# Chapter 43
## The X11 Procedure

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview: X11 Procedure</td>
<td>3086</td>
</tr>
<tr>
<td>Getting Started: X11 Procedure</td>
<td>3086</td>
</tr>
<tr>
<td>Basic Seasonal Adjustment</td>
<td>3087</td>
</tr>
<tr>
<td>X-11-ARIMA</td>
<td>3090</td>
</tr>
<tr>
<td>Syntax: X11 Procedure</td>
<td>3091</td>
</tr>
<tr>
<td>Functional Summary</td>
<td>3092</td>
</tr>
<tr>
<td>PROC X11 Statement</td>
<td>3093</td>
</tr>
<tr>
<td>ARIMA Statement</td>
<td>3094</td>
</tr>
<tr>
<td>BY Statement</td>
<td>3097</td>
</tr>
<tr>
<td>ID Statement</td>
<td>3097</td>
</tr>
<tr>
<td>MACURVES Statement</td>
<td>3097</td>
</tr>
<tr>
<td>MONTHLY Statement</td>
<td>3097</td>
</tr>
<tr>
<td>OUTPUT Statement</td>
<td>3101</td>
</tr>
<tr>
<td>PDWEIGHTS Statement</td>
<td>3102</td>
</tr>
<tr>
<td>QUARTERLY Statement</td>
<td>3103</td>
</tr>
<tr>
<td>SSPAN Statement</td>
<td>3105</td>
</tr>
<tr>
<td>TABLES Statement</td>
<td>3106</td>
</tr>
<tr>
<td>VAR Statement</td>
<td>3106</td>
</tr>
<tr>
<td>Details: X11 Procedure</td>
<td>3106</td>
</tr>
<tr>
<td>Historical Development of X-11</td>
<td>3106</td>
</tr>
<tr>
<td>Implementation of the X-11 Seasonal Adjustment Method</td>
<td>3108</td>
</tr>
<tr>
<td>Computational Details for Sliding Spans Analysis</td>
<td>3111</td>
</tr>
<tr>
<td>Data Requirements</td>
<td>3114</td>
</tr>
<tr>
<td>Missing Values</td>
<td>3114</td>
</tr>
<tr>
<td>Prior Daily Weights and Trading-Day Regression</td>
<td>3114</td>
</tr>
<tr>
<td>Adjustment for Prior Factors</td>
<td>3115</td>
</tr>
<tr>
<td>The YRAHEADOUT Option</td>
<td>3116</td>
</tr>
<tr>
<td>Effect of Backcast and Forecast Length</td>
<td>3116</td>
</tr>
<tr>
<td>Details of Model Selection</td>
<td>3117</td>
</tr>
<tr>
<td>OUT= Data Set</td>
<td>3119</td>
</tr>
<tr>
<td>The OUTSPAN= Data Set</td>
<td>3120</td>
</tr>
<tr>
<td>OUTSTB= Data Set</td>
<td>3120</td>
</tr>
<tr>
<td>OUTTDR= Data Set</td>
<td>3121</td>
</tr>
<tr>
<td>Printed Output</td>
<td>3122</td>
</tr>
<tr>
<td>ODS Table Names</td>
<td>3132</td>
</tr>
<tr>
<td>Examples: X11 Procedure</td>
<td>3137</td>
</tr>
</tbody>
</table>
Overview: X11 Procedure

The X11 procedure, an adaptation of the U.S. Bureau of the Census X-11 Seasonal Adjustment program, seasonally adjusts monthly or quarterly time series. The procedure makes additive or multiplicative adjustments and creates an output data set containing the adjusted time series and intermediate calculations.

The X11 procedure also provides the X-11-ARIMA method developed by Statistics Canada. This method fits an ARIMA model to the original series, then uses the model forecast to extend the original series. This extended series is then seasonally adjusted by the standard X-11 seasonal adjustment method. The extension of the series improves the estimation of the seasonal factors and reduces revisions to the seasonally adjusted series as new data become available.

The X11 procedure incorporates sliding spans analysis. This type of analysis provides a diagnostic for determining the suitability of seasonal adjustment for an economic series.

Seasonal adjustment of a series is based on the assumption that seasonal fluctuations can be measured in the original series, \( O_t, t = 1, \ldots, n \), and separated from trend cycle, trading-day, and irregular fluctuations. The seasonal component of this time series, \( S_t \), is defined as the intrayear variation that is repeated constantly or in an evolving fashion from year to year. The trend cycle component, \( C_t \), includes variation due to the long-term trend, the business cycle, and other long-term cyclical factors. The trading-day component, \( D_t \), is the variation that can be attributed to the composition of the calendar. The irregular component, \( I_t \), is the residual variation. Many economic time series are related in a multiplicative fashion (\( O_t = S_t C_t D_t I_t \)). A seasonally adjusted time series, \( C_t I_t \), consists of only the trend cycle and irregular components.

Getting Started: X11 Procedure

The most common use of the X11 procedure is to produce a seasonally adjusted series. Eliminating the seasonal component from an economic series facilitates comparison among consecutive months or quarters. A plot of the seasonally adjusted series is often more informative about trends or location in a business cycle than a plot of the unadjusted series.

The following example shows how to use PROC X11 to produce a seasonally adjusted series, \( C_t I_t \), from an original series \( O_t = S_t C_t D_t I_t \).
In the multiplicative model, the trend cycle component $C_t$ keeps the same scale as the original series $O_t$, while $S_t$, $D_t$, and $I_t$ vary around 1.0. In all printed tables and in the output data set, these latter components are expressed as percentages, and thus will vary around 100.0 (in the additive case, they vary around 0.0).

The naming convention used in PROC X11 for the tables follows the original U.S. Bureau of the Census X-11 Seasonal Adjustment program specification (Shiskin, Young, and Musgrave 1967). Also, see the section “Printed Output” on page 3122. This convention is outlined in Figure 43.1.

The tables corresponding to parts A – C are intermediate calculations. The final estimates of the individual components are found in the D tables: table D10 contains the final seasonal factors, table D12 contains the final trend cycle, and table D13 contains the final irregular series. If you are primarily interested in seasonally adjusting a series without consideration of intermediate calculations or diagnostics, you only need to look at table D11, the final seasonally adjusted series.

For further details about the X-11-ARIMA tables, see Ladiray and Quenneville (2001).

---

**Basic Seasonal Adjustment**

Suppose you have monthly retail sales data starting in September 1978 in a SAS data set named SALES. At this point you do not suspect that any calendar effects are present, and there are no prior adjustments that need to be made to the data.

In this simplest case, you need only specify the DATE= variable in the MONTHLY statement, which associates a SAS date value to each observation. To see the results of the seasonal adjustment, you must request table D11, the final seasonally adjusted series, in a TABLES statement.

```sas
data sales;
  input sales @@;
  date = intnx('month', '01sep1978'd, _n_-1);
  format date monyy7.;
datalines;
112 118 132 129 121 135 148 148 136 119 104 118
... more lines ...
/*---- X-11 ARIMA ----*/
proc x11 data=sales;
  monthly date=date;
  var sales;
  tables d11;
run;
```
Chapter 43: The X11 Procedure

Seasonal Adjustment of Sales

X-11 Seasonal Adjustment Program
U. S. Bureau of the Census
Economic Research and Analysis Division
November 1, 1968

The X-11 program is divided into seven major parts.

<table>
<thead>
<tr>
<th>Part</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Prior adjustments, if any</td>
</tr>
<tr>
<td>B.</td>
<td>Preliminary estimates of irregular component weights and regression trading day factors</td>
</tr>
<tr>
<td>C.</td>
<td>Final estimates of above</td>
</tr>
<tr>
<td>D.</td>
<td>Final estimates of seasonal, trend-cycle and irregular components</td>
</tr>
<tr>
<td>E.</td>
<td>Analytical tables</td>
</tr>
<tr>
<td>F.</td>
<td>Summary measures</td>
</tr>
<tr>
<td>G.</td>
<td>Charts</td>
</tr>
</tbody>
</table>

Series - sales
Type of run: multiplicative seasonal adjustment.
Selected Tables or Charts.

Sigma limits for graduating extreme values are 1.5 and 2.5
Irregular values outside of 2.5-sigma limits are excluded from trading day regression

Figure 43.2 Basic Seasonal Adjustment

<table>
<thead>
<tr>
<th>D11 Final Seasonally Adjusted Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1978</td>
</tr>
<tr>
<td>1979</td>
</tr>
<tr>
<td>1980</td>
</tr>
<tr>
<td>1984</td>
</tr>
<tr>
<td>1988</td>
</tr>
<tr>
<td>1989</td>
</tr>
<tr>
<td>1990</td>
</tr>
</tbody>
</table>

Total: 40324 Mean: 280.03 S.D.: 111.31
You can compare the original series, table B1, and the final seasonally adjusted series, table D11, by plotting them together. These tables are requested and named in the OUTPUT statement.

```sas
proc x11 data=sales noprint;
    monthly date=date;
    var sales;
    output out=out b1=sales d11=adjusted;
run;

proc sgplot data=out;
    series x=date y=sales / markers
        markerattrs=(color=red symbol='asterisk')
        lineattrs=(color=red)
        legendlabel="original";
    series x=date y=adjusted / markers
        markerattrs=(color=blue symbol='circle')
        lineattrs=(color=blue)
        legendlabel="adjusted";
    yaxis label='Original and Seasonally Adjusted Time Series';
run;
```

**Figure 43.3** Plot of Original and Seasonally Adjusted Data
Chapter 43: The X11 Procedure

X-11-ARIMA

An inherent problem with the X-11 method is the revision of the seasonal factor estimates as new data become available. The X-11 method uses a set of centered moving averages to estimate the seasonal components. These moving averages apply symmetric weights to all observations except those at the beginning and end of the series, where asymmetric weights have to be applied. These asymmetric weights can cause poor estimates of the seasonal factors, which then can cause large revisions when new data become available.

While large revisions to seasonally adjusted values are not common, they can happen. When they do happen, it undermines the credibility of the X-11 seasonal adjustment method.

A method to address this problem was developed at Statistics Canada (Dagum 1980, 1982a). This method, known as X-11-ARIMA, applies an ARIMA model to the original data (after adjustments, if any) to forecast the series one or more years. This extended series is then seasonally adjusted, allowing symmetric weights to be applied to the end of the original data. This method was tested against a large number of Canadian economic series and was found to greatly reduce the amount of revisions as new data were added.

The X-11-ARIMA method is available in PROC X11 through the use of the ARIMA statement. The ARIMA statement extends the original series either with a user-specified ARIMA model or by an automatic selection process in which the best model from a set of five predefined ARIMA models is used.

The following example illustrates the use of the ARIMA statement. The ARIMA statement does not contain a user-specified model, so the best model is chosen by the automatic selection process. Forecasts from this best model are then used to extend the original series by one year. The following partial listing shows parameter estimates and model diagnostics for the ARIMA model chosen by the automatic selection process.

```
proc x11 data=sales;
  monthly date=date;
  var sales;
  arima;
run;
```

Figure 43.4 X-11-ARIMA Model Selection

Monthly Retail Sales Data (in $1000)

The X11 Procedure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t Value</th>
<th>Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>MU</td>
<td>0.0001728</td>
<td>0.0009596</td>
<td>0.18</td>
<td>0</td>
</tr>
<tr>
<td>MA1,1</td>
<td>0.3739984</td>
<td>0.0893427</td>
<td>4.19</td>
<td>1</td>
</tr>
<tr>
<td>MA1,2</td>
<td>0.0231478</td>
<td>0.0892154</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>MA2,1</td>
<td>0.5727914</td>
<td>0.0790835</td>
<td>7.24</td>
<td>12</td>
</tr>
</tbody>
</table>
Table D11 (final seasonally adjusted series) is now constructed using symmetric weights on observations at the end of the actual data. This should result in better estimates of the seasonal factors and, thus, smaller revisions in Table D11 as more data become available.

### Syntax: X11 Procedure

The X11 procedure uses the following statements:

```plaintext
PROC X11 options;
   ARIMA options;
   BY variables;
   ID variables;
   MACURVES option;
   MONTHLY options;
   OUTPUT OUT=dataset options;
   PDWEIGHTS option;
   QUARTERLY options;
   SSPAN options;
   TABLES tablenames;
   VAR variables;
```

Either the MONTHLY or QUARTERLY statement must be specified, depending on the type of time series data you have. The PDWEIGHTS and MACURVES statements can be used only with the MONTHLY statement. The TABLES statement controls the printing of tables, while the OUTPUT statement controls the creation of the OUT= data set.
# Functional Summary

The statements and options controlling the X11 procedures are summarized in the following table.

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Set Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>specify input data set</td>
<td>PROC X11</td>
<td>DATA=</td>
</tr>
<tr>
<td>write the trading-day regression results to an output data set</td>
<td>PROC X11</td>
<td>OUTTDR=</td>
</tr>
<tr>
<td>write the stable seasonality test results to an output data set</td>
<td>PROC X11</td>
<td>OUTSTB=</td>
</tr>
<tr>
<td>write table values to an output data set</td>
<td>OUTPUT</td>
<td>OUT=</td>
</tr>
<tr>
<td>add extrapolated values to the output data set</td>
<td>PROC X11</td>
<td>OTEX</td>
</tr>
<tr>
<td>add year ahead estimates to the output data set</td>
<td>PROC X11</td>
<td>YRAHEADOUT</td>
</tr>
<tr>
<td>write the sliding spans analysis results to an output data set</td>
<td>PROC X11</td>
<td>OUTSPAN=</td>
</tr>
<tr>
<td><strong>Printing Control Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suppress all printed output</td>
<td>PROC X11</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>suppress all printed ARIMA output</td>
<td>ARIMA</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>print all ARIMA output</td>
<td>ARIMA</td>
<td>PRINTALL</td>
</tr>
<tr>
<td>print selected tables and charts</td>
<td>TABLES</td>
<td>PRINTOUT=</td>
</tr>
<tr>
<td>print selected groups of tables</td>
<td>MONTHLY</td>
<td>PRINTOUT=</td>
</tr>
<tr>
<td>print selected groups of charts</td>
<td>MONTHLY</td>
<td>CHARTS=</td>
</tr>
<tr>
<td>print selected groups of charts</td>
<td>QUARTERLY</td>
<td>CHARTS=</td>
</tr>
<tr>
<td>print preliminary tables associated with ARIMA processing</td>
<td>ARIMA</td>
<td>PRINTFP</td>
</tr>
<tr>
<td>specify number of decimals for printed tables</td>
<td>MONTHLY</td>
<td>NDEC=</td>
</tr>
<tr>
<td>specify number of decimals for printed tables</td>
<td>QUARTERLY</td>
<td>NDEC=</td>
</tr>
<tr>
<td>suppress all printed SSPAN output</td>
<td>SSPAN</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>print all SSPAN output</td>
<td>SSPAN</td>
<td>PRINTALL</td>
</tr>
<tr>
<td><strong>Date Information Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>specify a SAS date variable</td>
<td>MONTHLY</td>
<td>DATE=</td>
</tr>
<tr>
<td>specify the beginning date</td>
<td>MONTHLY</td>
<td>START=</td>
</tr>
<tr>
<td>specify the ending date</td>
<td>MONTHLY</td>
<td>END=</td>
</tr>
<tr>
<td>specify beginning year for trading-day regression</td>
<td>MONTHLY</td>
<td>TDCOMPUTE=</td>
</tr>
</tbody>
</table>
### Declaring the Role of Variables

- specify BY-group processing: `BY`
- specify the variables to be seasonally adjusted: `VAR`
- specify identifying variables: `ID`
- specify the prior monthly factor: `MONTHLY` `PMFACTOR=`

### Controlling the Table Computations

- use additive adjustment: `MONTHLY` `ADDITIVE`
- specify seasonal factor moving average length: `MACURVES` `ADDITIVE`
- specify the extreme value limit for trading-day regression: `MONTHLY` `EXCLUDE=`
- specify the lower bound for extreme irregulars: `MONTHLY` `FULLWEIGHT=`
- specify the upper bound for extreme irregulars: `MONTHLY` `ZEROWEIGHT=`
- include the length-of-month in trading-day regression: `MONTHLY` `LENGTH`
- specify trading-day regression action: `MONTHLY` `TDREGR=`
- compute summary measure only: `MONTHLY` `SUMMARY`
- modify extreme irregulars prior to trend cycle estimation: `MONTHLY` `TRENDADJ` `QUARTERLY` `TRENDADJ`
- specify moving average length in trend cycle estimation: `MONTHLY` `TRENDMA=` `QUARTERLY` `TRENDMA=`
- specify weights for prior trading-day factors: `PDWEIGHTS`

---

**PROC X11 Statement**

```sas
PROC X11 options;
```

The following options can appear in the PROC X11 statement:

**DATA= SAS-data-set**

specifies the input SAS data set used. If it is omitted, the most recently created SAS data set is used.

