

TRAMO DOCUMENTATION: CHANGES IN THE PROGRAM

by
V́ctor Ǵmez

November 1998

Contents

1	Changes for EUROSTAT	3
1.1	Fixing of ARIMA Parameters in Regarima Models	3
1.2	Fixing of Regression Parameters in Regarima Models	4
1.3	Treatment of Different Kinds of Holidays in TRAMO	6
2	General Changes in TRAMO	9
2.1	Negative INT2 and NBACK parameters	9
2.2	OUT = ± 2	10
2.3	TRAMO -t	10
2.4	Test for the Log-level Specification	10

In this document, the changes performed in the TRAMO program during the year 1998 are described. These changes can be divided into two main categories. The first one responds to the demands of EUROSTAT for the interface DEMETRA, while the second consists of some bug corrections and some small changes in the code to improve the program performance.

In the following, I will briefly describe the two parts, listing the subroutines which are new or have been affected by the changes.

1 Changes for EUROSTAT

The changes in the TRAMO program performed for EUROSTAT refer to the fixing of parameters in Regarima models and the treatment of different kinds of holidays according to EUROSTAT's specifications.

The fixing of parameters in Regarima models is accomplished by the appropriate setting of the parameters in the namelist INPUT and the corresponding namelists REG. A new option of the IREG parameter is used to specify the fixing of regression parameters, namely a negative value of IREG.

Instructions to specify the four kinds of holidays considered by EUROSTAT are entered to the TRAMO program by means of a namelist REG and a new option for the IUSER parameter, namely IUSER=-3.

1.1 Fixing of ARIMA Parameters in Regarima Models

The ARIMA model parameters are entered by means of the arrays PHI, BPHI, TH, and BTH. The arrays JPR, JPS, JQR and JQS are used to fix the ARIMA model parameters when the model is estimated using any method (IFILT = 2,3,4,1) different from the Hannan-Rissanen's method, henceforth referred to as the HR method. Note that the HR method is used to obtain initial estimates for the estimation of the ARIMA model when IFILT = 2,3,4,1, and to estimate the ARIMA model parameters in the automatic outlier detection and correction (AODC) stage when IMVX=0 has been specified.

If no ARIMA model estimation is to be performed when IFILT = 2,3,4,1, the specification INIT=2 is used. If no estimation of the ARIMA model parameters is desired during AODC, in addition to fixing the ARIMA model parameters by means of the arrays PHI, JPR, BPHI, JPS, etc., the speci-

fication IMVX=1 must be used. For example, if the ARIMA model is the following airline model

$$\nabla\nabla_{12}z_t = (1 - .4B)(1 - .6B^{12})a_t, \quad (1)$$

an appropriate specification to fix the parameters in this model in all stages of the program run, included the AODC stage, would be:

```
$INPUT
  IMEAN=0
  TH(1)=-.4 JQR(1)=1
  BTH(1)=-.6 JQS(1)=1
  INIT=2 IMVX=1
  INT1=125 IATIP=1
$
```

This specification would tell the program to perform AODC in the span $(125, N)$, where N is the number of observations, keeping fixed all ARIMA parameters.

1.2 Fixing of Regression Parameters in Regarima Models

To fix the regression parameters in Regarima models, the parameter IREG should be set equal to minus the number of fixed regression parameters, included the mean, the missing observation, the Trading Day, the Leap Year, the Easter and the outlier parameters. At the same time, the array RG should contain the -IREG values of the fixed regression parameters. For example, let us suppose that there are six regression variables in a Regarima model, where the ARIMA model is an airline model with mean and the estimated regression coefficients are .1, -.25, -.31, .45, -.6, and .7. The regression variables are supposed to be as follows. The first regression parameter variable is the mean. The second regression variable is a dummy variable for a missing value. The third regression variable is an intervention variable, which is an impulse at observation 28, the effect of which should be assigned to the irregular component in SEATS. The fourth regression variable is a Trading Day variable, and the fifth and sixth regression variables are the outliers 25 AO and 34 LS. Then, the following specification could be used in the TRAMO

input file to fix the regression parameters when a search for outliers is desired in the span $(125, N)$, where N is the sample size:

```
$INPUT
  ITRAD=1
  INT1=125 IATIP=1
  IREG=-4 RG(1)=.1 RG(2)=-.25 RG(3)=-.31
           RG(4)=.45 RG(5)=-.6 RG(6)=.7
$
$REG IUSER=2 NSER=2 $
      25 AO 34 LS
$REG ISEQ=1 REGEFF=3$
      28 1
```

The regression parameters may change during the AODC stage because multiple regressions are performed to discard spurious outliers. They may also change if the sample size used to compute them is different from the present sample size. However, the array RG is used to restore the values of the fixed regression parameters after the estimation of the ARIMA models or the search for outliers are finished.

