RELIABILITY OF THE AUTOMATIC IDENTIFICATION OF ARIMA MODELS IN PROGRAM TRAMO

Agustín Maravall

Bank of Spain

Roberto López Pavón and Domingo Pérez Cañete

Indra (*)

(*) external collaboration with the Bank of Spain
Address correspondence to: Agustín Maravall, Research Department, Bank of Spain, Alcalá 48, 28014 Madrid, Spain (maravall@bde.es)

Abstract

In so far that –as Hawking and Mlodinow state– “there can be no model-independent test of reality,” time series analysis applied to large sets of series needs an automatic model identification procedure, and seasonal adjustment should not be an exception. In fact, the so-called “ARIMA model based” seasonal adjustment method (as enforced in programs TRAMO and SEATS) is at present widely used throughout the world by data producers and analysts. The paper analyses the results of the automatic identification of ARIMA models of program TRAMO. Specifically, the question addressed is the following. Given that many ARIMA models are possible, how likely is it that (default) use of TRAMO yields a satisfactory result? Important requirements are proper detection of seasonality, of non-stationarity (i.e., of the proper combination of unit autoregressive roots), and of the stationary ARMA structure, and eventual identification of either the correct model, or a relatively close one that provides zero-mean normally identically independently distributed residuals and good out-of-sample forecasts. A comparison with the default AMI procedure in the present X12-ARIMA and DEMETRA+ programs (based on older versions of TRAMO) is made.

The simulation exercise shows a satisfactory performance of the default automatic TRAMO procedure applied to very large sets of series; certainly it can also provide good benchmark or starting points when a careful manual identification is intended.

1 INTRODUCTION

Seasonality, i.e., the seasonal component of a time series, is never directly observed, nor does it have a generally accepted and precise definition, and these limitations obscure proper treatment and analysis. In the early 80's, a seasonal adjustment based on minimum mean squared error (MMSE) estimation of unobserved components in linear stochastic time series models –namely, ARIMA models- was proposed by Hillmer and Tiao (1982) and Burman (1980). The approach came to be known as the ARIMA-model-based (AMB) seasonal adjustment. The proposal seemed interesting because it would provide the analyst with a precise definition of seasonality (as well as of the other unobserved components) by means of a model consistent with the model identified for the observed series. The approach would further permit model-derived diagnostics and parametric inference. Some extensions of the approach are found in, for example, Pierce (1979), Bell and Hillmer (1984), Maravall (1987), Gómez and Maravall (2001), Bell and Martin (2004), and McElroy (2008). However, application of the approach when many series are to be treated was discarded because it seemed to imply exceedingly heavy computational and time series analyst resources. Besides, many series need some preadjustment before they can be assumed the output of an ARIMA process: perhaps the series requires some transformation (such as the log), non-periodic calendar effects –such as TD- may need removal, and the series may be contaminated by outliers and/or by other special effects. Because not all users need to be time series modeling experts, and because –even if they are- the number of series that need treatment may be too big, an automatic model identification (AMI) procedure is needed. The procedure should address both preadjustment of the series and identification of the ARIMA model.

In the 90's, Gómez and Maravall presented a first version of two linked programs that enforced the AMB approach and contained an AMI option. The first program, TRAMO ("Time series Regression with ARIMA Noise, Missing Observations and Outliers") performed preadjustment and ARIMA model identification. The second program, SEATS ("Signal Extraction in ARIMA Time Series") decomposed the series into unobserved components and, in particular, performed seasonal adjustment (Gómez and Maravall, 1996).

The two programs are widely used throughout the world, most notably at statistical offices, central banks, and agencies involved with analysis and production of economic data; see, for example, European Statistical System (2009) and United Nations (2011). Together with X12, they are part of the new X13-ARIMA-SEATS program (U.S. Census Bureau, 2013), and of the Eurostat-supported program DEMETRA+ (S. Grudkowska, 2013).

Over the years, the empirical performance of TRAMO and SEATS has been discussed, and a large-scale analysis of their (early) AMI performance is contained in Fischer and Planas (2000). (This work had lead to a recommendation for its use in official production; Eurostat 1998). New versions of the programs have just been released, and the new version of TRAMO (to be referred to as TRAMO+) incorporates modifications and improvements in the AMI procedure. In what follows, the performance of this procedure is analyzed in terms of the following questions: if a series has been generated by an ARIMA model, will AMI properly detect presence/absence of seasonality, stationarity or the lack thereof (i.e., unit roots), the ARMA structure (i.e., model orders)? Will the identified model provide normally, identically, independently distributed (n.i.i.d.) residuals? Will the out-of-sample forecast performance be acceptable?

Program TSW+ (the Windows version of TRAMO-SEATS+) has been used, in all cases in an entirely automatic mode.

2 SUMMARY OF THE AUTOMATIC IDENTIFICATION PROCEDURE

2.1 The Regression-ARIMA model

Let the observed time series be $z = (z_{t_1}, z_{t_2}, \ldots, z_{t_m})$ where $1 = t_1 < t_2 < \cdots < t_m = T$. (There may be missing observations and the original observations may have been log transformed.) The Reg-ARIMA model is

$$z_t = y_t' \beta + x_t \tag{1}$$

where y_t is a matrix with n regression variables, and β is the vector with the regression coefficients. The variable x_t follows a (possibly nonstationary) ARIMA model. Hence, in (1), $y_t' \beta$ represents the deterministic component, and x_t the stochastic one.

