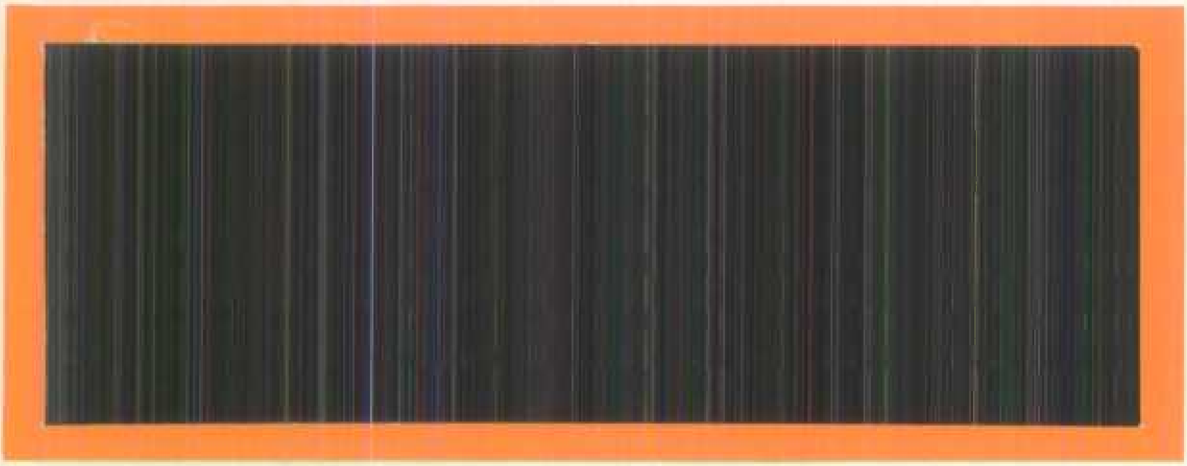
Statistics Canada Statistique Canada



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-86-017E

TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY BRANCH

USER'S MANUAL FOR PROC X11ARIMA

by

Misako Kozono Gratton and Estela Bee 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 Do not quote without authors' permission.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.

DEVELOPMENT OF THE X-11-ARIMA SEASONAL ADJUSTMENT PROGRAM

E.B. Dagum, Statistics Canada

PARSER DEVELOPMENT

M. Gratton in consultation with D. Delong, SAS Institute

FORTRAN INTERFACE DEVELOPMENT

M. Gratton, Statistics Canada

R. May, NORTAK SOFTWARE LTD.

T. Sutanto, NORTAK SOFTWARE LTD.

H. Fung, Statistics Canada

METHODOLOGICAL AND TECHNICAL CONTRIBUTION

P.A. Cholette, Statistics Canada

E.B. Dagum, Statistics Canada

H. Fung, Statistics Canada

M. Gratton, Statistics Canada

J. Higginson, Statistics Canada

M. Morry, Statistics Canada

A. Papineau, Statistics Canada

D. Piper, Statistics Canada

S. Tamasi, Statistics Canada

TABLE OF CONTENTS

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.

BACKGROUND

The main objective of developing the PROC X11ARIMA is to offer SAS users the facilities available in the X-11-ARIMA Seasonal Adjustment Method (Dagum, 1980). This method is used by organizations worldwide as a tool for seasonal adjustment of monthly and quarterly time series. In particular, a significant number of statistical offices has adopted it as the official method of seasonal adjustment for socio-economic time series.

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 (ref. PROC X11 in SAS/ETS User's Guide) 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.

Optionally, the procedure can produce a seasonally adjusted composite either directly or indirectly from a group of seasonally unadjusted components. Thus, the procedure functions exactly like the original X-11-ARIMA Seasonal Adjustment Program. However, in place of its control cards, appropriate SAS statements are used to describe user's problems. In terms of reading input data, the procedure expects a SAS data set, which can be created using appropriate SAS statements. Since SAS can read almost any data set in an IBM OS environment, reading input data is expected to be simplified considerably.

This manual assumes that the reader is familiar with SAS and the notation used in either Dagum (1980) or Gratton and Dagum (1986). Although the manual is presently available only in English, French translation will be available in the future. Research papers on the related subject have been published in several languages including English, French and Japanese. The interested persons may contact:

Dr. E.B. Dagum
Director
Time Series Research and Analysis Division
Statistics Canada
Ottawa, Ontario, Canada
K1A OT6

Telephone (613) 990-9876

PROC X11ARIMA was developed using the interface conventions specified in SAS Programmer's Guide - 1981 Edition. It is running at Statistics Canada under SAS Release 82.4. It also runs under SAS Version 5.