**OUTEXTRAP**

adds the extra observations used in ARIMA processing to the output data set.

When ARIMA forecasting/backcasting is requested, extra observations are appended to the ends of the series, and the calculations are carried out on this extended series. The appended observations are not normally written to the OUT= data set. However, if OUTEXTRAP is specified, these extra observations are written to the output data set. If a DATE= variable is specified in the MONTHLY/QUARTERLY statement, the date variable is extrapolated to identify forecasts/backcasts. The OUTEXTRAP option can be abbreviated as OUTEX.
**Chapter 43: The X11 Procedure**

**NOPRINT**
suppresses any printed output. The NOPRINT option overrides any PRINTOUT=, CHARTS=, or TABLES statement and any output associated with the ARIMA statement.

**OUTSPAN= SAS-data-set**
specifies the output data set to store the sliding spans analysis results. Tables A1, C18, D10, and D11 for each span are written to this data set. See the section “The OUTSPAN= Data Set” on page 3120 for details.

**OUTSTB= SAS-data-set**
specifies the output data set to store the stable seasonality test results (table D8). All the information in the analysis of variance table associated with the stable seasonality test is contained in the variables written to this data set. See the section “OUTSTB= Data Set” on page 3120 for details.

**OUTTDR= SAS-data-set**
specifies the output data set to store the trading-day regression results (tables B15 and C15). All the information in the analysis of variance table associated with the trading-day regression is contained in the variables written to this data set. This option is valid only when TDREGR=PRINT, TEST, or ADJUST is specified in the MONTHLY statement. See the section “OUTTDR= Data Set” on page 3121 for details.

**YRAHEADOUT**
adds one-year-ahead forecast values to the output data set for tables C16, C18, and D10. The original purpose of this option was to avoid recomputation of the seasonal adjustment factors when new data became available. While computing costs were an important factor when the X-11 method was developed, this is no longer the case and this option is obsolete. See the section “The YRAHEADOUT Option” on page 3116 for details.

---

**ARIMA Statement**

```
ARIMA options ;
```

The ARIMA statement applies the X-11-ARIMA method to the series specified in the VAR statement. This method uses an ARIMA model estimated from the original data to extend the series one or more years. The ARIMA statement options control the ARIMA model used and the estimation, forecasting, and printing of this model.

There are two ways of obtaining an ARIMA model to extend the series. A model can be given explicitly with the MODEL= and TRANSFORM= options. Alternatively, the best-fitting model from a set of five predefined models is found automatically whenever the MODEL= option is absent. See the section “Details of Model Selection” on page 3117 for details.

**BACKCAST= n**
specifies the number of years to backcast the series. The default is BACKCAST= 0. See the section “Effect of Backcast and Forecast Length” on page 3116 for details.
**CHICR= value**
specifies the criteria for the significance level for the Box-Ljung chi-square test for lack of fit when testing the five predefined models. The default is CHICR= 0.05. The CHICR= option values must be between 0.01 and 0.90. The hypothesis being tested is that of model adequacy. Nonrejection of the hypothesis is evidence for an adequate model. Making the CHICR= value smaller makes it easier to accept the model. See the section “Criteria Details” on page 3118 for further details on the CHICR= option.

**CONVERGE= value**
specifies the convergence criterion for the estimation of an ARIMA model. The default value is 0.001. The CONVERGE= value must be positive.

**FORECAST= n**
specifies the number of years to forecast the series. The default is FORECAST= 1. See the section “Effect of Backcast and Forecast Length” on page 3116 for details.

**MAPECR= value**
specifies the criteria for the mean absolute percent error (MAPE) when testing the five predefined models. A small MAPE value is evidence for an adequate model; a large MAPE value results in the model being rejected. The MAPECR= value is the boundary for acceptance/rejection. Thus a larger MAPECR= value would make it easier for a model to pass the criteria. The default is MAPECR= 15. The MAPECR= option values must be between 1 and 100. See the section “Criteria Details” on page 3118 for further details on the MAPECR= option.

**MAXITER= n**
specifies the maximum number of iterations in the estimation process. MAXITER must be between 1 and 60; the default value is 15.

**METHOD= CLS**
**METHOD= ULS**
**METHOD= ML**
specifies the estimation method. ML requests maximum likelihood, ULS requests unconditional least squares, and CLS requests conditional least squares. METHOD=CLS is the default. The maximum likelihood estimates are more expensive to compute than the conditional least squares estimates. In some cases, however, they can be preferable. For further information on the estimation methods, see “Estimation Details” on page 272 in Chapter 8, “The ARIMA Procedure.”

**MODEL= ( P=n1 Q=n2 SP=n3 SQ=n4 DIF=n5 SDIF=n6 < NOINT > < CENTER >)**
specifies the ARIMA model. The AR and MA orders are given by P=n1 and Q=n2, respectively, while the seasonal AR and MA orders are given by SP=n3 and SQ=n4, respectively. The lag corresponding to seasonality is determined by the MONTHLY or QUARTERLY statement. Similarly, differencing and seasonal differencing are given by DIF=n5 and SDIF=n6, respectively.

For example

```plaintext
arima model=( p=2 q=1 sp=1 dif=1 sdif=1 );
```
specifies a (2,1,1)(1,1,0)s model, where s, the seasonality, is either 12 (monthly) or 4 (quarterly). More examples of the MODEL= syntax are given in the section “Details of Model Selection” on page 3117.
**NOINT**
suppresses the fitting of a constant (or intercept) parameter in the model. (That is, the parameter \(\mu\) is omitted.)

**CENTER**
centers each time series by subtracting its sample mean. The analysis is done on the centered data. Later, when forecasts are generated, the mean is added back. Note that centering is done after differencing. The CENTER option is normally used in conjunction with the NOCONSTANT option of the ESTIMATE statement.

For example, to fit an AR(1) model on the centered data without an intercept, use the following ARIMA statement:

\[
\text{arima model=( p=1 center noint )};
\]

**NOPRINT**
suppresses the normal printout generated by the ARIMA statement. Note that the effect of specifying the NOPRINT option in the ARIMA statement is different from the effect of specifying the NOPRINT in the PROC X11 statement, since the former only affects ARIMA output.

**OVDIFCR= value**
specifies the criteria for the over-differencing test when testing the five predefined models. When the MA parameters in one of these models sum to a number close to 1.0, this is an indication of over-parameterization and the model is rejected. The \(\text{OVDIFCR=}\) value is the boundary for this rejection; values greater than this value fail the over-differencing test. A larger \(\text{OVDIFCR=}\) value would make it easier for a model to pass the criteria. The default is \(\text{OVDIFCR=} 0.90\). The \(\text{OVDIFCR=}\) option values must be between 0.80 and 0.99. See the section “Criteria Details” on page 3118 for further details on the \(\text{OVDIFCR=}\) option.

**PRINTALL**
provides the same output as the default printing for all models fit and, in addition, prints an estimation summary and chi-square statistics for each model fit. See “Printed Output” on page 3122 for details.

**PRINTFP**
prints the results for the initial pass of X11 made to exclude trading-day effects. This option has an effect only when the TDREGR= option specifies ADJUST, TEST, or PRINT. In these cases, an initial pass of the standard X11 method is required to get rid of calendar effects before doing any ARIMA estimation. Usually this first pass is not of interest, and by default no tables are printed. However, specifying PRINTFP in the ARIMA statement causes any tables printed in the final pass to also be printed for this initial pass.

**TRANSFORM= (LOG) | LOG**

**TRANSFORM= ( constant ** power )**
The ARIMA statement in PROC X11 allows certain transformations on the series before estimation. The specified transformation is applied only to a user-specified model. If TRANSFORM= is specified and the MODEL= option is not specified, the transformation request is ignored and a warning is printed.

The LOG transformation requests that the natural log of the series be used for estimation. The resulting forecast values are transformed back to the original scale.
A general power transformation of the form \( X_t \rightarrow (X_t + a)^b \) is obtained by specifying

\[
\text{transform} = (a ** b)
\]

If the constant \( a \) is not specified, it is assumed to be zero. The specified ARIMA model is then estimated using the transformed series. The resulting forecast values are transformed back to the original scale.

**BY Statement**

\[
\text{BY variables ;}
\]

A BY statement can be used with PROC X11 to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input DATA= data set to be sorted in order of the BY variables.

**ID Statement**

\[
\text{ID variables ;}
\]

If you are creating an output data set, use the ID statement to put values of the ID variables, in addition to the table values, into the output data set. The ID statement has no effect when an output data set is not created. If the DATE= variable is specified in the MONTHLY or QUARTERLY statement, this variable is included automatically in the OUTPUT data set. If no DATE= variable is specified, the variable _DATE_ is added.

The date variable (or _DATE_) values outside the range of the actual data (from ARIMA forecasting or backcasting, or from YRAHEADOUT) are extrapolated, while all other ID variables are missing.

**MACURVES Statement**

\[
\text{MACURVES month=option . . . ;}
\]

The MACURVES statement specifies the length of the moving-average curves for estimating the seasonal factors for any month. This statement can be used only with monthly time series data.

The \( \text{month=option} \) specifications consist of the month name (or the first three letters of the month name), an equal sign, and one of the following option values:

- '3' specifies a three-term moving average for the month
- '3X3' specifies a three-by-three moving average
- '3X5' specifies a three-by-five moving average
- '3X9' specifies a three-by-nine moving average
- STABLE specifies a stable seasonal factor (average of all values for the month)

For example, the statement
uses a three-term moving average to estimate seasonal factors for January, a $3 \times 3$ (a three-term moving average of a three-term moving average) for February, a $3 \times 5$ (a three-term moving average of a five-term moving average) for March, and a $3 \times 9$ (a three-term moving average of a nine-term moving average) for April.

The numeric values used for the weights of the various moving averages and a discussion of the derivation of these weights are given in Shiskin, Young, and Musgrave (1967). A general discussion of moving average weights is given in Dagum (1985).

If the specification for a month is omitted, the X11 procedure uses a three-by-three moving average for the first estimate of each iteration and a three-by-five average for the second estimate.

**MONTHLY Statement**

**MONTHLY options ;**

The MONTHLY statement must be used when the input data to PROC X11 are a monthly time series. The MONTHLY statement specifies options that determine the computations performed by PROC X11 and what is included in its output. Either the DATE= or START= option must be used.

The following options can appear in the MONTHLY statement.

**ADDITIVE**

performs additive adjustments. If the ADDITIVE option is omitted, PROC X11 performs multiplicative adjustments.

**CHARTS= STANDARD**

**CHARTS= FULL**

**CHARTS= NONE**

specifies the charts produced by the procedure. The default is CHARTS=STANDARD, which specifies 12 monthly seasonal charts and a trend cycle chart. If you specify CHARTS=FULL (or CHARTS=ALL), the procedure prints additional charts of irregular and seasonal factors. To print no charts, specify CHARTS=NONE.

The TABLES statement can also be used to specify particular monthly charts to be printed. If no CHARTS= option is given, and a TABLES statement is given, the TABLES statement overrides the default value of CHARTS=STANDARD; that is, no charts (or tables) are printed except those specified in the TABLES statement. However, if both the CHARTS= option and a TABLES statement are given, the charts corresponding to the CHARTS= option and those requested by the TABLES statement are printed.

For example, suppose you wanted only charts G1, the final seasonally adjusted series and trend cycle, and G4, the final irregular and final modified irregular series. You would specify the following statements:

```
monthly date=date;
tables g1 g4;
```
DATE= variable
specifies a variable that gives the date for each observation. The starting and ending dates are obtained from the first and last values of the DATE= variable, which must contain SAS date values. The procedure checks values of the DATE= variable to ensure that the input observations are sequenced correctly. This variable is automatically added to the OUTPUT= data set if one is requested and extrapolated if necessary. If the DATE= option is not specified, the START= option must be specified.

The DATE= option and the START= and END= options can be used in combination to subset a series for processing. For example, suppose you have 12 years of monthly data (144 observations, no missing values) beginning in January 1970 and ending in December 1981, and you wanted to seasonally adjust only six years beginning in January 1974. Specifying

```
monthly date=date start=jan1974 end=dec1979;
```

would seasonally adjust only this subset of the data. If instead you wanted to adjust the last eight years of data, only the START= option is needed:

```
monthly date=date start=jan1974;
```

END= mmmyyyy
specifies that only the part of the input series ending with the month and year given be adjusted (for example, END=DEC1970). See the DATE= variable option for using the START= and END= options to subset a series for processing.

EXCLUDE= value
excludes from the trading-day regression any irregular values that are more than value standard deviations from the mean. The EXCLUDE=value must be between 0.1 and 9.9, with the default value being 2.5.

FULLWEIGHT= value
assigns weights to irregular values based on their distance from the mean in standard deviation units. The weights are used for estimating seasonal and trend cycle components. Irregular values less than the FULLWEIGHT= value (in standard deviation units) are assigned full weights of 1, values that fall between the ZEROWEIGHT= and FULLWEIGHT= limits are assigned weights linearly graduated between 0 and 1, and values greater than the ZEROWEIGHT= limit are assigned a weight of 0.

For example, if ZEROWEIGHT=2 and FULLWEIGHT=1, a value 1.3 standard deviations from the mean would be assigned a graduated weight. The FULLWEIGHT=value must be between 0.1 and 9.9 but must be less than the ZEROWEIGHT=value. The default is FULLWEIGHT=1.5.

LENGTH
includes length-of-month allowance in computing trading-day factors. If this option is omitted, length-of-month allowances are included with the seasonal factors.

NDEC= n
specifies the number of decimal places shown in the printed tables in the listing. This option has no effect on the precision of the variable values in the output data set.
PMFACTOR= \texttt{variable}

specifies a variable containing the prior monthly factors. Use this option if you have previous knowledge of monthly adjustment factors. The PMFACTOR= option can be used to make the following adjustments:

- adjust the level of all or part of a series with discontinuities
- adjust for the influence of holidays that fall on different dates from year to year, such as the effect of Easter on certain retail sales
- adjust for unreasonable weather influence on series, such as housing starts
- adjust for changing starting dates of fiscal years (for budget series) or model years (for automobiles)
- adjust for temporary dislocating events, such as strikes

See the section "Prior Daily Weights and Trading-Day Regression" on page 3114 for details and examples using the PMFACTOR= option.

PRINTOUT= \texttt{STANDARD | LONG | FULL | NONE}

specifies the tables to be printed by the procedure. If the PRINTOUT=STANDARD option is specified, between 17 and 27 tables are printed, depending on the other options that are specified. PRINTOUT=LONG prints between 27 and 39 tables, and PRINTOUT=FULL prints between 44 and 59 tables. Specifying PRINTOUT=NONE results in no tables being printed; however, charts are still printed. The default is PRINTOUT=STANDARD.

The TABLES statement can also be used to specify particular monthly tables to be printed. If no PRINTOUT= option is specified, and a TABLES statement is given, the TABLES statement overrides the default value of PRINTOUT=STANDARD; that is, no tables (or charts) are printed except those given in the TABLES statement. However, if both the PRINTOUT= option and a TABLES statement are specified, the tables corresponding to the PRINTOUT= option and those requested by the TABLES statement are printed.

START= \texttt{mmmyyyy}

adjusts only the part of the input series starting with the specified month and year. When the DATE= option is not used, the START= option gives the year and month of the first input observation — for example, START=JAN1966. START= must be specified if DATE= is not given. If START= is specified (and no DATE= option is given), and an OUT= data set is requested, a variable named \texttt{DATE}_ is added to the data set, giving the date value for each observation. See the DATE= \texttt{variable} option for using the START= and END= options to subset a series.

SUMMARY

specifies that the data are already seasonally adjusted and the procedure is to produce summary measures. If the SUMMARY option is omitted, the X11 procedure performs seasonal adjustment of the input data before calculating summary measures.

