THE POWER TO KNOW.

SAS/ETS® 14.3 User's Guide

The X13 Procedure

This document is an individual chapter from *SAS/ETS® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/ETS® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.3 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

September 2017

Chapter 46
The X13 Procedure

Contents

Overview: X13 Procedure

The X13 procedure is an adaptation of the US Bureau of the Census X-13ARIMA-SEATS seasonal adjustment program (US Bureau of the Census 2013c). The X-13ARIMA-SEATS program was developed by the Time Series Staff of the Statistical Research Division, US Census Bureau, by incorporating the SEATS method into the X-12-ARIMA seasonal adjustment program. The X-12-ARIMA seasonal adjustment program contains components developed from Statistics Canada's X-11-ARIMA program (US Bureau of the Census 2010). The X-12-ARIMA automatic modeling method and the SEATS method are based on the work of Gómez and Maravall (1997a, b).

The new X-13ARIMA-SEATS program incorporates the X-12-ARIMA functionality. It also incorporates improvements on X-12-ARIMA methods. Because the X-12-ARIMA methods and improvements are available in X-13ARIMA-SEATS, the new X13 procedure and the existing X12 procedure use the same X-13ARIMA-SEATS methodology, and PROC X12 and PROC X13 are aliases for the same procedure.

The version of PROC X13 documented here was produced by converting the US Census Bureau's FORTRAN code to the SAS development language and adding typical SAS procedure syntax. This conversion work was performed by SAS and resulted in the X13 procedure. Although several features were added during the conversion, credit for the statistical aspects and general methodology of the X13 procedure belongs to the US Census Bureau.

The X13 procedure seasonally adjusts monthly or quarterly time series. The procedure makes additive or multiplicative adjustments and creates an output data set that contains the adjusted time series and intermediate calculations.

The X-13ARIMA-SEATS program includes the X-12-ARIMA program, which combines the capabilities of the X-11 program (Shiskin, Young, and Musgrave 1967) and the X-11-ARIMA/88 program (Dagum 1988) and also introduces some new features (Findley et al. 1998). One of the main enhancements in the X-12-ARIMA program involves the use of a regARIMA model, a regression model with ARIMA (autoregressive integrated moving average) errors. Thus, the X-12-ARIMA program contains methods developed by both the US Census Bureau and Statistics Canada. In addition, the X-12-ARIMA automatic modeling routine is based on the TRAMO (time series regression with ARIMA noise, missing values, and outliers) method (Gómez and Maravall 1997a, b). The four major components of the X-12-ARIMA program are regARIMA modeling, model diagnostics, seasonal adjustment that uses enhanced X-11 methodology, and post-adjustment diagnostics. Statistics Canada's X-11 method fits an ARIMA model to the original series, and then uses the model forecasts 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.

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 consistently or evolves slowly from year to year (Hillmer and Tiao 1982). The trend cycle component, C_t, includes variation that is attributed 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 ($O_t = S_t C_t D_t I_t$). Other economic series are related in an additive fashion ($O_t = S_t + C_t + D_t + I_t$). A seasonally adjusted time series, $C_t I_t$ or $C_t + I_t$, consists of only the trend cycle and irregular components. For more information about the X-11 seasonal adjustment method, see Ladiray and Quenneville (2001).

Graphics are now available with the X13 procedure. For more information, see the section "ODS Graphics" on page 3355.

Getting Started: X13 Procedure

The most common use of the X13 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 X13 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 displayed tables, these latter components are expressed as percentages and thus vary around 100.0 (in the additive case, they vary around 0.0). However, in the output data set, the data displayed as percentages in the displayed output are expressed as the decimal equivalent and thus vary around 1.0 in the multiplicative case.

The naming convention used in PROC X13 for the tables follows the convention used in the Census Bureau's X-13ARIMA-SEATS program; see *X-13ARIMA-SEATS Reference Manual* (US Bureau of the Census 2013c), *X-13ARIMA-SEATS Quick Reference for DOS* (US Bureau of the Census 2013a), and *X-13ARIMA-SEATS Quick Reference for UNIX/Linux* (US Bureau of the Census 2013b). Also see the section "Displayed Output, ODS Table Names, and OUTPUT Tablename Keywords" on page 3350. The table names are outlined in Table 46.15.

The tables that correspond to parts A through 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 need to look only at Table D11, the final seasonally adjusted series. Tables in part E contain information about extreme values and changes in the original and seasonally adjusted series. The tables in part F are seasonal adjustment quality measures. Spectral analysis is performed in part G. For more information about the tables produced by the X11 statement, see Ladiray and Quenneville (2001).

Basic Seasonal Adjustment

Suppose that 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 PROC X13 statement and request seasonal adjustment in the X11 statement as shown in the following statements:

```
data sales;
   set sashelp.air;
   sales = air;
   date = intnx( 'month', '01sep78'd, _n_-1 );
   format date monyy.;
run;

proc x13 data=sales date=date;
   var sales;
   x11;
   ods select d11;
run;
```

The results of the seasonal adjustment are in Table D11 (the final seasonally adjusted series) in the displayed output shown in Figure 46.1.

Figure 46.1 Basic Seasonal Adjustment

The X13 Procedure

Table D 11: Final Seasonally Adjusted Data
For Variable sales

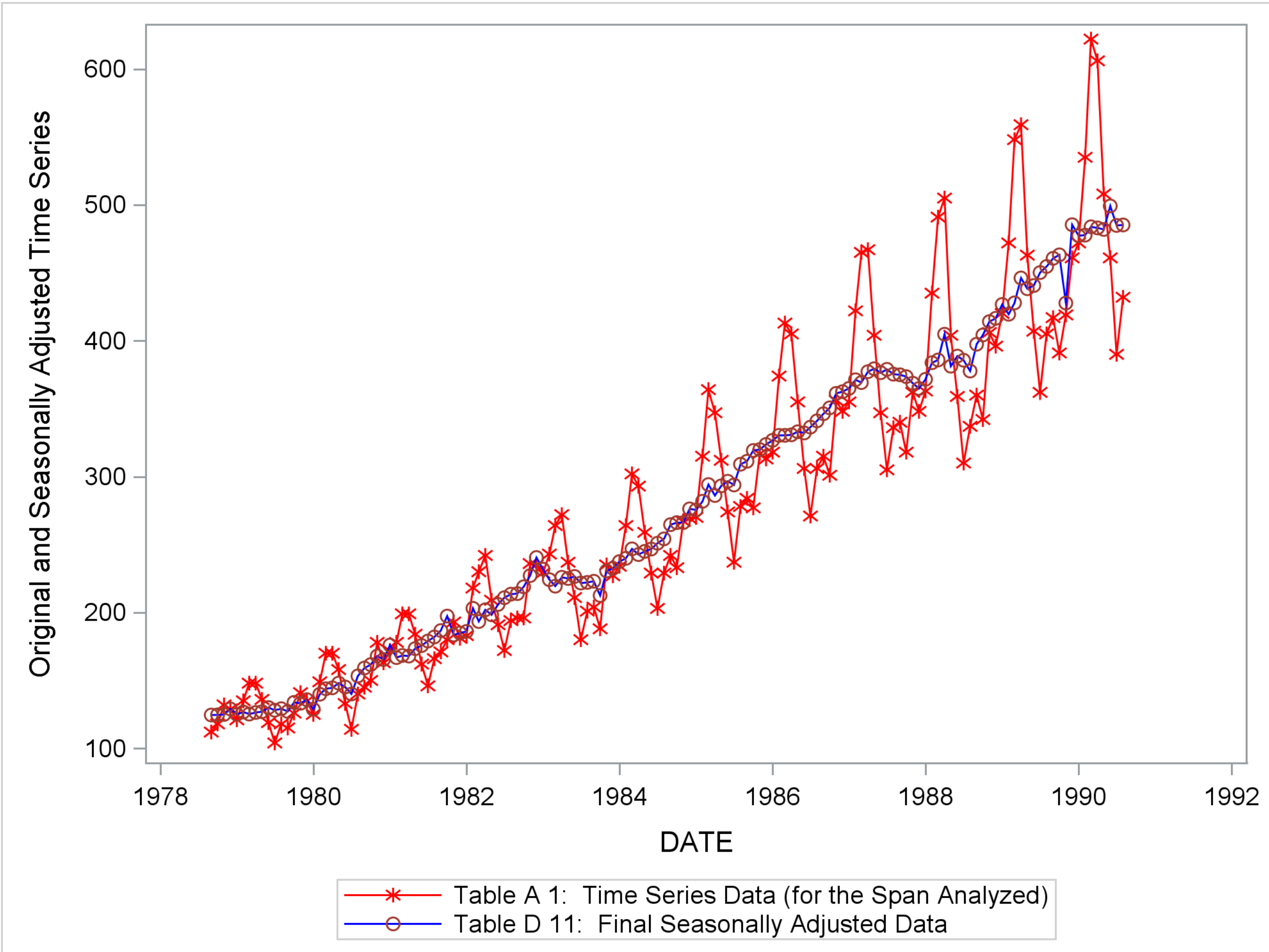
Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Total
1978	.	.	.	.	.	.	.	.	124.560	124.649	124.920	129.002	503.131
1979	125.087	126.759	125.252	126.415	127.012	130.041	128.056	129.165	127.182	133.847	133.199	135.847	1547.86
1980	128.767	139.839	143.883	144.576	148.048	145.170	140.021	153.322	159.128	161.614	167.996	165.388	1797.75
1981	175.984	166.805	168.380	167.913	173.429	175.711	179.012	182.017	186.737	197.367	183.443	184.907	2141.71
1982	186.080	203.099	193.386	201.988	198.322	205.983	210.898	213.516	213.897	218.902	227.172	240.453	2513.69
1983	231.839	224.165	219.411	225.907	225.015	226.535	221.680	222.177	222.959	212.531	230.552	232.565	2695.33
1984	237.477	239.870	246.835	242.642	244.982	246.732	251.023	254.210	264.670	266.120	266.217	276.251	3037.03
1985	275.485	281.826	294.144	286.114	293.192	296.601	293.861	309.102	311.275	319.239	319.936	323.663	3604.44
1986	326.693	330.341	330.383	330.792	333.037	332.134	336.444	341.017	346.256	350.609	361.283	362.519	4081.51
1987	364.951	371.274	369.238	377.242	379.413	376.451	378.930	375.392	374.940	373.612	368.753	364.885	4475.08
1988	371.618	383.842	385.849	404.810	381.270	388.689	385.661	377.706	397.438	404.247	414.084	416.486	4711.70
1989	426.716	419.491	427.869	446.161	438.317	440.639	450.193	454.638	460.644	463.209	427.728	485.386	5340.99
1990	477.259	477.753	483.841	483.056	481.902	499.200	484.893	485.245	.	.	.	.	3873.15
Avg	277.330	280.422	282.373	286.468	285.328	288.657	288.389	291.459	265.807	268.829	268.774	276.446	

Total: 40323 Mean: 280.02 S.D.: 111.31
Min: 124.56 Max: 499.2

You can compare the original series (Table A1) and the final seasonally adjusted series (Table D11) by plotting them together as shown in Figure 46.2. These tables are requested in the OUTPUT statement and are written to the OUT= data set. Note that the default variable name used in the output data set is the input variable name followed by an underscore and the corresponding table name.

```
proc x13 data=sales date=date noprint;
   var sales;
   x11;
   output out=out a1 d11;
run;

proc sgplot data=out;
   series x=date y=sales_A1 / name = "A1" markers
                              markerattrs=(color=red symbol='asterisk')
                              lineattrs=(color=red);
   series x=date y=sales_D11 / name= "D11" markers
                               markerattrs=(symbol='circle')
                               lineattrs=(color=blue);
   yaxis label='Original and Seasonally Adjusted Time Series';
run;
```

Figure 46.2 Plot of Original and Seasonally Adjusted Data



Syntax: X13 Procedure

The X13 procedure uses the following statements:

PROC X13 *options* ;
    VAR *variables* ;
    BY *variables* ;
    ID *variables* ;
    EVENT *variables* < */ options* > ;
    USERDEFINED *variables* ;
    TRANSFORM *options* ;
    ADJUST *option* ;
    IDENTIFY *options* ;
    PICKMDL *options* ;
    AUTOMDL *options* ;
    OUTLIER *options* ;
    REGRESSION *options* ;
    INPUT *variables* < */ options* > ;
    ARIMA *option* ;
    ESTIMATE *options* ;
    X11 *options* ;
    FORECAST *options* ;
    CHECK *options* ;
    SEATSDECOMP OUT= *SAS-data-set* < *options* > ;
    OUTPUT OUT= *SAS-data-set* < **YEARSEAS** > *tablename1 tablename2 . . .* ;
    TABLES *tablename1 tablename2 . . . options* ;

The statements used by PROC X13 perform basically the same function as the Census Bureau's X-13ARIMA-SEATS specs (specifications). *Specs* are used in X-13ARIMA-SEATS to control the computations and output. The PROC X13 statement performs some of the same functions as the Series spec in the Census Bureau's X-13ARIMA-SEATS software. The ADJUST statement performs some of the same functions as the Transform spec. The ARIMA, AUTOMDL, CHECK, ESTIMATE, FORECAST, IDENTIFY, OUTLIER, PICKMDL, REGRESSION, TRANSFORM, and X11 statements are designed to perform the same functions as the corresponding X-13ARIMA-SEATS specs, although full compatibility is not yet available. The Census Bureau documentation *X-13ARIMA-SEATS Reference Manual* (US Bureau of the Census 2009) provides added insight to the functionality of these statements. The SEATSDECOMP statement provides a SEATS (signal extraction in ARIMA time series) seasonal decomposition for the B1 series that uses the same ARIMA model as is used to model the series. For more information about SEATS, see Gómez and Maravall (1997a, b).

Functional Summary

Table 46.1 summarizes the statements and options that control the X13 procedure.

Table 46.1 Functional Summary

Description	Statement	Option
Data Set Options		
Specifies the auxiliary data set	PROC X13	AUXDATA=
Specifies the input data set	PROC X13	DATA=
Specifies the user-defined event definition data set	PROC X13	INEVENT=
Specifies regression and ARIMA information	PROC X13	MDLINFOIN=
Outputs regression and ARIMA information	PROC X13	MDLINFOOUT=
Writes summary statistics to an output data set	PROC X13	OUTSTAT=
Writes table values to an output data set	OUTPUT	OUT=
Appends forecasts to the OUTPUT OUT= data set	X11 or FORECAST	OUTFORECAST
Prefixes backcasts to the OUTPUT OUT= data set	FORECAST	OUTBACKCAST
Display Control Options		
Suppresses all displayed output	PROC X13	NOPRINT
Specifies the plots to be displayed	PROC X13	PLOTS=
Specifies the type of spectral plot to be displayed	PROC X13	PERIODOGRAM
Specifies the series for spectral analysis	PROC X13	SPECTRUMSERIES=
Displays automatic model information	AUTOMDL	PRINT=
Specifies the number of lags in regARIMA model residuals ACF and PACF tables and plots	CHECK	MAXLAG=
Displays regARIMA model residuals information	CHECK	PRINT=
Displays the iterations history	ESTIMATE	ITPRINT
Displays information about restarted iterations	ESTIMATE	PRINTERR
Specifies the differencing used in the ARIMA model identification ACF and PACF tables and plots	IDENTIFY	DIFF=
Specifies the seasonal differencing used in the ARIMA model identification ACF and PACF tables and plots	IDENTIFY	SDIFF=
Specifies the number of lags in ARIMA model identification ACF and PACF tables and plots	IDENTIFY	MAXLAG=
Displays regression model parameter estimates	IDENTIFY	PRINTREG

Table 46.1 *continued*

Description	Statement	Option
Requests tables that are not displayed by default	TABLES	
Specifies that the summary line not be displayed	TABLES	NOSUM
Date Information Options		
Specifies the date variable	PROC X13	DATE=
Specifies the date of the first observation	PROC X13	START=
Specifies the beginning or ending date or both of the subset	PROC X13	SPAN=
Specifies the interval of the time series	PROC X13	INTERVAL=
Specifies the interval of the time series	PROC X13	SEASONS=
Declaring the Role of Variables		
Specifies BY-group processing	BY	
Specifies identifying variables	ID	
Specifies the variables to be seasonally adjusted	VAR	
Specifies the user-defined variables that are available for regression	USERDEFINED	
Controlling the Table Computations		
Suppresses trimming of leading and trailing missing values (if they exist)	PROC X13	NOTRIMMISS
Transforms or prior-adjusts the series	TRANSFORM	FUNCTION=
Transforms or prior-adjusts the series	TRANSFORM	POWER=
Adjusts the series by using a predefined adjustment variable	ADJUST	PREDEFINED=
Specifies the likelihood function to be used for estimating AR and MA parameters	ESTIMATE	EXACT=
Specifies the maximum number of iterations for estimating AR and MA parameters	ESTIMATE	MAXITER
Specifies the convergence tolerance for nonlinear estimation	ESTIMATE	TOL=
Specifies size of forecast confidence limits	FORECAST	ALPHA=
Specifies the number of backcasts by which to extend the series for seasonal adjustment	FORECAST	NBACKCAST=
Specifies the number of forecasts by which to extend the series for seasonal adjustment	FORECAST	LEAD=
Specifies that one-step-ahead forecasts be computed	FORECAST	OUT1STEP

Table 46.1 *continued*

Description	Statement	Option
Specifying Outlier Detection Options		
Specifies automatic outlier detection	OUTLIER	
Specifies the span for outlier detection	OUTLIER	SPAN=
Specifies the outlier types to be detected	OUTLIER	TYPE=
Specifies the critical values for outlier detection	OUTLIER	CV=
Specifies the critical values for AO outlier detection	OUTLIER	AOCV=
Specifies the critical values for LS outlier detection	OUTLIER	LSCV=
Specifies the critical values for TC outlier detection	OUTLIER	TCCV=
Specifies the alpha value for outlier detection	OUTLIER	ALPHA=
Specifies the method for calculating the critical value for outlier detection based on the alpha value	OUTLIER	CVMETHOD=
Specifies the number of level-shift outliers to consider for forming a temporary level-shift	OUTLIER	LSRUN=
Specifies the rate of decay for temporary change outliers	OUTLIER	TCRATE=
Specifies the method of adding outliers at each iteration	OUTLIER	METHOD=
Specifies the difference in critical values for almost outliers	OUTLIER	ALMOST=
Specifying the Regression Model		
Specifies regression variables to be selected using an AIC-based test	REGRESSION	AICTEST=
Specifies predefined regression variables	REGRESSION	PREDEFINED=
Specifies user-defined regression variables	REGRESSION	USERVAR=
Specifies user-defined regression variables	INPUT	
Specifies user defined event regression variables	EVENT	
Specifies the method used to calculate the means for the Easter regression variable	REGRESSION	EASTERMEANS=
Specifies which types of regression effects are not to be removed before seasonal adjustment	REGRESSION	NOAPPLY=
Specifying the ARIMA Model		
Uses the X-13ARIMA-SEATS TRAMO-based method to choose a model	AUTOMDL	
Chooses a regARIMA model from a set that you specify	PICKMDL	
Specifies the ARIMA part of the model	ARIMA	MODEL=

Table 46.1 *continued*

Description	Statement	Option
Specifying Automatic Model Detection Options		
Specifies the maximum orders of ARMA polynomials	AUTOMDL	MAXORDER=
Specifies the maximum orders of differencing	AUTOMDL	MAXDIFF=
Specifies the estimation method for identifying difference orders	AUTOMDL	DIFFID=
Specifies the maximum number of iterations for exact likelihood for DIFFID=EXACTFIRST	AUTOMDL	DIFFIDITER=
Specifies the fixed orders of differencing	AUTOMDL	DIFFORDER=
Suppresses fitting of a constant parameter	AUTOMDL	NOINT
Specifies the preference for balanced models	AUTOMDL	BALANCED
Specifies Hannan-Rissanen initial estimation	AUTOMDL	HRINITIAL
Specifies default model acceptance based on Ljung-Box Q	AUTOMDL	ACCEPTDEFAULT
Specifies the acceptance value for Ljung-Box Q	AUTOMDL	LJUNGBOXLIMIT=
Specifies the percentage by which to reduce the outlier critical value	AUTOMDL	REDUCECV=
Specifies the critical value for ARMA coefficients	AUTOMDL	ARMACV=
Model Diagnostics		
Examines the regARIMA model residuals	CHECK	
Specifying Seasonal Adjustment Options		
Specifies seasonal adjustment	X11	
Specifies the mode of seasonal adjustment decomposition	X11	MODE=
Specifies the seasonal filter	X11	SEASONALMA=
Specifies the sigma limits	X11	SIGMALIM=
Specifies the Henderson trend filter	X11	TRENDMA=
Specifies the D11 calculation method	X11	TYPE=
Specifies the adjustment factors to remove from final seasonally adjusted series	X11	FINAL=
Specifies a method for reconciling the seasonally adjusted series to the original series	X11	FORCE=
Specifies that SEATS seasonal decomposition be output to a data set	SEATSDECOMP	OUT=

PROC X13 Statement

PROC X13 *options* **;**

The PROC X13 statement provides information about the time series to be processed by PROC X13. Either the DATE= or the START= option must be specified. If both options are specified, then a syntax error results and the X13 procedure is not executed.

The original series is displayed in Table A1. If there are missing values in the original series and a regARIMA model is specified or automatically selected, then Table MV1 is displayed. Table MV1 contains the original series with missing values replaced by the predicted values from the fitted model. If outliers are identified and Table A19 is added in the TABLES statement, then the outlier adjusted series is displayed in Table A19. Table B1 is displayed when the original data are altered (for example, through an ARIMA model estimation, prior adjustment factor, or regression) or the series is extended with forecasts.

Although the X-13ARIMA-SEATS method handles missing values, there are some restrictions. In order for PROC X13 to process the series, no month or quarter can contain missing values for all years. For instance, if the third quarter contained only missing values for all years, then processing is skipped for that series. In addition, if more than half the values for a month or a quarter are missing, then a warning message is displayed in the log file, and other errors might occur later in processing. If a series contains many missing values, other methods of missing value replacement should be considered prior to seasonally adjusting the series.

You can specify the following *options* in the PROC X13 statement:

AUXDATA=*SAS-data-set*

specifies an auxiliary input data set that contains user-defined variables, which are specified in the INPUT statement, the USERVAR= option in the REGRESSION statement, or the USERDEFINED statement. The AUXDATA= data set can also contain the date variable, which is specified in the DATE= option in the PROC X13 statement. If the date variable is present, then the date variable is used to align the observations in the auxiliary data set to the observations in the series that is being processed. The date values must be sorted in ascending order with no gaps or duplications, and the interval must match the interval of the series. If the date variable is not present or valid, then observations in the auxiliary data set are matched by observation number to the series that is being processed. The auxiliary data set does not support BY-group processing. The variables in the auxiliary data set are applied to all BY groups, where the dates of the BY group correspond to the dates of the auxiliary data set. Example 46.11 shows the use of the AUXDATA= data set.

DATA=*SAS-data-set*

specifies the input SAS data set to use. If this option is omitted, the most recently created SAS data set is used.

DATE=*variable*

DATEVAR=*variable*

specifies a variable that gives the date for each observation. Unless specified in the SPAN= option, the starting and ending dates are obtained from the first and last values of the BY group for the DATE= variable, which must contain SAS date or datetime values. The procedure checks values of the DATE= variable to ensure that the input observations are sequenced correctly in ascending order. If the INTERVAL= option or the SEASONS= option is specified, the values of the date variable must be consistent with the specified seasonality or interval. If neither the INTERVAL= option nor the

SEASONS= option is specified, then the procedure tries to determine the type of data from the values of the date variable. This variable is automatically added to the OUT= data set if a data set is requested in an OUTPUT statement, and the date values for the variable are extrapolated if necessary. If the DATE= option is not specified, the START= option must be specified.

INEVENT=*SAS-data-set*

specifies the input data set that defines any user-defined event variables. This option can be omitted if events are not specified or if only SAS predefined events are specified in an EVENT statement. For more information about the format of this data set, see the section "INEVENT= Data Set" on page 3360.

INTERVAL=*interval*

specifies the frequency of the input time series. If the input data consist of quarterly observations, then INTERVAL=QTR should be used. If the input data consist of monthly observations, then INTERVAL=MONTH should be used. If the INTERVAL= option is not specified and SEASONS=4, then INTERVAL=QTR is assumed; likewise, SEASONS=12 implies INTERVAL=MONTH. If both the INTERVAL= option and the SEASONS= option are specified, the values should not be conflicting. If neither the INTERVAL= option nor the SEASONS= option is specified and the START= option is specified, then the data are assumed to be monthly. If a date variable is specified using the DATE= option, it is not necessary to specify the INTERVAL= option or the SEASONS= option; however, if specified, the values of the INTERVAL= option or the SEASONS= option should not be in conflict with the values of the date variable. For more information about intervals, see Chapter 4, "Date Intervals, Formats, and Functions."

MDLINFOIN=*SAS-data-set*

specifies an optional input data set that contains model information that overrides information that is contained in one or more of the TRANSFORM, REGRESSION, ARIMA, and AUTOMDL statements. The *SAS-data-set* can contain BY-group, series names, and other information. For more information about this data set, see the section "MDLINFOIN= and MDLINFOOUT= Data Sets" on page 3358.

You can supply the following model information in *SAS-data-set*:

- a single model for each series that is used to forecast the series.
- multiple models for each series. If multiple models are specified for a series, the PICKMDL method is used to select from among the candidate models, and the selected model will be used to generate the forecasts. For more information, see the "PICKMDL Model Selection" on page 3348.

The MDLINFOIN= data set can include a variable that identifies different models. All observations that have the same value for the model identification variable are considered to be relevant to the same model. A single model can be considered to consist of all the observations for a BY group that consists of the BY variables (if any), the _NAME_ variable if it exists, and the model identification variable (whose default is _MODEL_). Even if the PICKMDL statement is not specified, but the MDLINFOIN= data set contains a _MODEL_ variable and more than one model for a series, then the PICKMDL method is automatically invoked to choose a model for that series.

MDLINFOOUT=*SAS-data-set*

specifies the optional output data set that contains the transformation, regression, and ARIMA information related to each seasonally adjusted series. The data set is sorted by the BY-group variables, if any, and by series names. The MDLINFOOUT= data set can be used as input for the MDLINFOIN=

option. For more information, see the section "MDLINFOIN= and MDLINFOOUT= Data Sets" on page 3358.

NOPRINT

suppresses any printed output.

NOTRIMMISS

suppresses the default, by which leading and trailing missing values are trimmed from each variable listed (or implied) in the VAR statement. If you specify the NOTRIMMISS option, PROC X13 treats leading and trailing missing values in the same manner as it treats embedded missing values. For information about the treatment of embedded missing values, see the section "Missing Values" on page 3340. Missing values are not supported in the regression variables that you specify in the REGRESSION, INPUT, or USERDEFINED statement; therefore, leading and trailing missing values are always trimmed from user-defined regressors even if you specify NOTRIMMISS.

OUTSTAT=*SAS-data-set*

specifies an optional output data set which contains the summary statistics that related to each seasonally adjusted series. The data set is sorted by the BY-group variables, if any, and by series names. For more information, see the section "OUTSTAT= Data Set" on page 3362.

PERIODOGRAM

specifies that the PERIODOGRAM rather than the spectrum of the series be plotted in the G tables and plots. If PERIODOGRAM is not specified, then the spectrum is plotted in the G tables.

PLOTS< (*global-plot-options***) > < =** *plot-request < (options) >* **>**

PLOTS< (*global-plot-options***) > < = (***plot-request < (options) >* **< . . .** *plot-request < (options) >* **>) >**

controls the plots that are produced through ODS Graphics. When you specify only one plot request, you can omit the parentheses around the plot request.

Following are some examples of the PLOTS= option:

```
plots=none
plots=all
plots=residual(none)
plots(only)=(series(acf pacf) residual(hist))
```

ODS Graphics must be enabled before you request plots. For example:

```
ods graphics on;

proc x13 data=sales date=date;
   var sales;
   identify diff=(0,1) sdiff=(0,1);
run;
```

Since no specific plot is requested in this program, the default plots associated with the PROC X13 and IDENTIFY statements are produced.

For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*). If you have enabled ODS Graphics but do not specify any specific plot request, then the default plots that are associated with each of the PROC X13 statements used in the program are produced. Line printer plots are suppressed when ODS Graphics is enabled.

If NONE is specified in an option, then no plots are produced for that option. If ALL is specified without NONE in an option, then all plots are produced for that option.

Global Plot Options

The *global-plot-options* apply to all relevant plots that are generated by the X13 procedure. The following *global-plot-option* is supported:

ONLY
suppresses the default plots. Only the plots specifically requested are produced.

Specific Plot Options

The following list describes the specific plots and their *options*:

ALL
produces all plots that are appropriate for the particular analysis.

NONE
suppresses all plots.

ADJUSTED(< *sa-plot-options* **>)**
SA(< *sa-plot-options* **>)**
produces plots of the seasonally adjusted series that results from the decomposition specified in the X11 statement. The SPECTRUM plot is produced by default.

The following *sa-plot-options* are available:

ALL
produces all seasonally adjusted plots.

NONE
suppresses all seasonally adjusted plots.

SPECTRUM
produces the spectral plot of Table G1. Table G1 is calculated based on the modified seasonally adjusted series (Table E2). The data are first-differenced and transformed as specified in the TRANSFORM statement. By default, the type of spectral estimate used to calculate the spectral plot is the spectrum. If the PERIODOGRAM option is specified in the PROC X13 statement, then the periodogram of the series is used to calculate the spectral plot.

FORECAST(< *forecast-plot-options* **>)**
produces the regARIMA model forecast plots if the FORECAST statement is specified. The FORECAST plot is produced by default. The following *forecast-plot-options* are available:

ALL
produces all the forecast plots that are appropriate for the particular analysis.

FORECAST
plots the actual time series and its one-step-ahead forecast over the historical period, and plots the forecast and its confidence bands over the forecast horizon. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts.

FORECASTONLY
plots the forecast and its confidence bands over the forecast horizon only.

MODELFORECASTS
plots the one-step-ahead model forecast and its confidence bands in the historical period; plots the forecast and its confidence bands over the forecast horizon. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts.

MODELS
plots the one-step-ahead model forecast and its confidence bands in the historical period. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts.

NONE
suppresses all the forecast plots.

TRANSFORECAST
plots the transformed time series and its one-step-ahead forecast over the historical period; plots the forecast and its confidence bands over the forecast horizon. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts. The TRANSFORECAST plot is available only if the data have been transformed using the TRANSFORM statement.

TRANSFORECASTONLY
plots the forecast of the transformed series and its confidence bands over the forecast horizon only. The TRANSFORECASTONLY plot is available only if the data have been transformed using the TRANSFORM statement.

TRANSMODELFORECASTS
plots the one-step-ahead model forecast of the transformed series and its confidence bands in the historical period; plots the forecast and its confidence bands over the forecast horizon. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts. The TRANSMODELFORECASTS plot is available only if the data have been transformed using the TRANSFORM statement.

TRANSMODELS
plots the one-step-ahead model forecast of the transformed series and its confidence bands in the historical period. The OUT1STEP option must be specified in the FORECAST statement in order for the X13 procedure to calculate the one-step-ahead forecasts. The TRANSMODELS plot is available only if the data have been transformed using the TRANSFORM statement.

IRREGULAR(< *ic-plot-options* **>)**

IC(< *ic-plot-options* **>)**

produces plots of the irregular series that results from the decomposition specified in the X11 statement. The SPECTRUM plot is produced by default.

The following *ic-plot-options* are available:

ALL

produces all irregular plots.

NONE

suppresses all irregular plots.

SPECTRUM

produces the spectral plot of Table G2. Table G2 is calculated based on the modified irregular series (Table E3). The data are first-differenced and transformed as specified in the TRANSFORM statement. By default, the type of spectral estimate used to calculate the spectral plot is the spectrum. If the PERIODOGRAM option is specified in the PROC X13 statement, then the periodogram of the series is used to calculate the spectral plot.

RESIDUAL(< *residual-plot-options* **>)**

produces the regARIMA model residual series plots if the CHECK statement is specified. The ACF, PACF, HIST, SQACF, and SPECTRUM plots are produced by default. The following *residual-plot-options* are available:

ACF

produces the plot of residual autocorrelations.

ALL

produces all the residual diagnostics plots that are appropriate for the particular analysis.

HIST

produces the histogram of the residuals and also the residual outliers and residual statistics tables that describe the residual histogram.

NONE

suppresses all the residual diagnostics plots.

PACF

produces the plot of residual partial-autocorrelations if PRINT=PACF is specified in the CHECK statement.

SPECTRUM

produces the spectral plot of Table GRs. Table GRs is calculated based on the regARIMA model residual series. By default, the type of spectral estimate used to calculate the spectral plot is the spectrum. If the PERIODOGRAM option is specified in the PROC X13 statement, then the periodogram of the series is used to calculate the spectral plot.

SQACF

produces the plot of squared residual autocorrelations.

SERIES(< *series-plot-options* **>)**

produces plots that are associated with the identification stage of the modeling. The ACF, PACF, and SPECTRUM plots are produced by default. The following *series-plot-options* are available:

ACF

produces the plot of autocorrelations.

ALL

produces all the plots that are associated with the identification stage.

NONE

suppresses all plots that are associated with the identification stage.

PACF

produces the plot of partial-autocorrelations.

SPECTRUM

produces the spectral plot of Table G0. Table G0 is calculated based on either Table A1, A19, B1, or E1, as specified by the SPECTRUMSERIES= option. The original data are first-differenced and transformed as specified in the TRANSFORM statement. By default, the type of spectral estimate that is used to calculate the spectral plot is the spectrum. If the PERIODOGRAM option is specified in the PROC X13 statement, then the periodogram of the series is used to calculate the spectral plot.

SEASONS=*number*

specifies the number of observations in a seasonal cycle. If the SEASONS= option is not specified and INTERVAL=QTR, then SEASONS=4 is assumed. If the SEASONS= option is not specified and INTERVAL=MONTH, then SEASONS=12 is assumed. If the SEASONS= option is specified, its value should not conflict with the values of the INTERVAL= option or the values of the date variable. For more information, see the descriptions of the START=, DATE=, and INTERVAL= options.

SPAN=(*mmmyy* **,***mmmyy* **)**

SPAN=(*'yyQq'* **,***'yyQq'* **)**

specifies the dates of the first and last observations to define a subset for processing. A single date in parentheses is interpreted to be the starting date of the subset. To specify only the ending date, use SPAN=(,*mmmyy*). If the starting or ending date is omitted, then the first or last date, respectively, of the input data set or BY group is assumed. Because the dates are input as strings and the quarterly dates begin with a numeric character, the specification for a quarterly date must be enclosed in quotation marks. A four-digit year can be specified; if a two-digit year is specified, the value specified in the YEARCUTOFF= SAS system option applies.