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** option is specified, computations appropriate to composite series will be carried out. The option should be omitted, unless compositing is required.

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;

This statement is necessary only if **EXTRAPOLATION=** (in the **MONTHLY** or **QUARTERLY** statement) is either 2, 4, 6 or 8. This statement is used to specify an ARIMA model to be used by the extrapolation procedure.

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.
AMODEL=(p,d,q)(P,D,Q)	specifies the ARIMA (p,d,q)(P,D,Q)s. The p (0<=p<=4) ordinary autoregressive parameters phi1, phi2,...phip are defined, the d (0<=d<=4) ordinary differences are applied, and the q (0<=q<=4) ordinary moving average parameters theta1, theta2,...,thetaq are defined in the model. Similary, the P (0<=P<=4) seasonal autoregressive parameters PHI1, PHI2, ..., PHIP are defined, the D (0<=D<=4) seasonal differences applied, and the Q (0<=Q<=4) seasonal moving average parameters THETA1, THETA2, ...,THETAQ are defined in the model. (An abbreviation **AM=** may be used in place of **AMODEL=**).
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=**. (An abbreviation DET= may be used in place of DETERM=).
INITV=(c0,c1,...,ck)	The c0 specifies the initial value for either phi0 or theta0 depending on i of **DETERM=i**. The c1, c2,...,ck

	(k<=9) respectively specifies the initial values for: phi1, phi2, ..., phip, theta1, theta2, ..., thetaq, PHI1, PHI2, ..., PHIP, THETA1, THETA2,...., THETAQ. If none is specified, the default value 0.1 is used. (An abbreviation **INIT=** may be used in place of **INITV=**).
ORDER=(h1,h2,...,hi)	specifies the i-th (0<=i<=10) order of B (backshift operator) corresponding to: phi1, phi2, ..., phip, theta1, theta2, ..., thetaq, PHI1, PHI2, ..., PHIP, THETA1, THETA2, ..., THETAQ. If omitted, the following orders are used: 1, 2, ..., p, 1, 2, ..., q, 1, 2, ..., P, 1, 2, ..., Q. (An abbreviation **ORD=** may be used in place of **ORDER=**).
PTRANS=(a,b)	specifies the additive constant a and power b for the transformation (Z+a)**b where Z is the variable name of the series to be transformed, and b the exponent. If b=0, (Z+a) will be transformed using natural logarithm. If no transformation is required, the parameter should be omitted. (An abbreviation **PT=** may be used in place of **PTRANS=**).
MAX_ITR=i	specifies the maximum number of iterations desired. If this option is omitted, 30 becomes the default value. (An abbreviation **MAX=** may be used in place of **MAX_ITR=**).
SERIES_ID=variable	specifies the name of the variable to which the **ARIMA**

statement applies. This parameter is meaningful only when the **COMPOSITE** option is in effect. (In place of **SERIES_ID** , **SER_ID** or **S_ID** may be used as an abbreviation).

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.

In the **MONTHLY** statement, **START=** and **END=** are mandatory, while others are optional.

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. (An abbreviation ST= may be used in place of **START=**).
END=mmmyy	specifies that the series ends at mmmyy. END=DEC70 denotes the series ending December 1970. (An abbreviation EN= may be used in place of **END=**).
TYPE=ADDITIVE	specifies additive adjustment,
TYPE=**MULTIPLICATIVE**	specifies multiplicative adjustment,
TYPE=LOGARITHMIC	specifies logarithmic adjustment. If TYPE= is omitted, the multiplicative adjustment is performed. (An abbreviation TY= may be used in place of **TYPE=**).
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). (An abbreviation **ND=** may be used in place of **NDEC=**).

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. If **PRINTOUT=** is omitted, **STANDARD** is assumed. (An abbreviation **PR=** may be used in place of **PRINTOUT=**).
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. (An abbreviation of **CH=** may be used in place of **CHARTS=**).
LSIGMA=x.x	specifies 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. (An abbreviation **LS=** may be used in place of **LSIGMA=**).
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. (An abbreviation **US=** may be used in place of **USIGMA=**).
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. Default is **TDREGR=NONE**. (An abbreviation **TDR=** may be used in place of **TDREGR=**).
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. (An abbreviation **TDA=** may be used in place of **TDAPPLY=**).
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. (An abbreviation **TDC=** may be used in place of **TDCOMPUTE=**).
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. (An abbreviation **TR=** may be used in place of **TRENDMA=**).
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. Default is 12-term centred moving average. (An abbreviation PTR= may be used in place of **PTRENDMA=**).

