Statistics Canada Statistique Canada

Methodology Branch

Time Series Research and Analysis Division

Direction de la méthodologie

Division de la recherche
et de l'analyse des chroniques

Canada

[redacted]

WORKING PAPER TSRA-86-004E

TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY BRANCH

SPECIFICATION FOR PROC X11ARIMA

by

M. Gratton and E.B. Dagum

This is a preliminary version. Do not quote without authors' permission.
Comments are welcome.

Résumé

La procédure X11ARIMA est une adaptation du programme de désaisonnalisation X-11-ARMMI développé à Statistique Canada par Dagum (1980). La procédure peut répéter toutes les opérations disponibles dans "PROC X11" en plus des options additionnelles suivantes: (a) modélisation ARMMI (Box et Jenkins, 1970) pour l'extrapolation des séries; (b) traitement des séries de composition par addition, soustraction, multiplication et division; et (c) un ensemble de statistique de contrôle pour évaluer la fiabilité des résultats.

PREFACE

This document was prepared for the following four purposes:

(1) A user of the X-11-ARIMA program can verify that all of its features are accurately translated into a set of SAS statements described in SPECIFICATIONS Section;

(2) The SPECIFICATIONS may be scrutinized by experts to ensure that:

  (a) syntax and semantics are reasonable;

  (b) the relationship between PROC X11 and PROC X11ARIMA is well defined;

(3) The development team will use the document as the target for implementation;

(4) A user's manual will be created based on this document.

ACKNOWLEDGEMENT

The authors wish to acknowledge the contributions made by David Delong, SAS Institute, Pierre Cholette, Helen Fung, John Higginson, Marietta Morry, and Alfred Papineau, Time Series Research and Analysis Division, Statistics Canada.

TABLE OF CONTENTS

PREFACE

THE X11ARIMA PROCEDURE

THE X11ARIMA PROCEDURE

ABSTRACT

The X11ARIMA procedure is an adaptation of the X-11-ARIMA Seasonal Adjustment Program developed at Statistics Canada by Dagum (1980). The procedure offers all the functions available in the PROC X11 plus additional features that include: (a) ARIMA modelling (Box and Jenkins, 1970) for extrapolation of the series; (b) the composition of series by addition, subtraction, multiplication and division of several components; and (c) a set of control statistics to assess the reliability of the output.

INTRODUCTION

This section was extracted from Dagum (1980) for the sole purpose of explaining the SPECIFICATION section. For a more complete discussion, readers are referred to the above document (available from Statistics Canada, Catalogue No. 12-564E (in English) or No. 12-564F (in French)).

Functionally speaking, the X11ARIMA consists of the following three parts:

1. Modelling the original series by autoregressive integrated moving average processes (ARIMA models) of the Box and Jenkins (1970) types.
2. Extrapolating one year of unadjusted data at each end of the series from ARIMA models that fit and project the original series well. This operation, called "forecasting" and "backcasting" is designed to extend the

observed series at both ends.

3. Seasonally adjusting the extended or original series with various seasonal and trend-cycle moving averages. When the ARIMA extrapolation option is not used, these moving averages are close to those of the X-11 Variant of the Census Method II developed by Shiskin, Young and Musgrave (1967). In addition, the user has the option of applying a centred 24-term filter instead of the centred 12-term moving averge for the preliminary estimation of the trend-cycle. This new filter gives better results for series strongly affected by short cycles (less than three years) or sudden changes in trend.

ARIMA Models and Extrapolation

ARIMA models bring together two basic concepts in extrapolating: autoregression and moving averages. ARIMA is an acronym for AutoRegressive Integrated Moving Average. It signifies that stationary models which are fitted to the differenced data must be summed or "integrated" to provide models for the non-stationary data.

Following the symbolic notation of Box and Jenkins (1970), a multiplicative seasonal ARIMA model may be expressed as ARIMA (p,d,q)(P,D,Q)s

where d is the order of the ordinary differences and D is the order of the seasonal differences applied to the original series to make it stationary;

p and P are the orders of ordinary and seasonal auto-regressive processes, respectively;

q and Q are the orders of ordinary and seasonal moving average processes, respectively;

s is the seasonal periodicity (s=12 for monthly data and s=4 for quarterly data).

If a first difference (d=1) is applied to the original series Z_t, the new series W_t is

$$W_t = Z_t - Z_{t-1}$$

Let B denote the lag (backshift) operation such that:

$$B^n Z_t = Z_{t-n}$$

then

$$W_t = (1-B)Z_t$$

For more complex cases of deterministic or stochastic instability, higher order differences are applied. If a seasonal difference (D=1) is applied, then the transformed series is

$$W_t = Z_t - Z_{t-s} = (1-B^s)Z_t$$

If ordinary and seasonal autoregressive processes are of order one (p=1 and P=1), then the parameters ϕ and Φ affect Z_{t-1} and Z_{t-s}, respectively. The parameters ϕ and Φ measure respectively the impact of Z_{t-1} and Z_{t-s} on the dependent variable Z_t.

If ordinary and seasonal moving average processes are of order 1 (q=1 and Q=1), then the parameters θ and Θ affect the residuals a_{t-1} and a_{t-s}, respectively. The parameters θ and Θ measure respectively the impact of a_{t-1} and a_{t-s} on the dependent variable Z_t.

