UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09-77
Statistics and Econometrics Series 23
December 2009

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

GRAPHICAL IDENTIFICATION OF TAR MODELS

Miguel Ángel Bermejo*, Daniel Peña and Ismael Sánchez***

Abstract

This paper proposes an automatic procedure to identify Threshold Autoregressive models and specify the threshold values. The proposed procedure is based on recursive estimation of arranged autoregression. The main advantage of the proposed procedure over its competitors is that the threshold values are automatically detected. The performance of the proposed procedure is evaluated using simulations and real data.

Keywords: Nonlinear time series; Recursive estimation; Arranged autoregression; TAR models; Nonlinearity test.

* Department of Statistics, Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganés (Madrid), Spain, e-mail: miguelangel.bermejo@uc3m.es and ismael.sanchez@uc3m.es.
** Department of Statistics, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain e-mail: daniel.pena@uc3m.es

Acknowledgements: Work supported by Ministerio de Educación y Ciencia, project SEJ2004-03303 and European SAFEWIND-EU project 213740.

Graphical Identification of TAR models

Miguel Ángel Bermejo, Daniel Peña, and Ismael Sánchez

December 1, 2009

Abstract

This paper proposes an automatic procedure to identify Threshold Autoregressive models and specify the threshold values. The proposed procedure is based on recursive estimation of arranged autoregression. The main advantage of the proposed procedure over its competitors is that the threshold values are automatically detected. The performance of the proposed procedure is evaluated using simulations and real data.

Key words: Nonlinear time series; Recursive estimation; Arranged autoregression; TAR models; Nonlinearity test.

1 Introduction

In time series analysis there is an extensive literature on models that allow changes in the structure of the parameters. In this article, we have focused on models where the change in the structure of the parameters is caused by a lagged value of the time series. These models are called self-exciting Threshold Autoregressive (TAR) models. TAR models were proposed by Tong(1978, 1983) and Tong and Lim(1980). A time series, y_t is a TAR(k;p,d) model if it follows the model

$$y_t = \Phi_0^{(j)} + \sum_{i=1}^{p} \Phi_i^{(j)} X_t + e_t^{(j)}, \qquad r_{j-1} \leq y_{t-d} < r_j, \tag{1}$$

where $j = 1, \ldots, k$. The integer k is the number of regimes, the matrix $X_t = [1_n | y_{t-1} | \ldots | y_{t-p}]$ is a set of explanatory variables, y_{t-d} is the threshold variable and $r_1, \ldots, r_{k-1}$ are the threshold values.

Moreover d is called delay parameter. In each regime, $e_t^{(j)}$ is a sequence of independents and identically distributed (i.i.d.) random variables with zero mean, and finite and constant deviation $\sigma^{(j)}$.

There are in the literature two main approaches to identify a TAR model. The first approach uses Maximum likelihood (ML) scores. Chan(1990) and Chan and Tong(1990) developed the null distribution of the ML test using a Gaussian process and found it to be non-standard. Hansen(1999) used asymptotic and bootstrap distributions to overcome this problem. If the threshold values $r_1, \ldots, r_{k-1}$ where known, ML tests would supply the most powerful test. However, that is not the case in a practical situation. In practice, the threshold value is a nuisance parameter which is not identified under the null hypothesis. This problem has a negative impact in the efficiency of the procedures. To circumvent this problem, ML tests need to assume certain range of possible threshold values. As a results, ML tests need both intensive computational methods and non-standard reference distributions.

The second main approach to detect a TAR model is by means of Portmanteau tests based on the predictive residuals of some arranged autoregressions. If the model is linear, the sequence of predictive residuals of the arranged autoregression have known properties. Petruccelli and Davis (1996) proposed a CUSUM-type test using these predictive residuals that is sensitive to the presence of a TAR structure. Tsay(1989) considered a variant of this idea that is based on a standard F test. The advantage of this second approach is that, as opposed to the ML score test approach, we do not need to know the threshold values to make the test. However, the tests do not provide any information about the threshold values. Those values are eventually needed to estimate the TAR model. Tsay (1989) proposed some approximate graphical methods using a scatterplot to detect manually the threshold values. However, some more accurate procedures to estimate the threshold values are needed.

We have used the idea of arranged autoregression to develop a graphical procedure based on the recursive and time-varying estimation of the parameters. The proposed procedure will allow us to detect TAR models and also to estimate the threshold values. We show that the proposed porcedure has a superior identification performance than previous proposals.

The article is organized as follows. In section 2 we introduce the arranged autoregression and we discuss when it is possible in time series. Section 3 introduces notation and discusses the recursive estimation method. Section 4 gives the proposed graphical procedure, which is called Arranged Re-

cursive Least Squares (ARLS). Moreover we illustrate the advantage of the ARLS tool. In section 5 is detailed an automatic procedure. Finally, section 6 applies the automatic-ARLS to real data.