If B denotes the backward shift operator, such that $B^j z_t = z_{t-j}$, the ARIMA model for x_t is of the type

$$v_t = \delta(B)\, x_t\,, \tag{2}$$

$$\phi(B)\,[v_t - \mu_v] = \theta(B)\, a_t\,, \qquad a_t \sim \text{niid}\,(0, V_a), \tag{3}$$

where v_t is the stationary transformation of x_t, μ_v its mean, $\delta(B)$ contains regular and seasonal differences; $\phi(B)$ is a stationary autoregressive (AR) polynomial in B; $\theta(B)$ is an invertible moving average (MA) polynomial in B. For seasonal series, the polynomials typically have a "multiplicative" structure. Letting s denote the number of observations per year, in TRAMO+, the polynomials in B factorize as

$$\delta(B) = (1 - B)^d \, (1 - B^s)^{d_s} = \nabla^d \, \nabla_s^{d_s}$$

where ∇ and ∇_s are the regular and seasonal differences, and

$$\phi(B) = \phi_p(B) \, \Phi_{p_s}(B^s) = \left(1 + \phi_1 B + \cdots + \phi_p B^p\right) (1 + \phi_S B^s) \tag{4}$$

$$\theta(B) = \theta_q(B) \, \Theta_{q_s}(B^s) = \left(1 + \theta_1 B + \cdots + \theta_q B^q\right) (1 + \theta_S B^s) \tag{5}$$

Stationarity and invertibility imply that all the roots of the polynomials in B in the right-hand-side of (4) and (5) lie outside the unit circle. In what follows, the variable x_t will be asumed centred around its mean and the general expression for the model will be the ARIMA (p, d, q) (p_s, d_s, q_s) $_s$ model:

$$\phi_p(B) \, \Phi_{p_s}(B^s) \, \nabla^d \, \nabla_s^{d_s} \, x_t = \theta_q(B) \, \Theta_{q_s}(B^s) \, a_t \, , \tag{6}$$

where $p, q = 0, 1, 2, 3$; $d = 0, 1, 2$; $d_s, p_s, q_s = 0, 1$.

In what follows, the only regression variables will be the outliers that may have been automatically identified by the program run in a default mode. Three types of possible outliers are considered: additive outlier (AO), i.e., a single spike; transitory change (TC), i.e., a spike that take some time to return to the previous level; and level shift (LS), i.e., a step function. TRAMO+ will pre-test for the log/level transformation and perform automatic ARIMA model identification joint with automatic outlier detection, estimate by exact maximum likelihood the model, interpolate missing values, and forecast the series.

2.2 Automatic Model Identification in the Presence of Outliers

The algorithm iterates between the following two stages.

1. Automatic outlier detection and correction: The procedure is based on Tsay (1986) and Chen and Liu (1993) with some modifications (see Gómez and Maravall, 2001). At each stage, given the ARIMA model, outliers are detected one by one, and eventually jointly estimated.
2. Automatic model identification: TRAMO+ proceeds in two steps: First, it identifies the differencing polynomial $\delta(B)$ that contains the unit roots. Second, it identifies the ARMA model, i.e, $\phi(B)$, $\Phi(B^s)$, $\theta(B)$, and $\Theta(B^s)$. A pre-test for possible presence of seasonality determines the default model, used at the

beginning of AMI and at some intermediate stages (as a benchmark comparison). For seasonal series the default model is the so-called “Airline model” (Box-Jenkins, 1970), given by the equation

$$\nabla \nabla_s \; x_t = (1+\theta_1 B)\,(1+\theta_S B^s)\, a_t \,, \tag{7}$$

i.e., the IMA $(0,1,1)\,(0,1,1)_s$ model. For nonseasonal series the default model is

$$\nabla x_t = (1+\theta B) + \mu \,, \tag{8}$$

i.e., the IMA $(1,1)$ plus mean model.

Identification of the ARIMA model is performed with the series corrected for the outliers detected at that stage. If the model changes, the automatic detection and correction of outliers is performed again from the beginning.

Identification of the Nonstationary polynomial $\boldsymbol{\delta}(\mathbf{B})$

To determine the appropriate differencing of the series, we discard standard unit root tests. First, when MA roots are not negligeable, the standard tests have low power. Second, a run of AMI for a single series may try thousands of models, where the next try depends on previous results. There is, thus, a serious data mining problem: the size of the test is a function of prior rejections and acceptances, and its correct value is not known.

We follow an alternative approach that relies on the superconsistency results of Tiao and Tsay (1983), and Tsay (1984). Sequences of multiplicative AR(1) and ARMA(1,1) are estimated, and instead of a fictitious size, the following value is fixed “a priori”: How large the modulus of an AR root should be in order to accept it as 1? By default, in the sequence of AR(1) and ARMA(1,1) estimations, when the modulus of the AR parameter is above 0.91 and 0.97, respectively, it is made 1. Unit AR roots are identified one by one; for MA roots invertibility is strictly imposed.

Identification of the stationary ARMA model: ϕ(B) and θ(B)

Identification of the stationary part of the model attempts to minimize the Bayesian information criterion given by

$$BIC_{P,Q} = \ln\left(\hat{\sigma}^2_{P,Q}\right) + (P+Q)\,\frac{\ln(N-D)}{N-D}\,.$$

where $P = p + p_s$, $Q = q + q_s$, and $D = d + d_s$. The search is done sequentially: for fixed regular polynomials, the seasonal ones are obtained, and viceversa. A more complete description of the AMI procedure and of the estimation algorithms can be found in Gómez and Maravall (1993, 1994, 2001a); Gómez, Maravall, and Peña (1999); and Maravall and Pérez (2011).