TDCOMPUTE= \texttt{year}

uses the part of the input series beginning with January of the specified year to derive trading-day weights. If this option is omitted, the entire series is used.
OUTPUT Statement

OUTPUT OUT= SAS-data-set tablename=var1 var2 . . . ;

The OUTPUT statement creates an output data set containing specified tables. The data set is named by the OUT= option.
Chapter 43: The X11 Procedure

OUT= SAS-data-set
If OUT= is omitted, the SAS System names the new data set by using the DATAn convention.

For each table to be included in the output data set, write the X11 table identification keyword, an equal sign, and a list of new variable names:

tablename = var1 var2 ...

The tablename keywords that can be used in the OUTPUT statement are listed in the section “Printed Output” on page 3122. The following is an example of a VAR statement and an OUTPUT statement:

```
var z1 z2 z3;
output out=out_x11 b1=s d11=w x y;
```

The variable s contains the table B1 values for the variable z1, while the table D11 values for variables z1, z2, and z3 are contained in variables w, x, and y, respectively. As this example shows, the list of variables following a tablename= keyword can be shorter than the VAR variable list.

In addition to the variables named by tablename =var1 var2 ..., the ID variables, and BY variables, the output data set contains a date identifier variable. If the DATE= option is given in the MONTHLY or QUARTERLY statement, the DATE= variable is the date identifier. If no DATE= option is given, a variable named _DATE_ is the date identifier.

PDWEIGHTS Statement

```
PDWEIGHTS day=w . . . ;
```

The PDWEIGHTS statement can be used to specify one to seven daily weights. The statement can only be used with monthly series that are seasonally adjusted using the multiplicative model. These weights are used to compute prior trading-day factors, which are then used to adjust the original series prior to the seasonal adjustment process. Only relative weights are needed; the X11 procedure adjusts the weights so that they sum to 7.0. The weights can also be corrected by the procedure on the basis of estimates of trading-day variation from the input data.

See the section “Prior Daily Weights and Trading-Day Regression” on page 3114 for details and examples using the PDWEIGHTS statement.

Each day=w option specifies a weight (w) for the named day. The day can be any day, Sunday through Saturday. The day keyword can be the full spelling of the day, or the three-letter abbreviation. For example, SATURDAY=1.0 and SAT=1.0 are both valid. The weights w must be a numeric value between 0.0 and 10.0.

The following is an example of a PDWEIGHTS statement:

```
pdweights sun=.2 mon=.9 tue=1 wed=1 thu=1 fri=.8 sat=.3;
```

Any number of days can be specified with one PDWEIGHTS statement. The default weight value for any day that is not specified is 0. If you do not use a PDWEIGHTS statement, the program computes daily weights if TDREGR=ADJUST is specified. See Shiskin, Young, and Musgrave (1967) for details.
QUARTERLY Statement

QUARTERLY options;

The QUARTERLY statement must be used when the input data are quarterly time series. This statement includes options that determine the computations performed by the procedure and what is in the printed output. The DATE= option or the START= option must be used.

The following options can appear in the QUARTERLY statement.

ADDITIVE
   performs additive adjustments. If this option is omitted, the procedure performs multiplicative adjustments.

CHARTS= STANDARD
CHARTS= FULL
CHARTS= NONE

specifies the charts to be produced by the procedure. The default value is CHARTS=STANDARD, which specifies four quarterly seasonal charts and a trend cycle chart. If you specify CHARTS=FULL (or CHARTS=ALL), the procedure prints additional charts of irregular and seasonal factors. To print no charts, specify CHARTS=NONE. The TABLES statement can also be used to specify particular charts to be printed. The presence of a TABLES statement overrides the default value of CHARTS=STANDARD; that is, if a TABLES statement is specified, and no CHARTS=option is specified, no charts (nor tables) are printed except those given in the TABLES statement. However, if both the CHARTS= option and a TABLES statement are given, the charts corresponding to the CHARTS= option and those requested by the TABLES statement are printed.

For example, suppose you wanted only charts G1, the final seasonally adjusted series and trend cycle, and G4, the final irregular and final modified irregular series. This is accomplished by specifying the following statements:

   quarterly date=date;
   tables g1 g4;

DATE= variable

specifies a variable that gives the date for each observation. The starting and ending dates are obtained from the first and last values of the DATE= variable, which must contain SAS date values. The procedure checks values of the DATE= variable to ensure that the input observations are sequenced correctly. This variable is automatically added to the OUTPUT= data set if one is requested, and extrapolated if necessary. If the DATE= option is not specified, the START= option must be specified.

The DATE= option and the START= and END= options can be used in combination to subset a series for processing. For example, suppose you have a series with 10 years of quarterly data (40 observations, no missing values) beginning in ‘1970Q1’ and ending in ‘1979Q4’, and you want to seasonally adjust only four years beginning in ‘1974Q1’ and ending in ‘1977Q4’. Specifying

   quarterly date=variable start='1974q1' end='1977q4';

seasonally adjusts only this subset of the data. If instead you wanted to adjust the last six years of data, only the START= option is needed:
quarterly date=variable start='1974q1';

END= 'yyyyQq'
specifies that only the part of the input series ending with the quarter and year given be adjusted (for example, END=’1973Q4’). The specification must be enclosed in quotes and q must be 1, 2, 3, or 4. See the DATE= variable option for using the START= and END= options to subset a series.

FULLWEIGHT= value
assigns weights to irregular values based on their distance from the mean in standard deviation units. The weights are used for estimating seasonal and trend cycle components. Irregular values less than the FULLWEIGHT= value (in standard deviation units) are assigned full weights of 1, values that fall between the ZEROWEIGHT= and FULLWEIGHT= limits are assigned weights linearly graduated between 0 and 1, and values greater than the ZEROWEIGHT= limit are assigned a weight of 0.

For example, if ZEROWEIGHT=2 and FULLWEIGHT=1, a value 1.3 standard deviations from the mean would be assigned a graduated weight. The default is FULLWEIGHT=1.5.

NDEC= n
specifies the number of decimal places shown on the output tables. This option has no effect on the precision of the variables in the output data set.

PRINTOUT= STANDARD
PRINTOUT= LONG
PRINTOUT= FULL
PRINTOUT= NONE
specifies the tables to print. If PRINTOUT=STANDARD is specified, between 17 and 27 tables are printed, depending on the other options that are specified. PRINTOUT=LONG prints between 27 and 39 tables, and PRINTOUT=FULL prints between 44 and 59 tables. Specifying PRINTOUT=NONE results in no tables being printed. The default is PRINTOUT=STANDARD.

The TABLES statement can also specify particular quarterly tables to be printed. If no PRINTOUT= is given, and a TABLES statement is given, the TABLES statement overrides the default value of PRINTOUT=STANDARD; that is, no tables (or charts) are printed except those given in the TABLES statement. However, if both the PRINTOUT= option and a TABLES statement are given, the tables corresponding to the PRINTOUT= option and those requested by the TABLES statement are printed.

START= 'yyyyQq'
adjusts only the part of the input series starting with the quarter and year given. When the DATE= option is not used, the START= option gives the year and quarter of the first input observation (for example, START=’1967Q1’). The specification must be enclosed in quotes, and q must be 1, 2, 3, or 4. START= must be specified if the DATE= option is not given. If START= is specified (and no DATE= is given), and an OUTPUT= data set is requested, a variable named _DATE_ is added to the data set, giving the date value for a given observation. See the DATE= option for using the START= and END= options to subset a series.

SUMMARY
specifies that the input is already seasonally adjusted and that the procedure is to produce summary measures. If this option is omitted, the procedure performs seasonal adjustment of the input data before calculating summary measures.
TRENDADJ modifies extreme irregular values prior to computing the trend cycle estimates. If this option is omitted, the trend cycle is computed without modification for extremes.

ZEROWEIGHT=value assigns weights to irregular values based on their distance from the mean in standard deviation units. The weights are used for estimating seasonal and trend cycle components. Irregular values beyond the standard deviation limit specified in the ZEROWEIGHT= option are assigned zero weights. Values that fall between the two limits (ZEROWEIGHT= and FULLWEIGHT=) are assigned weights linearly graduated between 0 and 1. For example, if ZEROWEIGHT=2 and FULLWEIGHT=1, a value 1.3 standard deviations from the mean would be assigned a graduated weight. The default is ZEROWEIGHT=2.5.

The ZEROWEIGHT option can be used in conjunction with the FULLWEIGHT= option to adjust outliers from a monthly or quarterly series. See Example 43.3 later in this chapter for an illustration of this use.

### SSPAN Statement

The SSPAN statement applies sliding spans analysis to determine the suitability of seasonal adjustment for an economic series.

The following options can appear in the SSPAN statement:

- **NDEC=n** specifies the number of decimal places shown on selected sliding span reports. This option has no effect on the precision of the variables values in the OUTSPAN output data set.

- **CUTOFF=value** gives the percentage value for determining an excessive difference within a span for the seasonal factors, the seasonally adjusted series, and month-to-month and year-to-year differences in the seasonally adjusted series. The default value is 3.0. The use of the CUTOFF=value in determining the maximum percent difference (MPD) is described in the section “Computational Details for Sliding Spans Analysis” on page 3111. Caution should be used in changing the default CUTOFF=value. The empirical threshold ranges found by the U.S. Census Bureau no longer apply when value is changed.

- **TDCUTOFF=value** gives the percentage value for determining an excessive difference within a span for the trading-day factors. The default value is 2.0. The use of the TDCUTOFF=value in determining the maximum percent difference (MPD) is described in the section “Computational Details for Sliding Spans Analysis” on page 3111. Caution should be used in changing the default TDCUTOFF=value. The empirical threshold ranges found by the U.S. Census Bureau no longer apply when the value is changed.

- **NOPRINT** suppresses all sliding span reports. See “Computational Details for Sliding Spans Analysis” on page 3111 for more details on sliding span reports.
**PRINT**

prints the summary sliding span reports S 0 through S 6.E.

**PRINTALL**

prints the summary sliding spans report S 0 through S 6.E, along with detail reports S 7.A through S 7.E.

**TABLES Statement**

```
TABLES tablenames ;
```

The TABLES statement prints the tables specified in addition to the tables that are printed as a result of the PRINTOUT= option in the MONTHLY or QUARTERLY statement. Table names are listed in Table 43.4 later in this chapter.

To print only selected tables, omit the PRINTOUT= option in the MONTHLY or QUARTERLY statement and list the tables to be printed in the TABLES statement. For example, to print only the final seasonal factors and final seasonally adjusted series, use the statement

```
tables d10 d11;
```

**VAR Statement**

```
VAR variables ;
```

The VAR statement is used to specify the variables in the input data set that are to be analyzed by the procedure. Only numeric variables can be specified. If the VAR statement is omitted, all numeric variables are analyzed except those appearing in a BY or ID statement or the variable named in the DATE= option in the MONTHLY or QUARTERLY statement.

**Details: X11 Procedure**

**Historical Development of X-11**

This section briefly describes the historical development of the standard X-11 seasonal adjustment method and the later development of the X-11-ARIMA method. Most of the following discussion is based on a comprehensive article by Bell and Hillmer (1984), which describes the history of X-11 and the justification of using seasonal adjustment methods, such as X-11, given the current availability of time series software. For further discussions about statistical problems associated with the X-11 method, see Ghysels (1990).

Seasonal adjustment methods began to be developed in the 1920s and 1930s, before there were suitable analytic models available and before electronic computing devices were in existence. The lack of any suitable model led to methods that worked the same for any series — that is, methods that were not model-based and
that could be applied to any series. Experience with economic series had shown that a given mathematical form could adequately represent a time series only for a fixed length; as more data were added, the model became inadequate. This suggested an approach that used moving averages. For further analysis of the properties of X-11 moving averages, see Cleveland and Tiao (1976).

The basic method was to break up an economic time series into long-term trend, long-term cyclical movements, seasonal movements, and irregular fluctuations.

Early investigators found that it was not possible to uniquely decompose the trend and cycle components. Thus, these two were grouped together; the resulting component is usually referred to as the “trend cycle component.”

It was also found that estimating seasonal components in the presence of trend produced biased estimates of the seasonal components, but, at the same time, estimating trend in the presence of seasonality was difficult. This eventually lead to the iterative approach used in the X-11 method.

Two other problems were encountered by early investigators. First, some economic series appear to have changing or evolving seasonality. Secondly, moving averages were very sensitive to extreme values. The estimation method used in the X-11 method allows for evolving seasonal components. For the second problem, the X-11 method uses repeated adjustment of extreme values.

All of these problems encountered in the early investigation of seasonal adjustment methods suggested the use of moving averages in estimating components. Even with the use of moving averages instead of a model-based method, massive amounts of hand calculations were required. Only a small number of series could be adjusted, and little experimentation could be done to evaluate variations on the method.

With the advent of electronic computing in the 1950s, work on seasonal adjustment methods proceeded rapidly. These methods still used the framework previously described; variants of these basic methods could now be easily tested against a large number of series.

Much of the work was done by Julian Shiskin and others at the U.S. Bureau of the Census beginning in 1954 and culminating after a number of variants into the X-11 Variant of the Census Method II Seasonal Adjustment Program, which PROC X11 implements.

References for this work during this period include Shiskin and Eisenpress (1957), Shiskin (1958), and Marris (1961). The authoritative documentation for the X-11 Variant is in Shiskin, Young, and Musgrave (1967). This document is not equivalent to a program specification; however, the FORTRAN code that implements the X-11 Variant is in the public domain. A less detailed description of the X-11 Variant is given in US Bureau of the Census (1969).

**Development of the X-11-ARIMA Method**

The X-11 method uses symmetric moving averages in estimating the various components. At the end of the series, however, these symmetric weights cannot be applied. Either asymmetric weights have to be used, or some method of extending the series must be found.

While various methods of extending a series have been proposed, the most important method to date has been the X-11-ARIMA method developed at Statistics Canada. This method uses Box-Jenkins ARIMA models to extend the series.

The Time Series Research and Analysis Division of Statistics Canada investigated 174 Canadian economic series and found five ARIMA models out of twelve that fit the majority of series well and reduced revisions for the most recent months. References that give details of various aspects of the X-11-ARIMA methodology...

**Differences between X11ARIMA/88 and PROC X11**

The original implementation of the X-11-ARIMA method was by Statistics Canada in 1980 (Dagum 1980), with later changes and enhancements made in 1988 (Dagum 1988). The calculations performed by PROC X11 differ from those in X11ARIMA/88, which will result in differences in the final component estimates provided by these implementations.

There are three areas where Statistics Canada made changes to the original X-11 seasonal adjustment method in developing X11ARIMA/80 (Monsell 1984). These are (a) selection of extreme values, (b) replacement of extreme values, and (c) generation of seasonal and trend cycle weights.

These changes have not been implemented in the current version of PROC X11. Thus the procedure produces results identical to those from previous versions of PROC X11 in the absence of an ARIMA statement.

Additional differences can result from the ARIMA estimation. X11ARIMA/88 uses conditional least squares (CLS), while CLS, unconditional least squares (ULS) and maximum likelihood (ML) are all available in PROC X11 by using the METHOD= option in the ARIMA statement. Generally, parameters estimates will differ for the different methods.

**Implementation of the X-11 Seasonal Adjustment Method**

The following steps describe the analysis of a monthly time series using multiplicative seasonal adjustment. Additional steps used by the X-11-ARIMA method are also indicated. Equivalent descriptions apply for an additive model if you replace *divide* with *subtract* where applicable.

In the multiplicative adjustment, the original series \( O_t \) is assumed to be of the form

\[
O_t = C_t S_t I_t P_t D_t
\]

where \( C_t \) is the trend cycle component, \( S_t \) is the seasonal component, \( I_t \) is the irregular component, \( P_t \) is the prior monthly factors component, and \( D_t \) is the trading-day component.

The trading-day component can be further factored as

\[
D_t = D_{t,r} D_{t,r,t},
\]

where \( D_{t,r,t} \) are the trading-day factors derived from the prior daily weights, and \( D_{t,r,t} \) are the residual trading-day factors estimated from the trading-day regression. For further information about estimating trading day variation, see Young (1965).