Hence, for ARIMA models, the variable Z_t is a function of lagged dependent variables and lagged residuals. For example, a special multiplicative model, ARIMA $(0,1,1)(0,1,1)_4$ can be written as:

$$(1-B)(1-B^4)Z_t = (1- \quad B)(1- \quad B^4)a_t$$

$$\text{or} \quad Z_t = Z_{t-1} + Z_{t-4} - Z_{t-5} + a_t - \theta a_{t-1} - \Theta a_{t-4} + \theta \Theta a_{t-5}$$

Selection of ARIMA Models

Based on the criteria given in Dagum (1981) for fitting and extrapolation, three ARIMA models have been incorporated for the automatic selection (the procedure tests for the most appropriate model). For this option, the series must be at least five years long. For series longer than 15 years, only the last 15 years will be used in the ARIMA fit and extrapolation. If the test is positive, the model chosen is the one that gives the smallest average extrapolation error. Then the unadjusted series is automatically extended with one year of extrapolated data and seasonally adjusted. If the test is negative, a message is given indicating that extrapolated values have not been incorporated into the unadjusted series.

Alternatively, models can be selected and supplied by the user in the form of ARIMA(p,d,q)(P,D,Q)s and desired logarithmic or power transformations.

If no model is accepted by the automatic option or imposed by the user, the procedure becomes almost equivalent to PROC X11 with added features developed at Statistics Canada. The major distinctions are due to a different treatment of extreme values and the fact that the end-weights of the seasonal and trend-cycle moving averages are not truncated at three digits but given in single precision.

Main Steps in Producing a Seasonally Adjusted Series

The X-11-ARIMA assumes that the main components of a time series O follow a multiplicative model (arithmentic mean of seasonal factors is 1), additive model (arithmetic mean of seasonal factors is 0), or logarithmic model (geometric mean of seasonal factors is 1). Each model can be expressed as:

(a) $O = C\ S\ I$ (multiplicative model)

(b) $O = C + S + I$ (additive model)

(c) $\log O = \log C + \log S + \log I$ (log additive model)

where O : original series

C : trend cycle component

I : irregular component

S : seasonal component

The estimation is made with different kinds of moving averages that are applied sequentially in 13 steps repeated twice. If prior factors (e.g. monthly prior factors, quarterly prior factors, prior daily weights etc.) are applicable, the series is adjusted using prior factors and then the components are estimated.

The main steps to produce a seasonally adjusted series are as follows (under the assumption of a multiplicative decomposition):

1. Compute the ratios between the original series and a centred 12-term moving average (2 x 12 m.a., that is a 2-term average of a 12-term average) as a first estimate of the seasonal and irregular components (SI).

2. Apply a weighted 5-term moving average (3 x 3 m.a.) to the seasonal-irregular ratios (SI) of each month separately, to obtain a preliminary estimate of the seasonal factors.

3. Compute a centred 12-term moving average of the preliminary factors found in step 2 for the entire series. To obtain the six missing values at either end of this average, repeat the first (last) available moving average value six times. Adjust the factors to add to 12 (approximately) over any 12-month period by dividing the centred 12-term average into the factors.

4. Divide the seasonal factor estimates into the seasonal irregular (SI) ratios to obtain an estimate of the irregular component.

5. Compute a moving five-year standard deviation sigma (s) of the estimates of the irregular component and test the irregulars in the central year of the five-year period against 2.5s. Remove values beyond 2.5s as extreme and recompute the moving five-year s. Assign a zero weight to irregulars beyond 2.5s and a weight of 1 (full weight) to irregulars within 1.5s.

Assign a linearly graduated weight between 0 and 1 to irregulars between 2.5s and 1.5s.

6. For the first two years, the sigma limits computed for the third year are used; and for the last two years, the sigma limits computed for the third-from-last year are used. To replace an extreme ratio in either of the two beginning or ending years, the average of the ratio times its weight and the three nearest full-weight ratios for that month is taken.

7. Apply a weighted 5-term moving average to the SI ratios with extreme values replaced, for each month separately, to estimate preliminary seasonal factors.

8. Repeat step 3, applied to the factors estimated in step 7.

9. To obtain a preliminary seasonally adjusted series divide 8 into the original series.

10. Apply a 9-, 13-, or 23-term Henderson moving average to the seasonally adjusted series and divide the resulting trend-cycle into the original series to give a second estimate of the SI ratios. (In the first iteration, only the 13-term Henderson is applied.)

11. Apply a weighted 7-term moving average (3 × 5 m.a.) to each month's SI ratios separately, to obtain a second estimate of the seasonal component.

12. Repeat step 3.

13. Divide 11 into the original series to obtain the seasonally adjusted series.

The Seasonal Adjustment of Composite Series

In the previous discussion, the 13 steps apply to a typical seasonal adjustment of an individual time series. The following discussion considers the seasonal

adjustment of composite series.

