

# **SAS/ETS<sup>®</sup> 14.2 User's Guide**

## **The X11 Procedure**

This document is an individual chapter from *SAS/ETS® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/ETS® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

#### **SAS/ETS® 14.2 User's Guide**

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

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

November 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to <http://support.sas.com/thirdpartylicenses>.

# Chapter 43

## The X11 Procedure

### Contents

---

Overview: X11 Procedure . . . . .	<b>3182</b>
Getting Started: X11 Procedure . . . . .	<b>3182</b>
Basic Seasonal Adjustment . . . . .	3183
X-11-ARIMA . . . . .	3186
Syntax: X11 Procedure . . . . .	<b>3187</b>
Functional Summary . . . . .	3188
PROC X11 Statement . . . . .	3189
ARIMA Statement . . . . .	3190
BY Statement . . . . .	3193
ID Statement . . . . .	3193
MACURVES Statement . . . . .	3193
MONTHLY Statement . . . . .	3194
OUTPUT Statement . . . . .	3198
PDWEIGHTS Statement . . . . .	3198
QUARTERLY Statement . . . . .	3199
SSPAN Statement . . . . .	3201
TABLES Statement . . . . .	3202
VAR Statement . . . . .	3202
Details: X11 Procedure . . . . .	<b>3203</b>
Historical Development of X-11 . . . . .	3203
Implementation of the X-11 Seasonal Adjustment Method . . . . .	3204
Computational Details for Sliding Spans Analysis . . . . .	3208
Data Requirements . . . . .	3210
Missing Values . . . . .	3210
Prior Daily Weights and Trading-Day Regression . . . . .	3210
Adjustment for Prior Factors . . . . .	3211
The YRAHEADOUT Option . . . . .	3212
Effect of Backcast and Forecast Length . . . . .	3212
Details of Model Selection . . . . .	3213
OUT= Data Set . . . . .	3216
The OUTSPAN= Data Set . . . . .	3216
OUTSTB= Data Set . . . . .	3216
OUTTDR= Data Set . . . . .	3217
Printed Output . . . . .	3218
ODS Table Names . . . . .	3229
Examples: X11 Procedure . . . . .	<b>3234</b>

Example 43.1: Component Estimation—Monthly Data . . . . .	3234
Example 43.2: Components Estimation—Quarterly Data . . . . .	3238
Example 43.3: Outlier Detection and Removal . . . . .	3240
References . . . . .	<b>3242</b>

---

---

## Overview: X11 Procedure

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

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

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

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

---

## Getting Started: X11 Procedure

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

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

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

The naming convention used in PROC X11 for the tables follows the original U.S. Bureau of the Census X-11 Seasonal Adjustment program specification (Shiskin, Young, and Musgrave 1967). Also, see the section “Printed Output” on page 3218. This convention is outlined in Figure 43.1.

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

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

---

## Basic Seasonal Adjustment

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

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

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

... more lines ...

/---- X-11 ARIMA ----*/
proc x11 data=sales;
    monthly date=date;
    var sales;
    tables d11;
run;
```

**Figure 43.1** Basic Seasonal Adjustment**The X11 Procedure****Seasonal Adjustment of - sales**


---

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

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

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.

Selected Tables or Charts.

Sigma limits for graduating extreme values are 1.5 and 2.5

Irregular values outside of 2.5-sigma limits are excluded from trading day regression

---

**Figure 43.2** Basic Seasonal Adjustment

D11 Final Seasonally Adjusted Series													
Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Total
1978	.	.	.	.	.	.	.	.	123.507	125.776	124.735	129.870	503.887
1979	124.935	126.533	125.282	125.650	127.754	129.648	127.880	129.285	126.562	134.905	133.356	136.117	1547.91
1980	128.734	139.542	143.726	143.854	148.723	144.530	140.120	153.475	159.281	162.128	168.848	165.159	1798.12
1981	176.329	166.264	167.433	167.509	173.573	175.541	179.301	182.254	187.448	197.431	184.341	184.304	2141.73
1982	186.747	202.467	192.024	202.761	197.548	206.344	211.690	213.691	214.204	218.060	228.035	240.347	2513.92
1983	233.109	223.345	218.179	226.389	224.249	227.700	222.045	222.127	222.835	212.227	230.187	232.827	2695.22
1984	238.261	239.698	246.958	242.349	244.665	247.005	251.247	253.805	264.924	266.004	265.366	277.025	3037.31
1985	275.766	282.316	294.169	285.034	294.034	296.114	294.196	309.162	311.539	319.518	318.564	323.921	3604.33
1986	325.471	332.228	330.401	330.282	333.792	331.349	337.095	341.127	346.173	350.183	360.792	362.333	4081.23
1987	363.592	373.118	368.670	377.650	380.316	376.297	379.668	375.607	374.257	372.672	368.135	364.150	4474.13
1988	370.966	384.743	386.833	405.209	380.840	389.132	385.479	377.147	397.404	403.156	413.843	416.142	4710.89
1989	428.276	418.236	429.409	446.467	437.639	440.832	450.103	454.176	460.601	462.029	427.499	485.113	5340.38
1990	480.631	474.669	486.137	483.140	481.111	499.169	485.370	485.103	.	.	.	.	3875.33
Avg	277.735	280.263	282.435	286.358	285.354	288.638	288.683	291.413	265.728	268.674	268.642	276.442	

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

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

```

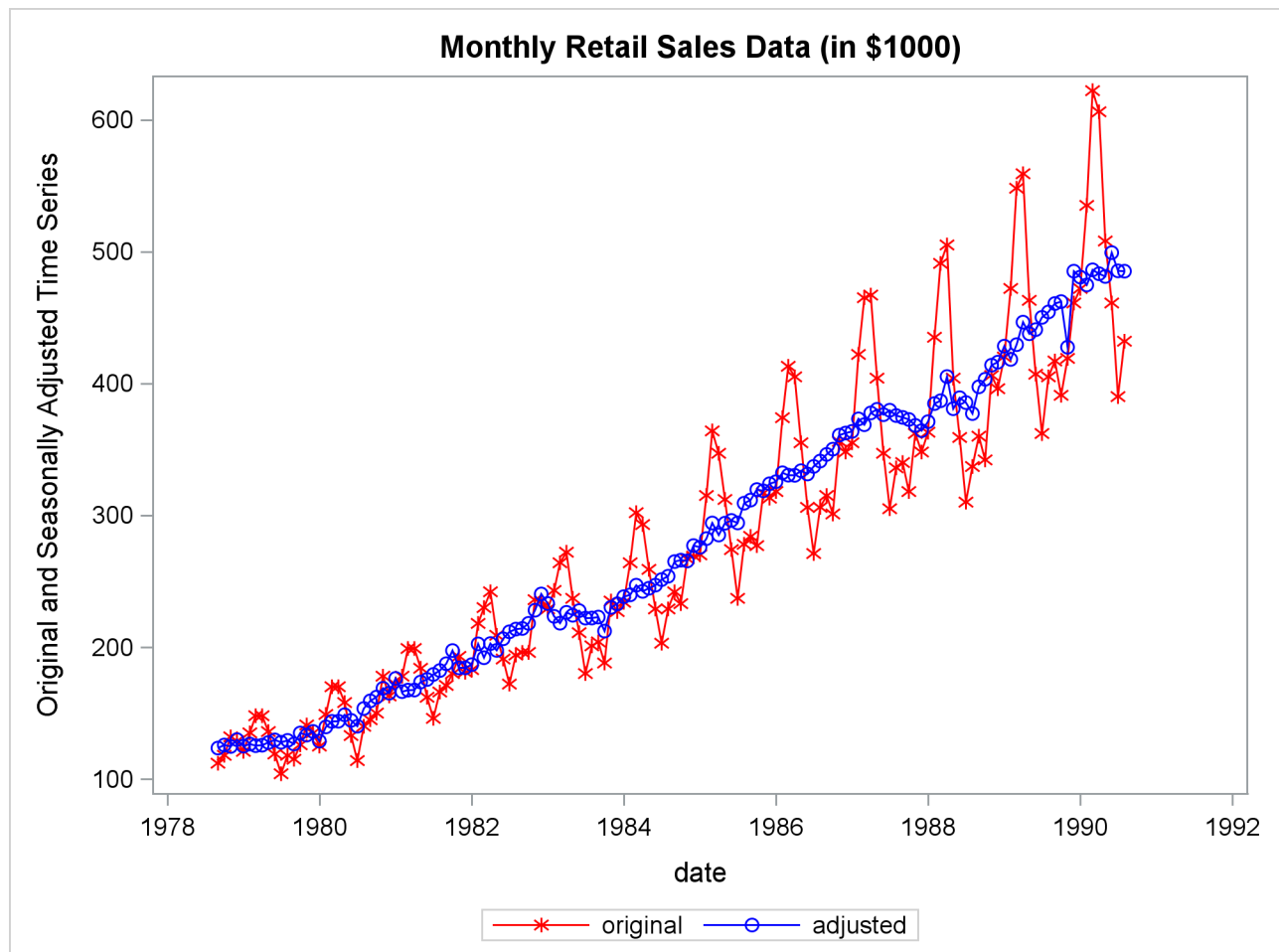
title 'Monthly Retail Sales Data (in $1000)';

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

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

```

**Figure 43.3** Plot of Original and Seasonally Adjusted Data



## X-11-ARIMA

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

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

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

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

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

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

**Figure 43.4** X-11-ARIMA Model Selection  
Monthly Retail Sales Data (in \$1000)

The X11 Procedure				
Seasonal Adjustment of - sales				
Conditional Least Squares Estimation				
Parameter	Estimate	Approx. 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



**Figure 43.4** *continued*

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

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

---

## Syntax: X11 Procedure

The X11 procedure uses the following statements:

```

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 [Table 43.1](#).

