Chapter 21
The X11 Procedure

Chapter Table of Contents

1152

Chapter 21
The X11 Procedure

Overview

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

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

The X11 procedure incorporates Sliding Spans Analysis. This type of analysis pro-
vides a diagnostic for determining the suitability of seasonal adjustment for an eco-
nomic series.

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

Getting Started

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

The following example shows how to use PROC X11 to produce a seasonally adjusted series, $C_t I_t$ from an original series $O_t = S_t C_t D_t I_t$.

In the multiplicative model, the trend cycle component C_t keeps the same scale as the original series O_t, while S_t, D_t, and I_t vary around 1.0. In all printed tables and in the output data set, these latter components are expressed as percentages, and thus will vary around 100.0 (in the additive case, they vary around 0.0).

The naming convention used in PROC X11 for the tables follows the original U.S. Bureau of the Census X-11 Seasonal Adjustment program specification; refer the U.S. Bureau of the Census, 1967, and "Printed Output" later in this chapter. This convention is outlined in Figure 21.1.

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

Basic Seasonal Adjustment

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

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

```
data sales;
   input sales @@;
   date = intnx( 'month', '01sep1978'd, _n_-1 );
   format date monyy7.;
   datalines;
run;

proc x11 data=sales;
   monthly date=date;
   var sales;
   tables d11;
run;
```

```
                    The X11 Procedure

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

The X-11 program is divided into seven major parts.
 Part          Description
  A.  Prior adjustments, if any
  B.  Preliminary estimates of irregular component weights
          and regression trading day factors
  C.  Final estimates of above
  D.  Final estimates of seasonal, trend-cycle and
          irregular components
  E.  Analytical tables
  F.  Summary measures
  G.  Charts


                     Series - sales
             Period covered - 9/1978 to 8/1990
       Type of run: multiplicative seasonal adjustment.
                  No printout.   No charts.
Sigma limits for graduating extreme values are 1.5 and 2.5
Irregular values outside of 2.5-sigma limits are excluded
                from trading day regression
```

```
                        The X11 Procedure

                   Seasonal Adjustment of - sales

                D11 Final Seasonally Adjusted Series
Year      JAN        FEB        MAR        APR        MAY        JUN

1978       .          .          .          .          .          .
1979    124.935    126.533    125.282    125.650    127.754    129.648
1980    128.734    139.542    143.726    143.854    148.723    144.530
1981    176.329    166.264    167.433    167.509    173.573    175.541
1982    186.747    202.467    192.024    202.761    197.548    206.344
1983    233.109    223.345    218.179    226.389    224.249    227.700
1984    238.261    239.698    246.958    242.349    244.665    247.005
1985    275.766    282.316    294.169    285.034    294.034    296.114
1986    325.471    332.228    330.401    330.282    333.792    331.349
1987    363.592    373.118    368.670    377.650    380.316    376.297
1988    370.966    384.743    386.833    405.209    380.840    389.132
1989    428.276    418.236    429.409    446.467    437.639    440.832
1990    480.631    474.669    486.137    483.140    481.111    499.169
--------------------------------------------------------------------
Avg     277.735    280.263    282.435    286.358    285.354    288.638


                D11 Final Seasonally Adjusted Series
Year      JUL        AUG        SEP        OCT        NOV        DEC      Total

1978       .          .       123.507    125.776    124.735    129.870    503.887
1979    127.880    129.285    126.562    134.905    133.356    136.117    1547.91
1980    140.120    153.475    159.281    162.128    168.848    165.159    1798.12
1981    179.301    182.254    187.448    197.431    184.341    184.304    2141.73
1982    211.690    213.691    214.204    218.060    228.035    240.347    2513.92
1983    222.045    222.127    222.835    212.227    230.187    232.827    2695.22
1984    251.247    253.805    264.924    266.004    265.366    277.025    3037.31
1985    294.196    309.162    311.539    319.518    318.564    323.921    3604.33
1986    337.095    341.127    346.173    350.183    360.792    362.333    4081.23
1987    379.668    375.607    374.257    372.672    368.135    364.150    4474.13
1988    385.479    377.147    397.404    403.156    413.843    416.142    4710.89
1989    450.103    454.176    460.601    462.029    427.499    485.113    5340.38
1990    485.370    485.103       .          .          .          .       3875.33
---------------------------------------------------------------------
Avg     288.683    291.413    265.728    268.674    268.642    276.442

             Total:  40324  Mean:  280.03  S.D.:  111.31
```

Figure 21.1. Basic Seasonal Adjustment

You can compare the original series, table B1, and the final seasonally adjusted series, table D11 by plotting them together. These tables are requested and named in the OUTPUT statement.

```
    title 'Monthly Retail Sales Data (in $1000)';

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

    symbol1 i=join v='star';
    symbol2 i=join v='circle';
    legend1 label=none value=('original' 'adjusted');
```

```
proc gplot data=out;
   plot sales    * date = 1
        adjusted * date = 2 / overlay legend=legend1;
run;
```
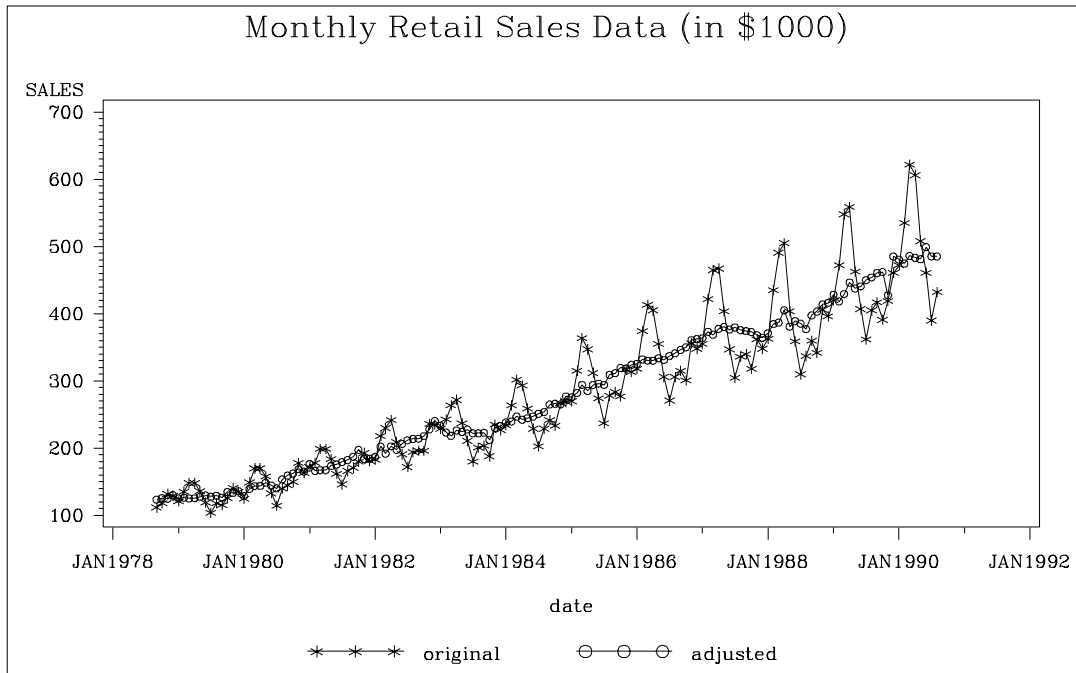
Figure 21.2. Plot of Original and Seasonally Adjusted Data

X-11-ARIMA

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

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

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

1157

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

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

```
proc x11 data=sales;
   monthly date=date;
   var sales;
   arima;
run;
```
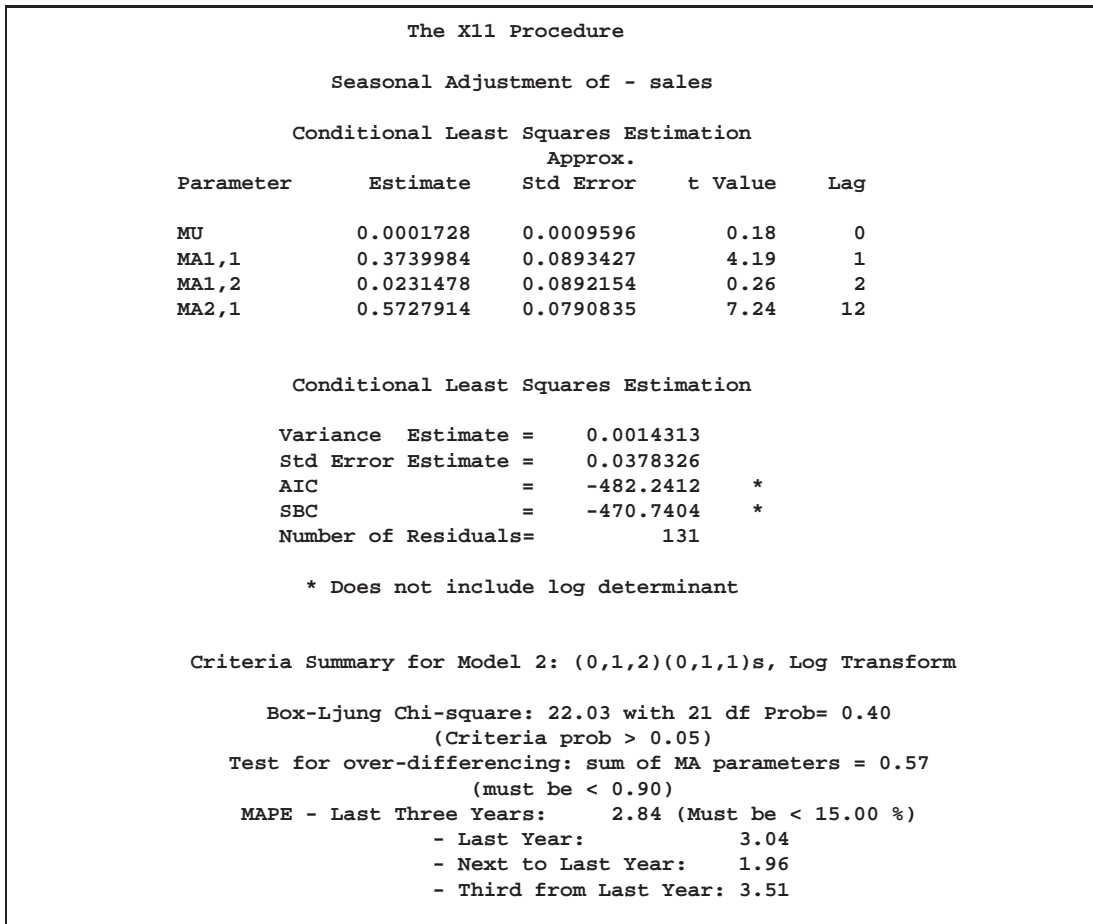
```
                    The X11 Procedure

              Seasonal Adjustment of - sales

           Conditional Least Squares Estimation
                          Approx.
     Parameter      Estimate     Std Error     t Value     Lag

     MU            0.0001728     0.0009596        0.18       0
     MA1,1         0.3739984     0.0893427        4.19       1
     MA1,2         0.0231478     0.0892154        0.26       2
     MA2,1         0.5727914     0.0790835        7.24      12


           Conditional Least Squares Estimation

           Variance  Estimate =     0.0014313
           Std Error Estimate =     0.0378326
           AIC               =     -482.2412     *
           SBC               =     -470.7404     *
           Number of Residuals=           131

             * Does not include log determinant


   Criteria Summary for Model 2: (0,1,2)(0,1,1)s, Log Transform

        Box-Ljung Chi-square: 22.03 with 21 df Prob= 0.40
                   (Criteria prob > 0.05)
     Test for over-differencing: sum of MA parameters = 0.57
                     (must be < 0.90)
     MAPE - Last Three Years:     2.84 (Must be < 15.00 %)
                   - Last Year:            3.04
                   - Next to Last Year:    1.96
                   - Third from Last Year: 3.51
```

Figure 21.3. X-11-ARIMA Model Selection

1158

Table D11 (final seasonally adjusted series) is now constructed using symmetric weights on observations at the end of the actual data. This should result in better estimates of the seasonal factors and, thus, smaller revisions in D11 as more data become available.