By composite series is here understood a series that results from the addition, subtraction, multiplication and/or division of several component series. These component series can enter into the composite with equal or different weights. Because of non-linearities involved in the process of composing the series by multiplication and division and in their seasonal adjustment method, the direct and indirect seasonally adjusted composites are usually different. The direct seasonal adjustment consists of making the composite of the unadjusted components, and then seasonally adjusting the composite series. The indirect seasonal adjustment consists of first seasonally adjusting each component series and then obtaining the seasonally adjusted composite series by implication. In order to decide whether the composite series should be seasonally adjusted using the direct or the indirect procedure the criterion of smoothness is often used. A classical measure of the degree of roughness or lack of smoothness in a seasonally adjusted composite series is the sum of squares of the first difference of the series. That is:

$$R_1 = \sum_t (\hat{X}_t - \hat{X}_{t-1})^2$$

where $\hat{X}_t$ is the series in question. The larger R_1 the rougher the series $\hat{X}_t$ or, equivalently the less smooth. The rationale of this measure is that the first difference filter removes most of the variations of long periodicities (trend and cycle). Lothian and Morry (1977) have found that the R_1 measure is related to the magnitude of the revisions in the seasonally adjusted series. The implicit definition of smoothness of R_1, however, excludes cycles of short periodicities and to compensate for this a new measure of roughness R based on the 13-term Henderson filter is given in Dagum (1979). The R_2 measure is:

$$R_2 = \sum_t (\hat{X}_t - H\hat{X}_t)^2 = \sum_t \{(I-H)\hat{X}_t\}^2$$

where (I-H) is the complement of the Henderson filter.

These two measures, expressed as averages and, in percentages when the composition is multiplicative, have been incorporated in PROC X11ARIMA and used for the direct and the indirect seasonal adjustment of composite series. Generally, both

measures give consistent results in favouring one procedure over the other from the viewpoint of smoothness. However, this consistency is not present when the composite series are strongly affected by cyclical variations of short periodicity and, in such cases, R_2 should be preferred in deciding which of the two procedures gives the smoothest seasonally adjusted data.

SPECIFICATIONS

The following statements are used to invoke and control **PROC X11ARIMA**.

```
    PROC X11ARIMA options;
         VAR variables;
         ID  variables;
         BY  variables;
         ARIMA options;
         MONTHLY  options;
         PDWEIGHTS options;
         MACURVES  options;
         QUARTERLY  options;
         COMPOSE options;
         OUTPUT OUT=SAS dataset Table1=v1 Table2=v2 . . .;
```

Either the **MONTHLY** or the **QUARTERLY** statement must be specified, according to the type of time series data. **PDWEIGHTS** can be used with the **MONTHLY** but not **QUARTERLY** statement.

PROC X11ARIMA Statement

`PROC X11ARIMA options;`

The following options may appear in this statement:

DATA=SAS dataset	specifies the name of the SAS data set to be used by **PROC X11ARIMA**. If it is omitted, the most recently created SAS data set is used.
COMPOSITE	specifies that each of the series identified on the VAR

statement is a component of a composite series. When the **COMPOSITE** statement is specified, computations appropriate to composite series will be carried out. For an ordinary series, this statement should be omitted.

VAR Statement

VAR variables;

The statement specifies the variables (names of the series) in the input data set that are to be analyzed by the procedure in the statement. Only numeric variables may be specified.

ID Statement

ID variables;

This statement is useful if an output data set is created since ID variables will be stored along with adjusted time series. They have no effect when an output data set is not created.

BY Statement

BY variables;

A BY statement may be used with **PROC X11ARIMA** to obtain separate analyses on observations in groups defined by the **BY** vairables. When a **BY** statement appears, the procedure expects the input data set to be sorted in order of **BY** variables.

ARIMA Statement

ARIMA options;

The following options may be used with the **ARIMA** statement:

AUTOMATIC requests automatic selection of ARIMA models by the

	procedure. If the test is positive, extrapolated values are incorporated into unadjusted series.
NONE	specifies no ARIMA option. This is almost equivalent to the PROC X11 with additional development incorporated at Statistics Canada. This is also equivalent to no **ARIMA** statement being specified.
USER	requests the user selected model to be used. In this case, complete information on the model must be specified using the **ARIMA** statement.
SMALL_P=i	specifies p in the ARIMA (p,d,q)(P,D,Q)s. Associated with p, i (0<=i<=4) autoregressive parameters phi1, phi2,...phii are defined in the model.
SMALL_D=j	specifies d in the ARIMA (p,d,q)(P,D,Q)s. The number of ordinary differences applied is j (0<=j<=4).
SMALL_Q=k	specifies q in the ARIMA (p,d,q)(P,D,Q)s. Associated with q, k (0<=k<=4) moving average parameters theta1, theta2,...,thetak are defined in the model.
BIG_P=m	specifies P in the ARIMA (p,d,q)(P,D,Q)s. Associated with P, m (0<=m<=4) seasonal autoregressive parameters PHI1, PHI2, ..., PHIm are defined in the model.
BIG_D=n	specifies D in the ARIMA (p,d,q)(P,D,Q)s. The number of seasonal differences applied is n (0<=n<=4).
BIG_Q=r	specifies Q in the ARIMA (p,d,q)(P,D,Q)s. Associated with Q, r (0<=r<=4) seasonal moving average parameters THETA1, THETA2, ...,THETAr are defined in the model.
DETERM=i	indicates the nature of the deterministic constant term in the model. If i=1, phi0 is in the model, and i=2, theta0. If i=0, it has the same effect as omitting **DETERM=**.
INIT_VALUE0=c0	specifies the initial value for either phi0 or theta0

depending on i of **DETERM**=i.