**Table 43.1** Functional Summary

Description	Statement	Option
<b>Data Set Options</b>		
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=
<b>Printing Control Options</b>		
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 QUARTERLY	PRINTOUT= PRINTOUT=
Print selected groups of charts	MONTHLY QUARTERLY	CHARTS= CHARTS=
Print preliminary tables associated with ARIMA processing	ARIMA	PRINTFP
Specify number of decimals for printed tables	MONTHLY QUARTERLY	NDEC= NDEC=
Suppress all printed SSPAN output	SSPAN	NOPRINT
Print all SSPAN output	SSPAN	PRINTALL
<b>Date Information Options</b>		
Specify a SAS date variable	MONTHLY QUARTERLY	DATE= DATE=
Specify the beginning date	MONTHLY QUARTERLY	START= START=
Specify the ending date	MONTHLY QUARTERLY	END= END=
Specify beginning year for trading-day regression	MONTHLY	TDCOMPUTE=

**Table 43.1** *continued*

Description	Statement	Option
<b>Declaring the Role of Variables</b>		
Specify BY-group processing	BY	
Specify the variables to be seasonally adjusted	VAR	
Specify identifying variables	ID	
Specify the prior monthly factor	MONTHLY	PMFACTOR=
<b>Controlling the Table Computations</b>		
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=
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	MONTHLY	TRENDADJ
	QUARTERLY	TRENDADJ
Specify moving average length in trend	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 to the ends of the series, and the calculations are carried out on this extended series. The appended observations are not normally written to the OUT= data set. However, if OUTEXTRAP is specified, these extra observations are written to the output data set. If a DATE= variable is specified in the MONTHLY/QUARTERLY statement, the date variable is extrapolated to identify forecasts/backcasts. The OUTEXTRAP option can be abbreviated as OUTEX.

**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. For more information, see the section “[The OUTSPAN= Data Set](#)” on page 3216.

**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. For more information, see the section “[OUTSTB= Data Set](#)” on page 3216.

**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. For more information, see the section “[OUTTDR= Data Set](#)” on page 3217.

**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. For more information, see the section “[The YRAHEADOUT Option](#)” on page 3212.

---

## 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. For more information, see the section “[Details of Model Selection](#)” on page 3213.

**BACKCAST=n**

specifies the number of years to backcast the series. The default is BACKCAST=0. For more information, see the section “[Effect of Backcast and Forecast Length](#)” on page 3212.

**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. For more information about the CHICR= option, see the section “[Criteria Details](#)” on page 3214.

**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. For more information, see the section “[Effect of Backcast and Forecast Length](#)” on page 3212.

**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. For more information about the MAPECR= option, see the section “[Criteria Details](#)” on page 3214.

**MAXITER=n**

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

**METHOD=CLS****METHOD=ULS****METHOD=ML**

specifies the estimation method. ML requests maximum likelihood, ULS requests unconditional least squares, and CLS requests conditional least squares. METHOD=CLS is the default. The maximum likelihood estimates are more expensive to compute than the conditional least squares estimates. In some cases, however, they can be preferable. For further information about the estimation methods, see the section “[Estimation Details](#)” on page 248 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, respectively. The lag corresponding to seasonality is determined by the MONTHLY or QUARTERLY statement. Similarly, differencing and seasonal differencing are given by DIF=n5 and SDIF=n6, respectively.

For example,

```
arima model=( p=2 q=1 sp=1 dif=1 sdif=1 );
```

specifies a (2,1,1)(1,1,0)<sub>s</sub> model, where *s*, the seasonality, is either 12 (monthly) or 4 (quarterly). For more examples of the MODEL= syntax, see the section “[Details of Model Selection](#)” on page 3213.

**NOINT**

suppresses the fitting of a constant (or intercept) parameter in the model. (That is, the parameter  $\mu$  is omitted.)

**CENTER**

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

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

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

**NOPRINT**

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

**OVDIFCR=value**

specifies the criteria for the over-differencing test when testing the five predefined models. When the MA parameters in one of these models sum to a number close to 1.0, this is an indication of over-parameterization and the model is rejected. The 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. For more information about the OVDIFCR= option, see the section “[Criteria Details](#)” on page 3214.

**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. For more information, see the “[Printed Output](#)” on page 3218.

**PRINTFP**

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

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

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

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

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

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

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

## BY Statement

**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 or 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)

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  $3 \times 3$  (a three-term moving average of a three-term moving average) for February, a  $3 \times 5$  (a three-term moving average of a five-term moving average) for March, and a  $3 \times 9$  (a three-term moving average of a nine-term moving average) for April.

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

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

---

## MONTHLY Statement

### **MONTHLY** options ;

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

The following options can appear in the MONTHLY statement:

#### **ADDITIVE**

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

#### **CHARTS=STANDARD**

#### **CHARTS=FULL**

#### **CHARTS=NONE**

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

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

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



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

**DATE=***variable*

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

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

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

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

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

**END=***mmmyyyy*

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

**EXCLUDE=***value*

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

**FULLWEIGHT=***value*

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

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

**LENGTH**

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

**NDEC=*n***

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

**PMFACTOR=*variable***

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

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

For more information and examples using the PMFACTOR= option, see the section “[Prior Daily Weights and Trading-Day Regression](#)” on page 3210.

**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 PRINTOUT=NONE results in no tables being printed; however, charts are still printed. The default is PRINTOUT=STANDARD.

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

**START=*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. For information about using the START= and END= options to subset a series, see the DATE= *variable* option.

**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.

**TD COMPUTE=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.

**TD REGR=NONE | PRINT | ADJUST | TEST**

specifies the treatment of trading-day regression. TD REGR=NONE omits the computation of the trading-day regression. TD REGR=PRINT computes and prints the trading-day regressions but does not adjust the series. TD REGR=ADJUST computes and prints the trading-day regression and adjusts the irregular components to obtain preliminary weights. TD REGR=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 TD REGR=NONE.

For more information and examples using the TD REGR= option, see the section “[Prior Daily Weights and Trading-Day Regression](#)” on page 3210.

If ARIMA processing is requested, any value of TD REGR other than the default TD REGR=NONE will cause PROC X11 to perform an initial pass (see the section “[Details: X11 Procedure](#)” on page 3203 and the PRINTFP option).

The significance level reported in 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 the  $F$  test is computed. Hence the distribution of the trading-day regression  $F$  statistics differs from an exact  $F$ ; for more information, see Cleveland and Devlin (1980).

**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 about the number of terms in the moving average, see Shiskin, Young, and Musgrave (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 value 1.3 standard deviations from the mean would be assigned a graduated weight. The ZEROWEIGHT=value must be between 0.1 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. For an illustration of this use, see [Example 43.3](#) later in this chapter.

## 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 by 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 the section “[Printed Output](#)” on page 3218. The following is an example of a VAR statement and an OUTPUT statement:

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

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

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

## PDWEIGHTS Statement

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

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

For more information and examples using the PDWEIGHTS statement, see the section “[Prior Daily Weights and Trading-Day Regression](#)” on page 3210.

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. For more information, see Shiskin, Young, and Musgrave (1967).

---

## QUARTERLY Statement

### QUARTERLY *options* ;

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

The following options can appear in the QUARTERLY statement:

#### ADDITIVE

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

#### CHARTS=STANDARD

#### CHARTS=FULL

#### CHARTS=NONE

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

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

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

#### DATE=*variable*

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

The DATE= option and the START= and END= options can be used in combination to subset a series for processing. For example, suppose you have a series with 10 years of quarterly data (40 observations,

no missing values) beginning in '1970Q1' and ending in '1979Q4', and you want to seasonally adjust only four years beginning in '1974Q1' and ending in '1977Q4'. Specifying

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

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

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

#### **END='yyyyQq'**

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

#### **FULLWEIGHT=value**

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

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

#### **NDEC=n**

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

#### **PRINTOUT=STANDARD**

#### **PRINTOUT=LONG**

#### **PRINTOUT=FULL**

#### **PRINTOUT=NONE**

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

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

#### **START='yyyyQq'**

adjusts only the part of the input series starting with the quarter and year given. When the DATE= option is not used, the START= option gives the year and quarter of the first input observation (for

example, `START='1967Q1'`). The specification must be enclosed in quotes, and  $q$  must be 1, 2, 3, or 4. `START=` must be specified if the `DATE=` option is not given. If `START=` is specified (and no `DATE=` is given), and an `OUTPUT=` data set is requested, a variable named `_DATE_` is added to the data set, giving the date value for a given observation. For information about using the `START=` and `END=` options to subset a series, see the `DATE=` option.

