09/2007

Large Scale Applied Time Series Analysis with Program TSW (TRAMO-SEATS for Windows)

Agustín Maravall
Research Department
Bank of Spain
Alcalá, 48
28015 Madrid
Tel:+ 34 91 338 5476
Fax:+ 34 91 338 5678
e-mail address:maravall@bde.es

Abstract

The demostration will center on the application of program TSW to a large set of monthly time series. TSW is a Windows interface of updated versions of programs TRAMO (Time series Regression with Arima noise, Missing values, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series).

The program estimates a general regression-ARIMA model, and computes forecasts and interpolators for possibly nonstationary series, with any sequence of missing observations, and in the presence of outliers. The program contains an option for automatic model identification, automatic detection and correction of several types of outliers, and for pretesting and estimation of Calendar-based effects. Several types of intervention or regression variables can also be included.

Next, the program estimates and forecasts the trend, seasonal, calendar, transitory, and noise components in the series, using signal extraction techniques applied to ARIMA models. The program contains a part on diagnosis and on inference, and an analysis of the properties of the estimators and of the estimation and forecasting errors. The last part of the output is oriented towards its use in short term economic policy and monitoring.

TRAMO contains an extension (TERROR, or Tramo for ERRORs) to the problem of quality control of data in large data bases of time series; SEATS can be applied for estimation of long-term trends and (business) cycles.

The programs can efficiently and reliably handle, in an entirely automatic manner, applications to sets of many thousand series. They are already being used intensively in research, data producing agencies, policy making institutions, and business. (Perhaps the most widely used application is Seasonal Adjustment.) They are freely available, together with documentation, at the Bank of Spain web site (www.bde.es).

1. BRIEF DESCRIPTION OF TSW

Program TSW is a Windows interface of updated versions of programs TRAMO and SEATS (Gómez and Maravall, 1996) developed at the Bank of Spain (Caporello and Maravall, 2004), for applied time series analysis. The programs and associated manuals and documentation are freely available at the Bank of Spain web site (www.bde.es)

The program estimates and forecasts regression models with errors that follow ARIMA processes.

There may be:

- missing observations in the series,
- contamination by outliers,
- contamination by other special (deterministic) effects.

An important case of the latter is the trading day (TD) effect, caused by the different distribution of week-days in different months, and Easter effect (EE), which captures the moving dates of Easter.

If B : lag operator, $B\,x(t) = x(t-1)$,

given the observations $y = [y(t_1), y(t_2), \ldots, y(t_m)]$

where $0 < t_1 < \ldots < t_m$,

the model can be expressed as

$$y(t) = \sum_{i=1}^{n_{out}} \omega_i \lambda_i(B)\; d_i(t) + \sum_{i=1}^{n_c} \alpha_i\, cal_i(t) + \sum_{i=1}^{n_{reg}} \beta_i\, reg_i(t) + x(t), \qquad (2.1)$$

where

$d_i(t)$: dummy variable that indicates the position of the i-th outlier, $\lambda_i(B)$: polynomial in B reflecting the outlier dynamic pattern,

cal_i : calendar-type variable,

reg_i : regression or intervention variable,

$x(t)$: ARIMA error,

ω_i : instant i-th outlier effect,

α_i and β_i : coefficients of the calendar and regression-intervention variables, respectively,

n_{out} , n_c and n_{reg} : total number of variables entering each summation term in (2.1).

In compact notation, (2.1) can be rewritten as

$$y(t) = z'(t)\, b + x(t) \quad , \qquad (2.2)$$

where b : vector with the ω, α and β coefficients,

$z'(t)$: matrix with the columns containing the variables in the 3 summation terms of (2.1).

ARIMA model for $x(t)$:

$$\phi(B)\; \delta(B)\; x(t) = \theta(B)\; a(t) \quad , \qquad (2.3)$$

where $a(t)$: white-noise $(0,\ V_a)$ innovation.

$\phi(B)$, $\delta(B)$, and $\theta(B)$: finite polynomials in B. The first one contains the stationary autoregressive (AR) roots, $\delta(B)$ contains the nonstationary AR roots, and $\theta(B)$ is an invertible moving average (MA) polynomial.

s: number of observations per year.

The polynomials assume the multiplicative form

$$\delta(B) = \nabla^{d} \nabla_{s}^{d_s} ,$$

$$\phi(B) = \left(1+\phi_1 B + ... + \phi_p B^p\right) \left(1+\Phi_1 B^s\right),$$

$$\theta(B) = \left(1+\theta_1 B + ... + \theta_q B^q\right) \left(1+\Theta_1 B^s\right),$$

where $\nabla = 1 - B$ and $\nabla_s = 1 - B^s$.

This model will be refered to as the ARIMA $(p,d,q)\,(p_s,d_s,q_s)_s$ model.

The model consisting of (2.2) and (2.3) will be called a regression(reg)-ARIMA model.

When used automatically, the program

- tests for the log/level transformation,
- for the presence of calendar effects,
- detects and corrects for three types of outliers:
  Additive Outliers (AO), Transitory Changes (TC), and Level Shifts (LS);
- identifies and estimates by maximum likelihood the reg-ARIMA model (uses conditional likehood).
  Write the model in compact notation as
  $y(t) = z'(t)\, b + x(t)$.
  b is concentrated out of the likelihood and estimation is iterative:
- Conditional on b $\rightarrow$ ARIMA (MLE),
- Conditional on ARIMA $\rightarrow$ b (GLS).
- interpolates missing values,
- computes forecasts of the preadjustment component $z'(t)\, b$ and of the ARIMA series $x(t)$ in (2.2).

2. <u>SUMMARY OF AUTOMATIC PROCEDURE</u>

Pretest for the Log-level Specification

The test consists of direct comparison of the BICs of the default model in levels and in logs (with a proper correction).

Pretest for Trading Day and Easter Effects

Regressions using the default model for the noise and, if the model is subsequently changed, the test is redone.

(Slight bias towards underdetection.)

A Remark on the use of the default (Airline) model

Pretesting and the starting point of AMI depend heavily on the Airline model

$$\nabla \nabla_{12} z_t = \left(1+\theta_1 B\right) \left(1+\theta_{12} B^{12}\right) a_t + \mu$$

(monthly series)

Three important reasons:

- Many studies have shown it is appropiate for large number of real series (40-60%)
- The Airline model approximates well many other models.
- Excellent “benchmark” model.