INIT_VALUE1=c1 respectively specifies the initial values for:

.   phi1, phi2, ..., phii,
.   theta1, theta2,..., thetak,
.   PHI1, PHI2, ..., PHIm,
.   THETA1, THETA2,..., THETAr.

INIT_VALUEx=cx The total number of initial values must not exceed 10: i+k+m+r<=10 without constant term or i+k+k+r<=9 with constant term. If none is specified, the default value 0.1 is used.

ORDER1=h1 specifies the order of B (backshift operator) corresponding to:

ORDER2=h2   phi1, phi2, ..., phii,
.   theta1, theta2, ..., thetaj,
.   PHI1, PHI2, ..., PHIm,
.   THETA1, THETA2, ..., THETAr.

ORDERx=hx If omitted, the following orders are used:

1, 2, ..., i,
1, 2, ..., j,
1, 2, ..., k,
1, 2, ..., m.

ADD_CON=a specifies the additive constant for the transformation (Z+a)**b where Z is the variable name of the series to be transformed, and b the exponent.

POWER=b specifies the exponent b of the power transformation (Z+a)**b. If b=0, (Z+a) will be transformed using natural logarithm.

MAX_ITR=i specifies the maximum number of iterations desired. If this option is omitted, 30 becomes the default value.

SERIES_ID=variable specifies the name of the variable to which the **ARIMA**

statement applies. This option is only permitted when the **COMPOSITE** option is in effect.

MONTHLY Statement

MONTHLY options;

The MONTHLY statement must be used when the series to be analyzed is a monthly time series. There must be at least 36 observations (3 years) and no more than 360 observations (30 years) in the input data set, or within a BY group if it is used.

The following options may appear in the **MONTHLY** statement:

START=mmmyy	specifies that the series starts at month mmm (e.g. JAN) and year yy (e.g. 66). START=JAN66 denotes the series starting January 1966.
END=mmmyy	specifies that the series ends at mmmyy. END=DEC70 denotes the series ending December 1970.
TYPE=ADDITIVE	specifies additive adjustment,
TYPE=MULTIPLICATIVE	specifies multiplicative adjustment,
TYPE=LOGARITHMIC	specifies logarithmic adjustment. If TYPE= is omitted, the multiplicative adjustment is performed.
NDEC=n	requests n decimal places to be printed for most output tables. (In the multiplicative model, trading-day adjustment factors on Tables C16 and C18, seasonal factors on Table D10 and combined factors on Table D16 are shown with standard output only).
ADJTOTAL	requests the adjustment of the yearly total of the seasonally adjusted series equal to that of the original series.
SUMMARY	requests the development of the summary measures (estimates of the trend-cycle, irregular, I/C, MCD, and

residual trading- day and seasonal variation) from a seasonally adjusted. If **SUMMARY** is omitted, composite series are seasonally adjusted and then summary measures are developed.

PRINTOUT=STANDARD	requests standard printout. From 19 to 31 tables are printed depending on which other options are selected.
PRINTOUT=BRIEF	requests brief printout. From 3 to 5 tables are printed (A1, B1, D10, D11 and D16).
PRINTOUT=ANALYSIS	requests analysis printout. From 7 to 13 tables are printed (A1, D, E, and F tables).
PRINTOUT=SHORT	requests short printout. From 7 to 13 tables are printed (mainly D and F tables).
PRINTOUT=LONG	requests long printout. From 28 to 42 tables are printed.
PRINTOUT=FULL	requests full printout. From 45 to 62 tables are printed.
CHARTS=STANDARD	requests standard charts. The original series, 12 monthly seasonal charts and the trend-cycle charts are printed.
CHARTS=NONE	requests no charts.
CHARTS=ALL	requests all charts, i.e. 12 monthly seasonal charts and charts of the original series, trend cycle, irregular, seasonal factors and the Kolmogorov-Smirnov cumulative periodogram.
LSIGMA=x.x	spcefies the lower sigma limit for graduating extreme values in estimating seasonal and trend-cycle componenets. Irregulars will be assigned full weights within the lower sigma limit entered here. If it is omitted, full weights will be assigned within 1.5 s.
USIGMA=x.x	specifies the upper sigma limit for graduating extreme values. Irregulars will be assigned zero weights outside the upper sigma limit entered here. If it is omitted, zero weights will be assigned to irregular values outside 2.5 s.

LENGTH	requests the length of month variation allowance in the prior and/or trading-day factors rather than in the seasonal factors. Divisors for all month are 30.4375, the average length of a month. If the **LENGTH** is omitted, the length-of-month variations are included in the seasonal factors. Divisors used in the construction of monthly weights are 31, 30 and 28.25 for 31 and 30 day months and February, respectively.
TDREGR=NONE	will exclude the computation of trading-day regression.
TDREGR=PRINT	will compute and print the results.
TDREGR=TEST	will compute the trading-day regression and print the results and use the estimates only if they explain significant variation on the basis of an F test.
TDREGR=ADJUST	will compute the trading-day regression, print the results and adjust the series by the regression estimates. If prior factors have been supplied, they will be corrected on the basis of these estimates.
TDAPPLY=yy	applies the trading-day regression estimates only to the part of the series beginning with January of the year yy. If prior weights are supplied, adjustment is made to the part of the series preceding this date by the prior weights only, and the part of the series from this date to the end is adjusted by the prior weights corrected by the regression estimates. If yy is not specified, the trading-day regression estimates or prior trading-day weights corrected by regression estimates will be applied to the entire series.
TDCOMPUTE=yy	derives the estimates of the trading-day weights using only the part of the series beginning with January of the year enterd here as input to the regression. If it is not specified, estimates of the trading-day weights are

derived using the entire series as input to the regression.