3 PERFORMANCE OF AMI ON SIMULATED SERIES

3.1 Simulation of the series

Monthly series of n.i.i.d.(0,1) innovations $[a_t]$ were simulated in MATLAB, and $(d + d_s)$ arbitrary starting conditions were set (see Bell, 1984). For 50 ARIMA models, 500 series with 120 observations ("short" series) and 500 "long" series with 240 observations were generated. Thus two sets of 25000 series each were obtained. Each set was divided into three subsets as follows:

- The first subset is formed by 8500 series simulated with **Airline-type models**, as in (7). The combinations of MA parameters $(\theta_1,\ \theta_s)$ were $(-0.9, -0.7), (-0.8, -0.4), (-0.7, -0.3), (-0.6, -0.4)$, $(-0.6, 0)$, $(-0.5, -0.95), (-0.5, -0.5)$, $(-0.4, -0.6), (-0.4, 0), (-0.3, -0.7)$, $(0, -0.7), (0, -0.5)$, $(0.3, -0.6), (0.3, 0), (0.4, -0.8)$, and $(0.5, -0.6)$.
- The second set contains 8000 series simulated from the following **non-seasonal** models. **Stationary models:** $x_t = a_t$; $(1 - 0.7\,B)\ x_t = a_t$; $x_t = (1 + 0.6\,B^2)\ a_t$; $(1 - 0.8\,B)\ x_t = (1 - 0.5\,B)\ a_t$; $(1 - B +$ *0.6 B2 xt=at ;1−0.41 B−0.37 B2 xt=1−0.30 B ; 1+0.3 B2−0.5 B3 xt=at* . **Non-stationary models:** $\nabla\, x_t = (1 - 0.7\,B)\ a_t$; $\nabla\, x_t = (1 - 0.3\,B)\ a_t$; $\nabla\, x_t = a_t$; $(1 -$ *0.7 B∇ xt=at ; (1−0.6 B)∇xt=1+0.5 B+0.7 B2 at* ; *(1−0.40 B+0.42 B2)* $\nabla x_t = a_t$; $(1 + 0.7\,B^{12})\,\nabla x_t = a_t$; $\nabla^2\, x_t = (1 - 0.8\,B)\,a_t$; $\nabla^2\, x_t = (1 - 0.31\,B + 0.36\,B^2)\,a_t$.
- The third set is formed by 8500 seasonal series not of the Airline-type; it will be referred to as the "**Other-seasonal models**" set. **Stationary models:** $(1 - 0.6\,B)\,(1 - 0.6\,B^{12})\ x_t = a_t$; $(1 - 0.8\,B^{12})\ x_t = (1 - 0.4\,B^{12})\ a_t$; $(1 - 0.7\,B)\ (1 - 0.85\,B^{12})\,x_t = (1 - 0.3\,B)\ a_t$; $(1 - 0.7\,B^{12})\,\nabla x_t = (1 - 0.4\,B + 0.7\,B^2)\,a_t$. **Non-stationary models:** $\nabla_{12}\, x_t = (1 - 0.5\,B^{12})\ a_t$; $(1 - 1.4\,\mathrm{B} + 0.7\,\mathrm{B}^2)\,\nabla_{12}\,\mathrm{x_t} = (1 - 0.5\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 + 0.4\,B^{12})\,\nabla_{12}\, x_t = (1 - 0.5\,B^{12})\ a_t$; $\nabla\,\nabla_{12}\,\mathrm{x_t} = (1 - 0.23\,\mathrm{B} - 0.19\,\mathrm{B}^2)\,(1 - 0.56\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 - 0.5\,\mathrm{B}^{12})\,\nabla\,\nabla_{12}\,\mathrm{x_t} = (1 - 0.4\,\mathrm{B})\ \mathrm{a_t}$; $(1 - 0.4\,\mathrm{B})\,\nabla\,\nabla_{12}\,\mathrm{x_t} = (1 + 0.4\,\mathrm{B} + 0.4\,\mathrm{B}^2)(1 - 0.4\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 - 0.3\,B)\,\nabla\,\nabla_{12}\,x_t = (1 -$

$0.6\, B^{12})\ a_t;\ (1 + 0.3\,\mathrm{B})\, \nabla\, \nabla_{12}\, \mathrm{x_t} = (1 - 0.6\,\mathrm{B})\, (1 - 0.3\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 + 0.4\, B^{12})\, \nabla\, \nabla_{12}\, x_t = (1 - 0.5\, B)\, (1 - 0.5\, B^{12})\, a_t$;
$(1 - 0.6\,\mathrm{B} + 0.5\,\mathrm{B}^2)\, \nabla\, \nabla_{12}\, \mathrm{x_t} = (1 - 0.8\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 + 0.5\,\mathrm{B} - 0.3\,\mathrm{B}^3)\, \nabla\, \nabla_{12}\, \mathrm{x_t} =$
$(1 - 0.4\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 + 0.1\,\mathrm{B} - 0.17\,\mathrm{B}^2 - 0.34\,\mathrm{B}^3)\, \nabla\, \nabla_{12}\, \mathrm{x_t} = (1 - 0.48\,\mathrm{B}^{12})\ \mathrm{a_t}$; $(1 + 0.4\, B^{12})\, \nabla^2\, \nabla_{12}\, x_t = (1 - 0.4\, B)\ a_t$.

Therefore, 16% of the models are stationary (40% of them seasonal), and 84% are non-stationary (75% of them seasonal). The models' orders cover the following ranges:

$p = 0, 1, 2, 3;\ \ d = 0, 1, 2;\ \ q = 0, 1, 2;\ \ p_s = 0, 1;\ \ d_s = 0, 1;\ \ q_s = 0, 1;$

so that the maximum order of differencing is $\nabla^2\, \nabla_{12}$ and 384 models are possible. Factorizing the AR polynomials, real and complex roots are present, with varying moduli and frequencies. In particular, identification of unit roots implies identification of one of the pairs (d, d_{12}) = (0,0), (1,0), (2,0), (0,1), (1,1), and (2,1).

