month" effect. We have to put that back in. Thus, the regression run is

$$M_t^* I_t - M_t = Z_t \beta + u_t$$

where $M_t$ is the number of days in the month, and $M_t^*$ is the average number of days in that month (that is, 28.25 for February, $M_t$ for all other months). $I_t$ is multiplied by the average for February because the seasonal factor averages both leap years and non-leap years, so the length of month factor that's been removed with the seasonal factor depends upon the average number of days.

After running the regression as described above, the regression factor is computed as

$$I_{R,t} = \frac{M_t + Z_t \hat{\beta}}{M_t^*}$$

When the absolute month-to-month changes in the irregular component of the data are less than about 8 per cent, this procedure will produce estimates which are useful in the removal of trading day variations. It will not be very successful with data which is more irregular.

## Extension with Forecasts

The `EXTENSION` option provides a series of forecasts to extend the raw data. The extension series has to be in the same terms as the actual data. If you forecast the log series, you need to anti-log it before providing it to **`X11`**.

All the calculations described earlier are done with the original series appended with the extension. However, reports and the outputs from the `ADJUSTED` and `FACTORS` options are restricted to the range of the actual data.

## Quarterly Data

When doing a seasonal adjustments with quarterly data, X11 proceeds exactly as described above, except for a few differences in the ways that averages and weights are computed:

- For tables B2, C2, and D2, a four-term moving average is used instead of a 12-term average to estimate the trend-cycle.
- For tables B7, C7, D7, and D12, a 5-term Henderson average is used initially, and a 5 or 9 is used in the recalculation.

## Other Modes (Additive, Log-Additive, Pseudo-Additive)

### MODE=ADDITIVE

In working with components, division is replaced with subtraction and multiplication with addition. The outlier adjustment series at the end of Parts B and C is

$$I_O = (1-w)I_4$$

Regression effects are estimated using $I_t = Z_t\beta + u_t$ and trading day regression length of month adjustment is handled by adding an extra "leap year February" dummy to regressor list.

### MODE=LOGADDITIVE

For the most part, this is an additive decomposition applied to the log of input series. However, all components are anti-logged before being displayed or output. The final trend cycle estimate is "bias-adjusted" since the mean of the log isn't the same as the log of the mean. The trading day length of month adjustment is handled by running the regression in the form:

$$m_t^* I_t + m_t^* - m_t = Z_t\beta + u_t$$

### MODE=PSEUDOADDITIVE

The changes to the formulas are too numerous to list, since the formula for extracting a component will depend upon both the source and the target. For instance, taking the cycle out of the data to create the *SI* component requires division, while taking the seasonal out of the *SI* component requires subtraction.

## Print Levels Chart

The following is a list of the data tables produced by the **X11** instruction. The number in parentheses indicates which print level (selected via the PRINT option) results in a particular table being printed out. PRINT=SHORT is print level 1, PRINT=STANDARD is level 2, PRINT=LONG is level 3, and PRINT=FULL is level 4, while PRINT=NONE (the default setting) is level zero.

For example, table B3 will only be printed if you use PRINT=FULL, while table B1 will be printed for any setting other than PRINT=NONE.

B1. (Prior Adjusted) Original Series (1,2,3,4)
B2. Preliminary trend cycle, B iteration (3,4)
B3. Unmodified SI ratios, B iteration (4)
B4. Replacement values for extremes of SI component (4)
B5. Preliminary seasonal factors, B iteration (4)
B6. Preliminary seasonally adjusted series, B iteration (4)
B7. Preliminary trend cycle, B iteration (4)
B8. Unmodified SI ratios, B iteration (4)
B9. Replacement values for extremes of SI component (4)
B10. Seasonal factors, B iteration (3,4)
B11. Seasonally adjusted data, B iteration (4)
B13. Irregular component, B iteration (3,4)
B16. Adjustment Factors Derived From Regression Coefficients (3,4)
B17. Preliminary Weights for Irregular Component (3,4)
B19. Adjusted Original Series (4)
B20. Preliminary extreme value adjustment factors (4)
C1. Modified original data, C iteration(3,4)
C2. Preliminary trend cycle, C iteration (4)
C4. Modified SI ratios, C iteration(4)
C5. Preliminary seasonal factors, C iteration (4)
C6. Preliminary seasonally adjusted series, C iteration(4)
C7. Preliminary trend cycle, C iteration (3,4)
C9. Modified SI Ratios (4)
C10. Seasonal factors, C iteration (3,4)
C11. Seasonally adjusted data, C iteration (4)
C13. Irregular component, C iteration (3,4)
C16. Adjustment Factors Derived From Regression Coefficients (1,2,3,4)
C17. Final Weights for Irregular Component (1,2,3,4)
C19. Adjusted Original Series (4)
C20. Preliminary extreme value adjustment factors (4)
D1. Modified original data, D iteration (3,4)
D2. Preliminary trend cycle, D iteration (4)
D4. Modified SI ratios, D iteration (4)
D5. Preliminary seasonal factors, D iteration (4)
D6. Preliminary seasonally adjusted series, D iteration (4)

D7. Preliminary trend cycle, D iteration (3,4)
D8. Final unmodified SI ratios (1,2,3,4)
D9. Final replacement values for SI ratios (1,2,3,4)
D10. Final seasonal factors (1,2,3,4)
D11. Final seasonally adjusted data (1,2,3,4)
D12. Final trend cycle (2,3,4)
D13. Final irregular component (2,3,4)
D20. Final Combined Adjustment Factors (2,3,4)

## Bibliography

Bell, W. R., and S. C. Hillmer (1984). “Issues Involved with the Seasonal Adjustment of Economic Time Series.” *Journal of Business and Economic Statistics*, Vol. 2, pp. 291-319.

Bureau of the Census. *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. Technical Paper No. 15.

Dagum, E. B. (1978). “Modelling, Forecasting, and Seasonally Adjusting Economic Time Series with the X-11 ARIMA Method,” *The Statistician*, Vol. 27, pp. 203-216.