Syntax

The X11 procedure uses the following statements:

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

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

Functional Summary

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

Description	Statement	Option
Data Set Options		
specify input data set	PROC X11	DATA=
write the trading-day regression results to an output data set	PROC X11	OUTTDR=
write the stable seasonality test results to an output data set	PROC X11	OUTSTB=
write table values to an output data set	OUTPUT	OUT=
add extrapolated values to the output data set	PROC X11	OUTEX
add year ahead estimates to the output data set	PROC X11	YRAHEADOUT
write the sliding spans analysis results to an output data set	PROC X11	OUTSPAN=

Description	Statement	Option
Printing Control Options		
suppress all printed output	PROC X11	NOPRINT
suppress all printed ARIMA output	ARIMA	NOPRINT
print all ARIMA output	ARIMA	PRINTALL
print selected tables and charts	TABLES	
print selected groups of tables	MONTHLY	PRINTOUT=
	QUARTERLY	PRINTOUT=
print selected groups of charts	MONTHLY	CHARTS=
	QUARTERLY	CHARTS=
print preliminary tables associated with ARIMA processing	ARIMA	PRINTFP
specify number of decimals for printed tables	MONTHLY	NDEC=
	QUARTERLY	NDEC=
suppress all printed SSPAN output	SSPAN	NOPRINT
print all SSPAN output	SSPAN	PRINTALL
Date Information Options		
specify a SAS Date variable	MONTHLY	DATE=
	QUARTERLY	DATE=
specify the beginning date	MONTHLY	START=
	QUARTERLY	START=
specify the ending date	MONTHLY	END=
	QUARTERLY	END=
specify beginning year for trading-day regression	MONTHLY	TDCOMPUTE=
Declaring the Role of Variables		
specify BY-group processing	BY	
specify the variables to be seasonally adjusted	VAR	
specify identifying variables	ID	
specify the prior monthly factor	MONTHLY	PMFACTOR=
Controlling the table computations		
use additive adjustment	MONTHLY	ADDITIVE
	QUARTERLY	ADDITIVE
specify seasonal factor moving average length	MACURVES	
specify the extreme value limit for trading-day regression	MONTHLY	EXCLUDE=
specify the lower bound for extreme irregulars	MONTHLY	FULLWEIGHT=
	QUARTERLY	FULLWEIGHT=
specify the upper bound for extreme irregulars	MONTHLY	ZEROWEIGHT=
	QUARTERLY	ZEROWEIGHT=

Description	Statement	Option
include the length-of-month in trading-day regression	MONTHLY	LENGTH
specify trading-day regression action	MONTHLY	TDREGR=
compute summary measure only	MONTHLY	SUMMARY
	QUARTERLY	SUMMARY
modify extreme irregulars prior to trend cycle estimation	MONTHLY	TRENDADJ
	QUARTERLY	TRENDADJ
specify moving average length in trend cycle estimation	MONTHLY	TRENDMA=
	QUARTERLY	TRENDMA=
specify weights for prior trading-day factors	PDWEIGHTS	

PROC X11 Statement

> **PROC X11** *options;*

The following options can appear in the PROC X11 statement:

DATA= *SAS-data-set*

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

OUTEXTRAP

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

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

NOPRINT

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

OUTSPAN= *SAS-data-set*

Specifies the output data set to store the sliding spans analysis results. Tables A1, C18, D10 and D11 for each span are written to this data set. See "OUTSPAN Data Set" later in this chapter for details.

OUTSTB= *SAS-data-set*

Specifies the output data set to store the stable seasonality test results (table D8). All the information in the analysis of variance table associated with the stable seasonality

test is contained in the variables written to this data set. See "OUTSTB Data Set" later in this chapter for details.

OUTTDR= *SAS-data-set*

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

YRAHEADOUT

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

ARIMA Statement

ARIMA *options;*

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

There are two ways of obtaining an ARIMA model to extend the series. A model can be given explicitly with the MODEL= and TRANSFORM= options. Alternatively, the best fitting model from a set of five predefined models is found automatically whenever the MODEL= option is absent. See "Automatic Model Selection" later in this chapter for details.

BACKCAST= *n*

Specifies the number of years to backcast the series. The default is BACKCAST= 0. See "Effect of Backcast and Forecast Length" later in this chapter for details.

CHICR= *value*

specifies the criteria for the significance level for the Box-Ljung chi-square test for lack of fit when testing the five predefined models. The default is CHICR= 0.05. The CHICR= option values must be between 0.01 and 0.90. The hypothesis being tested is that of model adequacy. Nonrejection of the hypothesis is evidence for an adequate model. Making the CHICR= value smaller makes it easier to accept the model. See "Criteria Details" later in this chapter for further details on the CHICR= option.

CONVERGE= *value*

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

FORECAST= *n*

> Specifies the number of years to forecast the series. The default is FORECAST= 1. See "Effect of Backcast and Forecast Length" later in this chapter for details.

MAPECR= *value*

> specifies the criteria for the Mean Absolute Percent Error (MAPE) when testing the five predefined models. A small MAPE value is evidence for an adequate model; a large MAPE value results in the model being rejected. The MAPECR= value is the boundary for acceptance/rejection. Thus a larger MAPECR= value would make it easier for a model to pass the criteria. The default is MAPECR= 15. The MAPECR= option values must be between 1 and 100. See "Criteria Details" later in this chapter for further details on the MAPECR= option.

MAXITER= *n*

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

METHOD= CLS
METHOD= ULS
METHOD= ML

> specifies the estimation method. ML requests maximum likelihood, ULS requests unconditional least-squares, and CLS requests conditional least-squares. METHOD=CLS is the default. The maximum likelihood estimates are more expensive to compute than the conditional least-squares estimates. In some cases, however, they may be preferable. For further information on the estimation methods, see "Estimation Details" in Chapter 7, "The ARIMA Procedure."

MODEL= (P=_n1_ **Q=**_n2_ **SP=**_n3_ **SQ=**_n4_ **DIF=**_n5_ **SDIF=**_n6_ **<NOINT> <CENTER>)**

> specifies the ARIMA model. The AR and MA orders are given by P=_n1_ and Q=_n2_ respectively, while the seasonal AR and MA orders are given by SP=_n3_ and SQ=_n4_. The lag corresponding to seasonality is determined by the MONTHLY or QUARTERLY statement. Similarly, differencing and seasonal differencing are given by DIF=_n5_ and SDIF=_n6_ respectively.
>
> For example
>
> ```
> arima model=( p=2 q=1 sp=1 dif=1 sdif=1 );
> ```
>
> specifies a $(2,1,1)(1,1,0)s$ model, where *s*, the seasonality is either 12 (monthly) or 4 (quarterly). More examples of the MODEL= syntax is given in the "Automatic Model Selection" section.

NOINT

> suppresses the fitting of a constant (or intercept) parameter in the model. (That is, the parameter μ is omitted.)

CENTER

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

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

```
arima model=( p=1 center noint );
```

NOPRINT

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

OVDIFCR= *value*

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

PRINTALL

provides the same output as the default printing for all models fit and, in addition, prints an estimation summary and chi-square statistics for each model fit. See "Printed Output" later in this chapter for details.

PRINTFP

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

TRANSFORM= (LOG) | LOG
TRANSFORM= (*constant ** power* **)**

The ARIMA statement in PROC X11 allows certain transformations on the series before estimation. The specified transformation is applied only to a user-specified model. If TRANSFORM= is specified without a MODEL=, the transformation request is ignored and a warning is printed.

The LOG transformation requests that the natural log of the series be used for estimation. The resulting forecasted values are transformed back to the original scale.

A general power transformation of the form $X_t \rightarrow (X_t + a)^b$ is obtained by specifying

```
transform= ( a ** b )
```

1164

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

BY Statement

BY *variables;*

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

ID Statement

ID *variables;*

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

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

MACURVES Statement

MACURVES *month=option ...;*

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

The *month=option* specifications consist of the month name (or the first three letters of the month name), an equal sign, and one of the following option values.

'3'   specifies a three-term moving average for the month

'3X3'   specifies a three-by-three moving average

'3X5'   specifies a three-by-five moving average

'3X9'   specifies a three-by-nine moving average

STABLE   specifies a stable seasonal factor (average of all values for the month)

1165

For example, the statement

```
macurves jan='3' feb='3x3' march='3x5' april='3x9';
```

uses a three-term moving average to estimate seasonal factors for January, a 3x3 (a three-term moving average of a three term moving average) for February, a 3x5 (a three-term moving average of a five-term moving average) for March, and a 3x9 (a three-term moving average of a nine-term moving average) for April.

The numeric values used for the weights of the various moving averages and a discussion of the derivation of these weights are given in U.S. Bureau of Census, 1967. A general discussion of moving average weights is given in Dagum, 1985.

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

MONTHLY Statement

MONTHLY *options;*

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

The following options can appear in the MONTHLY statement:

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

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

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

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

```
monthly date=date;
tables g1 g4;
```

DATE= *variable*

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

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

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

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

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

END= *mmmyyyy*

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

EXCLUDE= *value*

excludes from the trading-day regression any irregular values that are more than *value* standard deviations from the mean. The EXCLUDE=value must be between .1 and 9.9, with the default value being 2.5.

FULLWEIGHT= *value*

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

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

LENGTH

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

NDEC= *n*

specifies the number of decimal places shown on the printed tables on the listing. This option has no effect on the precision of the variables values in the output data set.

PMFACTOR= *variable*

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

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

See "Prior Daily Weights and Trading-Day Regression" in the "Details" section later in this chapter for details and examples using the PMFACTOR= option.