The complete set contains many models often found in practice. Non-seasonal series are possibly overrepresented, yet it was thought important to detect reliably which series have seasonality and which ones do not. Some models with awkward structures are also included. As a simple example, the model with seasonal orders $(1, 0, 0)_{12}$ and seasonal AR polynomial $(1 + \phi_{12}\, B^{12})$ with $\phi_{12} > 0$ displays spectral holes at seasonal frequencies. Not being associated with seasonality, nor with trend-cycle, the spectral peaks will generate a transitory component. Such an AR structure may appear , for example, when modeling SA series: the spectral holes induced by seasonal adjustment are associated with negative autocorrelation for seasonal lags in the seasonally adjusted series.

3.2 AMI results

TSW+ was applied to the simulated series in automatic mode with no trading-day pre-testing.

3.2.1 Preadjustment

Log-level test

The 50000 simulated series were exponentiated, then the log/level (likelihood ratio) test was applied to the total 100000 series. Table 1 presents the results.

Table 1: Errors in Log/Level test (in % of series)

	Series is in levels		Series is in logs	
	120	240	120	240
Airline model	0.1	0.0	0.2	0.0
Other-seasonal	0.4	0.1	1.1	0.1
Non-seasonal	0.0	0.0	1.6	1.0
Total Average	**0.2**	**0.0**	**1.0**	**0.4**

The test is accurate (averaging all groups, the error percentage is 0.4%), and shows a slight bias that favors levels. It can be seen that most errors occur for models with $d = 2$ (often, appropriate for models with smooth trend-cycle component).

Seasonality detection

Next, the series were pre-tested for possible presence of seasonality. The pre-test is based on four separate checks. One is a χ^2_{11} non-parametric rank test similar to the one in Kendall and Ord (1990), one checks the autocorrelations for seasonal lags (12 and 24) in the line of Pierce (1978), and uses a χ^2_2; one is an F-test for the significance of seasonal dummy variables similar to the one in Lytras, Feldspauch, and Bell (2007), and one is a test for the presence of peaks at seasonal frequencies in the spectrum of the differenced series. The first 3 tests are applied at the 99% critical value. The fourth test combines the results of two spectrum estimates: one, obtained with an AR(30) fit in the spirit of X12-ARIMA (Findley et al., 1998); the second is a non-parametric Tuckey-type estimator, as in Jenkins and Watts (1968), approximated by an F distribution.

The results of the 4 tests have to be combined into a single answer: is there seasonality in the series. The tests are first applied to the original series, and determine the starting model in AMI. Once the series has been corrected for outliers, the tests are applied again to the “linearized” series; these are the results reported in Table 2. The first 4 columns show the percentage of series (in each of the 6 groups) for which the tests have made an error (not detecting seasonality when there is some, or detecting seasonality when there is none). Leaving aside the Airline model case, for which all tests are close to perfect, in all other case the spectral test performs worse. The “overall test” in column 5 combines the results of the previous four tests, assigning weights broadly in accordance with their relative performance: more weight is given to the autocorrelation and F tests, and little weight is given to the spectral one. The overall test assumes seasonality even when the evidence is weak; its role is to orient AMI, but the final decision as to whether seasonality is present in the series is made by the AMI itself, i.e., by the final model obtained by TRAMO+, and the way it is decomposed by SEATS+.

The errors in this final decision are displayed in the last column of Table 2. It is seen that the test implied by AMI outperforms all other tests, including the overall one. On average, for the short series, this final test fails one out of 200 cases; when the series is long, the proportion becomes 1 out of 500, well below the 1% critical value used in the individual tests.

It can be seen that most errors in Table 2 are failures of the test to detect highly stationary seasonality, and that a slight overdetection of seasonality in non-seasonal series is also present.

Table 2: Errors in the detection-of-seasonality-in-series tests (in % of series in group)

		Non-parametric test	Auto-correlation test	Spectral test	F-test	Overall test	Model produced by AMI
Airline model	120	0.0	0.0	0.0	0.1	0.0	0.0
	240	0.0	0.0	0.0	0.1	0.0	0.0
Other-seasonal models	120	2.9	0.2	6.2	2.0	0.1	0.3
	240	1.9	0.0	4.9	1.4	0.0	0.0
Non-seasonal models	120	1.5	1.6	2.5	0.8	2.1	1.3
	240	1.9	1.8	3.1	0.7	2.4	0.7
Total	**120**	**1.5**	**0.6**	**2.9**	**1.0**	**0.7**	**0.5**
	240	**1.2**	**0.6**	**2.6**	**0.7**	**0.8**	**0.2**

Outlier detection

No outliers were added to the simulated series and hence detected outliers can be seen as spurious; the average number detected per series is shown in Table 3.

Table 3: Average number of outliers per series

	120	240
Airline model	0.18	0.11
Other-seasonal	0.16	0.10
Non-seasonal	0.17	0.09
Total	**0.17**	**0.10**

However, because our interest is automatic use of TRAMO+ with parameters set at default values and this use includes automatic outlier detection and correction, outliers may show up in the identified model. These spurious outliers may cause some distortion in AMI, yet this distortion is likely to be minor. (In 9000000 generated observations some are bound to exceed the critical value for outlier detection. For example, approximating the probability of detecting an AO (the most frequently detected type of outlier) in random samples of 120 and 240 observations, or a LS in random walks of equal lengths, the results of Table 3 are below the proportion of outliers that could be expected from the critical values used in the outlier detection tests (close to 3.5) and the fact that 3 types of outliers are tested for each observation. The numbers in Table 3 can be seen as type I errors of the test; they are seen to decrease significantly for the long series, in accordance with the increase in accuracy of AMI.