**Additional Steps When Using the X-11-ARIMA Method**

The X-11-ARIMA method consists of extending a given series by an ARIMA model and applying the usual X-11 seasonal adjustment method to this extended series. Thus in the simplest case in which there are no prior factors or calendar effects in the series, the ARIMA model selection, estimation, and forecasting are performed first, and the resulting extended series goes through the standard X-11 steps described in the next section.
Implementation of the X-11 Seasonal Adjustment Method

If prior factor or calendar effects are present, they must be eliminated from the series before the ARIMA estimation is done because these effects are not stochastic.

Prior factors, if present, are removed first. Calendar effects represented by prior daily weights are then removed. If there are no further calendar effects, the adjusted series is extended by the ARIMA model, and this extended series goes through the standard X-11 steps without repeating the removal of prior factors and calendar effects from prior daily weights.

If further calendar effects are present, a trading-day regression must be performed. In this case it is necessary to go through an initial pass of the X-11 steps to obtain a final trading-day adjustment. In this initial pass, the series, adjusted for prior factors and prior daily weights, goes through the standard X-11 steps. At the conclusion of these steps, a final series adjusted for prior factors and all calendar effects is available. This adjusted series is then extended by the ARIMA model, and this extended series goes through the standard X-11 steps again, without repeating the removal of prior factors and calendar effects from prior daily weights and trading-day regression.

The Standard X-11 Seasonal Adjustment Method

The standard X-11 seasonal adjustment method consists of the following steps. These steps are applied to the original data or the original data extended by an ARIMA model.

1. In step 1, the data are read, ignoring missing values until the first nonmissing value is found. If prior monthly factors are present, the procedure reads prior monthly $P_t$ factors and divides them into the original series to obtain $O_t/P_t = C_t S_t I_t D_{tr,t} D_{r,t}$.

   Seven daily weights can be specified to develop monthly factors to adjust the series for trading-day variation, $D_{tr,t}$; these factors are then divided into the original or prior adjusted series to obtain $C_t S_t I_t D_{r,t}$.

2. In steps 2, 3, and 4, three iterations are performed, each of which provides estimates of the seasonal $S_t$, trading-day $D_{tr,t}$, trend cycle $C_t$, and irregular components $I_t$. Each iteration refines estimates of the extreme values in the irregular components. After extreme values are identified and modified, final estimates of the seasonal component, seasonally adjusted series, trend cycle, and irregular components are produced. Step 2 consists of three substeps:

   a) During the first iteration, a centered, 12-term moving average is applied to the original series $O_t$ to provide a preliminary estimate $\hat{C}_t$ of the trend cycle curve $C_t$. This moving average combines 13 (a 2-term moving average of a 12-term moving average) consecutive monthly values, removing the $S_t$ and $I_t$. Next, it obtains a preliminary estimate $\hat{S}_t I_t$ by

   $$\hat{S}_t I_t = \frac{O_t}{\hat{C}_t}$$

   b) A moving average is then applied to the $\hat{S}_t I_t$ to obtain an estimate $\hat{S}_t$ of the seasonal factors. $\hat{S}_t I_t$ is then divided by this estimate to obtain an estimate $\hat{I}_t$ of the irregular component. Next, a moving standard deviation is calculated from the irregular component and is used in assigning a weight to each monthly value for measuring its degree of extremeness. These weights are used to modify extreme values in $\hat{S}_t I_t$. New seasonal factors are estimated by applying a moving average to the modified value of $\hat{S}_t I_t$. A preliminary seasonally adjusted series is obtained by dividing the original series by these new seasonal factors. A second estimate of the trend cycle is obtained by applying a weighted moving average to this seasonally adjusted series.
Chapter 43: The X11 Procedure

c) The same process is used to obtain second estimates of the seasonally adjusted series and improved estimates of the irregular component. This irregular component is again modified for extreme values and then used to provide estimates of trading-day factors and refined weights for the identification of extreme values.

3. Using the same computations, a second iteration is performed on the original series that has been adjusted by the trading-day factors and irregular weights developed in the first iteration. The second iteration produces final estimates of the trading-day factors and irregular weights.

4. A third and final iteration is performed using the original series that has been adjusted for trading-day factors and irregular weights computed during the second iteration. During the third iteration, PROC X11 develops final estimates of seasonal factors, the seasonally adjusted series, the trend cycle, and the irregular components. The procedure computes summary measures of variation and produces a moving average of the final adjusted series.

Sliding Spans Analysis

The motivation for sliding spans analysis is to answer the question, When is a economic series unsuitable for seasonal adjustment? There have been a number of past attempts to answer this question: stable seasonality $F$ test; moving seasonality $F$ test, $Q$ statistics, and others.

Sliding spans analysis attempts to quantify the stability of the seasonal adjustment process, and hence quantify the suitability of seasonal adjustment for a given series.

It is based on a very simple idea: for a stable series, deleting a small number of observations should not result in greatly different component estimates compared with the original, full series. Conversely, if deleting a small number of observations results in drastically different estimates, the series is unstable. For example, a drastic difference in the seasonal factors (Table D10) might result from a dominating irregular component or sudden changes in the seasonally component. When the seasonal component estimates of a series is unstable in this manner, they have little meaning and the series is likely to be unsuitable for seasonal adjustment.

Sliding spans analysis, developed at the Statistical Research Division of the U.S. Census Bureau (Findley et al. 1990; Findley and Monsell 1986), performs a repeated seasonal adjustment on subsets or spans of the full series. In particular, an initial span of the data, typically eight years in length, is seasonally adjusted, and the Tables C18, the trading-day factors (if trading-day regression performed), D10, the seasonal factors, and D11, the seasonally adjusted series are retained for further processing. Next, one year of data is deleted from the beginning of the initial span and one year of data is added. This new span is seasonally adjusted as before, with the same tables retained. This process continues until the end of the data is reached. The beginning and ending dates of the spans are such that the last observation in the original data is also the last observation in the last span. This is discussed in more detail in the following paragraphs.

The following notation for the components or differences computed in the sliding spans analysis follows Findley et al. (1990). The meaning for the symbol $X_t(k)$ is component $X$ in month (or quarter) $t$, computed from data in the $k$th span. These components are now defined.

- Seasonal Factors (Table D10): $S_t(k)$
- Trading-Day Factors (Table C18): $TD_t(k)$
- Seasonally Adjusted Data (Table D11): $SA_t(k)$
Computational Details for Sliding Spans Analysis

**Month-to-Month Changes in the Seasonally Adjusted Data:** $MM_t(k)$

**Year-to-Year Changes in the Seasonally Adjusted Data:** $YY_t(k)$

The key measure is the maximum percent difference across spans. For example, consider a series that begins in January 1972, ends in December 1984, and has four spans, each of length 8 years (see Figure 1 in Findley et al. (1990), p. 346). Consider $S_t(k)$ the seasonal factor (Table D10) for month $t$ for span $k$, and let $N_t$ denote the number of spans containing month $t$; that is,

$$N_t = \{ k : \text{span } k \text{ contains month } t \}$$

In the middle years of the series there is overlap of all four spans, and $N_t$ will be 4. The last year of the series will have only one span, while the beginning can have 1 or 0 spans depending on the original length.

Since we are interested in how much the seasonal factors vary for a given month across the spans, a natural quantity to consider is

$$\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)$$

In the case of the multiplicative model, it is useful to compute a percentage difference; define the maximum percentage difference (MPD) at time $t$ as

$$MPD_t = \frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)}$$

The seasonal factor for month $t$ is then unreliable if $MPD_t$ is large. While no exact significance level can be computed for this statistic, empirical levels have been established by considering over 500 economic series (Findley et al. 1990; Findley and Monsell 1986). For these series it was found that for four spans, stable series typically had less than 15% of the MPD values exceeding 3.0%, while in marginally stable series, between 15% and 25% of the MPD values exceeded 3.0%. A series in which 25% or more of the MPD values exceeded 3.0% is almost always unstable.

While these empirical values cannot be considered an exact significance level, they provide a useful empirical basis for deciding if a series is suitable for seasonal adjustment. These percentage values are shifted down when fewer than four spans are used.

---

**Computational Details for Sliding Spans Analysis**

**Length and Number of Spans**

The algorithm for determining the length and number of spans for a given series was developed at the U.S. Bureau of the Census, Statistical Research Division. A summary of this algorithm is as follows.

First, an initial length based on the MACURVE month=option specification is determined, and then the maximum number of spans possible using this length is determined. If this maximum number exceeds four, set the number of spans to four. If this maximum number is one or zero, there are not enough observations to perform the sliding spans analysis. In this case a note is written to the log and the sliding spans analysis is skipped for this variable.
If the maximum number of spans is two or three, the actual number of spans used is set equal to this maximum. Finally, the length is adjusted so that the spans begin in January (or the first quarter) of the beginning year of the span.

The remainder of this section gives the computation formulas for the maximum percentage difference (MPD) calculations along with the threshold regions.

**Seasonal Factors (Table D10)**

For the additive model, the MPD is defined as

$$\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)$$

For the multiplicative model, the MPD is

$$MPD_t = \frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)}$$

A series for which less than 15% of the MPD values of D10 exceed 3.0% is stable; between 15% and 25% is marginally stable; and greater than 25% is unstable. Span reports S 2.A through S 2.C give the various breakdowns for the number of times the MPD exceeded these levels.

**Trading Day Factor (Table C18)**

For the additive model, the MPD is defined as

$$\max_{k \in N_t} TD_t(k) - \min_{k \in N_t} TD_t(k)$$

For the multiplicative model, the MPD is

$$MPD_t = \frac{\max_{k \in N_t} TD_t(k) - \min_{k \in N_t} TD_t(k)}{\min_{k \in N_t} TD_t(k)}$$

The U.S. Census Bureau currently gives no recommendation concerning MPD thresholds for the trading-day factors. Span reports S 3.A through S 3.C give the various breakdowns for MPD thresholds. When TDREGR=NONE is specified, no trading-day computations are done, and this table is skipped.

**Seasonally Adjusted Data (Table D11)**

For the additive model, the MPD is defined as

$$\max_{k \in N_t} SA_t(k) - \min_{k \in N_t} SA_t(k)$$

For the multiplicative model, the MPD is

$$MPD_t = \frac{\max_{k \in N_t} SA_t(k) - \min_{k \in N_t} SA_t(k)}{\min_{k \in N_t} SA_t(k)}$$

A series for which less than 15% of the MPD values of D11 exceed 3.0% is stable; between 15% and 25% is marginally stable; and greater than 25% is unstable. Span reports S 4.A through S 4.C give the various breakdowns for the number of times the MPD exceeded these levels.
Month-to-Month Changes in the Seasonally Adjusted Data

Some additional notation is needed for the month-to-month and year-to-year differences. Define $N_{1t}$ as

$$N_{1t} = \{k : \text{span } k \text{ contains month } t \text{ and } t - 1\}$$

For the additive model, the month-to-month change for span $k$ is defined as

$$MM_t(k) = SA_t - SA_{t-1}$$

while for the multiplicative model

$$MM_t(k) = \frac{SA_t - SA_{t-1}}{SA_{t-1}}$$

Since this quantity is already in percentage form, the MPD for both the additive and multiplicative model is defined as

$$MPD_t = max_{k \in N_{1t}} MM_t(k) - min_{k \in N_{1t}} MM_t(k)$$

The current recommendation of the U.S. Census Bureau is that if 35% or more of the MPD values of the month-to-month differences of D11 exceed 3.0%, then the series is usually not stable; 40% exceeding this level clearly marks an unstable series. Span reports S 5.A.1 through S 5.C give the various breakdowns for the number of times the MPD exceeds these levels.

Year-to-Year Changes in the Seasonally Adjusted Data

First define $N_{12t}$ as

$$N_{12t} = \{k : \text{span } k \text{ contains month } t \text{ and } t - 12\}$$

(Appropriate changes in notation for a quarterly series are obvious.)

For the additive model, the month-to-month change for span $k$ is defined as

$$YY_t(k) = SA_t - SA_{t-12}$$

while for the multiplicative model

$$YY_t(k) = \frac{SA_t - SA_{t-12}}{SA_{t-12}}$$

Since this quantity is already in percentage form, the MPD for both the additive and multiplicative model is defined as

$$MPD_t = max_{k \in N_{12t}} YY_t(k) - min_{k \in N_{12t}} YY_t(k)$$

The current recommendation of the U.S. Census Bureau is that if 10% or more of the MPD values of the month-to-month differences of D11 exceed 3.0%, then the series is usually not stable. Span reports S 6.A through S 6.C give the various breakdowns for the number of times the MPD exceeds these levels.
Data Requirements

The input data set must contain either quarterly or monthly time series, and the data must be in chronological order. For the standard X-11 method, there must be at least three years of observations (12 for quarterly time series or 36 for monthly) in the input data sets or in each BY group in the input data set if a BY statement is used.

For the X-11-ARIMA method, there must be at least five years of observations (20 for quarterly time series or 60 for monthly) in the input data sets or in each BY group in the input data set if a BY statement is used.

Missing Values

Missing values at the beginning of a series to be adjusted are skipped. Processing starts with the first nonmissing value and continues until the end of the series or until another missing value is found.

Missing values are not allowed for the DATE= variable. The procedure terminates if missing values are found for this variable.

Missing values found in the PMFACTOR= variable are replaced by 100 for the multiplicative model (default) and by 0 for the additive model.

Missing values can occur in the output data set. If the time series specified in the OUTPUT statement is not computed by the procedure, the values of the corresponding variable are missing. If the time series specified in the OUTPUT statement is a moving average, the values of the corresponding variable are missing for the first \( n \) and last \( n \) observations, where \( n \) depends on the length of the moving average. Additionally, if the time series specified is an irregular component modified for extremes, only the modified values are given, and the remaining values are missing.

Prior Daily Weights and Trading-Day Regression

Suppose that a detailed examination of retail sales at ZXY Company indicates that certain days of the week have higher amounts of sales. In particular, Thursday, Friday, and Saturday have approximately twice the amount of sales as Monday, Tuesday, and Wednesday, and no sales occur on Sunday. This means that months with five Saturdays would have higher amounts of sales than months with only four Saturdays.

This phenomenon is called a calendar effect; it can be handled in PROC X11 by using the PDWEIGHTS (prior daily weights) statement or the TDREGR=option (trading-day regression). The PDWEIGHTS statement and the TDREGR=option can be used separately or together.

If the relative weights are known (as in the preceding) it is appropriate to use the PDWEIGHTS statement. If further residual calendar variation is present, TDREGR=ADJUST should also be used. If you know that a calendar effect is present, but know nothing about the relative weights, use TDREGR=ADJUST without a PDWEIGHTS statement.

In this example, it is assumed that the calendar variation is due to both prior daily weights and residual variation. Thus both a PDWEIGHTS statement and TDREGR=ADJUST are specified.
Note that only the relative weights are needed; in the actual computations, PROC X11 normalizes the weights to sum to 7.0. If a day of the week is not present in the PDWEIGHTS statement, it is given a value of zero. Thus “sun=0” is not needed.

```plaintext
proc x11 data=sales;
   monthly date=date tdregr=adjust;
   var sales;
   tables a1 a4 b15 b16 c14 c15 c18 d11;
   pdweights mon=1 tue=1 wed=1 thu=2 fri=2 sat=2;
   output out=x11out a1=a1 a4=a4 b1=b1 c14=c14
        c16=c16 c18=c18 d11=d11;
run;
```

Tables of interest include A1, A4, B15, B16, C14, C15, C18, and D11. Table A4 contains the adjustment factors derived from the prior daily weights; Table C14 contains the extreme irregular values excluded from trading-day regression; Table C15 contains the trading-day-regression results; Table C16 contains the monthly factors derived from the trading-day regression; and Table C18 contains the final trading-day factors derived from the combined daily weights. Finally, Table D11 contains the final seasonally adjusted series.

### Adjustment for Prior Factors

Suppose now that a strike at ZXY Company during July and August of 1988 caused sales to decrease an estimated 50%. Since this is a one-time event with a known cause, it is appropriate to prior adjust the data to reflect the effects of the strike. This is done in PROC X11 through the use of PMFACTOR=varname (prior monthly factor) in the MONTHLY statement.

In the following example, the PMFACTOR variable is named PMF. Since the estimate of the decrease in sales is 50%, PMF has a value of 50.0 for the observations corresponding to July and August 1988, and a value of 100.0 for the remaining observations.

