May 1998

PROGRAM TERROR

An application of TRAMO to the detection
of errors in time series

Víctor **Gómez** * and Agustín **Maravall** **

with the programming help of Gianluca **Caporello**.

* Ministerio de Economía y Hacienda, Madrid
** Banco de España, Madrid

terror\terrnote.ts

1. INTRODUCTION

This note contains a brief description of program TERROR (Tramo for ERRORS). For a more complete description of the procedures, see the manual of program TRAMO ("**T**ime series **R**egression with **A**rima noise, **M**issing observations, and **O**utliers" by V. Gómez and A. Maravall.)

Program TERROR is designed to handle large sets of time series with a monthly or lower frequency of observation.

For each series, the program automatically identifies an ARIMA model and detects and corrects for several types of outliers. (It also interpolates missing observations if there are any.) Next, the one-period-ahead forecast of the series is computed and compared with the new observation (this new observation is not used for estimation). In brief, when the forecast error is, in absolute value, larger than some a priori specified limit, the new observation is identified as a possible error.

The model obtained by TERROR for each series to compute the forecast is of the following type.

Given the vector of observations:

$$z = (z_{t_1}, \dots, z_{t_M}) \quad (1)$$

where $0 < t_1 < \dots < t_M$, the program fits the model

$$z_t = y'_t \, \omega + v_t , \quad (2)$$

where $\omega = (\omega_1,\dots,\omega_n)'$ is a vector of regression coefficients, $y'_t = (y_{1t},\dots, y_{nt})$ denotes n regression variables, and v_t follows the general ARIMA process

$$\phi(B)\, \delta(B)\, v_t = \Theta(B)\, a_t , \quad (3)$$

where B is the backshift operator; $\phi(B)$, $\delta(B)$, and $\Theta(B)$ are finite polynomials in B, and a_t is assumed a n.i.i.d. $(0, \sigma_a^2)$ white-noise innovation.

The polynomial $\delta(B)$ contains the unit roots associated with differencing (regular and seasonal), $\phi(B)$ is the polynomial with the stationary autoregresive roots (and the complex unit roots, if present), and $\Theta(B)$ denotes the (invertible) moving average polynomial. In TERROR, they assume the following multiplicative form:

$$\delta(B) = (1 - B)^d (1 - B^s)^D$$

$$\phi(B) = (1 + \phi_1 B + ... + \phi_p B^p)(1 + \phi_1 B^s + ... + \phi_P B^{s \times P})$$

$$\Theta(B) = (1 + \Theta_1 B + ... + \Theta_q B^q)(1 + \Theta_1 B^s + ... + \Theta_Q B^{s \times Q}),$$

where s denotes the number of observations per year. The model may contain a constant μ, equal to the mean of the differenced series $\delta(B) z_t$. In practice, this parameter is estimated as one of the regression parameters in (2).

The regression variables are outliers automatically identified by the program. Thus, if there are J outliers, occurring at periods $T_1,...,T_j$, the regression term can be expressed as

$$y'_t \omega = \sum_{i=1}^{J} \omega_i \lambda_i(B) I_t^{(Ti)},$$

where $I^{(Ti)}_t$ is a dummy variable equal to 1 when $t=T_i$, and zero otherwise. The polynomial $\lambda_i(B)$ specifies the type of outlier detected. Three types are considered:

Additive outlier: $\lambda(B) = 1$

Level shift : $\lambda(B) = 1/\nabla$

Transitory change: $\lambda(B) = 1/(1 - \delta B)$,

where $0 < \delta < 1$, and by default $\delta = .7$.

The additive outlier represents a one-period spike, the level shift represents a step function, and the transitory change is as an additive outlier that gradually disappears over several periods.

2. INPUT FILE AND INPUT PARAMETERS

2.1 Parameters fixed in TRAMO

In the DOS version, the **input file** is of the type ITER = 2, i.e., the same input namelist applies to all series. The program fixes several parameters in TRAMO, namely the following ones.

Parameter Meaning

Automatic ARIMA Model Identification

IDIF = 3 Used for automatic model identification. The program searches first for regular differences up to order 2 and for seasonal differences up to order 1. Then, it continues with the identification of an ARMA model for the differenced series.

INIC = 3 The program searches for regular polynomials up to order 3, and for seasonal polynomials up to order 1.

Automatic detection and correction of three types of outliers (level shifts, transitory changes, and Additive Outliers)

AIO = 2 Additive Outliers, Transitory Changes, and Level Shifts are considered.

IATIP = 1 Automatic detection and correction of outliers is performed.
After correcting for the outliers found in the first round, the program performs a new automatic model identification, and a new search for outliers if the model has been changed. In this second round, the critical value VA is reduced by the fraction .14.
If the second round does not provide a satisfactory model, a third round is carried out. (The model obtained with automatic identification is always compared with the default model.)

