

March 2004

**A TOOL FOR QUALITY CONTROL  
OF TIME SERIES DATA  
Program TERROR**

Gianluca **Caporello** and Agustín **Maravall**

Bank of Spain

Thanks are due to the research and secretarial help of Jorge Carrillo and Nieves Morales.

## **Abstract**

A program is presented for error detection in incoming data for data bases of time series. It provides a sensible answer to the question: when new data is reported, which of the new observations are likely to contain an error? The observation is suspicious if it is far from what could have been expected from the past history of the variable.

Without considering the new observation, and using a fully automatic procedure, a REGARIMA model is identified for the series, which may have missing observations and be contaminated by outliers. Calendar and, more generally, regression effects can also be present. The new observation is compared to the (out-of-sample) forecast obtained with the model and, if the forecast error is clearly unreasonable, the series is considered suspicious

The program, named TERROR, is a particular application of program TRAMO, freely available at the Bank of Spain web site ([www.bde.es](http://www.bde.es)). The program is reliable and efficient for very large scale applications.

**Keywords:** Forecasting; ARIMA models; Outliers; Data Editing; Error Detection; Computer Software.

**JEL numbers:** C22, C53, C80, C87.

## CONTENTS

- 1 INTRODUCTION
- 2 THE METHODOLOGY
- 3 PROGRAM TERROR: INPUT PARAMETERS
  - 3.1 Fixed Configuration
  - 3.2 Other Input Parameters from TRAMO
  - 3.3 Easter and Trading Day Effects
  - 3.4 New Input Parameters
  - 3.5 Summary of the Results for the Individual Series
  - 3.6 Summary Results for the Set of Series
- 4 INPUT FILE
- 5 MAIN OUTPUT FILE
- 6 AN EXAMPLE
- 7 REFERENCES

## 1. INTRODUCTION

Parallel to the increase in size and information of statistical data bases, there has been a growing concern with the control of the data quality. I center on data bases of time series and consider the following basic quality control problem: **When the data for the present period becomes available, which of the new observations are likely to be erroneous?** What we have in mind are relatively large errors, such as mistaking Belgium for Belize, apples for apricots, not using the correct period, misplacing a decimal, etc. Although not much is known, there is some evidence that the frequency of errors in large statistical data bases can be relatively large (3%, for example, may not be unusual.)

Some economic data bases (at statistical and economic agencies, or big companies) can be very large, and may contain several hundreds of thousands of series that are periodically updated. Quality control of newly reported data relies heavily on “manual” inspection, with the help of some simple measures, such as comparing the last two period-to-period growth, or the annual differences, checking if the new observation is an overall maximum or minimum, etc. These measures do not consider the full information contained in the series, but just a few values. Moreover, they can be strongly affected by seasonality, noise, or special events. Finally, manual (non-automatic) control limits strongly the number of series that can be checked, and hence the control has to be made at a relatively high level of aggregation.

Although “a priori” information can be of help (for example, some change in regulation, or a big bank merger) it would be convenient to dispose of an automatic method, reliable and efficient in large scale applications, that would consider the full information in the time series, and permit control at a more disaggregate level.

Such an automatic method needs to use a sensible and well-defined criterion for judging when a new observation is likely to be an error. An obvious criterion is the following. A new observation is judged suspicious when it is very far from what could have been expected from looking at its past history.

The idea of detecting errors by means of simple time-series models has been around for some time (see, for example, Alwan and Roberts (1988), Lucas and Saccuci (1990), and Montgomery and Mastrangelo (1991),) but the few available applications use very simple models (for example, an AR model with one or two parameters) or ad-hoc simple formulas (such as an Exponentially Weighted Moving Average scheme). Use of properly identified ARIMA models (Box and Jenkins, 1970) was judged too slow and complex, and much too demanding in terms of time-series analysts. Adding the fact that many series are contaminated by outliers (reflecting sometimes past errors that were not corrected; see Barnett and Lewis (1998),) and may contain holes due to missing observations, the problem of

properly modeling large sets of series is further complicated.

Gómez and Maravall (1992-2001) developed a methodology (and an associated program) for automatic (or manual) identification of ARIMA models, when observations may be missing and the series may be contaminated by outliers and by special effects (in particular, calendar effects). The automatic performance of program TRAMO (“Time Series Regression with ARIMA noise, Missing values, and Outliers”) has been intensively tested and it has proved fast and reliable; see, for example, the large scale study in Fischer and Planas (2000). TRAMO’s automatic model identification procedure is in fact being adapted and incorporated to the official X12 ARIMA US-Bureau of the Census program (see, for example, Monsell (2002).)

Program TERROR is simply an application of TRAMO, executed in an automatic manner to the problem of quality control in time series (with several possible options). The program identifies a REG-ARIMA model for each series (perhaps with missing values, outliers, and some deterministic effects) and obtains the standardized forecast error for the period associated with the new data (which is not considered in the process of model identification and forecasting). TERROR is designed to handle large sets of time series with a monthly or lower frequency of observation.

This paper contains a brief description and an application of TERROR. Needless to say, TERROR is meant to be a useful tool, that can be complemented with whatever additional criterion judged relevant. Program TERROR (and TRAMO) can be freely downloaded from the Bank of Spain web site ([www.bde.es](http://www.bde.es)).

## 2. THE METHODOLOGY

For each series, the program automatically identifies an ARIMA model, detects and corrects for several types of outliers, and, if appropriate, estimates calendar effects. It also interpolates missing observations if there are any. Next, the one-period-ahead out-of-sample forecast of the series is computed and compared with the new observation. 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. Summary results for all series and for the aggregate set are also provided. The program runs in a fully automatic manner, though several options are nevertheless available. The next paragraphs adapt the description of TRAMO in Gómez and Maravall (1996) to the case of the TERROR application.

Let  $z_t$  denotes one of the series in the set subject to quality control. 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 regression model

$$z_t = y_t' \beta + x_t, \quad (2)$$

where  $\beta = (\beta_1, \dots, \beta_n)'$  is a vector of regression coefficients,  $y_t' = (y_{1t}, \dots, y_{nt})$  denotes  $n$  regression variables, and  $x_t$  follows the general ARIMA process

$$\phi(B)\delta(B)x_t = \theta(B)a_t, \quad (3)$$

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

It should be noted that  $t_1$  and  $t_M$  may differ for different series in the set. The series may have very different number of observations and start or finish at different periods.

The polynomial  $\delta(B)$  contains the unit roots associated with differencing (regular and seasonal),  $\phi(B)$  is the polynomial with the stationary autoregressive roots, and  $\theta(B)$  denotes the (invertible) moving average polynomial. These polynomials take, in general, the following multiplicative form:

$$\begin{aligned} \delta(B) &= (1-B)^d (1-B^s)^D \\ \phi(B) &= (1 + \phi_1 B + \dots + \phi_p B^p) (1 + \Phi_1 B^s + \dots + \Phi_P B^{s \times P}) \\ \theta(B) &= (1 + \theta_1 B + \dots + \theta_q B^q) (1 + \Theta_1 B^s + \dots + \Theta_Q B^{s \times Q}), \end{aligned}$$

where  $s$  denotes the number of observations per year. The model may contain a constant  $\mu$ , 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).

Outliers that have been automatically identified by the program are also treated as regression variables. Thus, if there are  $J$  outliers, occurring at periods  $T_1, \dots, T_J$ , the outlier regression term can be expressed as

$$y_t' \omega = \sum_{i=1}^J \omega_i \lambda_i(B) I_t^{(T_i)},$$

where  $I_t^{(T_i)}$  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:  $\lambda(B) = 1 / \Delta$
- Transitory change:  $\lambda(B) = 1 / (1 - \delta B)$ ,

where  $\Delta = 1 - B$ ,  $0 < \delta < 1$ , 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. If appropriate for the group of series under consideration, the program can also handle automatic treatment (pre-testing and estimation) of Calendar effects such as Trading Day, Easter, Leap Year, and Holiday effects. Several specifications are available for the Trading Day variable. Calendar effects are also modeled as regression variables.

The program automatically,

- 1) identifies model (2) - (3), and estimates by exact maximum likelihood (or unconditional least squares) its parameters;
- 2) detects and corrects for several types of outliers and, perhaps, special effects;
- 3) yields optimal interpolators of the missing observations with their associated MSE;
- 4) computes optimal forecasts for the series, together with their MSE;

The basic methodology followed in TRAMO is described in Gómez and Maravall (1992, 1994, 1996, 2001), and Gómez, Maravall and Peña (1999). The program is aimed at monthly or lower frequency data, the maximum number of observations is 600 and the minimum depends on the periodicity of the data (in particular, 16 for quarterly and 36 for monthly data).

Estimation of the regression parameters (including calendar variables, outliers, and the missing observations among the initial values of the series) and of the ARIMA model parameters, is made by concentrating the former out of the likelihood (default). Several algorithms are available for computing the likelihood (i.e., the nonlinear sum of squares to be minimized.) When the differenced series can be used, the algorithm of Morf, Sidhu and Kailath (1974) is employed, with a simplification similar to that of Mélard (1984), but also extended to multiplicative seasonal moving average models. For the nondifferenced series, the Kalman filter is used (see Anderson and Moore, 1979). By default, the exact maximum likelihood method is employed (see Gómez and Maravall, 1994, and Gómez, Maravall and Peña, 1999). Nonlinear maximization of the likelihood function and computation of the parameter estimates standard errors is made using Marquardt's method and first numerical derivatives.

Estimation of regression parameters is made by using first the Cholesky decomposition of the inverse error covariance matrix to transform the regression equation (the Kalman filter provides an efficient algorithm to compute the variables in this transformed regression). Then, the resulting least squares problem is solved by applying the QR algorithm, where the Householder orthogonal transformation is used. This procedure yields an efficient and numerically stable method to compute GLS estimators of the regression parameters. Forecasting is made with the Kalman filter applied to the original series; for the problem of building initial conditions on a nonstationary situation, see Gómez and Maravall (1993).