This prior adjustment on SALES is performed by replacing SALES with the calculated value (SALES/PMF) * 100.0. A value of 100.0 for PMF leaves SALES unchanged, while a value of 50.0 for PMF doubles SALES. This value is the estimate of what SALES would have been without the strike. The following example shows how this prior adjustment is accomplished.

```plaintext
data sales2;
   set sales;
   if '01jul1988'd <= date <= '01aug1988'd then pmf = 50;
   else pmf = 100;
run;

proc x11 data=sales2;
   monthly date=date pmfactor=pmf;
   var sales;
   tables a1 a2 a3 d11;
   output out=x11out a1=a1 a2=a2 a3=a3 d11=d11;
run;
```

Table A2 contains the prior monthly factors (the values of PMF), and Table A3 contains the prior adjusted series.
The **YRAHEADOUT Option**

For monthly data, the YRAHEADOUT option affects only Tables C16 (regression trading-day adjustment factors), C18 (trading-day factors from combined daily weights), and D10 (seasonal factors). For quarterly data, only Table D10 is affected. Variables for all other tables have missing values for the forecast observations. The forecast values for a table are included only if that table is specified in the OUTPUT statement.

Tables C16 and C18 are calendar effects that are extrapolated by calendar composition. These factors are independent of the data once trading-day weights have been calculated. Table D10 is extrapolated by a linear combination of past values. If \( N \) is the total number of nonmissing observations for the analysis variable, this linear combination is given by

\[
D10_t = \frac{1}{2}(3 \times D10_{t-12} - D10_{t-24}), \quad t = N + 1, \ldots, N + 12
\]

If the input data are monthly time series, 12 extra observations are added to the end of the output data set. (If a BY statement is used, 12 extra observations are added to the end of each BY group.) If the input data are a quarterly time series, four extra observations are added to the end of the output data set. (If a BY statement is used, four extra observations are added to each BY group.)

The DATE= variable (or _DATE_) is extrapolated for the extra observations generated by the YRAHEADOUT option, while all other ID variables will have missing values.

If ARIMA processing is requested, and if both the OUTEXTRAP and YRAHEADOUT options are specified in the PROC X11 statement, an additional 12 (or 4) observations are added to the end of output data set for monthly (or quarterly) data after the ARIMA forecasts, using the same linear combination of past values as before.

---

**Effect of Backcast and Forecast Length**

Based on a number of empirical studies (Dagum 1982a, b, c; Dagum and Laniel 1987), one year of forecasts minimize revisions when new data become available. Two and three years of forecasts show only small gains.

Backcasting improves seasonal adjustment but introduces permanent revisions at the beginning of the series and also at the end for series of length 8, 9, or 10 years. For series shorter than 7 years, the advantages of backcasting outweigh the disadvantages (Dagum 1988).

Other studies (Pierce 1980; Bobbit and Otto 1990; Buszuwski 1987) suggest “full forecasting”—that is, using enough forecasts to allow symmetric weights for the seasonal moving averages for the most current data. For example, if a \( 3 \times 9 \) seasonal moving average was specified for one or more months by using the MACURVES statement, five years of forecasts would be required. This is because the seasonal moving averages are performed on calendar months separately, and the \( 3 \times 9 \) is an 11-term centered moving average, requiring five observations before and after the current observation. Thus

```
macurves dec='3x9';
```

would require five additional December values to compute the seasonal moving average.
Details of Model Selection

If an ARIMA statement is present but no MODEL= is given, PROC X11 estimates and forecasts five predefined models and selects the best. This section describes the details of the selection criteria and the selection process.

The five predefined models used by PROC X11 are the same as those used by X11ARIMA/88 from Statistics Canada. These particular models, shown in Table 43.2, were chosen on the basis of testing a large number of economics series (Dagum 1988) and should provide reasonable forecasts for most economic series.

The five predefined models used by PROC X11 are the same as those used by X11ARIMA/88 from Statistics Canada. These particular models, shown in Table 43.2, were chosen on the basis of testing a large number of economics series (Dagum 1988) and should provide reasonable forecasts for most economic series.

Table 43.2 Five Predefined Models

<table>
<thead>
<tr>
<th>Model #</th>
<th>Specification</th>
<th>Multiplicative</th>
<th>Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1,1)(0,1,1)s</td>
<td>log transform</td>
<td>no transform</td>
</tr>
<tr>
<td>2</td>
<td>(0,1,2)(0,1,1)s</td>
<td>log transform</td>
<td>no transform</td>
</tr>
<tr>
<td>3</td>
<td>(2,1,0)(0,1,1)s</td>
<td>log transform</td>
<td>no transform</td>
</tr>
<tr>
<td>4</td>
<td>(0,2,2)(0,1,1)s</td>
<td>log transform</td>
<td>no transform</td>
</tr>
<tr>
<td>5</td>
<td>(2,1,2)(0,1,1)s</td>
<td>no transform</td>
<td>no transform</td>
</tr>
</tbody>
</table>

The selection process proceeds as follows. The five models are estimated and one-step-ahead forecasts are produced in the order shown in Table 43.2. As each model is estimated, the following three criteria are checked:

- The mean absolute percent error (MAPE) for the last three years of the series must be less than 15%.
- The significance probability for the Box-Ljung chi-square for up to lag 24 for monthly (8 for quarterly) must greater than 0.05.
- The over-differencing criteria must not exceed 0.9.

The descriptions of these three criteria are given in the section “Criteria Details” on page 3118. The default values for these criteria are those used by X11ARIMA/88 from Statistics Canada; these defaults can be changed by the MAPECR=, CHICR=, and OVDIFCR= options.

A model that fails any one of these three criteria is excluded from further consideration. In addition, if the ARIMA estimation fails for a given model, a warning is issued, and the model is excluded. The final set of all models considered consists of those that pass all three criteria and are estimated successfully. From this set, the model with the smallest MAPE for the last three years is chosen.

If all five models fail, ARIMA processing is skipped for the variable being processed, and the standard X-11 seasonal adjustment is performed. A note is written to the log with this information.

The chosen model is then used to forecast the series one or more years (determined by the FORECAST= option in the ARIMA statement). These forecasts are appended to the original data (or the prior and calendar-adjusted data).

If a BACKCAST= option is specified, the chosen model form is used, but the parameters are reestimated using the reversed series. Using these parameters, the reversed series is forecast for the number of years specified by the BACKCAST= option. These forecasts are then reversed and appended to the beginning of the original series, or the prior and calendar-adjusted series, to produce the backcasts.
Note that the final selection rule (the smallest MAPE using the last three years) emphasizes the quality of the forecasts at the end of the series. This is consistent with the purpose of the X-11-ARIMA methodology, which is to improve the estimates of seasonal factors and thus minimize revisions to recent past data as new data become available.

**Criteria Details**

**Mean Absolute Percent Error (MAPE)**

For the MAPE criteria testing, only the last three years of the original series (or prior and calendar adjusted series) is used in computing the MAPE.

Let $y_t$, $t = 1, \ldots, n$, be the last three years of the series, and denote its one-step-ahead forecast by $\hat{y}_t$, where $n = 36$ for a monthly series and $n = 12$ for a quarterly series.

With this notation, the MAPE criteria are computed as

$$ MAPE = \frac{100}{n} \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{|y_t|} $$

**Box-Ljung Chi-Square**

The Box-Ljung chi-square is a lack-of-fit test based on the model residuals. This test statistic is computed using the Ljung-Box formula

$$ \chi^2_m = n(n + 2) \sum_{k=1}^{m} \frac{r_k^2}{(n-k)} $$

where $n$ is the number of residuals that can be computed for the time series, and

$$ r_k = \frac{\sum_{t=1}^{n-k} a_t a_{t+k}}{\sum_{t=1}^{n} a_t^2} $$

where the $a_t$'s are the residual sequence. This formula has been suggested by Ljung and Box (1978) as yielding a better fit to the asymptotic chi-square distribution. Some simulation studies of the finite sample properties of this statistic are given by Davies, Triggs, and Newbold (1977) and by Ljung and Box (1978).

For monthly series, $m = 24$, while for quarterly series, $m = 8$.

**Over-Differencing Test**

From Table 43.2 you can see that all models have a single seasonal MA factor and at most two nonseasonal MA factors. Also, all models have seasonal and nonseasonal differencing. Consider model 2 applied to a monthly series $y_t$ with $E(y_t) = \mu$:

$$(1 - B^1)(1 - B^{12})(y_t - \mu) = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_3 B^{12})a_t$$

If $\theta_3 = 1.0$, then the factors $(1 - \theta_3 B^{12})$ and $(1 - B^{12})$ will cancel, resulting in a lower-order model.

Similarly, if $\theta_1 + \theta_2 = 1.0$,

$$(1 - \theta_1 B - \theta_2 B^2) = (1 - B)(1 - \alpha B)$$

for some $\alpha \neq 0.0$. Again, this results in cancellation and a lower-order model.
Since the parameters are not exact, it is not reasonable to require that
\[ \theta_3 < 1.0 \text{ and } \theta_1 + \theta_2 < 1.0 \]

Instead, an approximate test is performed by requiring that
\[ \theta_3 \leq 0.9 \text{ and } \theta_1 + \theta_2 \leq 0.9 \]

The default value of 0.9 can be changed by the OVDIFCR= option. Similar reasoning applies to the other models.

**ARIMA Statement Options for the Five Predefined Models**

Table 43.3 lists the five predefined models and gives the equivalent MODEL= parameters in a PROC X11 ARIMA statement.

In all models except the fifth, a log transformation is performed before the ARIMA estimation for the multiplicative case; no transformation is performed for the additive case. For the fifth model, no transformation is done for either case.

The multiplicative case is assumed in the following table. The indicated seasonality \( s \) in the specification is either 12 (monthly) or 4 (quarterly). The MODEL statement assumes a monthly series.

<table>
<thead>
<tr>
<th>Model</th>
<th>ARIMA Statement Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0,1,1)(0,1,1)s)</td>
<td>MODEL=(Q=1 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG</td>
</tr>
<tr>
<td>((0,1,2)(0,1,1)s)</td>
<td>MODEL=(Q=2 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG</td>
</tr>
<tr>
<td>((2,1,0)(0,1,1)s)</td>
<td>MODEL=(P=2 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG</td>
</tr>
<tr>
<td>((0,2,2)(0,1,1)s)</td>
<td>MODEL=(Q=2 SQ=1 DIF=2 SDIF=1) TRANSFORM=LOG</td>
</tr>
<tr>
<td>((2,1,2)(0,1,1)s)</td>
<td>MODEL=(P=2 Q=2 SQ=1 DIF=1 SDIF=1)</td>
</tr>
</tbody>
</table>

**OUT= Data Set**

The OUT= data set specified in the OUTPUT statement contains the BY variables, if any; the ID variables, if any; and the DATE= variable if the DATE= option is given, or _DATE_ if the DATE= option is not specified.

In addition, the variables specified by the option

\[ \text{tablename =var1 var2 \ldots varn} \]

are placed in the OUT= data set. A list of tables available for monthly and quarterly series is given later, in Table 43.4.
The OUTSPAN= Data Set

The OUTSPAN= option is specified in the PROC statement, and writes the sliding spans results to the specified output data set. The OUTSPAN= data set contains the following variables:

- **A1**, a numeric variable that is a copy of the original series truncated to the current span. Note that overlapping spans will contain identical values for this variable.
- **C18**, a numeric variable that contains the trading-day factors for the seasonal adjustment for the current span.
- **D10**, a numeric variable that contains the seasonal factors for the seasonal adjustment for the current span.
- **D11**, a numeric variable that contains the seasonally adjusted series for the current span.
- **DATE**, a numeric variable that contains the date within the current span.
- **SPAN**, a numeric variable that contains the current span. The first span is the earliest span—that is the one with the earliest starting date.
- **VARNAMES**, a character variable containing the name of each variable in the VAR list. A separate sliding spans analysis is performed on each variable in the VAR list.

OUTSTB= Data Set

The output data set produced by the OUTSTB= option of the PROC X11 statement contains the information in the analysis of variance on table D8 (Final Unmodified S-I Ratios). This analysis of variance, following table D8 in the printed output, tests for stable seasonality (Shiskin, Young, and Musgrave 1967, Appendix A). These data contain the following variables:

- **VARNAMES**, a character variable containing the name of each variable in the VAR list.
- **TABLE**, a character variable specifying the table from which the analysis of variance is performed. When ARIMA processing is requested, and two passes of X11 are required (when TDREGR=PRINT, TEST, or ADJUST), Table D8 and the stable seasonality test are computed twice: once in the initial pass, then again in the final pass. Both of these computations are put in the OUTSTB data set and are identified by D18.1 and D18.2, respectively.
- **SOURCE**, a character variable corresponding to the “source” column in the analysis of variance table following Table D8.
- **SS**, a numeric variable containing the sum of squares associated with the corresponding source term.
- **DF**, a numeric variable containing the degrees of freedom associated with the corresponding source term.
• MS, a numeric variable containing the mean square associated with the corresponding source term. MS is missing for the source term “Total”

• F, a numeric variable containing the $F$ statistic for the “Between” source term. F is missing for all other source terms.

• PROBF, a numeric variable containing the significance level for the $F$ statistic. PROBF is missing for the source terms “Total” and “Error.”

---

**OUTTDR= Data Set**

The trading-day regression results (tables B15 and C15) are written to the OUTTDR= data set, which contains the following variables:

• **VARNAME**, a character variable containing the name of the VAR variable being processed

• **TABLE**, a character variable containing the name of the table. It can have only the value B15 (Preliminary Trading-Day Regression) or C15 (Final Trading-Day Regression).

• **_TYPE_**, a character variable whose value distinguishes the three distinct table format types. These types are (a) the regression, (b) the listing of the standard error associated with length-of-month, and (c) the analysis of variance. The first seven observations in the OUTTDR data set correspond to the regression on days of the week; thus the **_TYPE_** variable is given the value “REGRESS” (day-of-week regression coefficient). The next four observations correspond to 31-, 30-, 29-, and 28-day months and are given the value **_TYPE_=LOM_STD** (length-of-month standard errors). Finally, the last three observations correspond to the analysis of variance table, and **_TYPE_=ANOVA**.

• **PARM**, a character variable, further identifying the nature of the observation. PARM is set to blank for the three **_TYPE_=ANOVA** observations.

• **SOURCE**, a character variable containing the source in the regression. This variable is missing for all **_TYPE_=REGRESS** and **_TYPE_=LOM_STD**.

• **CWGT**, a numeric variable containing the combined trading-day weight (prior weight + weight found from regression). The variable is missing for all **_TYPE_=LOM_STD** and **_TYPE_=ANOV A**.

• **PRWGT**, a numeric variable containing the prior weight. The prior weight is 1.0 if PDWEIGHTS are not specified. This variable is missing for all **_TYPE_=LOM_STD** and **_TYPE_=ANOV A**.

• **COEFF**, a numeric variable containing the calculated regression coefficient for the given day. This variable is missing for all **_TYPE_=LOM_STD** and **_TYPE_=ANOV A**.

• **STDERR**, a numeric variable containing the standard errors. For observations with **_TYPE_=REGRESS**, this is the standard error corresponding to the regression coefficient. For observations with **_TYPE_=LOM_STD**, this is standard error for the corresponding length-of-month. This variable is missing for all **_TYPE_=ANOV A**.

• **T1**, a numeric variable containing the $t$ statistic corresponding to the test that the combined weight is different from the prior weight. This variable is missing for all **_TYPE_=LOM_STD** and **_TYPE_=ANOV A**.
• T2, a numeric variable containing the \( t \) statistic corresponding to the test that the combined weight is different from 1.0. This variable is missing for all _TYPE_=LOM_STD and _TYPE_=ANOVA.

• PROBT1, a numeric variable containing the significance level for \( t \) statistic T1. The variable is missing for all _TYPE_=LOM_STD and _TYPE_=ANOVA.

• PROBT2, a numeric variable containing the significance level for \( t \) statistic T2. The variable is missing for all _TYPE_=LOM_STD and _TYPE_=ANOVA.

• SS, a numeric variable containing the sum of squares associated with the corresponding source term. This variable is missing for all _TYPE_=REGRESS and LOM_STD.

• DF, a numeric variable containing the degrees of freedom associated with the corresponding source term. This variable is missing for all _TYPE_=REGRESS and LOM_STD.

• MS, a numeric variable containing the mean square associated with the corresponding source term. This variable is missing for the source term ‘Total’ and for all _TYPE_=REGRESS and LOM_STD.