Pre-test for log level specification

LAM = -1 The program tests for the log-level specification.
The test is based, first, on the slope (b) of a range-mean regression, "trimmed" to avoid outlier distortion. This slope b is compared to a constant (β), close to zero, that depends on the number of observations and on the value of RSA. When the results of the regression are unclear, the value of LAM is chosen according to the BIC of the default model, using both specifications.

Parameter	Meaning

Additional parameters that are fixed:

NBACK = -1	The last observation is omitted from the end of the series. The model is estimated for the shorter series and the one-period-ahead forecast error is computed for the last period (without reestimation of the model) as well as the associated t-value.
INTERP = 1	Interpolates missing values (if any)
ITER = 2	The same input namelist applies to all series in the file.

2.2 Other input parameters from TRAMO

The TRAMO parameters that are not part of the fixed input configuration can be entered if one wishes to modify their default value. The most relevant ones are the following:

Parameter — **Meaning**

MQ = Number of observations per year (12 for monthly, 6 for bimonthly, 4 for quarterly, 1 for annual, and so on).
Default = 12

VA = Used to set the critical value for outlier detection.
The default values are as follows:

$NZ \leq 50$,	VA = 3
$50 < NZ \leq 250$,	VA = 3.5
$250 < NZ \leq 500$,	VA = 3.8
$500 < NZ$	VA = 4

INT2 = Two parameters, INT1 and INT2, can be used to define the interval (INT1, INT2) over which outliers have to be searched. By default

INT1 = 1; INT2 = NZ.

A facility has been introduced to avoid outlier correction in the last observations of the series, which may cause instability in the series forecasts. In particular:

Parameter	Meaning

When INT2 = -k < 0, outliers are automatically detected and corrected in the interval (INT1, NZ - k). Then, the detection procedure is applied to the last k observations, and if some outlier is detected a warning is printed, but no correction is made. Example: if NZ = 47 and INT2 = -2, outliers are detected and corrected for the interval (1, 45). For the last 2 observations they are only detected. Since NBACK = -1, observation 47 is deleted from the sample estimation period, and hence in this case INT2 = -2 implies that outliers, if detected, will not be corrected for the last observation of the sample estimation period (observation 46).
Detection and correction of outliers is made with a procedure based on those of Tsay (1986) and Chen and Liu (1993). This procedure has some notable features (namely, exact residuals are used; estimation, based on two simple regressions, is fast; the parameter estimates are modified at each iteration; and it uses a robust method to incorporate or reject outliers which is similar to the stepwise regression procedure to select the "best" regression equation).

In automatic outlier identification and correction and automatic model identification, the sequence of actions is as follows. After obtaining the degrees of differencing automatically (IDIF = 3), the program obtains first a model using the BIC criterion and then performs the automatic identification and correction of outliers using the previously identified model. There are two rounds: first, the program uses this model to correct the series for the outlier effects. Then, it returns to model identification and further outlier identification and correction, using the model obtained in the second round and a critical level equal to .86 times that of the first round.

INCON = 0	Exact maximum likelihood estimation
= 1	Unconditional least squares.
	Default = 0
IMVX = 0	The fast method of Hannan-Rissanen is used for parameter estimation in the automatic detection and correction of outliers
= 1	Maximum likelihood estimation is used for parameter estimation in the automatic detection and correction of outliers
	Default = 0

2.3 New input parameters

TERROR is mostly controlled by the parameter SENS, which determines how sensitive should be the detection of errors, according to the following values.

SENS = 0 Low sensitivity;
= 1 Medium sensitivity;
= 2 High sensitivity.

The default value is 1.

The parameter SENS sets two parameters, k_1 and k_2.

Let t = forecast error/standard derivation of residuals. Then:

If, for a particular series,

$|t| > k_2$, the new observation in the series is classified as "likely" to contain an error.

If

$k_1 < |t| \le k_2$, the new observation is classified as containing a "possible" error.

If

$|t| \le k_1$, the new observation is accepted as without error.

The values of k_1 and k_2 for the different levels of sensitivity are as follows:

SENS = 0	$k_1 = 4$	$k_2 = 5$
SENS = 1	$k_1 = 3.9$	$k_2 = 4.42$
SENS = 2	$k_1 = 3.3$	$k_2 = 3.9$

These values can be changed. To do that, set

SENS $\ge$ 3,

then you can enter the new values of k_1 and/or k_2.

3. **INPUT FILE**

The input file should be named serie, and should have the structure of an ITER = 2 file (see TRAMO User Instructions). In particular, the line with the input namelist should only appear once, after the first serie in the file. As described above, all parameters are fixed or set by default and the INPUT NAMELIST should only contain the parameters (not fixed) that depart from the default value.

The input file consists of a sequence of series; for each series, the maximum numbers of observations is 600, and the minimum 15. Each serie should be entered as follows:

First line: J TITLE (no more than 72 characters)

Second line: NZ NYEAR NPER (free format)

Third line et seq.: Z(I) : I = 1,..., NZ (free format), where J refers to the order of the series in the file (J = 10 denotes the tenth series), NZ is the number of observations, NYEAR is the starting year, and NPER is the starting period of the year (if series is monthly, 1 for January, 2 for February, and so on), Z(.) is the array (or matrix) of observations. For each missing observation, the code -99999. must be entered.