PRINTOUT= STANDARD | LONG | FULL | NONE

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

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

START= *mmmyyyy*

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

1168

SUMMARY

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

TDCOMPUTE= *year*

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

TDREGR= NONE | PRINT | ADJUST | TEST

specifies the treatment of trading-day regression. The value NONE omits the computation of the trading-day regression. The value PRINT computes and prints the trading-day regressions but does not adjust the series. The value ADJUST computes and prints the trading-day regression and adjusts the irregular components to obtain preliminary weights. The value TEST adjusts the final series if the trading-day regression estimates explain significant variation on the basis of an *F* test (or residual trading-day variation if prior weights are used). The default is TDREGR=NONE.

See "Prior Daily Weights and Trading-Day Regression" in the "Details" section later in this chapter for details and examples using the TDREGR= option.

If ARIMA processing is requested, any value of TDREGR other than the default TDREGR=NONE will cause PROC X11 to perform an initial pass (see the "Details" section and the PRINTFP option).

The signifigance level reported Table C15 should be viewed with caution. The dependent variable in the trading day regression is the irregular component formed by an averaging operation. This induces a correlation in the dependent variable and hence in the residuals from which which the F-test is computed. Hence the distribution of the trading day regression F-statistics differs from an exact F; see Cleveland and Devlin, 1980 for details.

TRENDADJ

modifies extreme irregular values prior to computing the trend cycle estimates in the first iteration. If the TRENDADJ option is omitted, the trend cycle is computed without modifications for extremes.

TRENDMA= 9 | 13 | 23.

specifies the number of terms in the moving average to be used by the procedure in estimating the variable trend cycle component. The value of the TRENDMA= option must be 9, 13, or 23. If the TRENDMA= option is omitted, the procedure selects an appropriate moving average. For information concerning the number of terms in the moving average, see U.S. Bureau of the Census (1967).

ZEROWEIGHT= *value*

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

1169

value 1.3 standard deviations from the mean would be assigned a graduated weight. The ZEROWEIGHT=value must be between .1 and 9.9 but must be greater than the FULLWEIGHT=value. The default is ZEROWEIGHT=2.5.

The ZEROWEIGHT option can be used in conjunction with the FULLWEIGHT= option to adjust outliers from a monthly or quarterly series. See Example 21.3 later in this chapter for an illustration of this use.

OUTPUT Statement

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

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

OUT= *SAS-data-set*
If OUT= is omitted, the SAS System names the new data set using the DATA*n* convention.

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

tablename = var1 var2 ...

The *tablename* keywords that can be used in the OUTPUT statement are listed in "Printed Output" later in this chapter. The following is an example of a VAR and OUTPUT statement.

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

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

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

PDWEIGHTS Statement

> **PDWEIGHTS** *day=w ... ;*

The PDWEIGHTS statement can be used to specify one to seven daily weights. The statement can only be used with monthly series using the multiplicative model. These

weights are used to compute prior trading-day factors, which are then used to adjust the original series prior to the seasonal adjustment process. Only relative weights are needed; the X11 procedure adjusts the weights so that they sum to 7.0. The weights can also be corrected by the procedure on the basis of estimates of trading-day variation from the input data.

See "Prior Daily Weights and Trading-Day Regression" in the "Details" section later in this chapter for details and examples using the PDWEIGHTS statement.

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

The following is an example of a PDWEIGHTS statement.

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

Any number of days can be specified with one PDWEIGHTS statement. The default weight value for any day that is not specified is 0. If you do not use a PDWEIGHTS statement, the program computes daily weights if TDREGR=ADJUST is specified. Refer to U.S. Bureau of the Census (1967) for details.

QUARTERLY Statement

> **QUARTERLY** *options;*

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

The following options can appear in the QUARTERLY statement.

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

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

the TABLES statement. However, if both the CHARTS= option and a TABLES statement are given, the charts corresponding to the CHARTS= option and those requested by the TABLES statement are printed.

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

```
quarterly date=date;
tables g1 g4;
```

DATE= *variable*

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

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

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

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

```
quarterly date=variable start='1974q1';
```

END= *'yyyyQq'*

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

FULLWEIGHT= *value*

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

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

NDEC= *n*

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

PRINTOUT= STANDARD
PRINTOUT= LONG
PRINTOUT= FULL
PRINTOUT= NONE

> specifies the tables to print. If PRINTOUT=STANDARD is specified, between 17 and 27 tables are printed, depending on the other options that are specified. PRINT-OUT=LONG prints between 27 and 39 tables, and PRINTOUT=FULL prints between 44 and 59 tables. Specifying PRINTOUT=NONE results in no tables being printed. The default is PRINTOUT=STANDARD.
>
> The TABLES statement can also specify particular quarterly tables to be printed. If no PRINTOUT= is given, and a TABLES statement is given, the TABLES statement overrides the default value of PRINTOUT=STANDARD; that is, no tables (or charts) are printed except those given in the TABLES statement. However, if both the PRINTOUT= option and a TABLES statement are given, the tables corresponding to the PRINTOUT= option and those requested by the TABLES statement are printed.

START= *'yyyyQq'*

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

SUMMARY

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

TRENDADJ

> modifies extreme irregular values prior to computing the trend cycle estimates. If this option is omitted, the trend cycle is computed without modification for extremes.

ZEROWEIGHT= *value*

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

The ZEROWEIGHT option can be used in conjunction with the FULLWEIGHT= option to adjust outliers from a monthly or quarterly series. See Example 21.3 later in this chapter for an illustration of this use.

SSPAN Statement

SSPAN *options ;*

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

The following options can appear in the SSPAN Statement:

NDEC= *n*

specifies the number of decimal places shown on selected Sliding Span reports. This option has no effect on the precision of the variables values in the OUTSPAN output data set.

CUTOFF= *value*

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

TDCUTOFF= *value*

gives the percentage value for determining an excessive difference within a span for the trading day factors. The default value is 2.0. The use of the TDCUTOFF=value in determining the maximum percent difference (MPD) is described in "Computational Details" later in this section. Caution should be used in changing the default TDCUTOFF=value. The empirical threshold ranges found by the Census Bureau no longer apply when value is changed.

NOPRINT

suppresses all sliding spans reports.

PRINT

prints the summary sliding spans reports S 0 - S 6.E.

PRINTALL

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

TABLES Statement

TABLES *tablenames;*

The TABLES statement prints the tables specified in addition to the tables that are printed as a result of the PRINTOUT= option in the MONTHLY or QUARTERLY statement. Table names are listed in Table 21.3 later in this chapter.

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

```
tables d10 d11;
```

VAR Statement

VAR *variables;*

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

1175

Details

Historical Development of X-11

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

Seasonal adjustment methods began development in the 1920s and 1930s before there were suitable analytic models available and before electronic computing devices were developed. The lack of any suitable model led to methods that worked the same for any series, that is, methods that were not model-based and that could be applied to any series. Experience with economic series had shown that a given mathematical form could adequately represent a time series only for a fixed length; as more data was added, the model became inadequate. This suggested an approach using moving averages.

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

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

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

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

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

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

1176

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

References for this work during this period include Shiskin and Eisenpress (1957), Shiskin (1958), and Marris (1960). The authoritative documentation for the X-11 Variant is in U.S. Bureau of the Census (1967). This document is not equivalent to a program specification; however the FORTRAN code implementing the X-11 Variant is in the public domain. A less detailed description of the X-11 Variant is given in U.S. Bureau of the Census (1969).

Development of the X-11-ARIMA Method

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

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

The Time Series Research and Analysis Division of Statistic Canada investigated 174 Canadian economic series and found five ARIMA models out of twelve that fit the majority of series well and reduced revisions for the most recent months. References giving details of various aspects of the X-11-ARIMA methodology include Dagum (1980, 1982a, 1982b, 1983, 1988), Laniel (1985), Lothian and Morry (1978a), and Huot,Chui, Higginson, and Gait (1986).

Differences between X11ARIMA/88 and PROC X11

The original implementation of the X-11-ARIMA method was by Statistics Canada in 1980 (Dagum, 1980; X11ARIMA/80), with later changes and enhancements made in 1988 (Dagum, 1988; X11ARIMA/88). The calculations performed by PROC X11 differ from those in X11ARIMA/88, which will result in differences in the final component estimates provided by these implementations.

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

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

Additional differences can result from the ARIMA estimation. X11ARIMA/88 uses Conditional Least Squares (CLS), while CLS, Unconditional Least Squares (ULS) and Maximum Likelihood (ML) are all available in PROC X11 by using the METHOD= option on the ARIMA statement. Generally, parameters estimates will differ for the different methods.

Implementation of the X-11 Seasonal Adjustment Method

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

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

$$O_t = C_t S_t I_t P_t D_t,$$

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

The trading-day component can be further factored as

$$D_t = D_{r,t} D_{tr,t},$$

where $D_{tr,t}$ are the trading-day factors derived from the prior daily weights, and $D_{r,t}$ are the residual trading-day factors estimated from the trading-day regression.

Additional steps when using the X-11-ARIMA method

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

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

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

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

The Standard X-11 Seasonal Adjustment Method

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

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

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

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

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

   $$\widehat{S_t I_t} = \frac{O_t}{\hat{C}_t}$$

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

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

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

4. A third and final iteration is performed using the original series that has been adjusted for trading-day factors and irregular weights computed during the second iteration. During the third iteration, PROC X11 develops final estimates of

seasonal factors, the seasonally adjusted series, the trend cycle, and the irregular components. The procedure computes summary measures of variation and produces a moving average of the final adjusted series.

Sliding Spans Analysis

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

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

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

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

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

Seasonal Factors (Table D10): $S_t(k)$

Trading Day Factor (Table C18): $TD_t(k)$

Seasonally Adjust Data (Table D11): $SA_t(k)$

Month-to-month changes in the Seasonally Adjust Data: $MM_t(k)$

Year-to-Year changes in the Seasonally Adjust Data: $YY_t(k)$

The key measure is the maximum percent difference across spans. For example, consider a series beginning in JAN72, ending in DEC84, and having four spans, each

1180

of length 8 years (see Figure 1. in Findley et al., 1990, page 346). Consider $S_t(k)$ the seasonal factor (table D10) for month t for span k, and let N_t denote the number of spans containing month t, i.e.,

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

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

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

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

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

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

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

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

Computation Details for Sliding Spans Analysis

Length and Number of Spans

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

First, an initial length based on MACURVE specification is determined, then the maximum number of spans possible using this length is determined. If this maximum number exceed four, set the number of spans to four. If this maximum number is one or zero, there is not enough observations to perform the sliding spans analysis. In this case a note is written to the log and the sliding spans analysis is skipped for this variable.

If the maximum number of spans is two or three, the actual number of spans used is set equal to this maximum. Finally, the length is adjusted so that the spans begin in January (or the first quarter) of the beginning year of the span.

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

Seasonal Factors (Table D10): $S_t(k)$

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

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

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

Trading Day Factor (Table C18): $TD_t(k)$

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

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

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

Seasonally Adjust Data (Table D11): $SA_t(k)$

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

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

A series for which less than 15% of the MPD values of D11 exceed 3.0% is stable; between 15% and 25% is marginally stable; and greater than 25% unstable. Span

reports S 4.A - S 4.C give the various breakdowns for the number of times the MPD exceeded these levels.