Automatic Model Identification in the Presence of Outliers

The algorithm iterates between the following two stages

1. Automatic outlier detection and correction
2. Automatic model identification

The first model used is the default model.

At each step, the series is corrected for the outliers and other regression effects present at the time, and a new AMI is performed.

If the model changes, the automatic detection and correction of outliers is performed again from the beginning.

(a) AUTOMATIC OUTLIER DETECTION AND CORRECTION

1. Assume that we know the location (t=T), but not the type of outlier.

For T = T, we compute: $\hat{\omega}_{AO}(T), \hat{\omega}_{TC}(T), \hat{\omega}_{LS}(T)$

$\hat{\tau}_{AO}(T), \hat{\tau}_{TC}(T), \hat{\tau}_{LS}(T)$

and $\lambda_T = \max\left\{ \left|\hat{\tau}_{AO}(T)\right|, \left|\hat{\tau}_{TC}(T)\right|, \left|\hat{\tau}_{LS}(T)\right| \right\}$

Use $\lambda_T > C$ to test for significance.

C: An "a priori" set critical value.

2. If we don't know the timing of the outlier, we compute λ_t for t = 1, … , N and use

$\lambda = \max_t \lambda_t = \left|\hat{\tau}_{tp}(T)\right|$

If $\lambda > C$, there is an outlier of type tp (AO, TC, LS) at T.
We correct for this outlier, and start the process again to see if there is another outlier.
Outliers are removed one by one, until we obtain $\lambda_T < C$.
Now we proceed to joint estimation of the multiple outliers:

$z_t^* = z_t + \sum_{i=1}^{k} \omega_i \upsilon_i(B) I_{t_j}(t_j)$

We have to perform multiple regressions to avoid (as much as possible) masking effects.

(b) AUTOMATIC ARIMA MODEL IDENTIFICATION

Suppose the series $\{x_t\}$ follows the model

$\phi(B)\delta(B)x_t = \theta(B)a_t, \qquad a_t \overset{i.i.d.}{\sim} N(0,\sigma^2).$

TRAMO proceeds in two steps:

1. First, identify $\delta(B)$ (unit roots)
2. Second, identify the ARMA model, i.e, $\phi(B)$ and $\theta(B)$.

Identification of the Nonstationary polynomial δ (B)

To determine the appropriate differencing of the series, we discard unit root testing.
Problems with Unit Root Testing:

- When regular and seasonal U.R. may be present, available tests have low power.

For example, in

$\nabla z_t = (1-.8B)a_t$,

$\nabla_{12} z_t = (1-.8B^{12})a_t$,

U.R. would most likely be rejected due to the large MA root.

- Besides, in AMI + AODC, one may try thousands of models, where: next try depends on previous results.

→ Serious DATA MINING problem: the size of the test is a function of prior rejections and acceptances.

→ No way of knowing the true size of the test.

We follow an alternative approach:
Decide "a priori", instead of a fictitious size, the following value:
How large the modulus of a root should be in order to accept it as 1 (unit root)?

For AR and MA roots the criterion is different (roughly: unit AR roots are O.K.; unit MA roots are avoided.)

For AR root:
mod > .95 $\Rightarrow$ root made unit root.
For MA root:
mod > .99 $\Rightarrow$ mod. of root made .99,
so as to have invertibility.
With .99 no numerical problems appear.

We use some very useful results (Tiao, Tsay) on superconvengence of unit AR roots.
Example:
Let true model be the AR (3):

$$(1-.5B)\nabla^2 z_t = a_t$$

Because of superconsistency of U.R. estimators:

* If we estimate simply an AR (1) $\rightarrow$ the first U.R. (∇) is likely to be captured.
* If we estimate, again, an AR (1) on the previous residuals, $\rightarrow$ the second U.R. (∇) is likely to be captured.
* Alternatively, if we start by estimating an AR (2), both U.R. (∇^2) are likely to be captured.
* Further increases in p (the order of the AR), or further AR(1) fits to residuals will not point to a U.R.

The previous results extend in a straightforward manner to SEASONAL U.R.
TRAMO uses these results.
First, the model AR (2) AR_s (1) with mean,

$$\left(1+\phi_1 B+\phi_2 B^2\right)\left(1+\phi_s B^s\right)\left(z_t-\mu\right)=a_t \quad ,$$

is estimated.

As already mentioned, if the modulus of an MA root is relatively large, the bias in the estimator of the AR parameter can be large, and the U.R. can be missed.

Therefore, after detecting U.R. with AR fits, TRAMO uses ARMA (1,1) fits to detect U.R. that might not have been captured because of ignoring possibly large MA roots.

Hence, after the pure AR fit, TRAMO fits models of the form ARMA (1,1) $ARMA_s$ (1,1) with mean

$$(1+\phi\, B)(1+\Phi B^s)(z_t-\mu)=(1+\theta B)(1+\Theta B^s)a_t \quad .$$

The residuals of the last estimated model are used for a pre-test to specify a mean or not.
(Unit roots are identified 1 by 1.)

<u>Identification of the stationary ARMA model: ϕ(B) and θ(B)</u>

The program selects the orders (p,q), where $p=dg\{\phi(B)\}$ and $q=dg\{\theta(B)\}$, corresponding to the lowest $BIC_{p,q}$, where

$$BIC_{p,q}=\ln\left(\hat{\sigma}^2_{p,q}\right)+(p+q)\frac{\ln(N-d)}{N-d}\ .$$

It searches for models of the form

$$\phi_{p_r}(B)\phi_{p_s}(B^s)x_t=\theta_{q_r}(B)\theta_{q_s}(B^s)a_t$$

over the range

$0\le p_r, q_r \le 3, \qquad 0 \le p_s, q_s \le 2$ (1 if used with SEATS)

This is done sequentially (for fixed regular polynomials, the seasonal ones are obtained, and viceversa). The search favors parsimony and "balanced" models (similar AR and MA orders).

Remark:
When analyzing series with care, TRAMO may suggest a few models (perhaps 2 or 3) all of which could be reasonably acceptable.
When used with SEATS, looking for, among these models, the one that provides a more satisfactory decomposition (e.g.,a better seasonal adjustment) may provide additional tools for the choice.