Missing observations can be handled in two equivalent ways. The first one is an extension to nonstationary models of the skipping approach of Jones (1980), and is described in Gómez and Maravall (1994). In this case, interpolation of missing values is made by a simplified Fixed Point Smoother, and yields identical results to Kohn and Ansley (1986). The second one consists of assigning an arbitrary value and specifying an additive outlier for each missing observation. The interpolator is the difference between the tentative value and the estimated regression parameter and coincides with the interpolator obtained with the skipping approach (the likelihood is corrected so that it becomes that of the skipping approach; see Gómez, Maravall and Peña (1999) for more details.) Mean squared errors of the forecasts and interpolations are obtained following the approach of Kohn and Ansley (1985).

The program has a facility for detecting outliers and for removing their effect; the outliers can be entered by the user or they can be automatically detected by the program, using an approach based on those of Tsay (1986) and Chen and Liu (1993). The outliers are detected one by one, as proposed by Tsay (1986), and eventually multiple regressions are used, as in Chen and Liu (1993), to detect the ones that may be spurious. The procedure used to incorporate or reject outliers is similar to the stepwise regression procedure for selecting the "best" regression equation.

In brief, regression parameters are initialized by OLS and the ARIMA model parameters are first estimated with two regressions, as in Hannan and Risannen (1982). Next, the Kalman filter and the QR algorithm provide new regression parameter estimates and regression residuals. For each observation, t-tests are computed for several types of outliers. If there are outliers whose absolute t-values are greater than a pre-selected critical level  $V_A$ , the one with the greatest absolute t-value is selected. Otherwise, the series is free from outlier effects and the algorithm stops.

If some outlier has been detected, the series is corrected by its effect and the ARMA model parameters are re-estimated. Finally, a multiple regression is performed using the Kalman filter and the QR algorithm. If there are some outliers whose absolute t-values are below the critical level  $V_A$ , the one with the lowest absolute t-value is removed from the regression and the multiple regression reestimated. Next, using the regression residuals provided by the last multiple regression, t-test are computed for the types of outlier considered and for each observation. If there are now outliers whose absolute t-values are greater than the critical level  $V_A$ , the one with the greatest absolute t-value is selected and the algorithm goes on to the estimation of the ARMA model parameters to iterate. Otherwise, the algorithm stops. A notable feature of this algorithm is that all calculations are based on linear regression techniques, which reduces computational time. By default, three types of outliers are considered: additive outlier, level shift, and transitory change.

The program also contains a facility to pretest for the log-level specification (based on a comparison of the BIC using both specifications) and, if appropriate, for the possible presence of Calendar effects, such as Trading Day and Easter effects (the pretests are made with regressions using the default model for the noise and, if the model is subsequently changed, the test is redone.) The program further performs an automatic identification of the ARIMA model. This is done in two steps. The first one yields the nonstationary polynomial  $\bar{d}(B)$  of model (3). This is done by iterating on a sequence of AR and ARMA(1,1) models (with mean), which have a multiplicative structure when the data is seasonal. The procedure is based on results of Tiao and Tsay (1983), and Tsay (1984). Regular and seasonal differences are obtained, up to a maximum order of  $\Delta^2 \Delta_s$ , where  $\Delta = 1 - B$  and  $\Delta_s = 1 - B^s$ .

The second step identifies an ARMA model for the stationary series (corrected for outliers and regression-type effects) following the Hannan-Rissanen procedure, with an improvement which consists of using the Kalman filter instead of zeros to calculate the first residuals in the computation of the estimator of the variance of the innovations of model (3). For the general multiplicative model

$$\phi_p(B) \Phi_P(B^s) x_t = \theta_q(B) \Theta_Q(B^s) a_t,$$

the search is made over the range  $0 \leq (p, q) \leq 3$ ,  $0 \leq (P, Q) \leq 2$ . This is done sequentially (for fixed regular polynomials, the seasonal ones are obtained, and viceversa,) and the final orders of the polynomials are chosen according to the BIC criterion, with some possible constraints aimed at increasing parsimony and favoring "balanced" models (similar AR and MA orders).

Finally, the program combines the facilities for automatic detection and correction of outliers and automatic ARIMA model identification just described in an efficient way, so that it can perform automatic model identification of a nonstationary series in the presence of outliers and missing observations (perhaps with some regression effects added).

The default model in TERROR is the so-called Airline Model, popularized by Box and Jenkins (1970). The model is given by the equation

$$\Delta \Delta_s x_t = (1 + \theta_1 B)(1 + \theta_s B^s) a_t, \quad (4)$$

with  $-1 \leq (\theta_1, \theta_s) \leq 1$ . It is often found appropriate for many series (see the large-scale study in Fischer and Planas (2000)), and displays many convenient features (see, for example, Maravall (1998)); in particular it encompasses many other models, including models with close to deterministic trend or seasonality, or models without seasonality. For very short series, for which the automatic model identification is unreliable, TERROR relies heavily on the Airline model specification.

### **3. PROGRAM TERROR: INPUT PARAMETERS**

#### **3.1 FIXED CONFIGURATION**

TERROR can be executed in the DOS version of TRAMO by setting the input parameter TERROR = 1, or in the Windows program TSW by selecting in the main window the TERROR option. For large scale applications the program should be run as a batch file, as can be done with the DOS version. When the input parameter TERROR is set equal to 1 (or the TERROR option chosen) a particular configuration of TRAMO, mostly based on the automatic model identification and outlier detection and correction procedures, is specified. This input configuration will be applied to each time series and includes the following features.

##### **Automatic ARIMA Model Identification**

a) Automatic identification of the unit roots. The program searches 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. (Equivalent parameter in TRAMO: IDIF = 3).

b) Automatic identification of the stationary ARMA model. The program searches for regular polynomials up to order 3, and for seasonal polynomials up to order 2. (Equivalent parameter in TRAMO: INIC = 3).

##### **Automatic detection and correction of outliers**

Automatic detection and correction of Additive Outliers, Transitory Changes, and Level Shifts 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. If the second round does not provide a satisfactory model, a third round is carried out. (Equivalent parameters in TRAMO: AIO = 2, IATIP = 1).

##### **Pre-test for log/level specification**

The program tests for the log-level specification by comparing the Bayesian Information Criteria (BIC) of the default model using both specifications. (Equivalent parameter in TRAMO: LAM = -1).

##### **Interpolation**

Interpolates missing values (if any). (Equivalent parameter in TRAMO: INTERP = 1).

### Computation of the Forecasts and Forecast errors

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 and forecast error are computed for the last period (without reestimation of the model) as well as the t-value associated to the error. (Equivalent parameter in TRAMO: NBACK = -1).

Finally, the same input namelist applies to all series in the file. (Equivalent parameter in TRAMO: ITER = 2).

### 3.2 OTHER INPUT PARAMETERS FROM 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:

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:

$$\begin{array}{ll} NZ \leq 50, & VA = 3 \\ 50 < NZ \leq 450, & VA = 3 + .0025 (NZ - 50) \\ 450 < NZ & VA = 4 \end{array}$$

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

When  $INT2 = -k < 0$ , outliers are automatically detected and corrected in the interval (INT1,  $NZ - k$ ). The detection procedure is also 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  $INT2 = -3$ , outliers are detected and corrected for the interval (1, 44). For the last 3 observations they are only detected. In **TERROR**, since  $NBACK = -1$ , observation 47 is deleted from the sample estimation period, and hence in this case  $INT2 = -3$  implies that outliers, if detected, will not be corrected for the last two observations of the estimation period (observations 45 and 46). **Default** value (in **TERROR**)  $INT2 = -3$ .

In automatic outlier identification and correction and automatic model identification, the sequence of actions is as follows. After obtaining the degrees of differencing automatically, the program finds a model using the BIC criterion, and then performs the automatic identification and correction of outliers with this model. There are two iteration rounds: first, the program finds a model and corrects 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. If the model obtained is not found satisfactory, a third round is carried out with a slightly reduced critical level.

INCON = 0 Exact maximum likelihood estimation. (**Default**).  
 = 1 Unconditional least squares.

IMVX = 0 The fast method of Hannan-Rissanen is used for parameter estimation in the automatic detection and correction of outliers. (**Default**).  
 = 1 Maximum likelihood estimation is used for parameter estimation in the automatic detection and correction of outliers.

### 3.3 EASTER AND TRADING DAY EFFECTS

If the group of series to be treated by TERROR is suspicious of being possibly affected by Easter or Trading Day effects, automatic pre-testing and eventual estimation of the effect, can be incorporated to TERROR. As in TRAMO, this is controlled by the following input parameters.

IEAST = 0 No Easter effect. (**Default**).  
 = 1 Easter effect adjustment.  
 = -1 The program pretests for Easter effect and, if detected, adjusts for it.

IDUR = k Duration of Easter affected period (in days). (**Default = 6**).

ITRAD = 0 No Trading Day effect is estimated. (**Default**).  
 = 1 # of (M, T, W, Th, F) - # (Sat, Sun) x (5/2). [One-variable specification].  
 = 6 # M - # Sun. # T - # Sun, ..., # Sat - # Sun. [Six-variable specification]

- = 7 As the previous case, but with Leap-Year correction. [Seven-variable specification]
- = -1 As ITRAD = 1, but a pretest is made
- = -6 As ITRAD = 6, but a pretest is made.
- = -7 As ITRAD = 7, but a pretest is made.

For very short series the value ITRAD = -1 is recommended.

### 3.4 NEW INPUT PARAMETERS

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

- SENS = 0 Low sensitivity;
- = 1 Medium sensitivity; (default value)
- = 2 High sensitivity.

The parameter SENS sets two parameters,  $k_1$  and  $k_2$ .