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 cf 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 are not used for the calculations of the summary measures. (An abbreviation EXT= may be replaced by **EXTRAPOLATION=**).

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. (An abbreviation PM= may be used in place of PMFACTOR=).

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. (In place of SERIES_ID= , either SER_ID or S_ID may be used as an abbreviation).

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. (In place of **SERIES_ID=** , either **SER_ID** or **S_ID** may be used as an abbreviation).

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. (In place of **SERIES_ID** , either **SER_ID** or **S_ID** may be used as an abbreviation).

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.

In the **QUARTERLY** statement, **START=** and **END=** are mandatory, while others are optional.

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'. (An abbreviation **ST=** may be used in place of **START=**).
END='yyQi'	specifies that the series ends in year yy and quarter i. END='80Q4' denotes the series ending in the 4th quarter 1980. (An abbreviation **EN=** may be used in place of **END=**).
TYPE=ADDITIVE	specifies additive adjustment.
TYPE=MULTIPLICATIVE	specifies multiplicative adjustment.
TYPE=LOGARITHMIC	specifies logarithmic adjustment. If TYPE= is omitted, the multiplicative adjustment is performed. (An abbreviation TY= may be used in place of TYPE=).
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. (An abbreviation ND= may be used in place of NDEC=).
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).
PRINTDUT=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. If PRINTOUT= is omitted, STANDARD is assumed. (An abbreviation PR= may be used in place of PRINTOUT=).
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. (An abbreviation **CH=** may be used in place of **CHARTS=**).
LSIGMA=x.x	specifies 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. (An abbreviation **LS=** may be used in place of **LSIGMA=**).
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. (An abbreviation **US=** may be used in place of **USIGMA=**).
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. (An abbreviation **TR=** may be used in place of **TRENDMA=**).
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. (An abbreviation **PTR=** may be used in place of **PTRENDMA=**).

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 are not used for the calculations of the summary measures. (An abbreviation **EXT=** may be used in place of **EXTRAPOLATION=**).

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. (An abbreviation **PQ=** may be used in place of **PQFACTOR=**).
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. (In place of **SERIES_ID=** , either **SER_ID=** or **S_ID=** may be used as an abbreviation).

COMPOSE Statement

COMPOSE options;

The statement is optionally used to specify: (a) a component series, (b) corresponding operation (add, subtract, multiply, divide) 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 the **COMPOSE** statement must be used for each of the monthly or quarterly series entering as a component. In addition, the **COMPOSE**

statement is required for the resulting composite series. (see Notes on **COMPOSITE** option). The default values for **COMPOSE** statement are **C_OPERATION=ADD** and **C_WEIGHT**=1.0.

SERIES_ID=variable	specifies the name of the variable corresponding to the component series. The **MONTHLY** and other statements must precede the **COMPOSE** statement so that a group of statements corresponding to a component will be arranged with having the same **SERIES_ID**. After all components are entered, the resulting composite series will be identified as **C_RESULTS**. (In place of **SERIES_ID** , either **SER_ID** or **S_ID** may be used as an abbreviation).
C_OPERATION=ADD	specifies that the series identified by **SERIES_ID=** enters into the composite through addition.
C_OPERATION=SUB	specifies that the series identified by **SERIES_ID=** enters into the composite through subtraction.
C_OPERATION=MULT	specifies that the series identified by **SERIES_ID=** enters into the composite through multiplication.
C_OPERATION=DIV	specifies that the series identified by **SERIES_ID=** enters into the composite through division.
C_OPERATION=END	specifies that all components are entered, and that the last group of statements refers to the resulting composite series. (An abbreviation **C_OP=** may be used in place of **C_OPERATION=**).
C_WEIGHT=xx.xxx	specifies a constant xx.xxxx by which the component series is multiplied before the composite operation is performed. (An abbreviation **C_W=** may be used in place of **C_WEIGHT=**).

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 **QUARTERLY** 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.