3. DECOMPOSITION OF THE SERIES AND SEASONAL AJUSTMENT

Next, the program uses the AMB methodology to estimate unobserved components in x(t). The unobserved components are:

- the trend-cycle, $p(t)$,
- seasonal, $s(t)$,
- transitory, $c(t)$,
- irregular, $u(t)$,

components, as in

$$x(t) = p(t) + c(t) + u(t) + s(t) = n(t) + s(t) \quad , \tag{2.4}$$

where $n(t)$ denotes the SA-series.

Broadly, the trend-cycle captures the spectral peak around the zero frequency, the seasonal component captures the peaks around the seasonal frequencies, the irregular component picks up white-noise variation, and the transitory component captures transitory variation that differs from white noise.
From the ARIMA model for the series, the models for the components are derived.

Typically, for the trend-cycle and seasonal component,

$$\nabla^{D} p(t) = w_p(t) \quad , \qquad D = d + d_s \quad , \tag{2.5}$$

$$S\, s(t) = w_s(t) \quad , \qquad S = 1 + B + \ldots + B^{s-1} \quad , \tag{2.6}$$

where $w_p(t)$ and $w_s(t)$ are stationary ARMA processes.

The transitory component is also a stationary ARMA process.
The irregular component is white noise.
The processes $w_p(t)$, $w_s(t)$, $c(t)$, and $u(t)$ are assumed to be mutually uncorrelated.

Aggregation of the models for p, s, c, and u yields the ARIMA model (2.3) for the series $x(t)$.

The model for the SA-series is obtained through aggregation of the models for $p(t)$, $c(t)$, and $u(t)$. Its basic structure is also of the type (2.5), with p replaced by n.
It is well-known that, in (2.4), there are a variety of ways in which the additive white noise can be assigned to the components. Identification is achieved in SEATS by imposing the "canonical condition", whereby all additive white noise is assigned to the irregular component. In this way, the variance of the later is maximized, and the rest of the components are as stable as possible, given the stochastic features of the series.
The component estimator and forecast are obtained by means of a Wiener-Kolmogorov type filter as the MMSE estimators (under the normality assumption, equal to the conditional expectation) of the signal given the observed series.
The filter is two-sided, centered, symmetric, and convergent, and can be given a simple analytical representation.
Let $x(t)$ follow the model

$$\phi(B)\ x(t) = \theta(B)\ a(t) \ , \ \ a(t) \sim wn(0, V_a) \ , \tag{2.7}$$

where $\phi(B)$ also contains now the unit roots.

Consider the decomposition of x(t) into "signal plus non-signal" as in x (t) = s (t) + n (t).
Let the model for the signal be

$$\phi_s(B)\ s(t) = \theta_s(B)\ a_s(t) \quad , \qquad a_s(t) \sim wn\ (0, V_s) \quad ,$$

where $\phi_s(B)$ will contain the roots of $\phi(B)$ associated with the component s.
Denote by $\phi_n(B)$ the polynomial in B with the roots of $\phi(B)$ that are not in $\phi_s(B)$ (that is, the AR roots of the non-signal).
Then, if $F = B^{-1}$ denotes the forward operator (such that F x (t) = x (t+1),) for a doubly infinite series, the WK filter to estimate the signal is given by

$$\nu_s(B,F) = \frac{V_s}{V_a}\ \frac{\theta_s(B)\ \phi_n(B)}{\theta(B)}\ \frac{\theta_s(F)\ \phi_n(F)}{\theta(F)} \quad , \tag{2.8}$$

or, equivalently, by the ACF of the stationary ARMA model

$$\theta(B)\ z(t) = \left[\theta_s(B)\ \phi_n(B)\right] b(t)\ ,\ b(t) \sim wn\ (0, V_s / V_a)\ .$$

The estimator of the signal is obtained through

$$\hat{s}(t) = \nu_s(B,F)\ x(t) \quad . \tag{2.9}$$

In practice, one deals with a finite series: [x(1), x(2), … , x(T)].
Given that the WK filter converges, for long-enough series, the estimator of the signal for the mid-years of the sample can be considered to be equal to the final estimator (that is, the one that would be obtained with the doubly infinite series).
More generally, given the series [x(1), … , x(T)], the MMSE estimators and forecasts of the components are obtained applying the two-sided WK filter to the series extended at both ends with forecasts and backcasts.
It is possible, however, to obtain the full effect of the doubly infinite filter with just a small number of forecasts and backcasts.
The model-based framework is exploited to provide (SE) of the estimators and forecasts (as well as of the rates of growth).
Being obtained by using forecasts, the component estimators at the end points of the series are preliminary, and will suffer revisions as future data become available.
The model-based framework is also exploited to analyze revisions (size and speed of convergence) and to provide further elements of interest to short-term monitoring.

4. <u>THE APPLICATION: AUTOMATIC MODEL IDENTIFICATION AGGREGATE RESULTS</u>

(a) In-Sample Fit

The example consist of 500 monthly exports and imports series of 15 European Union countries, spanning the period January 1995 – January/February 2005 (121/122 observations). All series were treated with the input parameter RSA = 4, which implies automatic testing for the log/level transformation, for the possible presence of a deterministic mean, and for the possible presence of a TD effect - with the parsimonious specification "working/non-working days" – and of EE. Automatic identification of the (p,d,q) $(p_s,d_s,q_s)_{12}$ orders of the ARIMA model, joint with automatic identification of the 3 types of outliers already described, is also performed. Next, the complete model is estimated, the models and filters for the components are derived, and the components are estimated and forecasted; approximated estimation and forecasting SE are also provided. Execution time of the full set - in a standard portable PC-takes 2-3 min.
The output file containing the aggregate results for the set is displayed next.