Month-to-month changes in the Seasonally Adjust Data: MM$_t$(k)

Some additional notation is needed for the month-to-month and year-to-year differences. Define N1$_t$ by

$$N1_t = \{k : \ span \ k \ contains \ month \ t \ and \ t - 1\}$$

For the additive model the month-to-month change for span k is defined by

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

while for the multiplicative model

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

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

$$MPD_t = max_{k\epsilon N1_t} MM_t(k) - min_{k\epsilon N1_t} MM_t(k)$$

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

Year-to-year changes in the Seasonally Adjust Data: YY$_t$(k)

First define $N12_t$ by

$$N12_t = \{k : \ span \ k \ contains \ month \ t \ and \ t - 12\}$$

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

For the additive model the month-to-month change for span k is defined by

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

while for the multiplicative model

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

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

$$MPD_t = max_{k \epsilon N1_t} YY_t(k) - min_{k \epsilon N1_t} YY_t(k)$$

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

Data Requirements

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

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

Missing Values

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

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

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

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

Prior Daily Weights and Trading-Day Regression

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

This phenomenon is called a calendar effect; it can be handled in PROC X11 by using the PDWEIGHTS (Prior Daily WEIGHTS) statement or the TDREGR=option

1184

(Trading-Day REGRession). The PDWEIGHTS statement and the TDREGR=option can be used separately or together.

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

In this example, it is assumed that the calendar variation is due to both prior daily weights and residual variation. Thus both a PDWEIGHTS statement and TDREGR= ADJUST are specified.

Note that only the relative weights are needed; in the actual computations, PROC X11 normalizes the weights to sum to 7.0. If a day of the week is not present in the PDWEIGHTS statement, it is given a value of zero. Thus "sun=0" is not needed.

```
proc x11 data=sales;
   monthly date=date tdregr=adjust;
   var sales;
   tables a1 a4 b15 b16 C14 C15 c18 d11;
   pdweights mon=1 tue=1 wed=1 thu=2 fri=2 sat=2;
   output out=x11out a1=a1 a4=a4 b1=b1 c14=c14
                     c16=c16 c18=c18 d11=d11;
run;
```

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

Adjustment for Prior Factors

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

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

This prior adjustment to SALES is performed to SALES by computing comp (SALES/PMF) * 100.0. A value of 100.0 for PMF leaves SALES unchanged, while a value of 50.0 for PMF doubles SALES. This value is the estimate of what SALES

would have been without the strike. The following example shows how this prior adjustment is accomplished.

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

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

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

The YRAHEADOUT Option

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

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

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

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

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

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

1186

Effect of Backcast and Forecast Length

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

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

Other studies (Pierce, 1980, Bobbit and Otto, 1990, Buszuwski, 1987) suggest "full forecasting"; that is, using enough forecasts to allow symmetric weights for the seasonal moving averages for the most current data. For example, if a 3x9 seasonal moving average was specified for one or more months using the MACURVES statement, five years of forecasts would be required. This is because the seasonal moving averages are performed on calendar months separately, and the 3x9 is an eleven-term centered moving average, requiring five observations before and after the current observation. Thus

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

would require five additional December values to compute the seasonal moving average.

Details of Model Selection

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

The five predefined models used by PROC X11 are the same as those used by X11ARIMA/88 from Statistics Canada. These particular models, shown in Table 21.1 were chosen on the basis of testing a large number of economics series (Dagum, 1988) and should provide reasonable forecasts for most economic series.

Table 21.1. Five Predefined Models

Model #	Specification	Multiplicative	Additive
1	(0,1,1)(0,1,1)s	log transform	no transform
2	(0,1,2)(0,1,1)s	log transform	no transform
3	(2,1,0)(0,1,1)s	log transform	no transform
4	(0,2,2)(0,1,1)s	log transform	no transform
5	(2,1,2)(0,1,1)s	no transform	no transform

The selection process proceeds as follows. The five models are estimated and one-step-ahead forecasts are produced in the order shown in Table 21.1. As each model is estimated the following three criteria are checked:

- The Mean Absolute Percent Error (MAPE) for the last three years of the series must be less than 15 %.

1187

- The significance probability for the Box-Ljung Chi-square for up to lag 24 for monthly (8 for quarterly) must greater than 0.05.

- The over-differencing criteria must not exceed 0.9.

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

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

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

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

If a BACKCAST= is specified, the chosen model form is used, but the parameters are reestimated using the reversed series. Using these parameters, the reversed series is forecasted for the number of years specified by the BACKCAST= option. These forecasts are then reversed and appended to the beginning of the original series, or the prior and calendar-adjusted series, to produce the backcasts.

Note that the final selection rule (the smallest MAPE using the last three years) emphasizes the quality of the forecasts at the end of the series. This is consistent with the purpose of the X-11-ARIMA methodology, namely, to improve the estimates of seasonal factors and thus minimize revisions to recent past data as new data become available.

Criteria Details

The Mean Absolute Percent Error (MAPE)

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

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

With this notation, the MAPE criteria is computed as

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

Box-Ljung Chi-Square

The Box-Ljung Chi-Square is a lack of fit test using the model residuals. This test statistic is computed using the Ljung-Box formula

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

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

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

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

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

Over-Differencing Test

From Table 21.1 you can se that all models have a single seasonal MA factor and at most two nonseasonal MA factors. Also, all models have seasonal and nonseasonal differencing. Consider model 2 applied to a monthly series y_t with $E(y_t) = \mu$:

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

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

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

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

for some $\alpha \neq 0.0$. Again, this results in cancellation and a lower order model.

Since the parameters are not exact, it is not reasonable to require that

$$\theta_3 < 1.0 \text{ and } \theta_1 + \theta_2 < 1.0$$

Instead, an approximate test is performed by requiring that

$$\theta_3 \leq 0.9 \text{ and } \theta_1 + \theta_2 \leq 0.9$$

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

ARIMA Statement Options for the Five Predefined Models

The following table lists the five predefined models and gives the equivalent MODEL= parameters in a PROC X11 ARIMA statement.

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

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

Table 21.2. ARIMA Statements Options for Predefined Models

Model	ARIMA Statement Options
(0,1,1)(0,1,1)s	MODEL=(Q=1 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG
(0,1,2)(0,1,1)s	MODEL=(Q=2 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG
(2,1,0)(0,1,1)s	MODEL=(P=2 SQ=1 DIF=1 SDIF=1) TRANSFORM=LOG
(0,2,2)(0,1,1)s	MODEL=(Q=2 SQ=1 DIF=2 SDIF=1) TRANSFORM=LOG
(2,1,2)(0,1,1)s	MODEL=(P=2 Q=2 SQ=1 DIF=1 SDIF=1)

OUT= Data Set

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

In addition, the variables specified by the option

> *tablename =var1 var2 . . . varn*

are placed in the OUT= data set. A list of tables available for monthly and quarterly series is given in Table 21.3.

The OUTSPAN Data Set

OUTSPAN= *SAS-data-set*

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

- A1, a numeric variable that is a copy of the original series truncated to the current span. Note that overlapping spans will contain identical values for this variable.

- C18, a numeric variable that contains the Trading Day Factors for the seasonal adjustment for the current span.

- D10, a numeric variable that contains the Seasonal Factors for the seasonal adjustment for the current span.

- D11, a numeric variable that contains the Seasonally Adjusted Series for the current span.

1190

- DATE, a numeric variable that contains the date within the current span.

- SPAN, a numeric variable that contains the current span. The first span is the earliest span, i.e., the one with the earliest begin date.

- VARNAME, a character variable containing the name of each variable in the VAR list. A separate sliding spans analysis is performed on each variable in the VAR list.

OUTSTB= Data Set

The output data set produced by the OUTSTB= option of the PROC X11 statement contains the information in the analysis of variance on table D8 (Final Unmodified S-I Ratios). This analysis of variance, following table D8 in the printed output, tests for stable seasonality (refer to U.S. Bureau of the Census, 1967, Appendix A). The variables in this data are:

- VARNAME, a character variable containing the name of each variable in the VAR list.

- TABLE, a character variable specifying the table from which the analysis of variance is performed. When ARIMA processing is requested, and two passes of X11 are required (when TDREGR=PRINT, TEST, or ADJUST), Table D8 and the stable seasonality test are computed twice; once in the initial pass, then again in the final pass. Both of these computations are put in the OUTSTB data set and are identified by D18.1 or D18.2 respectively.

- SOURCE, a character variable corresponding to the "source" column in the Analysis of Variance table following Table D8.

- SS, a numeric variable containing the sum of squares associated with the corresponding source term.

- DF, a numeric variable containing the degrees of freedom associated with the corresponding source term.

- MS, a numeric variable containing the mean square associated with the corresponding source term. MS is missing for the source term "Total."

- F, a numeric variable containing the F statistic for the "Between" source term. F will be missing for all other source terms.

- PROBF, a numeric variable containing the significance level for the F statistic. PROBF is missing for the source term "Total" and "Error."

OUTTDR= Data Set

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

- VARNAME, a character variable containing the name of the VAR variable being processed.

1191

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

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

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

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

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

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

- COEFF, a numeric variable containing the calculated regression coefficient for the given day. This variable is missing for all _TYPE_ = LOM_STD and _TYPE_ = ANOVA .

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

- T1, a numeric variable containing the *t*-statistic corresponding to the test that the combined weight is different from the prior weight. This variable is missing for all _TYPE_ = LOM_STD and _TYPE_ = ANOVA .

- T2, a numeric variable containing the *t*-statistic corresponding to the test that the combined weight is different from 1.0 . This variable is missing for all _TYPE_ = LOM_STD and _TYPE_ = ANOVA.

- PROBT1, a numeric variable containing the significance level for *t*-statistic T1. The variable is missing for all _TYPE_ = LOM_STD and _TYPE_ = ANOVA.

- PROBT2, a numeric variable containing the significance level for *t*-statistic T2. The variable is missing for all _TYPE_ = LOM_STD and _TYPE_ = ANOVA .

1192

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

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

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

- F, a numeric variable containing the F statistic for the Regression source term. The variable is missing for the source terms Total and Error, and for all _TYPE_ = REGRESS and LOM_STD.

- PROBF, a numeric variable containing the significance level for the F statistic. This variable is missing for the source term Total and Error and for all _TYPE_ = REGRESS and LOM_STD.

Printed Output

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

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

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

Table 21.3 describes the individual tables and charts. Most tables apply both to quarterly and monthly series. Those that apply only to a monthly time series are indicated by an "M" in the notes section, while "P" indicates the table is not a time series, and is only printed, not output to the OUT= data set.