Notes on COMPOSITE Option

In the COMPOSITE option, a set of series with homogenous length is seasonally adjusted individually and entered into the composite operation. Suppose k monthly series S1, S2, ..., Sk are to form a composite series. The following diagram shows how each component can be entered:

```
               PROC X11ARIMA COMPOSITE ...;
               VAR S1, S2, ..., Sk;
          ---    MONTHLY ... SERIES_ID=S1 ...;
           |       .     ... SERIES_ID=S1 ...;
           |       .
          S1       .         (other optional statements) ...;
           |       .
           |       .
          ---    COMPOSE ... SERIES_ID=S1 ...;
          ---
           |       .
          S2       .
           |       .
          ---
          ---    MONTHLY ... SERIES_ID=Sk ...;
           |       .     ... SERIES_ID=Sk ...;
           |       .
          Sk       .         (other optional statements) ...;
           |       .
           |       .
          ---    COMPOSE ... SERIES_ID=Sk ...;
                 MONTHLY ...      ...      ...;
                   .
                   .
                   .         (other optional statements) ...;
                   .
```

```
COMPOSE ... C_OPERATION=END;
```

The above diagram shows that SAS statements must be entered in groups beginning with MONTHLY or QUARTERLY statement and ending with COMPOSE statement. The user is warned that only limited error checking is possible to ensure that a group of statement is valid semantically. Hence, the results must be checked carefully to ensure that the all series are processed as intended.

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.

Subsequently, the procedure has run successfully under SAS Version 5 at Statistics Canada.

Depending upon the future needs, the procedure may be adapted to CMS Operating System. 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=TISE.SASX4.PROC,DISP=SHR
//      EXEC SAST,DSN='TISE.SAS4G.LOAD'
//FT03F001 DD SYSOUT=*,DCB=(BLKSIZE=141,LRECL=137,RECFM=FBA)
//SYSIN DD *
           .
           .
           .
          (SAS statements)

         DATA steps and PROC steps

/*
```

The above JCL will invoke PROC X11ARIMA whose FORTRAN source has been compiled

using IBM FORTRAN G1 compiler.

Alternatively,

```
            //job card information
            //PROCLIB DD DSN=TISE.SASX4.PROC,DISP=SHR
            //   EXEC SAST,DSN='TISE.SAS4H.LOAD'
            //FT03F001 DD SYSOUT=*,DCB=(BLKSIZE=141,LRECL=137,RECFM=FBA)
            //SYSIN DD *
                    .
                    .
                    .

                  (SAS statements)

                  Data steps and PROC steps
            /*
```

The above JCL will invoke PROC X11ARIMA whose FORTRAN source has been compiled using IBM FORTRAN H Extended compiler.

SPECIAL CONSIDERATION FOR BETA TEST VERSION

In developing X11ARIMA, no change was made to the computaional routines in the X-11-ARIMA Program. Therefore, for most cases, there is no discrepancy between the results from PROC X11ARIMA and the X-11-ARIMA Seasonal Adjustment Porgram. Testing so far has turned up cases where one of the ARIMA parameters being estimated differently by 0.001 - 0.002.

The procedure was first developed using FORTRAN G1 compiler. It was changed to the H EXTENDED Compiler without any effect on the computational results. There seems to be some reduction in CPU time under H EXTENDED compiler.

Not enough test results are available at present to identify an exact cause or causes for the discrepancies in ARIMA parameter estimation. However, it has been known that the process is very sensitive to "environmental changes". One of such changes is the fact that SAS usually keeps input data in double precision while single precision is used by the X-11-ARIMA Seasonal Adjustment Program. This can affect data with a large number of significant digits.

Other discrepancies are all related to the difference in a manner of input and output. In the beta test version, OUTPUT statement is not accepted. In the printed output, heading of the tables may be similar but not identical to the original program. These will be cleaned up along with the bugs reported from the beta test version.

Reports on problems and suggestions for improvement are greatly appreciated. They should be sent to Dr. E.B. Dagum at the address and the telephone number given at the start of this manual.

EXAMPLES

The next three examples will produce output correspondent to the results shown in Dagum (1980). These test data are distributed with the distribution tape. Therefore, the user interested in executing these examples should contact the installer of PROC X11ARIMA.

Example for Monthly Series

```
//SASTESTM JOB ... ... Necesary Job Information ... ...
//PROCLIB DD DSN=TISE.SASX4.PROC,DISP=SHR
//S3  EXEC SAST,DSN='TISE.SAS4GX.LOAD'
//FT03F001 DD SYSOUT=*,DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)
//SYSIN DD *
DATA A;
  INFILE IN;
  INPUT ID $ 1-6 YR 7-8  9 (X1-X12) (6.0);
PROC PRINT;
DATA B;
  ARRAY A(I) X1-X12;
  SET A;
  DO I=1 TO 12;
  FINALM=A;
  OUTPUT;
  END;
  KEEP ID YR FINALM;
  LABEL FINALM
        =FINANCE,CHECKS CASHED IN CLEARING CTRS;
PROC PRINT;
PROC X11ARIMA;
    SERIES FINALM;
    MONTHLY START=JAN68 END=DEC77 TDREGR=TEST
            PRINTOUT=FULL CHARTS=ALL EXTRAPOLATION=1;
    PDWEIGHTS MON=1.272 TUES=1.071 WED=1.674 THUR=1.090
```