### 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. For an illustration of this use, see [Example 43.3](#) later in this chapter.

---

## 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-year differences in the seasonally adjusted series. The default value is 3.0. The use of the `CUTOFF=`value in determining the maximum percent difference (MPD) is described in the section “[Computational Details for Sliding Spans Analysis](#)” on page 3208. Caution should be used in changing the default `CUTOFF=`value. The empirical threshold ranges found by the U.S. Census Bureau no longer apply when value is changed.

**TDCUTOFF=value**

gives the percentage value for determining an excessive difference within a span for the trading-day factors. The default value is 2.0. The use of the TDCUTOFF=value in determining the maximum percent difference (MPD) is described in the section “[Computational Details for Sliding Spans Analysis](#)” on page 3208. Caution should be used in changing the default TDCUTOFF=value. The empirical threshold ranges found by the U.S. Census Bureau no longer apply when the value is changed.

**NOPRINT**

suppresses all sliding span reports. For more information about sliding span reports, see the section “[Computational Details for Sliding Spans Analysis](#)” on page 3208.

**PRINT**

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

**PRINTALL**

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

---

## TABLES Statement

**TABLES** *tablenames* ;

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

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

```
tables d10 d11;
```

---

## VAR Statement

**VAR** *variables* ;

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



---

## Details: X11 Procedure

---

### Historical Development of X-11

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

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

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

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

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

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

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

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

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

References for this work during this period include Shiskin and Eisenpress (1957), Shiskin (1958), and Marris (1961). The authoritative documentation for the X-11 Variant is in Shiskin, Young, and Musgrave (1967). This document is not equivalent to a program specification; however, the FORTRAN code that implements

the X-11 Variant is in the public domain. A less detailed description of the X-11 Variant is given in US Bureau of the Census (1969).

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

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

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

The Time Series Research and Analysis Division of Statistics Canada investigated 174 Canadian economic series and found five ARIMA models out of twelve that fit the majority of series well and reduced revisions for the most recent months. References that give details of various aspects of the X-11-ARIMA methodology include Dagum (1980, 1982a, c, 1983, 1988), Laniel (1985), Lothian and Morry (1978a), and Huot et al. (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), with later changes and enhancements made in 1988 (Dagum 1988). The calculations performed by PROC X11 differ from those in X11ARIMA/88, which will result in differences in the final component estimates provided by these implementations.

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

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

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

---

## Implementation of the X-11 Seasonal Adjustment Method

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

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

$$O_t = C_t S_t I_t P_t D_t$$

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

The trading-day component can be further factored as

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

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. For further information about estimating trading day variation, see Young (1965).

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

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

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

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

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

### The Standard X-11 Seasonal Adjustment Method

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

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

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

2. In steps 2, 3, and 4, three iterations are performed, each of which provides estimates of the seasonal  $S_t$ , trading-day  $D_{r,t}$ , 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 have been a number of past attempts to answer this question: stable seasonality  $F$  test; moving seasonality  $F$  test,  $Q$  statistics, and others.

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

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

Sliding spans analysis, developed at the Statistical Research Division of the U.S. Census Bureau (Findley et al. 1990; Findley and Monsell 1986), performs a repeated seasonal adjustment on subsets or spans of the

full series. In particular, an initial span of the data, typically eight years in length, is seasonally adjusted, and the Tables C18, the trading-day factors (if trading-day regression performed), D10, the seasonal factors, and D11, the seasonally adjusted series are retained for further processing. Next, one year of data is deleted from the beginning of the initial span and one year of data is added. This new span is seasonally adjusted as before, with the same tables retained. This process continues until the end of the data is reached. The beginning and ending dates of the spans are such that the last observation in the original data is also the last observation in the last span. This is discussed in more detail in the following paragraphs.

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

- Seasonal Factors (Table D10):  $S_t(k)$
- Trading-Day Factors (Table C18):  $TD_t(k)$
- Seasonally Adjusted Data (Table D11):  $SA_t(k)$
- Month-to-Month Changes in the Seasonally Adjusted Data:  $MM_t(k)$
- Year-to-Year Changes in the Seasonally Adjusted Data:  $YY_t(k)$

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

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

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

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

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

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

$$\text{MPD}_t = \frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)}$$

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

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

## Computational Details for Sliding Spans Analysis

### Length and Number of Spans

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

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

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

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

### Seasonal Factors (Table D10)

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

$$\text{MPD}_t = \frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)}$$

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

### Trading Day Factor (Table C18)

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

$$\text{MPD}_t = \frac{\max_{k \in N_t} TD_t(k) - \min_{k \in N_t} TD_t(k)}{\min_{k \in N_t} TD_t(k)}$$

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

### Seasonally Adjusted Data (Table D11)

For the additive model, the MPD is defined as

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

For the multiplicative model, the MPD is

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

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

### Month-to-Month Changes in the Seasonally Adjusted Data

Some additional notation is needed for the month-to-month and year-to-year differences. Define  $N1_t$  as

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

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

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

while for the multiplicative model

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

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

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

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

### Year-to-Year Changes in the Seasonally Adjusted Data

First define  $N12_t$  as

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

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

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

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

while for the multiplicative model

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

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

$$\text{MPD}_t = \max_{k \in N_{1t}} YY_t(k) - \min_{k \in N_{1t}} YY_t(k)$$

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

---

## Data Requirements

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

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

---

## Missing Values

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

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

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

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

---

## Prior Daily Weights and Trading-Day Regression

Suppose that a detailed examination of retail sales at ZXY Company indicates that certain days of the week have higher amounts of sales. In particular, Thursday, Friday, and Saturday have approximately twice the



amount of sales as Monday, Tuesday, and Wednesday, and no sales occur on Sunday. This means that months with five Saturdays would have higher amounts of sales than months with only four Saturdays.

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

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

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

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

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

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

---

## Adjustment for Prior Factors

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

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

This prior adjustment on SALES is performed by replacing SALES with the calculated value  $(\text{SALES}/\text{PMF}) * 100.0$ . 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 sales2;
  set sales;
  if '01jul1988'd <= date <= '01aug1988'd then pmf = 50;
  else pmf = 100;
run;