SERIES IN FILE : 500
SERIES PROCESSED : 500

SUMMARY RESULTS

TABLE 1 : GENERAL FEATURES

	# of series	%
Levels	36	7.20
Logs	464	92.80
Regular Diff.	450	90.00
Seasonal Diff.	377	75.40
Stationary	22	4.40
Non Stationary	478	95.60
Purely Regular	78	15.60
Nz Too Small for complete AMI	0	0.00
Airline Model (Default)	266	53.20

TABLE 2 : DIFFERENCES

# of series with	D = 0	D = 1	D = 2	Total
BD = 0	22	101	0	123
BD = 1	28	348	1	377
Total	50	449	1	500

TABLE 3 : ARMA PARAMETERS

% of series with	P	Q	BP	BQ	
0	71.4	26.4	92.8	21.2	
1	19.2	70.4	7.2	78.8	
2	7.2	2.2	0.0	0.0	
3	2.2	1.0	0.0	0.0	
Total > 0	28.6	73.6	7.2	78.8	
					Total
Average # of param. per series	0.4	0.8	0.1	0.8	2.0

TABLE 4 : MISSING VALUES AND REGRESSION

Outliers					
	MO	AO	TC	LS	Tot
% of series with	0.2	43.6	29.2	28.0	64.4
average # per series	4.0	0.8	0.4	0.4	1.6
maximum # per series	4	21	4	5	24
minimum # per series	0	0	0	0	0

	Calendar Var.		
	TD	EE	Tot
% of series with	76.2	13.2	77.0

It is seen that close to 93% of the series are modeled in the logs and 4% are found stationary. For 16% of them the model has no seasonal structure. The default Airline model is found appropriate for approximately 53% of the series.
Concerning unit roots, the transformation $\nabla\nabla_{12}$ is chosen for close to 70% of the series; for 20% of them only ∇ is needed, and for 6% only ∇_{12} is needed. A slightly smaller percentage requires no differencing, and for one series the transformation $\nabla^2\nabla_{12}$ seems appropriate. As for the ARMA parameters, the average number per series is 2.0, implying thus highly parsimonious models. The models are dominated by IMA(1,1) regular and seasonal structures; 30% require adding an AR structure, in 1/3 of the cases with $p = 2$ or 3. A few series present $q = 2$ or 3, and a stationary AR seasonal structure is only needed for about 7% of the series.
Slightly more than 1/3 of the series do not need outlier correction, and the average number of outliers per series is 1.6, a relatively small number. The AO type is the most dominant, with the rest evenly distributed between LS and TC outliers.
TD is detected in over 75% of the series; EE is far less significant (roughly, 13%).

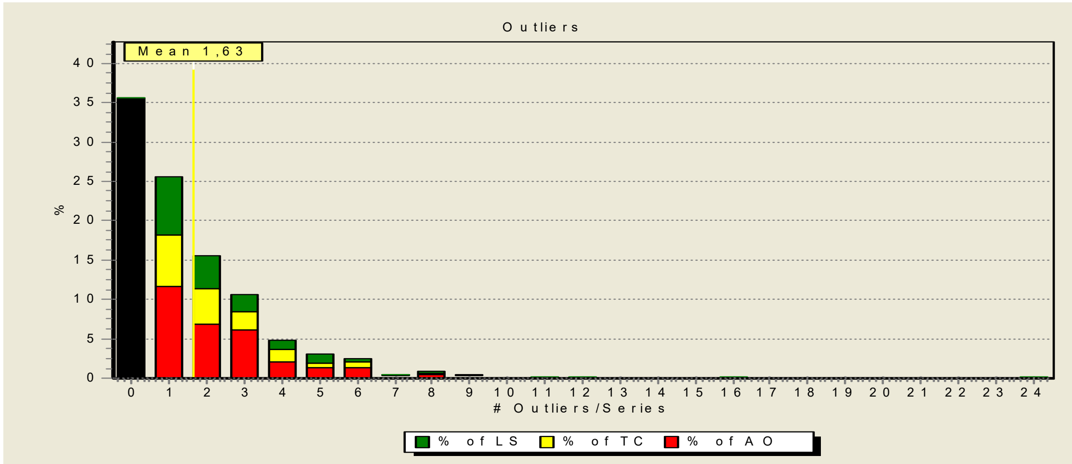
Finally, the results of several (approximate) tests are given. Q is the Ljung-Box test for residual autocorrelation (in our case, $\chi 2$ with approximately 22 df), N is the Behra-Jarque test for Normality of the residuals ($\chi 2$ with 2 df), SK and Kur are t-test for skewness and kurtosis in the residuals, QS is the Pierce test for residual seasonality ($\chi 2$ with 2 df), Q2 is the McLeod and Li test for linearity in the residuals (in our case, $\chi 2$ with, approx. 24 df) and Runs is a t- test for randomness in the signs of the residuals.
The aggregate results pass comfortably all the previous tests at the 99% level. Only the empirical size of Q2 is very slightly below the theoretical size.
Two additional tests for in-sample fit have been added. One is a nonparametric test for the presence of seasonality which is applied to the model residuals. At the 99% significance level, out of the 500 series, only 6 display (borderline) evidence of residual seasonality, clearly in line with the test size. The other test looks at the residuals for the first and second half of the series, and tests for the equality of the two means and variances. Again, at the 99% significance level, 8 series fail the test, slightly above the 5 that could be expected.
In summary, the previous evidence points toward a close to perfect aggregate performance of the AMI results.