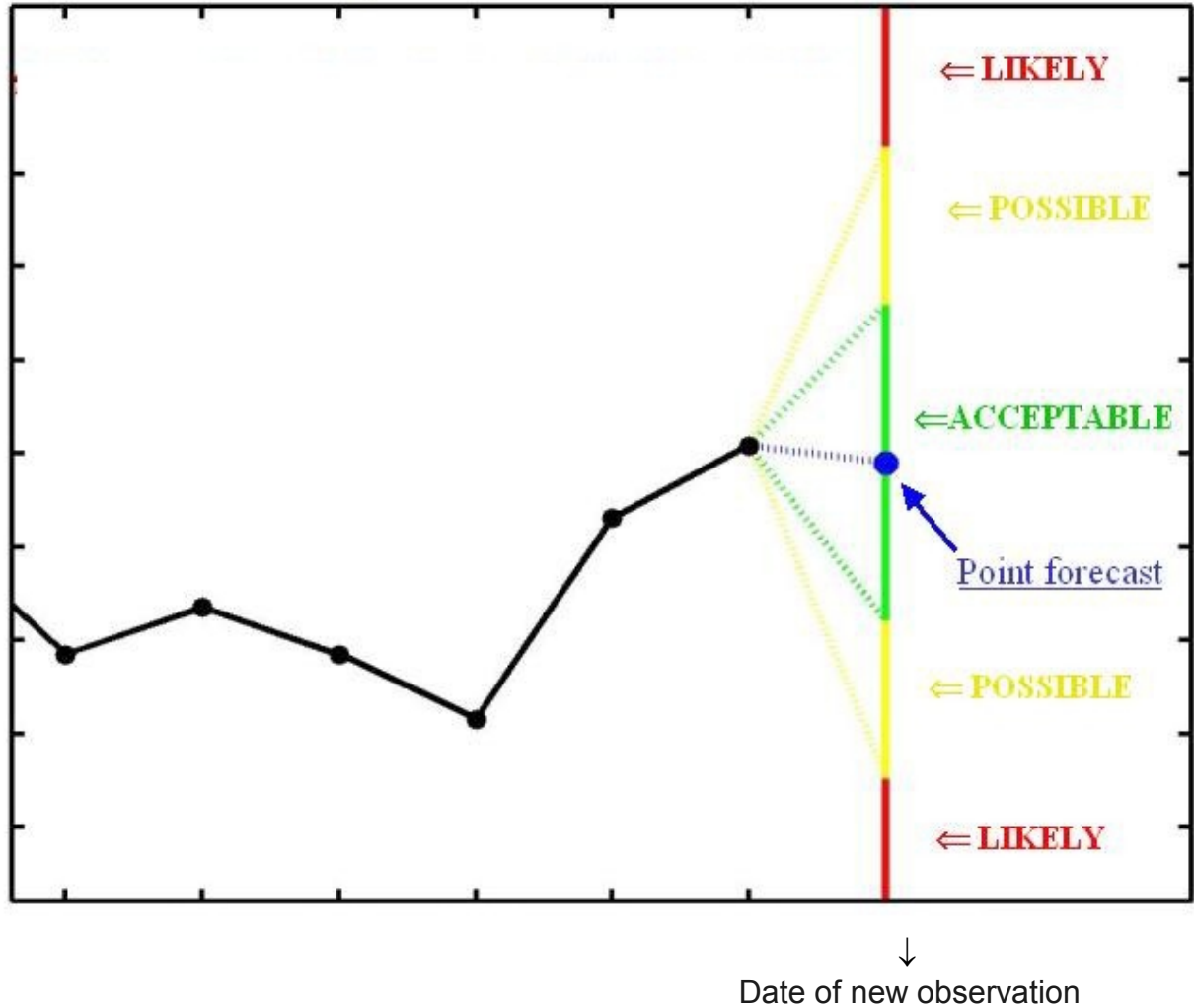
Let  $t$  = forecast error/standard deviation of residuals. Then:

If, for a particular series,

- $|t| > k_2$ , the new observation in the series is classified as "likely" to contain an error.
- If  $k_1 < |t| \leq k_2$ , the new observation is classified as containing a "possible" error.
- If  $|t| \leq k_1$ , the new observation is accepted as without error.

Figure 1

## Likely and Possible Errors



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

SENS = 0	$k_1 = 5$	$k_2 = 6$
SENS = 1	$k_1 = 4$	$k_2 = 5$
SENS = 2	$k_1 = 3$	$k_2 = 4$

These values can be changed. By setting

SENS  $\geq 3$ ,

one can enter the new values of  $k_1$  and/or  $k_2$ . For a given data set,

- larger values of the  $k$ 's imply that less elements will appear in the set of likely or possible errors.

- The values of the k's should be large enough to avoid serious "size" problems that would yield too many spurious errors. For example, if  $k_2 = 2$ , in 100 000 series, 5 000 would be expected to be spuriously detected as errors, clearly an unacceptable number.
- The k-values should depend on the type of series treated. For example, series with unstable models and large forecast errors will require larger values of the k's.
- Series that have not been subject to careful control of errors in the past are likely to require higher k's than series that have been looked at with more care, and whose quality is higher.

### **Avoiding overdetection of irrelevant series.**

By looking at the standardized forecast errors, all series in the set are treated as equally important. Yet one series can be in the order of 10, while other can be in the order of  $10^9$ . Often, it is desirable to "ignore" very small components (series of little interest) that can show nevertheless very large standardized forecast. The following parameter allows the program to ignore these irrelevant series.

MINABS = 0    parameter is inactive (all series will be considered). **Default.**  
           = k    (real # > 0 ). If, for a particular series, the absolute value of the forecast error (in original units) is  $< k$ , the series is not considered in the test for possible or likely errors.

### **3.5 SUMMARY OF THE RESULTS FOR THE INDIVIDUAL SERIES**

When the parameter

NMATRIX = 1    (**default** value) summary results for the individual series are produced in the form of several matrices.  
               = 0    no summary results for the individual series are produced.

The summary results consist of 5 matrices. For each matrix, each row refers to a particular series. The columns of the matrices are the following:

### Fitted Model (“tfit”)

Nz:	Number of observations in series.
Lam:	0 if logs have been taken; 1 if levels.
Mean:	0 if model has no mean; 1 if it has a mean.
p,d,q,bp,bd,bq:	orders (P, D, Q) (BP, BD, BQ) <sub>s</sub> of the fitted ARIMA model.
SE(res):	Standard Error of Residuals.
Q-val:	Ljung-Box-Pierce Q statistics for residual autocorrelation. (1)
N-test:	Bowman-Shenton test for Normality of the residuals. (2)
SK(t):	t-value for Ho: Skewness of residuals = 0
Kur(t):	t-value for Ho: Kurtosis of residuals = 3
QS:	Pierce Qs-test for seasonal autocorrelation in residuals. (2), (3)
Q2:	Q-statistics for autocorrelation in squared residuals. (1)
Runs:	t-test for runs (randomness) in signs of residuals.

(1)  $\chi_k^2$  where k = degrees of freedom. For Q, k = number of AC included in the test (default = 24 for monthly data, 16 for quarterly, ...) minus the number of ARIMA model parameter estimates. For Q2, the latter parameter estimates are ignored.

(2)  $\chi_2^2$ .

(3) when the lag-12 autocorrelation is negative, QS is unrelated to seasonality and no number is printed.

### ARMA Parameters (“tarmapar”)

The order is the following.

Estimate of the regular AR polynomial  $(1 + \phi_1 B + \phi_2 B^2 + \phi_3 B^3)$

Estimate of the seasonal AR polynomial  $(1 + \phi_s B^s)$

Estimate of the regular MA polynomial  $(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)$

Estimate of the seasonal MA polynomial  $(1 + \theta_s B^s)$

The associated t-values are also given

### Deterministic Effect ( total ) (“tdeterm”)

TD:	Number of Trading Day variables.
EE:	Presence / absence of Easter effect.
# OUT:	Total number of Outliers
AO:	Number of Additive Outliers
TC:	Number of Transitory Change outliers.
LS:	Number of Level Shift outliers.
REG:	Number of ( additional) regression variables.
MO:	Number of missing observations

### Calendar Effect (“tcalend”)

TD1, ..., TD6: Estimators of Trading Day variable effects.  
LY: Estimator of Leap-Year effect.  
EE: Estimator of Easter effect.

The associated t-values are also provided.

### Outliers (“toutlier”)

Detected and corrected outliers are listed; first Additive Outliers, then, Transitory Changes, and finally, Level Shifts. For each outlier, the date and associated t-value are given.

## 3.6 SUMMARY RESULTS FOR THE SET OF SERIES

When the input file contains a very large number of series, inspection of the results for each one of the series may be of little interest or unfeasible. It is possible however to obtain aggregate results for the full set of series through the parameter MODELSUMM. When

MODELSUMM = 1 , an aggregate summary is produced. **(Default)**  
= 0 , no aggregate summary is produced.

The summary consists of several tables and figures.

The first table presents, for the full set of series, the number of series (and the % this number implies)

- for which logs or levels are chosen,
- that need regular differencing,
- that need seasonal differencing,
- that are stationary or not,
- that are purely regular (no seasonal polynomials in model),
- that follow the default (Airline) model.

In the second table, for all possible combinations of regular and seasonal differences that the program considers, namely,

no differencing,  $\nabla$ ,  $\nabla^2$ ,  $\nabla_s$ ,  $\nabla\nabla_s$ ,  $\nabla^2\nabla_s$ ,

the number of series falling into each combination is given.

The third table gives information on the orders of the stationary ARMA model (the orders of the AR and MA, regular and seasonal, polynomials.)

The fourth table gives information on Missing Observations, Outliers (for the 3 types), Calendar effects (Trading Day and Easter) and possible additional regression variables.