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

```

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

---

## The YRAHEADOUT Option

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

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

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

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

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

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

---

## Effect of Backcast and Forecast Length

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

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

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

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

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

## Details of Model Selection

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

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

**Table 43.2** 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 43.2. As each model is estimated, the following three criteria are checked:

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

The descriptions of these three criteria are given in the section “[Criteria Details](#)” on page 3214. The default values for these criteria are those used by X11ARIMA/88 from Statistics Canada; these defaults can be changed by the MAPECR=, CHICR=, and OVDIFCR= options.

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

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

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

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

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

## Criteria Details

### Mean Absolute Percent Error (MAPE)

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

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

With this notation, the MAPE criteria are computed as

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

### Box-Ljung Chi-Square

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

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

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 (1978) as yielding a better fit to the asymptotic chi-square distribution. Some simulation studies of the finite sample properties of this statistic are given by Davies, Triggs, and Newbold (1977) and by Ljung and Box (1978).

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

### Over-differencing Test

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

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

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

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

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

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

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

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

Instead, an approximate test is performed by requiring that

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

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

### ARIMA Statement Options for the Five Predefined Models

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

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

The multiplicative case is assumed in Table 43.3. The indicated seasonality  $s$  in the specification is either 12 (monthly) or 4 (quarterly). The MODEL statement assumes a monthly series.

**Table 43.3** 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 later, in Table 43.4.

---

## The OUTSPAN= Data Set

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

- A1, a numeric variable that is a copy of the original series truncated to the current span. Note that overlapping spans will contain identical values for this variable.
- C18, a numeric variable that contains the trading-day factors for the seasonal adjustment for the current span
- D10, a numeric variable that contains the seasonal factors for the seasonal adjustment for the current span
- D11, a numeric variable that contains the seasonally adjusted series for the current span
- DATE, a numeric variable that contains the date within the current span
- SPAN, a numeric variable that contains the current span. The first span is the earliest span—that is, the one with the earliest starting date.
- 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 (Shiskin, Young, and Musgrave 1967, Appendix A). These data contain the following variables:

- 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 PROC X11 are required (when TDREGR=PRINT, TEST, or ADJUST), Table D8 and the stable seasonality test are computed twice: once in the initial pass, then again in the final pass. Both of these computations are put in the OUTSTB data set and are identified by D18.1 and D18.2, respectively.
- **SOURCE**, a character variable corresponding to the “source” column in the analysis of variance table following Table D8
- **SS**, a numeric variable containing the sum of squares associated with the corresponding source term
- **DF**, a numeric variable containing the degrees of freedom associated with the corresponding source term
- **MS**, a numeric variable containing the mean square associated with the corresponding source term. MS is missing for the source term “Total.”
- **F**, a numeric variable containing the  $F$  statistic for the “Between” source term. F is missing for all other source terms.
- **PROBF**, a numeric variable containing the significance level for the  $F$  statistic. PROBF is missing for the source terms “Total” and “Error.”

---

## OUTTDR= Data Set

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

- **VARNAME**, a character variable containing the name of the VAR variable being processed
- **TABLE**, a character variable containing the name of the table. It can have only the value B15 (Preliminary Trading-Day Regression) or C15 (Final Trading-Day Regression).
- **\_TYPE\_**, a character variable whose value distinguishes the three distinct table format types. These types are (a) the regression, (b) the listing of the standard error associated with length-of-month, and (c) the analysis of variance. The first seven observations in the OUTTDR= data set correspond to the regression on days of the week; thus the \_TYPE\_ variable is given the value “REGRESS” (day-of-week regression coefficient). The next four observations correspond to 31-, 30-, 29-, and 28-day months and are given the value \_TYPE\_=LOM\_STD (length-of-month standard errors). Finally, the last three observations correspond to the analysis of variance table, and \_TYPE\_=ANOVA.
- **PARM**, a character variable, further identifying the nature of the observation. PARM is set to blank for the three \_TYPE\_=ANOVA observations.
- **SOURCE**, a character variable containing the source in the regression. This variable is missing for all \_TYPE\_=REGRESS and 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`.
- SS, a numeric variable containing the sum of squares associated with the corresponding source term. This variable is missing for all `_TYPE_=REGRESS` and `LOM_STD`.
- DF, a numeric variable containing the degrees of freedom associated with the corresponding source term. This variable is missing for all `_TYPE_=REGRESS` and `LOM_STD`.
- MS, a numeric variable containing the mean square associated with the corresponding source term. This variable is missing for the source term 'Total' and for all `_TYPE_=REGRESS` and `LOM_STD`.
- F, a numeric variable containing the  $F$  statistic for the 'Regression' source term. The variable is missing for the source terms 'Total' and 'Error' and for all `_TYPE_=REGRESS` and `LOM_STD`.
- PROBF, a numeric variable containing the significance level for the  $F$  statistic. This variable is missing for the source term 'Total' and 'Error' and for all `_TYPE_=REGRESS` and `LOM_STD`.

---

## Printed Output

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

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



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

Table 43.4 describes the individual tables and charts. Most tables apply to both 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 that the table is not a time series, and is only printed, not output to the OUT= data set.

**Table 43.4** 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 and 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	

**Table 43.4** *continued*

<b>Table</b>	<b>Description</b>	<b>Notes</b>
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
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. For information about specifying individual tables, see the section “[TABLES Statement](#)” on page 3202. The following list gives the tables printed for each value of the PRINTOUT= option:

STANDARD (26 tables)	A1–A4, B1, C13–C19, D8–D13, E1–E6, F1, F2
LONG (40 tables)	A1–A5, A13–A15, B1, B2, B7, B10, B13–B15, C1, C7, C10, C13–C19, D1, D7–D11, D13, E1–E6, F1, F2
FULL (62 tables)	A1–A5, A13–A15, B1–B11, B13–B19, C1–C11, C13–C19, D1, D2, D4–D12, E1–E6, F1, F2

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

## The CHARTS= Option

Of the four charts listed in [Table 43.4](#), G1 and G2 are printed by default (CHARTS=STANDARD). Charts G3 and G4 are printed when CHARTS=FULL is specified. For information about specifying individual charts, see the section “[TABLES Statement](#)” on page 3202.

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

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

### Motivation

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