1193

Table 21.3. Table Names and Descriptions

Table	Description	Notes
A1	original series	M
A2	prior monthly adjustment factors	M
A3	original series adjusted for prior monthly factors	M
A4	prior trading-day adjustments	M
A5	prior adjusted or original series	M
A13	ARIMA forecasts	
A14	ARIMA backcasts	
A15	prior adjusted or original series extended by arima backcasts, forecasts	
B1	prior adjusted or original series	
B2	trend cycle	
B3	unmodified seasonal-irregular (S-I) ratios	
B4	replacement values for extreme S-I ratios	
B5	seasonal factors	
B6	seasonally adjusted series	
B7	trend cycle	
B8	unmodified S-I ratios	
B9	replacement values for extreme S-I ratios	
B10	seasonal factors	
B11	seasonally adjusted series	
B13	irregular series	
B14	extreme irregular values excluded from trading-day regression	M
B15	preliminary trading-day regression	M,P
B16	trading-day adjustment factors	M
B17	preliminary weights for irregular components	
B18	trading-day factors derived from combined daily weights	M
B19	original series adjusted for trading-day and prior variation	M
C1	original series modified by preliminary weights and adjusted for trading-day and prior variation	
C2	trend cycle	
C4	modified S-I ratios	
C5	seasonal factors	
C6	seasonally adjusted series	
C7	trend cycle	
C9	modified S-I ratios	
C10	seasonal factors	
C11	seasonally adjusted series	
C13	irregular series	
C14	extreme irregular values excluded from trading-day regression	M
C15	final trading-day regression	M,P
C16	final trading-day adjustment factors derived from regression coefficients	M
C17	final weight for irregular components	
C18	final trading-day factors derived from combined daily weights	M
C19	original series adjusted for trading-day and prior variation	M

1194

Table 21.3. (continued)

Table	Description	Notes
D1	original series modified for final weights and adjusted for trading-day and prior variation	
D2	trend cycle	
D4	modified S-I ratios	
D5	seasonal factors	
D6	seasonally adjusted series	
D7	trend cycle	
D8	final unmodified S-I ratios	
D9	final replacement values for extreme S-I ratios	
D10	final seasonal factors	
D11	final seasonally adjusted series	
D12	final trend cycle	
D13	final irregular series	
E1	original series with outliers replaced	
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	
F1	MCD moving average	
F2	summary measures	P
G1	chart of final seasonally adjusted series and trend cycle	P
G2	chart of S-I ratios with extremes, S-I ratios without extremes, and final seasonal factors	P
G3	chart of S-I ratios with extremes, S-I ratios without extremes, and final seasonal factors in calendar order	P
G4	chart of final irregular and final modified irregular series	P

The PRINTOUT= Option

The PRINTOUT= option controls printing for groups of tables. See the "TABLES Statement" in this chapter for details on specifying individual tables. The following list gives the tables printed for each value of the PRINTOUT= option.

STANDARD (26 tables) A1-A4, B1, C13-C19, D8-D13, E1-E6, F1, F2.

LONG (40 tables) A1-A5, A13-A15, B1, B2, B7, B10, B13-B15, C1, C7, C10, C13-C19, D1, D7-D11, D13, E1-E6, F1, F2.

FULL (62 tables) A1-A5, A13-A15, B1-B11, B13-B19, C1-C11, C13-C19, D1, D2, D4-D12, E1-E6, F1, F2.

The actual number of tables printed depends on the options and statements specified. If a table is not computed, it is not printed. For example, if TDREGR=NONE is specified, none of the tables associated with the trading-day are printed.

1195

The CHARTS= Option

Of the four charts listed in Table 21.3, G1 and G2 are printed by default (CHARTS=STANDARD). Charts G3 and G4 are printed when CHARTS=FULL is specified. See the "TABLES Statement" later in this chapter for details in specifying individual charts.

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

Past releases of PROC X11 printed the "Stable Seasonality Test" after Table D8. Two additional tests have been added and are printed the just after the "Stable Seasonality Test". The motivation, interpretation, and statistical details of all these tests are now given.

Motivation

The seasonal component of this time series, S_t, is defined as the intrayear variation that is repeated constantly (stable) or in an evolving fashion from year to year (moving seasonality).

To determine if stable seasonality if present in a series, PROC X11 computes a one-way analysis of variance using the seasons (months or quarters) as the factor on the Final Unmodified SI Ratios (Table D8). This is the appropriate table to use since the removal of the trend-cycle is equivalent to detrending. PROC X11 prints this test, labeled "Stable Seasonality Test" immediatley after the Table D8. This test has not changed from previous releases.

The X11 seasonal adjustment method allows for slowing evolving seasonality. PROC X11 now computes and prints a test for seasonality when it is evolving or moving. The test is a two-way analysis of variance using months (or quarters) and years. As in the "Stable Seasonality Test", this analysis of variance is performed on the Final Unmodified SI Ratios (Table D8). PROC X11 prints this test, labeled "Moving Seasonality Test" after the "Stable Seasonality Test".

The final new test that PROC X11 computes is a combined or joint test of both stable and moving seasonality. This test combines the two F-tests previously described, along with the Kruskal-Wallis Chi-squared test for the stable seasonality to determine "identifiable" seasonality. This test, labeled "Combined Test for the Presence of Identifiable Seasonality", is printed after the "Moving Seasonality Test".

Interpretation and Statistical Details

The "Stable Seasonality Test" is a one-way analysis of variance on the "Final Unmodified SI Ratios" with seasons (months or quarters) as the factor.

To determine if stable seasonality if present in a series, PROC X11 computes a one-way analysis of variance using the seasons (months or quarters) as the factor on the Final Unmodified SI Ratios (Table D8). This is the appropriate table to use since the removal of the similar to detrending.

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

1196

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

The F-test for moving seasonality is performed by a two-way analysis of variance. The two factors are seasons (months or quarters) and years. The years effect is tested separately; the null hypothesis is no effect due to years after accounting for variation due to months or quarters.)

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

The test for identifiable seasonality is performed by combining the F-tests for stable and moving seasonality, along with a Kruskal-Wallis test for stable seasonality. The description below is based on Dagum, (1980); for further details, see Lothian and Morry, 1978b.

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.

1) If the null hypothesis in the moving seasonality tests is not rejected at the 0.10 % level (one thousandths percent), the seasonality is not identifiable.

2) If the null hypothesis in 1) is rejected, but the moving seasonality null hypothesis is not rejected at the 5.0% level, then compute the following quantities:

$$T_1 = \frac{7}{F_m - F_s},$$

$$T_2 = \frac{3F_m}{F_s}.$$

Let T denote the simple average of T_1 and T_2:

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

if $T \geq 1.0$, the null hypothesis of identifiable seasonality *not* present is accepted.

3) If the moving seasonality f-test based on F_M passes, but one of the two statistics based on the T's fails, or the Kruskal-Wallis Chi-squared test fails at the 1% level, the then PROC X11 prints "Identifiable Seasonality Probably Present"

4) If the F_S, F_M and the Kruskal-Wallis Chi-squared test pass, then the null hypothesis (of identifiable seasonality *not* present if rejected, and PROC X11 prints "Identifiable Seasonality Present".

Tables Written to the OUT= data set

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

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

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

Printed Output Generated by Sliding Spans Analysis

Table S 0.A

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

Table S 0.B

Table S 0.B gives the summary of the two f-tests performed during the standard X11 seasonal adjustments for stable and moving seasonality on table D8, the final SI ratios. These tests are described in "Printed Output" in the "PROC X11" chapter.

Table S 1.A

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

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

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

$$MPD_j = \frac{max_{k=1,..,4}\bar{S}_j(k) - min_{k=1,..,4}\bar{S}_j(k)}{min_{k=1,..,4}\bar{S}_j(k)}$$

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

Table S 1.B

Table S 1.B gives a summary of range measures for each span. The first column, Range Means, is calculated by computing the maximum and minimum over all months or quarters in a span, the taking the difference. The next column is the range ratio means, which is simply the ratio of the previously described maximum and min-

imum. The next two columns are the minimum and maximum seasonal factors over the entire span, while the range sf column is the difference of these. Finally, the last column is the ratio of the Max SF and Min SF columns.

Breakdown Tables

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

Table S 2.A.1

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

Table S 2.A.2

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

Table S 2.A.3

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

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

Consider Table D10, the seasonal factors in a sliding spans analysis that uses 4 spans. For a given observation t, looking across the 4 spans, we can easily pick out large differences if they occur. More subtle differences can occur when estimates go from above to below (or vice versa) a base level. In the case of multiplicative model, the seasonal factors have a base level of 100.0. So it is useful to enumerate those instances where both a large change occurs (an MPD value exceeding 3.0%) and a change of sign (with respect to the base) occur.

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

Another behavior of interest occurs when component estimates increase then decrease (or vice versa) across spans for a given observation. Using the example above, the seasonal factors at observation t could first increase, then decrease across the 4 spans.

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

Again, consider Table D10, the seasonal factors in a sliding spans analysis that uses 4 spans. For a given observation t, (containing at least three spans), note the level of D10 for the first span. Continue across the spans until a difference of 1.0% or greater occurs (or no more spans are left), noting whether the difference is up or down. If the difference is up, continue until a difference of 1.0% or greater occurs downward (or no more spans are left). If such an up-down combination occurs, the observation is counted as an up-down turning point. A similar description occurs for a down-up turning point. Tables S 7.A-S 7.E, described below, show the occurrence of turning points, indicating whether up-down or down-up. Note that it requires at least three spans to test for a turning point. Hence Tables S 2.A.3 - S 6.A.3 show a reduced number in the "Turning Point" row for the "Total Tested" column, and in Tables S 7.A - S 7.E, the turning points symbols can only occur where three or more spans overlap.

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

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

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

The footnote to Table S 2.A.3 gives the Census Bureau recommendation for thresholds, as described in "Computational Details" earlier in this section.

Table S 2.B

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

Table S 2.C

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

Tables S 3.A.1 - S 3.A.3

These table relate to the Trading Day Factors (Table C18), and follow the same format as Tables S 2.A.1-S 2.A.3. The only difference between these tables and S 2.A.1-S 2.A.3 is the default cutoff value of 2.0% instead of the 3.0% used for the Seasonal Factors.

Tables S 3.B, S 3.C

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

Tables S 4.A.1 - S 4.A.3

These table relate to the Seasonally Adjusted Series (Table D11), and follow the same format as Tables S 2.A.1-S 2.A.3. The same default cutoff value of 3.0% is used.

Tables S 4.B, S 4.C

These tables, applied to the Seasonally Adjusted Series (Table D11) are the same format as tables S 2.B - S 2.C.

Tables S 5.A.1 - S 5.A.3

These table relate to the Month-to-Month (or Quarterly-to-Quarterly) differences in the Seasonally Adjusted Series, and follow the same format as Tables S 2.A.1-S 2.A.3. The same default cutoff value of 3.0% is used.

Tables S 5.B, S 5.C

These tables, applied to the Month-to-Month (or Quarterly-to-Quarterly) differences in the Seasonally Adjusted Series, are the same format as tables S 2.B - S 2.C. The same default cutoff value of 3.0% is used.