The fifth table presents some summary statistics (mean, standard deviation, maximum and minimum values) for the length of the series, the number of parameters and the number of outliers per model, and a sequence of diagnostics (residual autocorrelation; skewness, kurtosis, and normality of the residuals; residual seasonality; possible nonlinearity; randomness of signs in residuals.)

For all the diagnostics, the number of cases for which the statistics falls in the 1% rejection region is also given, as well as the percent of series that pass the tests.

The figures are, in essence, histograms showing the distribution of the length of the series; the distribution of the ARMA parameters; the distribution of the outliers (classified by type); and the comparison between the histogram of each diagnostic statistics and its asymptotic distribution.

#### 4. INPUT FILE

An Excel macro (**Xlsts.xls**) is provided with the programs (in [www.bde.es](http://www.bde.es)) that transforms an Excel file in practically any format, into an appropriate Excel and/or text input file for TERROR, either through TSW or through TRAMO-DOS. The appropriate formats are as follows.

##### 1. Excel

Assume there are  $k$  series ( $k = 1, 2, \dots$ ). Then the  $j$ -th column of the file ( $j = 1, \dots, k$ ) should have the following structure:

Row 1:  $j$ -th series name

Row 2: NZ SY SP MQ ,

where **NZ** = number of observations (including missing values).

**SY** = Starting year.

**SP** = Starting period of the year (for monthly data, 1 if the first observation is a January, 2 if a February, etc.)

**MQ** = number of observations per year (12 if series is monthly, 4 if quarterly, ...)

The first three parameters may vary for the different series in the file; in order to obtain sensible aggregate results, **MQ** should be the same for all

series.

Row 3 and those following: Observations. Row 3 will contain the most distant observations, Row (NZ + 2) will contain the most recent ones, read from left to right.

Missing observations should be entered as -99999.

Example of EXCEL Input File:

The screenshot shows a Microsoft Excel spreadsheet titled "Ejinputfile.xls". The spreadsheet has columns labeled A through I and rows numbered 1 through 48. The data is organized as follows:

	A	B	C	D	E	F	G	H	I
1	indi1	indi2	indi3	indi4	indi5	indi6			
2	40 198114	46 198114	38 198114	46 198114	44 198114	36 198114			
3	36,0	0,0	60,0	60,0	96,0	227,3			
4	106,7	-90,2	140,0	140,0	-99999,0	519,4			
5	212,4	-43,2	240,0	240,0	-99999,0	838,7			
6	352,3	-64,8	270,0	270,0	-99999,0	1182,1			
7	528,1	-85,4	330,1	330,3	772,8	1550,9			
8	740,4	-45,0	418,0	410,9	-99999,0	1890,7			
9	988,9	24,0	523,5	511,4	-99999,0	2237,6			
10	1272,4	-49,1	546,0	540,6	-99999,0	-99999,0			
11	1593,1	40,1	624,5	602,3	2257,7	2859,0			
12	1949,1	51,3	698,5	682,6	-99999,0	3205,2			
13	2341,8	134,2	818,6	783,9	-99999,0	3602,4			
14	2770,5	228,4	841,9	813,5	-99999,0	4050,5			
15	3235,6	327,4	926,5	875,8	4489,5	4545,2			
16	3737,2	392,9	1009,0	956,8	-99999,0	5161,0			
17	4274,9	441,1	1133,0	1058,5	-99999,0	5864,1			
18	4849,3	373,2	1161,0	1089,2	-99999,0	6656,5			
19	5460,8	298,6	1266,4	1152,7	7025,8	7491,1			
20	6108,6	263,0	1355,2	1234,2	-99999,0	8357,4			
21	6790,8	232,6	1430,5	1334,5	-99999,0	-99999,0			
22	7509,1	185,4	1459,3	1365,6	-99999,0	10093,1			
23	8263,0	200,6	1554,0	1429,0	10017,5	10927,3			
24	9052,7	241,5	1639,0	1510,6	-99999,0	11731,1			
25	9880,3	178,8	1748,8	1613,0	-99999,0	12547,9			
26	10743,5	83,7	1775,0	1643,9	-99999,0	13400,2			
27	11640,8	169,9	1825,6	1705,7	13636,3	14295,7			
28	12573,2	227,2	1890,4	1786,6	14690,8	15280,3			
29	13541,5	275,5	1998,4	1889,2	15815,4	16326,2			
30	14546,7	373,2	2029,9	1921,2	16949,8	17479,5			
31	15587,3	460,0	2074,6	1982,8	18121,9	18686,5			
32	16664,1	555,0	2141,5	2064,0	19360,6	19987,4			
33	17776,7	634,9	2243,9	2166,7	20655,5	21324,9			
34	18924,4	586,2	2268,2	2198,0	21778,8	22697,8			
35	20109,8	669,6	2331,8	2261,5	23111,2	24102,8			
36	21330,9	767,3	2394,7	2342,8	24493,0	25531,0			
37	22587,7	859,9	2487,3	2445,3	25934,8	26958,6			
38	23880,8	958,9	2513,6	2477,3	27353,3	28406,5			
39	25209,3	1037,6	2569,1	2540,5	28816,1				
40	26573,4	1054,4	2624,7	2621,6	30252,5				
41	27974,4	1154,1		2725,2	31865,4				
42	29414,2	1120,3		2760,2	33316,3				
43		1215,4		2825,1	34963,7				
44		1314,0		2906,3	36629,7				
45		1413,9		3010,0	38387,3				
46		1464,8		3044,7	40060,2				
47		1417,2		3111,8					
48		1377,7		3194,7					

2. The input file for DOS is a standard multiple-series TRAMO input file. It should be named *serie*, and placed in the directory TRAMO. It consists of a sequence of series. Each series should be entered as follows:

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

Second line:                NZ    SY    SP                    MQ (free format)

Third and next lines:      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), and NZ, SY, and SP are as before. Z(.) is the array (or matrix) of observations (if a matrix, read row by row).

For each missing observation, the code -99999. must be entered.

The first series (and only the first series) is followed by the INPUT namelist, entered as follows:

    \$INPUT ERROR = 1, (if desired) other parameter = parameter value, ... \$,  
where only non-fixed, non-default parameter values need to be entered. Executing TRAMO, the ERROR application is obtained.

An example of a DOS input file is presented. The file consists of quarterly series. The fixed configuration implied by ERROR will apply. The only parameters that depart from the default value are

MQ = 4 , which specifies that the series is quarterly,

INT2 = 0, which specifies that outlier detection and correction should be performed for the complete estimation period (from 1 to NZ-1).

Notice that the third series contains the value -99999, indicating missing values.

## EXAMPLE OF AN INPUT FILE

```

1      SERIESONE
48 1986 1 4
      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=0 TERROR = 1 $
2      SERIESTWO
48 1986 1 4
      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      SERIESTHREE
47 1986 2 4
      10910      -99999.    10542
      10640      -99999.    -99999.    14060
      12938      -99999.    -99999.    11574
      -99999.    13572     -99999.    -99999.
      12389      11901     -99999.    15877
      -99999.    18517     -99999.    19784
      20751      25736     -99999.    48067
      56396      63402     -99999.    62345
      56001      47451     -99999.    35479
      -99999.    28343     -99999.    27825
      26382      25333     -99999.    22783
      26999      28101      28883     37710

4      SERIES FOUR
29 1989 4 4
      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

```

## 5. MAIN OUTPUT FILE

When **TERROR** is executed, the results of the quality control check are contained in the file **list.out** (located in the subdirectory **TRAMO\OUTPUT**.) The file lists, first, the parameters used in the application; then, the series that have been found suspicious. Each line contains the name of the series, the date of the observation for which the test has been made, the new reported value and what would have been its forecast one period ago, the forecast error, its standard deviation, the associated t-value, and the classification of the new observation as a “Likely” or “Possible” error.

As already mentioned, when **NMATRIX = 1**, several files are produced with matrices where each row contains the summary results for each series in the set. Additionally, when **MODELSUMM = 1**, several tables and figures summarize the aggregate results for the full set of series.

## 6. AN EXAMPLE

To illustrate the output files and figures produced by the program, 200 real monthly time series were arbitrarily selected. The series have different lengths (from 42 to 522 observations), belong to different countries and different periods, and represent different types of economic variables (monetary and financial variables, foreign trade variables, production and price indices, a variety of short-term indicators, etc.) They were moved in time so that the last observation was always June 2002. (Given that the series in the set are unlikely to have a large error, in order to obtain a better illustration, for some of the series the last observation was perturbed: for randomly selected 25 pairs of consecutive series, the last observations were exchanged.) Then, the last observation was considered a newly reported value, and **TERROR** was applied to detect possible or likely errors in the 200 new observations. The set is used simply to illustrate the program output; a more detailed discussion of an application can be found in Luna and Maravall (1999).

Table 1 contains the main results of the test (the file **list.out**); tables 2, 3, and 4, the summary statistics obtained when **NMATRIX = 1** (in this example, the matrices **tfit**, **tdeterm**, and **tout**) for a subset of 30 series.