EXCLUDE=x.x	specifies that the irregular values beyond x.x s limit are to be excluded. If it is not specified, the default value is 2.5 s.
TRENDMA=9 TRENDMA=13 TRENDMA=23	specifies the selection of a moving average for variable trend-cycle routine. Either a 9-term, 13-term or 23-term Henderson may be selected. If it is omitted, an appropriate selection will be made from the three listed here.
TRENDADJ	Modification of extreme values may be made before computing the trend-cycle estimate. This adjustment for extremes substantially reduces the effect of major prolonged strikes or similar irregular occurences on the Table B7 and subsequent trend-cycle estimates. Care should be exercised in its use, however, since for some series the estimates near sharp business cycle peaks or troughs will be similarly affected. If omitted, computes the B7 trend-cycle curve without strike adjustment.
PTRENDMA=12 PTRENDMA=24	specifies the 12-term or 24-term centered moving average, respectively. This option allows preliminary estimation of the trend-cycle component.
EXTRAPOLATION=0 EXTRAPOLATION=1 EXTRAPOLATION=2 EXTRAPOLATION=3 EXTRAPOLATION=4 EXTRAPOLATION=5 EXTRAPOLATION=6 EXTRAPOLATION=7 EXTRAPOLATION=8	The **EXTRAPOLATION=** specifies one year of forecasts and backcasts using **ARIMA** extrapolated values at the beginning (backcasts) and or at the end (forecasts) of the series and it can be applied only to series of at least five complete years. For series longer than 15 complete years, only the last 15 years will be used to fit the model. The extrapolation values are printed only in TABLE B1 and not used for the calculations of the summary measures.

If **EXTRAPOLATION=0** is specified, there will be no extrapolation.

If **EXTRAPOLATION=1** is specified, three **ARIMA** models will be automaticcally fitted to the unadjusted series. The model giving the smallest average extrapolation error for the last three years is chosen to produce one year of extrapolated values at both ends of the series. None of the models is selected and, therefore, no extrapolation is made if : (a) the absolute average error for the last three years is greater than 12% for the forecasts or 18% for the backcasts; or (b) the chi-squared probability is smaller than 10% or (c) there are signs of over-differencing. If the above criteria failed marginally, the user can still apply the model that gives the smallest extrapolation error by resubmitting the series with the option where the user provides his own model.

If **EXTRAPOLATION=2** is specified, a model chosen by the user will be fitted to the unadjusted series and the extrapolated values will be used even if the model does not pass the above acceptance criteria. The user must choose a model through the **ARIMA statement** specifying (p,d,q)(P,D,Q)s.

If **EXTRAPOLATION=3** is specified, the effect is similar but the extreme values of the original series are automatically replaced by their corresponding function values of the **ARIMA** model chosen. This option should be used when the unadjusted series is strongly affected by outliers to avoid a bad extrapolation and a poor estimation of the seasonal factors. The replacement of the extreme values is not made for 2(p+Pxs+d+Dxs) observations at the beginning of the series. This

means that for a monthly ARIMA (0,1,1)(0,1,1)s model, no replacement of extremes is made for the first 26 months of the series.

If **EXTRAPOLATION=4** is specified, the effect is similar to **EXTRAPOLATION=2** but the extreme values of the original series are automatically replaced by their corresponding function values of the ARIMA model chosen. This option should be used when the unadjusted series is strongly affected by outliers to avoid a bad extrapolation and a poor estimation of the seasonal factors. The replacement of the extreme values is not made for 2(p+Pxs+d+Dxs) observations at the beginning of the series. This means that for a monthly ARIMA (0,1,1)(0,1,1)s model, no replacement of extremes is made for the first 26 months of the series.

If **EXTRAPOLATION=5** is specified, the effect is similar to **EXTRAPOLATION=1** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=6** is specified, the effect is similar to **EXTRAPOLATION=2** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=7** is specified, the effect is similar to **EXTRAPOLATION=3** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=8** is specified, the effect is similar to **EXTRAPOLATION=4** but the **ARIMA** model is used to generate only forecasts.

PMFACTOR=variable specifies Prior **Monthly Adjustment Factors**. This option is used to specify whether or not a prior adjustment is required. The prior factors are divided into the original

	data, before the multiplicative or logarithmic seasonal adjustment process. They are subtracted from the original series before an additive adjustment. If this option is omitted, no prior monthly adjustment will take place.
SERIES_ID=variable	specifies the name of the variable to which the **MONTHLY** statement applies. This option is only permitted when the **COMPOSITE** option is in effect.