SPECTRUMSERIES=*table-name*

specifies the table name of the series that is used in the spectrum of the original series (Table G0). The table names that can be specified are A1, A19, B1, or E1. The default is B1.

START=*mmmyy*

START=*'yyQq'*

STARTDATE=*mmmyy*

STARTDATE=*'yyQq'*

specifies the date of the first observation. Unless the SPAN= option is used, the starting and ending dates are the dates of the first and last observations, respectively. Either this option or the DATE= option is required. When using this option, use either the INTERVAL= option or the SEASONS= option to specify monthly or quarterly data. If neither the INTERVAL= option nor the SEASONS= option is present, monthly data are assumed. Because the dates are input as strings and the quarterly dates begin with a numeric character, the specification for a quarterly date must be enclosed in quotation marks. A four-digit year can be specified; if a two-digit year is specified, the value specified in the YEARCUTOFF= SAS system option applies. When using the START= option with BY processing, the start date is applied to the first observation in each BY group.

ADJUST Statement

ADJUST *option* **;**

The ADJUST statement adjusts the series for leap year and length-of-period factors prior to estimating a regARIMA model. The “Prior Adjustment Factors” table is associated with the ADJUST statement.

The following *option* can appear in the ADJUST statement:

PREDEFINED=LOM | LOQ | LPYEAR

specifies length-of-month adjustment, length-of-quarter adjustment, or leap year adjustment. PREDEFINED=LOM and PREDEFINED=LOQ are equivalent because the actual adjustment is determined by the interval of the time series. Also, because leap year adjustment is a limited form of length-of-period adjustment, only one type of predefined adjustment can be specified. The PREDEFINED= option should not be used in conjunction with PREDEFINED=TD or PREDEFINED=TD1COEF in the REGRESSION statement or MODE=ADD or MODE=PSEUDOADD in the X11 statement. PREDEFINED=LPYEAR cannot be specified unless the series is log transformed.

If the series is to be transformed by using a Box-Cox or logistic transformation, the series is first adjusted according to the ADJUST statement, and then it is transformed.

In the case of a length-of-month adjustment for the series with observations Y_t, each observation is first divided by the number of days in that month, m_t, and then multiplied by the average length of month (30.4375), resulting in $(30.4375 \times Y_t)/m_t$. Length-of-quarter adjustments are performed in a similar manner, resulting in $(91.3125 \times Y_t)/q_t$, where q_t is the length in days of quarter t.

Forecasts of the transformed and adjusted data are transformed and adjusted back to the original scale for output.

ARIMA Statement

ARIMA *option* **;**

The ARIMA statement specifies the ARIMA part of the regARIMA model. This statement defines a pure ARIMA model if no REGRESSION statements, INPUT statements, or EVENT statements are specified. The ARIMA part of the model can include multiplicative seasonal factors.

The following *option* can appear in the ARIMA statement:

MODEL=((*p d q*) (*P D Q*)*s*)

specifies the ARIMA model. The format follows standard Box-Jenkins notation (Box, Jenkins, and Reinsel 1994). The nonseasonal AR and MA orders are given by *p* and *q*, respectively, while the seasonal AR and MA orders are given by *P* and *Q*. The number of differences and seasonal differences are given by *d* and *D*, respectively. The notation (*p d q*) and (*P D Q*) can also be specified as (*p*, *d*, *q*) and (*P*, *D*, *Q*). The maximum lag of any AR or MA parameter is 36. The maximum value of a difference order, *d* or *D*, is 144. All values for *p*, *d*, *q*, *P*, *D*, and *Q* should be nonnegative integers. The seasonality parameter, *s*, should be a positive integer. If *s* is omitted, it is set equal to the value that is specified in the SEASONS= option in the PROC X13 statement.

For example, the following statements specify an ARIMA (2,1,1)(1,1,0)12 model:

```
proc x13 data=ICMETI seasons=12 start=jan1968;
   arima model=((2,1,1)(1,1,0));
```

AUTOMDL Statement

AUTOMDL *options* **;**

The AUTOMDL statement invokes the automatic model selection procedure of the X-13ARIMA-SEATS method. This method is based largely on the TRAMO (time series regression with ARIMA noise, missing values, and outliers) method by Gómez and Maravall (1997a, b). If the AUTOMDL statement is used without the OUTLIER statement, then only missing values regressors are included in the regARIMA model. If both the AUTOMDL and the OUTLIER statements are used, then both missing values regressors and regressors for automatically identified outliers are included in the regARIMA model. For more information about missing value regressors, see the section “Missing Values” on page 3340.

If both the AUTOMDL statement and the ARIMA statement are present, the ARIMA statement is ignored. The ARIMA statement specifies the model, but the AUTOMDL statement allows the X13 procedure to select the model. If the AUTOMDL statement is specified and a data set is specified in the MDLINFOIN= option in the PROC X13 statement, then the AUTOMDL statement is ignored if the specified data set contains a model specification for the series. If no model for the series is specified in the MDLINFOIN= data set, the AUTOMDL or ARIMA statement is used to determine the model. Thus, it is possible to give a specific model for some series and automatically identify the model for other series by using both the MDLINFOIN= option and the AUTOMDL statement.

The AUTOMDL statement cannot be specified when the PICKMDL statement is also specified. The AUTOMDL and PICKMDL statements each specify different methods of automatic model selection. So, either one method must be used or the other method must be used to select a model.

When the AUTOMDL statement is specified, the X13 procedure compares a model selected using a TRAMO method to a default model. The TRAMO method is implemented first, and involves two parts: identifying the orders of differencing and identifying the ARIMA model. The table “ARIMA Estimates for Unit Root Identification” provides details about the identification of the orders of differencing, and the table “Results of Unit Root Test for Identifying Orders of Differencing” shows the orders of differencing selected by TRAMO. The table “Models Estimated by Automatic ARIMA Model Selection Procedure” provides details regarding the TRAMO automatic model selection, and the table “Best Five ARIMA Models Chosen by Automatic Modeling” ranks the best five models estimated using the TRAMO method. The “Comparison of Automatically Selected Model and Default Model" table compares the model selected by the TRAMO method to a default model. At this point in the processing, if the default model is selected over the TRAMO model, then PROC X13 displays a note. No note is displayed if the TRAMO model is selected. The Ljung-Box *Q* statistic is then checked for acceptance, and the results are displayed in the "Check of the Residual Ljung-Box Q Statistic" table. The initial model selected at this point is displayed in the "Initial Automatic Model Selection" table. PROC X13 then performs final checks for unit roots, overdifferencing, and insignificant ARMA coefficients. The results of the final checks are displayed in the "Final Checks for Identified Model" table, which also indicates changes to the model order if the orders are changed. The last table, "Final Automatic Model Selection," shows the results of the automatic model selection; if the orders have been altered during the final checks, the Orders Altered column displays a value of Yes. An example of the automatic modeling selection procedure is shown in Example 46.4.

The following *options* can appear in the AUTOMDL statement:

ACCEPTDEFAULT

specifies that the default model be chosen if its Ljung-Box *Q* is acceptable.

ARMACV=*value*

specifies the threshold value for the *t* statistics that are associated with the highest-order ARMA coefficients. As a check of model parsimony, the parameter estimates and *t* statistics of the highest-order ARMA coefficients are examined to determine whether the coefficient is insignificant. An ARMA coefficient is considered to be insignificant if the *t* value that is displayed in the table "Exact ARMA Maximum Likelihood Estimation" is below the value specified in the ARMACV= option and the absolute value of the parameter estimate is reliably close to zero. The absolute value is considered to be *reliably close to zero* if it is below 0.15 for 150 or fewer observations or is below 0.1 for more than 150 observations. If the highest-order ARMA coefficient is found to be insignificant, then the order of the ARMA model is reduced. For example, if AUTOMDL identifies a (3 1 1)(0 0 1) model and the parameter estimate of the seasonal MA lag of order 1 is –0.09 and its *t* value is –0.55, then the ARIMA model is reduced to at least (3 1 1)(0 0 0). After the model is reestimated, the check for insignificant coefficients is performed again. If ARMACV=0.54 is specified in the preceding example, then the coefficient is not found to be insignificant and the model is not reduced.

If a constant is allowed in the model and if the *t* value associated with the constant parameter estimate is below the ARMACV= critical value, then the constant is considered to be insignificant and is removed from the model. Note that if a constant is added to or removed from the model and then the ARIMA model changes, then the *t* statistic for the constant parameter estimate also changes. Thus, changing the ARMACV= value does not necessarily add or remove a constant term from the model.

The value specified in the ARMACV= option should be greater than zero. The default value is 1.0.

BALANCED

specifies that the automatic modeling procedure prefer balanced models over unbalanced models. A balanced model is one in which the sum of the AR, seasonal AR, differencing, and seasonal differencing orders equals the sum of the MA and seasonal MA orders. Specifying BALANCED gives the same preference as the TRAMO program. If BALANCED is not specified, all models are given equal consideration.

DIFFID=CONDITIONAL | EXACT | EXACTFIRST (Experimental)

specifies the estimation to be used in automatic difference identification when Hannen-Rissanen fails. You can specify the following values:

CONDITIONAL uses conditional likelihood estimation.

EXACT uses exact likelihood estimation.

EXACTFIRST attempts to estimate the parameters by using exact likelihood for the first *diffiditer* iterations, where *diffiditer* is specified in the DIFFIDITER= option. If the estimation does not converge within *diffiditer* iterations, then conditional likelihood is used to estimate the parameters.

The effects of this option are displayed in the Estimation Method column in the "ARIMA Estimates for Unit Root Identification" table. By default, DIFFID=EXACTFIRST.

DIFFIDITER=*diffiditer* (Experimental)

specifies the maximum number of exact likelihood estimation iterations when DIFFID=EXACTFIRST is specified. If the number of iterations exceeds *diffiditer*, then conditional likelihood is used to estimate the remaining iterations. The default value is 500; this default differs from the default value of 200 in the US Census Bureau's implementation of X-13ARIMA-SEATS.

DIFFORDER=(*nonseasonal-order***,** *seasonal-order***)**

specifies the fixed orders of differencing to be used in the automatic ARIMA model identification procedure. When the DIFFORDER= option is used, only the AR and MA orders are automatically identified. Acceptable values for the regular (nonseasonal) differencing orders are 0, 1, and 2; acceptable values for the seasonal differencing orders are 0 and 1. If the MAXDIFF= option is also specified, then the DIFFORDER= option is ignored. There are no default values for DIFFORDER. If neither the DIFFORDER= option nor the MAXDIFF= option is specified, then the default is MAXDIFF=(2,1).

HRINITIAL

specifies that Hannan-Rissanen estimation be done before exact maximum likelihood estimation to provide initial values. If the HRINITIAL option is specified, then models for which the Hannan-Rissanen estimation has an unacceptable coefficient are rejected.

LJUNGBOXLIMIT=*value*

specifies acceptance criteria for the confidence coefficient of the Ljung-Box Q statistic. If the Ljung-Box Q for a final model is greater than this value, the model is rejected, the outlier critical value is reduced, and outlier identification is redone with the reduced value. For more information, see the REDUCECV option. The value specified in the LJUNGBOXLIMIT= option must be greater than 0 and less than 1. The default value is 0.95.

MAXDIFF=(*nonseasonal-order***,** *seasonal-order***)**

specifies the maximum orders of regular and seasonal differencing for the automatic identification of differencing orders. When MAXDIFF is specified, the differencing orders are identified first, and then the AR and MA orders are identified. Acceptable values for the regular differencing orders are 1 and 2. The only acceptable value for the seasonal differencing order is 1. If both the MAXDIFF= option and the DIFFORDER option= are specified, then the DIFFORDER= option is ignored. If neither the DIFFORDER= nor the MAXDIFF= option is specified, the default is MAXDIFF=(2,1).

MAXORDER=(*nonseasonal-order***,** *seasonal-order***)**

specifies the maximum orders of nonseasonal and seasonal ARMA polynomials for the automatic ARIMA model identification procedure. The maximum order for the nonseasonal ARMA parameters is 4, and the maximum order for the seasonal ARMA is 2.

NOINT

suppresses the fitting of a constant or intercept parameter in the model.

PRINT=(*option-list*)

specifies the tables to be displayed in the output. You can specify one or more of the following *options* (parentheses are optional; use a space between *options*):

NONE suppresses all automatic modeling output.

ALL includes all automatic modeling tables in the output if NONE is not specified in the *option-list*.

ONLY specifies that only the listed tables be output.

AUTOCHOICE displays the tables titled "Comparison of Automatically Selected Model and Default Model" and "Final Automatic Model Selection." The "Comparison of Automatically Selected Model and Default Model" table compares a default model to the model chosen by the TRAMO-based automatic modeling method. The "Final Automatic Model Selection" table indicates which model has been chosen automatically. These tables are output by default unless NONE or ONLY is specified in the *option-list*.

AUTOCHOICEMDL displays the table "Models Estimated by Automatic ARIMA Model Selection Procedure." This table summarizes the various models that were considered by the TRAMO automatic model selection method and their measures of fit.

AUTOLJUNGBOX displays the table "Check of the Residual Ljung-Box Q Statistic." This table is displayed only if the model is not accepted because the Ljung-Box Q statistic is greater than the acceptance limit. The details of the test and the changes made either to the model or to the model selection method are displayed.

BEST5MODEL displays the table "Best Five ARIMA Models Chosen by Automatic Modeling." This table ranks the five best models that were considered by the TRAMO automatic modeling method.

FINALCHECKS displays the table "Final Checks for Identified Model." This table displays the results of the final checks for model adequacy. The final checks can result in the orders of the initially identified model being altered. Any order changes or changes in the constant term are included in this table. This table is output by default unless NONE or ONLY is specified in the *option-list*.

INITCHOICEMDL displays the table "Initial Automatic Model Selection." The "Comparison of Automatically Selected Model and Default Model" table compares a default model to the model chosen by the TRAMO-based automatic modeling method. The chosen model can then be altered if the model fails the Ljung-Box Q statistic test. The "Initial Automatic Model Selection" table indicates which model has been chosen automatically after the Ljung-Box Q statistic test. This table is output by default unless NONE or ONLY is specified in the *option-list*.

UNITROOTTEST displays the table titled "Results of Unit Root Test for Identifying Orders of Differencing." This table displays the orders that were automatically selected by the AUTOMDL statement. Unless the nonseasonal and seasonal differences are specified using the DIFFORDER= option, the AUTOMDL statement automatically identifies the orders of differencing. This table is output by default unless NONE or ONLY is specified in the *option-list*.

UNITROOTTESTMDL displays the table titled "ARIMA Estimates for Unit Root Identification." This table summarizes the various models that were considered by the TRAMO automatic selection method while it identified the orders of differencing and the statistics associated with those models. The unit root identification method first attempts to obtain the coefficients by using the Hannan-Rissanen method. If Hannan-Rissanen estimation cannot be performed, the algorithm attempts to obtain the coefficients by using conditional likelihood estimation.

By default, PRINT=(UNITROOTTEST AUTOCHOICE INITCHOICEMDL FINALCHECKS).

REDUCECV=*value*

specifies the percentage by which the outlier critical value be reduced when a final model is found to have an unacceptable confidence coefficient for the Ljung-Box Q statistic. This value should be between 0 and 1. The default value is 0.14286.

BY Statement

BY *variables* **;**

A BY statement can be used with PROC X13 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.

CHECK Statement

CHECK *options* **;**

The CHECK statement produces statistics for diagnostic checking of residuals from the estimated regARIMA model.

The following tables that are associated with diagnostic checking are displayed in the output: "Autocorrelation of regARIMA Model Residuals," "Partial Autocorrelation of regARIMA Model Residuals," "Autocorrelation of Squared regARIMA Model Residuals," "Outliers of the Unstandardized Residuals," "Summary Statistics for the Unstandardized Residuals," "Normality Statistics for regARIMA Model Residuals," and "Table G Rs: 10*LOG(SPECTRUM) of the regARIMA Model Residuals." If ODS graphics is enabled, the following plots that are associated with diagnostic checking output are produced: the autocorrelation function (ErrorACF) plot of the residuals, the partial autocorrelation function (ErrorPACF) plot of the residuals, the autocorrelation function (SqErrorACF) plot of the squared residuals, a histogram (ResidualHistogram) of the residuals, and a spectral plot (SpectralPlot) of the residuals. For more information about controlling the display of plots, see the PLOTS=RESIDUAL option in the PROC X13 statement.

The residual histogram displayed by the X13 procedure shows the distribution of the unstandardized, uncentered regARIMA model residuals; the residual histogram displayed by the US Census Bureau's X-13ARIMA-SEATS seasonal adjustment program displays standardized and mean-centered residuals.

The following *options* can appear in the CHECK statement:

MAXLAG=*value*

specifies the number of lags for the residual sample autocorrelation function (ACF) and partial autocorrelation function (PACF). The default is 36 for monthly series and 12 for quarterly series. The minimum value for MAXLAG= is 1.

For the table "Autocorrelation of Squared regARIMA Model Residuals" and the corresponding SqErrorACF plot, the maximum number of lags calculated is 12 for monthly series and 4 for quarterly series. The MAXLAG= option can only reduce the number of lags for this table and plot.

PRINT=(*option-list***)**

specifies the diagnostic checking tables to be displayed. You can specify one or more of the following *options* (parentheses are optional; use a space between *options*):

NONE suppresses diagnostic checking output. If PRINT=NONE is specified and no other PRINT= option is specified, then none of the tables that are associated with diagnostic checking are displayed. However, PRINT=NONE has no effect if other PRINT= options are specified in the CHECK statement.

ALL specifies that all tables related to diagnostic checking be displayed.

ACF displays the table titled "Autocorrelation of regARIMA Model Residuals."

ACFSQUARED displays the table titled "Autocorrelation of Squared regARIMA Model Residuals."

NORM displays the table titled "Normality Statistics for regARIMA Model Residuals." Measures of normality included in this table are skewness, Geary's *a* statistic, and kurtosis.

PACF displays the table titled "Partial Autocorrelation of regARIMA Model Residuals."

RESIDUALOUTLIER displays the table titled "Outliers of the Unstandardized Residuals" if the residuals contain outliers. You can specify this option either as PRINT=RESIDUALOUTLIER or PRINT=RESOUTLIER.

RESIDUALSTATISTICS displays the table titled "Summary Statistics for the Unstandardized Residuals." You can specify this option either as PRINT=RESIDUALSTATISTICS or PRINT=RESSTAT.

SPECRESIDUAL displays the table titled "Table G Rs: 10*LOG(SPECTRUM) of the regARIMA Model Residuals."

By default, PRINT=(ACF ACFSQUARED NORM RESIDUALOUTLIER RESIDUALSTATISTICS SPECRESIDUAL).

ESTIMATE Statement

ESTIMATE *options* **;**

The ESTIMATE statement estimates the regARIMA model. The regARIMA model is specified by the REGRESSION, INPUT, EVENT, and ARIMA statements or by the MDLINFOIN= data set in the PROC

X13 statement. Estimation output includes point estimates and standard errors for all estimated AR, MA, and regression parameters; the maximum likelihood estimate of the variance σ^2; t statistics for individual regression parameters; χ^2 statistics for assessing the joint significance of the parameters associated with certain regression effects (if included in the model); and likelihood-based model selection statistics (if the exact likelihood function is used). The regression effects for which χ^2 statistics are produced are fixed seasonal effects.

Tables displayed in the output associated with estimation are "Exact ARMA Likelihood Estimation Iteration Tolerances," "Average Absolute Percentage Error in within-Sample Forecasts," "ARMA Iteration History," "AR/MA Roots," "Exact ARMA Likelihood Estimation Iteration Summary," "Regression Model Parameter Estimates," "Chi-Squared Tests for Groups of Regressors," "Exact ARMA Maximum Likelihood Estimation," and "Estimation Summary."

The following *options* can appear in the ESTIMATE statement:

EXACT=ARMA | MA | NONE (Experimental)

specifies the likelihood function for estimation, likelihood evaluation, and forecasting. You can specify the following values:

ARMA uses the likelihood function that is exact for both AR and MA parameters.

MA uses the likelihood function that is exact for MA parameters, but conditional for AR parameters.

NONE uses the likelihood function that is conditional for both AR and MA parameters.

The ARMA estimation iterations are displayed in the "Iteration History" table, which is available when the ITPRINT option is specified. By default, EXACT=ARMA.

ITPRINT

displays the "Iteration History" table. This table includes detailed output for estimation iterations, including log-likelihood values, parameters, counts of function evaluations, and iterations. It is useful to examine the "Iteration History" table when errors occur within estimation iterations. By default, only successful iterations are displayed, unless the PRINTERR option is specified. An unsuccessful iteration is an iteration that is restarted due to a problem such as a root inside the unit circle. Successful iterations have a status of 0. If restarted iterations are displayed, a note at the end of the table gives definitions for status codes that indicate a restarted iteration. For restarted iterations, the number of function evaluations and the number of iterations is –1, which is displayed as missing. If regression parameters are included in the model, then both IGLS and ARMA iterations are included in the table. The number of function evaluations is a cumulative total.

MAXITER=*value*

specifies the maximum number of iterations used in estimating the AR and MA parameters. For models that include regression variables, this limit applies to the total number of ARMA iterations over all iterations of the iterative generalized least squares (IGLS) algorithm. For models without regression variables, *value* is the maximum number of iterations allowed for the set of ARMA iterations. By default, MAXITER=1500.

PRINTERR

causes restarted iterations to be included in the "Iteration History" table if ITPRINT is specified; creates the "Restarted Iterations" table if ITPRINT is not specified. Whether or not PRINTERR is specified, a WARNING message is printed to the log file if any iteration is restarted during estimation.

TOL=*value*

specifies the convergence tolerance for the nonlinear estimation. Absolute changes in the log-likelihood are compared to the TOL= value to check convergence of the estimation iterations. For models with regression variables, the TOL= value is used to check convergence of the IGLS iterations (where the regression parameters are reestimated for each new set of AR and MA parameters). For models without regression variables, there are no IGLS iterations, and the TOL= value is then used to check convergence of the nonlinear iterations that are used to estimate the AR and MA parameters. The default value is TOL=0.00001. The minimum tolerance value is a positive value based on the machine precision and the length of the series. If a tolerance less than the minimum supported value is specified, an error message is displayed and the series is not processed.

EVENT Statement

EVENT *variables* < */ options* > **;**

The EVENT statement specifies events to be included in the regression portion of the regARIMA model. Multiple EVENT statements can be specified. Dummy variable values for EVENT variables are generated by the X13 procedure, however, the EVENT variables are input as user-defined regression effects to the X-13ARIMA-SEATS method. Thus, the EVENT variables are treated in the same manner as it treats variables specified in the USERVAR= option in the REGRESSION statement. If a MDLINFOIN= data set is not specified in the PROC X13 statement, then all variables specified in the EVENT statements are applied to all BY groups and all time series that are processed. If a MDLINFOIN= data set is specified, then the EVENT statements apply only if no regression information for the BY group and series is available in the MDLINFOIN= data set. The events specified in the EVENT statements either must be SAS predefined events or must be defined in the data set specified in the INEVENT= option in the PROC X13 statement. For a summary of SAS predefined events, see the section "SAS Predefined Events" on page 3341.

The EVENT statement can also be used to include outlier, level-shift, and temporary change regressors that are available as predefined US Census Bureau variables in the X-13ARIMA-SEATS program. For example, the following statements specify an additive outlier in January 1970 and a level-shift that begins in July 1971:

```
proc x13 data=ICMETI seasons=12 start=jan1968;
   event AO01JAN1970D CBLS01JUL1971D;
```

The following statements specify an additive outlier in the second quarter 1970 and a temporary change that begins in the fourth quarter 1971:

```
proc x13 data=ICMETI seasons=4 start='1970q1';
   event AO01APR1970D TC01OCT1971D;
```

The following *options* can appear in the EVENT statement:

B=(*value* < **F** > *. . .* **)**

specifies initial or fixed values for the EVENT parameters in the order in which they appear in *variables*. Each B= list applies to the variable list that immediately precedes the slash.

For example, the following statements set an initial value of 1 for the event, x:

```
event y ;
event x / b=1 2 ;
```

In this example, the B= option applies only to the second EVENT statement. The value 2 is discarded because there is only one variable in the variable list.

To assign an initial value of 1 to the y regressor and 2 to the x regressor, use the following statements:

```
event y / b=1;
event x / b=2 ;
```

An **F** immediately following the numerical value indicates that this is not an initial value, but a fixed value. For an example that uses fixed parameters, see Example 46.8. In PROC X13, individual parameters can be fixed while other parameters in the same model are estimated.

USERTYPE=(*values*)

enables a user-defined variable to be processed in the same manner as a US Census predefined variable. You can specify the following *values*: AO, CONSTANT, EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, LS, RP, SCEASTER, SEASONAL, TC, TD, TDSTOCK, THANKS, or USER. For example, the US Census Bureau EASTER(w) regression effects are included in the "RegARIMA Holiday Component" table (A7). Specify USERTYPE=EASTER to include an event variable that is processed exactly as the US Census predefined EASTER(w) variable, including inclusion in the A7 table. The NOAPPLY= option in the REGRESSION statement also changes the processing of variables based on the USERTYPE= value. Table 46.4 shows the regression types that are associated with each regression effects table.

Each USERTYPE= list applies to the variable list that immediately precedes the slash. The same rules for assigning B= values to regression variables apply for USERTYPE= options. For example, the following statements specify that the event in the variable MyEaster be processed exactly as the US Census predefined LOM variable:

```
event MyLOM;
event MyEaster / usertype=LOM EASTER;
```

In this example, the USERTYPE= option applies only to the MyEaster variable in the second EVENT statement. The USERTYPE value **EASTER** is discarded because there is only one variable in the variable list.

To assign the USERTYPE value **LOM** to the MyLOM variable and **EASTER** to the MyEaster variable, use the following statements:

```
event MyLOM / usertype=LOM;
event MyEaster / usertype=EASTER;
```

The following USERTYPE= options specify that the regression effect be removed from the seasonally adjusted series: EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, SCEASTER, SEASONAL, TD, TDSTOCK, THANKS, and USER. When a regression effect is removed from the seasonally adjusted series, the level (mean) of the seasonally adjusted series can be altered. It is often desirable to use a zero-mean (mean-adjusted) regressor for effects that are to be removed from the seasonally adjusted series. For an example showing the effects of specifying a zero-mean regressor, see Example 46.6.

FORECAST Statement

FORECAST *options* **;**

The FORECAST statement uses the estimated model to forecast the time series. The output contains point forecasts and forecast statistics for the transformed and original series. Whenever forecasts or backcasts (or both) are generated and seasonal adjustment is performed, the forecasts and backcasts are appended to the original series, and the seasonal adjustment procedures are applied to the forecast or backcast (or both) extended series. If the FORECAST statement is not specified, but a regARIMA model is specified using either the ARIMA or AUTOMDL statement, then the series is extended one year ahead by default.

Tables that contain forecasts, standard errors, and confidence limits are displayed in association with the FORECAST statement. If the data are transformed, then two tables are displayed: one table for the original data, and one table for the transformed data. Data from these tables can be output to a SAS data set using ODS. The auxiliary variable _SCALE_ is included in forecast data sets that are output using ODS. The value of _SCALE_ is “Original” or “Transformed” to indicate the scale of the data. The auxiliary variable can also be used in ODS SELECT and ODS OUTPUT statements. For example, you can specify the following statements to output the forecasts on the original scale to a data set forecasts and the forecasts on the transformed scale to a data set Tforecasts:

```
ods output Original.ForecastCL=forecasts;
ods output Transformed.ForecastCL=Tforecasts;
```

The following *options* can appear in the FORECAST statement:

ALPHA=α

specifies the size of the upper and lower confidence limits, which are calculated as $1 - \alpha$, where α must be between 0 and 1. By default, ALPHA=0.05, which produces 95% confidence intervals.

LEAD=*value*

specifies the number of periods ahead to forecast for regARIMA extension of the series. The default is the number of periods in a year (4 or 12), and the maximum is 120. Setting LEAD=0 specifies that the series not be extended by forecasts for seasonal adjustment. The LEAD= value also controls the number of forecasts that are displayed in Table D10.A. However, if the series is not extended by forecasts (LEAD=0), then the default year of forecasts is displayed in Table D10.A. Forecast values in Table D10.A are calculated using the method shown on page 148 of Ladiray and Quenneville (2001) based on values that are displayed in Table D10. The regARIMA forecasts affect the D10.A forecasts only indirectly through the impact of the regARIMA forecasts on the seasonal factors that are shown in Table D10. If the SEATSDECOMP statement is specified, then *value* is increased to the minimum required for SEATS decomposition. For more information, see the section “SEATS Decomposition” on page 3349.

NBACKCAST=*value*

BACKCAST=*value*

NBACK=*value*

specifies the number of periods to backcast for regARIMA extension of the series. The default is NBACKCAST=0, which specifies that the series not be extended with backcasts. The maximum number of backcasts is 120. When the OUTBACKCAST option is specified, the NBACKCAST= value also controls the number of backcasts that are output to the OUT= data set specified in the OUTPUT statement. If the SEATSDECOMP statement is specified, then *value* is increased to the minimum required for SEATS decomposition. For more information, see the section "SEATS Decomposition" on page 3349.

OUT1STEP

specifies that the one-step-ahead forecasts be computed and displayed in addition to the multistep forecasts. The default is to compute and display only the multistep forecasts beginning at the forecast horizon.

OUTBACKCAST

OUTBKCAST

determines whether backcasts are included in certain tables sent to the output data set. If OUTBACKCAST is specified, then backcast values are included in the output data set for tables A6, A7, A8, A9, A10, B1, D10, D10B, D10D, D16, D16B, and D18. The default is not to include backcasts.

OUTFCST

OUTFORECAST

determines whether forecasts are included in certain tables sent to the output data set. If OUTFORECAST is specified, then forecast values are included in the output data set for Tables A6, A7, A8, A9, A10, B1, D10, D10B, D10D, D16, D16B, D18, and E18. The default is not to include forecasts. The OUTFORECAST option can be specified in either the X11 statement or the FORECAST statement with identical results.

ID Statement

ID *variables* **;**

If you are creating an output data set, use the ID statement to copy values of the ID variables, in addition to the table values, into the output data set. Or, if the VAR statement is omitted, all numeric variables that are not identified as BY variables, ID variables, the DATE= variable, or user-defined regressors are processed as time series. The ID statement has no effect when a VAR statement is specified and an output data set is not created. If the DATE= variable is specified in the PROC X13 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 forecasting) are extrapolated, while all other ID variables are missing in the forecast horizon.

IDENTIFY Statement

IDENTIFY *options* **;**

The IDENTIFY statement produces plots of the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) for identifying the ARIMA part of a regARIMA model. The sample ACF and PACF are produced for all combinations of the nonseasonal and seasonal differences of the data specified by the DIFF= and SDIFF= options.

The original series is first transformed as specified in the TRANSFORM statement.

If the model includes a regression component (specified using the REGRESSION, INPUT, and EVENT statements or the MDLINFOIN= data set in the PROC X13 statement), both the transformed series and the regressors are differenced at the highest order that is specified in the DIFF= and SDIFF= option. The parameter estimates are calculated using the differenced data. Then the undifferenced regression effects (with the exception of a constant term) are removed from the undifferenced data to produce undifferenced regression residuals. The ACFs and PACFs are calculated for the specified differences of the undifferenced regression residuals.

If the model does not include a regression component, then the ACFs and PACFs are calculated for the specified differences of the transformed data.

Tables displayed in association with identification are "Autocorrelation of Model Residuals" and "Partial Autocorrelation of Model Residuals." If the model includes a regression component (specified using the REGRESSION, INPUT, and EVENT statements or the MDLINFOIN= data set in the PROC X13 statement), then the "Regression Model Parameter Estimates" table is also displayed if the PRINTREG option is specified.

The following *options* can appear in the IDENTIFY statement:

DIFF=(*order***,** *order***,** *order***)**

specifies orders of nonseasonal differencing to use in model identification. The value 0 specifies no differencing, the value 1 specifies one nonseasonal difference $(1 - B)$, the value 2 specifies two nonseasonal differences $(1 - B)^2$, and so forth. The ACFs and PACFs are produced for all orders of nonseasonal differencing specified, in combination with all orders of seasonal differencing that are specified in the SDIFF= option. The default is DIFF=(0). You can specify up to three values for nonseasonal differences.

MAXLAG=*value*

specifies the number of lags for the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the regression residuals for model identification. The default is 36 for monthly series and 12 for quarterly series. MAXLAG applies to both tables and plots. The minimum value for MAXLAG= is 1.

PRINTREG

causes the "Regression Model Parameter Estimates" table to be printed if the REGRESSION statement is present. By default, this table is not printed.

SDIFF=(*order***,** *order***,** *order***)**

specifies orders of seasonal differencing to use in model identification. The value 0 specifies no seasonal differencing, the value 1 specifies one seasonal difference $(1 - B^s)$, the value 2 specifies two seasonal differences $(1 - B^s)^2$, and so forth. The value for s corresponds to the period specified in the SEASONS= option in the PROC X13 statement. The value of the SEASONS= option is supplied explicitly or is implicitly supplied through the INTERVAL= option or the values of the DATE= variable. The ACFs and PACFs are produced for all orders of seasonal differencing specified, in combination with all orders of nonseasonal differencing specified in the DIFF= option. The default is SDIFF=(0). You can specify up to three values for seasonal differences.

For example, the following statement produces ACFs and PACFs for two levels of differencing: $(1-B)$ and $(1 - B)(1 - B^s)$:

```
identify diff=(1) sdiff=(0, 1);
```

INPUT Statement

INPUT *variables* < */ options* > **;**

The INPUT statement specifies variables in the DATA= or AUXDATA= data set (which are specified in the PROC X13 statement) that are to be used as regressors in the regression portion of the regARIMA model. The variables in the data set should contain the values for each observation that define the regressor. Past values of regression variables should also be included in the DATA= or AUXDATA= data set if the time series listed in the VAR statement is to be extended with regARIMA backcasts. Similarly, future values of regression variables should also be included in the DATA= or AUXDATA= data set if the time series listed in the VAR statement is to be extended with regARIMA forecasts.