**TABLE 1: RESULTS OF THE TEST (file "list.out")**

DATE : 2002-09-26 14:51:20

Input Parameters :

int2=-3 terror= 1 k1= 4.000 k2= 5.000

SERIES TITLE	Date	New Value	Forecast	Diff.	StdDev	T-Value	Results
Series 25	06-2002	13.90000	32.47640	-18.57640	3.474482	-5.346524	Likely
Series 28	06-2002	-5.700000	11.87520	-17.57520	2.215275	-7.933642	Likely
Series 36	06-2002	5.162784	5.223978	-.0611942	0.0152308	-4.017802	Possible
Series 76	06-2002	17.83883	14.55002	3.288818	0.7272889	4.522024	Possible
Series 90	06-2002	23.69046	26.96759	-3.277133	0.6237600	-5.253837	Likely
Series 110	06-2002	4.739614	4.807734	-.0681206	0.0157679	-4.320194	Possible
Series 140	06-2002	94.81960	3.959579	90.86002	3.657171	24.84434	Likely
Series 153	06-2002	67.93730	-21.27136	89.20866	5.209442	17.12442	Likely
Series 162	06-2002	665.0469	834.5731	-169.5262	37.98918	-4.462487	Possible
Series 167	06-2002	129.2012	46.15271	83.04849	4.026240	20.62681	Likely
Series 168	06-2002	-752.6359	-601.7935	-150.8424	15.91433	-9.478405	Likely
Series 169	06-2002	6.614892	6.685780	-.0708876	0.0062549	-11.33313	Likely
Series 171	06-2002	-807.7597	-833.2131	25.45344	6.323477	4.025228	Possible
Series 176	06-2002	880.9004	887.3626	-6.462161	1.530289	-4.222837	Possible

Summary Statistics

200 Series were tested.

6 Releases were suspicious (possibly wrong).

8 Releases were very suspicious (likely wrong).

0 Series produced a Run-Time EXCEPTION.

0 Series did not match TERROR memory constraints.

186 Series passed the plausibility tests.



SUMMARY RESULTS FOR THE INDIVIDUAL SERIES

**TABLE 2. MODEL FITTED AND DIAGNOSTICS (file tfit)**

n	Nz	Lam	Mean	P	D	Q	BP	BD	BQ	SE(res)	BIC	Q-val	N-test	SK(t)	KUR(t)	QS	Q2	RUNS
1	42	0	1	0	1	1	0	1	1	0.0137032	-8.33656	8.925	1.42	-0.536	-1.06	1.52	10.37	-0.786
2	42	0	0	0	1	1	0	1	0	0.0213987	-7.60621	8.139	1.40	1.12	0.381	1.52	8.365	-0.770
3	44	0	0	1	2	0	0	1	1	0.0134521	-8.30121	6.841	1.65	0.962	0.853	1.52	7.520	0.786
4	60	0	1	0	1	1	0	1	1	0.0017026	-12.5690	22.36	6.59	-2.30	1.15	3.74	20.44	-0.302
5	60	1	1	0	1	0	0	1	1	0.1631566	-3.50408	24.85	5.76	-2.24	0.864	1.94	29.99	-1.21
6	80	0	0	1	1	0	0	1	1	0.0047282	-10.6122	8.158	0.351	0.584	0.098	0.704	21.24	0.496
7	83	0	1	3	1	1	0	0	0	0.0074806	-9.54232	29.06	2.88	-1.08	-1.31	0.144	20.97	-0.227
8	102	1	0	0	1	1	0	1	1	947.9065	13.7873	21.83	0.220	-0.267	-0.385	2.51	16.65	-0.214
9	102	1	0	0	0	0	0	1	1	1697.962	14.9135	13.83	0.344	-0.407	-0.423	0.466	15.32	-0.853
10	102	1	0	3	0	0	0	1	1	2088.101	15.4438	18.87	0.649	-0.061	-0.804	0.815	37.61	1.28
11	112	1	0	0	1	1	0	1	1	484.8674	12.4770	28.85	7.00	-0.690	2.55	-	40.16	0.817
12	120	1	0	0	1	1	0	1	1	1.088913	0.239301	20.14	0.374	0.479	-0.380	0.902	21.69	-0.976
13	120	1	0	0	1	1	0	0	0	0.9008736	-0.176862	21.28	3.74	1.10	1.59	-	19.96	-0.925
14	120	1	0	0	1	1	0	0	0	0.9744592	-0.019826	25.69	2.21	0.611	-1.36	-	19.77	0.370
15	120	1	0	1	1	0	0	0	0	0.9863089	0.004348	14.64	0.397	0.036	-0.629	-	13.70	0.370
16	144	0	0	0	1	1	0	1	1	0.0017695	-12.5851	22.16	14.0	1.21	3.54	-	16.25	1.24
17	155	0	0	0	1	1	0	1	1	0.1170087	-4.23510	20.81	0.735	-0.230	-0.826	0.101	29.59	-0.169
18	180	0	0	0	2	1	0	0	0	0.0032096	-11.4126	27.53	1.15	1.05	0.231	-	54.49	0.303
19	240	1	0	0	1	1	0	1	1	6.451238	4.28677	24.09	32.8	2.18	5.30	-	24.87	0.711
20	240	1	0	0	1	1	0	0	0	4.604365	3.37078	22.25	41.9	3.60	5.38	-	86.51	-2.42
21	267	0	0	0	1	1	0	1	1	0.1101631	-4.34004	11.82	0.412	-0.467	0.441	0.434	25.14	0.632
22	292	1	0	0	2	1	0	1	0	25.67014	6.73953	35.18	2.22	0.973	1.13	-	57.24	0.989
23	292	1	0	0	2	1	0	1	1	18.39150	5.99004	23.81	41.3	1.11	6.33	-	42.03	-0.122
24	317	0	1	0	1	1	0	0	0	0.0282578	-6.96743	34.73	1.19	0.011	1.09	-	19.07	1.15
25	327	0	0	1	1	1	0	1	1	0.0338586	-6.65015	50.60	1.75	-0.997	-0.868	7.49	29.90	0.114
26	329	0	1	1	1	0	1	0	1	0.0056447	-10.1935	38.90	4.81	2.16	-0.392	-	15.20	-
27	340	1	0	0	2	1	0	0	0	4.063632	2.88981	27.78	7.66	2.17	1.72	-	25.00	0.110
28	491	0	1	3	1	0	0	0	1	0.0092988	-9.14423	32.10	0.959	0.157	-0.967	6.30	29.21	-0.552
29	503	0	0	3	1	1	0	1	1	0.0226754	-7.37198	40.72	9.17	-0.229	3.02	-	70.17	-0.643
30	503	1	1	0	1	1	0	0	0	0.7844712	-3.340115	28.11	0.716	-0.813	-0.234	-	60.43	0.363

**TABLE 3. DETERMINISTIC EFFECTS (file tdeterm)**

n	TD	EE	#OUT	AO	TC	LS	REG	MO
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	2	0	0	2	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	1	0	1	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0
16	0	0	1	0	0	1	0	0
17	0	0	0	0	0	0	0	0
18	0	0	2	0	1	1	0	0
19	0	0	27	22	5	0	0	0
20	0	0	16	11	4	1	0	0
21	0	0	2	1	1	0	0	0
22	0	0	14	6	1	7	0	0
23	0	0	8	2	2	4	0	0
24	0	0	9	7	2	0	0	0
25	0	0	5	3	1	1	0	0
26	0	0	7	3	1	3	0	0
27	0	0	5	2	1	2	0	0
28	0	0	15	4	3	8	0	0
29	0	0	14	7	5	2	0	0
30	0	0	12	8	0	4	0	0

**TABLE 4. OUTLIERS (file "toutliers")**

1	
2	
3	LS01(1098, 5.46) LS02(1198, 3.28)
4	
5	
6	
7	AO01(1298, -3.02)
8	
9	
10	
11	TC01(1290, 4.09)
12	
13	
14	
15	
16	LS01(0192, 7.15)
17	
18	TC01(0783, -3.68) LS01(0186, 4.33)
19	AO01(0598, 49.82) AO02(0597, 26.09) AO03(0797, 29.45) AO04(0793, 23.52) AO05(0997, 18.74) AO06(0691, 20.40) AO07(0893, 12.53) AO08(0994, 9.90) AO09(0999, 8.42) .....
20	AO01(0597, 66.94) AO02(0807, 22.36) AO03(1298, 28.09) AO04(1007, 15.20) AO05(0298, 14.33) AO06(1202, 12.49) AO07(1198, 10.02) AO08(1199, 8.01) AO09(0790, 6.33) .....
21	AO01(0875, -4.78) TC01(0181, -3.34)
22	AO01(1291, 10.91) AO02(0984, 7.03) AO03(1288, 5.01) AO04(1076, -4.08) AO05(1192, -3.14) AO06(0573, -2.95) TC01(0576, -4.20) LS01(0579, -5.80) LS02(0976, -5.02) .....
23	AO01(1291, 8.96) AO02(0984, 5.10) TC01(1288, 4.41) TC02(0576, -3.67) LS01(0579, -6.23) LS02(0387, 3.56) LS03(1176, 3.34) LS04(0871, 3.76)
24	AO01(0701, -6.38) AO02(0501, -5.39) AO03(0789, 5.21) AO04(0101, -3.77) AO05(0884, -3.75) AO06(0878, -3.72) AO07(1292, -3.50) TC01(0100, -5.07) TC02(0183, 3.93)
25	AO01(0874, 6.00) AO02(0391, -3.03) AO03(0486, 2.97) TC01(0174, 4.16) LS01(0167, -3.58)
26	AO01(0363, 5.09) AO02(1285, -3.53) AO03(0668, 3.16) TC01(0576, 8.24) LS01(0663, -4.96) LS02(1174, 3.81) LS03(0177, 3.31)
27	AO01(0487, 4.40) AO02(0187, 4.11) TC01(1292, 3.85) LS01(1196, -4.96) LS02(0391, 3.62)
28	AO01(0881, -4.17) AO02(0862, -4.35) AO03(1076, -4.10) AO04(0380, 2.95) TC01(0865, -4.59) TC02(0873, -4.00) TC03(1082, -3.47) LS01(0963, 4.48) LS02(0175, -3.46) .....
29	AO01(1185, 5.71) AO02(0486, 6.57) AO03(0286, 6.57) AO04(1194, -4.55) AO05(1200, 4.40) AO06(0489, 4.03) AO07(0197, -3.67) TC01(0185, 4.71) TC02(0795, -6.08) .....
30	AO01(0272, -10.06) AO02(0179, -9.81) AO03(0478, 6.52) AO04(0685, 4.40) AO05(0577, 4.38) AO06(0187, -4.03) AO07(0489, 3.95) AO08(0690, 3.23) LS01(0174, -7.33) .....