PDWEIGHTS Statement

PDWEIGHTS options;

The **PDWEIGHTS** statement can be used to specify one to seven Prior Daily Weights. This option is available only with multiplicative or logarithmic adjustment for monthly time series data. Seven daily weights may be entered to adjust the series for trading-day variation prior to the seasonal adjustment process. The seven weights are combined to yield the prior trading-day adjustment factors shown in Table A4. The weights are adjusted to total 7.0. These weights may be modified by the trading-day regression routine. The option below may appear on the **PDWEIGHTS** statement:

day=x.xxx	specifies a weight x.xxx for a given day. The day is any day of the week and x.xxx must be a numeric value between 0.000 and 9.999. (e.g. SUNDAY=3.500). Any number of days can be specified with one **PDWEIGHTS** statement. The default weight value for any day that is not specefied is zero.
SERIES_ID=variable	specifies the name of the variable to which the **PDWEIGHTS** statement applies. This option is only permitted when the **COMPOSITE** option is in effect.

MACURVES Statement

```
MACURVES options;
```

For monthly or quarterly series, the **MACURVES** statement can be used to select the length of moving average curves for estimating the seasonal factors for a given month or quarter. If the **MACURVES** statement is omitted, a 3-by-3 moving average is selected for the first estimate of the seasonals in each iteration and a 3 by 5 in the final estimate. However, for series shorter than five complete years, the program chooses only the stable seasonality option and the user has no control over it. The following options may appear in the **MACURVES** statement:

AUTOSEL	specifies the selection of different moving averages for each month or quarter. The user has no control over the selection procedure.
month=specification	specifies the moving average of a given month for monthly data. The month (or the first three letters) is followed by an equal sign, and then one of the values given below: '3x3' specifies a 3 by 3 moving average. '3x5' specifies a 3 by 5 moving average. '3x9' specifies a 3 by 9 moving average. STABLE specifies stable seasonal factor (average of all values for the month). For example, FEB='3x3' selects 3 by 3 moving average in all iterations. If the specification for the month is omitted, a 3 by 3 in the first iteration is followed by 3 by 5 in the second iteration unless the series is shorter than five complete years.
quarter=specification	specifies the moving average of a given quarter for quarterly data. For example, Q1='3x3' selects 3 by 3 moving average in all iterations. If the specification for a quarter is omitted, a 3 by 3 in the first iteration

is followed by 3 by 5 in the second iteration unless the series is shorter than five years.

SERIES_ID=variable	specifies the name of the variable to which the **MACURVES** statement applies. This option is only permitted when the **COMPOSITE** option is in effect.

QUARTERLY Statement

QUARTERLY options;

The **QUARTERLY** statement must be used when the series to be analyzed is a quarterly time series. There must be at least 12 observations (3 years) and no more than 120 observations (30 years) in the input data set, or within a BY group if it is used.

The following options may appear in the **QUARTERLY** statement:

START='yyQi'	specifies that the series starts in year yy and quarter i. The quarter must be specified as Q1, Q2, Q3 or Q4 and entire specification enclosed in single quotes, e.g. START='79Q1'.
END='yyQi'	specifies that the series ends in year yy and quarter i. END='80Q4' denotes the series ending in the 4th quarter 1980.
TYPE=ADDITIVE	specifies additive adjustment,
TYPE=MULTIPLICATIVE	specifies multiplicative adjustment,
TYPE=LOGARITHMIC	specifies logarithmic adjustment. If TYPE= is omitted, the multiplicative adjustment is performed.
NDEC=n	requests n decimal places to be printed for most output tables. In the multiplicative version, seasonal factors on Table D10 are shown with two decimal places in the regular output only. Tables of SI ratios are shown with one decimal place.

ADJTOTAL	requests the adjustment of the yearly total of the seasonally adjusted series to be equal to that of the original series.
SUMMARY	requests the development of the summary measures (estimates of the trend-cycle, irregular, I/C, QCD, and residual seasonal variation) from a seasonally adjusted. If **SUMMARY** is omitted, composite series are seasonally adjusted and then summary measures are developed.
PRINTOUT=STANDARD	requests standard printout. From 19 to 31 tables are printed depending on other options selected.
PRINTOUT=BRIEF	requests brief printout. From 3 to 4 tables are printed (A1, D10, and D11).
PRINTOUT=ANALYSIS	requests analysis printout. From 7 to 13 tables are printed (A1, D, E, and F tables).
PRINTOUT=SHORT	requests short printout. From 7 to 13 tables are printed (mainly D and F tables).
PRINTOUT=LONG	requests long printout. From 28 to 42 tables are printed.
PRINTOUT=FULL	requests full printout. From 45 to 62 tables are printed.
CHARTS=STANDARD	requests standard charts. The original series, 4 quarterly seasonal charts and the trend-cycle charts are printed.
CHARTS=NONE	requests no charts.
CHARTS=ALL	requests all charts, i.e. 4 quarterly seasonal charts and charts of the original series, trend cycle, irregular, seasonal factors and the Kolmogorov-Smirnov cumulative periodogram.
LSIGMA=x.x	spcefies lower sigma limit for graduating extreme values in estimating seasonal and trend-cycle components. Irregulars will be assigned full weights within the lowere sigma limit entered here. If it is omitted, full weights