You can specify multiple INPUT statements. If you do not specify a MDLINFOIN= data set in the PROC X13 statement, then all variables listed in the INPUT statements are applied to all BY groups and all time series that are processed. If you specify a MDLINFOIN= data set, then the INPUT statements apply only if no regression information for the BY group and series is available in the MDLINFOIN= data set.

The INPUT statement provides the same functionality as the USERVAR= option in the REGRESSION statement. For more information about specifying user-defined regression variables, see the section "User-Defined Regression Variables" on page 3344, Example 46.6, and Example 46.11.

The following *options* can appear in the INPUT statement:

B=(*value* < **F** > *. . .* **)**

specifies initial or fixed values for the regression parameters in the order in which they appear in *variables*. Each B= list applies to the variable list that immediately precedes the slash.

For example, the following statements set an initial value of 1 for the user-defined regressor, x:

```
input y ;
input x / b=1 2 ;
```

In this example, the B= option applies only to the second INPUT statement. The value 2 is discarded because there is only one variable in the variable list.

To assign an initial value of 1 to the y regressor and 2 to the x regressor, use the following statements:

```
input y / b=1;
input x / b=2 ;
```

An **F** immediately following the numerical value indicates that this is not an initial value, but a fixed value. For an example that uses fixed parameters, see Example 46.8. In PROC X13, individual parameters can be fixed while other parameters in the same model are estimated.

USERTYPE=(*values***)**

enables a user-defined variable to be processed in the same manner as a US Census predefined variable. You can specify the following *values*: AO, CONSTANT, EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, LS, RP, SCEASTER, SEASONAL, TC, TD, TDSTOCK, THANKS, or USER. For example, the US Census Bureau EASTER(w) regression effects are included the "RegARIMA Holiday Component" table (A7). Specify USERTYPE=EASTER to include a user-defined variable that is processed exactly as the US Census predefined EASTER(w) variable, including inclusion in the A7 table. The NOAPPLY= option in the REGRESSION statement also changes the processing of variables based on the USERTYPE= value. Table 46.4 shows the regression types that are associated with each regression effects table.

Each USERTYPE= list applies to the variable list that immediately precedes the slash. The same rules for assigning B= values to regression variables apply for USERTYPE= options. For example, the following statements specify that the user-defined regressor in the variable MyEaster be processed exactly as the US Census predefined LOM variable:

```
input MyLOM;
input MyEaster / usertype=LOM EASTER;
```

In this example, the USERTYPE= option applies only to the MyEaster variable in the second INPUT statement. The USERTYPE value EASTER is discarded because there is only one variable in the variable list.

To assign the USERTYPE value LOM to the MyLOM variable and EASTER to the MyEaster variable, use the following statements:

```
input MyLOM / usertype=LOM;
input MyEaster / usertype=EASTER;
```

The following USERTYPE= options specify that the regression effect be removed from the seasonally adjusted series: EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, SCEASTER, SEASONAL, TD, TDSTOCK, THANKS, and USER. When a regression effect is removed from the seasonally adjusted series, the level (mean) of the seasonally adjusted series can be altered. It is often desirable to use a zero-mean (mean-adjusted) regressor for effects that are to be removed from the seasonally adjusted series. For an example that specifies a zero-mean regressor, see Example 46.6.

OUTLIER Statement

OUTLIER *options* ;

The OUTLIER statement specifies that the X13 procedure perform automatic detection of additive point outliers, temporary change outliers, level-shifts, or any combination of the three when using the specified model. After outliers are identified, the appropriate regression variables are incorporated into the model as “Automatically Identified Outliers,” and the model is reestimated. This procedure is repeated until no additional outliers are found.

The OUTLIER statement also identifies potential outliers and lists them in the “Potential Outliers” table in the displayed output. Potential outliers are identified by decreasing the critical value by the value that is specified in the ALMOST= option.

In the output, the initial critical values used for outlier detection in a given analysis are displayed in the table “Critical Values to Use in Outlier Detection.” Outliers that are detected and incorporated into the model are displayed in the output in the table “Regression Model Parameter Estimates,” where the regression variable is listed as “Automatically Identified.”

You can specify the following *options*:

ALMOST=*value*

specifies the difference between the critical value for an automatically identified outlier and a potential outlier that is "almost" identified. *value* is subtracted from the critical value that is used to identify outliers to form a critical value that more aggressively identifies potential outliers. Potential outliers are not included in the regARIMA model. However, potential outliers are displayed in the "Potential Outliers" table. *value* must be greater than 0. By default, ALMOST=0.5.

ALPHA=*value*

specifies the significance level to use for outlier identification, where critical values are calculated based on *value*. Any critical value that is specified in the CV=, AOCV=, LSCV=, or TCCV= option overrides the critical values that are calculated based on this option. *value* must be greater than 0 and less than or equal to 0.1. If you do not specify this option or the CV= option, the X-13ARIMA-SEATS method calculates the default initial critical value by assuming ALPHA=0.05.

AOCV=*value*

specifies a critical value to use for additive point outliers. If you specify this option, it overrides any default initial critical value for AO outliers. For more information, see the CV= option.

CV=*value*

specifies a default initial critical value to use for detecting all types of outliers. The absolute value of the t statistic that is associated with an outlier parameter estimate is compared with *value* to determine the significance of the outlier. If you do not specify this option, then the default initial critical value is computed based on the ALPHA= option, the CVMETHOD= option, and the number of observations for the model span that is used in the analysis. Table 46.2 shows initial critical values for various series lengths, which are based on the default values of the ALPHA= option and CVMETHOD= option. Increasing the critical value decreases the sensitivity of the outlier detection routine and can reduce the number of observations that are treated as outliers. The automatic model identification process might decrease the critical value by a certain percentage if the automatic model identification process fails to identify an acceptable model.

Table 46.2 Default Critical Values for Outlier Identification

Number of Observations	Outlier Critical Value
1	1.96
2	2.24
3	2.44
4	2.62
5	2.74
6	2.84
7	2.92
8	2.99
9	3.04
10	3.09
11	3.13
12	3.16
24	3.42
36	3.55
48	3.63

Table 46.2 *continued*

Number of Observations	Outlier Critical Value
72	3.73
96	3.80
120	3.85
144	3.89
168	3.92
192	3.95
216	3.97
240	3.99
264	4.01
288	4.03
312	4.04
336	4.05
360	4.07

CVMETHOD=CORRECTED | LJUNG

specifies the method to use to calculate the default initial critical value, based on the ALPHA= value and the number of observations for the model span that is used in the analysis. You can specify the following values:

CORRECTED uses a method that is a modification of the Ljung method in which critical values are interpolated based on the number of observations in the model span.

LJUNG uses a method that is based on the asymptotic formula described in Ljung (1993).

By default, CVMETHOD=CORRECTED.

LSCV=*value*

specifies a critical value to use for level-shift outliers. If you specify this option, it overrides any default initial critical value for LS outliers. For more information, see the CV= option.

LSRUN=*value*

specifies the maximum number of successive level-shift outliers to combine to form a temporary level-shift. Valid *values* for this option are 0 to 5, inclusive. If LSRUN=0 or LSRUN=1, no temporary level-shifts are evaluated. The evaluation of the temporary level-shifts is displayed in the "Tests for Cancellation of Level Shifts" table. By default, LSRUN=0.

METHOD=ADDALL | ADDONE

specifies whether to add outliers one at a time for each model estimation iteration or to add all outliers at once for each model estimation iteration. You can specify the following values:

ADDALL includes all significant outliers as regressors in the model, and then reestimates the model.

ADDONE adds the most significant outlier as a regressor in the model, and then reestimates the model.

For both methods, all candidate points for outliers are evaluated at each iteration and model estimation iterations continue until no remaining outliers are identified. By default, METHOD=ADDONE.

SPAN=(*mmmyy* **,***mmmyy* **)**

SPAN=(*'yyQq'* **,***'yyQq'* **)**

specifies the dates of the first and last observations to define a subset for searching for outliers. A single date in parentheses is interpreted to be the starting date of the subset. To specify only the ending date, use SPAN=(,*mmmyy*) or SPAN=(,*'yyQq'*). If the starting or ending date is omitted, then the first or last date, respectively, of the input data set or BY group is assumed. Because the dates are input as strings and the quarterly dates begin with a numeric character, the specification for a quarterly date must be enclosed in quotation marks. A four-digit year can be specified. If a two-digit year is specified, the value specified in the YEARCUTOFF= SAS system option applies.

TCCV=*value*

specifies a critical value to use for temporary change (TC) outliers. If you specify this option, it overrides any default initial critical value for TC outliers. For more information, see the CV= option.

TCRATE=*value*

specifies the rate of decay for temporary change outliers. *value* must be greater than 0 and less than 1. The default value is $(0.7)^{\frac{12}{period}}$, where *period* is the number of observations in one year.

TYPE=NONE

TYPE=(*outlier types***)**

lists the outlier types to be detected by the automatic outlier identification method. TYPE=NONE turns off outlier detection. The valid outlier types are AO, LS, and TC. The default is TYPE=(AO LS).

OUTPUT Statement

OUTPUT OUT=*SAS-data-set* < **YEARSEAS** > *tablename1 tablename2 . . .* **;**

The OUTPUT statement creates an output data set that contains specified tables. The data set is named by the OUT= option.

OUT=*SAS-data-set*

names the data set to contain the specified tables. If the OUT= option is omitted, the data set is named using the default DATA*n* convention.

YEARSEAS

YRSEAS

specifies that two additional variables be added to the OUT= data set. The two additional variables are the variables _YEAR_ and _SEASON_. The variable _YEAR_ contains the year of the date identifying the observation. The variable _SEASON_ contains the month for monthly data, or quarter for quarterly data, of the date that identifies the observation. For monthly data, the value of _SEASON_ is between 1 and 12. For quarterly data, the value of _SEASON_ is between 1 and 4. The _YEAR_ and _SEASON_ variables are useful when creating seasonal plots.

tablename1 tablename2 . . .

specify X13 *tablenames* that correspond to the title label used by the US Census Bureau X-13ARIMA-SEATS software. Specify one *tablename* for each table to be included in the output data set. Currently available *tablenames* are A1, A2, A6, A7, A8, A8AO, A8LS, A8TC, A9, A10, A19, B1, B7, B13, B17, B20, C1, C17, C20, D1, D7, D8, D8B, D8BX, D8BO, D8BL, D9, D10, D10B, D10D, D11, D11A, D11F, D11R, D12, D13, D16, D16B, D18, E1, E2, E3, E5, E6, E6A, E6R, E7, E8, E18, and MV1. Specifying D8B is equivalent to specifying D8, D8BX, D8BO, and D8BL because Table D 8.B displays the D8 series along with labels for extremes (D8BX), outliers (D8BO), and level shifts (D8BL). If no table is specified in the OUTPUT statement, Table A1 is output to the OUT= data set by default.

The *tablenames* that can be used in the OUTPUT statement are listed in the section "Displayed Output, ODS Table Names, and OUTPUT Tablename Keywords" on page 3350. The following is an example of a VAR statement and an OUTPUT statement:

```
var sales costs;
output out=out_x13  b1 d11;
```

The default variable name used in the output data set is the input variable name followed by an underscore and the corresponding table name. The variable sales_B1 contains the Table B1 values for the variable sales, the variable costs_B1 contains the Table B1 values for the variable costs, the variable sales_D11 contains the Table D11 values for the variable sales, and the variable costs_D11 contains the Table D11 values for the variable costs. If necessary, the variable name is shortened so that the table name can be added. If the DATE= variable is specified in the PROC X13 statement, then that variable is included in the output data set; otherwise, a variable named _DATE_ is written to the OUT= data set as the date identifier.

PICKMDL Statement (Experimental)

PICKMDL *options* **;**

The PICKMDL statement enables you to specify a variety of options for the PICKMDL method. The PICKMDL method uses models that are specified in the MDLINFOIN= data set to choose a regARIMA model. If the MDLINFOIN= option is not specified, then the PICKMDL method chooses a model from the list shown in Table 46.14. Example 46.9 demonstrates the use of the PICKMDL statement.

The PICKMDL statement cannot be specified when the AUTOMDL statement is also specified. The AUTOMDL and PICKMDL statements each specify different methods of automatic model selection. So only one of these methods can be used to select a model.

For more information about using the US Census Bureau's PICKMDL method for model selection, see the section "PICKMDL Model Selection" on page 3348.

You can specify the following *options* in the PICKMDL statement:

ARIMAMISS= ARIMASTMT | ZEROORDERS

specifies the method for interpreting missing ARIMA information in a model that is present in the MDLINFOIN= data set. You can specify the following values:

ARIMASTMT interprets missing information as the model that is specified in the MODEL= option of the ARIMA statement. This option should not be specified if the MDLINFOOUT= data set from a previous X13 procedure call is being used to replicate previous results. However, the (0 0 0)(0 0 0) model is not always the most appropriate model to use as a default when no model has been specified. This option enables you to specify default model orders.

ZEROORDERS interprets missing information as the (0 0 0)(0 0 0) model. This method is compatible with the output from the MDLINFOOUT= option.

By default, ARIMAMISS=ZEROORDERS.

MDLVAR=*variable*
specifies the variable in the MDLINFOIN= data set that identifies the models. A model identification variable is not required in the data set if fewer than two models are specified for each series. By default, MDLVAR= _MODEL_.

METHOD= BEST | FIRST
specifies the method for choosing the regARIMA model. You can specify the following values:

BEST chooses the best model.

FIRST chooses the first acceptable model.

By default, METHOD=FIRST.

REGRESSION Statement

REGRESSION *regression-group-options* **;**

REGRESSION PREDEFINED= *variables* < / **B=(***value* < **F** > **. . .)** > **;**

REGRESSION USERVAR= *variables* < / **B=(***value* < **F** > **. . .) USERTYPE=(***values***)** > **;**

The REGRESSION statement includes regression variables in a regARIMA model or specifies regression variables whose effects are to be removed by the IDENTIFY statement to aid in ARIMA model identification. Include the PREDEFINED= option to select predefined regression variables. Include the USERVAR= option to specify user-defined regression variables.

Table 46.3 shows the X-13ARIMA-SEATS tables that contain regression factors. Tables A8AO, A8LS, and A8TC are available only when more than one outlier type is present in the model.

Table 46.3 X-13ARIMA-SEATS Regression Effects Tables

Table	Regression Effects
A6	Trading day effects
A7	Holiday effects including Easter, Labor Day, and Thanksgiving-Christmas
A8	Combined effects of outliers, level-shifts, ramps, and temporary changes
A8AO	Point outlier effects; available only when more than one outlier type is present in the model
A8LS	Level-shift and ramp effects; available only when more than one outlier type is present in the model

Table 46.3 *continued*

Table	Regression Effects
A8TC	Temporary change effects; available only when more than one outlier type is present in the model
A9	User-defined regression effects
A10	User-defined seasonal component effects

Missing values in the span of an input series automatically create missing value regressors. For more information about missing values, see the NOTRIMMISS option in the PROC X13 statement and the section "Missing Values" on page 3340.

Combining your model with additional predefined regression variables can result in a singularity problem. To successfully perform the regression if a singularity occurs, you might need to alter either the model or the choices of the regressors.

To seasonally adjust a series that uses a regARIMA model, the factors derived from regression are used as multiplicative or additive factors, depending on the mode of seasonal decomposition. Therefore, regressors that are appropriate to the mode of the seasonal decomposition should be defined, so that meaningful combined adjustment factors can be derived and adjustment diagnostics can be generated. For example, if a regARIMA model is applied to a log-transformed series, then the regression factors are expressed as ratios, which match the form of the seasonal factors that are generated by the multiplicative or log-additive adjustment modes. Conversely, if a regARIMA model is fit to the original series, then the regression factors are measured on the same scale as the original series, which matches the scale of the seasonal factors that are generated by the additive adjustment mode. Note that the default transformation (no transformation) and the default seasonal adjustment mode (multiplicative) are in conflict. Thus, when you specify the X11 statement and any of the REGRESSION, INPUT, or EVENT statements, you must also either use the TRANSFORM statement to specify a transformation or use the MODE= option in the X11 statement to specify a different mode to seasonally adjust the data that uses the regARIMA model.

According to Ladiray and Quenneville (2001), "X-12-ARIMA is based on the same principle [as the X-11 method] but proposes, in addition, a complete module, called Reg-ARIMA, that allows for the initial series to be corrected for all sorts of undesirable effects. These effects are estimated using regression models with ARIMA errors (Findley et al. [23])." The REGRESSION, INPUT, and EVENT statements specify these regression effects. Predefined effects that can be corrected in this manner are listed in the PREDEFINED= option. You can create your own definitions to remove other effects by using the USERVAR= option and the EVENT statement.

You can specify either the PREDEFINED= option or the USERVAR= option, but not both, in a single REGRESSION statement. You can use multiple REGRESSION statements.

You can specify the following *regression-group-options* in the REGRESSION statement. The *regression-group-options* apply to all regression variables in a regression group. For predefined regression variables, the regression group is predefined. For user-defined regression variables, you can specify the regression group in the USERTYPE= option.

AICTEST=(EASTER | TD | TD1COEF | TD1NOLPYEAR | TDNOLPYEAR | TDSTOCK | USER)

specifies that an AIC-based selection be used to determine whether a given set of regression variables are to be included with the specified regARIMA model. For example, if you specify a trading day model selection, then AIC values (with a correction for the length of the series, henceforth referred to as AICC) are derived for models with and without the specified trading day variable. By default, the model with a smaller AICC is used to generate forecasts, identify outliers, and so on. If you specify more than one type of regressor, the AIC tests are performed sequentially in this order: (a) trading day regressors, (b) Easter regressors, (c) user-defined regressors. If there are several variables of the same type (for example, several trading day regressors), then AIC-based selection is applied to them as a group. That is, either all variables of this type or none are included in the final model. If you do not specify this option, no automatic AIC-based selection is performed.

If you use the AUTOMDL statement to identify the model and you also specify this option, then this option affects the model selection process in the following manner:

- AIC-based selection tests are performed on the default model.
- A new series is created by removing the regression effects that are identified in the default model from the original series. The automatic model identification process attempts to identify a model that is based on the new series.
- After a model is automatically identified, AIC-based selection tests that use the automatically identified model are performed on the original series.
- The default model, including regressors that are identified by using AIC-based selection, is compared to the automatically identified model, which also might include regressors that are identified by using AIC-based selections. The regressors for the two models can differ.

For more information about the X-13ARIMA-SEATS automatic modeling method, see section 7.2 of the *X-13ARIMA-SEATS Reference Manual* (US Bureau of the Census 2009).

EASTERMEANS=(YR400 | YR500 | SPAN)

specifies how the monthly means, which are used to remove seasonality from the EASTER predefined regressor, are calculated. When PREDEFINED=EASTER(w) is specified in the REGRESSION statement, monthly means are computed internally over the 500-year range from 1600 to 2099 by default. These monthly means are then used to remove seasonality from the Easter effect prior to calculating the Easter regression coefficient. The EASTERMEANS= option is ignored if no predefined EASTER regressor is included in the regression model or if SCEASTER(w) is the only predefined Easter regressor specified. You can specify the following values:

SPAN computes short-term monthly means rather than long-term monthly means to remove seasonality in the Easter effect. In this case, the monthly means are computed over the same span of data that is used to calculate the coefficient of the EASTER(w) regressor.

YR400 computes monthly means over the 400-year range from 1583 to 1982. This method was used in earlier versions of the X-13ARIMA-SEATS methodology.

YR500 computes monthly means over the 500-year range from 1600 to 2099.

By default, EASTERMEANS=YR500.

NOAPPLY=(AO | HOLIDAY | LS | TC | TD | USER | USERSEASONAL)

specifies a list of the types of regression effects whose model-estimated values are not to be removed from the original series before performing the seasonal adjustment calculations that are specified by the X11 statement. The NOAPPLY= option applies to the regression component values displayed in the X11 seasonal adjustment method regARIMA component tables as shown in Table 46.4.

Table 46.4 NOAPPLY= Types and Regression Effects

NOAPPLY= Option	Regression Effects Table	Description
AO	A8AO	Point outliers
HOLIDAY	A7	Easter, Labor Day, and Thanksgiving-to-Christmas holiday effects
LS	A8LS	Level changes and ramps
TC	A8TC	Temporary changes
TD	A6	Trading day effects
USER	A9	User-defined regression effects
USERSEASONAL	A10	User-defined seasonal regression effects

You can specify the following regression variable specification options in the REGRESSION statement.

PREDEFINED=CONSTANT | EASTER(*value*) | LABOR(*value*) | LOM | LOMSTOCK | LOQ | LPYEAR
PREDEFINED=SCEASTER(*value*) | SEASONAL | SINCOS(*value . . .*) | TD | TD1COEF
PREDEFINED=TD1NOLPYEAR | TDNOLPYEAR | TDSTOCK(*value*) | THANK(*value*)

lists the predefined regression variables to be included in the model. Data values for these variables are calculated by the program, mostly as functions of the calendar. Table 46.5 gives definitions for the available predefined variables. The values LOM and LOQ are equivalent: the actual regression is controlled by the SEASONS= option in the PROC X13 statement. You can specify multiple predefined regression variables. The syntax for using both a length-of-month and a seasonal regression can be in one of the following forms:

```
regression predefined=lom seasonal;

regression predefined=(lom seasonal);

regression predefined=lom predefined=seasonal;
```

The following restrictions apply when you use more than one predefined regression variable:

- You can specify only one of TD, TDNOLPYEAR, TD1COEF, or TD1NOLPYEAR.
- You cannot specify LPYEAR with TD, TD1COEF, LOM, LOMSTOCK, or LOQ.

- You cannot specify LOM or LOQ with TD or TD1COEF.
- If you specify the SINCOS predefined regression variable, then you must also specify the INTERVAL= option or the SEASONS= option in the PROC X13 statement because there are restrictions on this regression variable that are based on the frequency of the data.

The predefined regression variables, EASTER, LABOR, SCEASTER, SINCOS, TDSTOCK, and THANK, require extra parameters. Only one TDSTOCK regressor can be implemented in the regression model. If you specify multiple TDSTOCK variables, PROC X13 uses the last TDSTOCK variable specified. For EASTER, LABOR, SCEASTER, SINCOS, and THANK, you can specify the variables with different parameters to implement multiple regressors in the model. For example, the following statement specifies two EASTER regressors with widths 7 and 14:

```
regression predefined=easter(7) easter(14);
```

For SINCOS, specifying a parameter includes both the sine and the cosine regressor except for the highest order allowed (2 for quarterly data and 6 for monthly data.) For quarterly data, the following statement is the most common use of the SINCOS variable; it includes three regressors in the model:

```
regression predefined=sincos(1,2);
```

For monthly data, the following statement is the most common use of the SINCOS variable; it includes 11 regressors in the model:

```
regression predefined=sincos(1,2,3,4,5,6);
```

Table 46.5 Predefined Regression Variables in X-13ARIMA-SEATS

Regression Effect	Variable Definitions
Trend constant CONSTANT	$(1-B)^{-d}(1-B^s)^{-D} I(t \geq 1)$ where $I(t \geq 1) = \begin{cases} 1 & \text{for } t \geq 1 \\ 0 & \text{for } t < 1 \end{cases}$
Easter holiday EASTER(w)	$E(w,t) = \frac{1}{w} \times n_t$ and n_t is the number of the w days before Easter that fall in month (or quarter) t. (Note: This variable is 0 except in February, March, and April (or first and second quarter). It is nonzero in February only for $w > 22$.) Restriction: $1 \leq w \leq 25$.
Labor Day LABOR(w)	$L(w,t) = \frac{1}{w} \times$ [no. of the w days before Labor Day that fall in month t] (Note: This variable is 0 except in August and September.) Restriction: $1 \leq w \leq 25$.
Length-of-month (monthly flow) LOM	$m_t - \bar{m}$ where m_t = length of month t (in days) and $\bar{m} = 30.4375$ (average length of month)
Stock length-of-month LOMSTOCK	$SLOM_t = \begin{cases} m_t - \bar{m} - \mu(l) & \text{for } t = 1 \\ SLOM_{t-1} + m_t - \bar{m} & \text{otherwise} \end{cases}$ where $\bar{m}$ and m_t are defined in LOM and $\mu(l) = \begin{cases} 0.375 & \text{when first February in series is a leap year} \\ 0.125 & \text{when second February in series is a leap year} \\ -0.125 & \text{when third February in series is a leap year} \\ -0.375 & \text{when fourth February in series is a leap year} \end{cases}$
Length-of-quarter (quarterly flow) LOQ	$q_t - \bar{q}$ where q_t = length of quarter t (in days) and $\bar{q} = 91.3125$ (average length of quarter)

Table 46.5 *continued*

Regression Effect	Variable Definitions
Leap year (monthly and quarterly flow) LPYEAR	$LY_t = \begin{cases} 0.75 & \text{in leap year February (first quarter)} \\ -0.25 & \text{in other Februaries (first quarter)} \\ 0 & \text{otherwise} \end{cases}$
Statistics Canada Easter (monthly or quarterly flow) SCEASTER(w)	If Easter falls before April w, let n_E be the number of the w days on or before Easter that fall in March. Then: $E(w,t) = \begin{cases} n_E/w & \text{in March} \\ -n_E/w & \text{in April} \\ 0 & \text{otherwise} \end{cases}$ If Easter falls on or after April w, then $E(w,t) = 0$. (Note: This variable is 0 except in March and April (or first and second quarter).) Restriction: $1 \leq w \leq 24$.
Fixed seasonal SEASONAL	$M_{1,t} = \begin{cases} 1 & \text{in January} \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases}$ $, \ldots, M_{11,t} = \begin{cases} 1 & \text{in November} \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases}$
Fixed seasonal SINCOS(j) SINCOS($j_1, \ldots, j_n$)	$\sin(w_j t), \cos(w_j t)$, where $w_j = 2\pi j/s$, $1 \leq j \leq s/2$, and s is the seasonal period (drop $\sin(w_j t) \equiv 0$ for $j = s/2$) Restrictions: $1 \leq j_i \leq s/2$, $1 \leq n \leq s/2$.
Trading day TD, TDNOLPYEAR	$T_{1,t} =$ (number of Mondays) – (number of Sundays) $, \ldots, T_{6,t} =$ (number of Saturdays) – (number of Sundays)
One coefficient trading day TD1COEF, TD1NOLPYEAR	(number of weekdays) $- \frac{5}{2}$(number of Saturdays and Sundays)

Table 46.5 *continued*

Regression Effect	Variable Definitions
Stock trading day TDSTOCK(w)	$D_{1,t} = \begin{cases} 1 & \tilde{w}\text{th day of month } t \text{ is a Monday} \\ -1 & \tilde{w}\text{th day of month } t \text{ is a Sunday} \\ 0 & \text{otherwise} \end{cases}$ $, \ldots, D_{6,t} = \begin{cases} 1 & \tilde{w}\text{th day of month } t \text{ is a Saturday} \\ -1 & \tilde{w}\text{th day of month } t \text{ is a Sunday} \\ 0 & \text{otherwise} \end{cases}$ where $\tilde{w}$ is the smaller of w and the length of month t. For end-of-month stock series, set w to 31; that is, specify TDSTOCK(31). Restriction: $1 \le w \le 31$.
Thanksgiving THANK(w)	ThC(w, t) = proportion of days from w days before Thanksgiving through December 24 that fall in month t (negative values of w indicate days after Thanksgiving). (Note: This variable is 0 except in November and December.) Restriction: $-8 \le w \le 17$.

USERVAR=(*variables*)

specifies variables in the DATA= or AUXDATA= data set (which are specified in the PROC X13 statement) that are to be used as regressors. The variables in the data set should contain the values for each observation that define the regressor. Regression variables should also include future values in the data set for the forecast horizon if the time series is to be extended with regARIMA forecasts. Regression variables should include past values if the time series is to be extended with regARIMA backcasts. Missing values are not permitted within the data span, including backcasts and forecasts, of the user-defined regressors. Example 46.6 shows how to create an input data set that contains both the series to be seasonally adjusted and a user-defined input variable. Example 46.11 shows how to create an auxiliary data set that contains a user-defined input variable. For more information about specifying user-defined regression variables, see the section "User-Defined Regression Variables" on page 3344.

All regression variables in the USERVAR= option apply to all time series to be seasonally adjusted unless the MDLINFOIN= data set specifies different regression information. You cannot specify the PREDEFINED= option and the USERVAR= option in the same REGRESSION statement; however, you can specify multiple REGRESSION statements.

You can specify the following *options* for individual regression variables. Individual regression variable options are specified in the PREDEFINED= and USERVAR= options after the slash. The B= option can be specified in both the PREDEFINED= and USERVAR= options. Because the regression group is predefined for predefined variables, you can specify the USERTYPE= option only in the USERVAR= option.

B=(*value* < **F** > *. . .* **)**

specifies initial or fixed values for the regression parameters in the order in which they appear in a PREDEFINED= or USERVAR= option. Each B= list applies to the PREDEFINED= or USERVAR= variable list that immediately precedes the slash.

For example, the following statements set an initial value of 1 for the user-defined regressor, x:

```
regression predefined=LOM ;
regression uservar=x / b=1 2 ;
```

In this example, the B= option applies only to the USERVAR= option. The value 2 is discarded because there is only one variable in the USERVAR= list.

To assign an initial value of 1 to the LOM regressor and 2 to the x regressor, use the following statements:

```
regression predefined=LOM / b=1;
regression uservar=x / b=2 ;
```

An F immediately following the numerical value indicates that this is not an initial value, but a fixed value. For an example that uses fixed parameters, see Example 46.8. In PROC X13, individual parameters can be fixed while other parameters in the same model are estimated.

USERTYPE=(*values***)**

enables a variable that you define to be processed in the same manner as a US Census predefined variable. You can specify the following *values*: AO, CONSTANT, EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, LS, RP, SCEASTER, SEASONAL, TC, TD, TDSTOCK, THANKS, or USER. For example, the US Census Bureau EASTER(w) regression effects are included the "RegARIMA Holiday Component" table (A7). Specify USERTYPE=EASTER to define a variable that is processed exactly as the US Census predefined EASTER(w) variable, including inclusion in the A7 table. Each USERTYPE= list applies to the USERVAR= variable list that immediately precedes the slash. USERTYPE= does not apply to US Census predefined variables.

The same rules for assigning B= values to regression variables apply for USERTYPE= options. For example, the following statements specify that the user-defined regressor in the variable MyEaster be processed exactly as the US Census predefined LOM variable:

```
regression uservar=MyLOM;
regression uservar=MyEaster / usertype=LOM EASTER;
```

In this example, the USERTYPE= option applies only to the MyEaster variable in the second REGRESSION statement. The USERTYPE value EASTER is discarded because there is only one variable in the USERVAR= list.

To assign the USERTYPE value LOM to the MyLOM variable and EASTER to the MyEaster variable, use the following statements:

```
regression uservar=MyLOM / usertype=LOM;
regression uservar=MyEaster / usertype=EASTER;
```

The following USERTYPE= options specify that the regression effect be removed from the seasonally adjusted series: EASTER, HOLIDAY, LABOR, LOM, LOMSTOCK, LOQ, LPYEAR, SCEASTER, SEASONAL, TD, TDSTOCK, THANKS, and USER. When a regression effect is removed from the seasonally adjusted series, the level (mean) of the seasonally adjusted series can be altered. It is often desirable to use a zero-mean (mean-adjusted) regressor for effects that are to be removed from the seasonally adjusted series. For an example that specifies a zero-mean regressor, see Example 46.6.

SEATSDECOMP Statement (Experimental)

SEATSDECOMP OUT= *SAS-data-set* < *options* > **;**

The SEATSDECOMP statement creates an output data set (named by the OUT= option) that contains the SEATS decomposition series.

The following is an example of a VAR statement and a SEATSDECOMP statement:

```
var sales costs;
seatsdecomp out=SEATS_DECOMP;
```

The default variable name used in the output data set is the input variable name followed by an underscore and the corresponding table name. Because the B1 series is used as the original input series for the SEATS decomposition, the output data set SEATS_DECOMP from the example will contain the seasonal decomposition variables in the following order:

sales_OS	contains the Table B1 values for the variable sales.
sales_SC	contains the SEATS decomposition seasonal component for the variable sales.
sales_TC	contains the SEATS trend component values for the variable sales.
sales_SA	contains the SEATS seasonally adjusted series for the variable sales.
sales_IC	contains the SEATS irregular component for the variable sales.
costs_OS	contains the Table B1 values for the variable costs.
costs_SC	contains the SEATS decomposition seasonal component for the variable costs.
costs_TC	contains the SEATS trend component values for the variable costs.
costs_SA	contains the SEATS seasonally adjusted series for the variable costs.
costs_IC	contains the SEATS irregular component for the variable costs.

If necessary, the variable name is shortened so that the component name can be added. If you specify the DATE= variable in the PROC X13 statement, then that variable is included in the output data set; otherwise, a variable named _DATE_ is written to the OUT= data set as the date identifier. For more information about the output data set, see the section "SEATSDECOMP OUT= Data Set" on page 3357.

You can specify the following *options* in the SEATSDECOMP statement:

LEAD=*value*

specifies the number of periods ahead to forecast for a regARIMA extension of the series. The default is twice the number of periods in a year (8 or 24), and the maximum is 120. In the SEATS computations, the number of backcasts and forecasts are the same, and the minimum number is also dependent on the ARIMA model orders. For more information, see the section "SEATS Decomposition" on page 3349. If you specify a LEAD= value that is less than the default, then the number of forecasts specified in the LEAD= option are displayed in the OUT= data set. If the value of the LEAD= option and NBACKCAST= options in the FORECAST statement are less than the required number for SEATS decomposition, then the values of the LEAD= and NBACKCAST= options in the FORECAST statement are increased.

NBACKCAST=*value*

BACKCAST=*value*

NBACK=*value*

specifies the number of periods to backcast for a regARIMA extension of the series. The default is twice the number of periods in a year (8 or 24), and the maximum is 120. In the SEATS computations, the number of backcasts and forecasts are the same, and the minimum number is also dependent on the ARIMA model orders. For more information, see the section "SEATS Decomposition" on page 3349. If you specify a NBACKCAST= value that is less than the default, then the number of backcasts specified in the NBACKCAST= option are displayed in the OUT= data set. If the value of the LEAD= option and NBACKCAST= option specified in the FORECAST statement are less than the required number for SEATS decomposition when SEATSDECOMP is specified, then the value of LEAD= and NBACKCAST= in the FORECAST statement will be increased.

OUT=*SAS-data-set*

names the data set to contain the SEATS decomposition series: original series, seasonal component, trend component, seasonally adjusted series, irregular component. If the OUT= option is omitted, the data set is named using the default DATA*n* convention.