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

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

PROC X11 next computes a nonparametric Kruskal-Wallis chi-squared test for stable seasonality, “Nonparametric Test for the Presence of Seasonality Assuming Stability.” The Kruskal-Wallis test is performed on the

ranks of the Final Unmodified SI Ratios (Table D8). For more information about the Kruskal-Wallis test, see Lehmann and D'Abrera (2006, pp. 204–210).

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

### **Interpretation and Statistical Details**

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

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

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

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

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

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

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

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

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

$$T_1 = \frac{7}{F_s}$$

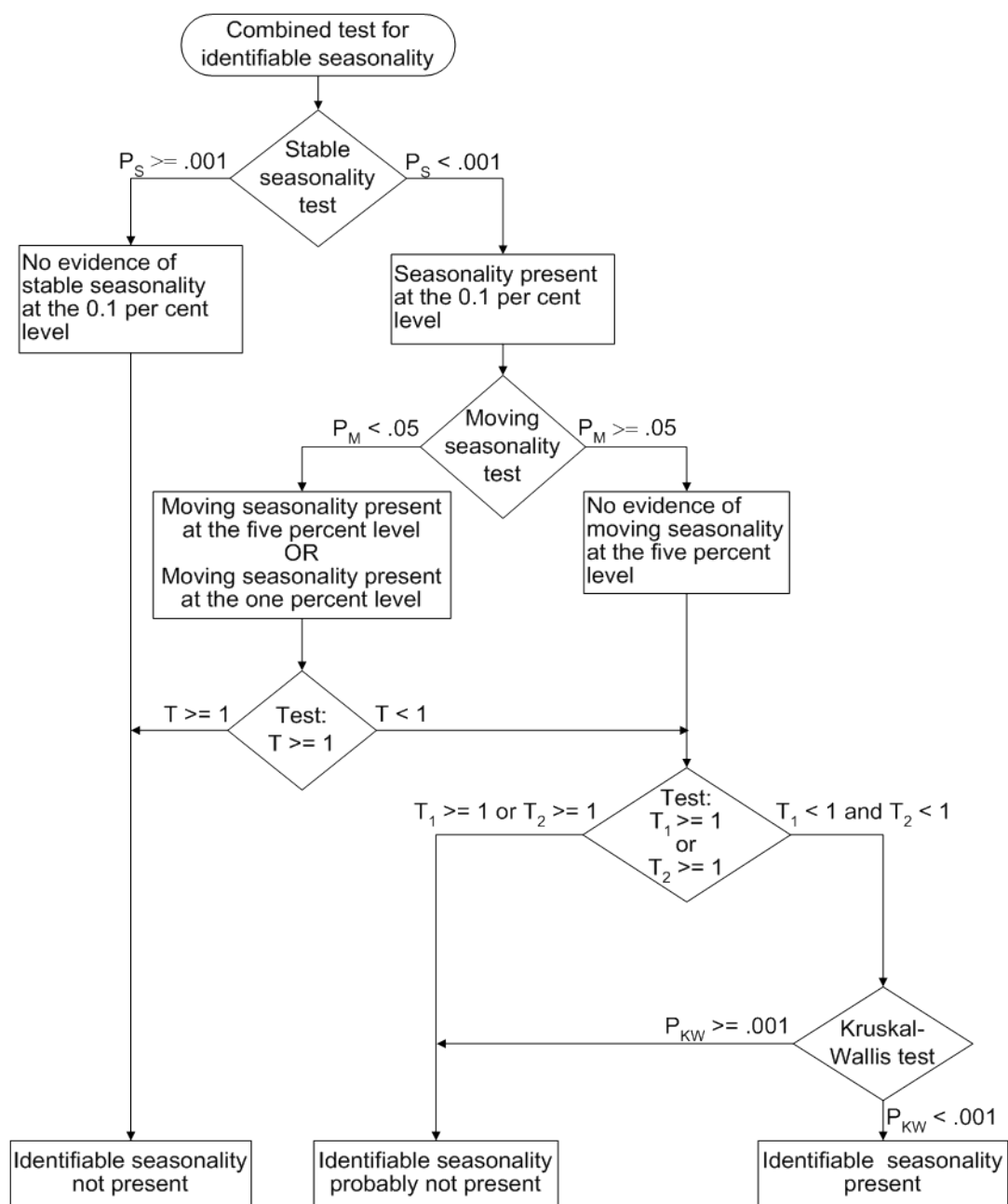
$$T_2 = \frac{3F_m}{F_s}$$

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

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

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

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

**Figure 43.5** Combined Seasonality Test Flowchart

## Tables Written to the OUT= Data Set

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

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

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

## Printed Output Generated by Sliding Spans Analysis

### Table S 0.A

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

### Table S 0.B

Table S 0.B gives the summary of the two  $F$  tests performed during the standard X11 seasonal adjustments for stable and moving seasonality on Table D8, the final SI ratios. These tests are described in the section “Printed Output” on page 3218.

### Table S 1.A

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

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

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

$$\text{MPD}_j = \frac{\max_{k=1,\dots,4} \bar{S}_j(k) - \min_{k=1,\dots,4} \bar{S}_j(k)}{\min_{k=1,\dots,4} \bar{S}_j(k)}$$

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

### Table S 1.B

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

### Breakdown Tables

Table S 2.A.1 begins the breakdown analysis for the various series considered in the sliding spans analysis. The key concept here is the MPD described earlier in the section “Table S 1.A” on page 3225 and in the section “Computational Details for Sliding Spans Analysis” on page 3208. For a month or quarter that appears in two or more spans, the maximum percentage difference is computed and tested against a cutoff level. If it exceeds this cutoff, it is counted as an instance of exceeding the level. It is of interest to see if such instances fall disproportionately in certain months and years. Tables S 2.A.1 through S 6.A.3 display this breakdown for all series considered.

#### Table S 2.A.1

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

#### Table S 2.A.2

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

#### Table S 2.A.3

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

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

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

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

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

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

Again, consider Table D10, the seasonal factors in a sliding spans analysis that uses four spans. For a given observation  $t$  (containing at least three spans), note the level of D10 for the first span. Continue across the spans until a difference of 1.0% or greater occurs (or no more spans are left), noting whether the difference is up or down. If the difference is up, continue until a difference of 1.0% or greater occurs downward (or no more spans are left). If such an up-down combination occurs, the observation is counted as an up-down turning point. A similar description occurs for a down-up turning point. Tables S 7.A through S 7.E, described



below, show the occurrence of turning points, indicating whether up-down or down-up. Note that it requires at least three spans to test for a turning point. Hence Tables S 2.A.3 through S 6.A.3 show a reduced number in the “Turning Point” row for the “Total Tested” column, and in Tables S 7.A through S 7.E, the turning points symbols can occur only where three or more spans overlap.

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

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

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

The footnote to Table S 2.A.3 gives the U.S. Census Bureau recommendation for thresholds, as described in the section “[Computational Details for Sliding Spans Analysis](#)” on page 3208.

#### **Table S 2.B**

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

#### **Table S 2.C**

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

#### **Tables S 3.A.1 through S 3.A.3**

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

#### **Tables S 3.B, S 3.C**

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

#### **Tables S 4.A.1 through S 4.A.3**

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

**Tables S 4.B, S 4.C**

These tables, applied to the seasonally adjusted series (Table D11), are the same format as Tables S 2.B through S 2.C.

**Tables S 5.A.1 through S 5.A.3**

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

**Tables S 5.B, S 5.C**

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

**Tables S 6.A.1 through S 6.A.3**

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

**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 through S 2.C. The same default cutoff value of 3.0% is used.

**Table S 7.A**

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

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

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

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

The next column is the maximum percentage difference (MPD). This quantity, described in the section “Computational Details for Sliding Spans Analysis” on page 3208, is the main computation for sliding spans analysis. A measure of how extreme the MPD value is given in the last column, the “Level of Excess” column. The symbols used and their meaning are described in Table S 2.A.3. If a given observation has exceeded the cutoff, the level of excess column is blank.

**Table S 7.B**

Table S 7.B gives the entire listing of the trading-day factors (Table C18) for each span. The format of this table is exactly like that of Table S 7.A.

**Table S 7.C**

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

**Table S 7.D**

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

**Table S 7.E**

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

## Printed Output from the ARIMA Statement

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

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

---

## ODS Table Names

PROC X11 assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 43.5.

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

**Table 43.5** ODS Tables Produced in PROC X11

ODS Table Name	Description	Option
<b>ODS Tables Created by the MONTHLY and QUARTERLY Statements</b>		
Preface	X11 Seasonal Adjustment Program information giving credits, dates, and so on	Always printed unless NOPRINT
A1	Table A1: original series	
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	
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 combined trading day weights	
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	

**Table 43.5** *continued*

ODS Table Name	Description	Option
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 combined trading day weights	
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	
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 adjusted 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	

**Table 43.5** *continued*

ODS Table Name	Description	Option
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	
StableSeasFTest	Moving seasonality test	
KruskalWallisTest	Nonparametric test for the presence of seasonality assuming stability	
CombinedSeasonalityTest	Summary of results and combined test for the presence of identifiable seasonality	
f2a	F2 summary measures, part 1	
f2b	F2 summary measures, part 2	
f2c	F2 summary measures, part 3	
f2d	I/C ratio for monthly/quarterly span	
f2f	Average % change with regard to sign and standard deviation over span	
E4	Differences or ratios of annual totals for original and adjusted series	
ChartG1	Chart G1	
ChartG2	Chart G2	

**Table 43.5** *continued*

ODS Table Name	Description	Option
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 quarter	
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	PRINTALL
SPR7_X	S 7.X sliding spans analysis	

## Examples: X11 Procedure

### Example 43.1: Component Estimation—Monthly Data

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

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

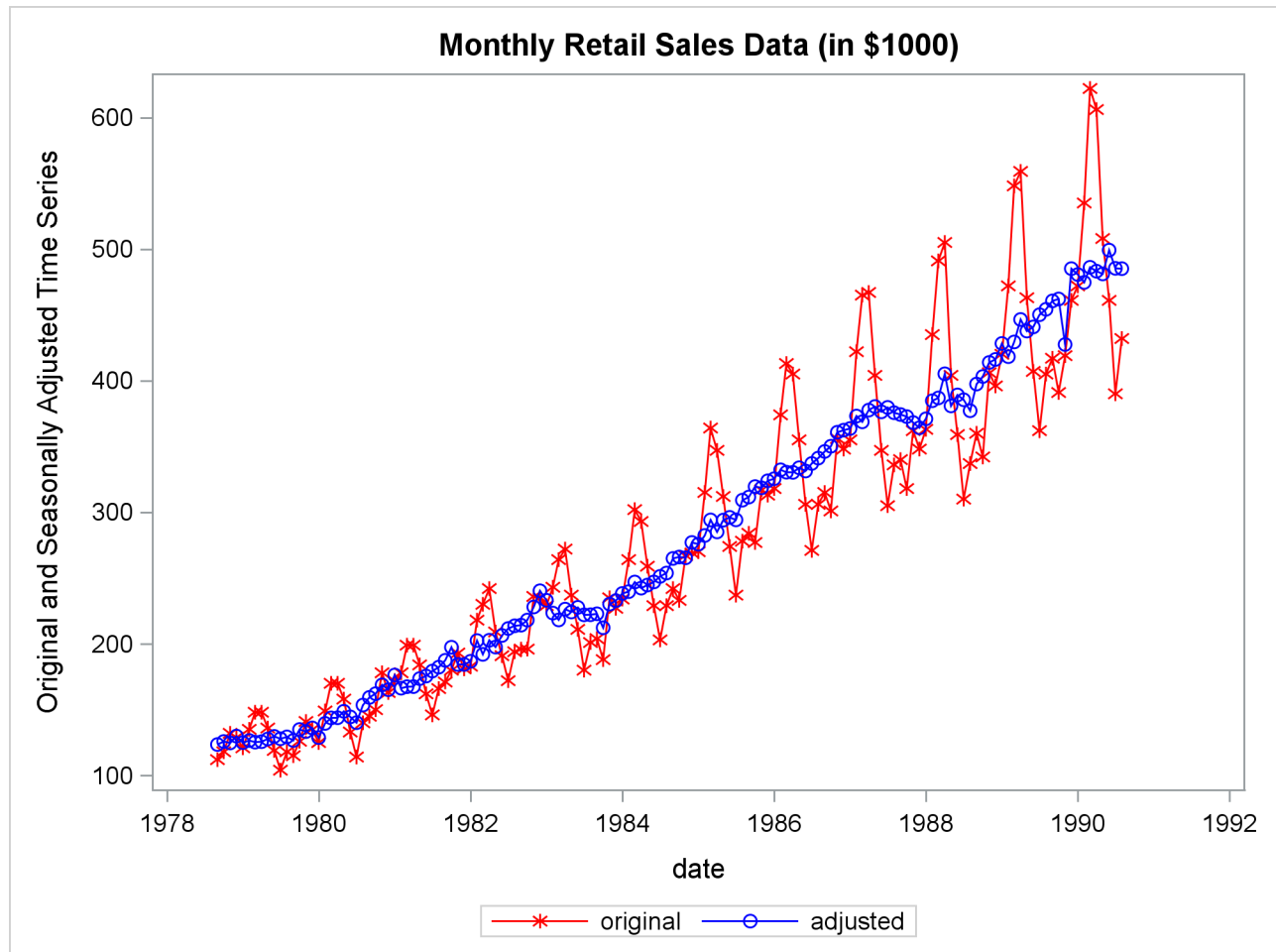
```
data sales;
  input sales @@;
  date = intnx( 'month', '01sep1978'd, _n_-1 );
  format date monyy7.;
datalines;
112 118 132 129 121 135 148 148 136 119 104 118