The first series is followed by the INPUT namelist, entered as follows:

$INPUT parameter = parameter value ... $,

where only non-fixed, non-default parameter values need to be entered. The format "$...$" can be replaced by others such as "&.../". The INPUT namelist is entered after the first series.

An example of an input file is presented. The fixed configuration of TERROR will apply. The only parameters that depart from their default value are

a) MQ = 4, which specifies that the series are quarterly ones (default is monthly).

b) INT2 = -2, which specifies that, for the last two observations of the series, outliers are only detected but not corrected.

Notice that the third series contains holes corresponding to missing values.

EXAMPLE OF INTPUT FILE

```
1        0012-ARGELIA,A14-ACT.TOTAL-PTS. + M.EXTRANJERA,TO
48 1986  1
              77485     72684     70303     66187
              64519     61316     59160     56160
              55040     52270     52005     48887
              47371     46850     47198     47239
              47345     46775     48923     50012
              58136     65971     67761     67319
              95307     89150     95727    122034
             143982    155044    168047    174882
             184586    187004    185379    180137
             164085    145485    137217    126161
             125174    120003    107599    103621
             106805    105384    100703     97806
$INPUT MQ=4 INT2=-2 $
2        0020-ANDORRA,A14-ACT.TOTAL-PTS. + M.EXTRANJERA,TO
48 1986  1
              12389      8227      8627     16277
              13296      5263      9281      9769
              10380     17787     11064      7650
              18701      8517      9156      9114
              20940     12368     11669     12704
              13029     13679     11033      9951
              10794      9565     10187     13187
              13498     19958     20268     30502
              40780     40261     31394     29658
              31537     38386     39725     33322
              38434     33550     36786     20566
              22848     23870     36492     33538

3        0024-ANGOLA,A14-ACT.TOTAL-PTS. + M.EXTRANJERA,TO
48 1986  1
              11818     10910      9857     10542
              10640    -99999    -99999     14060
              12938    -99999     13340     11574
              12038     13572    -99999    -99999
              12389     11901     13292     15877
              15177     18517     17014     19784
              20751     25736     39555     48067
              56396     63402     63886     62345
              56001     47451     39780     35479
              30810     28343     26995     27825
              26382     25333     24117     22783
              26999     28101     28883     37710
4        0032-ARGENTINA,A14-ACT.TOTAL-PTS. + M.EXTRANJERA,TO
48 1986  1
              95915     95132     90424     87995
              83089     78520     73128     70696
              69726     74866     79512     77214
              77841     74953     71186     66207
              66372     54533     56834     50212
              64001     63282     59404     60801
              63065     58750     60839     73021
              78635     77720     77284     99295
             106840    110325    101370     80381
              84523     85920     84708     96051
              97306    104271    114889    179567
             207081    244191    249611    366232
```

4. OUTPUT FILE

TERROR produces the output file list.out in the directory TERROR\OUTPUT. The file consists of two sets of pairs of numbers, each pair given in brackets. The first set contains the series for which $|t| > k_2$, that is the series "likely" to contain an error. The second set contains the series for which $k_1 < |t| \leq k_2$, that is the series for which an error is "possible".

Each bracket contains two numbers. The first one gives the position of the series in the list of series contained in the input file. The second number is the t value of the forecast error. Thus { 4, 6.32 } indicates that the 4th series is likely to have an error, since t=6.32.

An example of a list.out file is provided. It specifies that the program has been applied to a file with 436 series, with k_1 and k_2 set at their default values (the ones implied by SENS = 1). Of the 436 series, 28 are candidates for a likely error (the first candidate, is series 192, with the largest value of t, equal to 16.32). The set of possible errors provides 8 additional series.

All series were processed without error.

EXAMPLE OF THE OUTPUT FILE

```
SERIES

 436  K1=         3.900      K2=         4.420


Likely =
          {   4,    6.32},{  24,   -5.36},{  27,   10.75},{  32,    7.56},
          {  45,    5.74},{  52,    6.32},{  88,   11.28},{  92,   -8.13},
          {  95,    6.87},{ 106,    5.48},{ 114,   -5.55},{ 135,   12.81},
          { 140,    7.57},{ 143,    4.72},{ 163,    5.94},{ 172,   12.70},
          { 192,   16.32},{ 198,   14.14},{ 248,    5.41},{ 267,    6.92},
          { 281,    8.70},{ 283,    5.74},{ 308,   12.50},{ 345,    5.95},
          { 352,   13.96},{ 358,    4.64},{ 370,    4.51},{ 391,    7.69},


  AENP =     0.05


Possible =
          {  57,   -4.33},{ 112,    4.04},{ 132,   -3.94},{ 134,    3.99},
          { 160,   -4.04},{ 187,    4.20},{ 215,   -4.26},{ 266,   -4.00},


  AENL =     0.00

NDP Exception =
```