• F, a numeric variable containing the \( F \) statistic for the ‘Regression’ source term. The variable is missing for the source terms ‘Total’ and ‘Error’, and for all _TYPE_=REGRESS and LOM_STD.

• PROBF, a numeric variable containing the significance level for the \( F \) statistic. This variable is missing for the source term ‘Total’ and ‘Error’ and for all _TYPE_=REGRESS and LOM_STD.

**Printed Output**

The output from PROC X11, both printed tables and the series written to the OUT= data set, depends on whether the data are monthly or quarterly. For the printed tables, the output depends further on the value of the PRINTOUT= option and the TABLE statement, along with other options specified.

The printed output is organized into tables identified by a part letter and a sequence number within the part. The seven major parts of the X11 procedure are as follows:

- A prior adjustments (optional)
- B preliminary estimates of irregular component weights and regression trading-day factors
- C final estimates of irregular component weights and regression trading-day factors
- D final estimates of seasonal, trend cycle, and irregular components
- E analytical tables
- F summary measures
- G charts

*Table 43.4* describes the individual tables and charts. Most tables apply both to quarterly and monthly series. Those that apply only to a monthly time series are indicated by an “M” in the notes section, while “P” indicates the table is not a time series, and is only printed, not output to the OUT= data set.
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>original series</td>
<td>M</td>
</tr>
<tr>
<td>A2</td>
<td>prior monthly adjustment factors</td>
<td>M</td>
</tr>
<tr>
<td>A3</td>
<td>original series adjusted for prior monthly factors</td>
<td>M</td>
</tr>
<tr>
<td>A4</td>
<td>prior trading-day adjustments</td>
<td>M</td>
</tr>
<tr>
<td>A5</td>
<td>prior adjusted or original series</td>
<td>M</td>
</tr>
<tr>
<td>A13</td>
<td>ARIMA forecasts</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>ARIMA backcasts</td>
<td></td>
</tr>
<tr>
<td>A15</td>
<td>prior adjusted or original series extended by ARIMA backcasts and forecasts</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>prior adjusted or original series</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>unmodified seasonal-irregular (S-I) ratios</td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>replacement values for extreme S-I ratios</td>
<td></td>
</tr>
<tr>
<td>B5</td>
<td>seasonal factors</td>
<td></td>
</tr>
<tr>
<td>B6</td>
<td>seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>B7</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>B8</td>
<td>unmodified S-I ratios</td>
<td></td>
</tr>
<tr>
<td>B9</td>
<td>replacement values for extreme S-I ratios</td>
<td></td>
</tr>
<tr>
<td>B10</td>
<td>seasonal factors</td>
<td></td>
</tr>
<tr>
<td>B11</td>
<td>seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>B13</td>
<td>irregular series</td>
<td></td>
</tr>
<tr>
<td>B14</td>
<td>extreme irregular values excluded from trading-day regression</td>
<td>M</td>
</tr>
<tr>
<td>B15</td>
<td>preliminary trading-day regression</td>
<td>M,P</td>
</tr>
<tr>
<td>B16</td>
<td>trading-day adjustment factors</td>
<td>M</td>
</tr>
<tr>
<td>B17</td>
<td>preliminary weights for irregular components</td>
<td></td>
</tr>
<tr>
<td>B18</td>
<td>trading-day factors derived from combined daily weights</td>
<td>M</td>
</tr>
<tr>
<td>B19</td>
<td>original series adjusted for trading-day and prior variation</td>
<td>M</td>
</tr>
<tr>
<td>C1</td>
<td>original series modified by preliminary weights and adjusted for trading-day and prior variation</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>modified S-I ratios</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>seasonal factors</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>modified S-I ratios</td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>seasonal factors</td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>irregular series</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>extreme irregular values excluded from trading-day regression</td>
<td>M</td>
</tr>
<tr>
<td>C15</td>
<td>final trading-day regression</td>
<td>M,P</td>
</tr>
<tr>
<td>C16</td>
<td>final trading-day adjustment factors derived from regression coefficients</td>
<td>M</td>
</tr>
<tr>
<td>C17</td>
<td>final weight for irregular components</td>
<td></td>
</tr>
</tbody>
</table>
Table 43.4  continued

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C18</td>
<td>final trading-day factors derived from combined daily weights</td>
<td>M</td>
</tr>
<tr>
<td>C19</td>
<td>original series adjusted for trading-day and prior variation</td>
<td>M</td>
</tr>
<tr>
<td>D1</td>
<td>original series modified for final weights and adjusted for trading-day and prior variation</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>D4</td>
<td>modified S-I ratios</td>
<td></td>
</tr>
<tr>
<td>D5</td>
<td>seasonal factors</td>
<td></td>
</tr>
<tr>
<td>D6</td>
<td>seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>D7</td>
<td>trend cycle</td>
<td></td>
</tr>
<tr>
<td>D8</td>
<td>final unmodified S-I ratios</td>
<td></td>
</tr>
<tr>
<td>D9</td>
<td>final replacement values for extreme S-I ratios</td>
<td></td>
</tr>
<tr>
<td>D10</td>
<td>final seasonal factors</td>
<td></td>
</tr>
<tr>
<td>D11</td>
<td>final seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>D12</td>
<td>final trend cycle</td>
<td></td>
</tr>
<tr>
<td>D13</td>
<td>final irregular series</td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>original series with outliers replaced</td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>modified seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>modified irregular series</td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>ratios of annual totals</td>
<td>P</td>
</tr>
<tr>
<td>E5</td>
<td>percent changes in original series</td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>percent changes in final seasonally adjusted series</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>MCD moving average</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>summary measures</td>
<td>P</td>
</tr>
<tr>
<td>G1</td>
<td>chart of final seasonally adjusted series and trend cycle</td>
<td>P</td>
</tr>
<tr>
<td>G2</td>
<td>chart of S-I ratios with extremes, S-I ratios without extremes, and final seasonal factors</td>
<td>P</td>
</tr>
<tr>
<td>G3</td>
<td>chart of S-I ratios with extremes, S-I ratios without extremes, and final seasonal factors in calendar order</td>
<td>P</td>
</tr>
<tr>
<td>G4</td>
<td>chart of final irregular and final modified irregular series</td>
<td>P</td>
</tr>
</tbody>
</table>

The PRINTOUT= Option

The PRINTOUT= option controls printing for groups of tables. See the “TABLES Statement” on page 3106 for details on specifying individual tables. The following list gives the tables printed for each value of the PRINTOUT= option:

- STANDARD (26 tables) A1–A4, B1, C13–C19, D8–D13, E1–E6, F1, F2
- LONG (40 tables) A1–A5, A13–A15, B1, B2, B7, B10, B13–B15, C1, C7, C10, C13–C19, D1, D7–D11, D13, E1–E6, F1, F2
- FULL (62 tables) A1–A5, A13–A15, B1–B11, B13–B19, C1–C11, C13–C19, D1, D2, D4–D12, E1–E6, F1, F2
The actual number of tables printed depends on the options and statements specified. If a table is not computed, it is not printed. For example, if TDREGR\(=\)NONE is specified, none of the tables associated with the trading-day are printed.

**The CHARTS= Option**

Of the four charts listed in Table 43.4, G1 and G2 are printed by default (CHARTS=STANDARD). Charts G3 and G4 are printed when CHARTS=FULL is specified. See the “TABLES Statement” on page 3106 for details on specifying individual charts.

**Stable, Moving, and Combined Seasonality Tests on the Final Unmodified SI Ratios (Table D8)**

PROC X11 displays four tests used to identify stable seasonality and moving seasonality and to measure identifiable seasonality. These tests are displayed after Table D8. They are “Stable Seasonality Test,” “Moving Seasonality Test,” “Nonparametric Test for the Presence of Seasonality Assuming Stability,” and “Summary of Results and Combined Test for the Presence of Identifiable Seasonality.” The motivation, interpretation, and statistical details of all these tests are now given.

**Motivation**

The seasonal component of this time series, \(S_t\), is defined as the intrayear variation that is repeated constantly (stable) or in an evolving fashion from year to year (moving seasonality). If the increase in the seasonal factors from year to year is too large, then the seasonal factors will introduce distortion into the model. It is important to determine if seasonality is identifiable without distorting the series.

To determine if stable seasonality is present in a series, PROC X11 computes a one-way analysis of variance by using the seasons (months or quarters) as the factor on the Final Unmodified SI Ratios (Table D8). This is the appropriate table to use because the removal of the trend cycle is equivalent to detrending. PROC X11 prints this test, labeled “Stable Seasonality Test,” immediately after the Table D8.

The X11 seasonal adjustment method tests for moving seasonality. Moving seasonality can be a source of distortion when seasonal factors are used in the model. PROC X11 computes and prints a test for moving seasonality. The test is a two-way analysis of variance that uses months (or quarters) and years. As in the “Stable Seasonality Test,” this analysis of variance is performed on the Final Unmodified SI Ratios (Table D8). PROC X11 prints this test, labeled “Moving Seasonality Test,” after the “Stable Seasonality Test.”

PROC X11 next computes a nonparametric Kruskal-Wallis chi-squared test for stable seasonality. “Nonparametric Test for the Presence of Seasonality Assuming Stability.” The Kruskal-Wallis test is performed on the ranks of the Final Unmodified SI Ratios (Table D8). For further details about the Kruskal-Wallis test, see Lehmann and D’Abrera (2006, pp. 204–210).

The results of the preceding three tests are combined into a joint test to measure identifiable seasonality, “Summary of Results and Combined Test for the Presence of Identifiable Seasonality.” This test combines the two \(F\) tests previously described, along with the Kruskal-Wallis chi-squared test for stable seasonality, to determine “identifiable” seasonality. This test is printed after “Nonparametric Test for the Presence of Seasonality Assuming Stability.”

**Interpretation and Statistical Details**

The “Stable Seasonality Test” is a one-way analysis of variance on the “Final Unmodified SI Ratios” with seasons (months or quarters) as the factor.
To determine whether stable seasonality is present in a series, PROC X11 computes a one-way analysis of variance by using the seasons (months or quarters) as the factor on the Final Unmodified SI Ratios (Table D8). This is the appropriate table to use because the removal of the trend cycle is similar to detrending.

A large $F$ statistic and a small significance level are evidence that a significant amount of variation in the SI-ratios is due to months or quarters, which in turn is evidence of seasonality; the null hypothesis of no month/quarter effect is rejected.

Conversely, a small $F$ statistic and a large significance level (close to 1.0) are evidence that variation due to month or quarter could be due to random error, and the null hypothesis of no month/quarter effect is not rejected. The interpretation and utility of seasonal adjustment are problematic under such conditions.

The $F$ test for moving seasonality is performed by a two-way analysis of variance. The two factors are seasons (months or quarters) and years. The years effect is tested separately; the null hypothesis is no effect due to years after accounting for variation due to months or quarters. For further details about the moving seasonality test, see Lothian (1984a, b, 1978) and Higginson (1975).

The significance level reported in both the moving and stable seasonality tests are only approximate. Table D8, the Final Unmodified SI Ratios, is constructed from an averaging operation that induces a correlation in the residuals from which the $F$ test is computed. Hence the computed $F$ statistic differs from an exact $F$ statistic; see Cleveland and Devlin (1980) for details.

The test for identifiable seasonality is performed by combining the $F$ tests for stable and moving seasonality, along with a Kruskal-Wallis test for stable seasonality. The following description is based on Lothian and Morry (1978b); other details can be found in Dagum (1988, 1983).

Let $F_s$ and $F_m$ denote the $F$ value for the stable and moving seasonality tests, respectively. The combined test is performed as shown in Table 43.5 and as follows:

1. If the null hypothesis of no stable seasonality is not rejected at the 0.10% significance level ($P_S \geq 0.001$), then the series is considered to be nonseasonal. PROC X11 returns the conclusion, “Identifiable Seasonality Not Present.”

2. If the null hypothesis in step 1 is rejected, then PROC X11 computes the following quantities:

$$T_1 = \frac{7}{F_s}$$
$$T_2 = \frac{3F_m}{F_s}$$

Let $T$ denote the simple average of $T_1$ and $T_2$:

$$T = \frac{(T_1 + T_2)}{2}$$

If the null hypothesis of no moving seasonality is rejected at the 5.0% significance level ($P_M < 0.05$) and if $T \geq 1.0$, the null hypothesis of identifiable seasonality not present is not rejected and PROC X11 returns the conclusion, “Identifiable Seasonality Not Present.”

3. If the null hypothesis of identifiable seasonality not present has not been accepted, but $T_1 \geq 1.0$, $T_2 \geq 1.0$, or the Kruskal-Wallis chi-squared test fails to reject at the 0.10% significance level ($P_{KW} \geq 0.001$), then PROC X11 returns the conclusion “Identifiable Seasonality Probably Not Present.”
4. If the null hypotheses of no stable seasonality associated with the $F_S$ and Kruskal-Wallis chi-squared tests are rejected and if none of the combined measures described in steps 2 and 3 fail, then the null hypothesis of identifiable seasonality not present is rejected and PROC X11 returns the conclusion “Identifiable Seasonality Present.”

**Figure 43.5 Combined Seasonality Test Flowchart**
Tables Written to the OUT= Data Set

All tables that are time series can be written to the OUT= data set. However, depending on the specified options and statements, not all tables are computed. When a table is not computed, but is requested in the OUTPUT statement, the resulting variable has all missing values.

For example, if the PMFACTOR= option is not specified, Table A2 is not computed, and requesting this table in the OUTPUT statement results in the corresponding variable having all missing values.

The trading-day regression results, Tables B15 and C15, although not written to the OUT= data set, can be written to an output data set; see the OUTTDR= option for details.

Printed Output Generated by Sliding Spans Analysis

Table S 0.A
Table S 0.A gives the variable name, the length and number of spans, and the beginning and ending dates of each span.

Table S 0.B
Table S 0.B gives the summary of the two F tests performed during the standard X11 seasonal adjustments for stable and moving seasonality on Table D8, the final SI ratios. These tests are described in the section “Printed Output” on page 3122.

Table S 1.A
Table S 1.A gives the range analysis of seasonal factors. This includes the means for each month (or quarter) within a span, the maximum percentage difference across spans for each month, and the average. The minimum and maximum within a span are also indicated.

For example, for a monthly series and an analysis with four spans, the January row would contain a column for each span, with the value representing the average seasonal factor (Table D10) over all January calendar months occurring within the span. Beside each span column is a character column with either a MIN, MAX, or blank value, indicating which calendar month had the minimum and maximum value over that span.

Denote the average over the jth calendar month in span \( k; k = 1, \ldots, 4 \), by \( \bar{S}_j(k) \); then the maximum percent difference (MPD) for month \( j \) is defined by

\[
MPD_j = \frac{\max_{k=1,\ldots,4} \bar{S}_j(k) - \min_{k=1,\ldots,4} \bar{S}_j(k)}{\min_{k=1,\ldots,4} \bar{S}_j(k)}
\]

The last numeric column of Table S 1.A is the average value over all spans for each calendar month, with the minimum and maximum row flagged as in the span columns.

Table S 1.B
Table S 1.B gives a summary of range measures for each span. The first column, Range Means, is calculated by computing the maximum and minimum over all months or quarters in a span, then taking the difference. The next column is the range ratio means, which is simply the ratio of the previously described maximum and minimum. The next two columns are the minimum and maximum seasonal factors over the entire span, while the range sf column is the difference of these. Finally, the last column is the ratio of the Max SF and Min SF columns.
Breakdown Tables

Table S 2.A.1 begins the breakdown analysis for the various series considered in the sliding spans analysis. The key concept here is the MPD described above in the section “Table S 1.A” on page 3128 and in the section “Computational Details for Sliding Spans Analysis” on page 3111. For a month or quarter that appears in two or more spans, the maximum percentage difference is computed and tested against a cutoff level. If it exceeds this cutoff, it is counted as an instance of exceeding the level. It is of interest to see if such instances fall disproportionately in certain months and years. Tables S 2.A.1 through S 6.A.3 display this breakdown for all series considered.

Table S 2.A.1

Table S 2.A.1 gives the monthly (quarterly) breakdown for the seasonal factors (table D10). The first column identifies the month or quarter. The next column is the number of times the MPD for D10 exceeded 3.0%, followed by the total count. The last is the average maximum percentage difference for the corresponding month or quarter.