Note 1: The order of the fixed values entered by the user in the array RG should coincide with the order in which the regression variables are stored in TRAMO: the mean, dummies for missing observations (only with INTERP=2), indicator variables (or intervention variables, etc.), Trading Day variables, Leap Year variable, Easter variable and outlier variables. Thus, in the earlier example, .1 is the value of the mean, -.25 is the value of the missing value coefficient, -.31 the value of the intervention coefficient, .45 the value of the Trading Day coefficient, -.6 the value of the 25 AO outlier coefficient, and .7 the value of the 34 LS outlier coefficient.

Note 2: The white noise residuals reported by the program are obtained using the regression coefficients computed by the program, which, as mentioned earlier, may be different from the fixed regression values. However, the interpolated values, the linear series and the forecasts are computed using the fixed coefficients.

Note 3: It is not possible, in general, to fix some regression parameters and let other regression parameters vary. The only exception to this rule are the

outliers. It is possible to fix first some outliers and then perform a search for outliers in some span of the data, so that new outliers can be detected but not corrected ($INT2 < 0$) or detected and corrected ($INT2 < 0$).

Note 4: The difference between a positive IREG and a negative IREG is that minus the latter is equal to the former plus $IMEAN + \#$ of missing values (only with $INTERP=2$) + $ITRAD + IEAST + \#$ of fixed outliers.

Note 5: To fix outliers, only one namelist REG can be used. This applies to both positive and negative IREG. When IREG is positive, it must include the $\#$ of fixed outliers.

Note 6: When the user fixes the regression parameters, Trading Day variables are included ($ITRAD=1,2,6,7$), and an external file (namelist REG with $IUSER=-2$) or the EUROSTAT's specifications (namelist REG with $IUSER=-3$) are used to incorporate holidays that do not fall on Sunday ($ITRAD=1,2$), or on Saturday or Sunday ($ITRAD=6,7$), the negative value that the user assigns to the input parameter IREG must be such that $-IREG$ includes the $\#$ of Trading Day variables twice. For example, if there is only one regression variable in the model, which is a Trading Day variable ($ITRAD = 1$), the EUROSTAT's specifications (namelist REG with $IUSER=-3$) are used to incorporate holidays, and the Trading Day parameter is fixed, the user must enter $IREG=-2$.

1.3 Treatment of Different Kinds of Holidays in TRAMO

According to EUROSTAT, there are 4 kinds of holidays:

1. Fixed yearly repeated holidays (FYRH), e.g., 1st of January
2. Moving yearly repeated holidays (MYRH), e.g., Easter Monday
3. Fixed yearly repeated free weekdays (FYRFW), e.g., each third Wednesday in May
4. Fixed unrepeated holidays (FUH), e.g., 8 June 1975 (death of the President)

To compute the week day given any date, the algorithm provided by Jens Dossé is used in TRAMO. This algorithm is implemented in subroutine WEEKDAY. Using this algorithm, it is possible to compute the week day which corresponds to any FYRH or FUH.

The 13 different moving holidays considered by EUROSTAT are stored in TRAMO in a 199×13 matrix in subroutine GENHLVAR. This matrix contains the week days of the 13 moving holidays from 1901 to 2099. The order in which the moving holidays are stored is, using EUROSTAT's acronyms, the following:

AshWed MauThu GooFri EasMon PraFri AscThu WhiMon CorThu OLeMon

OGooFri OEaMon OEaTue OGhMon

The acronyms stand for:

AshWed : Ash Wednesday

MauThu : Maundy Thursday (Easter Thursday)

GooFri : Good Friday (Easter Friday)

EasMon : Easter Monday

PraFri : Common Prayer Day (Friday)

AscThu : Ascension Day (Thursday)

WhiMon : Whit Monday (Pentecost Monday)

CorThu : Corpus Christi (Thursday)

OLeMon : Orthodox Calendar -
Start of Lent (Clean Monday)

OGooFri : Orthodox Calendar -
Good Friday (Easter Friday)

OEaMon : Orthodox Calendar -
Easter Monday

OEaTue : Orthodox Calendar -
Easter Tuesday

OGhMon : Orthodox Calendar -
Holy Ghost (Pentecost Monday)

The previous 4 kinds of holidays are treated in TRAMO like the holidays read from an external file. That is, for each month, the number of working days is decreased and the number of non-working days is increased by the amount of the holidays.

To specify the previous 4 kinds of holidays, the input namelist should contain the parameter:

- a) when ITRAD = 6, 7,
IREG = \pm (# of regression variables +6),
- b) when ITRAD = 1, 2,
IREG = \pm (# of regression variables +1),

where # of regression variables includes those entered by the user and the ones generated by the program. If IREG < 0, it also includes the mean, missing observation (only if INTERP=2), Trading Day, Leap year, Easter, and outlier variables. Holidays are treated as non-working days. That is, for each month, the number of working days is decreased and the number of non-working days is increased by the amount of the holidays.