YEARSEAS

YRSEAS

specifies that two additional variables be added to the OUT= data set: _YEAR_ and _SEASON_. The variable _YEAR_ contains the year of the date that identifies the observation. The variable _SEASON_ contains the month for monthly data, or quarter for quarterly data, of the date that identifies the observation. For monthly data, the value of _SEASON_ is between 1 and 12. For quarterly data, the value of _SEASON_ is between 1 and 4. The _YEAR_ and _SEASON_ variables are useful when you create seasonal plots.

TABLES Statement

TABLES *tablename1 tablename2 . . . options* **;**

The TABLES statement enables you to alter the display of the PROC X13 tables. You can specify the display of tables that are not displayed by default by PROC X13, and the NOSUM option enables you to suppress the printing of the period summary line in the time series tables.

tablename1 tablename2 . . .

specifies X13 *tablenames* that correspond to the title label used by the US Census Bureau X-13ARIMA-SEATS software. For each table to be included in the displayed output, you must specify the X13 *tablename* keyword. Currently available tables are A19, B7, B13, B17, B20, C1, C20, D1, D7, E1, E2, and E3. Although these tables are not displayed by default, their values are sometimes useful in understanding the X-13ARIMA-SEATS method. For more information about the available tables, see the section "Displayed Output, ODS Table Names, and OUTPUT Tablename Keywords" on page 3350.

NOSUM

NOSUMMARY

NOSUMMARYLINE

applies to the tables available for output in the OUTPUT Statement. By default, these tables include a summary line that gives the average, total, or standard deviation for the historical data by period. The NOSUM option suppresses the display of the summary line in the listing. Also, if the tables are output with ODS, the summary line is not an observation in the data set. Thus, the output to the data set is only the time series, both the historical data and the forecast data, if available.

TRANSFORM Statement

TRANSFORM *options* **;**

The TRANSFORM statement transforms or adjusts the series prior to estimating a regARIMA model. With this statement, the series can be Box-Cox (power) transformed. The "Prior Adjustment Factors" table is associated with the TRANSFORM statement.

Only one of the following *options* can appear in the TRANSFORM statement:

POWER=*value*

transforms the input series, Y_t, by using a Box-Cox power transformation,

$$Y_t \rightarrow y_t = \begin{cases} \log(Y_t) & \lambda = 0 \\ \lambda^2 + (Y_t^\lambda - 1)/\lambda & \lambda \neq 0 \end{cases}$$

The power λ must be specified (for example, POWER=0.33). The default is no transformation ($\lambda = 1$); that is, POWER=1. The log transformation (POWER=0), square root transformation (POWER=0.5), and the inverse transformation (POWER=–1) are equivalent to the corresponding FUNCTION= option.

Table 46.6 Power Values Related to the Census Bureau Function Argument

FUNCTION=	Transformation	Range for Y_t	Equivalent Power Argument
NONE	Y_t	All values	POWER=1
LOG	$\log(Y_t)$	$Y_t > 0$ for all t	POWER=0
SQRT	$2(\sqrt{Y_t} - 0.875)$	$Y_t \geq 0$ for all t	POWER=0.5
INVERSE	$2 - \frac{1}{Y_t}$	$Y_t \neq 0$ for all t	POWER=–1
LOGISTIC	$\log(\frac{Y_t}{1-Y_t})$	$0 < Y_t < 1$ for all t	No equivalent

FUNCTION=NONE | LOG | SQRT | INVERSE | LOGISTIC | AUTO

specifies the transformation to be applied to the series prior to estimating a regARIMA model. The transformation used by FUNCTION=NONE, LOG, SQRT, INVERSE, or LOGISTIC is related to the POWER= option as shown in Table 46.6. FUNCTION=AUTO uses selection based on Akaike's information criterion (AIC) to decide between a log transformation and no transformation. The default is FUNCTION=NONE.

However, the FUNCTION= and POWER= options are not completely equivalent. In some cases, using the FUNCTION= option causes the program to automatically select other options. For example, FUNCTION=NONE causes the default mode to be MODE=ADD in the X11 statement. Also, the choice of transformation invoked by the FUNCTION=AUTO option can impact the default mode of the X11 statement.

There are restrictions on the value used in the POWER= and FUNCTION= options when preadjustment factors for seasonal adjustment are generated from a regARIMA model. When seasonal adjustment is requested with the X11 statement, any value of the POWER option can be used for the purpose of forecasting the series with a regARIMA model. However, this is not the case when factors generated from the regression coefficients are used to adjust either the original series or the final seasonally adjusted series. In this case, the only accepted transformations are the log transformation, which can be specified as POWER=0 for multiplicative or log-additive seasonal adjustments, and no transformation, which can be specified as POWER=1 for additive seasonal adjustments. If no seasonal adjustment is performed, any POWER transformation can be used. The preceding restrictions also apply when FUNCTION=NONE and FUNCTION=LOG are specified.

USERDEFINED Statement

USERDEFINED *variables* **;**

The USERDEFINED statement is used to identify the variables in the input data set or auxiliary data set that are available for user-defined regression. Only numeric variables can be specified. Specifying variables in the USERDEFINED statement does not include the variables as regressors. If a variable is specified in the INPUT statement or USERVAR= option in the REGRESSION statement, it is not necessary to include that variable in the USERDEFINED statement. However, if a variable is specified in the MDLINFOIN= data set in the PROC X13 statement and is not specified in an INPUT statement or in the USERVAR= option in the REGRESSION statement, then the variable should be specified in the USERDEFINED statement in order to make the variable available for regression.

VAR Statement

VAR *variables* **;**

The VAR statement specifies 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 that appear in a BY statement, ID statement, INPUT statement, or USERDEFINED statement; in the USERVAR= option in the REGRESSION statement; or in the DATE= option in the PROC X13 statement.

X11 Statement

X11 *options* **;**

The X11 statement is an optional statement for invoking seasonal adjustment by an enhanced version of the methodology of the US Census Bureau X-11 and X-11Q programs. You can control the type of seasonal adjustment decomposition calculated with the MODE= option. The output includes the final tables and diagnostics for the X-11 seasonal adjustment method listed in Table 46.7. Tables B7, B13, B17, B20, C1, E1, E2, E3, C20, D1, and D7 are not displayed by default; however, you can display these tables by requesting them in the TABLES statement.

Table 46.7 Tables Related to X11 Seasonal Adjustment

Table Name	Description
B1	Original series, adjusted for prior effects and forecast extended
B7	Preliminary trend-cycle, B iteration
B13	Irregular component, B iteration
B17	Preliminary weights for the irregular component
B20	Extreme values, B iteration
C1	Original series modified for outliers, trading day, and prior factors, C iteration
C17	Final weights for the irregular component
C20	Final extreme value adjustment factors
D1	Modified original data, D iteration
D7	Preliminary trend cycle, D iteration
D8	Final unmodified SI ratios (differences)
D8A	*F* tests for stable and moving seasonality, D8
D8B	Final unmodified SI ratios, with labels for outliers and extreme values
D9	Final replacement values for extreme SI ratios (differences), D iteration
D9A	Moving seasonality ratios for each period
SeasonalFilter	Seasonal filter statistics for Table D10
D10	Final seasonal factors
D10B	Seasonal factors, adjusted for user-defined seasonal
D10D	Final seasonal difference
D11	Final seasonally adjusted series
D11A	Final seasonally adjusted series with forced yearly totals
D11R	Rounded final seasonally adjusted series (with forced yearly totals)
TrendFilter	Trend filter statistics for Table D12
D12	Final trend cycle
D13	Final irregular component
D16	Combined seasonal and trading day factors
D16B	Final adjustment differences
D18	Combined calendar adjustment factors
E1	Original data modified for extremes
E2	Modified seasonally adjusted series
E3	Modified irregular series
E4	Ratio of yearly totals of original and seasonally adjusted series
E5	Percent changes (differences) in original series
E6	Percent changes (differences) in seasonally adjusted series
E6A	Percent changes (differences) in seasonally adjusted series with forced yearly totals (D11.A)
E6R	Percent changes (differences) in rounded seasonally adjusted series (D11.R)
E7	Percent changes (differences) in final trend component series
E8	Percent changes (differences) in original series adjusted for calendar factors (A18)
E18	Final adjustment ratios (original series to seasonally adjusted series)
F2A–F2I	X11 diagnostic summary
F3	Monitoring and quality assessment statistics
F4	Day of the week trading day component factors
G	Spectral plots

For more information about the X-11 seasonal adjustment diagnostics, see Shiskin, Young, and Musgrave (1967), Lothian and Morry (1978a), and Ladiray and Quenneville (2001).

You can specify the following *options* in the X11 statement:

FINAL=AO | LS | TC | USER |ALL

FINAL=(*options*)

lists the types of prior adjustment factors, obtained from the EVENT, REGRESSION, and OUTLIER statements, that are to be removed from the final seasonally adjusted series. Additive outliers are removed by specifying FINAL=AO. Level change and ramp outliers are removed by specifying FINAL=LS. Temporary change outliers are removed by specifying FINAL=TC. User-defined regressors or events (USERTYPE=USER) are removed by specifying FINAL=USER. All the preceding are removed by specifying FINAL=ALL or by specifying all the options in parentheses, FINAL=(AO LS TC USER). If this option is not specified, the final seasonally adjusted series contains these effects.

FORCE=TOTALS | ROUND | BOTH

specifies that the seasonally adjusted series be modified to: (a) force the yearly totals of the seasonally adjusted series and the original series to be the same (FORCE=TOTALS), (b) adjust the seasonally adjusted values for each calendar year so that the sum of the rounded seasonally adjusted series for any year equals the rounded annual total (FORCE=ROUND), or (c) first force the yearly totals, then round the adjusted series (FORCE=BOTH). When FORCE=TOTALS is specified, the differences between the annual totals is distributed over the seasonally adjusted values in a way that approximately preserves the month-to-month (or quarter-to-quarter) movements of the original series. For more information, see Huot (1975) and Cholette (1979). This forcing procedure is not recommended if the seasonal pattern is changing or if trading day adjustment is performed. Forcing the seasonally adjusted totals to be the same as the original series annual totals can degrade the quality of the seasonal adjustment, especially when the seasonal pattern is undergoing change. It is not natural if trading day adjustment is performed because the aggregate trading day effect over a year is variable and moderately different from zero.

MODE=ADD | MULT | LOGADD | PSEUDOADD

determines the mode of the seasonal adjustment decomposition to be performed. The four option choices correspond to additive, multiplicative, log-additive, and pseudo-additive decomposition, respectively. If this option is omitted, the procedure performs multiplicative adjustments. Table 46.8 shows the values of the MODE= option and the corresponding models for the original (O) and the seasonally adjusted (SA) series.

Table 46.8 Modes of Seasonal Adjustment and Their Models

Value of Mode Option	Name	Model for O	Model for SA
MULT	Multiplicative	$O = C \times S \times I$	$\mathrm{SA} = C \times I$
ADD	Additive	$O = C + S + I$	$\mathrm{SA} = C + I$
PSEUDOADD	Pseudo-additive	$O = C \times [S + I - 1]$	$\mathrm{SA} = C \times I$
LOGADD	Log-additive	$\log(O) = C + S + I$	$\mathrm{SA} = \exp(C + I)$

OUTFORECAST
OUTFCST

determines whether forecasts are included in certain tables sent to the output data set. If OUTFORECAST is specified, then forecast values are included in the output data set for Tables A6, A7, A8, A9, A10, B1, D10, D10B, D10D, D16, D16B, D18, and E18. The default is not to include forecasts. The OUTFORECAST option can be specified in either the X11 statement or the FORECAST statement with identical results.

SEASONALMA=S3X1 | S3X3 | S3X5 | S3X9 | S3X15 | STABLE | X11DEFAULT | MSR
SEASONALMA=*(filter-list-by-period)*

specifies which seasonal moving average (also called "seasonal filter") to use to estimate the seasonal factors. These seasonal moving averages are $n \times m$ *moving averages*, meaning that an n-term simple average is taken of a sequence of consecutive m-term simple averages. X11DEFAULT is the method used by the US Census Bureau's X-11-ARIMA program.

You can specify either a single filter option or a list. A single option indicates that all periods will use the same filter or the same method of identifying the filter. Alternately, you can specify the seasonal filters for each seasonal period by specifying SEASONALMA=*(filter-list-by-period)*, where *(filter-list-by-period)* lists the moving average filter for each period. For quarterly data, you must specify four filters; for monthly data, you must specify 12 filters. In the *filter-list-by-period*, you can specify S3X1, S3X3, S3X5, S3X9, or S3X15. For example, the following statement assigns a 3 × 1 moving average filter to the first quarter of a quarterly series and a 3 × 3 moving average to the second, third, and fourth quarters:

```
X11 SEASONALMA=( S3X1 S3X3 S3X3 S3X3 );
```

Table 46.9 describes the seasonal filter options available for the entire series:

Table 46.9 X-13ARIMA-SEATS Seasonal Filter Options and Descriptions

Filter Name	Description of Filter
S3X1	A 3 × 1 moving average
S3X3	A 3 × 3 moving average
S3X5	A 3 × 5 moving average
S3X9	A 3 × 9 moving average
S3X15	A 3 × 15 moving average
STABLE	Stable seasonal filter: a single seasonal factor for each calendar month or quarter is generated by calculating the simple average of all the values for each month or quarter (taken after detrending and outlier adjustment)
X11DEFAULT	Uses a 3 × 3 moving average to calculate the initial seasonal factors in each iteration and a 3 × 5 moving average to calculate the final seasonal factors
MSR	Filter chosen automatically by using the moving seasonality ratio of X-11-ARIMA/88 (Dagum 1988)

By default, SEASONALMA=MSR, which is the methodology of Statistic Canada's X-11-ARIMA/88 program.

SIGMALIM=(*lower limit***,** *upper limit* **)**

SIGMALIM=(*lower limit* **)**

SIGMALIM=(, *upper limit* **)**

specifies the lower and upper sigma limits in standard deviation units which are used to identify and down-weight extreme irregular values in the internal seasonal adjustment computations. One or both limits can be specified. The lower limit must be greater than 0 and not greater than the upper limit. If the lower sigma limit is not specified, then it defaults to a value of 1.5. The default upper sigma limit is 2.5. The comma must be used if the upper limit is specified.

Table 46.10 shows the effect of the SIGMALIM= option on the weights that are applied to the internal irregular values.

Table 46.10 Weights for Irregular Values

Weight	Sigma Limit
0	If $\frac{\lvert I_t - \mu \rvert}{\sigma_{1,I_t}} \geq$ *upper limit*
Partial weight	If *lower limit* $< \frac{\lvert I_t - \mu \rvert}{\sigma_{2,I_t}} <$ *upper limit*
1	If $\frac{\lvert I_t - \mu \rvert}{\sigma_{2,I_t}} \leq$ *lower limit*

In Table 46.10, μ is the theoretical mean of the irregular component, and σ_{1,I_t} and σ_{2,I_t} are the respective estimates of the standard deviation of the irregular component before and after extreme values are removed. The estimates of the standard deviation σ_{1,I_t} and σ_{2,I_t} vary with respect to t, and they are the same if no extreme values are removed. If they are different ($\sigma_{2,I_t} < \sigma_{1,I_t}$), then the first line in Table 46.10 is reevaluated with σ_{2,I_t}. In the special case where the lower limit equals the upper limit, the weight is 1 for $\frac{\lvert I_t - \mu \rvert}{\sigma_{2,I_t}} \leq$ *lower limit*, and 0 otherwise. For more information about how extreme irregular values are handled in the X11 computations, see Ladiray and Quenneville 2001, pp. 53–68, 122–125.

TRENDMA=*value*

specifies which Henderson moving average is used to estimate the final trend cycle. Any odd number greater than one and less than or equal to 101 can be specified (for example, TRENDMA=23). If the TRENDMA= option is not specified, the program selects a trend moving average based on statistical characteristics of the data. For monthly series, a 9-, 13-, or 23-term Henderson moving average is selected. For quarterly series, the program chooses either a 5- or a 7-term Henderson moving average.

TYPE=SA | SUMMARY | TREND

specifies the method used to calculate the final seasonally adjusted series (Table D11). The default method is TYPE=SA. This method assumes that the original series has not been seasonally adjusted. For method TYPE=SUMMARY, the trend cycle, irregular, trading day, and holiday factors are calculated, but not removed from the seasonally adjusted series. Thus, for TYPE=SUMMARY, Table D11 is the same as the original series. For TYPE=TREND, trading day, holiday, and prior adjustment factors are removed from the original series to calculate the seasonally adjusted series (Table D11) and also are used in the calculation of the final trend (Table D12).

Details: X13 Procedure

Data Requirements

The input data set must contain either quarterly or monthly time series, and the data must be sorted in chronological order within each BY group. For the standard X-13ARIMA-SEATS method, there must be at least three years of observations (12 for quarterly time series or 36 for monthly).

If an ARIMA model is specified in the ARIMA statement, AUTOMDL statement, PICKMDL statement, or the MDLINFOIN= data set, then more than three years of observations might be required in order to fit the ARIMA model and perform the computations associated with the seasonal decomposition and other diagnostics.

The minimum number of observations applies to each series listed in the VAR statement and within each BY group and is determined after any missing values are trimmed from the series.

Missing Values

PROC X13 can process a series with missing values.

Types of Missing Values

Missing values in a series are considered to be one of two types:

- A leading or trailing missing value occurs before the first nonmissing value or after the last nonmissing value, respectively, in the span of a series. The span of a series can be determined either explicitly by the SPAN= option or implicitly by the START= or DATE= option in the PROC X13 statement. By default, leading and trailing missing values are ignored. If you specify the NOTRIMMISS option in the PROC X13 statement, PROC X13 processes leading and trailing missing values according to the X-13ARIMA-SEATS missing value method.

- An embedded missing value occurs between the first nonmissing value and the last nonmissing value in the span of the series. PROC X13 processes embedded missing values according to the X-13ARIMA-SEATS missing value method.

X-13ARIMA-SEATS Missing Value Method

When the X-13ARIMA-SEATS method encounters a missing value, it inserts an additive outlier for the missing observation into the set of regression variables for the model of the series and then replaces the missing observation with a value large enough to be considered an outlier during model estimation. After the regARIMA model is estimated, the X-13ARIMA-SEATS method adjusts the original series by using factors that are generated from these missing value outlier regressors. The adjusted values are estimates of the missing values, and the adjusted series is displayed in Table MV1. The X-13ARIMA-SEATS missing value method requires the use of a regARIMA model to replace the missing values. Thus, either an ARIMA or AUTOMDL statement or the MDLINFOIN= option in the PROC X13 statement must be specified if there are embedded missing values in the time series.

SAS Predefined Events

SAS predefined events are summarized in this section. For complete details about SAS predefined events, see the section "EVENTKEY Statement" in *SAS Forecast Studio: User's Guide*.

Table 46.11 shows a summary of the SAS predefined event keywords. Table 46.12 lists the holiday date keywords that can be used as SAS predefined events. Table 46.13 lists the seasonal date keywords that can be used as SAS predefined events.

Table 46.11 Definitions for EVENTKEY Predefined Event Keywords

Variable Name or Variable Name Format	Description	Qualifier Options
AO<*obs*>OBS AO<*date*>D AO<*datetime*>DT	Outlier	TYPE=POINT VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=0)
LS<*obs*>OBS LS<*date*>D LS<*datetime*>DT	Level-shift	TYPE=LS VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=ALL)
TLS<*obs*>OBS<n> TLS<*date*>D<n> TLS<*datetime*>DT<n>	Temporary level-shift	TYPE=LS VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=<n>)
NLS<*obs*>OBS NLS<*date*>D NLS<*datetime*>DT	Negative level-shift	TYPE=LS VALUE=–1 BEFORE=(DURATION=0) AFTER=(DURATION=ALL)
CBLS<*obs*>OBS CBLS<*date*>D CBLS<*datetime*>DT	US Census Bureau level-shift	TYPE=LS VALUE=–1 SHIFT=–1 BEFORE=(DURATION=ALL) AFTER=(DURATION=0)
TC<*obs*>OBS TC<*date*>D TC<*datetime*>DT	Temporary change	TYPE=TC VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=ALL)
<date keyword>	Date pulse	TYPE=POINT VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=0) PULSE=DAY

Table 46.11 *continued*

Variable Name or Variable Name Format	**Description**	**Qualifier Options**
LINEAR QUAD CUBIC	Polynomial trends	TYPE=LIN TYPE=QUAD TYPE=CUBIC VALUE=1 BEFORE=(DURATION=ALL) AFTER=(DURATION=ALL) The default timing value is the 0 observation.
INVERSE LOG	Trends	TYPE=INV TYPE=LOG VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=ALL) The default timing value is the 0 observation.
<seasonal keywords>	Seasonal	TYPE=POINT PULSE= depends on keyword VALUE=1 BEFORE=(DURATION=0) AFTER=(DURATION=0) Timing values are based on keyword.

Table 46.12 Holiday Date Keywords and Definitions

Date Keyword	Definition
BOXING	December 26th
CANADA	July 1st
CANADAOBSERVED	July 1st, or July 2nd if July 1st is a Sunday
CHRISTMAS	December 25th
COLUMBUS	Second Monday in October
EASTER	Easter Sunday
FATHERS	Third Sunday in June
HALLOWEEN	October 31st
LABOR	First Monday in September
MLK	Third Monday in January
MEMORIAL	Last Monday in May
MOTHERS	Second Sunday in May
NEWYEAR	January 1st
THANKSGIVING	Fourth Thursday in November
THANKSGIVINGCANADA	Second Monday in October
USINDEPENDENCE	July 4th
USPRESIDENTS	Third Monday in February (since 1971)
VALENTINES	February 14th
VETERANS	November 11th
VETERANSUSG	Veterans Day date that is observed by US government for Monday–Friday schedule
VETERANSUSPS	Veterans Day date that is observed by US government for Monday–Saturday schedule (US Post Office)
VICTORIA	Monday on or preceding May 24th

Table 46.13 Seasonal Date Keywords and Definitions

Date Keyword	Definition
SECOND_1, . . . , SECOND_60	Specified second
MINUTE_1, . . . , MINUTE_60	Beginning of the specified minute
HOUR_1, . . . , HOUR_24	Beginning of the specified hour
SUNDAY, . . . , SATURDAY	All Sundays, and so on, in the time series
WEEK_1, . . . , WEEK_53	First day of the *n*th week of the year (PULSE=WEEK.n shifts this date for $n \neq 1$)
TENDAY_1, TENDAY_4, . . . , TENDAY_34	The first day of the month
TENDAY_2, TENDAY_5, . . . , TENDAY_35	The 11th day of the month
TENDAY_3, TENDAY_6, . . . , TENDAY_36	The 21st day of the month
SEMIMONTH_1, SEMIMONTH_3, . . . , SEMIMONTH_23	The first day of the month
SEMIMONTH_2, SEMIMONTH_4, . . . , SEMIMONTH_24	The 16th day of the month
JANUARY, . . . , DECEMBER	The first day of the specified month
QTR_1, QTR_2, QTR_3, QTR_4	The first date of the quarter indicated after the underscore (PULSE=QTR.n shifts this date for $n \neq 1$)
SEMIYEAR_1, SEMIYEAR_2	The first date of the semiyear (PULSE=SEMIYEAR.n shifts this date for $n \neq 1$)

User-Defined Regression Variables

The X-13ARIMA-SEATS method enables you to define regression variables to be included in the regARIMA model. A user-defined regression variable is composed of a value at each time series observation that you provide; the entire variable is implemented as a regressor in the regARIMA model. The regARIMA model is used in the seasonal decomposition process to extend the series prior to X11 decomposition. Because the X-13ARIMA-SEATS method does not impute, forecast, nor backcast user-defined regression variables, you must provide a nonmissing value at each observation in the span of the time series to be modeled and also provide a nonmissing value at each observation to be forecast or backcast.

A user-defined regression variable can be included in either the PROC X13 DATA= or AUXDATA= data set. You can supply the values for the user-defined regression variable by one of the following methods:

- You can include them in an auxiliary data set. The auxiliary data set should have a date variable that corresponds to the date variable in the DATA= data set. The name of the auxiliary data set is specified in the AUXDATA= option in the PROC X13 statement. The name of the date variable that exists in both the DATA= and AUXDATA= data sets is specified in the DATE= option in the PROC X13 statement. The observations in the auxiliary data set must span the entire series plus any forecast and backcast period.

- You can include them in the DATA= data set. Because the number of observations and the date values are exactly the same for both user-defined regressors and time series values, you need to include forecast and backcast values for user-defined regression variables beyond the span of the time series in one of the following ways:
  - You must specify leading missing values in the series to be seasonally adjusted for backcast periods. You must specify trailing missing values in the series to be seasonally adjusted for forecast periods. You must not use the NOTRIMMISS option in this case. The span of the series to be seasonally adjusted that is implied by trimming the leading and trailing missing values will be less than the span of the date values in the DATA= data set. Using this method, forecast error cannot be computed for the forecast and backcast periods.
  - You can use the SPAN= option in the PROC X13 statement to alter the span of the series to be seasonally adjusted to allow for backcast and forecast periods within the span of the date values in the DATA= data set. Using this method, forecast error can be computed for the forecast and backcast periods.

These methods of including user-defined regression variables in the regARMIA model are illustrated in Example 46.6 and Example 46.11.

If missing values for the user-defined regression variable are present within the span of the time series, including backcast and forecast observations, then an error message is displayed and the time series is not processed. If the span of the user-defined regression variable, or the span after leading and trailing missing values are trimmed, is not sufficient to cover the span of the series to be seasonally adjusted, including any backcasts and forecasts, then an error message is also displayed, and the time series is not processed.

Combined Test for the Presence of Identifiable Seasonality

The seasonal component of a 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 introduce distortion into the model. It is important to determine whether seasonality is identifiable without distorting the series.

For seasonality to be identifiable, the series should be identified as seasonal by using the "Test for the Presence of Seasonality Assuming Stability" and "Nonparametric Test for the Presence of Seasonality Assuming Stability." Also, since the presence of moving seasonality can cause distortion, it is important to evaluate the moving seasonality in conjunction with the stable seasonality to determine whether the seasonality is identifiable. The results of these tests are displayed in "F tests for Seasonality" (Table D8.A) in the X13 procedure.

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 follows (see also Figure 46.3):

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 X13 returns the conclusion, “Identifiable Seasonality Not Present.”

2. If the null hypothesis in step 1 is rejected, then PROC X13 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:

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

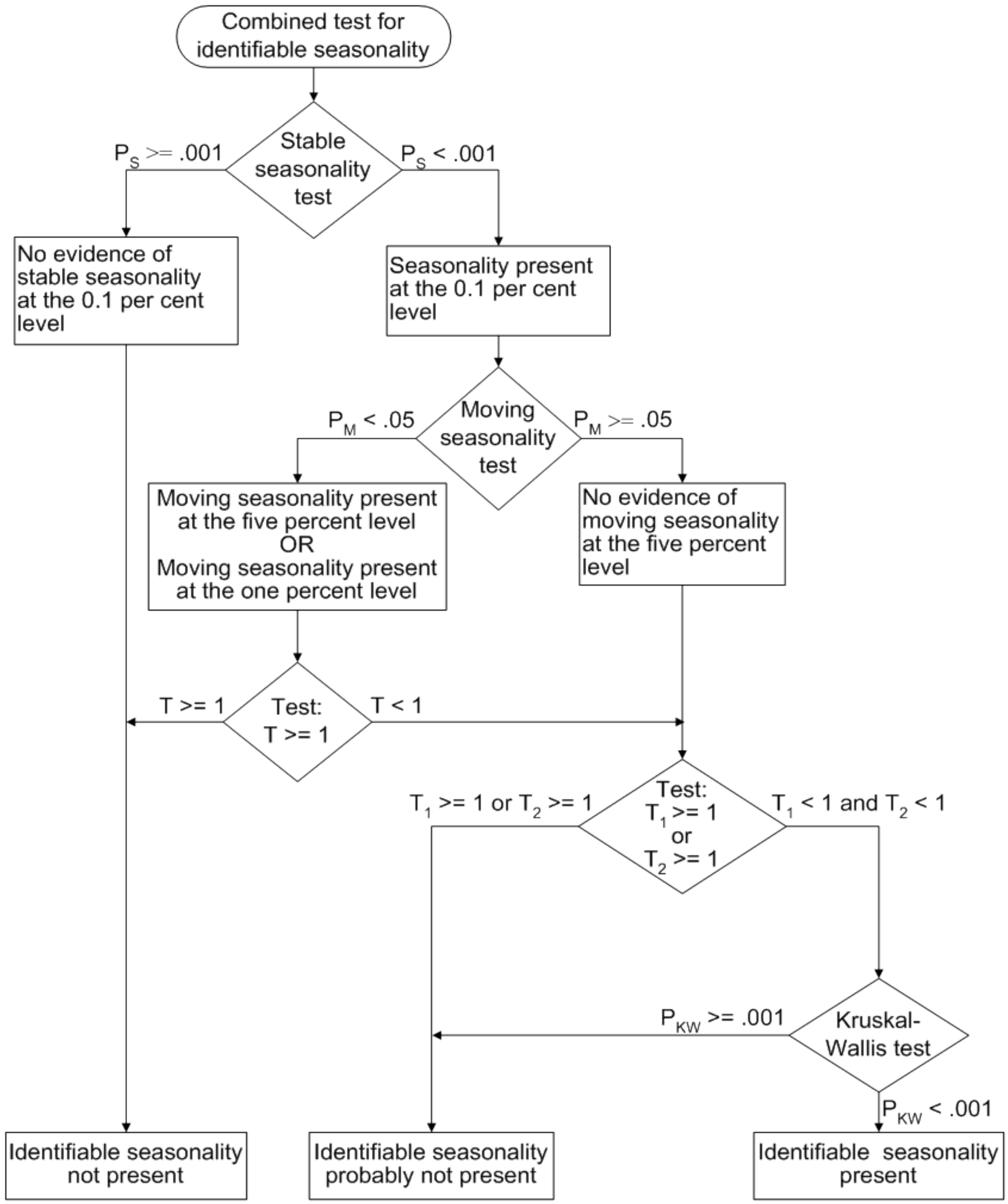
   If the null hypothesis of no moving seasonality is rejected at the 5.0% significance level ($P_M < 0.05$) and if $T \geq 1.0$, the null hypothesis of identifiable seasonality *not* present is not rejected and PROC X13 returns the conclusion, “Identifiable Seasonality Not Present.”

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 ($P_{\mathrm{KW}} \geq 0.001$), then PROC X13 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 X13 returns the conclusion “Identifiable Seasonality Present.”

Included in the displayed output of Table D8A is the table “Summary of Results and Combined Test for the Presence of Identifiable Seasonality.” This table displays the T_1, T_2, and T values and the significance levels for the stable seasonality test, the moving seasonality test, and the Kruskal-Wallis test. The last item in the table is the result of the combined test for identifiable seasonality.

Figure 46.3 Combined Seasonality Test Flowchart



Computations

For more information about the computations used in PROC X13, see the *X-13ARIMA-SEATS Reference Manual* (US Bureau of the Census 2009).

For more information about the X-11 method of decomposition, see *Seasonal Adjustment with the X-11 Method* (Ladiray and Quenneville 2001).

PICKMDL Model Selection

You can request that the X-13ARIMA-SEATS method select a model in a manner similar to the method used in X-11-ARIMA (Dagum 1988, 1983). Information about this model selection (PICKMDL) is based on the description in the *X-13ARIMA-SEATS Reference Manual* (US Bureau of the Census 2009). You can request the PICKMDL method in one of the following ways:

- by specifying the PICKMDL statement
- by specifying more than one value for the _MODEL_ variable in the MDLINFOIN= data set (subset by BY group and series)

The default settings for the PICKMDL automatic model selection method classify a model as acceptable if all of the following conditions are true:

- The absolute average percentage error of the extrapolated values within the last three years of data is less than 15%.
- The *p*-value is greater than 5% for the fitted model's Ljung-Box *Q* statistic test of the lack of correlation in the model's residuals.
- There are no signs of overdifferencing. Overdifferencing is indicated if the sum of the nonseasonal MA parameter estimates (for models with at least one nonseasonal difference) is greater than 0.9.

If a data set is specified in the MDLINFOIN= option and the data set contains more than one model for a series to be forecast, then the models described in the data set are candidates for the PICKMDL method of model selection. If the MDLINFOIN= option is not specified, then the candidate models are shown in Table 46.14, along with the order in which the models are considered. The order in which the model is considered is important when METHOD=FIRST is specified in the PICKMDL statement.

Table 46.14 PICKMDL Method Default ARIMA Models

Order of Candidate Model	ARIMA Model Orders
1	(0 1 1)(0 1 1)
2	(0 1 2)(0 1 1)
3	(2 1 0)(0 1 1)
4	(0 2 2)(0 1 1)
5	(2 1 2)(0 1 1)

No model is selected when none of the models in the MDLINFOIN= data set are acceptable. For more information about the output when no model is selected, see the section "Final Automatic Model Selection Table" on page 3353.

The regARIMA model consists of a transformation, a regression component, and an ARIMA model component. For each series, the following conditions hold:

- If no regression is specified in the MDLINFOIN= data set model but regressors are specified using the INPUT, EVENT, or REGRESSION statements, then the ARIMA models from the MDLINFOIN= data set are tested in conjunction with the regression variables specified in the INPUT, EVENT, and REGRESSION statements.
- If no ARIMA model is specified in the MDLINFOIN= data set but an ARIMA model is specified using an ARIMA statement or TRANSFORM statement, then the regression information from each model specified in the MDLINFOIN= data set is used in conjunction with the ARIMA model specified by the TRANSFORM and ARIMA statements.
- If no model information is specified in the MDLINFOIN= data set, then any model information specified by the TRANSFORM, INPUT, REGRESSION, EVENT, and ARIMA statements is used, and the PICKMDL statement is not in effect for that series.

SEATS Decomposition

PROC X13 can decompose the B1 series by using the SEATS decomposition method described in Gómez and Maravall (1997a, b). The SEATS decomposition method is planned for inclusion in the US Census Bureau's X13 program, which is not yet available for release.

The SEATS method requires the series to be extended with the same number of backcast and forecast observations. The number of observations backcast and forecast must meet the following minimum criteria:

- The number of forecast and backcast observations must be at least twice the number of observations in a year, with a minimum of 8.
- The number of forecast and backcast observations must be at least $2 \times (q + Q * s)$, where the ARIMA model used to extend the series is $(pdq)(PDQ)s$ in standard Box-Jenkins notation.
- The number of forecast and backcast observations must be at least $p + d + q + (P + D + Q) * s$, where the ARIMA model used to extend the series is $(pdq)(PDQ)s$ in standard Box-Jenkins notation.