Table S 2.A.2

Table S 2.A.2 gives the same information as Table S 2.A.1, but on a yearly basis.

Table S 2.A.3

The description of Table S 2.A.3 requires the definition of “Sign Change” and “Turning Point.”

First, some motivation. Recall that for a highly stable series, adding or deleting a small number of observations should not affect the estimation of the various components of a seasonal adjustment procedure.

Consider Table D10, the seasonal factors in a sliding spans analysis that uses four spans. For a given observation $t$, looking across the four spans, we can easily pick out large differences if they occur. More subtle differences can occur when estimates go from above to below (or vice versa) a base level. In the case of multiplicative model, the seasonal factors have a base level of 100.0. So it is useful to enumerate those instances where both a large change occurs (an MPD value exceeding 3.0%) and a change of sign (with respect to the base) occur.

Let $B$ denote the base value (which in general depends on the component being considered and the model type, multiplicative or additive). If, for span 1, $S_t(1)$ is below $B$ (i.e., $S_t(1) - B$ is negative) and for some subsequent span $k$, $S_t(k)$ is above $B$ (i.e., $S_t(k) - B$ is positive), then a positive “Change in Sign” has occurred at observation $t$. Similarly, if, for span 1, $S_t(1)$ is above $B$, and for some subsequent span $k$, $S_t(k)$ is below $B$, then a negative “Change in Sign” has occurred. Both cases, positive or negative, constitute a “Change in Sign”; the actual direction is indicated in tables S 7.A through S 7.E, which are described below.

Another behavior of interest occurs when component estimates increase then decrease (or vice versa) across spans for a given observation. Using the preceding example, the seasonal factors at observation $t$ could first increase, then decrease across the four spans.

This behavior, combined with an MPD exceeding the level, is of interest in questions of stability.

Again, consider Table D10, the seasonal factors in a sliding spans analysis that uses four spans. For a given observation $t$ (containing at least three spans), note the level of D10 for the first span. Continue across the spans until a difference of 1.0% or greater occurs (or no more spans are left), noting whether the difference is up or down. If the difference is up, continue until a difference of 1.0% or greater occurs downward (or no more spans are left). If such an up-down combination occurs, the observation is counted as an up-down turning point. A similar description occurs for a down-up turning point. Tables S 7.A through S 7.E, described
below, show the occurrence of turning points, indicating whether up-down or down-up. Note that it requires at least three spans to test for a turning point. Hence Tables S 2.A.3 through S 6.A.3 show a reduced number in the “Turning Point” row for the “Total Tested” column, and in Tables S 7.A through S 7.E, the turning points symbols can occur only where three or more spans overlap.

With these descriptions of sign change and turning point, we now describe Table S 2.A.3. The first column gives the type or category, the second column gives the total number of observations falling into the category, the third column gives the total number tested, and the last column gives the percentage for the number found in the category.

The first category (row) of the table is for flagged observations—that is, those observations where the MPD exceeded the appropriate cutoff level (3.0% is default for the seasonal factors). The second category is for level changes, while the third category is for turning points. The fourth category is for flagged sign changes—that is, for those observations that are sign changes, how many are also flagged. Note the total tested column for this category equals the number found for sign change, reflecting the definition of the fourth category.

The fifth column is for flagged turning points—that is, for those observations that are turning points, how many are also flagged.

The footnote to Table S 2.A.3 gives the U.S. Census Bureau recommendation for thresholds, as described in the section “Computational Details for Sliding Spans Analysis” on page 3111.

Table S 2.B
Table S 2.B gives the histogram of flagged for seasonal factors (Table D10) using the appropriate cutoff value (default 3.0%). This table looks at the spread of the number of times the MPD exceeded the corresponding level. The range is divided up into four intervals: 3.0%–4.0%, 4.0%–5.0%, 5.0%–6.0%, and greater than 6.0%. The first column shows the symbol used in Table S 7.A, the second column gives the range in interval notation, and the last column gives the number found in the corresponding interval. Note that the sum of the last column should agree with the “Number Found” column of the “Flagged MPD” row in Table S 2.A.3.

Table S 2.C
Table S 2.C gives selected percentiles for the MPD for the seasonal factors (Table D10).

Tables S 3.A.1 through S 3.A.3
These table relate to the trading-day factors (Table C18) and follow the same format as Tables S 2.A.1 through S 2.A.3. The only difference between these tables and Tables S 2.A.1 through S 2.A.3 is the default cutoff value of 2.0% instead of the 3.0% used for the seasonal factors.

Tables S 3.B, S 3.C
These tables, applied to the trading-day factors (Table C18), are the same format as Tables S 2.B through S 2.C. The default cutoff value is different, with corresponding differences in the intervals in S 3.B.

Tables S 4.A.1 through S 4.A.3
These tables relate to the seasonally adjusted series (Table D11) and follow the same format as Tables S 2.A.1 through S 2.A.3. The same default cutoff value of 3.0% is used.
Tables S 4.B, S 4.C
These tables, applied to the seasonally adjusted series (Table D11), are the same format as tables S 2.B through S 2.C.

Tables S 5.A.1 through S 5.A.3
These tables relate to the month-to-month (or quarter-to-quarter) differences in the seasonally adjusted series, and follow the same format as Tables S 2.A.1 through S 2.A.3. The same default cutoff value of 3.0% is used.

Tables S 5.B, S 5.C
These tables, applied to the month-to-month (or quarter-to-quarter) differences in the seasonally adjusted series, are the same format as tables S 2.B through S 2.C. The same default cutoff value of 3.0% is used.

Tables S 6.A.1 through S 6.A.3
These tables relate to the year-to-year differences in the seasonally adjusted series, and follow the same format as Tables S 2.A.1 through S 2.A.3. The same default cutoff value of 3.0% is used.

These tables, applied to the year-to-year differences in the seasonally adjusted series, are the same format as tables S 2.B through S 2.C. The same default cutoff value of 3.0% is used.

Table S 7.A
Table S 7.A gives the entire listing of the seasonal factors (Table D10) for each span. The first column gives the date for each observation included in the spans. Note that the dates do not cover the entire original data set. Only those observations included in one or more spans are listed.

The next $N$ columns (where $N$ is the number of spans) are the individual spans starting at the earliest span. The span columns are labeled by their beginning and ending dates.

Following the last span is the “Sign Change” column. As explained in the description of Table S 2.A.3, a sign change occurs at a given observation when the seasonal factor estimates go from above to below, or below to above, a base level. For the seasonal factors, 100.0 is the base level for the multiplicative model, 0.0 for the additive model. A blank value indicates no sign change, a “U” indicates a movement “upward” from the base level and a “D” indicates a movement “downward” from the base level.

The next column is the “Turning Point” column. As explained in the description of Table S 2.A.3, a turning point occurs when there is an upward then downward movement, or downward then upward movement, of sufficient magnitude. A blank value indicates no turning point, a “U-D” indicates a movement “upward then downward,” and a “D-U” indicates a movement “downward then upward.”

The next column is the maximum percentage difference (MPD). This quantity, described in the section “Computational Details for Sliding Spans Analysis” on page 3111, is the main computation for sliding spans analysis. A measure of how extreme the MPD value is given in the last column, the “Level of Excess” column. The symbols used and their meaning are described in Table S 2.A.3. If a given observation has exceeded the cutoff, the level of excess column is blank.
Table S 7.B
Table S 7.B gives the entire listing of the trading-day factors (Table C18) for each span. The format of this table is exactly like that of Table S 7.A.

Table S 7.C
Table S 7.C gives the entire listing of the seasonally adjusted data (Table D11) for each span. The format of this table is exactly like that of Table S 7.A except for the “Sign Change” column, which is not printed. The seasonally adjusted data have the same units as the original data; there is no natural base level as in the case of a percentage. Hence the sign change is not appropriate for D11.

Table S 7.D
Table S 7.D gives the entire listing of the month-to-month (or quarter-to-quarter) changes in seasonally adjusted data for each span. The format of this table is exactly like that of Table S 7.A.

Table S 7.E
Table S 7.E gives the entire listing of the year-to-year changes in seasonally adjusted data for each span. The format of this table is exactly like that of Table S 7.A.

Printed Output from the ARIMA Statement
The information printed by default for the ARIMA model includes the parameter estimates, their approximate standard errors, t ratios, and variances, the standard deviation of the error term, and the AIC and SBC statistics for the model. In addition, a criteria summary for the chosen model is given that shows the values for each of the three test criteria and the corresponding critical values.

If the PRINTALL option is specified, a summary of the nonlinear estimation optimization and a table of Box-Ljung statistics is also produced. If the automatic model selection is used, this information is printed for each of the five predefined models. Finally, a model selection summary is printed, showing the final model chosen.

ODS Table Names
PROC X11 assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table.

NOTE: For monthly and quarterly tables, use the ODS names MonthlyTables and QuarterlyTables; For brevity, only the MonthlyTables are listed here; the QuarterlyTables are simply duplicates. Printing of individual tables can be specified by using the TABLES table_name, which is not listed here. Printing groups of tables is specified in the MONTHLY and QUARTERLY statements by specifying the option PRINTOUT=NONE|STANDARD|LONG|FULL. The default is PRINTOUT=STANDARD.
## ODS Table Names

### Table 43.5  ODS Tables Produced in PROC X11

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the MONTHLY and QUARTERLY Statements</strong></td>
<td></td>
</tr>
<tr>
<td>Preface</td>
<td>X11 Seasonal Adjustment Program information giving credits, dates, and so on</td>
</tr>
<tr>
<td>A1</td>
<td>Table A1: original series</td>
</tr>
<tr>
<td>A2</td>
<td>Table A2: prior monthly</td>
</tr>
<tr>
<td>A3</td>
<td>Table A3: original series adjusted for prior monthly factors</td>
</tr>
<tr>
<td>A4</td>
<td>Table A4: prior trading day adjustment factors with and without length of month adjustment</td>
</tr>
<tr>
<td>A5</td>
<td>Table A5: original series adjusted for priors</td>
</tr>
<tr>
<td>B1</td>
<td>Table B1: original series or original series adjusted for priors</td>
</tr>
<tr>
<td>B2</td>
<td>Table B2: trend cycle—centered nn-term moving average</td>
</tr>
<tr>
<td>B3</td>
<td>Table B3: unmodified SI ratios</td>
</tr>
<tr>
<td>B4</td>
<td>Table B4: replacement values for extreme SI ratios</td>
</tr>
<tr>
<td>B5</td>
<td>Table B5: seasonal factors</td>
</tr>
<tr>
<td>B6</td>
<td>Table B6: seasonally adjusted series</td>
</tr>
<tr>
<td>B7</td>
<td>Table B7: trend cycle—Henderson curve</td>
</tr>
<tr>
<td>B8</td>
<td>Table B8: unmodified SI ratios</td>
</tr>
<tr>
<td>B9</td>
<td>Table B9: replacement values for extreme SI ratios</td>
</tr>
<tr>
<td>B10</td>
<td>Table B10: seasonal factors</td>
</tr>
<tr>
<td>B11</td>
<td>Table B11: seasonally adjusted series</td>
</tr>
<tr>
<td>B13</td>
<td>Table B13: irregular series</td>
</tr>
<tr>
<td>B15</td>
<td>Table B15: preliminary trading day regression</td>
</tr>
<tr>
<td>B16</td>
<td>Table B16: trading day adjustment factors derived from regression</td>
</tr>
<tr>
<td>B17</td>
<td>Table B17: preliminary weights for irregular component</td>
</tr>
<tr>
<td>B18</td>
<td>Table B18: trading day adjustment factors from combined weights</td>
</tr>
<tr>
<td>B19</td>
<td>Table B19: original series adjusted for preliminary combined trading day weights</td>
</tr>
<tr>
<td>C1</td>
<td>Table C1: original series adjusted for preliminary weights</td>
</tr>
<tr>
<td>C2</td>
<td>Table C2: trend cycle—centered nn-term moving average</td>
</tr>
<tr>
<td>ODS Table Name</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>C4</td>
<td>Table C4: modified SI ratios</td>
</tr>
<tr>
<td>C5</td>
<td>Table C5: seasonal factors</td>
</tr>
<tr>
<td>C6</td>
<td>Table C6: seasonally adjusted series</td>
</tr>
<tr>
<td>C7</td>
<td>Table C7 trend cycle—Henderson curve</td>
</tr>
<tr>
<td>C9</td>
<td>Table C9: modified SI ratios</td>
</tr>
<tr>
<td>C10</td>
<td>Table C10: seasonal factors</td>
</tr>
<tr>
<td>C11</td>
<td>Table C11: seasonally adjusted series</td>
</tr>
<tr>
<td>C13</td>
<td>Table C13: irregular series</td>
</tr>
<tr>
<td>C15</td>
<td>Table C15: final trading day regression</td>
</tr>
<tr>
<td>C16</td>
<td>Table C16: trading day adjustment factors derived from regression</td>
</tr>
<tr>
<td>C17</td>
<td>Table C17: final weights for irregular component</td>
</tr>
<tr>
<td>C18</td>
<td>Table C18: trading day adjustment factors from combined weights</td>
</tr>
<tr>
<td>C19</td>
<td>Table C19: original series adjusted for final combined trading day weights</td>
</tr>
<tr>
<td>D1</td>
<td>Table D1: original series adjusted for final weights nn-term moving average</td>
</tr>
<tr>
<td>D4</td>
<td>Table D4: modified SI ratios</td>
</tr>
<tr>
<td>D5</td>
<td>Table D5: seasonal factors</td>
</tr>
<tr>
<td>D6</td>
<td>Table D6: seasonally adjusted series</td>
</tr>
<tr>
<td>D7</td>
<td>Table D7: trend cycle—Henderson curve</td>
</tr>
<tr>
<td>D8</td>
<td>Table D8: final unmodified SI ratios</td>
</tr>
<tr>
<td>D10</td>
<td>Table D10: final seasonal factors</td>
</tr>
<tr>
<td>D11</td>
<td>Table D11: final seasonally adjusted series</td>
</tr>
<tr>
<td>D12</td>
<td>Table D12: final trend cycle—Henderson curve</td>
</tr>
<tr>
<td>D13</td>
<td>Table D13: final irregular series</td>
</tr>
<tr>
<td>E1</td>
<td>Table E1: original series modified for extremes</td>
</tr>
<tr>
<td>E2</td>
<td>Table E2: modified seasonally adjusted series</td>
</tr>
<tr>
<td>E3</td>
<td>Table E3: modified irregular series</td>
</tr>
<tr>
<td>E5</td>
<td>Table E5: month-to-month changes in original series</td>
</tr>
<tr>
<td>E6</td>
<td>Table E6: month-to-month changes in final seasonally adjusted series</td>
</tr>
<tr>
<td>F1</td>
<td>Table F1: MCD moving average</td>
</tr>
<tr>
<td>A13</td>
<td>Table A13: ARIMA forecasts</td>
</tr>
<tr>
<td>A14</td>
<td>Table A14: ARIMA backcasts</td>
</tr>
<tr>
<td>A15</td>
<td>Table A15: ARIMA extrapolation</td>
</tr>
<tr>
<td>ODS Table Name</td>
<td>Description</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>B14</td>
<td>Table B14: irregular values excluded from trading day regression</td>
</tr>
<tr>
<td>C14</td>
<td>Table C14: irregular values excluded from trading day regression</td>
</tr>
<tr>
<td>D9</td>
<td>Table D9: final replacement values</td>
</tr>
<tr>
<td>PriorDailyWgts</td>
<td>adjusted prior daily weights</td>
</tr>
<tr>
<td>TDR_0</td>
<td>final/preliminary trading day regression, part 1</td>
</tr>
<tr>
<td>TDR_1</td>
<td>final/preliminary trading day regression, part 2</td>
</tr>
<tr>
<td>StandErrors</td>
<td>standard errors of trading day adjustment factors</td>
</tr>
<tr>
<td>D9A</td>
<td>year-to-year change in irregular and seasonal components and moving seasonality ratio</td>
</tr>
<tr>
<td>StableSeasTest</td>
<td>stable seasonality test</td>
</tr>
<tr>
<td>StableSeasFTest</td>
<td>moving seasonality test</td>
</tr>
<tr>
<td>KruskalWallisTest</td>
<td>nonparametric test for the presence of seasonality assuming stability</td>
</tr>
<tr>
<td>CombinedSeasonalityTest</td>
<td>summary of results and combined test for the presence of identifiable seasonality</td>
</tr>
<tr>
<td>f2a</td>
<td>F2 summary measures, part 1</td>
</tr>
<tr>
<td>f2b</td>
<td>F2 summary measures, part 2</td>
</tr>
<tr>
<td>f2c</td>
<td>F2 summary measures, part 3</td>
</tr>
<tr>
<td>f2d</td>
<td>I/C ratio for monthly/quarterly span</td>
</tr>
<tr>
<td>f2f</td>
<td>average % change with regard to sign and standard deviation over span</td>
</tr>
<tr>
<td>E4</td>
<td>differences or ratios of annual totals for original and adjusted series</td>
</tr>
<tr>
<td>ChartG1</td>
<td>chart G1</td>
</tr>
<tr>
<td>ChartG2</td>
<td>chart G2</td>
</tr>
</tbody>
</table>
### Table 43.5 continued