Identification of the differencing polynomial (unit root detection)

An important part of AMI is identification of the non-stationary roots (i.e., the orders d and d_s of the regular and seasonal differencing). Given that $d = 0, 1, 2$, and $d_s = 0, 1$, six combinations (d, d_s) are possible. Table 4 displays the % of errors made when detecting unit roots in the six groups of series, short and long, separating the errors made in the detection of d from the error in the detection of d_s, and distinguishing between errors due to over –or to under- detection. First, it is seen that the results improve considerably for the long series: doubling the number of observations cuts –on average- in more than half the proportion of errors. In terms of identification of the complete differencing polynomial, the % of errors decrease from 5.6% (series with 120 observations) to 2.6% (series with 240 observations). For all groups, when present, seasonal unit AR roots are captured with less than 1% of errors; spurious seasonal unit roots occur with the same frequency (<1%), except for the case of "non-seasonal model" short series (2.3%).

As for regular unit AR roots, over-estimation of d in "Other seasonal models" and under estimation of d in the "Airline model" group for short series, present the highest proportion of errors (about 4% and 3.5%, respectively). In all other cases the percentage of errors is below 2.5% (short series) and 1.5% (long series).

Table 5 is as Table 4, but the errors are classified according to the orders of the differencing polynomial of the model generating the series. The largest proportion of errors is due to regular over-differencing of stationary series (i.e., the group with $d = d_s = 0$), and of short series with $(d = 0,\ d_s = 1)$, to regular under-differencing of short series with $(d = 2,\ d_s = 0)$, and to seasonal over-differencing of short stationary series. In all other cases the percentage of errors is in the range $(0 - 3.7\%)$.

Altogether, the proportion of successes in identifying correctly the complete differencing polynomial is 94.4% for the series with 120 observations, and 97.4% for those with 240 observations. Most of the errors concern regular differencing in short series, in particular stationary ones, and are concentrated in the series generated with models that have a large and positive AR real root (for example, 0.8, or 0.85). By default, when the estimated root in the model finally obtained with TRAMO's AMI is above 0.9 (regular roots) or 0.95 (seasonal roots), the programs sets it equal to 1 and re-estimates the model. Thus, when a stationary and non-stationary specification seem both possible, AMI tends to favor non-stationarity. This (slight) bias towards non-stationarity is justified by the fact that ARIMA models are basically short-term tools, that IMA(1,1) structures are more stable than ARMA(1,1) ones, and that non-stationary models tend to yield more regular seasonal component and smoother trend-cycle.

Table 4: Errors in unit root detection (in % of series grouped by type of model)

Group	# observ.	# of series in group	Regular unit roots (d)		Seasonal unit roots (d_s)		Complete differencing polynomial (d and/or d_s)
			Under-estimation	Over-estimation	Under-estimation	Over-estimation	
Airline model	120	8500	3.5	0.2	0.4	0	4.0
Other seasonal models	120	8500	1.6	4.2	0	0.6	6.3
Non-seasonal model	120	8000	2.4	2.1	0.8	2.3	6.8
Total	**120**	**25000**	**2.5**	**2.1**	**0.4**	**1.0**	**5.6**
Airline model	240	8500	0.6	0.1	0	0	0.7
Other seasonal models	240	8500	0.2	3.8	0	0.6	4.6
Non-seasonal model	240	8000	0.6	1.4	0	0.8	2.6
Total	**240**	**25000**	**0.4**	**1.7**	**0**	**0.4**	**2.6**

Table 5: Errors in differencing polynomials (in % of series in grouped by orders of differencing)

Simulated model		# of obs. in series	# of series in group	Errors in d		Errors in d_s		Total errors in diff. polynomial
Regular differences d	Seasonal differences d_s			Under-diff.	Over-diff.	Under-diff.	Over-diff.	
0	0	120	4500	---	7.4	---	4.4	10.4
		240	4500	---	6.6	---	1.7	7.8
1	0	120	4000	0.0	1.5	---	1.0	2.5
		240	4000	0.0	0.8	---	0.5	1.2
2	0	120	1000	12.8	---	---	0.6	13.1
		240	1000	1.8	---	---	1.2	3.0
0	1	120	1500	---	4.2	0.7	---	4.9
		240	1500	---	3.2	0.0	---	3.2
1	1	120	13500	3.7	0.5	0.6	---	4.6
		240	13500	0.8	0.4	0.0	---	1.2
2	1	120	500	3.2	---	1.2	---	4.4
		240	500	0.4	---	0.0	---	0.4
Total		**120**	**25000**	**2.5**	**2.1**	**0.4**	**1.0**	**5.6**
		240	**25000**	**0.5**	**1.7**	**0.0**	**0.4**	**2.6**

ARMA model parameters

Concerning the stationary ARMA model given by (3), Table 6 presents the average number of parameters per model. This number is remarkably close to the average number of parameters in the models used to generate the series.

Table 6: Average number of stationary parameters per series

	120	240	In simulation model
Airline model	1.9	1.8	1.7
Other-seasonal	2.4	2.6	2.6
Non-seasonal	1.5	1.5	1.5
Total	**1.93**	**1.97**	**1.94**

Identification of the ARIMA model orders

Next, exact identification of the ARIMA model orders $(p, d, q)(p_s, d_s, q_s)_{12}$ is considered. The first column of Table 7 shows the percentage of series in each group for which identification has produced the correct values for the six order parameters. It should be kept in mind that by default, the AMI of TRAMO+ considers 384 possible combinations of model orders. Some of these models are close and hence difficult to distinguish when the series is relatively short. Simple examples are the ARMA(1,1) and IMA(1,1) when the AR parameter is close to -1; the ARMA(1,1,1) and ARMA(2,1,0) when the roots of the AR(2) are real and not large; or the ARI(1,1) and IMA(1,1) models when the AR parameter is small in modulus.