Tables S 6.A.1 - S 6.A.3

These table relate to the Year-to-Year differences in the Seasonally Adjusted Series, and follow the same format as Tables S 2.A.1-S 2.A.3. The same default cutoff value of 3.0% is used.

Tables S 6.B, S 6.C

These tables, applied to the Year-to-Year differences in the Seasonally Adjusted Series, are the same format as tables S 2.B - S 2.C. The same default cutoff value of 3.0% is used.

Table S 7.A

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

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

Following the last span is the "Sign Change" column. As explained in the description of Table S 2.A.3, a sign change occurs at a given observation when the seasonal factor estimates go from above to below, or below to above, a base level. For the seasonal

1201

factors, 100.0 is the base level for the multiplicative model, 0.0 for the additive model. A blank value indicates no sign change, a "U" indicates a movement "upwards" from the base level and a "D" indicates a movement "downwards" from the base level.

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

The next column is the maximum percent difference (MPD). This quantity, described in "Computational Details" above, is the main computation for sliding spans analysis. A measure of how extreme the MPD value is given in the last column, the "Level of Excess" column. The symbols used and their meaning is described in Table S 2.A.3. If a given observation has exceeded the cutoff, the level of excess column is blank.

Table S 7.B

Table S 7.B gives the entire listing of the Trading Day Factors (Table C18) for each span. The format of this table is exactly like Table S 7.A.

Table S 7.C

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

Table S 7.D

Table S 7.D gives the entire listing of the Month-to-Month (or Quarter-to-Quarter) Changes in Seasonally Adjusted Data for each span. The format of this table is exactly like Table S 7.A.

Table S 7.E

Table S 7.E gives the entire listing of the Year-to-Year Changes in Seasonally Adjusted Data for each span. The format of this table is exactly like Table S 7.A.

Printed Output from the ARIMA Statement

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

If the PRINTALL option is specified, a summary of the Nonlinear Estimation Optimization and a table of Box-Ljung Statistics is also produced. If the automatic model selection is used, this information is printed for each of the five predefined models. Lastly, a Model Selection Summary is printed, showing the final model chosen.

ODS Table Names

PROC X11 assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 6, "Using the Output Delivery System."

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

Table 21.4. ODS Tables Produced in PROC X11

ODS Table Name	Description	Option
ODS Tables Created by the MONTHLY and QUARTERLY Statements		
Preface	X11 Seasonal Adjustment Program Information giving credits, dates, etc.	Always printed unless NOPRINT
A1	Table A1: OriginalSeries	
A2	Table A2: Prior Monthly	
A3	Table A3: Original Series Adjusted for Prior Monthly Factors	
A4	Table A4: Prior Trading Day Adjustment Factors With and Without Length of Month Adjustment	
A5	Table A5: Original Series Adjusted for Priors	
B1	Table B1: Original Series or Original Series Adjusted for Priors	
B2	Table B2: Trend Cycle - Centered nn-Term Moving Average	
B3	Table B3: Unmodified SI Ratios	
B4	Table B4: Replacement Values for Extreme SI Ratios	
B5	Table B5: Seasonal Factors	
B6	Table B6: Seasonally Adjusted Series	
B7	Table B7: Trend Cycle - Henderson Curve	
B8	Table B8: Unmodified SI Ratios	
B9	Table B9: Replacement Values for Extreme SI Ratios	
B10	Table B10: Seasonal Factors	
B11	Table B11: Seasonally Adjusted Series	
B13	Table B13: Irregular Series	

Table 21.4. (continued)

ODS Table Name	Description	Option
B15	Table B15: Preliminary Trading Day Regression	
B16	Table B16: Trading Day Adjustment Factors Derived from Regression	
B17	Table B17: Preliminary Weights for Irregular Component	
B18	Table B18: Trading Day Adjustment Factors from Combined Weights	
B19	Table B19: Original Series Adjusted for Preliminary Comb. TD. Wgts.	
C1	Table C1: Original Series Adjusted for Preliminary Weights	
C2	Table C2: Trend Cycle - Centered nn-Term Moving Average	
C4	Table C4: Modified SI Ratios	
C5	Table C5: Seasonal Factors	
C6	Table C6: Seasonally Adjusted Series	
C7	Table C7 Trend Cycle - Henderson Curve	
C9	Table C9: Modified SI Ratios	
C10	Table C10: Seasonal Factors	
C11	Table C11: Seasonally Adjusted Series	
C13	Table C13: Irregular Series	
C15	Table C15: Final Trading Day Regression	
C16	Table C16: Trading Day Adjustment Factors Derived from Regression	
C17	Table C17: Final Weights for Irregular Component	
C18	Table C18: Trading Day Adjustment Factors from Combined Weights	
C19	Table C19: Original Series Adjusted for Final Comb. TD. Wgts.	
D1	Table D1: Original Series Adjusted for Final Weights nn-Term Moving Average	
D4	Table D4: Modified SI Ratios	
D5	Table D5: Seasonal Factors	
D6	Table D6: Seasonally Adjusted Series	
D7	Table D7: Trend Cycle - Henderson Curve	
D8	Table D8: Final Unmodified SI Ratios	
D10	Table D10: Final Seasonal Factors	
D11	Table D11: Final Seasonally Adjusted Series	
D12	Table D12: Final Trend Cycle - Henderson Curve	
D13	Table D13: Final Irregular Series	

Table 21.4. (continued)

ODS Table Name	Description	Option
E1	Table E1: Original Series Modified for Extremes	
E2	Table E2: Modified Seasonally Adjusted Series	
E3	Table E3: Modified Irregular Series	
E5	Table E5: Month-to-Month Changes in Original Series	
E6	Table E6: Month-to-Month Changes in Final Seasonally Adj. Series	
F1	Table F1: MCD Moving Average	
A13	Table A13: ARIMA Forecasts	ARIMA statement
A14	Table A14: ARIMA Backcasts	ARIMA statement
A15	Table A15: Arima Extrapolation	ARIMA statement
B14	Table B14: Irregular Values Excluded from Trading Day Regression	
C14	Table C14: Irregular Values Excluded from Trading Day Regression	
D9	Table D9: Final Replacement Values	
PriorDailyWgts	Adjusted Prior Daily Weights	
TDR_0	Final/ Preliminary Trading Day Regression, part 1	MONTHLY only, TDREGR=ADJUST, TEST
TDR_1	Final/ Preliminary Trading Day Regression, part 2	MONTHLY only, TDREGR=ADJUST, TEST
StandErrors	Standard Errors of Trading Day Adjustment Factors	MONTHLY only, TDREGR=ADJUST, TEST
D9A	Year to Year Change in Irregular and Seasonal Components And Moving Seasonality Ratio	
StableSeasTest	Stable Seasonality Test	MONTHLY only
StableSeasFTest	Stable Seasonality Test	MONTHLY only
f2a	F2 Summary Measures, part 1	
f2b	F2 Summary Measures, part 2	
f2c	F2 Summary Measures, part 3	
f2d	I/C Ratio for Month/Quarterly Span	

Table 21.4. (continued)

ODS Table Name	Description	Option
f2f	Avg % Change with regard to Sign and Std. Over Span	
E4	Differences or Ratios of Annual Totals, Original and Adjusted Series	
ChartG1	Chart G1	
ChartG2	Chart G2	

ODS Tables Created by the ARIMA Statement

CriteriaSummary	Criteria Summary	ARIMA statement
ConvergeSummary	Convergence Summary	
ArimaEst	Arima estimation results, part 1	
ArimaEst2	Arima estimation results, part 2	
Model_Summary	Model Summary	
Ljung_BoxQ	Table of Ljung-Box Q Statistics	
A13	Table A13: ARIMA Forecasts	
A14	Table A14: ARIMA Backcasts	
A15	Table A15: Arima Extrapolation	

ODS Tables Created by the SSPAN Statement

SPR0A_1	S 0.A Sliding Spans Analysis, Number, Length of Spans	default printing
SpanDates	S 0.A Sliding Spans Analysis: Dates of Spans	
SPR0B	S 0.B Summary of F-tests for Stable and Moving Seasonality	
SPR1_1	S 1.A Range Analysis of Seasonal Factors	
SPR1_b	S 1.B Summary of Range Measures	
SPRXA	2XA.1 Breakdown of Differences by Month or Qtr	
SPRXB_2	S X.B Histogram of Flagged Observations	
SPRXA_2	S X.A.2 Breakdown of Differences by Year	
MpdStats	S X.C: Statistics for Maximum Percentage Differences	
S_X_A_3	S 2.X.3 Breakdown Summary of Flagged Observations	
SPR7_X	S 7.X Sliding Spans Analysis	PRINTALL

Examples

Example 21.1. Component Estimation - Monthly Data

This example computes and plots the final estimates of the individual components for a monthly series. In the first plot, Output 21.1.1 an overlaid plot of the original and seasonally adjusted data is produced. The trend in the data is more evident in the seasonally adjusted data than in the original data. This trend is even more clear in Output 21.1.3, the plot of Table D12, the trend cycle. Note that both the seasonal factors and the irregular factors vary around 100, while the trend cycle and the seasonally adjusted data are in the scale of the original data.