As was mentioned earlier, when `MODELSUMM=1` , an aggregate summary is provided. It consists of the following tables and figures.

Series in file: 200

Series processed: 200

Input parameters: `error= 1 modelsumm= 1 k1= 4.000 k2= 5.000`

Series monthly

GENERAL	# of series	%
Levels	143	71,5
Logs	57	28,5
Regular diff	164	82
Seasonal diff	110	55
Stationary	20	10
Non stationary	180	90
Purely regular	66	33
Default (Airline model)	53	26,5

DIFFERENCES	D=0	D=1	D=2	Total
BD=0	20	61	9	90
BD=1	16	73	21	110
-----	-----	-----	-----	-----
Total	36	134	30	200

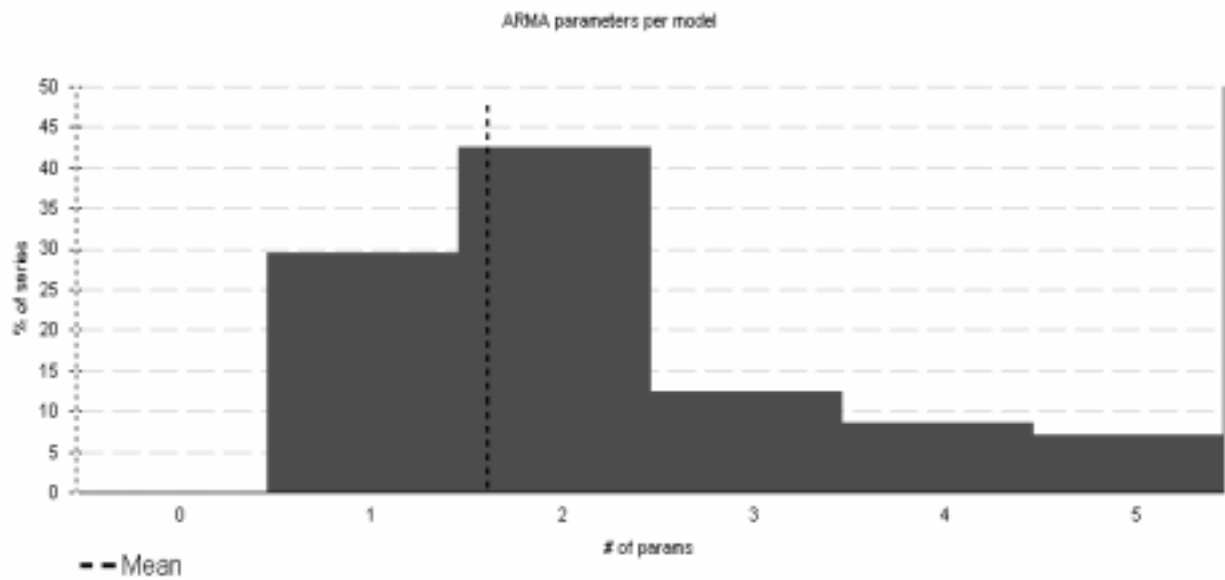
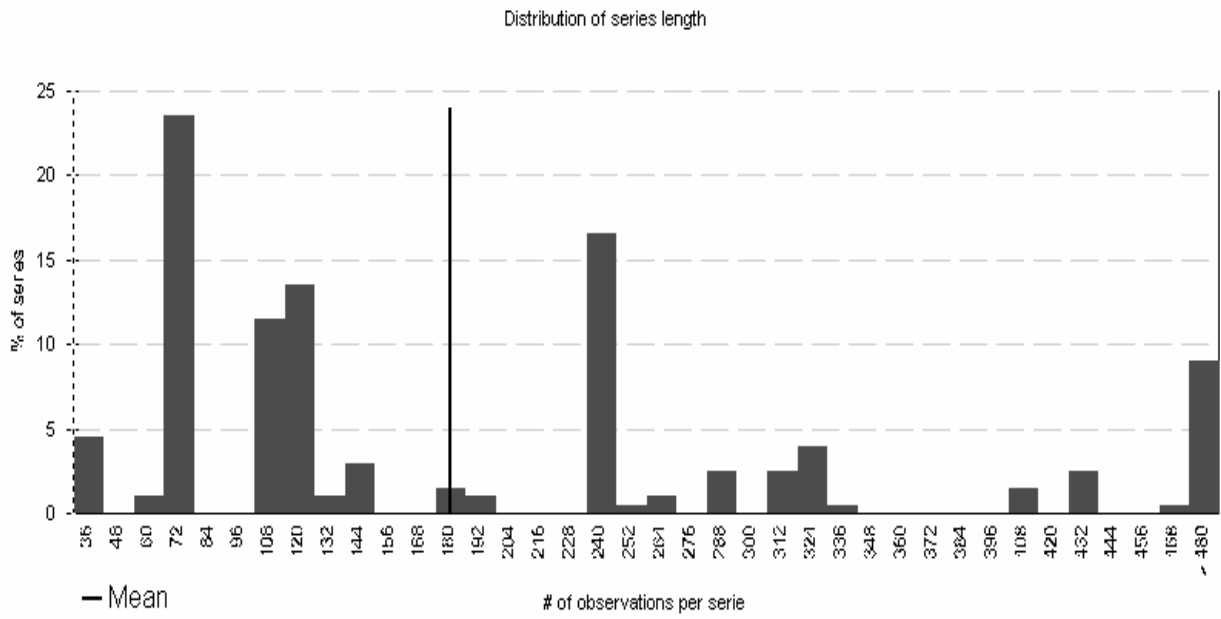
ARMA PARAMETER	P	Q	BP	BQ	
% of series with 0	60	26	92	41,5	
% of series with 1	21	66,5	8	58,5	
% of series with 2	8	4,5	0	0	
% of series with 3	11	3	0	0	
-----	-----	-----	-----	-----	
Total > 0	40	74	8	58,5	
-----	-----	-----	-----	-----	
Average # of param. per serie	0,7	0,8	0,1	0,6	2,2

SUMMARY	Mean	SD	Max	Min	% of series		
					Approx. 1% CV	Beyond 1% CV	that pass the test(99%)
Length	187,3	130,6	522	41	--	--	--
# of ARMA param. per serie	2,2	1,2	5	1	--	--	--
# of outliers per serie	5,1	7,9	29	0	--	--	--
Q-Val	24,3	9,4	68,02	6.393	40,3	4	96
N-Test	6,1	13	88	0,003	9,21	14	86
Skewness	0,2	1,5	6,83	-3,58	2,58	8,5	91,5
Kurtosis	0,6	1,8	8,91	-1,72	2,58	12	88
Qs	--	--	--	--	9,21	2,5	97,5
Q2	25,7	14	95,84	6.454	43	9	91
Runs	-0,1	0,9	2,36	-2,51	2,58	10,5	89,5

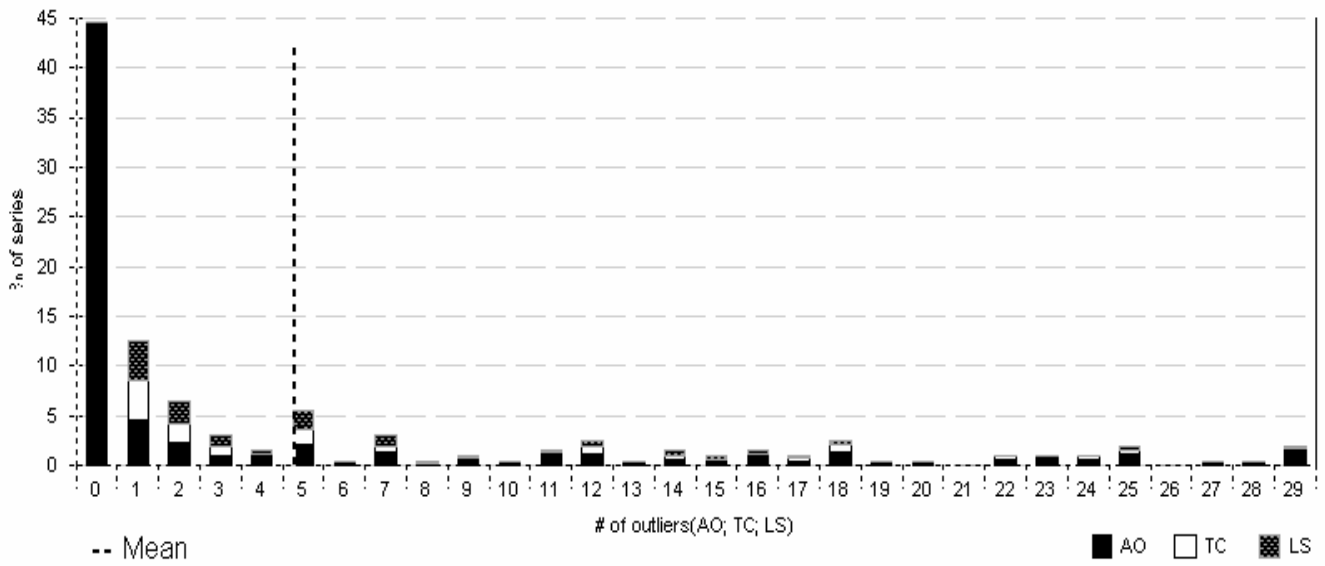
MO & REGRESSION	MO	AO	TC	LS	Tot. Outlier	TD	EE	Tot. Calendar
% of series with	0	43,5	37,5	33	55,5	0	0	0
Average # por serie*	--	3	1	1	5	--	--	--
Maximum #	--	28	17	17	29	--	--	--
Minimum #	--	0	0	0	0	--	--	--