The average group performance varies between a minimum of 1 out of 2 and a maximum of 6 out of 7 correct identifications of the complete model (short series with non-seasonal models and long series with Airline-type models, respectively). Averaging all groups, automatic default run of TRAMO+ yields the following results: the model is correctly identified 2/3 of the time for the series with 120 observations and 4/5 of the time for series with 240 observations.

Identification of unit roots is considerably accurate (the range of success varies between 93.2% and 99.3%). Therefore, most of the failures in the identification of the full model affect the smaller roots of the ARMA polynomials and, as seen in the next section, the effect of the misspecification is likely to be moderate.

Table 7: Correct Identification of the ARIMA model

	# obs. in series	Complete model orders			Differencing polynomial (d and d_s)		
		TSW+	X13A-S	Demetra+	**TSW+**	X13A-S	Demetra+
Airline-type models	120	**78.0**	68.6	71.6	**96.0**	94.8	96.4
	240	**85.6**	79.9	79.9	**99.3**	98.9	99.3
Other-seasonal models	120	**47.4**	37.3	43.3	**93.7**	86.5	85.9
	240	**71.7**	50.8	66.3	**97.4**	86.8	88.5
Non-seasonal models	120	**69.1**	37.8	54.0	**93.2**	70.4	76.6
	240	**79.4**	36.3	64.5	**95.4**	68.1	80.4
Total	**120**	**64.8**	**47.2**	**56.4**	**94.4**	**84.6**	**86.8**
	240	**78.9**	**54.5**	**70.3**	**97.5**	**86.3**	**89.6**

A remark on the default model

It is a well-known fact that, in practice, the default model (namely, the Airline model of equation (7)) provides a good fit to many economic time series. Table 8 presents the errors in model identification having to do with cases in which an Airline model is identified for a series generated with a different model, and in which the generating model was an Airline model, yet the identified model is not. Table 8 evidences that, contrary to an often expressed belief, there is no over-detection of Airline models; rather the opposite is true.

Table 8: Errors in Airline model detection (in % of series in group)

	120	240
Airline model	21.9	14.4
Other-seasonal	15.0	6.1
Non-seasonal	0.3	0.2
Total	**12.6**	**7.0**

A comparison of results

Up to the year 2011, programs TRAMO and SEATS (and TSW) maintained the basic structure of the Gómez and Maravall (1996) programs, and revisions were kept moderate. In the year 2001 work was started on new versions that corrected, completed, and extended the standard ones. This paper presents the AMI results of the new versions, TRAMO+, SEATS+, and TSW+.

The TRAMO and SEATS programs made available for the routine RegARIMA in X12-ARIMA and X13-ARIMA-SEATS, and for DEMETRA+, were older versions of the new programs that will eventually be updated (at least, partially). Over the last two years, considerable amount of work has been done on the AMI procedure. (Most notably, the old versions had a tendency to over-difference the series, over-detect outliers, and over-adjust for seasonality.)

To get a feeling for the differences in AMI between the older and present versions, X13-ARIMA-SEATS (release version 1.0, build 150) and DEMETRA+ (version 1.0.4.323) were applied to the set of 50000 series and compared to the results of TSW+ (version 750). Table 7 presents the comparison. It should be mentioned that the difference between the three AMIs is not simply due to revisions in the TRAMO+ versions. In both, DEMETRA+ and X12-ARIMA, when adopting TRAMO's AMI, some modifications were made. Still, Table 7 provides a fair idea of the effects of the TRAMO+ revisions on the AMI procedure, and of the relevance of updating older versions.

TSW+ yields the best results. For the Airline-type group, the percentage of correctly identified models increases by an amount between 6 and 10 percent points (p.p.). For the groups Other-Seasonal and Non-seasonal the improvement is considerably larger, most notably when the comparison is made with X12-ARIMA. (This reflects the fact that the TRAMO+ version in the present DEMETRA+ program is more recent.) Notice that the improvement is largest for the group of Non-seasonal models where the % of successful identification is between 15 and 40 p.p. higher for the case of TSW+. Further, contrary to the case of the Airline-model group, for the Other-seasonal and Non-seasonal groups, improvement in unit root detection accounts for an important fraction of the total improvement. In any event, identification of unit roots is always considerably more successful than identification of the stationary orders.

3.2.3 Model diagnostics

Residual diagnostics and out-of-sample performance

TSW+ offers two types of diagnostics. One is aimed at testing the n.i.i.d. assumption on the residuals; the other performs out-of-sample forecast tests. The Normality assumption is checked with the Behra-Jarque Normality test, plus the skewness and kurtosis t-tests; the autocorrelation test is the standard Ljung-Box test (with 24 autocorrelations); independence is further checked with a non-parametric t-test on randomness of the residual sign runs; and the identical distribution assumption is checked with the constant mean and variance test, that tests, first, for equality of means between the first and second half of the series residuals; if accepted, equality of variances is then tested. The out-of-sample checks are, first, a test whereby one-period-ahead forecast errors are sequentially computed for the model estimated for the series with the last 18 observations removed (with the model fixed), and an F-test compares the variance of these errors with the variance of the in-sample residuals. The second test computes the standardized out-of-sample one-period-ahead forecast error for each of the series in the group, and computes the proportion that lie beyond the 1% critical value of a t distribution. (The option TERROR, i.e., “TRAMO for errors,” applied to the full group, directly provides the answer.)