From Output 21.1.2 the seasonal component appears to be slowly increasing, while no apparent pattern exists for the irregular series in Output 21.1.4.

```
data sales;
   input sales @@;
   date = intnx( 'month', '01sep1978'd, _n_-1 );
   format date monyy7.;
   datalines;
112 118 132 129 121 135 148 148 136 119 104 118
115 126 141 135 125 149 170 170 158 133 114 140
145 150 178 163 172 178 199 199 184 162 146 166
171 180 193 181 183 218 230 242 209 191 172 194
196 196 236 235 229 243 264 272 237 211 180 201
204 188 235 227 234 264 302 293 259 229 203 229
242 233 267 269 270 315 364 347 312 274 237 278
284 277 317 313 318 374 413 405 355 306 271 306
315 301 356 348 355 422 465 467 404 347 305 336
340 318 362 348 363 435 491 505 404 359 310 337
360 342 406 396 420 472 548 559 463 407 362 405
417 391 419 461 472 535 622 606 508 461 390 432
run;

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

symbol1 i=join v='star';
symbol2 i=join v='circle';
legend1 label=none value=('original' 'adjusted');

proc gplot data=out;
   plot series * date = 1 d11 * date = 2
                       / overlay legend=lengend1;
run;

symbol1 i=join v=dot;
```
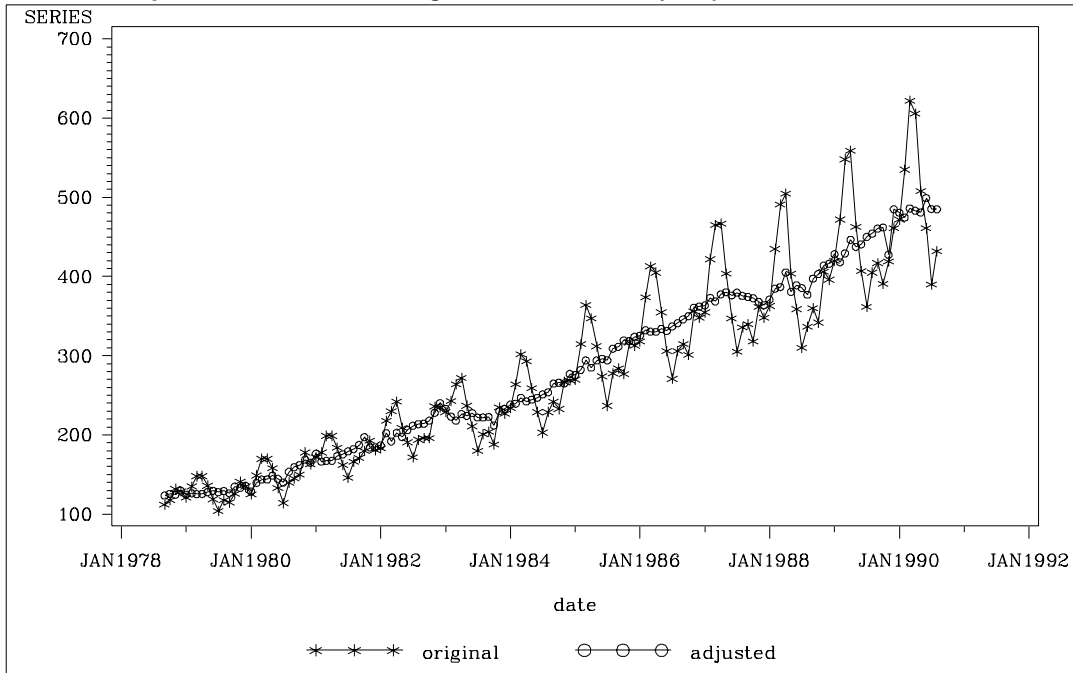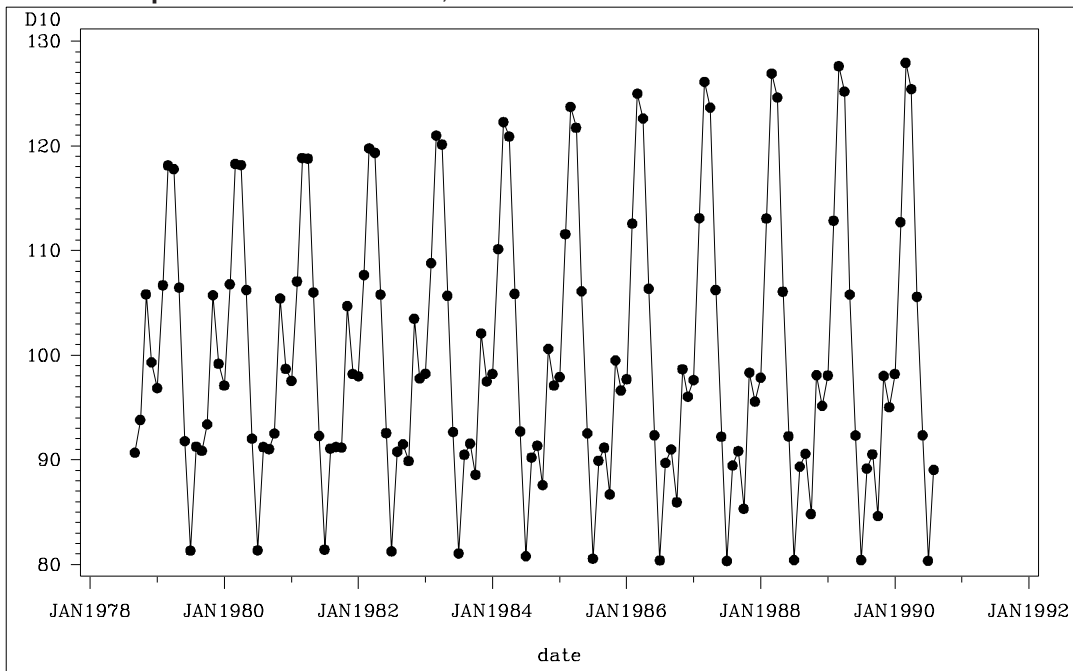
1207

```
proc gplot data=out;
   plot ( d10 d12 d13 ) * date;
run;
```
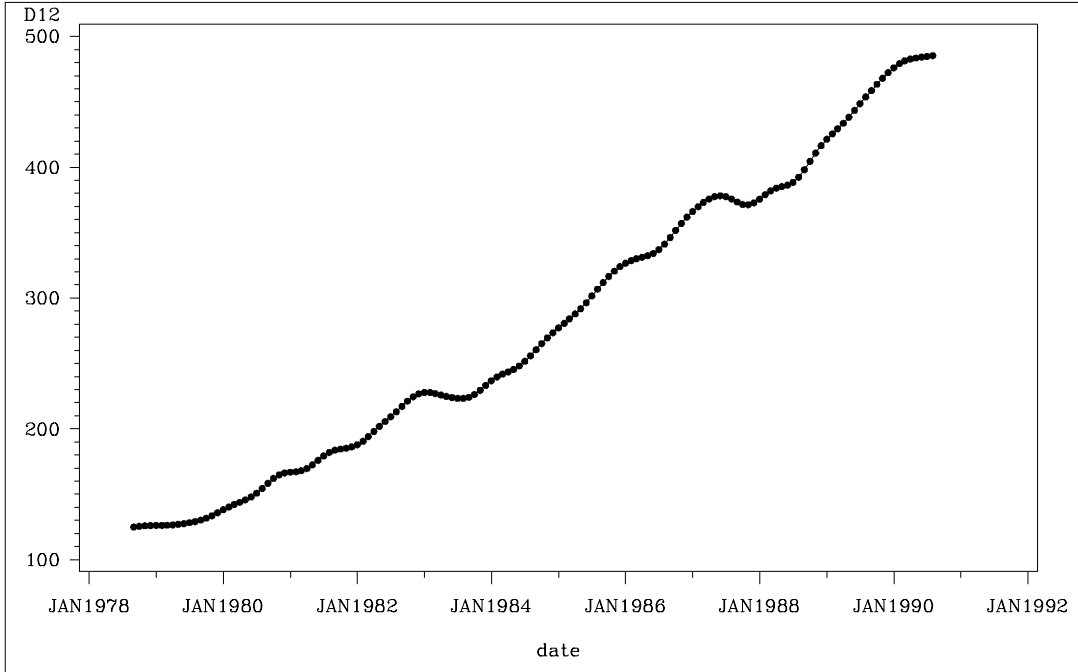
Output 21.1.1.   Plot of Original and Seasonally Adjusted Data
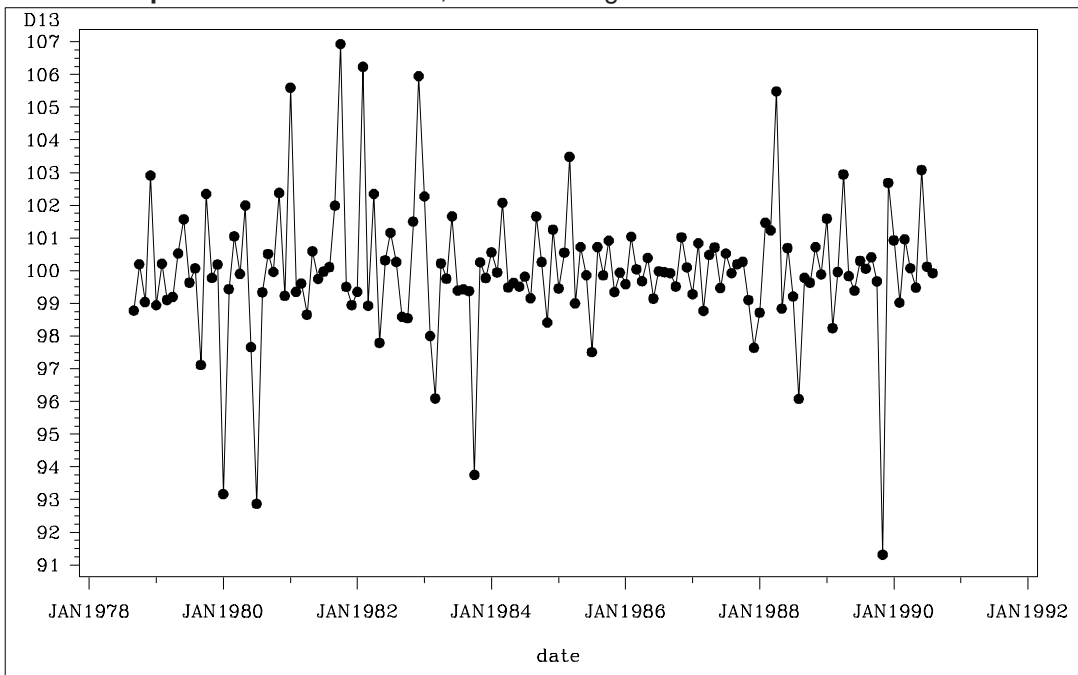

Output 21.1.2.   Plot of D10, the Final Seasonal Factors

1208

Output 21.1.3. Plot of D12, the Final Trend Cycle

Output 21.1.4. Plot of D13, the Final Irregular Series

Example 21.2. Components Estimation - Quarterly Data

This example is similar to Example 21.1, except quarterly data is used. Tables B1, the original series, and D11, the final seasonally adjusted series, are printed by the TABLES statement. The OUTPUT statement writes the listed tables to an output data set.

```
data quarter;
   input date yyq6. +1 fy35rr 5.2;
   format date yyq6.;
datalines;
1971Q1 6.59
1971Q2 6.01
1971Q3 6.51
1971Q4 6.18
1972Q1 5.52
1972Q2 5.59
1972Q3 5.84
1972Q4 6.33
1973Q1 6.52
1973Q2 7.35
1973Q3 9.24
1973Q4 10.08
1974Q1 9.91
1974Q2 11.15
1974Q3 12.40
1974Q4 11.64
1975Q1 9.94
1975Q2 8.16
1975Q3 8.22
1975Q4 8.29
1976Q1 7.54
1976Q2 7.44
1976Q3 7.80
1976Q4 7.28
run;

proc x11 data=quarter;
   var fy35rr;
   quarterly date=date;
   tables b1 d11;
   output out=out b1=b1 d10=d10 d11=d11 d12=d12 d13=d13;
run;
```