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the ARIMA Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CriteriaSummary</td>
<td>criteria summary</td>
<td>ARIMA statement</td>
</tr>
<tr>
<td>ConvergeSummary</td>
<td>convergence summary</td>
<td></td>
</tr>
<tr>
<td>ArimaEst</td>
<td>ARIMA estimation results, part 1</td>
<td></td>
</tr>
<tr>
<td>ArimaEst2</td>
<td>ARIMA estimation results, part 2</td>
<td></td>
</tr>
<tr>
<td>Model_Summary</td>
<td>model summary</td>
<td></td>
</tr>
<tr>
<td>Ljung_BoxQ</td>
<td>table of Ljung-Box Q statistics</td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>Table A13: ARIMA forecasts</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>Table A14: ARIMA backcasts</td>
<td></td>
</tr>
<tr>
<td>A15</td>
<td>Table A15: ARIMA extrapolation</td>
<td></td>
</tr>
<tr>
<td><strong>ODS Tables Created by the SSPAN Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPR0A_1</td>
<td>S 0.A sliding spans analysis, number, length of spans</td>
<td>default printing</td>
</tr>
<tr>
<td>SpanDates</td>
<td>S 0.A sliding spans analysis: dates of spans</td>
<td></td>
</tr>
<tr>
<td>SPR0B</td>
<td>S 0.B summary of F tests for stable and moving seasonality</td>
<td></td>
</tr>
<tr>
<td>SPR1_1</td>
<td>S 1.A range analysis of seasonal factors</td>
<td></td>
</tr>
<tr>
<td>SPR1_b</td>
<td>S 1.B summary of range measures</td>
<td></td>
</tr>
<tr>
<td>SPRXA</td>
<td>2XA.1 breakdown of differences by month or quarter</td>
<td></td>
</tr>
<tr>
<td>SPRXB_2</td>
<td>S X.B histogram of flagged observations</td>
<td></td>
</tr>
<tr>
<td>SPRXA_2</td>
<td>S X.A.2 breakdown of differences by year</td>
<td></td>
</tr>
<tr>
<td>MpdStats</td>
<td>S X.C: statistics for maximum percentage differences</td>
<td></td>
</tr>
<tr>
<td>S_X_A_3</td>
<td>S 2.X.3 breakdown summary of flagged observations</td>
<td></td>
</tr>
<tr>
<td>SPR7_X</td>
<td>S 7.X sliding spans analysis</td>
<td>PRINTALL</td>
</tr>
</tbody>
</table>
Example 43.1: Component Estimation—Monthly Data

This example computes and plots the final estimates of the individual components for a monthly series. In the first plot (Output 43.1.1), an overlaid plot of the original and seasonally adjusted data is produced. The trend in the data is more evident in the seasonally adjusted data than in the original data. This trend is even more clear in Output 43.1.3, the plot of Table D12, the trend cycle. Note that both the seasonal factors and the irregular factors vary around 100, while the trend cycle and the seasonally adjusted data are in the scale of the original data.

From Output 43.1.2 the seasonal component appears to be slowly increasing, while no apparent pattern exists for the irregular series in Output 43.1.4.

data sales;
  input sales @@;
  date = intnx( 'month', '01sep1978'd, _n_-1 );
  format date monyy7.;
datalines;
  112 118 132 129 121 135 148 148 136 119 104 118
  ... more lines ...

proc x11 data=sales noprint;
  monthly date=date;
  var sales;
  tables b1 d11;
  output out=out b1=series d10=d10 d11=d11
d12=d12 d13=d13;
run;

title 'Monthly Retail Sales Data (in $1000)';
proc sgplot data=out;
  series x=date y=series / markers
    markerattrs=(color=red symbol='asterisk')
    lineattrs=(color=red)
    legendlabel="original" ;
  series x=date y=d11 / markers
    markerattrs=(color=blue symbol='circle')
    lineattrs=(color=blue)
    legendlabel="adjusted" ;
yaxis label='Original and Seasonally Adjusted Time Series';
run;
Output 43.1.1 Plots of Original and Seasonally Adjusted Data

Monthly Seasonal Factors (in percent)

```
proc sgplot data=out;
   series x=date y=d10 / markers markerattrs=(symbol=CircleFilled) ;
run;
```

```
proc sgplot data=out;
   series x=date y=d12 / markers markerattrs=(symbol=CircleFilled) ;
run;
```

```
proc sgplot data=out;
   series x=date y=d13 / markers markerattrs=(symbol=CircleFilled) ;
run;
```
Output 43.1.2 Plot of D10, the Final Seasonal Factors
Output 43.1.3 Plot of D12, the Final Trend Cycle

Monthly Retail Sales Data (in $1000)
**Example 43.2: Components Estimation—Quarterly Data**

This example is similar to Example 43.1, except quarterly data are used. Tables B1, the original series, and D11, the final seasonally adjusted series, are printed by the TABLES statement. The OUTPUT statement writes the listed tables to an output data set.

```plaintext
data quarter;
  input date yyq6. +1 fy35rr 5.2;
  format date yyq6.;
datalines;
1971Q1 6.59
... more lines ...
```

---

**Output 43.1.4** Plot of D13, the Final Irregular Series

*Monthly Irregular Factors (in percent)*
title 'Monthly Retail Sales Data (in $1000)';
proc x11 data=quarter;
   var fy35rr;
   quarterly date=date;
   tables b1 d11;
   output out=out b1=b1 d10=d10 d11=d11 d12=d12 d13=d13;
run;

Output 43.2.1 X11 Procedure Quarterly Example

Monthly Retail Sales Data (in $1000)

The X11 Procedure

Seasonal Adjustment of - fy35rr

X-11 Seasonal Adjustment Program
U. S. Bureau of the Census
Economic Research and Analysis Division
November 1, 1968

The X-11 program is divided into seven major parts.

Part Description
A. Prior adjustments, if any
B. Preliminary estimates of irregular component weights and regression trading day factors
C. Final estimates of above
D. Final estimates of seasonal, trend-cycle and irregular components
E. Analytical tables
F. Summary measures
G. Charts

Series - fy35rr
Period covered - 1st Quarter 1971 to 4th Quarter 1976

<table>
<thead>
<tr>
<th>Year</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>5.520</td>
<td>5.590</td>
<td>5.840</td>
<td>6.330</td>
<td>23.280</td>
</tr>
<tr>
<td>1974</td>
<td>9.910</td>
<td>11.150</td>
<td>12.400</td>
<td>11.640</td>
<td>45.100</td>
</tr>
<tr>
<td>1975</td>
<td>9.940</td>
<td>8.160</td>
<td>8.220</td>
<td>8.290</td>
<td>34.610</td>
</tr>
<tr>
<td>1976</td>
<td>7.540</td>
<td>7.440</td>
<td>7.800</td>
<td>7.280</td>
<td>30.060</td>
</tr>
<tr>
<td>Avg</td>
<td>7.670</td>
<td>7.617</td>
<td>8.335</td>
<td>8.300</td>
<td></td>
</tr>
</tbody>
</table>

Total: 191.53  Mean: 7.9804  S.D.: 1.9424
Example 43.3: Outlier Detection and Removal

PROC X11 can be used to detect and replace outliers in the irregular component of a monthly or quarterly series.

The weighting scheme used in measuring the “extremeness” of the irregulars is developed iteratively; thus the statistical properties of the outlier adjustment method are unknown.

In this example, the data are simulated by generating a trend plus a random error. Two periods in the series were made “extreme” by multiplying one generated value by 2.0 and another by 0.10. The additive model is appropriate based on the way the data were generated. Note that the trend in the generated data was modeled automatically by the trend cycle component estimation.

The detection of outliers is accomplished by considering Table D9, the final replacement values for extreme S-I ratios. This table indicates which observations had irregular component values more than FULLWEIGHT= standard deviation units from 0.0 (1.0 for the multiplicative model). The default value of the FULLWEIGHT= option is 1.5; a larger value would result in fewer observations being declared extreme.

In this example, FULLWEIGHT=3.0 is used to isolate the extreme inflated and deflated values generated in the DATA step. The value of ZEROWEIGHT= must be greater than FULLWEIGHT; it is given a value of 3.5.

A plot of the original and modified series, Output 43.3.2, shows that the deviation from the trend line for the modified series is greatly reduced compared with the original series.

data a;
  retain seed 99831;
  do kk = 1 to 48;
    x = kk + 100 + rannor( seed );
    date = intnx( 'month', '01jan1970'd, kk-1 );
    if kk = 20 then x = 2 * x;
    else if kk = 30 then x = x / 10;
    output;
  end;
run;
proc x11 data=a;
  monthly date=date additive
    fullweight=3.0 zeroweight=3.5;
  var x;
  table d9;
  output out=b b1=original e1=e1;
run;

proc sgplot data=b;
  series x=date y=original / markers
    markerattrs=(color=red symbol='asterisk')
    lineattrs=(color=red)
    legendlabel="unmodified" ;
  series x=date y=e1 / markers
    markerattrs=(color=blue symbol='circle')
    lineattrs=(color=blue)
    legendlabel="modified" ;
  yaxis label='Original and Outlier Adjusted Time Series';
run;

**Output 43.3.1** Detection of Extreme Irregulars

**Monthly Retail Sales Data (in $1000)**

**The X11 Procedure**

**Seasonal Adjustment of - x**

<table>
<thead>
<tr>
<th>Year</th>
<th>JAN</th>
<th>FEB</th>
<th>MAR</th>
<th>APR</th>
<th>MAY</th>
<th>JUN</th>
<th>JUL</th>
<th>AUG</th>
<th>SEP</th>
<th>OCT</th>
<th>NOV</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1971</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>11.180</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1972</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-10.671</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1973</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
Output 43.3.2  Plot of Modified and Unmodified Values

Monthly Retail Sales Data (in $1000)

References


Subject Index

BY groups
X11 procedure, 3097

Census X-11 method, see X11 procedure
Census X-11 methodology
X11 procedure, 3109
combined seasonality test, 3125

data requirements
X11 procedure, 3114
differences with X11ARIMA/88
X11 procedure, 3108

ID variables
X11 procedure, 3097, 3099
irregular component
X11 procedure, 3086, 3091

Kruskal-Wallis test, 3125

model selection for X-11-ARIMA method
X11 procedure, 3117
moving seasonality test, 3125

output data sets
X11 procedure, 3119–3121
output table names
X11 procedure, 3132

printed output
X11 procedure, 3122

seasonal adjustment
time series data, 3086
X11 procedure, 3086, 3091
seasonal component
X11 procedure, 3086
seasonality tests, 3125
sliding spans analysis, 3110
stable seasonality test, 3125

time series data
seasonal adjustment, 3086
trading-day component
X11 procedure, 3086, 3091
trend cycle component
X11 procedure, 3086, 3091

X-11 ARIMA methodology
X11 procedure, 3108

X-11 seasonal adjustment method, see X11 procedure
X-11-ARIMA seasonal adjustment method, see X11 procedure

X11 procedure
BY groups, 3097
Census X-11 method, 3086
Census X-11 methodology, 3109
data requirements, 3114
differences with X11ARIMA/88, 3108
ID variables, 3097, 3099
irregular component, 3086, 3091
model selection for X-11-ARIMA method, 3117
output data sets, 3119–3121
output table names, 3132
printed output, 3122
seasonal adjustment, 3086, 3091
seasonal component, 3086
trading-day component, 3086, 3091
trend cycle component, 3086, 3091
X-11 ARIMA methodology, 3108
X-11 seasonal adjustment method, 3086
X-11-ARIMA seasonal adjustment method, 3086
Syntax Index

**ADDITIVE** option
- MONTHLY statement (X11), 3098
- QUARTERLY statement (X11), 3103

**ARIMA statement**
- X11 procedure, 3094

**BACKCAST=** option
- ARIMA statement (X11), 3094

**BY statement**
- X11 procedure, 3097

**CENTER** option
- ARIMA statement (X11), 3096

**CHARTS=** option
- MONTHLY statement (X11), 3098
- QUARTERLY statement (X11), 3103

**CHICR=** option
- ARIMA statement (X11), 3095

**CONVERGE=** option
- ARIMA statement (X11), 3095

**CUTOFF=** option
- SSPAN statement (X11), 3105

**DATA=** option
- PROC X11 statement, 3093

**DATE=** option
- MONTHLY statement (X11), 3099
- QUARTERLY statement (X11), 3103

**END=** option
- MONTHLY statement (X11), 3099
- QUARTERLY statement (X11), 3104

**EXCLUDE=** option
- MONTHLY statement (X11), 3099

**FORECAST=** option
- ARIMA statement (X11), 3095

**FULLWEIGHT=** option
- MONTHLY statement (X11), 3099
- QUARTERLY statement (X11), 3104

**ID statement**
- X11 procedure, 3097

**LENGTH** option
- MONTHLY statement (X11), 3099

**MACURVES statement**
- X11 procedure, 3097

**MAPECR=** option
- ARIMA statement (X11), 3095

**MAXITER=** option
- ARIMA statement (X11), 3095

**METHOD=** option
- ARIMA statement (X11), 3095

**MODEL=** option
- ARIMA statement (X11), 3095

**MONTHLY statement**
- X11 procedure, 3098

**NDEC=** option
- MONTHLY statement (X11), 3099
- QUARTERLY statement (X11), 3104
- SSPAN statement (X11), 3105

**NOINT** option
- ARIMA statement (X11), 3096

**NOPRINT** option
- ARIMA statement (X11), 3096
- PROC X11 statement, 3094
- SSPAN statement (X11), 3105

**OUT=** option
- OUTPUT statement (X11), 3102, 3119
- OUTEXTRAP option
- PROC X11 statement, 3093

**OUTPUT statement**
- X11 procedure, 3101

**OUTSPAN=** option
- PROC X11 statement, 3094, 3120
- VAR statement (X11), 3120

**OUTSTB=** option
- PROC X11 statement, 3094, 3120

**OUTTDR=** option
- PROC X11 statement, 3094, 3121

**OVDIFCR=** option
- ARIMA statement (X11), 3096

**PDWEIGHTS statement**
- X11 procedure, 3102

**PRINTER** option
- SSPAN statement (X11), 3106

**PRINTALL option**
- ARIMA statement (X11), 3096
- SSPAN statement (X11), 3106

**PRINTFP option**
- ARIMA statement (X11), 3096

**PRINTOUT=** option
MONTHLY statement (X11), 3100
QUARTERLY statement (X11), 3104
PROC X11 statement, 3093

QUARTERLY statement
X11 procedure, 3103

SSPAN statement
X11 procedure, 3105
START= option
MONTHLY statement (X11), 3100
QUARTERLY statement (X11), 3104
SUMMARY option
MONTHLY statement (X11), 3100
QUARTERLY statement (X11), 3104

TABLES statement
X11 procedure, 3106
TDCOMPUTE= option
MONTHLY statement (X11), 3100
TDCUTOFF= option
SSPAN statement (X11), 3105
TDREGR= option
MONTHLY statement (X11), 3101
TRANSFORM= option
ARIMA statement (X11), 3096
TRENDADJ option
MONTHLY statement (X11), 3101
QUARTERLY statement (X11), 3105
TRENDMA= option
MONTHLY statement (X11), 3101

VAR statement
X11 procedure, 3106

X11 procedure, 3091
syntax, 3091

YRAHEADOUT option
PROC X11 statement, 3094

ZEROWEIGHT= option
MONTHLY statement (X11), 3101
QUARTERLY statement (X11), 3105