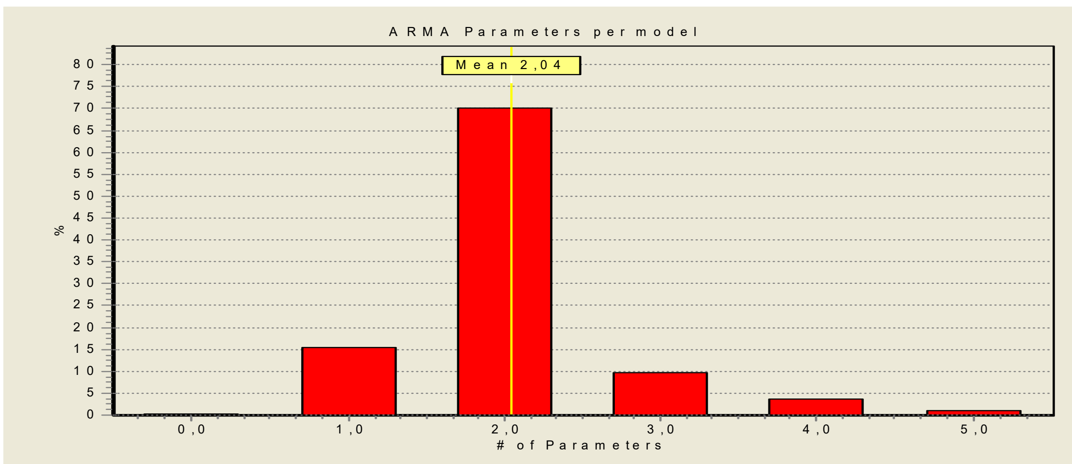
TABLE 5 : SUMMARY STATISTICS

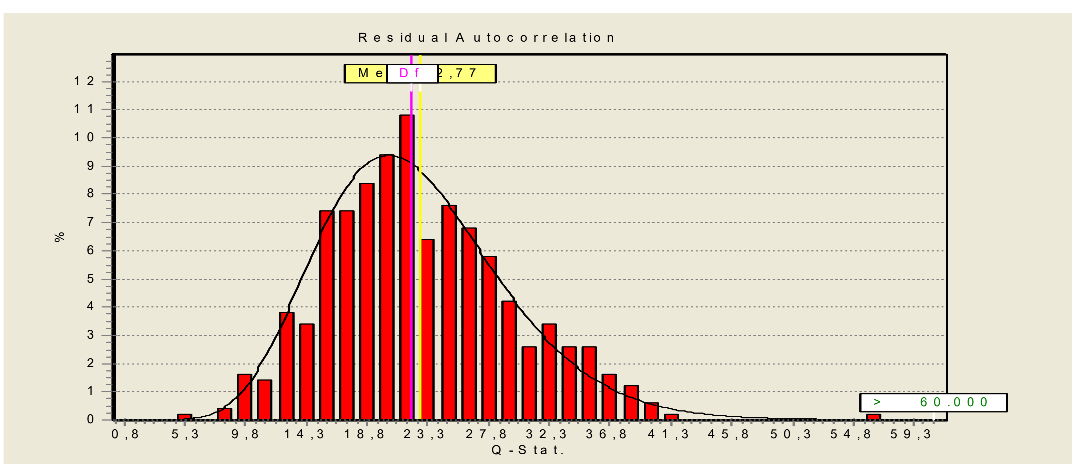
	Mean	Max	Approx 1% CV	Beyond 1% CV	% of series that pass the test (99%)
Length	121.8	122			
# of ARMA param. per serie	2.0	5			
# of outliers per serie	1.6	24			
Q	22.8	56.8	40.29	0.4	99.6
N	1.9	138.00	9.21	1.0	99.0
SK	0.1	5.8	2.58	0.6	99.4
Kur	0.0	10.2	2.58	1.2	98.8
QS			9.21	0.4	99.6
Q2	23.1	64.7	42.98	2.6	97.4
Runs	-0.1	2.4	2.58	0.2	99.8

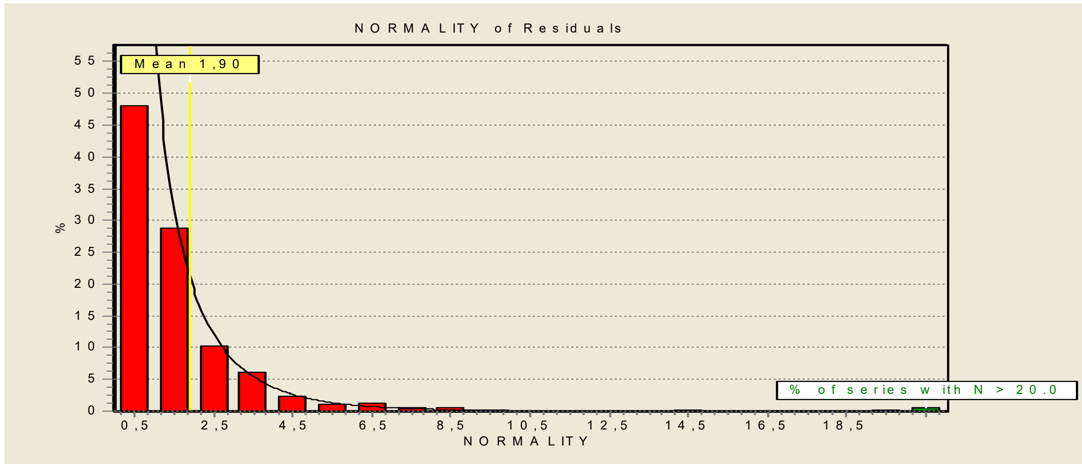
The "modal" model can be described as the Airline model for the logs and no mean, with one or two outliers (one of them an AO), and TD effect.
The model contains 5 or 6 parameters: θ_1 controls the stochastic character of the trend, θ_{12} controls that of the seasonal component, α captures the importance of the TD effect, σ_a reflects the overall forecasting accuracy, and ω_1 , (perhaps ω_2) the effects of the outlier (s).
For the sample size considered (121/122 obs.) this seems a sensible and parsimonious parametrization.

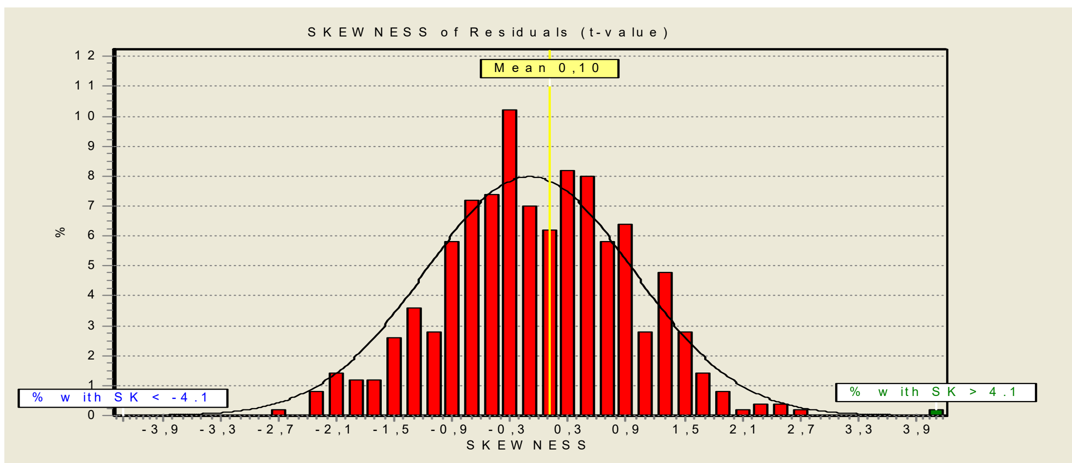
The following figures summarize the previous results. The first two figures display the relative frequency of series according to the number of outliers per series (AO, TC, and LS with different colours), and the number of ARMA parameters per series. The next seven figures compare the histograms and asymptotic distributions for each of the 7 (approximate) tests: Autocorrelation, Normality, Skewness, Kurtosis, non-Linearity, Randomness of the residuals, and residual seasonality. Very few anomalies are detected. Still, a few problematic series can be detected, such as the one with Q= 56.8, the one with QS=9.5, or the one with 24 outliers.