If you specify the SEATSDECOMP statement and the number of forecasts or backcasts (either the default number or the number you specify) is not sufficient for SEATS decomposition, then the number of forecasts or backcasts is increased to the minimum required.

Displayed Output, ODS Table Names, and OUTPUT Tablename Keywords

The options specified in PROC X13 control both the tables produced by the procedure and the tables available for output to the OUT= data set specified in the OUTPUT statement.

The displayed output is organized into tables identified by a part letter and a sequence number within the part. The seven major parts of the X13 procedure are as follows:

A prior adjustments and regARIMA components (optional)

B preliminary estimates of irregular component weights and trading day regression factors (X-11 method)

C final estimates of irregular component weights and trading day regression factors

D final estimates of seasonal, trend cycle, and irregular components

E analytical tables

F summary measures

G charts

Table 46.15 describes the individual tables and charts. "P" indicates that the table is only displayed and is not available for output to the OUT= data set. Data from displayed tables can be extracted into data sets by using the Output Delivery System (ODS). For more information about the SAS Output Delivery System, see the *SAS Output Delivery System: User's Guide*. For more information about the features of the ODS Graphics system, including the many ways that you can control or customize the plots that are produced by SAS procedures, see Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

When tables available through the OUTPUT statement are output using ODS, the summary line is included in the ODS output by default. The summary line gives the average, standard deviation, or total by each period. The value –1 for YEAR indicates that the summary line is a total; the value –2 for YEAR indicates that the summary line is an average; and the value –3 for YEAR indicates that the line is a standard deviation. The value of YEAR for historical and forecast values is greater than or equal to zero. Thus, a negative value indicates a summary line. You can suppress the summary line altogether by specifying the NOSUM option in the TABLES statement. However, the NOSUM option also suppresses the display of the summary line in the displayed table.

"T" indicates that the table is available using the OUTPUT statement, but is not displayed by default; you must request that these tables be displayed by using the TABLES Statement. If there is no notation in the "Notes" column, then the table is available directly using the OUTPUT statement, without specifying the TABLES statement. If a table is not computed, then it is not displayed; if it is requested in the OUTPUT statement, then the variable in the OUT= data set contains missing values. The actual number of tables displayed depends on the options and statements specified.

Table 46.15 Table Names and Descriptions

Table	Description	Notes
Tables Associated with Model Order Identification		
ModelDescription	Regression model used in ARIMA model identification	P
ACF	Autocorrelation function	P
PACF	Partial autocorrelation function	P

Table 46.15 *continued*

Table	Description	Notes
Tables Associated with Automatic Modeling		
UnitRootTestModel	ARIMA estimates for unit root identification	P
UnitRootTest	Results of unit root test for identifying orders of differencing	P
AutoChoiceModel	Models estimated by automatic ARIMA model selection procedure	P
AutoLjungBox	Check of the residual Ljung-Box Q statistic	P
Best5Model	Best five ARIMA models chosen by automatic modeling	P
AutomaticModelChoice	Comparison of automatically selected model and default model	P
InitialModelChoice	Initial automatic model selection	P
FinalModelChecks	Final checks for identified model	P
FinalModelChoice	Final automatic model selection	P
Diagnostic Tables		
ErrorACF	Autocorrelation of regARIMA model residuals	P
ErrorPACF	Partial autocorrelation of regARIMA model residuals	P
SqErrorACF	Autocorrelation of squared regARIMA model residuals	P
ResidualOutliers	Outliers of the unstandardized residuals	P
ResidualStatistics	Summary statistics for the unstandardized residuals	P
NormalityStatistics	Normality statistics for regARIMA model residuals	P
G	Spectral analysis of regARIMA model residuals	P
Modeling Tables		
MissingExtreme	Extreme or missing values	P
ARMAIterationTolerances	Exact ARMA likelihood estimation iteration tolerances	P
IterHistory	ARMA iteration history	P
OutlierDetection	Critical values to use in outlier detection	P
PotentialOutliers	Potential outliers	P
TLSTest	Tests for cancellation of level-shifts	P
ARMAIterationSummary	Exact ARMA likelihood estimation iteration summary	P
ModelDescription	Model description for regARIMA model estimation	P
RegParameterEstimates	Regression model parameter estimates	P
RegressorGroupChiSq	Chi-squared tests for groups of regressors	P
ARMAParameterEstimates	Exact ARMA maximum likelihood estimation	P
AvgFcstErr	Average absolute percentage error in within-sample or without-sample forecasts or backcasts	P
Roots	Seasonal or nonseasonal AR or MA roots	P
MLESummary	Estimation summary	P
ForecastCL	Forecasts, standard errors, and confidence limits	P
MV1	Original series adjusted for missing value regressors	

Table 46.15 *continued*

Table	Description	Notes
Sequenced Tables		
A1	Original series	
A2	Prior-adjustment factors	
A6	RegARIMA trading day component	
A7	RegARIMA holiday component	
A8	RegARIMA combined outlier component	
A8AO	RegARIMA AO outlier component	
A8LS	RegARIMA level change outlier component	
A8TC	RegARIMA temporary change outlier component	
A9	RegARIMA user-defined regression component	
A10	RegARIMA user-defined seasonal component	
A19	RegARIMA outlier adjusted original data	T
B1	Prior-adjusted or original series	
B7	Preliminary trend-cycle, B iteration	T
B13	Irregular component, B iteration	T
B17	Preliminary weights for the irregular component	T
B20	Extreme values, B iteration	T
C1	Original series modified for outliers, trading day, and prior factors, C iteration	T
C17	Final weight for irregular components	
C20	Final extreme value adjustment factors	T
D1	Modified original data, D iteration	T
D7	Preliminary trend cycle, D iteration	T
D8	Final unmodified SI ratios	
D8A	Seasonality tests	P
D8B	Final unmodified SI ratios, with labels for outliers and extreme values	
D9	Final replacement values for extreme SI ratios	
D9A	Moving seasonality ratio	P
SeasonalFilter	Seasonal filter statistics for Table D10	P
D10	Final seasonal factors	
D10B	Seasonal factors, adjusted for user-defined seasonal	
D10D	Final seasonal difference	
D11	Final seasonally adjusted series	
D11A	Final seasonally adjusted series with forced yearly totals	
D11F	Factors applied to get adjusted series with forced yearly totals	
D11R	Rounded final seasonally adjusted series (with forced yearly totals)	
TrendFilter	Trend filter statistics for Table D12	P
D12	Final trend cycle	
D13	Final irregular series	
D16	Combined adjustment factors	

Table 46.15 *continued*

Table	Description	Notes
D16B	Final adjustment differences	
D18	Combined calendar adjustment factors	
E1	Original data modified for extremes	
E2	Modified seasonally adjusted series	
E3	Modified irregular series	
E4	Ratios of annual totals	P
E5	Percent changes in original series	
E6	Percent changes in final seasonally adjusted series	
E6A	Percent changes (differences) in seasonally adjusted series with forced yearly totals (D11.A)	
E6R	Percent changes (differences) in rounded seasonally adjusted series (D11.R)	
E7	Differences in final trend cycle	
E8	Percent changes (differences) in original series adjusted for calendar factors (A18)	
E18	Final adjustment ratios (original series to seasonally adjusted series)	
F2A-I	Summary measures	P
F3	Quality assessment statistics	P
F4	Day of the week trading day component factors	P
G	Spectral analysis	P

Final Automatic Model Selection Table

When the PICKMDL statement is specified and no model is selected, then the model in the "Final Automatic Model Selection" table is displayed as "(*, *, *) (*, *, *)" and an error message is displayed in both the log file and the output. If the "Final Automatic Model Selection" table is output to a data set, the model orders are output as –1, indicating the failure to select a model. For more information about PICKMDL model selection, see the section "PICKMDL Model Selection" on page 3348.

Table D 8.B

Table D8B displays the same series as Table D8. However, additional information is provided about the D8 series. The following values are displayed as labels for each observation of the series:

- The first label column indicates whether the D8 series value is extreme as determined by the X-11 extreme value method. An extreme observation is marked with an asterisk in the first label column. This data value is 0 or 1. If D8B or D8BX is specified in the OUTPUT statement, this value is output as the D8BX series to the data set that is specified in the OUT= option in the OUTPUT statement.

- The second label column contains the number of AO, TC, or RP outliers, if any, that affect the observation. This data value is 0 if no outliers affect the observation. Only the nonzero values are displayed in the table. If D8B or D8BO is specified in the OUTPUT statement, the number of outliers is output as the D8BO series to the data set that is specified in the OUT= option in the OUTPUT statement.
- The third label column indicates whether the observation is affected by level shift outliers as determined by an X-13ARIMA-SEATS method. This data value contains the number of level shifts that affect the observation. A nonzero value is displayed as "L". If D8B or D8BL is specified in the OUTPUT statement, the data values are output as the D8BL series to the data set that is specified in the OUT= option in the OUTPUT statement.

If any observations in Table D 8.B are affected by extremes, outliers, or level shifts, then notes that indicate the number of observations affected in each category are displayed at the end of the table.

Using Auxiliary Variables to Subset Output Data Sets

The X13 procedure can produce more than one table with the same name. For example, the following IDENTIFY statement produces ACF and PACF tables for two levels of differencing:

```
identify diff=(1) sdiff=(0, 1);
```

Auxiliary variables in the output data can be used to subset the data. In this example, the auxiliary variables Diff and SDiff specify the levels of regular and seasonal differencing that are used to compute the ACF. The following statements show how to retrieve the ACF results for the first differenced series:

```
ods select acf;
ods output acf=acf;
proc x13 data=sashelp.air date=date;
   identify diff=(1) sdiff=(0,1);
run;
title "Regular Difference=1 Seasonal Difference=0";
data acfd1D0;
   set acf(where=(Diff=1 and Sdiff=0));
run;
```

In addition to any BY variables, the auxiliary variables in the ACF and PACF data sets are _NAME_, _TYPE_, Transform, Adjust, Regressors, Diff, and SDiff. Auxiliary variables can be related to the group as shown in the Results Viewer (for example, BY variables, _NAME_, and _TYPE_). However, they can also be variables in the template where printing is suppressed by using PRINT=OFF. Auxiliary variables such as Transform, Adjust, and Regressors are not displayed in the ACF and PACF tables because similar information is displayed in the ModelDescription table that immediately precedes the ACF and PACF tables. The variables Diff and SDiff are not displayed because the levels of differencing are included in the title of the ACF and PACF tables.

The BY variables and the _NAME_ variable are available for all ODS OUTPUT data sets that are produced by the X13 procedure. The _TYPE_ variable is available for all ODS OUTPUT data sets that are produced during the model identification and model estimation stages. The _TYPE_ variable enables you to determine whether data in a table, such as the ModelDescription table, originated from the model identification stage or the model estimation stage.

The forecast data sets contain the auxiliary variable _SCALE_. The value of _SCALE_ is “Original” or “Transformed” to indicate the scale of the data. The auxiliary variable _SCALE_ is the same as the group in the Results Viewer. It is not displayed in the forecast tables because the table titles indicate the scale of the data.

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS” (*SAS/STAT User’s Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” in that chapter.

This section describes the use of ODS for creating graphics with the X13 procedure.

The graphs available through ODS Graphics are ACF plots, PACF plots, a residual histogram, spectral graphs, and forecasting plots. ACF and PACF plots for regARIMA model identification are not available unless the IDENTIFY statement is used. ACF plots, PACF plots, the residual histogram, and the residual spectral graph for diagnosis of the regARIMA model residuals are not available unless the CHECK statement is used. Forecasting plots are not available unless the FORECAST statement is used. A spectral plot of the original series is always available; however, additional spectral plots are provided when the X11 statement and CHECK statement are used. When ODS Graphics is not enabled, the ACF, PACF, and spectral analysis are displayed as columns of a table. The residual histogram is available only when ODS Graphics is enabled. To obtain a table that contains values related to the residual histogram, use the ODS OUTPUT statement.

ODS Graph Names

PROC X13 assigns a name to each graph it creates by using ODS. You can use these names to selectively reference the graphs. The names are listed in Table 46.16.

Table 46.16 ODS Graphs Produced by PROC X13

ODS Graph Name	Plot Description	PROC X13 PLOTS= Option
ACFPlot	Autocorrelation of regression residuals	SERIES(ACF)
ErrorACFPlot	Autocorrelation of regARIMA model residuals	RESIDUAL(ACF)
ErrorPACFPlot	Partial autocorrelation of regARIMA model residuals	RESIDUAL(PACF)
ForecastsOnlyPlot	Forecasts only of the original series	FORECAST(FORECASTONLY)
ForecastsOnlyPlot	Forecasts only of the transformed series	FORECAST(TRANSFORECASTONLY)
ForecastsPlot	Forecasts of the original series	FORECAST(FORECAST)

Table 46.16 *continued*

ODS Graph Name	Plot Description	PROC X13 PLOTS= Option
ForecastsPlot	Forecasts of the transformed series	FORECAST(TRANSFORECAST)
ModelForecastsPlot	Model and forecasts of the original series	FORECAST(MODELFORECASTS)
ModelForecastsPlot	Model and forecasts of the transformed series	FORECAST(TRANSMODELFORECASTS)
ModelPlot	Model of the original series	FORECAST(MODELS)
ModelPlot	Model of the transformed series	FORECAST(TRANSMODELS)
PACFPlot	Partial autocorrelation of regression residuals	SERIES(PACF)
ResidualHistogram	Distribution of regARIMA residuals	RESIDUAL(HIST)
SpectralPlot	Spectral plot of the seasonally adjusted series	ADJUSTED(SPECTRUM)
SpectralPlot	Spectral plot of irregular series	IRREGULAR(SPECTRUM)
SpectralPlot	Spectral plot of the regARIMA model residuals	RESIDUAL(SPECTRUM)
SpectralPlot	Spectral plot of the original series	SERIES(SPECTRUM)
SqErrorACFPlot	Autocorrelation of squared regARIMA model residuals	RESIDUAL(SQACF)

OUT= Data Set

You can use the OUTPUT statement to output the component series computed in the X-13ARIMA-SEATS decomposition.

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 specified or the variable _DATE_ if the DATE= option is not specified. If user-defined regressor variables or EVENT variables are specified, they are included. In addition, the various components specified by the table names in the OUTPUT statement are included in the OUT= data set.

The OUTPUT OUT= data set can contain the following variables:

BY variables
: are the BY variables used to subset the series by BY groups. The BY variables included in this data set match the BY variables, if any, used to process the series in the DATA= data set.

ID variables	enable the series observations to be identified using further information. The ID variables included in this data set match the ID variables, if any, specified in the ID statement and input from the DATA= data set.
DATE variable	is the time ID variable used to process the time series. It is either the variable specified in the DATE= option in the PROC X13 statement or the variable _DATE_ generated by the START= option in the PROC X13 statement.
YEAR variable	contains a value for the year of the date variable for the observation. This variable is included in the OUT= data set if YEARSEAS is specified in the OUTPUT statement.
SEASON variable	contains a value for the month or quarter of the date variable for the observation. This variable is included in the OUT= data set if YEARSEAS is specified in the OUTPUT statement.
User-defined variables	are variables specified in the INPUT statement or the USERVAR= option in the REGRESSION statement. The values of these variables are copied from the DATA= data set or from the AUXDATA= data set.
EVENT variables	variables specified in the EVENT statement. The values of these variables are computed based on the event definition and the dates of the time series observations.
Table variables	contains the data from the X-13ARIMA-SEATS decomposition tables: A1, A2, A6, A7, A8, A8AO, A8LS, A8TC, A9, A10, A19, B1, B7, B13, B17, B20, C1, C17, C20, D1, D7, D8, D8BX, D8BO, D8BL, D9, D10, D10B, D10D, D11, D11A, D11F, D11R, D12, D13, D16, D16B, D18, E1, E2, E3, E5, E6, E6A, E6R, E7, E8, E18, and MV1. The variable name used in the output data set is the input variable name followed by an underscore and the corresponding table name.

SEATSDECOMP OUT= Data Set

You can use the SEATSDECOMP statement to output the component series that is computed using the SEATS method of seasonal decomposition.

The OUT= data set specified in the SEATSDECOMP statement contains the BY variables (if any), the ID variables (if any), and either the DATE= variable if the DATE= option is specified or the variable _DATE_ if the DATE= option is not specified. If user-defined regressor variables or EVENT variables are specified, they are included. In addition, the five components computed by the SEATS decomposition method are included in the OUT= data set for each series.

The SEATSDECOMP OUT= data set can contain the following variables:

BY variables	are the BY variables used to subset the series by BY groups. The BY variables included in this data set match the BY variables (if any) that are used to process the series in the DATA= data set.
ID variables	enable the series observations to be identified using further information. The ID variables included in this data set match the ID variables (if any) that are specified in the ID statement and input from the DATA= data set.
DATE variable	is the time ID variable used to process the time series. It is either the variable specified in the DATE= option in the PROC X13 statement or the variable _DATE_ that is generated by the START= option in the PROC X13 statement.

YEAR variable — contains a value for the year of the date variable for the observation. This variable is included in the OUT= data set if YEARSEAS is specified in the OUTPUT statement.

SEASON variable — contains a value for the month or quarter of the date variable for the observation. This variable is included in the OUT= data set if YEARSEAS is specified in the OUTPUT statement.

User-defined variables — are variables specified in the INPUT statement or the USERVAR= option in the REGRESSION statement. The values of these variables are copied from the DATA= data set or from the AUXDATA= data set.

EVENT variables — are variables that are specified in the EVENT statement. The values of these are computed based on the event definition and the dates of the time series observations.

Component variables — contains the data from the SEATS decomposition tables. The variable name used in the output data set is the input variable name followed by an underscore and the corresponding table name.

<variable>_OS — contains the original series for SEATS decomposition. This is the B1 series from the X-13ARIMA-SEATS method.

<variable>_SC — contains the seasonal component series that is calculated by SEATS decomposition.

<variable>_TC — contains the trend component series that is calculated by SEATS decomposition.

<variable>_SA — contains the seasonally adjusted series that is calculated by SEATS decomposition.

<variable>_IC — contains the irregular series that is calculated by SEATS decomposition.

Special Data Sets

The X13 procedure can read a MDLINFOIN= input data set and output a MDLINFOOUT= data set. The structure of both of these data sets is the same. The difference is that when the MDLINFOIN= data set is read, only information relative to specifying a model is processed, whereas the MDLINFOOUT= data set contains the results of estimating a model. The X13 procedure can also read data sets that contain event definition data. The structure of these data sets is the same as in the SAS High-Performance Forecasting system.

MDLINFOIN= and MDLINFOOUT= Data Sets

The MDLINFOIN= and MDLINFOOUT= data sets can contain one or more of the following variables:

BY variables — enable the model information to be specified by BY groups. BY variables can be included in this data set that match the BY variables used to process the series. If no BY variables are included, then the models specified by _NAME_ in the MDLINFOIN= data set apply to all BY groups in the DATA= data set.

NAME — contains the variable name of the time series to which a particular model is to be applied. Omit the _NAME_ variable if you are specifying the same model for all series in a BY group.

MODEL — contains a name to identify the model for this observation. You can specify a name other than _MODEL_ in the MDLVAR= option in the PICKMDL statement. The _MODEL_ variable is an ID variable; all observations that have the same value of this variable belong to the same model. This variable is used to identify different model candidates when the PICKMDL method is used to choose a model; it is not needed if only a single model is specified.

MODELTYPE specifies whether the observation contains regression or ARIMA information. The value of _MODELTYPE_ should be either REG to supply regression information or ARIMA to supply model information. If valid regression information exists in the MDLINFOIN= data set for a BY group and series being processed, then the REGRESSION, INPUT, and EVENT statements are ignored for that BY group and series. Likewise, if valid ARIMA model information exists in the data set, then the AUTOMDL, ARIMA, and TRANSFORM statements are ignored. Valid values for the other variables in the data set depend on the value of the _MODELTYPE_ variable. Although other values of _MODELTYPE_ might be permitted in other SAS procedures, PROC X13 recognizes only REG and ARIMA.

MODELPART further qualifies the regression information in the observation. For _MODELTYPE_=REG, valid values of _MODELPART_ are INPUT, EVENT, and PREDEFINED. A value of INPUT indicates that this observation refers to the user-defined variable whose name is given in _DSVAR_. Likewise, a value of EVENT indicates that the observation refers to the SAS or user-defined event whose name is given in _DSVAR_. PREDEFINED indicates that the name given in _DSVAR_ is a predefined US Census Bureau variable. If only ARIMA model information is included in the data set (that is, all observations have _MODELTYPE_=ARIMA), then the _MODELPART_ variable can be omitted. For observations where _MODELTYPE_=ARIMA, valid values for _MODELPART_ are FORECAST, ".", or blank.

COMPONENT further qualifies the regression or ARIMA information in the observation. For _MODELTYPE_=REG, the only valid value of _COMPONENT_ is SCALE. For _MODELTYPE_= ARIMA, the valid values of _COMPONENT_ are TRANSFORM, CONSTANT, NONSEASONAL, and SEASONAL. TRANSFORM indicates that the observation contains the information that would be supplied in the TRANSFORM statement. CONSTANT is specified to control the constant term in the model. NONSEASONAL and SEASONAL refer to the AR, MA, and differencing terms in the ARIMA model.

PARMTYPE further qualifies the regression or ARIMA information in the observation. For _MODELTYPE_=REG, the value of _PARMTYPE_ is the same as the value of the USERTYPE= option in the REGRESSION statement. Since the USERTYPE= option applies only to user-defined events and variables, the value of _PARMTYPE_ does not alter processing in observations where _MODELPART_=PREDEFINED. However, it is consistent to use a value for _PARMTYPE_ that matches the US Census Bureau predefined variable. For the constant term in the model information, _PARMTYPE_ should be SCALE. For transformation information, the value of _PARMTYPE_ should be NONE, LOG, LOGIT, SQRT, or BOXCOX. For _MODELTYPE_=ARIMA, valid values of _PARMTYPE_ are AR, MA, and DIF.

DSVAR specifies the variable name associated with the current observation. For _MODELTYPE_=REG, the value of _DSVAR_ is the name of the user-defined variable, the event, or the US Census Bureau predefined variable. For _MODELTYPE_=ARIMA, _DSVAR_ should match the name of the series being processed. If the ARIMA model information applies to more than one series, then _DSVAR_ can be blank or ".", equivalently.

VALUE contains a numerical value that is used as a parameter for certain types of information. For example, the PREDEFINED=EASTER(6) option in the REGESSION statement is implemented in the MDLINFOIN= data set by using _DSVAR_=EASTER and

VALUE=6. For a BOXCOX transformation, _VALUE_ is set equal to the λ parameter value. For _COMPONENT_=SEASONAL, if _VALUE_ is nonmissing, then _VALUE_ is used as the seasonal period. If _VALUE_ is missing for _COMPONENT_=SEASONAL, then the seasonal period is determined by the interval of the series.

FACTOR
: applies only to the AR and MA portions of the ARIMA model. The value of _FACTOR_ identifies the factor of the given AR or MA term. Therefore, the value of _FACTOR_ is the same for all observations that are related to the same factor.

LAG
: identifies the degree for differencing and AR and MA lags. If _COMPONENT_=SEASONAL, then the value in _LAG_ is multiplied by the seasonal period indicated by the value of _VALUE_.

SHIFT
: contains the shift value for transfer functions. This value is not processed by PROC X13, but it might be processed by other procedures in which transfer functions can be specified.

NOEST
: indicates whether a parameter associated with the observation is to be estimated. For example, the NOINT option is indicated by _COMPONENT_=CONSTANT with _NOEST_=1 and _EST_=0. _NOEST_=1 indicates that the value in _EST_ is a fixed value. _NOEST_ pertains to the constant term, to AR and MA parameters, and to regression parameters.

EST
: contains an initial or fixed value for a parameter associated with the observation that is to be estimated. _NOEST_=1 indicates the value in _EST_ is a fixed value. _EST_ pertains to the constant term, to AR and MA parameters, and to regression parameters.

STDERR
: contains output information about estimated parameters. The variable _STDERR_ is not processed by the MDLINFOIN= data set for PROC X13. In the MDLINFOOUT= data set, _STDERR_ contains the standard error that pertains to the estimated parameter in the variable _EST_.

TVALUE
: contains output information about estimated parameters. The variable _TVALUE_ is not processed by the MDLINFOIN= data set for PROC X13. In the MDLINFOOUT= data set, _TVALUE_ contains the *t* value that pertains to the estimated parameter in the variable _EST_.

PVALUE
: contains output information about estimated parameters. The variable _PVALUE_ is not processed by the MDLINFOIN= data set for PROC X13. In the MDLINFOOUT= data set, _PVALUE_ contains the *p*-value that pertains to the estimated parameter in the variable _EST_.

LABEL
: contains a character string. The value of the variable _LABEL_ does not affect the model that is input when the data set is specified in the MDLINFOIN= option. The user can store any string in the variable _LABEL_. If a model is selected from the MDLINFOIN= data set, then the value of the variable _LABEL_ (if any) for the first observation corresponding to that model is output to the MDLINFOOUT= data set (if specified).

INEVENT= Data Set

The INEVENT= data set can contain the following variables. When a variable is omitted from the data set, that variable is assumed to have the default value for all observations. The default values are specified in the list.

NAME specifies the event variable name. _NAME_ is displayed with the case preserved. Since _NAME_ is a SAS variable name, the event can be referenced by using any case. The _NAME_ variable is required; there is no default.

CLASS specifies the class of event: SIMPLE, COMBINATION, PREDEFINED. The default for _CLASS_ is SIMPLE.

KEYNAME contains either a date keyword (SIMPLE EVENT), a predefined event variable name (PREDEFINED EVENT), or an event name (COMBINATION EVENT). All _KEYNAME_ values are displayed in upper case. However, if the _KEYNAME_ value refers to an event name, then the actual name can be of mixed case. The default for _KEYNAME_ is no keyname, designated by ".".

STARTDATE contains either the date timing value or the first date timing value to use in a do-list. The default for _STARTDATE_ is no date, designated by a missing value.

ENDDATE contains the last date timing value to use in a do-list. The default for _ENDDATE_ is no date, designated by a missing value.

DATEINTRVL contains the interval for the date do-list. The default for _DATEINTRVL_ is no interval, designated by ".".

STARTDT contains either the datetime timing value or the first datetime timing value to use in a do-list. The default for _STARTDT_ is no datetime, designated by a missing value.

ENDDT contains the last datetime timing value to use in a do-list. The default for _ENDDT_ is no datetime, designated by a missing value.

DTINTRVL contains the interval for the datetime do-list. The default for _DTINTRVL_ is no interval, designated by ".".

STARTOBS contains either the observation number timing value or the first observation number timing value to use in a do-list. The default for _STARTOBS_ is no observation number, designated by a missing value.

ENDOBS contains the last observation number timing value to use in a do-list. The default for _ENDOBS_ is no observation number, designated by a missing value.

OBSINTRVL contains the interval length of the observation number do-list. The default for _OBSINTRVL_ is no interval, designated by ".".

TYPE specifies the type of event. The valid values of _TYPE_ are POINT, LS, RAMP, TR, TEMPRAMP, TC, LIN, LINEAR, QUAD, CUBIC, INV, INVERSE, LOG, and LOGARITHMIC. The default for _TYPE_ is POINT.

VALUE specifies the value for nonzero observation. The default for _VALUE_ is 1.0.

PULSE specifies the interval that defines the units for the duration values. The default for _PULSE_ is no interval, designated by ".".

_DUR_BEFORE_ specifies the number of durations before the timing value. The default for _DUR_BEFORE_ is 0.

_DUR_AFTER_ specifies the number of durations after the timing value. The default for _DUR_AFTER_ is 0.

_SLOPE_BEF_ determines whether the curve is GROWTH or DECAY before the timing value for _TYPE_=RAMP, _TYPE_=TEMPRAMP, and _TYPE_=TC. Valid values are GROWTH and DECAY. The default for _SLOPE_BEF_ is GROWTH.

_SLOPE_AFT_ determines whether the curve is GROWTH or DECAY after the timing value for _TYPE_=RAMP, _TYPE_=TEMPRAMP, and _TYPE_=TC. Valid values are GROWTH and DECAY. The default for _SLOPE_AFT_ is GROWTH unless _TYPE_=TC; then the default is DECAY.

SHIFT specifies the number of _PULSE_= intervals to shift the timing value. The shift can be positive (forward in time) or negative (backward in time). If _PULSE_= is not specified, then the shift is in observations. The default for _SHIFT_ is 0.

TCPARM specifies the parameter for EVENT of TYPE=TC. The default for _TCPARM_ is 0.5.

RULE specifies the rule to use when combining events or when timing values of an event overlap. The valid values of _RULE_ are ADD, MAX, MIN, MINNZ, MINMAG, and MULT. The default for _RULE_ is ADD.

PERIOD specifies the frequency interval at which the event should be repeated. If this value is missing, then the event is not periodic. The default for _PERIOD_ is no interval, designated by ".".

LABEL specifies the label or description for the event. If a label is not specified, then the default label value is displayed as ".". For events that produce dummy variables, either the user-supplied label or the default label is used. For COMPLEX events, the _LABEL_ value is merely a description of the group of events.

OUTSTAT= Data Set

The OUTSTAT= data set can contain the following variables:

BY variables sorts the statistics into BY groups. BY variables are included in this data set that match the BY variables used to process the series.

NAME specifies the variable name of the time series to which the statistics apply.

STAT describes the statistic that is stored in VALUE or CVALUE. STAT takes the following values:

- Period — the period of the series, 4 or 12.
- Mode — the mode of the seasonal adjustment from the X11 statement. Possible values are ADD, MULT, LOGADD, and PSEUDOADD.
- Start — the beginning of the model span expressed as *monyyyy* for monthly series or *yyyy*Q*q* for quarterly series.
- End — the end of the model span expressed as *monyyyy* for monthly series or *yyyy*Q*q* for quarterly series.
- NbFcst — the number of forecast observations.
- SigmaLimLower — the lower sigma limit in standard deviation units.
- SigmaLimUpper — the upper sigma limit in standard deviation units.
- pLBQ_24 — the Ljung-Box Q statistic of the residuals at lag 24, for monthly series. Note that lag 12 (pLBQ_12) and lag 16 (pLBQ_16) are included in the data set for quarterly series.
- D8Fs — the stable seasonality F test value from Table D8.

D8Fm	the moving seasonality *F* test value from Table D8.
ISRatio	the final irregular-to-seasonal ratio from Table F 2.H.
SMA_ALL	the final seasonal moving average filter for all periods.
RSF	the residual seasonality *F* test value for Table D11 for the entire series.
RSF3	the residual seasonality *F* test value for Table D11 for the last three years.
RSFA	the residual seasonality *F* test value for Table D11.A for the entire series.
RSF3A	the residual seasonality *F* test value for Table D11.A for the last three years.
RSFR	the residual seasonality *F* test value for Table D11.R for the entire series.
RSF3R	the residual seasonality *F* test value for Table D11.R for the last three years.
TMA	the Henderson trend moving average filter selected.
ICRatio	the final irregular-to-trend cycle ratio from Table F 2.H.
E5sd	the standard deviation from Table E5.
E6sd	the standard deviation from Table E6.
E6Asd	the standard deviation from Table E6.A.
MCD	months of cyclical dominance.
Q	the overall level Q from Table F3.
Q2	Q overall level without M2 from Table F3.

FMT
: indicates whether the format is numeric or character. FMT="NUM" if the value is numeric and stored in the VALUE variable. FMT="CHAR" if the value is a string and stored in the CVALUE variable.

VALUE
: contains the numerical value of the statistic or missing if the statistic is of type character.

CVALUE
: contains the character value of the text statistic or blank if the statistic is of type numeric.

Examples: X13 Procedure

Example 46.1: ARIMA Model Identification

This example shows typical PROC X13 statements that are used for ARIMA model identification. This example invokes the X13 procedure and uses the TRANSFORM and IDENTIFY statements. It specifies the time series data, takes the logarithm of the series (TRANSFORM statement), and generates ACFs and PACFs for the specified levels of differencing (IDENTIFY statement). The ACFs and PACFs for DIFF=1 and SDIFF=1 are shown in Output 46.1.1, Output 46.1.2, Output 46.1.3, and Output 46.1.4. The data set is the same as in the section "Basic Seasonal Adjustment" on page 3288.