```
                   FRI=0.929 SAT=0.431 SUN=0.532;
     //IN DD DSN=TISE.SAS4G.CNTL(SASDATAM),DISP=SHR
```

Example for Quarterly series

```
     //SASTESTQ JOB ... ... Necesary Job Information ... ...
     //PROCLIB DD DSN=TISE.SASX4.PROC,DISP=SHR
     //S3  EXEC SAST,DSN='TISE.SAS4G.LOAD'
     //FT03F001 DD SYSOUT=*,DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)
     //SYSIN DD *
     DATA A;
       INFILE IN;
       INPUT ID $ 1-6 YR 7-8  9 X1-X4;
     PROC PRINT;
     DATA B;
       ARRAY Q(I) X1-X4;
       SET A;
       DO I=1 TO 4;
       FINALQ=Q;
       OUTPUT;
       END;
       KEEP ID YR FINALQ;
       LABEL FINALQ=FREIGHT AND SHIPPING PAYMENTS;
     PROC PRINT;
     PROC X11ARIMA;
         VAR FINALQ;
         QUARTERLY START='69Q1' END='78Q4' NDEC=1 TYPE=ADDITIVE
                   PRINTOUT=ANALYSIS CHARTS=STANDARD EXTRAPOLATION=1;
     //IN DD DSN=TISE.SAS4G.CNTL(SASDATAQ),DISP=SHR
```

Example for Composite Series

```
//SASTESTC JOB ... ... Necesary Job Information ... ...
//PROCLIB DD DSN=TISE.SASX4.PROC,DISP=SHR
//S3  EXEC SAST,DSN='TISE.SAS4G.LOAD'
//FT03F001 DD SYSOUT=*,DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)
//SYSIN DD *
DATA A;
  INFILE IN;
  INPUT ID $ 1-6 YR 7-8  9 (X1-X12) (6.0);
PROC PRINT;
DATA B;
  ARRAY A(I) X1-X12;
  ARRAY C1(J) Y1-Y144;
  ARRAY C2(J) Z1-Z144;
  DO L=1 TO 2;
  DO K=1 TO 12;
  SET A;
  DO I=1 TO 12;
  J=I+(K-1)*12;
  IF L=1 THEN C1=A;
  IF L=2 THEN C2=A;
  END;
  END;
  END;
  DO J=1 TO 144;
  M0030=C1;
  F0041=C2;
  OUTPUT;
  END;
  KEEP ID YR M0030 F0041;
  LABEL M0030
```

```
        =UNEMPLOYMENT MEN 16-19 YEARS;
  LABEL F0041
        =UNEMPLOYMENT WOMEN 16-19 YEARS;
PROC PRINT;
PROC X11ARIMA COMPOSITE;
    SERIES M0030 F0041;
    MONTHLY START=JAN67 END=DEC78 TYPE=ADDITIVE
            PRINTOUT=BRIEF CHARTS=NONE EXTRAPOLATION=2
                                     SERIES_ID=M0030;
    ARIMA USER AMODEL=(0,1,1)(0,1,1) SERIES_ID=M0030;
    COMPOSE                          SERIES_ID=M0030;
    MONTHLY START=JAN67 END=DEC78 TYPE=ADDITIVE
            PRINTOUT=BRIEF CHARTS=NONE EXTRAPOLATION=2
                                     SERIES_ID=F0041;
    ARIMA USER AMODEL=(0,1,1)(0,1,1) SERIES_ID=F0041;
    COMPOSE                          SERIES_ID=F0041;
    MONTHLY START=JAN67 END=DEC78 TYPE=ADDITIVE
            PRINTOUT=BRIEF CHARTS=NONE EXTRAPOLATION=2;
    ARIMA USER AMODEL=(0,1,1)(0,1,1);
    COMPOSE C_OP=END;
//IN DD DSN=TISE.SAS4G.CNTL(SASDATAC),DISP=SHR
```

REFERENCES

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

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).

Gratton, M. and Dagum E.B. (1986), Specification for PRDC X11ARIMA, Working Paper TSRA-86-004E, Statistics 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 Dffice, Washington, DC.