	will be assigned within 1.5 s.
USIGMA=x.x	specifies upper sigma limit for graduating extreme values. Irregulars will be assigned zero weights out side the upper sigma limit specified here. If it is omitted, zero weights will be assigned to irregulars outside 2.5s.
TRENDMA=5 **TRENDMA=7**	specifies the selection of a moving average for the variable trend-cycle routine. Either a 5-term or 7-term Henderson may be selected. If it is omitted, an appropriate selection will be made from the two listed above.
TRENDADJ	Modification of extreme values may be made before computing the trend-cycle estimate. This adjustment for extremes substantially reduces the effect of major prolonged strikes or similar irregular occurences on the B7 and subsequent trend-cycle estimates. Care should be exercised in its use, however, since for some series the estimates near sharp business cycle peaks or troughs will be similarly affected. If omitted, the B7 trend-cycle curve is computed without strike adjustment.
PTRENDMA=4 **PTRENDMA=8**	specifies 4-term or 8-term centered moving average, respectively. This option allows preliminary estimation of a trend-cycle component. If it is omitted, 4-term is assumed.
EXTRAPOLATION=0 **EXTRAPOLATION=1** **EXTRAPOLATION=2** **EXTRAPOLATION=3** **EXTRAPOLATION=4** **EXTRAPOLATION=5** **EXTRAPOLATION=6** **EXTRAPOLATION=7**	The **EXTRAPOLATION=** specifies one year of forecasts and backcasts using **ARIMA** extrapolated values at the beginning (backcasts) and or at the end (forecasts) of the series and it can be applied only to series of at least five complete years. For series longer than 15 complete years, only the last 15 years will be used to fit the model. The extrapolation values are printed only in TABLE B1 and not used for the

EXTRAPOLATION=8 calculations of the summary measures.

If **EXTRAPOLATION=0** is specified, there will be no extrapolation.

If **EXTRAPOLATION=1** is specified, three built-in **ARIMA** models will be automatically fitted to the unadjusted series. The model giving the smallest average extrapolation error for the last three years is chosen to produce one year of extrapolated values at both ends of the series. None of the models is selected and, therefore, no extrapolation is made if : (a) the absolute average error for the last three years is greater than 12% for the forecasts or 18% for the backcasts; or (b) the chi-squared probability is smaller than 10% or (c) there are signs of over-differencing. If the above criteria failed marginally, the user can still apply the model that gives the smallest extrapolation error by resubmitting the series with the option where the user provides his own model.

If **EXTRAPOLATION=2** is specified, a model chosen by the user will be fitted to the unadjusted series and the extrapolated values will be used even if the model does not pass the above acceptance criteria. (a model must be chosen through the **USER** options of the **ARIMA** statement).

If **EXTRAPOLATION=3** is specified, the effect is similar but the extreme values of the original series are automatically replaced by their corresponding function values of the **ARIMA** model chosen. This option should be used when the unadjusted series is strongly affected by outliers to avoid a bad extrapolation and a poor estimation of the seasonal factors. The replacement of

the extreme values is not made for 2(p+Pxs+d+Dxs) observations at the beginning of the series. This means that for a quarterly ARIMA (0,1,1)(0,1,1)s model, no replacement of extremes is made for the first 10 quarters of the series.

If **EXTRAPOLATION=4** is specified, the effect is similar to **EXTRAPOLATION=2** but the extreme values of the original series are automatically replaced by their corresponding function values of the ARIMA model chosen. This option should be used when the unadjusted series is strongly affected by outliers to avoid a bad extrapolation and a poor estimation of the seasonal factors. The replacement of the extreme values is not made for 2(p+Pxs+d+Dxs) observations at the beginning of the series. This means that for a quarterly ARIMA (0,1,1)(0,1,1)s model, no replacement of extremes is made for the first 10 quarters of the series.

If **EXTRAPOLATION=5** is specified, the effect is similar to **EXTRAPOLATION=1** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=6** is specified, the effect is similar to **EXTRAPOLATION=2** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=7** is specified, the effect is similar to **EXTRAPOLATION=3** but the **ARIMA** model is used to generate only forecasts.

If **EXTRAPOLATION=8** is specified, the effect is similar to **EXTRAPOLATION=4** but the **ARIMA** model is used to generate only forecasts.

PQFACTOR=variable specifies **Prior Quarterly Adjustment Factors**. This option

is used to specify whether or not a prior adjustment is required. The prior factors are divided into the original data, before the multiplicative or logarithmic seasonal adjustment process. They are subtracted from the original series before an additive adjustment. If this option is omitted, no prior quarterly adjustment will take place.

SERIES_ID=variable	specifies the name of the variable to which the **QUARTERLY** statement applies. This option is only permitted when the **COMPOSITE** option is in effect.

COMPOSE Statement

COMPOSE options;

The statement is optionally used to specify: (a) a component series, (b) corresponding operatin (add, subtract, multiply, devide) and (c) weight for the series, when the **COMPOSITE** option is in effect in the **PROC X11ARIMA** statement. If the **COMPOSITE** option is in effect but no **COMPOSE** statement is used, then the default operation (i.e. addition) and the default weight (1.0) are used for all the component series specified on the **VAR** statement. Any number of **COMPOSE** statement may be used. The following options may appear in the **COMPOSE** statement.

SERIES_ID=variable	specifies the name of the variable corresponding to the component series.
C_OPERATION=ADD	specifies that the series identified by C_SERIES= enters into the composite through addition.
C_OPERATION=SUB	specifies that the series identified by C_SERIES= enters into the composite through subtraction.
C_OPERATION=MULT	specifies that the series identified by C_SERIES= enters into the composite through multiplication.
C_OPERATION=DIV	specifies that the series identified by C_SERIES= enters into the composite through division.
C_WEIGHT=xx.xxx	specifies a constant xx.xxxx by which the component series

is multiplied before the composite operation is performed.