The graphical displays are available when ODS Graphics is enabled. For more information about the graphics available in the X13 procedure, see the section "ODS Graphics" on page 3355.

```
proc x13 data=sales date=date;
   var sales;
   transform power=0;
   identify diff=(0,1) sdiff=(0,1);
run;
```

Output 46.1.1 ACFs (Nonseasonal Order=1 Seasonal Order=1)

The X13 Procedure

Autocorrelation of Regression Residuals for ARIMA Model Identification

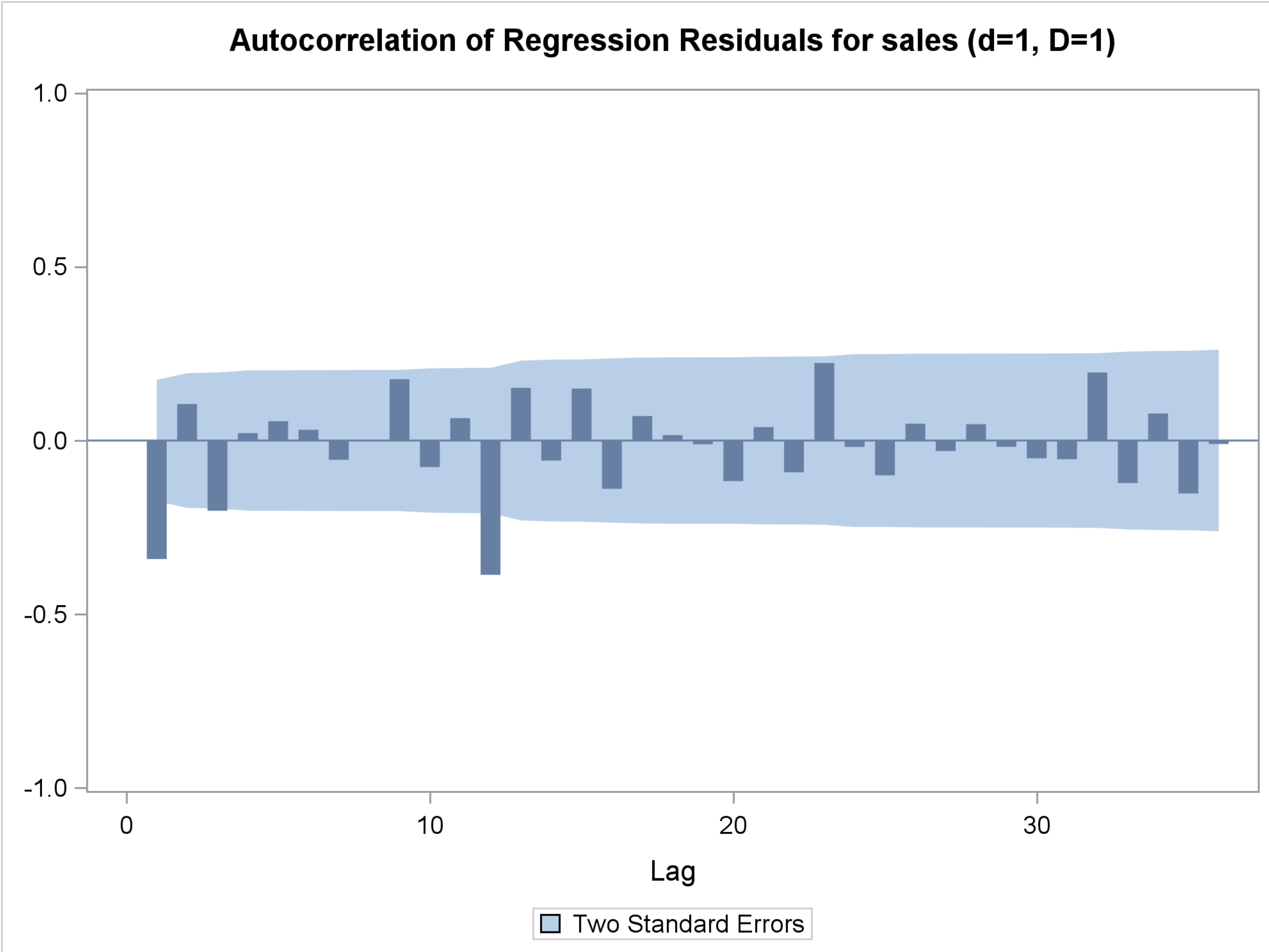
For Variable sales

Differencing: Nonseasonal Order=1 Seasonal Order=1

Lag	Correlation	Standard Error	Chi-Square	DF	Pr > ChiSq
1	-0.34112	0.08737	15.5957	1	<.0001
2	0.10505	0.09701	17.0860	2	0.0002
3	-0.20214	0.09787	22.6478	3	<.0001
4	0.02136	0.10101	22.7104	4	0.0001
5	0.05565	0.10104	23.1387	5	0.0003
6	0.03080	0.10128	23.2709	6	0.0007
7	-0.05558	0.10135	23.7050	7	0.0013
8	-0.00076	0.10158	23.7050	8	0.0026
9	0.17637	0.10158	28.1473	9	0.0009
10	-0.07636	0.10389	28.9869	10	0.0013
11	0.06438	0.10432	29.5887	11	0.0018
12	-0.38661	0.10462	51.4728	12	<.0001
13	0.15160	0.11501	54.8664	13	<.0001
14	-0.05761	0.11653	55.3605	14	<.0001
15	0.14957	0.11674	58.7204	15	<.0001
16	-0.13894	0.11820	61.6452	16	<.0001
17	0.07048	0.11944	62.4045	17	<.0001
18	0.01563	0.11975	62.4421	18	<.0001
19	-0.01061	0.11977	62.4596	19	<.0001
20	-0.11673	0.11978	64.5984	20	<.0001
21	0.03855	0.12064	64.8338	21	<.0001
22	-0.09136	0.12074	66.1681	22	<.0001
23	0.22327	0.12126	74.2099	23	<.0001
24	-0.01842	0.12436	74.2652	24	<.0001
25	-0.10029	0.12438	75.9183	25	<.0001
26	0.04857	0.12500	76.3097	26	<.0001
27	-0.03024	0.12514	76.4629	27	<.0001
28	0.04713	0.12520	76.8387	28	<.0001
29	-0.01803	0.12533	76.8943	29	<.0001
30	-0.05107	0.12535	77.3442	30	<.0001
31	-0.05377	0.12551	77.8478	31	<.0001
32	0.19573	0.12569	84.5900	32	<.0001
33	-0.12242	0.12799	87.2543	33	<.0001
34	0.07775	0.12888	88.3401	34	<.0001
35	-0.15245	0.12924	92.5584	35	<.0001
36	-0.01000	0.13061	92.5767	36	<.0001

Note: The P-values approximate the probability of observing a Chi-Square at least this large when the model fitted is correct. When DF is positive, small values of P, customarily those below 0.05, indicate model inadequacy.

Output 46.1.2 Plot for ACFs (Nonseasonal Order=1 Seasonal Order=1)



Output 46.1.3 PACFs (Nonseasonal Order=1 Seasonal Order=1)

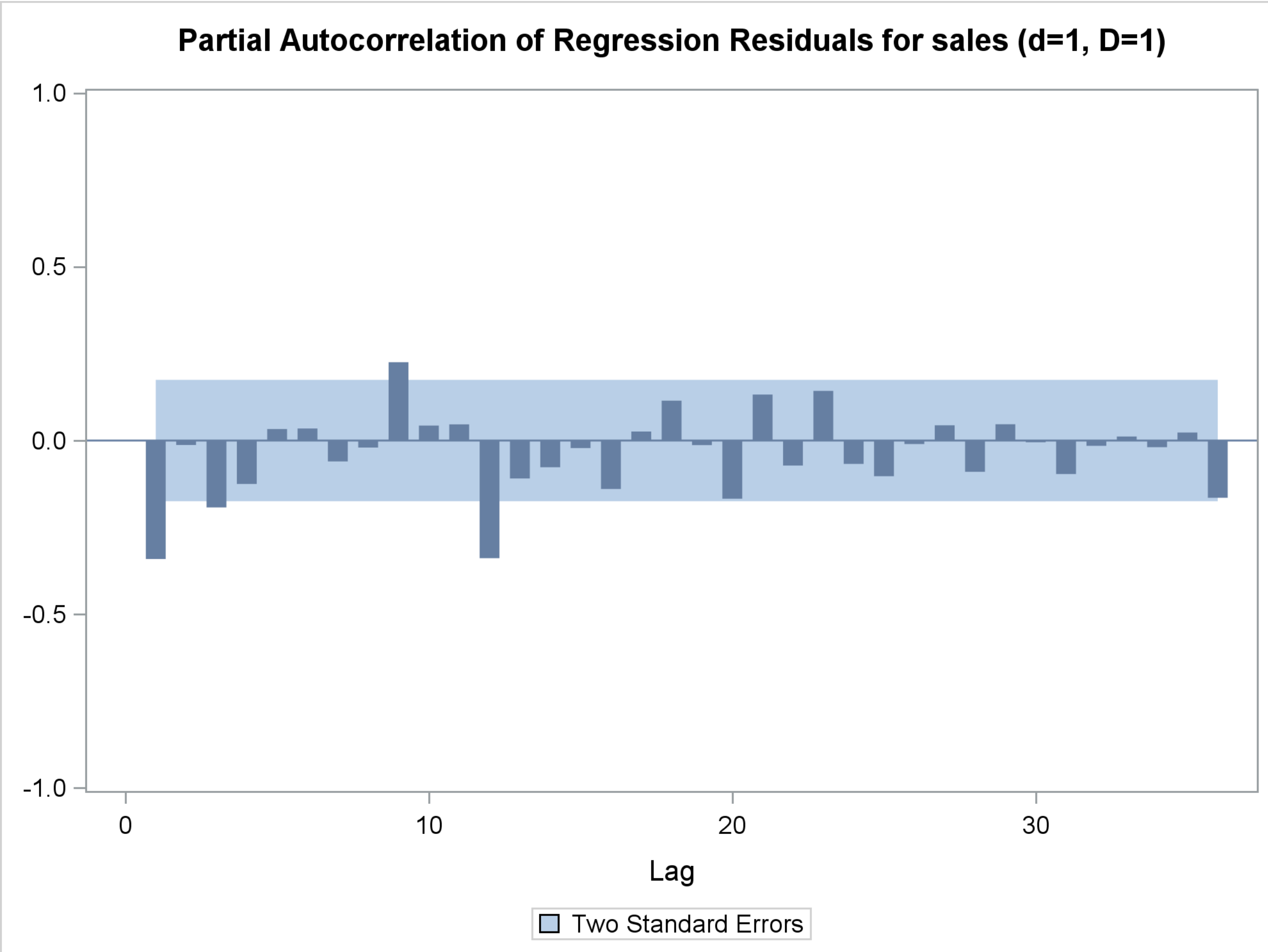
Partial Autocorrelations of Regression Residuals for ARIMA Model Identification

For Variable sales

Differencing: Nonseasonal Order=1 Seasonal Order=1

Lag	Correlation	Standard Error
1	-0.34112	0.08737
2	-0.01281	0.08737
3	-0.19266	0.08737
4	-0.12503	0.08737
5	0.03309	0.08737
6	0.03468	0.08737
7	-0.06019	0.08737
8	-0.02022	0.08737
9	0.22558	0.08737
10	0.04307	0.08737
11	0.04659	0.08737
12	-0.33869	0.08737
13	-0.10918	0.08737
14	-0.07684	0.08737
15	-0.02175	0.08737
16	-0.13955	0.08737
17	0.02589	0.08737
18	0.11482	0.08737
19	-0.01316	0.08737
20	-0.16743	0.08737
21	0.13240	0.08737
22	-0.07204	0.08737
23	0.14285	0.08737
24	-0.06733	0.08737
25	-0.10267	0.08737
26	-0.01007	0.08737
27	0.04378	0.08737
28	-0.08995	0.08737
29	0.04690	0.08737
30	-0.00490	0.08737
31	-0.09638	0.08737
32	-0.01528	0.08737
33	0.01150	0.08737
34	-0.01916	0.08737
35	0.02303	0.08737
36	-0.16488	0.08737

Output 46.1.4 Plot for PACFs (Nonseasonal Order=1 Seasonal Order=1)



Example 46.2: Model Estimation

After studying the output from Example 46.1 and identifying the ARIMA part of the model as, for example, (0 1 1)(0 1 1) 12, you can replace the IDENTIFY statement with the ARIMA and ESTIMATE statements as follows:

```
proc x13 data=sales date=date;
   var sales;
   transform power=0;
   arima model=( (0,1,1)(0,1,1) );
   estimate;
run ;
```

The parameter estimates and estimation summary statistics are shown in Output 46.2.1.

Output 46.2.1 Estimation Data

The X13 Procedure

Exact ARMA Likelihood Estimation Iteration Tolerances For Variable sales	
Maximum Total ARMA Iterations	1500
Convergence Tolerance	1.0E-05

Average absolute percentage error in within-sample forecasts: For Variable sales	
Last year:	2.81
Last-1 year:	6.38
Last-2 year:	7.69
Last three years:	5.63

Exact ARMA Likelihood Estimation Iteration Summary For Variable sales	
Number of ARMA iterations	6
Number of Function Evaluations	19

Exact ARMA Maximum Likelihood Estimation
For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	**1**	0.40181	0.07887	5.09	<.0001
Seasonal MA	**12**	0.55695	0.07626	7.30	<.0001

Estimation Summary For Variable sales	
Number of Observations	144
Number of Residuals	131
Number of Parameters Estimated	3
Variance Estimate	1.3E-03
Standard Error Estimate	3.7E-02
Standard Error of Variance	1.7E-04
Log likelihood	244.6965
Transformation Adjustment	-735.2943
Adjusted Log likelihood	-490.5978
AIC	987.1956
AICC (F-corrected-AIC)	987.3845
Hannan Quinn	990.7005
BIC	995.8211

Example 46.3: Seasonal Adjustment

Assuming that the model in Example 46.2 is satisfactory, a seasonal adjustment that uses forecast extension can be performed by adding the X11 statement to the procedure. By default, the data are forecast one year ahead at the end of the series.

```
ods output D8A#1=SalesD8A_1;
ods output D8A#2=SalesD8A_2;
ods output D8A#3=SalesD8A_3;
ods output D8A#4=SalesD8A_4;
proc x13 data=sales date=date;
   var sales;
   transform power=0;
   arima model=( (0,1,1)(0,1,1) );
   estimate;
   x11;
run;

title 'Stable Seasonality Test';
proc print data=SalesD8A_1 LABEL;
run;

title 'Nonparametric Stable Seasonality Test';
proc print data=SalesD8A_2 LABEL;
run;

title 'Moving Seasonality Test';
proc print data=SalesD8A_3 LABEL;
run;

title 'Combined Seasonality Test';
proc print data=SalesD8A_4 LABEL NOOBS;
   var _NAME_ Name1 Label1 cValue1;
run;
```

Table D8A, which contains the seasonality tests, is shown in Output 46.3.1.

Output 46.3.1 Table D8A as Displayed

The X13 Procedure

Table D 8.A: F-tests for Seasonality
For Variable sales

Test for the Presence of Seasonality Assuming Stability	Sum of Squares	DF	Mean Square	F-Value	
Between Months	23571.41	11	2142.855	190.9544	**
Residual	1481.28	132	11.22182		
Total	25052.69	143			

**** Seasonality present at the 0.1 percent level.**

Output 46.3.1 *continued*

Nonparametric Test for the Presence of Seasonality Assuming Stability		
Kruskal-Wallis Statistic	**DF**	**Probability Level**
131.9546	11	.00%

Seasonality present at the one percent level.

Moving Seasonality Test					
	Sum of Squares	**DF**	**Mean Square**	**F-Value**	
Between Years	259.2517	10	25.92517	3.370317	**
Error	846.1424	110	7.692204		

****Moving seasonality present at the one percent level.**

Summary of Results and Combined Test for the Presence of Identifiable Seasonality	
Seasonality Tests:	Probability Level
Stable Seasonality F-test	0.000
Moving Seasonality F-test	0.001
Kruskal-Wallis Chi-square Test	0.000
Combined Measures:	Value
T1 = 7/F_Stable	0.04
T2 = 3*F_Moving/F_Stable	0.05
T = (T1 + T2)/2	0.04
Combined Test of Identifiable Seasonality:	Present

The four ODS statements in the preceding example direct output from the D8A tables into four data sets: SalesD8A_1, SalesD8A_2, SalesD8A_3, and SalesD8A_4. It is best to direct the output to four different data sets because the four tables associated with Table D8A have varying formats. The ODS data sets are shown in Output 46.3.2, Output 46.3.3, Output 46.3.4, and Output 46.3.5.

Output 46.3.2 Table D8A Output in Data Set SalesD8A_1

Stable Seasonality Test

Obs	_NAME_	FT_SRC	Sum of Squares	DF	Mean Square	F-Value	FT_AST
1	sales	Between Months	23571.41	11	2142.855	190.9544	**
2	sales	Residual	1481.28	132	11.22182	.	
3	sales	Total	25052.69	143	.	.	

Output 46.3.3 Table D8A Output in Data Set SalesD8A_2

Nonparametric Stable Seasonality Test

Obs	_NAME_	Kruskal-Wallis Statistic	DF	Probability Level
1	sales	131.9546	11	.00%

Output 46.3.4 Table D8A Output in Data Set SalesD8A_3

Moving Seasonality Test

Obs	_NAME_	FT_SRC	Sum of Squares	DF	Mean Square	F-Value	FT_AST
1	sales	Between Years	259.2517	10	25.92517	3.370317	**
2	sales	Error	846.1424	110	7.692204	.	

Output 46.3.5 Table D8A Output in Data Set SalesD8A_4

Combined Seasonality Test

NAME	Name1	Label1	cValue1
sales		Seasonality Tests:	Probability Level
sales			
sales	P_STABLE	Stable Seasonality F-test	0.000
sales	P_MOV	Moving Seasonality F-test	0.001
sales	P_KW	Kruskal-Wallis Chi-square Test	0.000
sales			
sales		Combined Measures:	Value
sales			
sales	T1	T1 = 7/F_Stable	0.04
sales	T2	T2 = 3*F_Moving/F_Stable	0.05
sales	T	T = (T1 + T2)/2	0.04
sales			
sales	IDSeasTest	Combined Test of Identifiable Seasonality:	Present

Example 46.4: RegARIMA Automatic Model Selection

This example demonstrates regARIMA modeling and TRAMO-based automatic model selection, which is available with the AUTOMDL statement. ODS SELECT statements are used to limit the displayed output to the model selection and estimation stages. The same data set is used as in the previous examples.

```
title 'TRAMO Automatic Model Identification';
ods select UnitRootTestModel
           UnitRootTest
           AutoChoiceModel
           Best5Model
           AutomaticModelChoice
           InitialModelChoice
           FinalModelChecks
           FinalModelChoice
           AutomdlNote;
proc x13 data=sales date=date;
   var sales;
   transform function=log;
   regression predefined=td;
   automdl maxorder=(1,1)
           print=all;
   estimate;
   x11;
   output out=out a1 a2 a6 b1 c17 c20 d1 d7 d8 d9 d10
                  d11 d12 d13 d16 d18;
run;

proc print data=out(obs=21);
   title 'Output Variables Related to Trading Day Regression';
run;
```

The automatic model selection output is shown in Output 46.4.1, Output 46.4.2, and Output 46.4.3. The first table, “ARIMA Estimate for Unit Root Identification” in Output 46.4.1, gives details of the method that TRAMO uses to automatically select the orders of differencing. The second table, “Results of Unit Root Test for Identifying Orders of Differencing” in Output 46.4.1, shows that a regular difference order of 1 and a seasonal difference order of 1 has been determined by TRAMO. The third table, “Models Estimated by Automatic ARIMA Model Selection Procedure” in Output 46.4.2, shows all the models examined by the TRAMO-based method. The fourth table, “Best Five ARIMA Models Chosen by Automatic Modeling” in Output 46.4.3, shows the top five models in order of rank and their BIC2 statistic. The fifth table, “Comparison of Automatically Selected Model and Default Model” in Output 46.4.3, compares the model selected by the TRAMO model to the default regARIMA model of the X-13ARIMA-SEATS method. The sixth table, “Initial Automatic Model Selection” in Output 46.4.3, shows which model was selected between the two models that are compared in the table “Comparison of Automatically Selected Model and Default Model.” (When available, the table “Check of the Residual Ljung-Box Q Statistic” in Output 46.4.3 contains additional information about the initial model choice.) The seventh table, “Final Checks for Identified Model” in Output 46.4.3, displays the results of the final model checks for model adequacy. The eighth table, “Final Automatic Model Selection” in Output 46.4.3, shows which model was actually selected.

Output 46.4.1 Output from the AUTOMDL Statement

TRAMO Automatic Model Identification

The X13 Procedure

ARIMA Estimates for Unit Root Identification

For Variable sales

Model Number	Estimation Method	Estimated Model	ARMA Parameter	Estimate
1	H-R	(2, 0, 0) (1, 0, 0)	NS_AR_1	0.67540
	H-R	(2, 0, 0) (1, 0, 0)	NS_AR_2	0.28425
	H-R	(2, 0, 0) (1, 0, 0)	S_AR_12	0.91963
2	H-R	(1, 1, 1) (1, 0, 1)	NS_AR_1	0.13418
	H-R	(1, 1, 1) (1, 0, 1)	S_AR_12	0.98500
	H-R	(1, 1, 1) (1, 0, 1)	NS_MA_1	0.47884
	H-R	(1, 1, 1) (1, 0, 1)	S_MA_12	0.51726
3	H-R	(1, 1, 1) (1, 1, 1)	NS_AR_1	-0.39269
	H-R	(1, 1, 1) (1, 1, 1)	S_AR_12	0.06223
	H-R	(1, 1, 1) (1, 1, 1)	NS_MA_1	-0.09570
	H-R	(1, 1, 1) (1, 1, 1)	S_MA_12	0.58536

Results of Unit Root Test for Identifying Orders of Differencing

For Variable sales

Regular difference order	Seasonal difference order	Mean Significant
1	1	no

Output 46.4.2 Output from the AUTOMDL Statement

Models estimated by Automatic ARIMA Model Selection procedure For Variable sales					
		ARMA		Statistics of Fit	
Model Number	Estimated Model	Parameter	Estimate	BIC	BIC2
1	(3, 1, 0) (0, 1, 0)	NS_AR_1	-0.33524		
	(3, 1, 0) (0, 1, 0)	NS_AR_2	-0.05558		
	(3, 1, 0) (0, 1, 0)	NS_AR_3	-0.15649		
	(3, 1, 0) (0, 1, 0)			1024.469	-3.40549
2	(3, 1, 0) (0, 1, 1)	NS_AR_1	-0.33186		
	(3, 1, 0) (0, 1, 1)	NS_AR_2	-0.05823		
	(3, 1, 0) (0, 1, 1)	NS_AR_3	-0.15200		
	(3, 1, 0) (0, 1, 1)	S_MA_12	0.55279		
	(3, 1, 0) (0, 1, 1)			993.7880	-3.63970
3	(3, 1, 0) (1, 1, 0)	NS_AR_1	-0.38673		
	(3, 1, 0) (1, 1, 0)	NS_AR_2	-0.08768		
	(3, 1, 0) (1, 1, 0)	NS_AR_3	-0.18143		
	(3, 1, 0) (1, 1, 0)	S_AR_12	-0.47336		
	(3, 1, 0) (1, 1, 0)			1000.224	-3.59057
4	(3, 1, 0) (1, 1, 1)	NS_AR_1	-0.34352		
	(3, 1, 0) (1, 1, 1)	NS_AR_2	-0.06504		
	(3, 1, 0) (1, 1, 1)	NS_AR_3	-0.15728		
	(3, 1, 0) (1, 1, 1)	S_AR_12	-0.12163		
	(3, 1, 0) (1, 1, 1)	S_MA_12	0.47073		
	(3, 1, 0) (1, 1, 1)			998.0548	-3.60713
5	(0, 1, 0) (0, 1, 1)	S_MA_12	0.60446		
	(0, 1, 0) (0, 1, 1)			996.8560	-3.61628
6	(0, 1, 1) (0, 1, 1)	NS_MA_1	0.36272		
	(0, 1, 1) (0, 1, 1)	S_MA_12	0.55599		
	(0, 1, 1) (0, 1, 1)			986.6405	-3.69426
7	(1, 1, 0) (0, 1, 1)	NS_AR_1	-0.32734		
	(1, 1, 0) (0, 1, 1)	S_MA_12	0.55834		
	(1, 1, 0) (0, 1, 1)			987.1500	-3.69037
8	(1, 1, 1) (0, 1, 1)	NS_AR_1	0.17833		
	(1, 1, 1) (0, 1, 1)	NS_MA_1	0.52867		
	(1, 1, 1) (0, 1, 1)	S_MA_12	0.56212		
	(1, 1, 1) (0, 1, 1)			991.2363	-3.65918
9	(0, 1, 1) (0, 1, 0)	NS_MA_1	0.36005		
	(0, 1, 1) (0, 1, 0)			1017.770	-3.45663

Output 46.4.3 Output from the AUTOMDL Statement

Best Five ARIMA Models Chosen by Automatic Modeling

For Variable sales

Rank	Estimated Model	BIC2
1	(0, 1, 1) (0, 1, 1)	-3.69426
2	(1, 1, 0) (0, 1, 1)	-3.69037
3	(1, 1, 1) (0, 1, 1)	-3.65918
4	(0, 1, 0) (0, 1, 1)	-3.61628
5	(0, 1, 1) (0, 1, 0)	-3.45663

Comparison of Automatically Selected Model and Default Model

For Variable sales

		Statistics of Fit		
Source of Candidate Models	Estimated Model	Confidence Coefficient of the Ljung-Box Q Statistic	Residual Standard Error	Number of Outliers
Automatic Model Choice	(0, 1, 1) (0, 1, 1)	0.62561	0.03546	0
Airline Model (Default)	(0, 1, 1) (0, 1, 1)	0.62561	0.03546	0

Initial Automatic Model Selection

For Variable sales

Source of Model	Estimated Model
Automatic Model Choice	(0, 1, 1) (0, 1, 1)

Final Checks for Identified Model

For Variable sales

Test	Result	Model Change
Check for Unit Roots	No unit root.	No Change
Check for Nonseasonal Overdifferencing	Nonseasonal MA not within 0.001 of 1.0 - model passes.	No Change
Check for insignificant ARMA coefficients	No insignificant ARMA coefficients found.	No Change

Final Automatic Model Selection

For Variable sales

Source of Model	Orders Altered	Estimated Model
Automatic Model Choice	No	(0, 1, 1) (0, 1, 1)

Table 46.17 and Output 46.4.4 illustrate the regARIMA modeling method. Table 46.17 shows the relationship between the output variables in PROC X13 that results from a regARIMA model. Note that some of these formulas apply only to this example. Output 46.4.4 shows the values of these variables for the first 21 observations in the example.

Table 46.17 regARIMA Output Variables and Descriptions

Table	Title	Type	Formula
A1	Time series data (for the span analyzed)	Data	Input
A2	Prior-adjustment factors leap year (from trading day regression) adjustments	Factor	Calculated from regression
A6	RegARIMA trading day component leap year prior adjustments included from Table A2	Factor	Calculated from regression
B1	Original series (prior adjusted) (adjusted for regARIMA factors)	Data	B1 = A1/A6* *Because only TD specified
C17	Final weights for irregular component	Factor	Calculated using moving standard deviation
C20	Final extreme value adjustment factors	Factor	Calculated using C16 and C17
D1	Modified original data, D iteration	Data	D1 = B1/C20** D1 = C19/C20 **C19 = B1 in this example
D7	Preliminary trend cycle, D iteration	Data	Calculated using Henderson moving average
D8	Final unmodified SI ratios	Factor	D8 = B1/D7*** D8 = C19/D7 ***TD specified in regression
D9	Final replacement values for SI ratios	Factor	If C17 shows extreme values, D9 = D1/D7; D9 = . otherwise
D10	Final seasonal factors	Factor	Calculated using moving averages
D11	Final seasonally adjusted data (also adjusted for trading day)	Data	D11 = B1/D10**** D11 = C19/D10 ****B1 = C19 for this example
D12	Final trend cycle	Data	Calculated using Henderson moving average
D13	Final irregular component	Factor	D13 = D11/D12
D16	Combined adjustment factors (includes seasonal, trading day factors)	Factor	D16 = A1/D11
D18	Combined calendar adjustment factors (includes trading day factors)	Factor	D18 = D16/D10 D18 = A6***** *****Regression TD is the only calendar adjustment factor in this example

Output 46.4.4 Output Variables Related to Trading Day Regression

Output Variables Related to Trading Day Regression

Obs	DATE	sales_A1	sales_A2	sales_A6	sales_B1	sales_C17	sales_C20	sales_D1	sales_D7	sales_D8
1	SEP78	112	1.00000	1.01328	110.532	1.00000	1.00000	110.532	124.138	0.89040
2	OCT78	118	1.00000	0.99727	118.323	1.00000	1.00000	118.323	124.905	0.94731
3	NOV78	132	1.00000	0.98960	133.388	1.00000	1.00000	133.388	125.646	1.06161
4	DEC78	129	1.00000	1.00957	127.777	1.00000	1.00000	127.777	126.231	1.01225
5	JAN79	121	1.00000	0.99408	121.721	1.00000	1.00000	121.721	126.557	0.96179
6	FEB79	135	0.99115	0.99115	136.205	1.00000	1.00000	136.205	126.678	1.07521
7	MAR79	148	1.00000	1.00966	146.584	1.00000	1.00000	146.584	126.825	1.15580
8	APR79	148	1.00000	0.99279	149.075	1.00000	1.00000	149.075	127.038	1.17347
9	MAY79	136	1.00000	0.99406	136.813	1.00000	1.00000	136.813	127.433	1.07360
10	JUN79	119	1.00000	1.01328	117.440	1.00000	1.00000	117.440	127.900	0.91822
11	JUL79	104	1.00000	0.99727	104.285	1.00000	1.00000	104.285	128.499	0.81156
12	AUG79	118	1.00000	0.99678	118.381	1.00000	1.00000	118.381	129.253	0.91589
13	SEP79	115	1.00000	1.00229	114.737	0.98630	0.99964	114.778	130.160	0.88151
14	OCT79	126	1.00000	0.99408	126.751	0.88092	1.00320	126.346	131.238	0.96581
15	NOV79	141	1.00000	1.00366	140.486	1.00000	1.00000	140.486	132.699	1.05869
16	DEC79	135	1.00000	0.99872	135.173	1.00000	1.00000	135.173	134.595	1.00429
17	JAN80	125	1.00000	0.99406	125.747	0.00000	0.95084	132.248	136.820	0.91906
18	FEB80	149	1.02655	1.03400	144.100	1.00000	1.00000	144.100	139.215	1.03509
19	MAR80	170	1.00000	0.99872	170.217	1.00000	1.00000	170.217	141.559	1.20245
20	APR80	170	1.00000	0.99763	170.404	1.00000	1.00000	170.404	143.777	1.18520
21	MAY80	158	1.00000	1.00966	156.489	1.00000	1.00000	156.489	145.925	1.07239

Obs	sales_D9	sales_D10	sales_D11	sales_D12	sales_D13	sales_D16	sales_D18
1	.	0.90264	122.453	124.448	0.98398	0.91463	1.01328
2	.	0.94328	125.438	125.115	1.00258	0.94070	0.99727
3	.	1.06320	125.459	125.723	0.99790	1.05214	0.98960
4	.	0.99534	128.375	126.205	1.01720	1.00487	1.00957
5	.	0.97312	125.083	126.479	0.98896	0.96735	0.99408
6	.	1.05931	128.579	126.587	1.01574	1.04994	0.99115
7	.	1.17842	124.391	126.723	0.98160	1.18980	1.00966
8	.	1.18283	126.033	126.902	0.99315	1.17430	0.99279
9	.	1.06125	128.916	127.257	1.01303	1.05495	0.99406
10	.	0.91663	128.121	127.747	1.00293	0.92881	1.01328
11	.	0.81329	128.226	128.421	0.99848	0.81107	0.99727
12	.	0.91135	129.897	129.316	1.00449	0.90841	0.99678
13	0.88182	0.90514	126.761	130.347	0.97249	0.90722	1.00229
14	0.96273	0.93820	135.100	131.507	1.02732	0.93264	0.99408
15	.	1.06183	132.306	132.937	0.99525	1.06571	1.00366
16	.	0.99339	136.072	134.720	1.01004	0.99212	0.99872
17	0.96658	0.97481	128.996	136.763	0.94321	0.96902	0.99406
18	.	1.06153	135.748	138.996	0.97663	1.09762	1.03400
19	.	1.17965	144.295	141.221	1.02177	1.17814	0.99872
20	.	1.18499	143.802	143.397	1.00283	1.18218	0.99763
21	.	1.06005	147.624	145.591	1.01397	1.07028	1.00966

Example 46.5: Automatic Outlier Detection

This example demonstrates the use of the OUTLIER statement to automatically detect and remove outliers from a time series to be seasonally adjusted. The data set is the same as in the section "Basic Seasonal Adjustment" on page 3288 and the previous examples. Adding the OUTLIER statement to Example 46.3 requests that outliers be detected by using the default critical value as described in the section "OUTLIER Statement" on page 3317. The tables associated with outlier detection for this example are shown in Output 46.5.1. The first table shows the critical values; the second table shows that a single potential outlier was identified; the third table shows the estimates for the ARMA parameters. Since no outliers are included in the regression model, the "Regression Model Parameter Estimates" table is not displayed. Because only a potential outlier was identified, and not an actual outlier, in this case the A1 series and the B1 series are identical.

```
title 'Automatic Outlier Identification';
proc x13 data=sales date=date;
   var sales;
   transform function=log;
   arima model=( (0,1,1)(0,1,1) );
   outlier;
   estimate;
   x11;
   output out=nooutlier a1 b1 d10;
run ;
```

Output 46.5.1 PROC X13 Output When Potential Outliers Are Identified

Automatic Outlier Identification

The X13 Procedure

Critical Values to use in Outlier Detection For Variable sales	
Begin	SEP1978
End	AUG1990
Observations	144
Method	Add One
AO Critical Value	3.889838
LS Critical Value	3.889838

Note: The following time series values might later be identified as outliers when data are added or revised. They were not identified as outliers in this run either because their test t-statistics were slightly below the critical value or because they were eliminated during the backward deletion step of the identification procedure, when a non-robust t-statistic is used.

Potential Outliers For Variable sales

Type of Outlier	Date	t Value for AO	t Value for LS
AO	**NOV1989**	-3.48	-1.51

Output 46.5.1 *continued*

Exact ARMA Maximum Likelihood Estimation
For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	**1**	0.40181	0.07887	5.09	<.0001
Seasonal MA	**12**	0.55695	0.07626	7.30	<.0001

In the next example, reducing the critical value to 3.3 causes the outlier identification routine to more aggressively identify outliers as shown in Output 46.5.2. The first table shows the critical values. The second table shows that three additive outliers and a level-shift have been included in the regression model. The third table shows how the inclusion of outliers in the model affects the ARMA parameters.

```
proc x13 data=sales date=date;
   var sales;
   transform function=log;
   arima model=((0,1,1) (0,1,1));
   outlier cv=3.3;
   estimate;
   x11;
   output out=outlier a1 a8  a8ao a8ls b1 d10;
run;

proc print data=outlier(obs=45);
run;
```

Output 46.5.2 PROC X13 Output When Outliers Are Identified

Automatic Outlier Identification

The X13 Procedure

Critical Values to use in Outlier Detection For Variable sales	
Begin	SEP1978
End	AUG1990
Observations	144
Method	Add One
AO Critical Value	3.3
LS Critical Value	3.3

Regression Model Parameter Estimates

For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
Automatically Identified	AO JAN1981	Est	0.09590	0.02168	4.42	<.0001
	LS FEB1983	Est	-0.09673	0.02488	-3.89	0.0002
	AO OCT1983	Est	-0.08032	0.02146	-3.74	0.0003
	AO NOV1989	Est	-0.10323	0.02480	-4.16	<.0001

Exact ARMA Maximum Likelihood Estimation

For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	**1**	0.33205	0.08239	4.03	<.0001
Seasonal MA	**12**	0.49647	0.07676	6.47	<.0001

The first 45 observations of the A1, A8, A8AO, A8LS, B1, and D10 series are displayed in Output 46.5.3. You can confirm the following relationships from the data:

$$A8 = A8AO \times A8LS$$

$$B1 = A1/A8$$

The seasonal factors are stored in the variable sales_D10.