... more lines ...

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

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



**Output 43.1.1** Plots of Original and Seasonally Adjusted Data

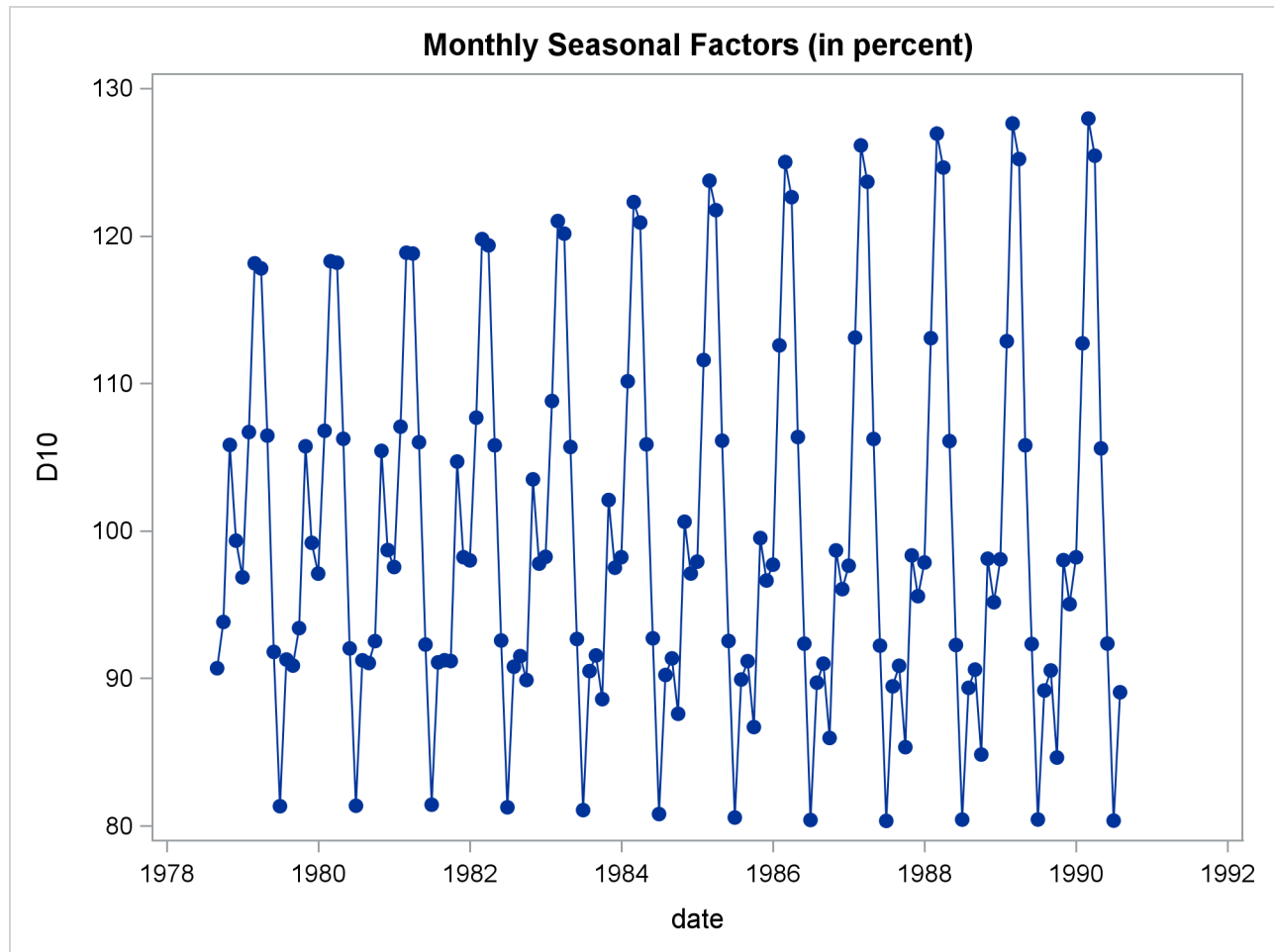
```

title 'Monthly Seasonal Factors (in percent)';
proc sgplot data=out;
    series x=date y=d10 / markers markerattrs=(symbol=CircleFilled) ;
run;

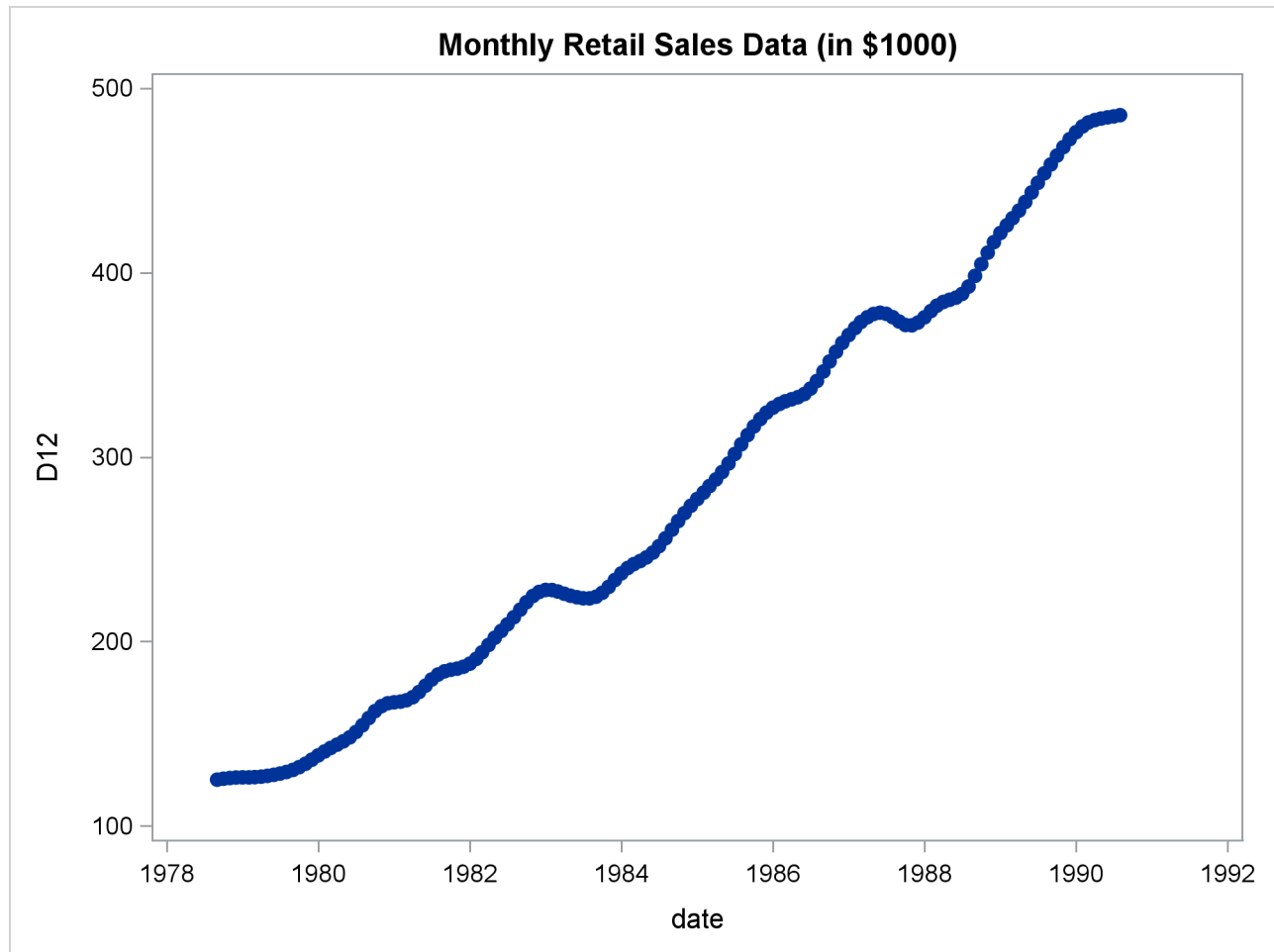
title 'Monthly Retail Sales Data (in $1000)';
proc sgplot data=out;
    series x=date y=d12 / markers markerattrs=(symbol=CircleFilled) ;
run;

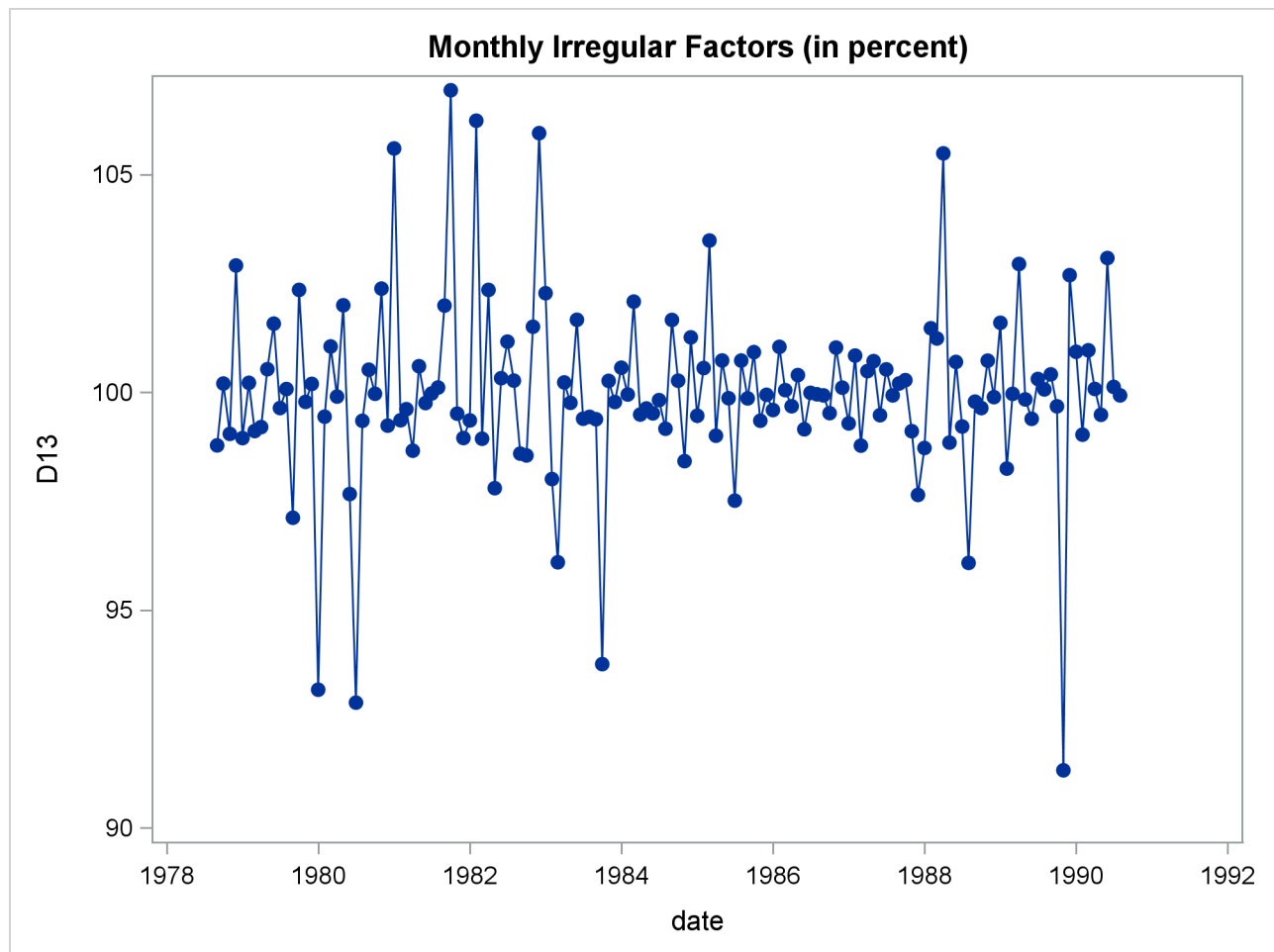
title 'Monthly Irregular Factors (in percent)';
proc sgplot data=out;
    series x=date y=d13 / markers markerattrs=(symbol=CircleFilled) ;
run;