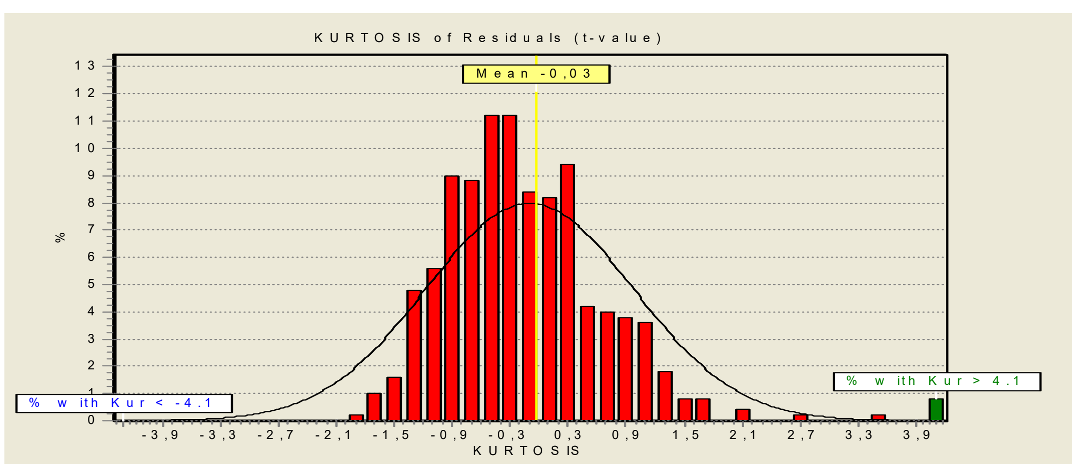

Outliers
Mean 1,63
%
Outliers/Series
% of LS
% of TC
% of AO

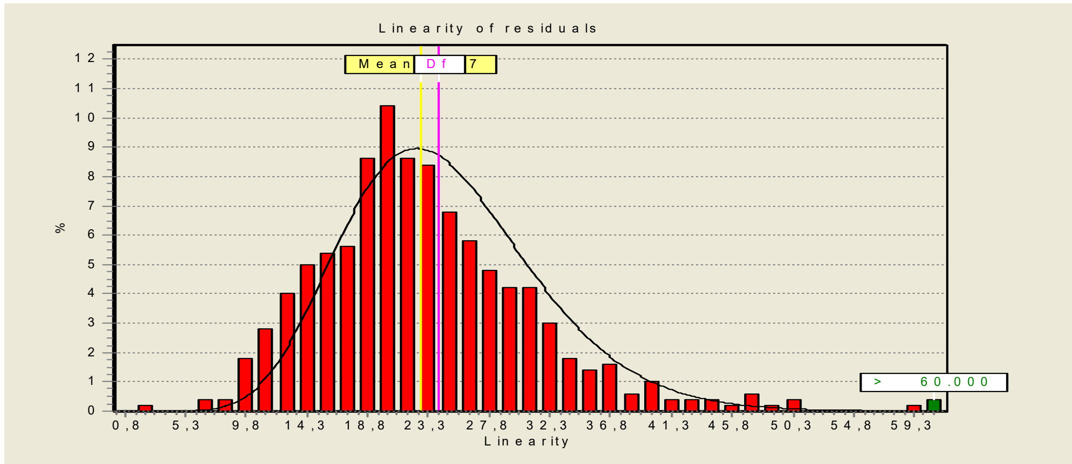

ARMA Parameters per model
Mean 2,04
%
# of Parameters

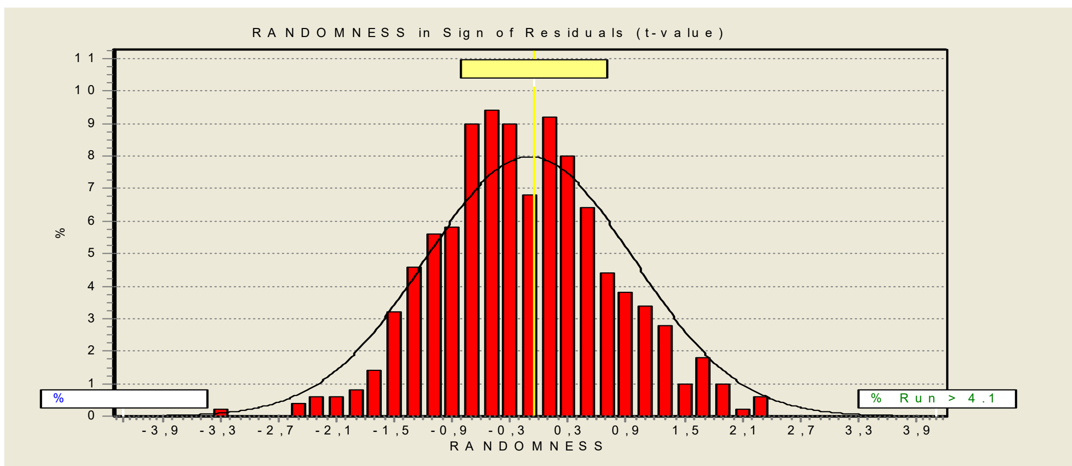

Residual Autocorrelation
%
Q-Stat.
> 60.000

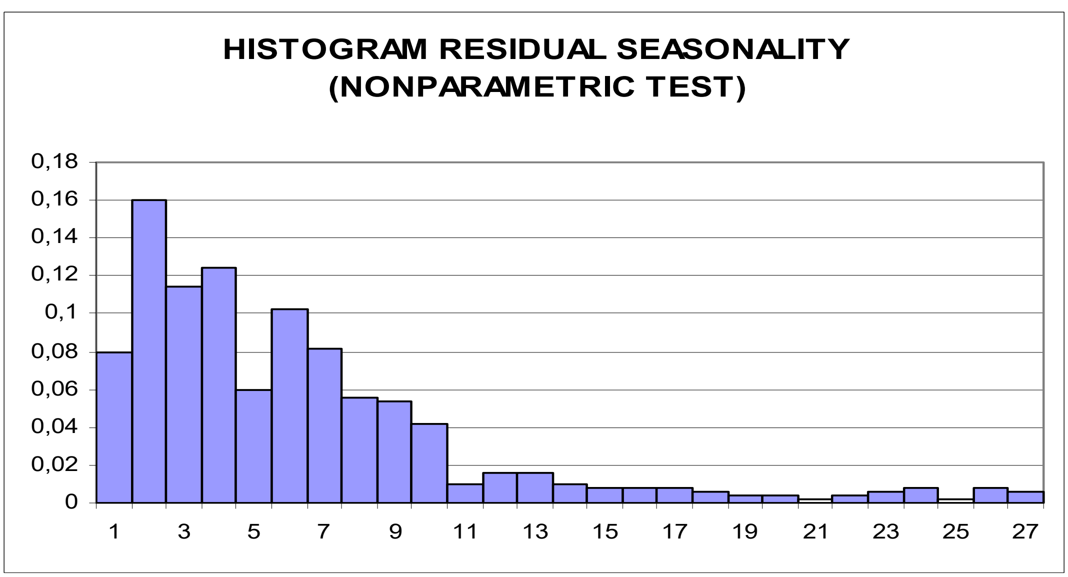

NORMALITY of Residuals
Mean 1,90
% of series with N > 20.0
%
NORMALITY
0,5 2,5 4,5 6,5 8,5 10,5 12,5 14,5 16,5 18,5
0 5 10 15 20 25 30 35 40 45 50 55


SKEWNESS of Residuals (t-value)
Mean 0,10
% with SK < -4.1
% with SK > 4.1
%
SKEWNESS
-3,9 -3,3 -2,7 -2,1 -1,5 -0,9 -0,3 0,3 0,9 1,5 2,1 2,7 3,3 3,9
0 1 2 3 4 5 6 7 8 9 10 11 12


KURTOSIS of Residuals (t-value)
Mean -0,03
% with Kur < -4.1
% with Kur > 4.1
%
KURTOSIS
-3,9 -3,3 -2,7 -2,1 -1,5 -0,9 -0,3 0,3 0,9 1,5 2,1 2,7 3,3 3,9
0 1 2 3 4 5 6 7 8 9 10 11 12 13


Linearity of residuals
Mean
Df
7
> 60.000
%
Linearity


RANDOMNESS in Sign of Residuals (t-value)
%
% Run > 4.1
%
RANDOMNESS


HISTOGRAM RESIDUAL SEASONALITY
(NONPARAMETRIC TEST)

(b) Out-of-Sample Forecasts

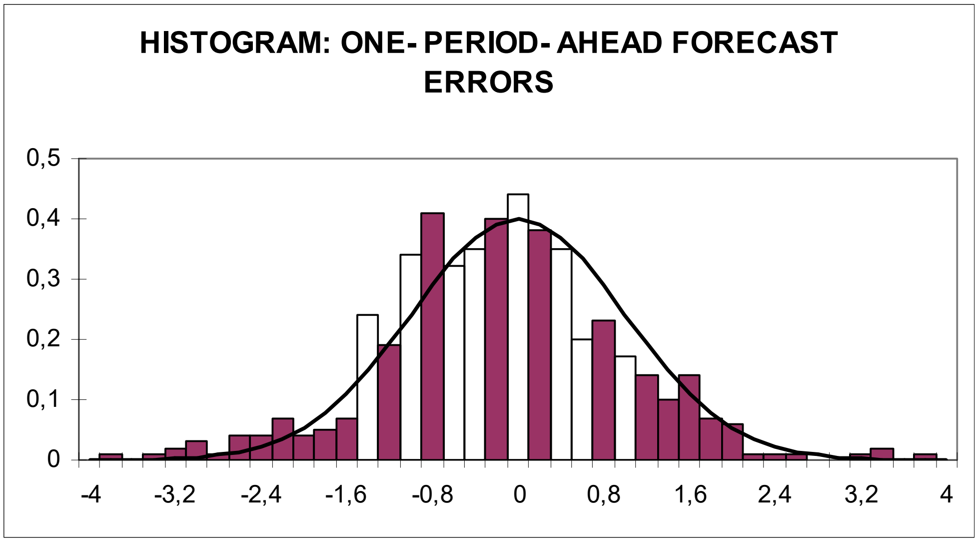
It is well-known that it is easier to get a good in-sample fit than a good out-of-sample forecasting. A fast and simple forecasting check can be done by running TSW with the input TERROR = 1, and setting k1 = 3 and k2 = 4 as thresholds for the t-values of the forecast errors (see Caporello and Maravall, 