Output 46.5.3 PROC X13 Output Series Related to Outlier Detection

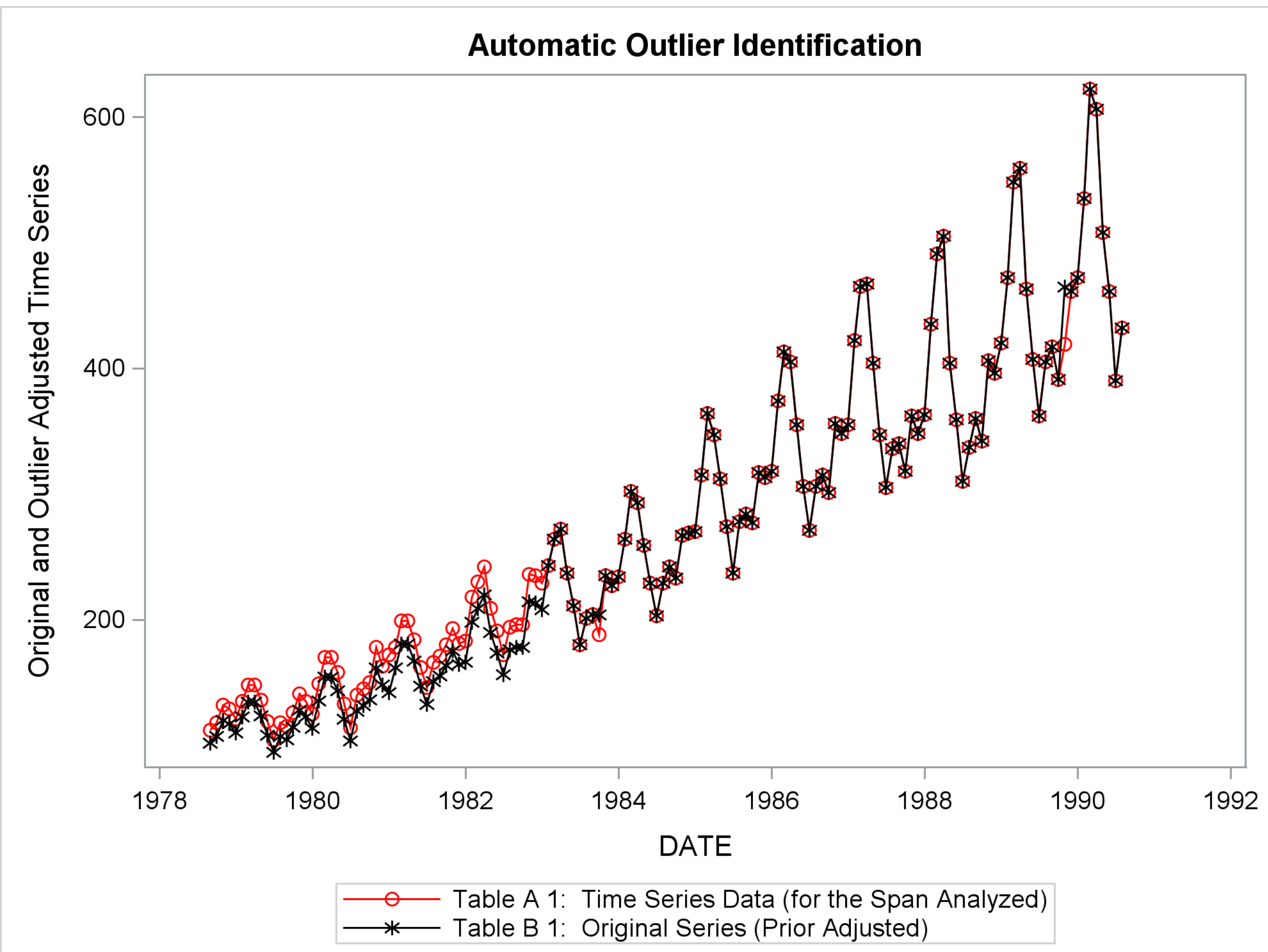
Automatic Outlier Identification

Obs	DATE	sales_A1	sales_A8	sales_A8AO	sales_A8LS	sales_B1	sales_D10
1	SEP78	112	1.10156	1.00000	1.10156	101.674	0.90496
2	OCT78	118	1.10156	1.00000	1.10156	107.121	0.94487
3	NOV78	132	1.10156	1.00000	1.10156	119.830	1.04711
4	DEC78	129	1.10156	1.00000	1.10156	117.107	1.00119
5	JAN79	121	1.10156	1.00000	1.10156	109.844	0.94833
6	FEB79	135	1.10156	1.00000	1.10156	122.553	1.06817
7	MAR79	148	1.10156	1.00000	1.10156	134.355	1.18679
8	APR79	148	1.10156	1.00000	1.10156	134.355	1.17607
9	MAY79	136	1.10156	1.00000	1.10156	123.461	1.07565
10	JUN79	119	1.10156	1.00000	1.10156	108.029	0.91844
11	JUL79	104	1.10156	1.00000	1.10156	94.412	0.81206
12	AUG79	118	1.10156	1.00000	1.10156	107.121	0.91602
13	SEP79	115	1.10156	1.00000	1.10156	104.397	0.90865
14	OCT79	126	1.10156	1.00000	1.10156	114.383	0.94131
15	NOV79	141	1.10156	1.00000	1.10156	128.000	1.04496
16	DEC79	135	1.10156	1.00000	1.10156	122.553	0.99766
17	JAN80	125	1.10156	1.00000	1.10156	113.475	0.94942
18	FEB80	149	1.10156	1.00000	1.10156	135.263	1.07172
19	MAR80	170	1.10156	1.00000	1.10156	154.327	1.18663
20	APR80	170	1.10156	1.00000	1.10156	154.327	1.18105
21	MAY80	158	1.10156	1.00000	1.10156	143.433	1.07383
22	JUN80	133	1.10156	1.00000	1.10156	120.738	0.91930
23	JUL80	114	1.10156	1.00000	1.10156	103.490	0.81385
24	AUG80	140	1.10156	1.00000	1.10156	127.093	0.91466
25	SEP80	145	1.10156	1.00000	1.10156	131.632	0.91302
26	OCT80	150	1.10156	1.00000	1.10156	136.171	0.93086
27	NOV80	178	1.10156	1.00000	1.10156	161.589	1.03965
28	DEC80	163	1.10156	1.00000	1.10156	147.972	0.99440
29	JAN81	172	1.21243	1.10065	1.10156	141.864	0.95136
30	FEB81	178	1.10156	1.00000	1.10156	161.589	1.07981
31	MAR81	199	1.10156	1.00000	1.10156	180.653	1.18661
32	APR81	199	1.10156	1.00000	1.10156	180.653	1.19097
33	MAY81	184	1.10156	1.00000	1.10156	167.036	1.06905
34	JUN81	162	1.10156	1.00000	1.10156	147.064	0.92446
35	JUL81	146	1.10156	1.00000	1.10156	132.539	0.81517
36	AUG81	166	1.10156	1.00000	1.10156	150.695	0.91148
37	SEP81	171	1.10156	1.00000	1.10156	155.234	0.91352
38	OCT81	180	1.10156	1.00000	1.10156	163.405	0.91632
39	NOV81	193	1.10156	1.00000	1.10156	175.206	1.03194
40	DEC81	181	1.10156	1.00000	1.10156	164.312	0.98879
41	JAN82	183	1.10156	1.00000	1.10156	166.128	0.95699
42	FEB82	218	1.10156	1.00000	1.10156	197.901	1.09125
43	MAR82	230	1.10156	1.00000	1.10156	208.795	1.19059
44	APR82	242	1.10156	1.00000	1.10156	219.688	1.20448
45	MAY82	209	1.10156	1.00000	1.10156	189.731	1.06355

From the two previous examples, you can examine how outlier detection affects the seasonally adjusted series. Output 46.5.4 shows a plot of A1 versus B1 in the series where outliers are detected. B1 has been adjusted for the additive outliers and the level-shift.

```
proc sgplot data=outlier;
   series x=date y=sales_A1 / name='A1' markers
                              markerattrs=(color=red symbol='circle')
                              lineattrs=(color=red);
   series x=date y=sales_B1 / name='B1' markers
                              markerattrs=(color=black symbol='asterisk')
                              lineattrs=(color=black);
   yaxis label='Original and Outlier Adjusted Time Series';
run;
```

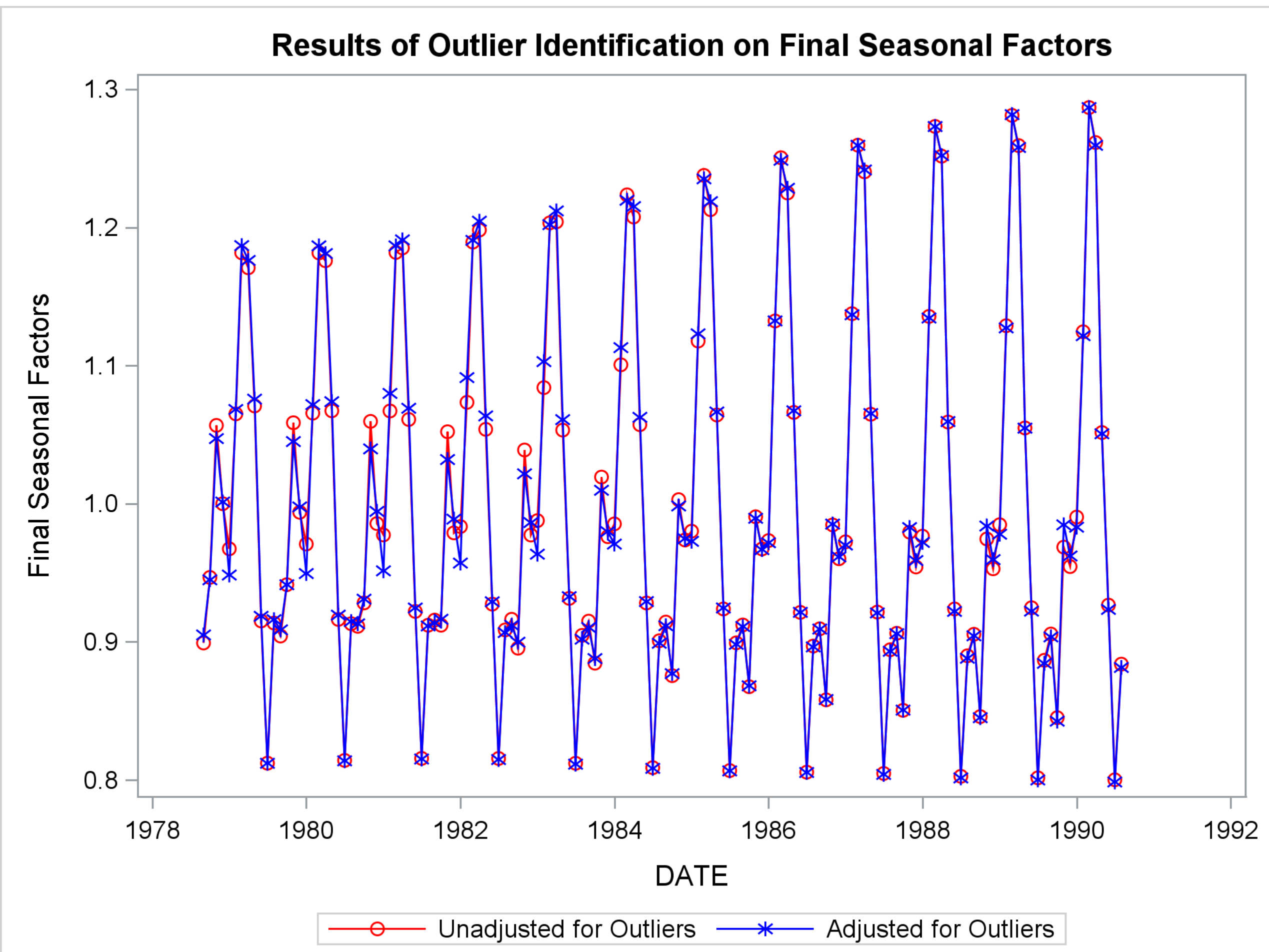
Output 46.5.4 Original Series and Outlier Adjusted Series



Output 46.5.5 compares the seasonal factors (Table D10) of the series unadjusted for outliers to the series adjusted for outliers. The seasonal factors are based on the B1 series.

```
data both;
   merge nooutlier(rename=(sales_D10=unadj_D10)) outlier;
run;

title 'Results of Outlier Identification on Final Seasonal Factors';
proc sgplot data=both;
   series x=date y=unadj_D10 / name='unadjusted' markers
                               markerattrs=(color=red symbol='circle')
                               lineattrs=(color=red)
                               legendlabel='Unadjusted for Outliers';
   series x=date y=sales_D10 / name='adjusted' markers
                               markerattrs=(color=blue symbol='asterisk')
                               lineattrs=(color=blue)
                               legendlabel='Adjusted for Outliers';
   yaxis label='Final Seasonal Factors';
run;
```

Output 46.5.5 Seasonal Factors Based on Original and Outlier Adjusted Series



Example 46.6: User-Defined Regressors

This example demonstrates the use of the USERVAR= option in the REGRESSION statement to include user-defined regressors in the regARIMA model. The user-defined regressors must be defined as nonmissing values for the span of the series being modeled plus any backcast or forecast values. Suppose you have the data set SALESDATA with 132 monthly observations beginning in January 1949.

```
title 'Data Set to be Seasonally Adjusted';
data salesdata;
   set sashelp.air(obs=132);
run;
```

Because the regARIMA model forecasts one year ahead, you must define the regressor for 144 observations that start in January 1949. You can construct a simple length-of-month regressor by using the following DATA step:

```
title 'User-defined Regressor for Data to be Seasonally Adjusted';
data regressors(keep=date LengthOfMonth);
   set sashelp.air;
   LengthOfMonth = INTNX('MONTH',date,1) - date;
run;
```

In this example, the two data sets are merged to use them as input to PROC X13. You can also use the AUXDATA= data set to input user-defined regressors. For more information, see Example 46.11. The BY statement is used to align the regressors with the time series by the time ID variable DATE.

```
title 'Data Set Containing Series and Regressors';
data datain;
   merge regressors salesdata;
   by date;
run;

proc print data=datain(firstobs=121);
run;
```

The last 24 observations of the input data set are displayed in Output 46.6.1. The regressor variable is defined for one year (12 observations) beyond the span of the time series to be seasonally adjusted.

Output 46.6.1 PROC X13 Input Data Set with User-Defined Regressor

Data Set Containing Series and Regressors

Obs	DATE	LengthOfMonth	AIR
121	JAN59	31	360
122	FEB59	28	342
123	MAR59	31	406
124	APR59	30	396
125	MAY59	31	420
126	JUN59	30	472
127	JUL59	31	548
128	AUG59	31	559
129	SEP59	30	463
130	OCT59	31	407
131	NOV59	30	362
132	DEC59	31	405
133	JAN60	31	.
134	FEB60	29	.
135	MAR60	31	.
136	APR60	30	.
137	MAY60	31	.
138	JUN60	30	.
139	JUL60	31	.
140	AUG60	31	.
141	SEP60	30	.
142	OCT60	31	.
143	NOV60	30	.
144	DEC60	31	.

The DATAIN data set is now ready to be used as input to PROC X13. The DATE= variable and the user-defined regressors are automatically excluded from the variables to be seasonally adjusted.

```
title 'regARIMA Model with User-defined Regressor';
proc x13 data=datain date=DATE interval=MONTH plots=none;
   transform function=log;
   regression uservar=LengthOfMonth / usertype=lom;
   automdl;
   x11;
   output out=out a1 d11;
run;
```

The parameter estimates for the regARIMA model are shown in Output 46.6.2

Output 46.6.2 PROC X13 Output for User-Defined Regression Parameter

regARIMA Model with User-defined Regressor

The X13 Procedure

Regression Model Parameter Estimates
For Variable AIR

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	LengthOfMonth	Est	0.04683	0.01834	2.55	0.0119

Exact ARMA Maximum Likelihood Estimation
For Variable AIR

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	**1**	0.33678	0.08506	3.96	0.0001
Seasonal MA	**12**	0.54078	0.07726	7.00	<.0001

Another way to include user-defined regressors in the regARIMA model is to specify the SPAN= option in the PROC X13 statement. The following user-defined regressor is similar to the one defined previously. However, this length-of-month regressor is mean adjusted. Using a zero-mean regressor prevents the regressor from altering the level of the series. In this instance, the series to be seasonally adjusted, AIR, and the regression variable, LengthOfMonth, have nonmissing observations at all time periods in the data set DATAIN.

```
title 'User-defined Regressor for Data to be Seasonally Adjusted, Mean Adjusted';
data datain(keep=date AIR LengthOfMonth);
   set sashelp.air;
   LengthOfMonth = INTNX('MONTH',date,1) - date - 30.4375;
run;
```

Because the default forecast period is one year ahead, the span of the series must be limited to one year before the end of the regression variable definition to forecast using the regression variable LengthOfMonth,

```
title 'regARIMA Model with Zero-Mean User-defined Regressor';
proc x13 data=datain date=DATE interval=MONTH span=(,DEC1959) plots=none;
   transform function=log;
   regression uservar=LengthOfMonth / usertype=lom;
   automdl;
   x11;
   output out=outzm a1 d11;
run;
```

The parameter estimates for the regARIMA model that are estimated using a zero-mean regressor are shown in Output 46.6.3

Output 46.6.3 PROC X13 Output for Zero-Mean User-Defined Regression Parameter

regARIMA Model with Zero-Mean User-defined Regressor

The X13 Procedure

Regression Model Parameter Estimates
For Variable AIR

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	LengthOfMonth	Est	0.04683	0.01834	2.55	0.0119

Exact ARMA Maximum Likelihood Estimation
For Variable AIR

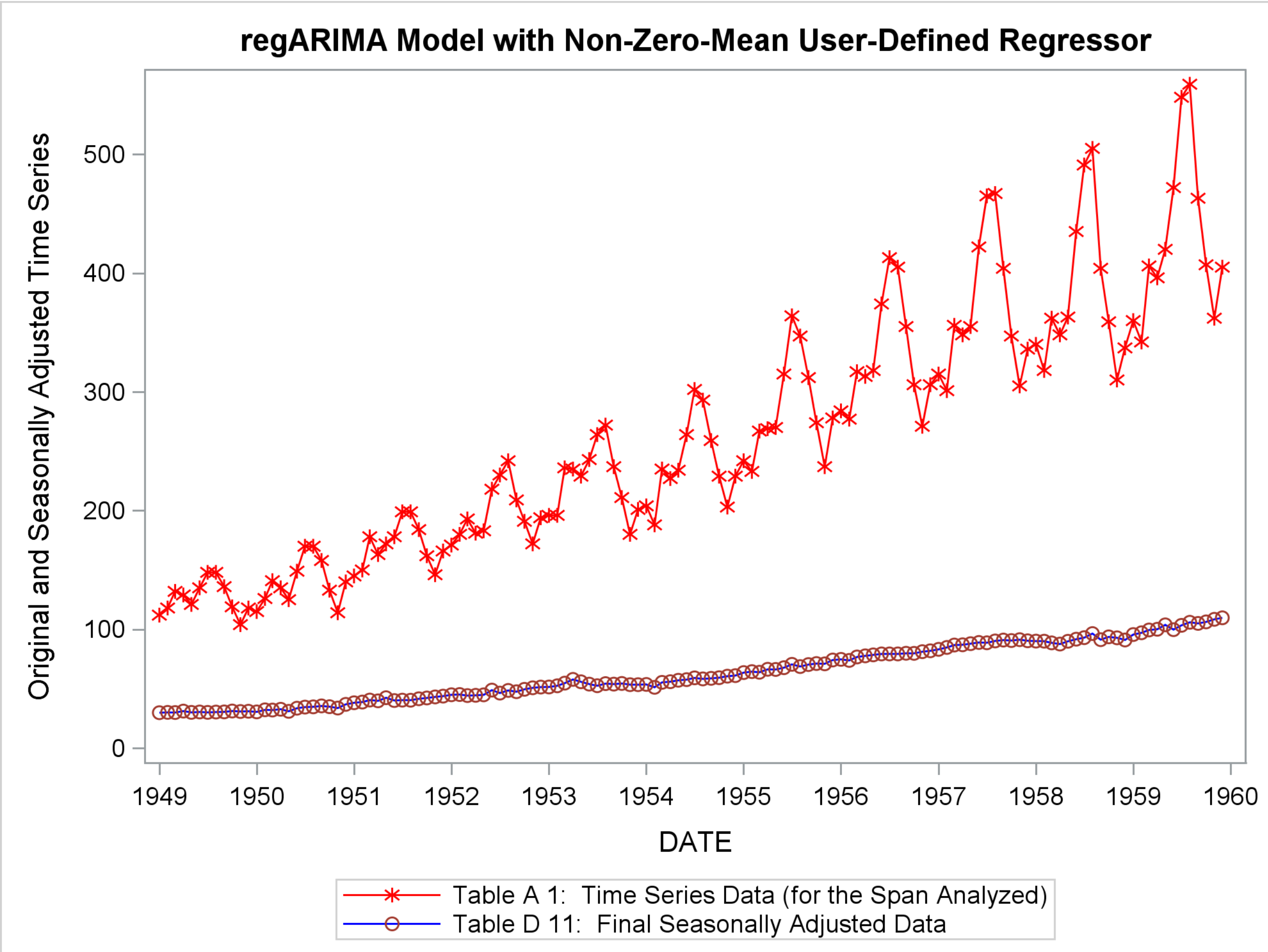
Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal MA	1	0.33678	0.08506	3.96	0.0001
Seasonal MA	12	0.54078	0.07726	7.00	<.0001

Specifying USERTYPE=LOM causes the regression effect to be removed from the seasonally adjusted series. The effect of the mean of the regression variable on the seasonally adjusted series can be seen by examining the plots of the original series and the seasonally adjusted series.

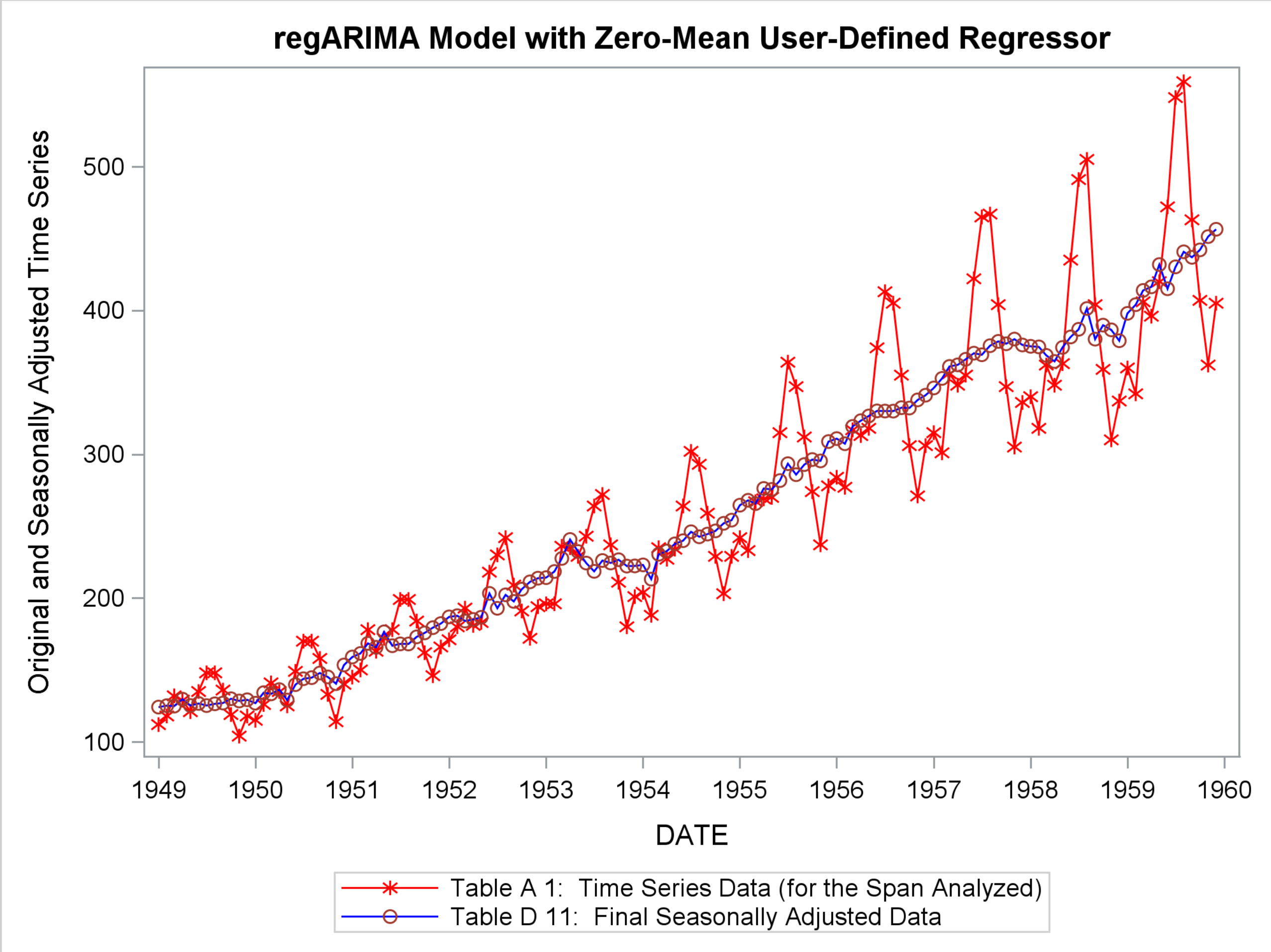
```
title 'regARIMA Model with Non-Zero-Mean User-Defined Regressor';
proc sgplot data=out;
   series x=date y=air_A1 / name = "A1" markers
                             markerattrs=(color=red symbol='asterisk')
                             lineattrs=(color=red);
   series x=date y=air_D11 / name= "D11" markers
                              markerattrs=(symbol='circle')
                              lineattrs=(color=blue);
   yaxis label='Original and Seasonally Adjusted Time Series';
run;

title 'regARIMA Model with Zero-Mean User-Defined Regressor';
proc sgplot data=outzm;
   series x=date y=air_A1 / name = "A1" markers
                             markerattrs=(color=red symbol='asterisk')
                             lineattrs=(color=red);
   series x=date y=air_D11 / name= "D11" markers
                              markerattrs=(symbol='circle')
                              lineattrs=(color=blue);
   yaxis label='Original and Seasonally Adjusted Time Series';
run;
```

The graph of the original and seasonally adjusted series in Output 46.6.4 shows that the level of the seasonally adjusted series has been altered due to the user-defined regressor. The graph of the original and seasonally adjusted series in Output 46.6.5 shows that the level of the seasonally adjusted series is the same as the original series since the user-defined regressor has zero-mean.

Output 46.6.4 Plot of Original and Seasonally Adjusted Data



Output 46.6.5 Plot of Original and Seasonally Adjusted Data (Zero-Mean Regressor)



When actual values are available for the forecast periods, information about forecast error is available in the output. Output 46.6.6 shows the table "Forecasts and Standard Errors of the Transformed Data on the Original Scale" for a series with missing values in the forecast period. Output 46.6.7 shows the table "Forecasts and Standard Errors of the Transformed Data on the Original Scale" for a series with actual values in the forecast period. Thus, it is more desirable to use SPAN= option to limit the span of a series if the actual values are available for the forecast period.

Output 46.6.6 PROC X13 Forecasts for Series Extended with Missing Values

Forecasts and Standard Errors of the Transformed Data

On the Original scale

For Variable AIR

Date	Forecast	Standard Error	95% Confidence Limits	
JAN1960	419.600	14.85053	391.509	449.705
FEB1960	416.480	19.05188	380.826	455.472
MAR1960	466.697	22.66762	424.402	513.208
APR1960	454.468	24.53242	408.951	505.051
MAY1960	473.876	27.91366	422.353	531.684
JUN1960	547.601	34.74893	483.769	619.855
JUL1960	623.318	42.20549	546.139	711.405
AUG1960	631.731	45.30824	549.231	726.623
SEP1960	527.221	39.81839	455.011	610.890
OCT1960	462.774	36.63020	396.605	539.984
NOV1960	407.155	33.64286	346.608	478.277
DEC1960	452.702	38.91914	382.913	535.212

Output 46.6.7 PROC X13 Forecasts for Series with Actual Values in Forecast Periods

Forecasts and Standard Errors of the Transformed Data

On the Original scale

For Variable AIR

Date	Data	Forecast	Forecast Error	Standard Error	t Value	95% Confidence Limits	
JAN1960	417.000	419.600	-2.600	14.85053	-0.18	391.509	449.705
FEB1960	391.000	416.480	-25.480	19.05188	-1.34	380.826	455.472
MAR1960	419.000	466.697	-47.697	22.66762	-2.10	424.402	513.208
APR1960	461.000	454.468	6.532	24.53242	0.27	408.951	505.051
MAY1960	472.000	473.876	-1.876	27.91366	-0.07	422.353	531.684
JUN1960	535.000	547.601	-12.601	34.74893	-0.36	483.769	619.855
JUL1960	622.000	623.318	-1.318	42.20549	-0.03	546.139	711.405
AUG1960	606.000	631.731	-25.731	45.30824	-0.57	549.231	726.623
SEP1960	508.000	527.221	-19.221	39.81839	-0.48	455.011	610.890
OCT1960	461.000	462.774	-1.774	36.63020	-0.05	396.605	539.984
NOV1960	390.000	407.155	-17.155	33.64286	-0.51	346.608	478.277
DEC1960	432.000	452.702	-20.702	38.91914	-0.53	382.913	535.212

Example 46.7: MDLINFOIN= and MDLINFOOUT= Data Sets

This example illustrates the use of MDLINFOIN= and MDLINFOOUT= data sets. Using the data set shown, PROC X13 step identifies the model with outliers as displayed in Output 46.7.1. Output 46.7.2 shows the data set that represents the chosen model.

```
data b1;
   input y @@;
   datalines;
   112 118 132 129
   121 135 148 148
   136 119 104 118
   115 126 141 135
   125 149 270 170
   158 133 114 140
;

title 'Model Identification Output to MDLINFOOUT= Data Set';
proc x13 data=b1 start='1980q1' interval=qtr MdlInfoOut=mdl;
   automdl;
   outlier;
run ;

proc print data=mdl;
run;
```

Output 46.7.1 Displayed Model Identification with Outliers

Model Identification Output to MDLINFOOUT= Data Set

The X13 Procedure

Critical Values to use in Outlier Detection

For Variable y

Begin	1980Q1
End	1985Q4
Observations	24
Method	Add One
AO Critical Value	3.419415
LS Critical Value	3.419415

Final Automatic Model Selection

For Variable y

Source of Model	Orders Altered	Estimated Model
Automatic Model Choice	No	(2, 1, 0) (0, 0, 0)

Regression Model Parameter Estimates

For Variable y

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
Automatically Identified	AO 1984Q3	Est	102.36589	5.96584	17.16	<.0001

Output 46.7.1 *continued*

Exact ARMA Maximum Likelihood Estimation
For Variable y

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal AR	1	0.40892	0.20213	2.02	0.0554
	2	-0.53710	0.20975	-2.56	0.0178

Output 46.7.2 PROC X13 MDLINFOOUT= Data Set Model with Outlier Detection

Model Identification Output to MDLINFOOUT= Data Set

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	y	REG	EVENT	SCALE	AO	AO01JUL1984D	.	.
2	y	ARIMA	FORECAST	NONSEASONAL	DIF	y	.	.
3	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1
4	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	102.366	5.96584	17.1587	0.000000		.	
2	1	.	.	.	.	.	.		.	
3	1	.	0	0.409	0.20213	2.0231	0.055385		.	
4	2	.	0	-0.537	0.20975	-2.5606	0.017830		.	

Suppose that after examining the output from the preceding example, you decide that an Easter regressor should be added to the model. The following statements create a data set with the model identified above and adds a US Census Bureau Predefined Easter(25) regressor. The new model data set to be used as input in the MDLINFOIN= option is displayed in the data set shown in Output 46.7.3.

```
data pluseaster;
   _NAME_ = 'y';
   _MODELTYPE_ = 'REG';
   _MODELPART_ = 'PREDEFINED';
   _COMPONENT_ = 'SCALE';
   _PARMTYPE_ = 'EASTER';
   _DSVAR_ = 'EASTER';
   _VALUE_ = 25;
run;

data mdlpluseaster;
   set mdl;
run;

proc append base=mdlpluseaster data=pluseaster force;
run;

proc print data=mdlpluseaster;
run;
```

Output 46.7.3 MDLINFOIN= Data Set Model with Easter(25) Regression Added

Model Identification Output to MDLINFOOUT= Data Set

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	y	REG	EVENT	SCALE	AO	AO01JUL1984D	.	.
2	y	ARIMA	FORECAST	NONSEASONAL	DIF	y	.	.
3	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1
4	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1
5	y	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	102.366	5.96584	17.1587	0.000000	.		
2	1	.	.	.	.	.	.	.		
3	1	.	0	0.409	0.20213	2.0231	0.055385	.		
4	2	.	0	-0.537	0.20975	-2.5606	0.017830	.		
5	.	.	.	.	.	.	.	.		

The following statements estimate the regression and ARIMA parameters by using the model described in the new data set mdlpluseaster. The results of estimating the new model are shown in Output 46.7.4.

```
proc x13 data=b1 start='1980q1' interval=qtr
   MdlInfoIn=mdlpluseaster MdlInfoOut=mdl2;
   estimate;
run;
```

Output 46.7.4 Estimate Model with Added Easter(25) Regression

Model Identification Output to MDLINFOOUT= Data Set

The X13 Procedure

Regression Model Parameter Estimates
For Variable y

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
Easter	Easter[25]	Est	6.73250	4.73335	1.42	0.1696
User Defined	AO01JUL1984D	Est	105.83795	6.12689	17.27	<.0001

Exact ARMA Maximum Likelihood Estimation
For Variable y

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal AR	**1**	0.45233	0.20676	2.19	0.0401
	2	-0.54855	0.21583	-2.54	0.0190

The new model estimation results are displayed in the data set mdl2 shown in Output 46.7.5.

```
proc print data=mdl2;
run;
```

Output 46.7.5 MDLINFOOUT= Data Set, Estimation of Model with Easter(25) Regression Added

Model Identification Output to MDLINFOOUT= Data Set

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	y	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.
2	y	REG	EVENT	SCALE	AO	AO01JUL1984D	.	.
3	y	ARIMA	FORECAST	NONSEASONAL	DIF	y	.	.
4	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1
5	y	ARIMA	FORECAST	NONSEASONAL	AR	y	.	1

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	6.733	4.73335	1.4224	0.16961		.	
2	.	.	0	105.838	6.12689	17.2743	0.00000		.	
3	1	.	.	.	.	.	.		.	
4	1	.	0	0.452	0.20676	2.1877	0.04014		.	
5	2	.	0	-0.549	0.21583	-2.5415	0.01899		.	