```

**Output 43.1.2** Plot of D10, the Final Seasonal Factors

**Output 43.1.3** Plot of D12, the Final Trend Cycle



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

### Example 43.2: Components Estimation—Quarterly Data

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

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

... more lines ...
```

```

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

```

### Output 43.2.1 X11 Procedure Quarterly Example

#### Monthly Retail Sales Data (in \$1000)

##### The X11 Procedure

##### Seasonal Adjustment of - fy35rr

---

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

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

Part      Description

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

Series - fy35rr

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

---

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**

**Output 43.2.2** X11 Procedure Quarterly Example, Table D11

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 43.3: Outlier Detection and Removal

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

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

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

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

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

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

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

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

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

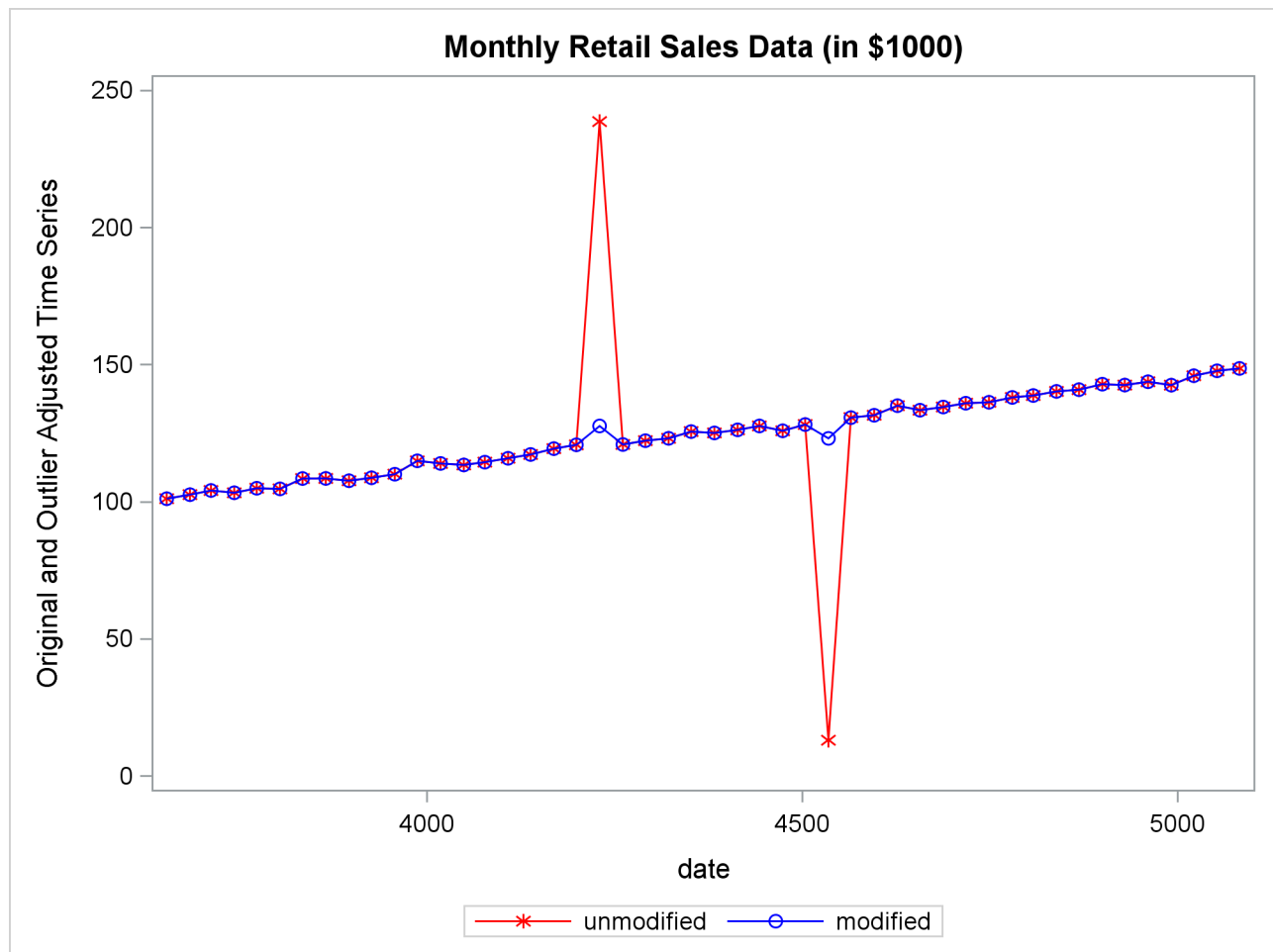
### Output 43.3.1 Detection of Extreme Irregulars

### Monthly Retail Sales Data (in \$1000)

## The X11 Procedure

### Seasonal Adjustment of - x

[illegible]

**Output 43.3.2** 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* 2:291–320.
- Bobbitt, L. G., and Otto, M. C. (1990). "Effects of Forecasts on the Revisions of Seasonally Adjusted Data Using the X-11 Adjustment Procedure." In *Proceedings of the Business and Economic Statistics Section*, 449–453. Alexandria, VA: American Statistical Association.
- Buszuwski, J. A. (1987). "Alternative ARIMA Forecasting Horizons When Seasonally Adjusting Producer Price Data with X-11-ARIMA." In *Proceedings of the Business and Economic Statistics Section*, 488–493. Alexandria, VA: American Statistical Association.
- Cleveland, W. P., and Tiao, G. C. (1976). "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program." *Journal of the American Statistical Association* 71:581–587.
- 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* 75:487–496.



- Dagum, E. B. (1980). *The X-11-ARIMA Seasonal Adjustment Method*. Ottawa: Statistics Canada.
- Dagum, E. B. (1982a). "The Effects of Asymmetric Filters on Seasonal Factor Revision." *Journal of the American Statistical Association* 77:732–738.
- Dagum, E. B. (1982b). "Revisions of Seasonally Adjusted Data Due to Filter Changes." In *Proceedings of the Business and Economic Section*, 39–45. Alexandria, VA: American Statistical Association.
- Dagum, E. B. (1982c). "Revisions of Time Varying Seasonal Filters." *Journal of Forecasting* 1:173–187.
- Dagum, E. B. (1983). *The X-11-ARIMA Seasonal Adjustment Method*. Technical Report 12-564E, Statistics Canada, Ottawa.
- Dagum, E. B. (1985). "Moving Averages." In *Encyclopedia of Statistical Sciences*, vol. 5, edited by S. Kotz, N. L. Johnson, and C. B. Read. New York: John Wiley & Sons.
- Dagum, E. B. (1988). *The X-11-ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual*. Ottawa: Statistics Canada.
- Dagum, E. B., and Laniel, N. (1987). "Revisions of Trend Cycle Estimators of Moving Average Seasonal Adjustment Method." *Journal of Business and Economic Statistics* 5:177–189.
- 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., and 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*, edited by M. R. Perryman and J. R. Schmidt, 195–228. Amsterdam: Kluwer-Nijhoff.
- 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* 85:345–355.
- 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* 8:145–152.
- Higginson, J. (1975). *An F Test for the Presence of Moving Seasonality When Using Census Method II-X-II Variant*. StatCan Staff Paper STC2102E, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- 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* 2:217–229.
- Ladiray, D., and Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method*. New York: Springer-Verlag.
- Laniel, N. (1985). "Design Criteria for the 13-Term Henderson End-Weights." Working paper, Methodology Branch, Statistics Canada, Ottawa.
- Lehmann, E. L., and D'Abrera, H. J. M. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Rev. ed. New York: Springer.
- 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. (1978). *The Identification and Treatment of Moving Seasonality in the X-11 Seasonal Adjustment Method*. StatCan Staff Paper STC0803E, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- Lothian, J. (1984a). *The Identification and Treatment of Moving Seasonality in the X-11-ARIMA Seasonal Adjustment Method*. StatCan Staff Paper, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- Lothian, J. (1984b). "The Identification and Treatment of Moving Seasonality in X-11-ARIMA." In *Proceedings of the Business and Economic Statistics Section*, 166–171. Alexandria, VA: American Statistical Association.
- Lothian, J., and Morry, M. (1978a). *Selection of Models for the Automated X-11-ARIMA Seasonal Adjustment Program*. StatCan Staff Paper STC1789, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- Lothian, J., and Morry, M. (1978b). *A Test for the Presence of Identifiable Seasonality When Using the X-11-ARIMA Program*. StatCan Staff Paper STC2118, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- Marris, S. N. (1961). "The Treatment of Moving Seasonality in Census Method II." In *Seasonal Adjustment on Electronic Computers*, 257–309. Paris: Organisation for Economic Co-operation and Development.
- Monsell, B. C. (1984). *The Substantive Changes in the X-11 Procedure of X-11-ARIMA*. SRD Research Report Census/SRD/RR-84/10, Statistical Research Division, US Bureau of the Census.