A namelist REG has to be added, with the parameter values:

IUSER = -3,

NSER = 6 or 1 and the value of ILONG. This namelist is followed by 5 lines. These lines contain:

Line 1 # of FYRH # of MYRH # of FYRFW # of FUH (free format)

Line 2 # of FYRH pairs of numbers (free format)

Line 3 # of MYRH numbers (free format)

Line 4 # of FYRFW pairs of numbers (free format)

Line 5 # of FUH 3-plus of numbers (free format)

where each pair of the second line is of the form 3 (day of month) 1 (month, January), each number of the third line refers to the column in the moving holiday matrix e.g., 4 (Easmon), each pair of numbers in the fourth line is of

the form 1 (day of the week, monday) 9 (month, September), and each 3-pla in the fifth line if of the form 8 (day of month) 6 (month, June) 1975 (year).

For example, if the user wants to enter the following:

6 FYRH: 1 Jan 12 Feb 21 Feb 4 Jul 11 Nov 25 Dec

2 MYRH: EasMon WhiMon

3 FYRFW: Mon Sep Mon Oct Tue Nov

1 FUH: 8 Jun 1975

the 5 lines after the namelist REG must be as follows:

6 2 3 1

1 1 12 2 21 2 4 7 11 11 25 12

4 7

1 9 1 10 2 11

8 6 1975

2 General Changes in TRAMO

In this section, I will describe the changes made in the source code in the TRAMO program during the year 1998 which do not refer to the changes made for EUROSTAT.

2.1 Negative INT2 and NBACK parameters

When $INT2 < 0$ or $NBACK < 0$, both parameters act independently.

- If $INT2 < 0$, the program searches for outliers in the span ($INT1$, $INT2 + NZ$), where NZ is the original number of observations, that is, it is the number of observations without decreasing by $-NBACK$ if $NBACK < 0$. The program detects outliers in the last $-INT2$ observation, but it does not correct for the effects of the detected outliers.
- If $NBACK < 0$, the sample used by the program for parameter estimation, AODC, etc., is (1, $NZ + NBACK$). An F-test is performed to compare the in-sample with the out-of-sample residuals.
- It may happen that $INT2 < 0$, $NBACK < 0$ and $NZ + NBACK < NZ + INT2$. For example, $NZ = 100$, $NBACK = -5$, and $INT2 = -1$. Then,

$NZ + NBACK = 95 < NZ + INT2 = 99$. In this case, the AODC is performed in the span ($INT1, NZ + NBACK$) and an outlier detection (without correction) is done in the last $-INT2$ observations.

2.2 OUT = ± 2

If $OUT = \pm 2$, in addition to the usual file TABLE.OUT, two more files are written in the OUTPUT directory. These files are NUMRES1.OUT and NUMRES2.OUT. These two last files contain numerical results only. The file NUMRES1.OUT has 19 columns containing:

```
ISERIE, DDFTIME, NZ, LAMDA, IMEAN, ITRAD, IEAST, P, D, Q,
BP, BD, BQ, NUMAO, NUMTC, NUMLS, RVR, PVBOX, PVNT
```

where ISERIE is the series number, DDFTIME is the elapsed time for that series, NUMAO is the number of identified AOs, NUMTC is the number of identified TCs, NUMLS is the number of identified LSs, RVR is the standard deviation of the residuals, PVBOX is the P-value of the Ljung–Box statistic, and PVNT is the P-value of the Jarque–Bera normality test.

The file NUMRES2.OUT has 8 columns containing the ARIMA parameters in the order: 3 columns for PHI, 1 column for BPHI, 3 columns for TH, and 1 column for BTH.

2.3 TRAMO -t

The line command TRAMO -t creates a file with the name VARIABLE.OUT containing the original series (1st column) and the regression variables (rest of the columns), in the order in which the regression variables are stored in TRAMO: mean, missing variable, indicator (or intervention), Trading Day, Leap year, Easter, and outlier variables.

2.4 Test for the Log-level Specification

The test for the log-level specification is performed now by means of the minimization of the likelihood only. The range-mean regression test is no longer used. A new parameter, FCT, has been included. This parameter is used to favor the specification of logs. The residual sum of squares (SSSL) of an airline model with mean fitted to the original series is compared with

FCT*SSCL, where SSCL is the residual sum of squares of an airline model with mean fitted to the logged series. If SSSL is less than FCT*SSCL, no logs are taken. The default value of FCT is $1/(1 - .0125)$.