OUTPUT Statement

```
OUTPUT OUT=SASdataset table=variable . . .;
```

The statement is used for creating an output data set. Any value of the time series calculated by the procedure may be written to the new data set. The following can be specified on the **OUTPUT** statement:

OUT=SAS dataset	specifies a name for the output data set. Following the appropriate rules for naming SAS data sets, a temporary or permanent data set may be created. Ommitting the OUT= statement is equivalent to specifying **OUT=_DATA_**.
table=variable ...	specifies the table for the time series to be included, and variables to be included in the output data set. The names used as variables must be found on the **VAR** list. If an **ID** statement or **BY** statement is used with the procedure, the value of the vairables appearing on these statements will appear in the output data set. The table numbers that may be used in the **OUTPUT** statement are given in the next section **NOTES ON USE**.

NOTES ON USE

Missing Values

PROC X11ARIMA does not permit missing values in the input data. If a missing value is found in the variable (series) being analysed, the procedure prints a message and goes on to the next series. If a prior factor (PMFACTOR= or QMFACTOR=) is specified, values of these variable are also checked for missing values. If a missing value is found in either variable (series), the procedure terminates the execution of current BY group.

Missing values may occur in the output data set. For example, if the moving averages are computed, the first and the last n observations may be missing (to be checked later)

Input Data Set

Unlike most other SAS procedures, **PROC X11** and **PROC X11ARIMA** limit the number of observations processable to 360 observations for the **MONTHLY** data and 120 for the **QAURTERLY** data. If "forecasting" or "backcasting" is in effect, the limit also applies, i.e. one or two years must be included within the 30-year limit.

Output Data Set

Any table identified below can be placed on the output data set so long as they are computed by the procedure and printed.

Printed Output

The printed output and corresponding options are summarized in the **Matrix of Options and Tables**. (The matrix to be attached later) The user may choose appropriate options to ensure that desired Tables are computed and placed on the output data set.

OPERATING SYSTEM AND JCL

The PROC X11ARIMA is intended to be used just like any other SAS procedure which runs under the IBM Operating System. The current release of PROC X11ARIMA is tested under IBM OS MVS/XA. The procedure uses SAS 82.4 Version FORTRAN interface conventions. In other words, the procedure is implemented using the information available in the following sources:

(a) SAS Programmer's Guide - 1981 Edition,
(b) SAS Views: Procedure Writing - 1981 Edition,
(c) SAS MACRO Library on the SAS 82.4 Release Distribution Tape.

When the Version 5 becomes available at Statistics Canada, it will be tested again but no major problem will likely occur.

Depending upon the future needs, the procedure may be adapted to CMS Operating System without problems. Since the CMS Operating System is not available at Statistics Canada, testing under CMS will wait until suitable opportunities become available.

To use the procedure under the IBM Operating System, the load module of the procedure must be installed and concatenated to STEPLIB. At Statistics Canada, the following alternatives are available for the batch job submission:

```
//job card information
//PROCLIB DD DSN=STC2.SAS.PROC,DISP=SHR
//      EXEC SAS
//STEPLIB DD DSN=TISE.SX11A.LOAD,DISP=SHR
//
//
//
//
          (SAS statements)

        DATA steps and PROC steps

/*
```

Alternatively,

```
//job card information
//PROCLIB DD DSN=TISE.SX11A.PROC,DISP=SHR
//   EXEC SAS

       (SAS statements)

       Data steps and PROC steps
/*
```

REFERENCES

Box, G.E.P. and Jenkins, G.M. (1970), Time Series Analysis Forecasting and Control, Holden-Day, San Fransisco.

Dagum, E.B. (1979), The Seasonal Adjustment of Economic Time Series Aggregates: A Case Study: The Unemployment Rate. Background Paper No. 31, The U.S. National Commission of Employment and Unemployment Statistics. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Dagum, E.B. (1980), The X-11-ARIMA Seasonal Adjustment Method, Statistics Canada Cat. No. 12-564E (in English), Cat. No. 12-564F (in French), Ottawa, Canada.

Dagum, E.B. (1981), "Diagnostic Checks in the ARIMA Models of the X-11-ARIMA Seasonal Adjustment Method," Time Series Analysis (O. Anderson and M.R. Perryman, eds.) North-Holland, pp. 133-145.

Lothian, J, and Morry, M. (1977), The Problem of Aggregation: Direct or Indirect Seasonal Adjustment, Research Paper, Time Series Research and Analysis Division, Statistics Canada, Ottawa, Canada.

SAS Institute (1982), SAS/ETS User's Guide: Econometric and Time Series Library - 1982 Edition, SAS Institute, Cary, NC.

SAS Institute (1981), SAS Programmer's Guide - 1981 Edition, SAS Institute, Cary, NC.

SAS Institute (1981), SAS Views: Procedure Writing - 1981 Edition, SAS Institute, Cary, NC.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967), The X-11 Variant of the Census Method II Seasonal Adjustment Program, Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce, U.S. Government Printing Office, Washington, DC.