- Pierce, D. A. (1980). "Data Revisions with Moving Average Seasonal Adjustment Procedures." *Journal of Econometrics* 14:95–114.
- Shiskin, J. (1958). "Decomposition of Economic Time Series." *Science* 128:1539–1546.
- Shiskin, J., and Eisenpress, H. (1957). "Seasonal Adjustments by Electronic Computer Methods." *Journal of the American Statistical Association* 52:415–449.
- Shiskin, J., Young, A. H., and Musgrave, J. C. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. Technical Report 15, US Department of Commerce, Bureau of the Census.
- US Bureau of the Census (1969). *X-11 Information for the User*. Washington, DC: US Government Printing Office.
- Young, A. H. (1965). *Estimating Trading Day Variation in Monthly Economic Time Series*. Technical Report 12, US Department of Commerce, Bureau of the Census, Washington, DC.

# Subject Index

- BY groups
  - X11 procedure, [3193](#)
- Census X-11 method, *see* X11 procedure
- Census X-11 methodology
  - X11 procedure, [3205](#)
- combined seasonality test, [3221](#)
- data requirements
  - X11 procedure, [3210](#)
- differences with X11ARIMA/88
  - X11 procedure, [3204](#)
- ID variables
  - X11 procedure, [3193](#), [3195](#)
- irregular component
  - X11 procedure, [3182](#), [3187](#)
- Kruskal-Wallis test, [3221](#)
- model selection for X-11-ARIMA method
  - X11 procedure, [3213](#)
- moving seasonality test, [3221](#)
- output data sets
  - X11 procedure, [3216](#), [3217](#)
- output table names
  - X11 procedure, [3229](#)
- printed output
  - X11 procedure, [3218](#)
- seasonal adjustment
  - time series data, [3182](#)
  - X11 procedure, [3182](#), [3187](#)
- seasonal component
  - X11 procedure, [3182](#)
- seasonality tests, [3221](#)
- sliding spans analysis, [3206](#)
- stable seasonality test, [3221](#)
- time series data
  - seasonal adjustment, [3182](#)
- trading-day component
  - X11 procedure, [3182](#), [3187](#)
- trend cycle component
  - X11 procedure, [3182](#), [3187](#)
- X-11 ARIMA methodology
  - X11 procedure, [3204](#)
- X-11 seasonal adjustment method, *see* X11 procedure
- X-11-ARIMA seasonal adjustment method, *see* X11 procedure
- X11 procedure
  - BY groups, [3193](#)
  - Census X-11 method, [3182](#)
  - Census X-11 methodology, [3205](#)
  - data requirements, [3210](#)
  - differences with X11ARIMA/88, [3204](#)
  - ID variables, [3193](#), [3195](#)
  - irregular component, [3182](#), [3187](#)
  - model selection for X-11-ARIMA method, [3213](#)
  - output data sets, [3216](#), [3217](#)
  - output table names, [3229](#)
  - printed output, [3218](#)
  - seasonal adjustment, [3182](#), [3187](#)
  - seasonal component, [3182](#)
  - trading-day component, [3182](#), [3187](#)
  - trend cycle component, [3182](#), [3187](#)
  - X-11 ARIMA methodology, [3204](#)
  - X-11 seasonal adjustment method, [3182](#)
  - X-11-ARIMA seasonal adjustment method, [3182](#)



# Syntax Index

- ADDITIVE option
  - MONTHLY statement (X11), 3194
  - QUARTERLY statement (X11), 3199
- ARIMA statement
  - X11 procedure, 3190
- BACKCAST= option
  - ARIMA statement (X11), 3190
- BY statement
  - X11 procedure, 3193
- CENTER option
  - ARIMA statement (X11), 3192
- CHARTS= option
  - MONTHLY statement (X11), 3194
  - QUARTERLY statement (X11), 3199
- CHICR= option
  - ARIMA statement (X11), 3191
- CONVERGE= option
  - ARIMA statement (X11), 3191
- CUTOFF= option
  - SSPAN statement (X11), 3201
- DATA= option
  - PROC X11 statement, 3189
- DATE= option
  - MONTHLY statement (X11), 3195
  - QUARTERLY statement (X11), 3199
- END= option
  - MONTHLY statement (X11), 3195
  - QUARTERLY statement (X11), 3200
- EXCLUDE= option
  - MONTHLY statement (X11), 3195
- FORECAST= option
  - ARIMA statement (X11), 3191
- FULLWEIGHT= option
  - MONTHLY statement (X11), 3195
  - QUARTERLY statement (X11), 3200
- ID statement
  - X11 procedure, 3193
- LENGTH option
  - MONTHLY statement (X11), 3195
- MACURVES statement
  - X11 procedure, 3193
- MAPECR= option
  - ARIMA statement (X11), 3191
- MAXITER= option
  - ARIMA statement (X11), 3191
- METHOD= option
  - ARIMA statement (X11), 3191
- MODEL= option
  - ARIMA statement (X11), 3191
- MONTHLY statement
  - X11 procedure, 3194
- NDEC= option
  - MONTHLY statement (X11), 3196
  - QUARTERLY statement (X11), 3200
  - SSPAN statement (X11), 3201
- NOINT option
  - ARIMA statement (X11), 3192
- NOPRINT option
  - ARIMA statement (X11), 3192
  - PROC X11 statement, 3190
  - SSPAN statement (X11), 3202
- OUT= option
  - OUTPUT statement (X11), 3198, 3216
- OUTEXTRAP option
  - PROC X11 statement, 3189
- OUTPUT statement
  - X11 procedure, 3198
- OUTSPAN= option
  - PROC X11 statement, 3190, 3216
  - VAR statement (X11), 3216
- OUTSTB= option
  - PROC X11 statement, 3190, 3216
- OUTTDR= option
  - PROC X11 statement, 3190, 3217
- OVDIFCR= option
  - ARIMA statement (X11), 3192
- PDWEIGHTS statement
  - X11 procedure, 3198
- PMFACTOR= option
  - MONTHLY statement (X11), 3196
- PRINT option
  - SSPAN statement (X11), 3202
- PRINTALL option
  - ARIMA statement (X11), 3192
  - SSPAN statement (X11), 3202
- PRINTFP option
  - ARIMA statement (X11), 3192
- PRINTOUT= option

MONTHLY statement (X11), 3196  
QUARTERLY statement (X11), 3200  
PROC X11 statement, 3189

QUARTERLY statement  
X11 procedure, 3199

SSPAN statement  
X11 procedure, 3201

START= option  
MONTHLY statement (X11), 3196  
QUARTERLY statement (X11), 3200  
SUMMARY option  
MONTHLY statement (X11), 3196  
QUARTERLY statement (X11), 3201

TABLES statement  
X11 procedure, 3202

TDCOMPUTE= option  
MONTHLY statement (X11), 3197

TDCUTOFF= option  
SSPAN statement (X11), 3202

TDREGR= option  
MONTHLY statement (X11), 3197

TRANSFORM= option  
ARIMA statement (X11), 3192

TRENDADJ option  
MONTHLY statement (X11), 3197  
QUARTERLY statement (X11), 3201

TRENDMA= option  
MONTHLY statement (X11), 3197

VAR statement  
X11 procedure, 3202

X11 procedure, 3187  
syntax, 3187

YRAHEADOUT option  
PROC X11 statement, 3190

ZEROWEIGHT= option  
MONTHLY statement (X11), 3197  
QUARTERLY statement (X11), 3201