The diagnostic checks for n.i.i.d. residuals are presented in Table 9. Each entry shows the % of series in the group that fail the test at the 1% size. All residual tests perform satisfactorily. The empirical size falls, in all cases, within the range $(0.2 - 1.3\%)$, with the N test at the top of the range, and randomness in signs and lack of autocorrelation lying at the bottom. For 32 of the 36 groups, the empirical sizes are smaller than the theoretical 1% one. Sample variation may cause that a slightly misspecified ARMA model produces slightly better diagnostics, and hence is selected by AMI. Because of this fact, a bias towards smaller empirical sizes in the in-sample tests for the simulated series could be expected. Given that the effect of sample variation should decrease with the length of the series, it seems reasonable that the long series –as seen in Table 9- are closer to the 1% (approximate) theoretical size.

However, the better performance of the misspecified model is unlikely to extend out of sample, so that the bias towards a smaller size induced by the sampling variation should be smaller in out-of-sample tests. In fact, as Table 10 shows the proportion of errors in the out-of-sample forecast tests lies in the interval $(1.1 - 2.3\%)$ for the short series. For the long series the interval becomes $(0.9 - 1.5\%)$, in agreement with the increased accuracy in model identification.

Table 9: Simulated series: model diagnostics; % of series in group that fail the test

		n.i.i.d. assumption on residuals					
	# obs. per series	Constant mean and variance	Autocorrelation	Random signs	Normality	Skewness	Kurtosis
Airline model (8500)	120	0.8	0.2	0.2	0.9	0.8	0.6
	240	0.8	0.3	0.2	1.3	0.8	0.8
Other-seasonal (8500)	120	0.6	0.3	0.3	1.2	0.8	0.8
	240	0.8	0.4	0.2	1.3	0.9	1.0
Non-seasonal (8000)	120	0.7	0.6	0.3	0.7	0.6	0.5
	240	0.8	0.6	0.2	0.7	0.7	0.5
Total (25000)	**120**	**0.7**	**0.4**	**0.3**	**1.0**	**0.7**	**0.6**
	240	**0.8**	**0.4**	**0.2**	**1.1**	**0.8**	**0.7**

Table 10: Out-of-sample forecast tests (% of series that fail the test)

	Out-of-sample forecast	
	F-test (18 final periods)	t-test (1-period-ahead)
Airline model (8500)	1.7	1.5
	1.1	1.2
Other-seasonal (8500)	2.3	1.2
	1.5	0.9
Non-seasonal (8000)	1.9	1.1
	1.1	0.9
Total (25000)	**2.0**	**1.3**
	1.2	**1.0**

Seasonality in residuals

When the models are to be used in seasonal adjustment, it is important to check for whether seasonality may still remain in the model residuals. Table 11 exhibits the % of series in each group that show evidence of seasonality according to the same set of tests as those in Table 2 with the exception of the F-test. For all cases, the frequency of detecting seasonality in the model residuals is –at most– 1 every 500 series; for the overall test, it is –at most– 1 every 1000.

Table 11: Seasonality and Calendar residual effects (% of residual series in group that show evidence)

	# obs. per series	Evidence of seasonality in residuals				Spectral evidence of TD effect in residuals
		Seasonal autocorrel.	Non-parametric test	Spectral evidence	Overall test	
Airline model (8500)	120	0.0	0.1	0.1	0.0	0.1
	240	0.0	0.0	0.1	0.0	0.1
Other-seasonal (8500)	120	0.1	0.1	0.1	0.1	0.1
	240	0.1	0.2	0.1	0.1	0.1
Non-seasonal (8000)	120	0.1	0.2	0.2	0.0	0.1
	240	0.1	0.2	0.2	0.1	0.1
Total (25000)	**120**	**0.1**	**0.1**	**0.2**	**0.0**	**0.1**
	240	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**

4 SUMMARY AND CONCLUSIONS

In so far as time series have different dynamic structures, an appropriate model for each series needs to be identified. Because all analysts need not be time series modeling experts, or because, even if they are, the number of series to be treated is too big -as is often the case in seasonal adjustment- an automatic model identification (AMI) procedure is required.

In this paper some evidence on the performance of the AMI procedure in TRAMO+ is discussed. The question addressed is: does the AMI procedure capture well series that follow ARIMA models? To answer the question, 50000 series that follow 50 different ARIMA models (stationary and non-stationary, seasonal and non-seasonal) were simulated. For each model, 500 series with 120 observations and 500 series with 240 observations were generated.

The series were exponentiated and the resulting 50000 series were added to the original ones; then, the log/level test was applied to the 100000 series. On average, an error is made every 250 series. As for the detection-of-seasonality sequence of tests, the final result yields, on average, one error for every 200 series (short series) and one error for every 500 series (long series). Further, the full model is correctly identified 2 out of 3 cases (short series) and 4 out of 5 cases (long series). The complete differencing polynomial (that allows for regular differencing of order 0, 1, or 2, and seasonal differencing of order 0 or 1) is correctly identified 94.4% of the time (short series) and 97.4% of the time (long series). Model diagnostics that test the n.i.i.d. assumption for the residuals are excellent (the size of the test is always 1%, and the empirical size is below 1.3% in all 36 groups), and the two out-of-sample forecast tests perform satisfactorily (between 1% and 2% of errors). Concerning seasonality, no seasonality and no evidence of trading day effect is left in the residuals (one error every 1000 series in about all groups). In conclusion, the AMI in TRAMO+ does a good job modeling series that follow ARIMA models.