1210

Output 21.2.1. Printed Output of PROC X11 Quarterly Example

```
                        The X11 Procedure

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

    The X-11 program is divided into seven major parts.
     Part         Description
      A.  Prior adjustments, if any
      B.  Preliminary estimates of irregular component weights
             and regression trading day factors
      C.  Final estimates of above
      D.  Final estimates of seasonal, trend-cycle and
             irregular components
      E.  Analytical tables
      F.  Summary measures
      G.  Charts


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




                        The X11 Procedure

                  Seasonal Adjustment of - fy35rr

                     B1 Original Series
      Year      1st       2nd       3rd       4th      Total

      1971     6.590     6.010     6.510     6.180     25.290
      1972     5.520     5.590     5.840     6.330     23.280
      1973     6.520     7.350     9.240    10.080     33.190
      1974     9.910    11.150    12.400    11.640     45.100
      1975     9.940     8.160     8.220     8.290     34.610
      1976     7.540     7.440     7.800     7.280     30.060
      ---------------------------------------------------------
      Avg      7.670     7.617     8.335     8.300

         Total:  191.53  Mean:  7.9804  S.D.:  1.9424
```

```
                        The X11 Procedure

                  Seasonal Adjustment of - fy35rr

                  D11 Final Seasonally Adjusted Series
      Year      1st       2nd       3rd       4th      Total

      1971     6.877     6.272     6.222     5.956     25.326
      1972     5.762     5.836     5.583     6.089     23.271
      1973     6.820     7.669     8.840     9.681     33.009
      1974    10.370    11.655    11.855    11.160     45.040
      1975    10.418     8.534     7.853     7.947     34.752
      1976     7.901     7.793     7.444     6.979     30.116
      ---------------------------------------------------------
      Avg      8.025     7.960     7.966     7.969

         Total:  191.51  Mean:  7.9797  S.D.:  1.9059
```

Example 21.3. Outlier Detection and Removal

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

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

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

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

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

A plot of the original and modified series, Output 21.3.2, shows that the deviation from the trend line for the modified series is greatly reduced compared with the original series.

```
data a;
   retain seed 99831;
   do kk = 1 to 48;
      x = kk + 100 + rannor( seed );
      date = intnx( 'month', '01jan1970'd, kk-1 );
      if kk = 20 then x = 2 * x;
      else if kk = 30 then x = x / 10;
      output;
      end;
run;

proc x11 data=a;
   monthly date=date additive
           fullweight=3.0 zeroweight=3.5;
   var x;
   table d9;
   output out=b b1=original e1=e1;
run;

symbol1 i=join v=star;
symbol2 i=join v=circle;
legend1 label=none value=('unmodified' 'modified');

proc gplot data= b;
```

1212

```
      plot original * date = 1 e1 * date = 2 / overlay legend=legend1;
      format date monyy7.;
   run;
```

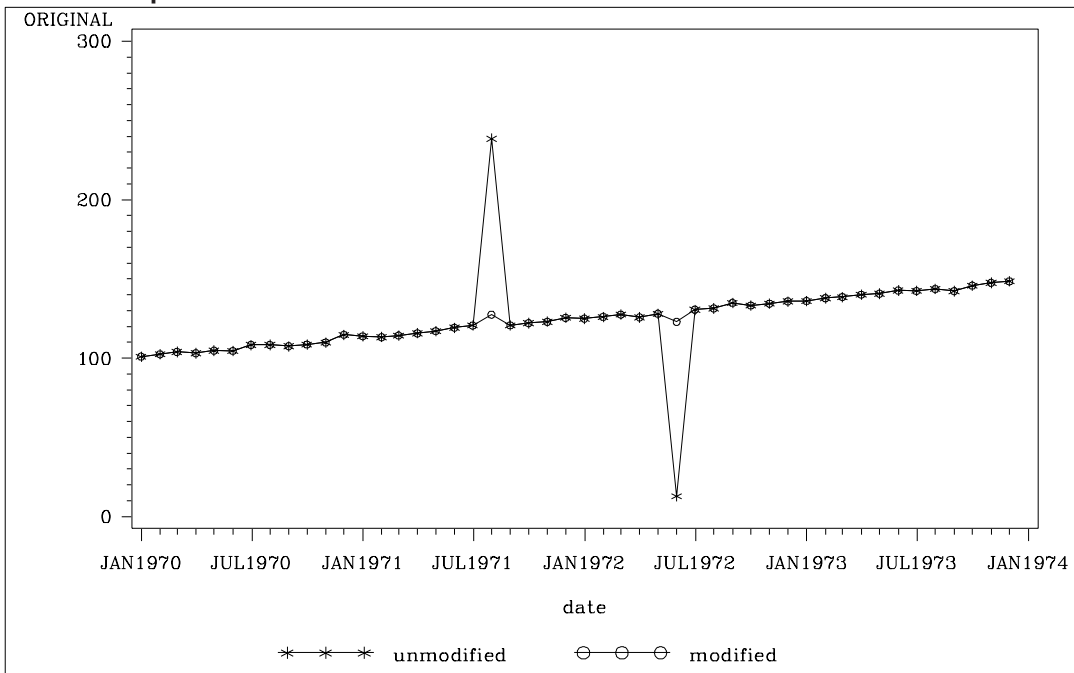**Output 21.3.1.** Detection of Extreme Irregulars

```
                          The X11 Procedure

                     Seasonal Adjustment of - x

            D9 Final Replacement Values for Extreme SI Ratios
    Year        JAN          FEB          MAR          APR          MAY          JUN

    1970         .            .            .            .            .            .
    1971         .            .            .            .            .            .
    1972         .            .            .            .            .         -10.671
    1973         .            .            .            .            .            .

            D9 Final Replacement Values for Extreme SI Ratios
    Year        JUL          AUG          SEP          OCT          NOV          DEC

    1970         .            .            .            .            .            .
    1971         .         11.180          .            .            .            .
    1972         .            .            .            .            .            .
    1973         .            .            .            .            .            .
```

Output 21.3.2. Plot of Modified and Unmodified Values

References

Bell, W.R., and Hillmer, S.C., (1984), "Issues Involved With the Seasonal Adjustment of Economic Time Series," *Journal of Business and Economic Statistics*, Vol. 2, No. 4.

Bobbit, L.G., and Otto, M.C., (1990), "Effects of Forecasts on the Revisions of Seasonally Adjusted Data Using the X-11 Adjustment Procedure," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 449-453.

Buszuwski, J.A., (1987), "Alternative ARIMA Forecasting Horizons When Seasonally Adjusting Producer Price Data with X-11-ARIMA," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 488-493.

Cleveland, W.S. and Devlin, S.J., (1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods," *Journal of the American Statistical Association*, Vol. 75, No. 371, pp. 487-496.

Cleveland, W.P. and Tiao, G.C. (1976), "Decomposition of Seasonal Time Series: A Model for Census X-11 Program," *Journal of the American Statistical Association*, Vol. 71, No. 355.

Dagum, E.B. (1980), *The X-11-ARIMA Seasonal Adjustment Method*, Statistics Canada.

Dagum, E.B. (1982a), "Revisions of Time Varying Seasonal Filters," *Journal of Forecasting*, Vol. 1, pp. 173-187.

Dagum, E.B. (1982b), "The Effects of Asymmetric Filters on Seasonal Factor Revision," *Journal of the American Statistical Association*, Vol. 77, No. 380, pp. 732-738.

Dagum, E.B. (1982c), "Revisions of Seasonally Adjusted Data Due to Filter Changes," *Proceedings of the Business and Economic Section, the American Statistical Association*, pp. 39-45.

Dagum, E.S. (1983), "The X-11-ARIMA Seasonal Adjustment Method," Ottawa: Statistics Canada

Dagum, E.S. (1985), "Moving Averages," in *Encyclopedia of Statistical Sciences*, Vol. 5, eds. S. Kotz and N. L. Johnson, New York: John Wiley & Sons, Inc.

Dagum, E.S. and Laniel, N. (1987), "Revisions of Trend Cycle Estimators of Moving Average Seasonal Adjustment Method," *Journal of Business and Economic Statistics*, Vol. 5, No. 2, pp. 177-189.

Dagum, E.B. (1988), *The X11ARIMA/88 Seasonal Adjustment Method*, Statistics Canada.

Davies, N., Triggs, C.M., and Newbold, P. (1977), "Significance Levels of the Box-Pierce Portmanteau Statistic in Finite Samples," *Biometrika*, 64, 517-522.

Findley, D.F., Monsell, B.C., Shulman, H.B., and Pugh, M.G., (1990), "Sliding Spans Diagnostics for Seasonal and Related Adjustments," *Journal of the American Statistical Association*, Vol 85, No. 410, pps 345-355

Findley, D.F., Monsell, B.C., (1986) "New Techniques for Determining if a Time Series can be Seasonally Adjusted Reliably, and Their Application to U. S. Foreign Trade Series," in *Regional Econometric Modeling*, eds. M.R. Perryman and J. R. Schmidt, Amsterdam: Kluwer-Nijhoff, pp. 195-228.

Ghysels, E. (1990), "Unit Root Tests and the Statistical Pitfalls of Seasonal Adjustment: The Case of U.S. Post War Real GNP," *Journal of Business and Economic Statistics*, Vol. 8, No. 2, pp. 145-152.

Hannan, E.J., (1963), "The Estimation of Seasonal Variation in Economic Time Series," *Journal of the American Statistical Association*, Vol. 58, No. 301.

Huot, G., Chui, L., Higginson, J., and Gait, N., (1986), "Analysis of Revisions in the Seasonal Adjustment of Data Using X11ARIMA Model-Based Filters," International Journal of Forecasting, Vol. 2, pp. 217-229.

Laniel, N. (1985) "Design Criteria for the 13-term Henderson End-Weights," Working Paper, Methodology Branch, Ottawa: Statistics Canada

Ljung, G.M. and Box, G.E.P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297-303.

Lothian, J. and Morry M., (1978a), "Selection of Models for the Automated X-11-ARIMA Seasonal Adjustment Program," Statistics Canada.

Lothian, J. and Morry M., (1978b), "A Test for the Presence of Identifiable Seasonality when using the X-11 Program" Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.

Marris, S.N., (1960), "The Treatment of Moving Seasonality in Census Method II," *Seasonal Adjustments on Electronic Computers*, Organization for Economic Cooperation and Development.

Monsell, B.C., (1984), "The Substantive Changes in the X-11 Procedure of X-11-ARIMA," Bureau of the Census, Statistical Research Division, SRD Research Report Number Census/SRD/RR-84/10.

Nettheim, N.F., (1965), "A Spectral Study of Overadjustment for Seasonality," *1964 Proceedings of the Business and Economics Section; American Statistical Association*; republished as the Bureau of the Census Working Paper No. 21, 1965.

Pierce, D.A., (1980), "Data Revisions with Moving Average Seasonal Adjustment Procedures," *Journal of Econometrics*, 14, 95-114.

Shiskin, J., and Eisenpress, H., (1957), "Seasonal Adjustment by Electronic Computer Methods," JASA, Vol. 52, No. 280.

Shiskin, J., (1958), "Decomposition of Economic Time Series," Science, Vol. 128, No. 3338.

U.S. Bureau of the Census, (1965). "Estimating Trading Day Variation in Monthly Economic Time Series," Technical Paper No. 12. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census, (1967). "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15, 1967 revision. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census (1969), *X-11 Information for the User*, U.S. Department of Commerce, Washington, DC: Government Printing Office.

SAS/ETS User's Guide, Version 8