2003). This implies applying the automatic procedure to all series without considering the last observation. The out-of-sample one-period-ahead forecasts are computed as well as their associated SE. Let the forecast for the last period (T) made at period (T-1) be y(T|T-1) and denote by σ (T|T-1) the associated SE. The standarized forecast error is given by ε (T|T-1) = [y(T) - y(T|T-1)] / σ (T|T-1).
The following tables present the series in the group that have

$$(4 > |\varepsilon| > 3) \text{ and } (|\varepsilon| > 4.)$$

SERIES TITLE	Date	Log(New Value)	Log(Forecast)	Diff.	StdDev	T-Value
44 - E15_0400_1_3	02-2005	11.94	12.58	-0.64	0.17	-3.59
143 - E15_1120_1_2	01-2005	13.03	12.88	0.15	0.04	3.39
144 - E15_1120_1_3	01-2005	12.85	13.34	-0.48	0.14	-3.39
154 - E15_1130_1_3	01-2005	12.89	13.33	-0.44	0.13	-3.21
179 - E15_1415_1_8	02-2005	6.73	6.96	-0.23	0.06	-3.74
182 - E15_1811_1_1	01-2005	12.78	13.01	-0.23	0.07	-3.10
209 - E15_5190_1_8	01-2005	8.26	8.15	0.10	0.03	3.24
248 - E15_5500_1_7	02-2005	11.41	12.27	-0.86	0.28	-3.07
15 - E15_0039_2_4	02-2005	9.51	9.09	0.42	0.13	3.14
95 - E15_1051_2_4	01-2005	9.70	10.33	-0.63	0.20	-3.03
143 - E15_1120_2_2	01-2005	12.50	12.28	0.22	0.06	3.61

Summary Statistics
500 Series were tested.
11 Releases exceeded the critical value t=3
0 Releases exceeded the critical value t=4
0 Series produced a Run-Time EXCEPTION.
489 Series passed the test: |t| < 3.