Example 46.8: Setting Regression Parameters

This example illustrates the use of fixed regression parameters in PROC X13. Suppose that you have the same data set as in the section "Basic Seasonal Adjustment" on page 3288. You can specify the following statements to use TRAMO to automatically identify a model that includes a US Census Bureau Easter(25) regressor:

```
title 'Estimate Easter(25) Parameter';
proc x13 data=sales date=date MdlInfoOut=mdlout1;
   var sales;
   regression predefined=easter(25);
   automdl;
run ;
```

The displayed results are shown in Output 46.8.1.

Output 46.8.1 Automatic Model ID with Easter(25) Regression

Estimate Easter(25) Parameter

The X13 Procedure

Regression Model Parameter Estimates
For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
Easter	Easter[25]	Est	-5.09298	3.50786	-1.45	0.1489

Output 46.8.1 *continued*

Exact ARMA Maximum Likelihood Estimation
For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal AR	**1**	0.62148	0.09279	6.70	<.0001
	2	0.23354	0.10385	2.25	0.0262
	3	-0.07191	0.09055	-0.79	0.4285
Nonseasonal MA	**1**	0.97377	0.03771	25.82	<.0001
Seasonal MA	**12**	0.10558	0.10205	1.03	0.3028

The MDLINFOOUT= data set, mdlout1, that contains the model and parameter estimates is shown in Output 46.8.2.

```
proc print data=mdlout1;
run;
```

Output 46.8.2 MDLINFOOUT= Data Set, Estimation of Automatic Model ID with Easter(25) Regression

Estimate Easter(25) Parameter

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.
2	sales	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
3	sales	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
4	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
5	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
6	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
7	sales	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
8	sales	ARIMA	FORECAST	SEASONAL	MA	sales	.	2

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	-5.09298	3.50786	-1.4519	0.14894		.	
2	1	.	.	.	.	.	.		.	
3	1	.	.	.	.	.	.		.	
4	1	.	0	0.62148	0.09279	6.6980	0.00000		.	
5	2	.	0	0.23354	0.10385	2.2488	0.02621		.	
6	3	.	0	-0.07191	0.09055	-0.7942	0.42851		.	
7	1	.	0	0.97377	0.03771	25.8240	0.00000		.	
8	1	.	0	0.10558	0.10205	1.0346	0.30277		.	

To fix the Easter(25) parameter while adding a regressor that is weighted according to the number of Saturdays in a month, either use the REGRESSION and EVENT statements or create a MDLINFOIN= data set. The following statements show the method for using the REGRESSION statement to fix the EASTER parameter and the EVENT statement to add the SATURDAY regressor. The output is shown in Output 46.8.3.

```
title 'Use SAS Statements to Alter Model';
proc x13 data=sales date=date MdlInfoOut=mdlout2grm;
   var sales;
   regression predefined=easter(25) / b=-5.029298 F;
   event Saturday;
   automdl;
run ;
```

Output 46.8.3 Automatic Model ID with Fixed Easter(25) and Saturday Regression

Use SAS Statements to Alter Model

The X13 Procedure

Regression Model Parameter Estimates
For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	Saturday	Est	3.23225	1.16701	2.77	0.0064
Easter	Easter[25]	Fixed	-5.02930	.	.	.

Exact ARMA Maximum Likelihood Estimation
For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal AR	**1**	-0.32506	0.08256	-3.94	0.0001

To fix the EASTER regressor and add the new SATURDAY regressor by using a DATA step, you can create the data set mdlin2 as shown. The data set mdlin2 is displayed in Output 46.8.4.

```
title 'Use a SAS DATA Step to Create a MdlInfoIn= Data Set';
data plusSaturday;
   _NAME_ = 'sales';
   _MODELTYPE_ = 'REG';
   _MODELPART_ = 'EVENT';
   _COMPONENT_ = 'SCALE';
   _PARMTYPE_ = 'USER';
   _DSVAR_ = 'SATURDAY';
run;

data mdlin2;
   set mdlout1;
   if ( _DSVAR_ = 'EASTER' ) then do;
       _NOEST_ = 1;
       _EST_ = -5.029298;
       end;
run;

proc append base=mdlin2 data=plusSaturday force;
run;

proc print data=mdlin2;
run;
```

Output 46.8.4 MDLINFOIN= Data Set, Fixed Easter(25) and Added Saturday Regression, Previously Identified Model

Use a SAS DATA Step to Create a MdlInfoIn= Data Set

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.
2	sales	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
3	sales	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
4	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
5	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
6	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
7	sales	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
8	sales	ARIMA	FORECAST	SEASONAL	MA	sales	.	2
9	sales	REG	EVENT	SCALE	USER	SATURDAY	.	.

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	1	-5.02930	3.50786	-1.4519	0.14894		.	
2	1	.	.	.	.	.	.		.	
3	1	.	.	.	.	.	.		.	
4	1	.	0	0.62148	0.09279	6.6980	0.00000		.	
5	2	.	0	0.23354	0.10385	2.2488	0.02621		.	
6	3	.	0	-0.07191	0.09055	-0.7942	0.42851		.	
7	1	.	0	0.97377	0.03771	25.8240	0.00000		.	
8	1	.	0	0.10558	0.10205	1.0346	0.30277		.	
9	.	.	.	.	.	.	.		.	

The data set mdlin2 can be used to replace the regression and model information contained in the REGRESSION, EVENT, and AUTOMDL statements. Note that the model specified in the mdlin2 data set is the same model as the automatically identified model. The following example uses the mdlin2 data set as input; the results are displayed in Output 46.8.5:

```
title 'Use Updated Data Set to Alter Model';
proc x13 data=sales date=date MdlInfoIn=mdlin2 MdlInfoOut=mdlout2DS;
   var sales;
   estimate;
run ;
```

Output 46.8.5 Estimate MDLINFOIN= File for Model with Fixed Easter(25) and Saturday Regression, Previously Identified Model

Use Updated Data Set to Alter Model

The X13 Procedure

Regression Model Parameter Estimates
For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	SATURDAY	Est	3.41762	1.07641	3.18	0.0019
Easter	Easter[25]	Fixed	-5.02930	.	.	.

Exact ARMA Maximum Likelihood Estimation
For Variable sales

Parameter	Lag	Estimate	Standard Error	t Value	Pr > |t|
Nonseasonal AR	**1**	0.62225	0.09175	6.78	<.0001
	2	0.30429	0.10109	3.01	0.0031
	3	-0.14862	0.08859	-1.68	0.0958
Nonseasonal MA	**1**	0.97125	0.03798	25.57	<.0001
Seasonal MA	**12**	0.11691	0.10000	1.17	0.2445

The following statements specify almost the same information as contained in the data set mdlin2. The ARIMA statement specifies the lags of the model. However, the initial AR and MA parameter values are the default. When using the mdlin2 data set as input, the initial values can be specified. The results are displayed in Output 46.8.6.

```
title 'Use SAS Statements to Alter Model';
proc x13 data=sales date=date MdlInfoOut=mdlout3grm;
   var sales;
   regression predefined=easter(25) / b=-5.029298 F;
   event Saturday;
   arima model=((3 1 1)(0 1 1));
   estimate;
run ;

proc print data=mdlout3grm;
run;
```

Output 46.8.6 MDLINFOOUT= Statement, Fixed Easter(25) and Added Saturday Regression, Previously Identified Model

Use SAS Statements to Alter Model

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	REG	EVENT	SCALE	USER	Saturday	.	.
2	sales	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.
3	sales	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
4	sales	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
5	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
6	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
7	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
8	sales	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
9	sales	ARIMA	FORECAST	SEASONAL	MA	sales	.	2

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	3.41760	1.07640	3.1750	0.00187		.	
2	.	.	1	-5.02930	.	.	.		.	
3	1	.	.	.	.	.	.		.	
4	1	.	.	.	.	.	.		.	
5	1	.	0	0.62228	0.09175	6.7825	0.00000		.	
6	2	.	0	0.30431	0.10109	3.0103	0.00314		.	
7	3	.	0	-0.14864	0.08859	-1.6779	0.09578		.	
8	1	.	0	0.97128	0.03796	25.5882	0.00000		.	
9	1	.	0	0.11684	0.10000	1.1684	0.24481		.	

The MDLINFOOUT= data set provides a method for comparing the results of the model identification. The data set mdlout3grm that results from using the MODEL= option in the ARIMA statement can be compared to the data set mdlout2DS that results from using the MDLINFOIN= data set with initial values for the AR and MA parameters. The mdlout2DS data set is shown in Output 46.8.7, and the results of the comparison are shown in Output 46.8.8. The slight difference in the estimated parameters can be attributed to the difference in the initial values for the AR and MA parameters.

```
proc print data=mdlout2DS;
run;
```

Output 46.8.7 MDLINFOOUT= Data Set, Fixed Easter(25) and Added Saturday Regression, Previously Identified Model

Use SAS Statements to Alter Model

Obs	_NAME_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	REG	EVENT	SCALE	USER	SATURDAY	.	.
2	sales	REG	PREDEFINED	SCALE	EASTER	EASTER	25	.
3	sales	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
4	sales	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
5	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
6	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
7	sales	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
8	sales	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
9	sales	ARIMA	FORECAST	SEASONAL	MA	sales	.	2

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	0	3.41762	1.07641	3.1750	0.00187		.	
2	.	.	1	-5.02930	.	.	.		.	
3	1	.	.	.	.	.	.		.	
4	1	.	.	.	.	.	.		.	
5	1	.	0	0.62225	0.09175	6.7817	0.00000		.	
6	2	.	0	0.30429	0.10109	3.0100	0.00314		.	
7	3	.	0	-0.14862	0.08859	-1.6776	0.09584		.	
8	1	.	0	0.97125	0.03798	25.5712	0.00000		.	
9	1	.	0	0.11691	0.10000	1.1691	0.24451		.	

```
title 'Compare Results of SAS Statement Input and MdlInfoIn= Input';
proc compare base= mdlout3grm compare=mdlout2DS;
var _EST_;
run ;
```

Output 46.8.8 Compare Parameter Estimates from Different MDLINFOOUT= Data Sets

Compare Results of SAS Statement Input and MdlInfoIn= Input

Value Comparison Results for Variables

Obs	Value of Parameter Estimate Base _EST_	Compare _EST_	Diff.	% Diff
1	3.4176	3.4176	0.0000227	0.000665
5	0.6223	0.6222	-0.000033	-0.005259
6	0.3043	0.3043	-0.000021	-0.006983
7	-0.1486	-0.1486	0.0000236	-0.0159
8	0.9713	0.9713	-0.000024	-0.002459
9	0.1168	0.1169	0.0000763	0.0653

Example 46.9: Creating an MDLINFO= Data Set for Use with the PICKMDL Statement

This example illustrates how you can create a data set for use in the PICKMDL statement that contains five commonly used ARIMA models:

- ARIMA (0 1 1)(0 1 1)s
- ARIMA (0 1 2)(0 1 1)s
- ARIMA (2 1 0)(0 1 1)s
- ARIMA (0 2 2)(0 1 1)s
- ARIMA (2 1 2)(0 1 1)s

The following macro code creates a MDLINFOIN= data set for a general ARIMA model:

```
%macro makemodel(name,p,d,q,sp,sd,sq,model);
   data "&name" (keep= _MODELTYPE_ _MODELPART_ _COMPONENT_
                       _DSVAR_ _PARMTYPE_ _FACTOR_ _LAG_
                         _LABEL_ );
      length _MODELTYPE_ _MODELPART_ _COMPONENT_ _DSVAR_
             _PARMTYPE_ $32;
      length _FACTOR_ _LAG_ 8;
      length _LABEL_ $32;

      _MODELTYPE_="ARIMA";
      _MODELPART_="FORECAST";
      _DSVAR_=".";

      _LABEL_="("||"&p"||" "||"&d"||" "||"&q"||")("||
              "&sp"||" "||"&sd"||" "||"&sq"||")s";

      /* nonseasonal AR factors */
      _COMPONENT_="NONSEASONAL";
      _PARMTYPE_="AR";
      _FACTOR_=1;
      do _LAG_=1 to &p;
         output;
         end;

      /* seasonal AR factors */
      _COMPONENT_="SEASONAL";
      _PARMTYPE_="AR";
      _FACTOR_=2;
      do _LAG_=1 to &sp;
         output;
         end;
```

```
      /* nonseasonal MA factors */
      _COMPONENT_="NONSEASONAL";
      _PARMTYPE_="MA";
      _FACTOR_=1;
      do _LAG_=1 to &q;
         output;
         end;

      /* seasonal MA factors */
      _COMPONENT_="SEASONAL";
      _PARMTYPE_="MA";
      _FACTOR_=2;
      do _LAG_=1 to &sq;
         output;
         end;

      /* nonseasonal DIF */
      _COMPONENT_="NONSEASONAL";
      _PARMTYPE_="DIF";
      _FACTOR_=1;
      _LAG_=1;
      do i_=1 to &d;
         output;
         end;

      /* seasonal DIF */
      _COMPONENT_="SEASONAL";
      _PARMTYPE_="DIF";
      _FACTOR_=2;
      _LAG_=1;
      do i_=1 to &sd;
         output;
         end;

   run;
   data sasuser.&name;
     length _MODEL_ $32;
     set "&name";
     _MODEL_ = "&model";
   run;

%mend makemodel;
```

The following SAS statements use the macro to generate a data set with some commonly used models for use in the PICKMDL statement:

```
%makemodel(x13mdl1,0,1,1,0,1,1,Model1);
%makemodel(x13mdl2,0,1,2,0,1,1,Model2);
%makemodel(x13mdl3,2,1,0,0,1,1,Model3);
%makemodel(x13mdl4,0,2,2,0,1,1,Model4);
%makemodel(x13mdl5,2,1,2,0,1,1,Model5);

data Models;
   length _NAME_ $32;
   set sasuser.x13mdl1 sasuser.x13mdl2 sasuser.x13mdl3
       sasuser.x13mdl4 sasuser.x13mdl5;
   _NAME_ = 'sales';
run;
```

The Models data set is shown in Output 46.9.1.

```
title '5 Commonly Used Models';
proc print data=Models;
run ;
```

Output 46.9.1 A Data Set That Contains Models for Use with the PICKMDL Statement

5 Commonly Used Models

Obs	_NAME_	_MODEL_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_DSVAR_
1	sales	Model1	ARIMA	FORECAST	NONSEASONAL	.
2	sales	Model1	ARIMA	FORECAST	SEASONAL	.
3	sales	Model1	ARIMA	FORECAST	NONSEASONAL	.
4	sales	Model1	ARIMA	FORECAST	SEASONAL	.
5	sales	Model2	ARIMA	FORECAST	NONSEASONAL	.
6	sales	Model2	ARIMA	FORECAST	NONSEASONAL	.
7	sales	Model2	ARIMA	FORECAST	SEASONAL	.
8	sales	Model2	ARIMA	FORECAST	NONSEASONAL	.
9	sales	Model2	ARIMA	FORECAST	SEASONAL	.
10	sales	Model3	ARIMA	FORECAST	NONSEASONAL	.
11	sales	Model3	ARIMA	FORECAST	NONSEASONAL	.
12	sales	Model3	ARIMA	FORECAST	SEASONAL	.
13	sales	Model3	ARIMA	FORECAST	NONSEASONAL	.
14	sales	Model3	ARIMA	FORECAST	SEASONAL	.
15	sales	Model4	ARIMA	FORECAST	NONSEASONAL	.
16	sales	Model4	ARIMA	FORECAST	NONSEASONAL	.
17	sales	Model4	ARIMA	FORECAST	SEASONAL	.
18	sales	Model4	ARIMA	FORECAST	NONSEASONAL	.
19	sales	Model4	ARIMA	FORECAST	NONSEASONAL	.
20	sales	Model4	ARIMA	FORECAST	SEASONAL	.
21	sales	Model5	ARIMA	FORECAST	NONSEASONAL	.
22	sales	Model5	ARIMA	FORECAST	NONSEASONAL	.

Obs	_PARMTYPE_	_FACTOR_	_LAG_	_LABEL_
1	MA	1	1	(0 1 1)(0 1 1)s
2	MA	2	1	(0 1 1)(0 1 1)s
3	DIF	1	1	(0 1 1)(0 1 1)s
4	DIF	2	1	(0 1 1)(0 1 1)s
5	MA	1	1	(0 1 2)(0 1 1)s
6	MA	1	2	(0 1 2)(0 1 1)s
7	MA	2	1	(0 1 2)(0 1 1)s
8	DIF	1	1	(0 1 2)(0 1 1)s
9	DIF	2	1	(0 1 2)(0 1 1)s
10	AR	1	1	(2 1 0)(0 1 1)s
11	AR	1	2	(2 1 0)(0 1 1)s
12	MA	2	1	(2 1 0)(0 1 1)s
13	DIF	1	1	(2 1 0)(0 1 1)s
14	DIF	2	1	(2 1 0)(0 1 1)s
15	MA	1	1	(0 2 2)(0 1 1)s
16	MA	1	2	(0 2 2)(0 1 1)s
17	MA	2	1	(0 2 2)(0 1 1)s
18	DIF	1	1	(0 2 2)(0 1 1)s
19	DIF	1	1	(0 2 2)(0 1 1)s
20	DIF	2	1	(0 2 2)(0 1 1)s
21	AR	1	1	(2 1 2)(0 1 1)s
22	AR	1	2	(2 1 2)(0 1 1)s

Output 46.9.1 *continued*

5 Commonly Used Models

Obs	_NAME_	_MODEL_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_DSVAR_
23	sales	Model5	ARIMA	FORECAST	NONSEASONAL	.
24	sales	Model5	ARIMA	FORECAST	NONSEASONAL	.
25	sales	Model5	ARIMA	FORECAST	SEASONAL	.
26	sales	Model5	ARIMA	FORECAST	NONSEASONAL	.
27	sales	Model5	ARIMA	FORECAST	SEASONAL	.

Obs	_PARMTYPE_	_FACTOR_	_LAG_	_LABEL_
23	MA	1	1	(2 1 2)(0 1 1)s
24	MA	1	2	(2 1 2)(0 1 1)s
25	MA	2	1	(2 1 2)(0 1 1)s
26	DIF	1	1	(2 1 2)(0 1 1)s
27	DIF	2	1	(2 1 2)(0 1 1)s

The following statements request that the PICKMDL method be used to choose a model from the list of models that are defined in the Models data set. The default METHOD=FIRST option chooses the first acceptable model. The chosen model is shown in the mdlchosen data set in Output 46.9.2.

```
proc x13 data=sales date=date mdlinfoin=Models mdlinfoout=mdlchosen;
   var sales;
   transform function=log;
   pickmdl method=first;
run;

title 'Chosen Model';
proc print data=mdlchosen;
run ;
```

Output 46.9.2 The Model Chosen from the Five Commonly Used Models

Chosen Model

Obs	_NAME_	_MODEL_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	MODEL1	ARIMA	FORECAST	TRANSFORM	LOG	sales	.	.
2	sales	MODEL1	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
3	sales	MODEL1	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
4	sales	MODEL1	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
5	sales	MODEL1	ARIMA	FORECAST	SEASONAL	MA	sales	.	2

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	.	.	.	.	.		.	(0 1 1)(0 1 1)s
2	1	.	.	.	.	.	.		.	(0 1 1)(0 1 1)s
3	1	.	.	.	.	.	.		.	(0 1 1)(0 1 1)s
4	1	.	0	0.40181	0.078870	5.09458	.000001192		.	(0 1 1)(0 1 1)s
5	1	.	0	0.55695	0.076255	7.30369	2.4359E-11		.	(0 1 1)(0 1 1)s

The following statements reverse the order of the models in the input data set. The default METHOD=FIRST option is used to select the model. The chosen model is shown in the mdlchosen data set in Output 46.9.3. With METHOD=FIRST, a different model is chosen because the order is changed.

```
data Models;
   length _NAME_ $32;
   set sasuser.x13mdl5 sasuser.x13mdl4 sasuser.x13mdl3
       sasuser.x13mdl2 sasuser.x13mdl1 ;
   _NAME_ = 'sales';
run;

proc x13 data=sales date=date mdlinfoin=Models mdlinfoout=mdlchosen;
   var sales;
   transform function=log;
   pickmdl method=first;
run;

title 'Chosen Model';
proc print data=mdlchosen;
run ;
```

Output 46.9.3 The Model Chosen from the Five Commonly Used Models, Reversed Order

Chosen Model

Obs	_NAME_	_MODEL_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	MODEL3	ARIMA	FORECAST	TRANSFORM	LOG	sales	.	.
2	sales	MODEL3	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
3	sales	MODEL3	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
4	sales	MODEL3	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
5	sales	MODEL3	ARIMA	FORECAST	NONSEASONAL	AR	sales	.	1
6	sales	MODEL3	ARIMA	FORECAST	SEASONAL	MA	sales	.	1

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	.	.	.	.	.		.	(2 1 0)(0 1 1)s
2	1	.	.	.	.	.	.		.	(2 1 0)(0 1 1)s
3	1	.	.	.	.	.	.		.	(2 1 0)(0 1 1)s
4	1	.	0	-0.36159	0.086055	-4.20188	0.00005		.	(2 1 0)(0 1 1)s
5	2	.	0	-0.06366	0.086141	-0.73905	0.46120		.	(2 1 0)(0 1 1)s
6	1	.	0	0.56109	0.072814	7.70588	0.00000		.	(2 1 0)(0 1 1)s

The following example shows the use of PICKMDL statement option METHOD=BEST to select the model. The chosen model is shown in the mdlchosen data set in Output 46.9.4. With METHOD=BEST, a different model is chosen than either of the previous models chosen. Because the order in which the models occur in the MDLINFOIN= data set affects model selection when METHOD=FIRST is specified, it is a common practice to list models from the simplest model to the most complex in the MDLINFOIN= data set that is used in conjunction with the PICKMDL statement.

```
proc x13 data=sales date=date mdlinfoin=Models mdlinfoout=mdlchosen;
   var sales;
   transform function=log;
   pickmdl method=best;
run;

title 'Chosen Model';
proc print data=mdlchosen;
run ;
```

Output 46.9.4 The Model Chosen from the Five Commonly Used Models, METHOD=BEST

Chosen Model

Obs	_NAME_	_MODEL_	_MODELTYPE_	_MODELPART_	_COMPONENT_	_PARMTYPE_	_DSVAR_	_VALUE_	_FACTOR_
1	sales	MODEL2	ARIMA	FORECAST	TRANSFORM	LOG	sales	.	.
2	sales	MODEL2	ARIMA	FORECAST	NONSEASONAL	DIF	sales	.	.
3	sales	MODEL2	ARIMA	FORECAST	SEASONAL	DIF	sales	.	.
4	sales	MODEL2	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
5	sales	MODEL2	ARIMA	FORECAST	NONSEASONAL	MA	sales	.	1
6	sales	MODEL2	ARIMA	FORECAST	SEASONAL	MA	sales	.	2

Obs	_LAG_	_SHIFT_	_NOEST_	_EST_	_STDERR_	_TVALUE_	_PVALUE_	_STATUS_	_SCORE_	_LABEL_
1	.	.	.	.	.	.	.		.	(0 1 2)(0 1 1)s
2	1	.	.	.	.	.	.		.	(0 1 2)(0 1 1)s
3	1	.	.	.	.	.	.		.	(0 1 2)(0 1 1)s
4	1	.	0	0.39613	0.086126	4.59937	0.00001		.	(0 1 2)(0 1 1)s
5	2	.	0	0.03961	0.086163	0.45966	0.64652		.	(0 1 2)(0 1 1)s
6	1	.	0	0.55903	0.076446	7.31277	0.00000		.	(0 1 2)(0 1 1)s

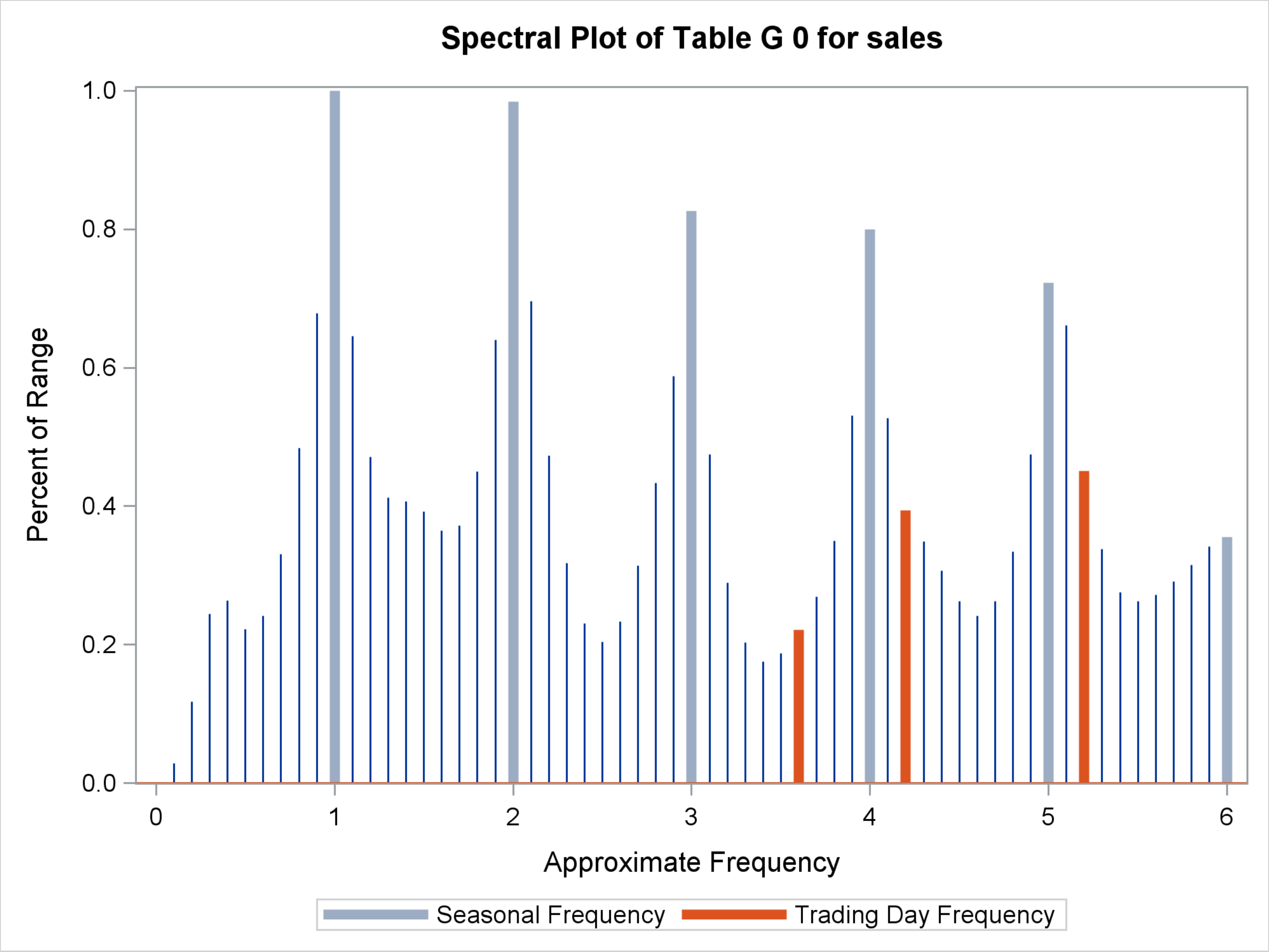
Example 46.10: Illustration of ODS Graphics

This example illustrates the use of ODS Graphics. Using the same data set as in the section "Basic Seasonal Adjustment" on page 3288 and the previous examples, a spectral plot of the original series is displayed in Output 46.10.1.

The graphical displays are available when ODS Graphics is enabled. For specific information about the graphics available in the X13 procedure, see the section "ODS Graphics" on page 3355.

```
proc x13 data=sales date=date;
   var sales;
run;
```

Output 46.10.1 Spectral Plot for Original Data



Example 46.11: AUXDATA= Data Set

This example demonstrates the use of the AUXDATA= data set to input user-defined regressors for use in the regARIMA model. User-defined regressors are often economic indicators, but in this example a user-defined regressor is generated in the following statements:

```
data auxreg(keep=date lengthofmonth);
   set sales;
   lengthofmonth = (INTNX('MONTH',date,1) - date) - (365/12);
   format date monyy.;
run;
```

When you use the AUXDATA= data set, it is not necessary to merge the user-defined regressor data set with the DATA= data set. The following statements input the regressor lengthofmonth in the data set auxreg. The regressor lengthofmonth is specified in the REGRESSION statement, and the data set auxreg is specified in the AUXDATA= option in the PROC X13 statement.

```
title 'Align lengthofmonth Regressor from Auxreg to First Three Years';
ods select regParameterEstimates;
proc x13 data=sales(obs=36) date=date auxdata=auxreg;
   var sales;
   regression uservar=lengthofmonth;
   arima model=((0 1 1) (0 1 1));
   estimate;
run;

title 'Align lengthofmonth Regressor from Auxreg to Second Three Years';
ods select regParameterEstimates;
proc x13 data=sales(firstobs=37 obs=72) date=date auxdata=auxreg;
   var sales;
   regression uservar=lengthofmonth;
   arima model=((0 1 1) (0 1 1));
   estimate;
run;
```

Output 46.11.1 and Output 46.11.2 display the parameter estimates for the two series.

Output 46.11.1 Using Regressors in the AUXDATA= Data for the First Three Years of Series

Align lengthofmonth Regressor from Auxreg to First Three Years

The X13 Procedure

Regression Model Parameter Estimates
For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	lengthofmonth	Est	2.98046	5.36251	0.56	0.5840

Output 46.11.2 Using Regressors in the AUXDATA= Data for the Second Three Years of Series

Align lengthofmonth Regressor from Auxreg to Second Three Years

The X13 Procedure

Regression Model Parameter Estimates
For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	lengthofmonth	Est	-0.51215	8.43145	-0.06	0.9521

The X13 procedure uses the date variable in the sales data set and the auxreg data set to align the user-defined regressors.

In the following example, the DATA= data set salesby contains BY groups. The X13 procedure aligns the regressor in the auxreg data set to each BY group in the salesby data set according to the variable date that is specified by the DATE= option in the PROC X13 statement. The variable date must be present in the auxreg data set to align the values.

```
data salesby;
     set sales(obs=72);
     if ( _n_ < 37 ) then by=1;
     else by=2;
run;
ods select regParameterEstimates;
title 'Align lengthofmonth Regressor from Auxreg to BY Groups';
proc x13 data=salesby date=date auxdata=auxreg;
   var sales;
   by by;
   regression uservar=lengthofmonth;
   arima model=((0 1 1) (0 1 1));
   estimate;
run;
```

The results in Output 46.11.3 match the previous results in Output 46.11.1 and Output 46.11.2.

Output 46.11.3 Using Regressors in the AUXDATA= Data with BY Groups

Align lengthofmonth Regressor from Auxreg to BY Groups

The X13 Procedure

by=1

Regression Model Parameter Estimates For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	lengthofmonth	Est	2.98046	5.36251	0.56	0.5840

Align lengthofmonth Regressor from Auxreg to BY Groups

The X13 Procedure

by=2

Regression Model Parameter Estimates For Variable sales

Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > |t|
User Defined	lengthofmonth	Est	-0.51215	8.43145	-0.06	0.9521

References

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

Cholette, P. A. (1979). *A Comparison and Assessment of Various Adjustment Methods of Sub-annual Series to Yearly Benchmarks*. StatCan Staff Paper STC2119, Seasonal Adjustment and Time Series Staff, Statistics Canada, Ottawa.

Dagum, E. B. (1983). *The X-11-ARIMA Seasonal Adjustment Method*. Technical Report 12-564E, Statistics Canada, Ottawa.

Dagum, E. B. (1988). *The X-11-ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual*. Ottawa: Statistics Canada.

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

Gómez, V., and Maravall, A. (1997a). *Guide for Using the Programs TRAMO and SEATS, Beta Version*. Madrid: Banco de España.

Gómez, V., and Maravall, A. (1997b). *Programs TRAMO and SEATS: Instructions for the User, Beta Version*. Madrid: Banco de España.

Hillmer, S. C., and Tiao, G. C. (1982). "An ARIMA-Model-Based Approach to Seasonal Adjustment." *Journal of the American Statistical Association* 77:63–70.

Huot, G. (1975). *Quadratic Minimization Adjustment of Monthly or Quarterly Series to Annual Totals*. StatCan Staff Paper STC2104, Statistics Canada, Seasonal Adjustment and Time Series Staff, Ottawa.

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

Ljung, G. M. (1993). "On Outlier Detection in Time Series." *Journal of the Royal Statistical Society, Series B* 55:559–567.

Lothian, J., and Morry, M. (1978a). *A Set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*. StatCan Staff Paper STC1788E, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.

Lothian, J., and Morry, M. (1978b). *A Test for the Presence of Identifiable Seasonality When Using the X-11-ARIMA Program*. StatCan Staff Paper STC2118, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.

Shiskin, J., Young, A. H., and Musgrave, J. C. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. Technical Report 15, US Department of Commerce, Bureau of the Census.

US Bureau of the Census (2009). *X-12-ARIMA Reference Manual, Version 0.3*. US Bureau of the Census, Washington, DC. `http://www.census.gov/ts/x12a/v03/x12adocV03.pdf`.

US Bureau of the Census (2010). *X-12-ARIMA Seasonal Adjustment Program, Version 0.3*. US Bureau of the Census, Washington, DC.

US Bureau of the Census (2013a). *X-13ARIMA-SEATS Quick Reference for DOS, Version 1.1*. US Bureau of the Census, Washington, DC. `http://www.census.gov/ts/x13as/pc/qrefX13ASpc.pdf`.

US Bureau of the Census (2013b). *X-13ARIMA-SEATS Quick Reference for UNIX/Linux, Version 1.1*. US Bureau of the Census, Washington, DC. `http://www.census.gov/ts/x13as/unix/qrefX13ASunix.pdf`.

US Bureau of the Census (2013c). *X-13ARIMA-SEATS Reference Manual, Version 1.1*. US Bureau of the Census, Washington, DC. `http://www.census.gov/ts/x13as/docX13ASHTML.pdf`.

Subject Index

Syntax Index