\* For column MO, only the series with MO's are considered

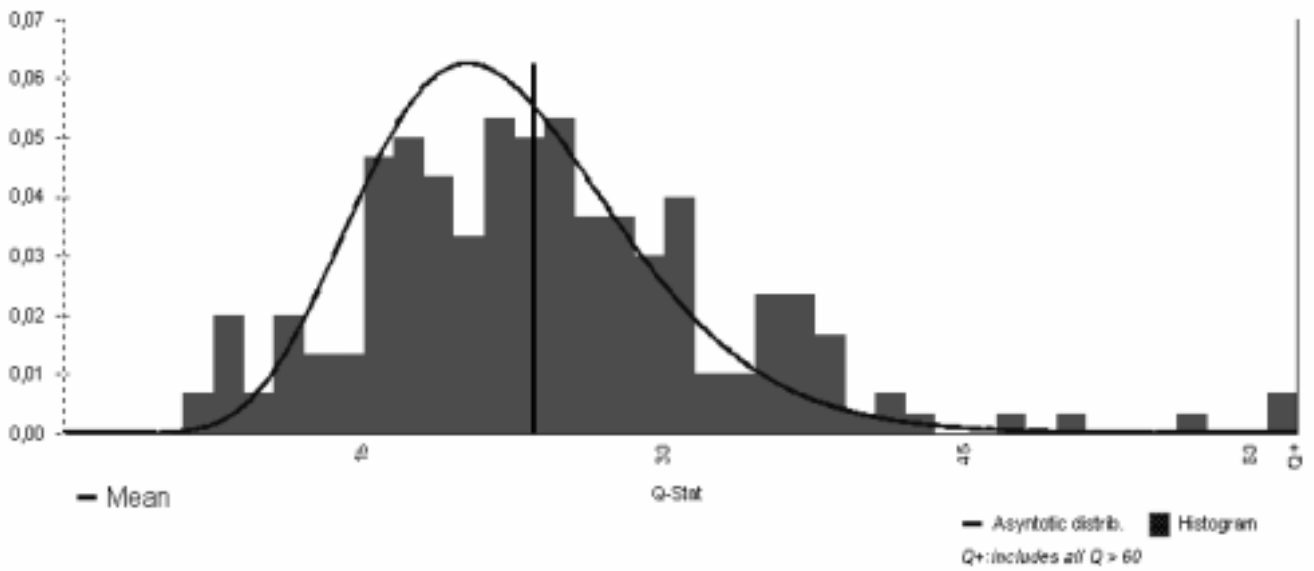
# FIGURES



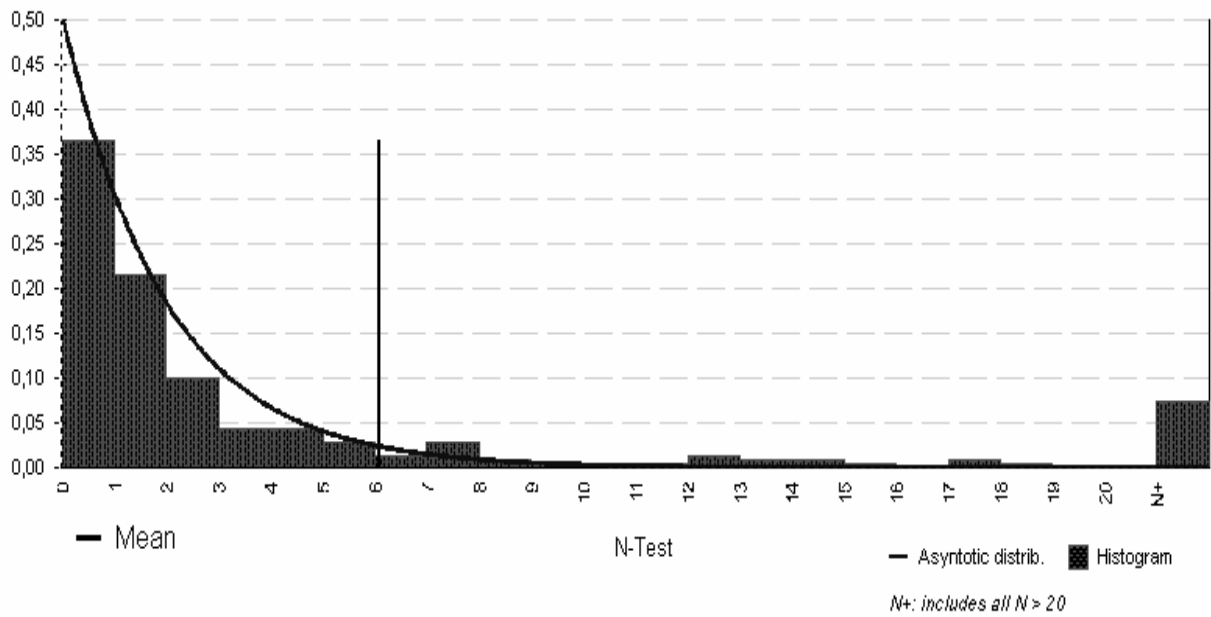
Distribution of outliers



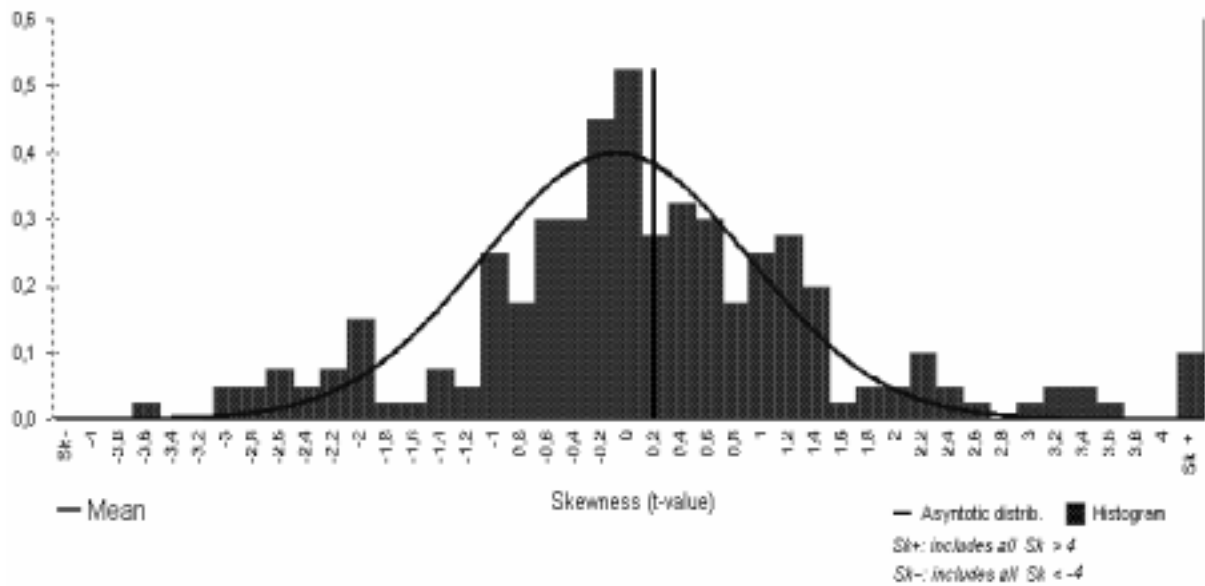
Residual autocorrelation (Q)

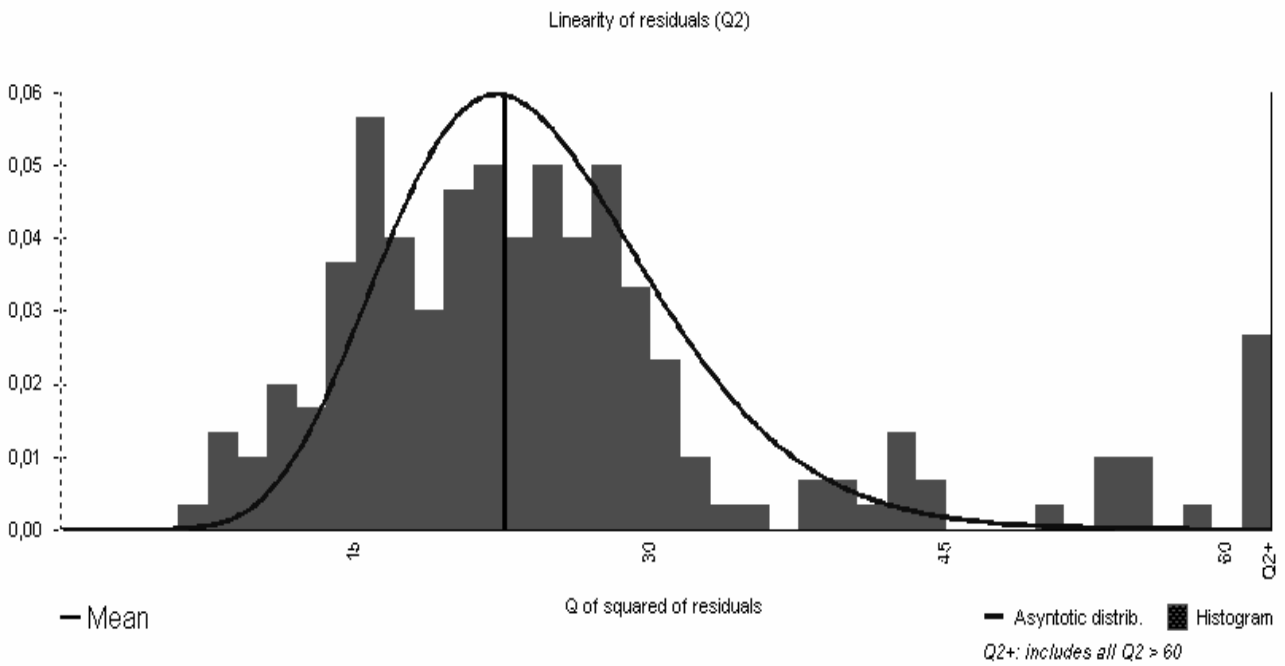
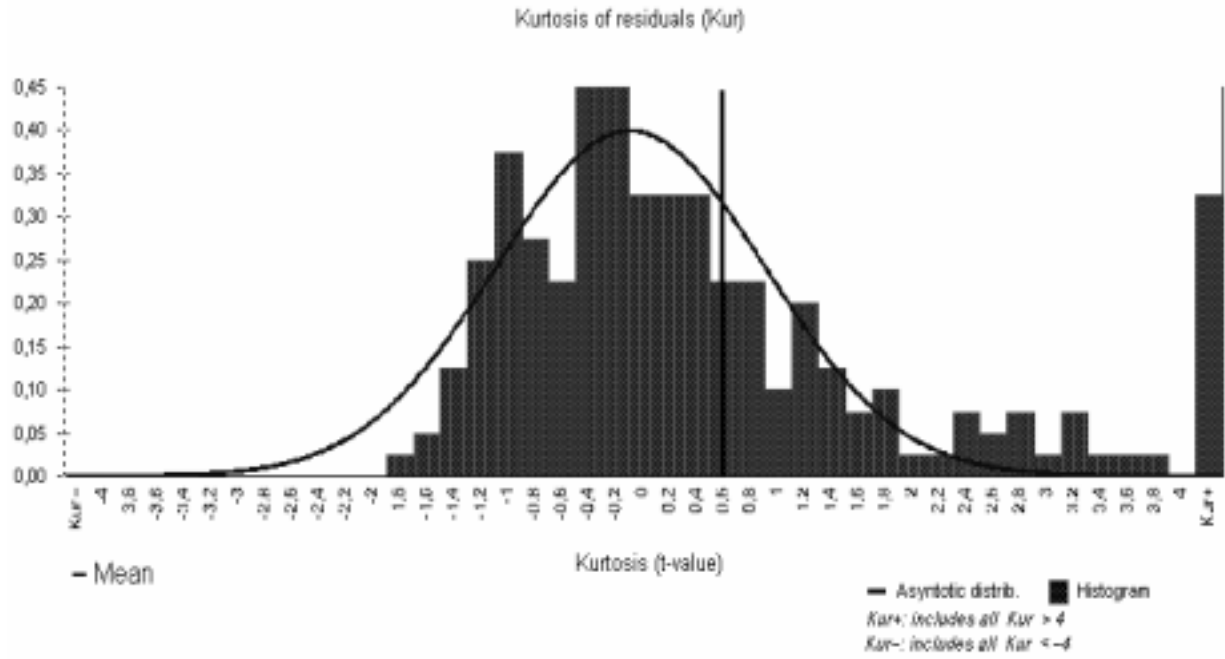