TRAMO+ has been applied in automatic mode with all parameters set at default values. The automatic procedure could be maintained while some parameters are changed to non-default values. For example, for series that fail the Normality test and have no outliers, lowering the critical value for outlier detection is likely to improve Normalilty at the price of some outlier. As another example, if favoring non-stationarity is desired, one may change the default critical value of the unit root parameters. Or, to get better results for the longer series, one may remove some of the early periods. But the purpose of this paper was to show the performance of TRAMO+ when run automatically by default on a large number of series.

REFERENCES

BELL, W.R. (1984), "Signal Extraction for Nonstationary Time Series", *Annals of Statistics* 12, 646-664.

BELL, W.R. and HILLMER, S.C. (1984), "Issues involved with the Seasonal Adjustment of Economic Time Series", *Journal of Business and Economic Statistics* 2, 291-320.

BELL, W.R. and MARTIN, D.E.K. (2004), "Computation of Asymmetric Signal Extraction Filters and Mean Squared Error for ARIMA Component Models", *Journal of Time Series Analysis* 25, 603-625.

BURMAN, J.P. (1980), "Seasonal Adjustment by Signal Extraction", *Journal of the Royal Statistical Society* A, 143, 321-337.

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.

EUROSTAT (2009), *European Statistical System Guidelines on Seasonal Adjustment*, Luxembourg: Office for Official Publications of the European Communities.

EUROSTAT (1998), *Seasonal Adjustment Methods: A Comparison*. Luxembourg: Office for Official Publications of the European Communities, Eurostat, September 1998.

FINDLEY, D.F., MONSELL, B.C., BELL, W.R., OTTO, M.C. and CHEN, B.C. (1998), "New Capabilities and Methods of the X12 ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 12, 127-177.
http://www.census.gov/srd/www/sapaper/jbes98_abs.html

FISCHER, B and PLANAS, C. (2000), "Large Scale Fitting of Regression Models with ARIMA Errors," *Journal of Official Statistics*, 16, 173-184. (See also Eurostat Working Paper no. 9/1998/A/8).

GÓMEZ, V. and MARAVALL, A. (2001), "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. (1996), "Programs TRAMO and SEATS. Instructions for the User", (with some updates), Working Paper 9628, Servicio de Estudios, Banco de España..
http://www.bde.es/webbde/SES/servicio/software/tramo/manualdos.pdf

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. (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. 39-62.

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*.

GRUDKOWSKA, S. (2012), *Demetra+ User Manual*, National Bank of Poland, March 2012.
http://www.cros-portal.eu/sites/default/files/Demetra%2B%20User%20Manual%20March%202012.pdf

HAWKING, S. and MLODINOW, L. (2010), *The Grand Design*, New York: Bantam Books.

HILLMER, S.C. and TIAO, G.C. (1982), "An ARIMA-Model Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association* 77, 63-70.

JENKINS, G.M. and WATTS, D.G. (1968), *Spectral Analysis and its Applications*, San Francisco: Holden Day.

KENDALL, M. and ORD, J.K. (1990), *Time Series*, London: Edward Arnold.

LYTRAS, D.P., FELDPAUSCH, R.M. and BELL, W.R. (2007), "Determining Seasonality: A Comparison of Diagnostics from X-12-ARIMA", U.S. Census Bureau.

http://www.census.gov/ts/papers/ices2007dpl.pdf

MARAVALL, A. (1987), "On Minimum Mean Squared Error Estimation of the Noise in Unobserved Component Models", *Journal of Business and Economic Statistics* 5, 115-120.

MARAVALL, A. and PÉREZ D. (2011), "Applying and Interpreting Model-Based Seasonal Adjustment. The Euro-Area Industrial Production Series", in W.R. Bell, S. H. Holan, and T.S. McElroy (eds.), *Economic Time Series: Modeling and Seasonality*, New York: CRC Press, 2012.

McELROY, T. S. (2008), “Matrix Formulas for Nonstationary Signal Extraction”, *Econometric Theory*.

PIERCE, D.A. (1979), "Signal Extraction Error in Nonstationary Time Series", *Annals of Statistics* 7, 1303-1320.

PIERCE, D.A. (1978), "Seasonal Adjustment when Both Deterministic and Stochastic Seasonality are Present", in Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Dept. of Commerce-Bureau of the Census, 242-269.

TSAY, R.S. (1986), "Time Series Model Specification in the Presence of Outliers", *Journal of the American Statistical Association* 81, 132-141.

TSAY, R.S. (1984), "Regression Models with Time Series Errors", *Journal of the American Statistical Association* 79, 118-124.

TSAY, R.S. and TIAO, G.C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Non-Stationary Arma Models", *Journal of the American Statistical Association* 79, 84-96.

UNITED NATIONS ECONOMIC COMISSION FOR EUROPE (2011), *Practical Guide to Seasonal Adjustment with Demetra+*, New York and Geneva: United Nations.

http://www.unece.org/fileadmin/DAM/stats/publications/Practical_Guide_to_Seasonal_Adjustment_final_web.pdf

U.S. CENSUS BUREAU (2012), *X-13ARIMA-SEATS Reference Manual, version 1.0*, Time Series Research Staff, Center for Statistical Research, Washington D.C.: US Census Bureau, July 2012.

http://www.census.gov/ts/x13as/docX13ASHTML.pdf

U.S. CENSUS BUREAU (2011), *X-12-ARIMA Reference Manual, version 0.3*, Time Series Research Staff, Center for Statistical Research, Washington D.C.: US Census Bureau, February 2011.

http://www.census.gov/ts/x12a/v03/x12adocV03.pdf