If the 500 values of ε were sampled from a N(0,1) distribution, the most likely number of ε's in the ranges (3 < $|\varepsilon|$ < 4) and ($|\varepsilon|$ > 4) would be 1 and 0, respectively.
The average number of outliers detected in the in-sample fit (1.6 per series) implies that about 8 outliers could be expected among the 500 ε's.
Of the 500 forecast errors obtained, 12 have fallen in the range
(3 < $|\varepsilon|$ < 4) and none in the range ($|\varepsilon|$ > 4).
Comparing (11 and 0) to the expected values (9 and 0), the out-of-sample forecasting performance is reasonably good.
Still, the previous table indicates that, at least, series 44 and 179 in the imports group, and series 143 in the exports group can be considered problematic.
It is a general result that, despite the fact that both groups of series (imports and exports) are considerably noisy, the behavior of exports tends to be a bit more regular than that of imports.
The histogram of the 500 standardized one – period – ahead forecast error is displayed next. Despite a slight negative bias, the histogram conforms well to the asymptotic N (0,1) distribution.



(c) Summary Results for Individual Series

Identification of additional series that are problematic from a fitting point of view can be done through the summary matrices output by the program, where each series occupies one row of the matrix. An Excel macro (also available at the Bank of Spain web site) reads these matrices and pick up the series that appear to be problematic, according to criteria that can be set by the user. These criteria may include additional considerations. Some important ones are: a) classifying a series as problematic when the number of outliers is above a certain threshold (for example, 4% of the observations); b) checking close to non-invertible models for controlling possible overdifferencing (in particular, values of the moving average parameter close to -1 in IMA(1,1) regular or seasonal structures); c) checking for possible underdetection of TD and EE (TSW is more concerned about spurious detection –more difficult to detect- and hence is slightly biased towards underdetection).

(d) Seasonal Adjustment

Besides the summary files containing the main results of the reg-ARIMA model adequacy (TRAMO part), the program also outputs summary results having to do with seasonal adjustment and the adequacy of the series decomposition (SEATS part). First, it indicates for which series the model obtained in AMI and used for computing the series forecasts has been changed automatically by the program to provide a better decomposition. The most important cause of a change is when the model does not provide an admissible decomposition. The new model used to adjust the series is shown. In our example, the model is changed for 25 series, and 11 of these changes (2% of the series in the set) are due to the absence of an admissible decomposition.

Then, a check is made on the accuracy of the spectral factorization that provides the component's models. No series in the set yields any problem in this respect. Next, a check is made to detect significant differences between the ACF and CCF of the theoretical estimators and the empirical estimates. In the example, 10% of the series show differences at the 5% significance level (work at the USBC indicates that these tests require some more refinement).

Next the standard deviations of the component innovations are presented. They permit to assess, for example, the relative stability of the seasonal component and trend-cycle. Approximate SE of the components estimators, of the revisions the concurrent estimator will suffer, and of some associated rates of growth are also provided. (Naturally, smaller SE are always preferable.) The speed of convergence of the concurrent estimator to the final one is displayed, as well as information concerning the significance of the detected seasonality in the final and preliminary estimators and in the next year of forecasts. Finally, when appropriate, the bias effect on the levels induced by the log transformation is also computed.

Additional important checks are being added to this summary output. Examples are a test for idempotency (seasonal adjustment of the adjusted series should basically reproduce the adjusted series), a check for seasonal overdifferencing, a check for highly stationary and unstable seasonality, and –most importantly- spectral domain tools similar to the ones in X12ARIMA that yield information on the gain and phase effect of preliminary estimators and on residual seasonality or calendar effects. Thus, focusing, for example, on the possible presence of residual seasonality, it can be detected through the QS test (in

TRAMO), the nonparametric test for the presence of seasonality in the residuals (in TRAMO), the tests for significance of seasonality: historical and preliminary estimates, and forecasts (in SEATS), the comparison of the theoretical and empirical variances of the (differenced) seasonally adjusted series and seasonal component (in SEATS), and two different estimators of the spectra of the residuals and (differenced) seasonally adjusted series (in future version of SEATS).

Final Remark:

Once problematic series have been identified, editing "Model ø" in the main window, alternative specifications can be tried until a final one is selected for each of the series. Then, clicking in "Model ++" in the main window, the set of final model can be saved (in a variety of manners) in order to avoid AMI every time a new observation becomes available.

REFERENCES

The following list contains some basic references having to do with TSW, TRAMO, SEATS, and TERROR. They contain, in turn, more detailed references. Additional references and documentation can be found in (www.bde.es→professionals →econometrics software).

CAPORELLO, G., and MARAVALL, A. (2004), "Program TSW: Revised Reference Manual", Working Paper 0408, Servicio de Estudios, Banco de España.

CAPORELLO, G. and MARAVALL, A. (2003), "A Tool for Quality Control of Time Series Data. Program TERROR", Occasional Paper 0301, Servicio de Estudios, Banco de España.

GÓMEZ, V. and MARAVALL, A. (2001a), "Automatic Modeling 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. and MARAVALL, A. (2001b), "Seasonal Adjustment and Signal Extraction in Economic Time Series", Ch.8 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. and MARAVALL, A. (1996), "Programs TRAMO and SEATS. Instructions for the User", (with some updates), Working Paper 9628, Servicio de Estudios, Banco de España.