### Normality of residuals (N)

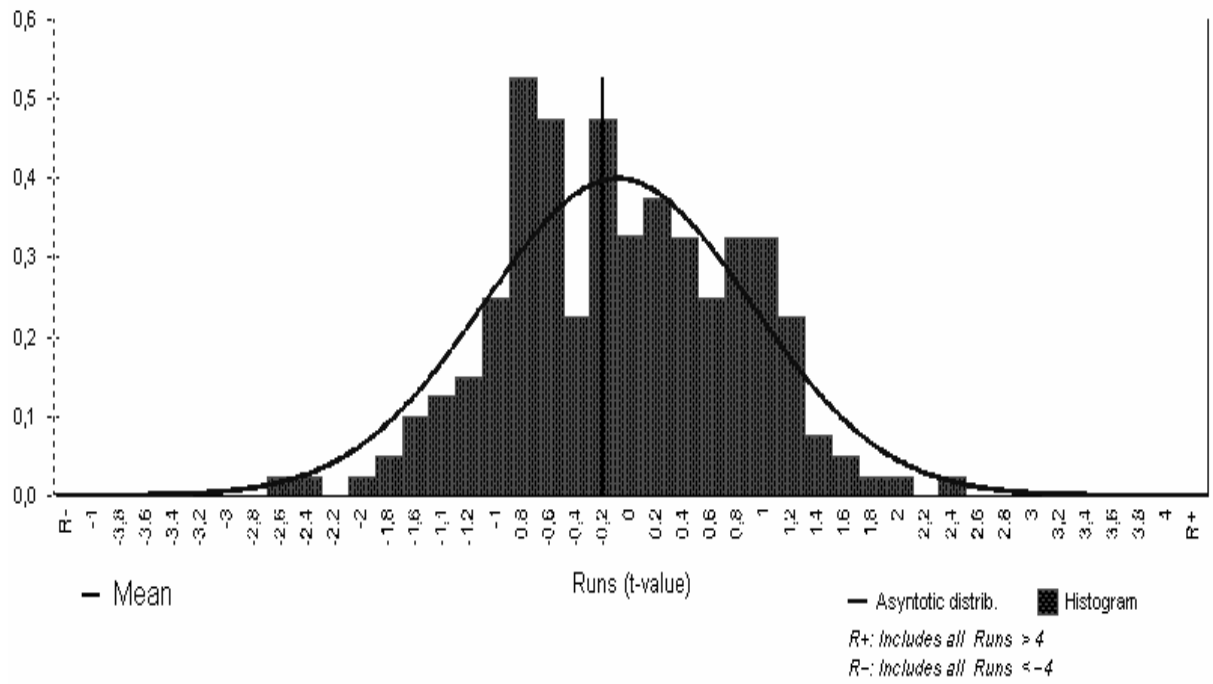


### Skewness of residuals (Sk)





### Randomness of signs in residuals (Runs)



## 7. REFERENCES

- ALWAN, L.C. and ROBERTS, H.V. (1988), "Time-Series Modeling for Statistical Process Control", *Journal of Business and Economic Statistics* 6, 87-95.
- ANDERSON, B. and MOORE, J. (1979), *Optimal Filtering*, New Jersey: Prentice Hall.
- BARNETT, V. and LEWIS, T. (1998), *Outliers in Statistical Data*, NY: J. Wiley and Sons.
- BOX, G.E.P. and JENKINS, G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- CAPORELLO, G., MARAVALL, A. and SÁNCHEZ, F.J. (2001), "Program TSW Reference Manual", Working Paper 0112, Servicio de Estudios, Banco de España.
- CHEN, C. and LIU, L.M. (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association* 88, 284-297.
- FISCHER, B. and PLANAS, C. (2000), "Large Scale Fitting of Regression Models with ARIMA Errors," *Journal of Official Statistics*, 16, 173-184.
- GÓMEZ, V. and MARAVALL, A. (1992), "Time Series Regression with ARIMA Noise and Missing Observations - Program TRAM", EUI Working Paper ECO No. 92/81, Department of Economics, European University Institute.
- GÓMEZ, V. and MARAVALL, A. (1993), "Initializing the Kalman Filter with Incompletely Specified Initial Conditions", in Chen, G.R. (ed.), *Approximate Kalman Filtering (Series on Approximation and Decomposition)*, London: World Scientific Publ. Co.
- GÓMEZ, V. and MARAVALL, A. (1994), "Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter", *Journal of the American Statistical Association* 89, 611-624.
- GÓMEZ, V. and MARAVALL, A. (1996), "Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User", Working Paper 9628, Servicio de Estudios, Banco de España.
- GÓMEZ, V. and MARAVALL, A. (2001), "Automatic Modelling Methods for Univariate Series", Ch.7 in Peña D., Tiao G.C. and Tsay, R.S. (eds.) *A Course in Time Series Analysis*, New York: J. Wiley and Sons.
- GÓMEZ, V., MARAVALL A., and PEÑA, D. (1999), "Missing Observations in ARIMA Models: Skipping Approach Versus Additive Outlier Approach", *Journal of Econometrics*, 88, 341-364.
- HANNAN, E.J. and RISSANEN, J. (1982), "Recursive Estimation of Mixed Autoregressive-Moving Average Order", *Biometrika* 69, 81-94.
- JONES, R. (1980) "Maximum Likelihood Fitting of Arma Models to Time Series With Missing Observations", *Technometrics* 22, 389-395.

- KOHN, R. and ANSLEY, C.F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models", *Biometrika* 72, 694-697.
- KOHN, R. and ANSLEY, C.F. (1986), "Estimation, Prediction and Interpolation for ARIMA Models with Missing Data", *Journal of the American Statistical Association* 81, 751-761.
- LUCAS, J.M. and SACCUCCI, M.S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements (with discussion)" *Technometrics* 32, 1-29.
- LUNA, C. and MARAVALL, A. (1999), "Un Nuevo Método para el Control de Calidad de los Datos en Series Temporales", *Boletín Económico*, Banco de España, May 1999, 37-44.
- MARAVALL, A. (1998), "Comments on the Seasonal Adjustment Program X12ARIMA", *Journal of Business and Economic Statistics*, 16, 155-160.
- MÉLARD, G. (1984), "A Fast Algorithm for the Exact Likelihood of Autoregressive-Moving Average Models", *Applied Statistics* 35, 104-114.
- MONSELL, BRIAN C. (2002) "An Update of the Development of the X-12-ARIMA Seasonal Adjustment Program", *Modeling Seasonality and Periodicity*, Proceedings of the 3<sup>rd</sup> International Symposium on Frontiers of Time Series Modeling, The Institute of Statistical Mathematics, Tokyo: Japan, January 2002.
- MONTGOMERY, D.C. and MASTRANGELO, C.M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data", *Journal of Quality Technology* 23, 179-204.
- MORF, M., SIDHU, G.S. and KAILATH, T. (1974), "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete-Time Systems", *IEEE Transactions on Automatic Control*, AC - 19, 315-323
- TIAO, G.C. and TSAY, R.S. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models", *The Annals of Statistics* 11, 856-871.
- TSAY, R.S. (1984), "Regression Models with Time Series Errors", *Journal of the American Statistical Association* 79, 118-124.
- TSAY, R.S. (1986), "Time Series Model Specification in the Presence of Outliers", *Journal of the American Statistical Association* 81, 132-141.