2 Arranged Autoregression

An AR(p) regression can be written as $y_t = X_t'\phi + a_t$ where $X_t = (1, y_{t-1}, ..., y_{t-p})'$ and $t = 1, 2, ..., n$. Following the notation in Tsay (1989), we refer to (y_t, X_t') as a case. We denote then an arranged autoregression as an autoregression with the cases rearranged based on a particular criteria. It is interesting to see that by rearranging cases, we still maintain the temporal structure of the series within the cases. Consequently, these arranged autoregressions keep the property of weakly exchangeability, in the sense that the vector of error terms a_t of any rearrangement still maintains its covariance matrix unaltered (Wedlin, 1998).

Let us define S as the set of all possible orders of the time index $t = 1, \ldots, n$, and s_i as the ith element of a random element of S. Let us denote as π_i as the ith element of the element of S corresponding to arranging the cases in ascending order of the threshold variable y_{t-d}. That is, π_i is the time index of the ith smallest element of $(y_h, ..., y_{n-d})$, where $h = \max(1, p+1-d)$. To ilustrate the arranged autoregressions we show a simple example. Let y_t be an AR(1) with time-varying parameter

$$y_t = \phi_{0t} + \phi_{1t} y_{t-1} + a_t. \tag{2}$$

Then, if we sort (y_t, y_{t-1}) using y_{t-1} as the threshold variable, we obtain the arranged autoregression

$$\begin{pmatrix} y_{\pi_1} \\ y_{\pi_2} \\ \vdots \\ y_{\pi_n} \end{pmatrix} = \begin{pmatrix} \phi_{0\pi_1} \\ \phi_{0\pi_2} \\ \vdots \\ \phi_{0\pi_n} \end{pmatrix} + \begin{pmatrix} \phi_{1\pi_1} y_{\pi_1 - 1} \\ \phi_{1\pi_2} y_{\pi_2 - 1} \\ \vdots \\ \phi_{1\pi_n} y_{\pi_n - 1} \end{pmatrix} + \begin{pmatrix} a_{\pi_1} \\ a_{\pi_2} \\ \vdots \\ a_{\pi_n} \end{pmatrix}. \tag{3}$$

A TAR(2;1,1) model is just a particular case of this example, where the time-varying parameter has two values. It is important to note that in a TAR(2;1,1) the sequence of parameters in (3) has a change point at the threshold value r. We will use this property to estimate the parameters in (3) using some time-adaptive procedure such that we can easily see a change in the estimated parameters at $t = r$. If the true model is linear, then the sequence of recursive estimates of (3) will have the same

properties as the time-adaptive estimation of an arranged autoregression using any random element from S. We need then to use a suitable time-adaptive estimation porcedure.

3 Recursive methods for the estimation of time-varying parameters

3.1 Weighted least squares

We define the arranged time series y_{s_t} as a time-varying rearranged AR(p) process with time-varying parameters

$$y_{s_t} = X'_{s_t}\phi_{s_t} + a_{s_t}; t = 1, 2, ..., n; \tag{4}$$

where, for any ordering s, belonging to the set S, a_{s_t} is a sequence of i.i.d. random variables such that $E[a_{s_t}] = 0$ and $E(a_{s_t}^2) = \sigma^2 < \infty$. The vector $X_{s_t} = (1, y_{s_{t-1}}, ..., y_{s_{t-p}})'$ is a set of explanatory variables that can be either deterministic or stochastic. The vector $\phi_{s_t} = (\phi_{0s_t}, \phi_{1s_t}, \ldots, \phi_{ps_t})'$ is the set of time-varying parameters that need to be estimated.

The Weighted Least Squares (WLS) estimator $\hat{\phi}_{s_t}$ is the solution of $\hat{\phi}_{s_t} = \arg\min_{\phi} C_{s_t}(\phi)$, where

$$C_{s_t}(\phi) = \sum_{j=1}^{t} \kappa(t, j)(y_{s_t} - X'_{s_t}\phi)^2, \tag{5}$$

where $\kappa(t, j)$ is the so-called forgetting profile. In this article we will use forgetting profiles of the type

$$\kappa(t, j) = \prod_{i=j+1}^{t} \lambda_t, \ j < t, \tag{6}$$

where $\kappa(t, t) = 1$ and λ_t is called the forgetting factor, and holds $0 \leq \lambda_t \leq 1$. The forgetting factor can either be constant, $\lambda_t = \lambda$, or time-varying. The forgetting factor causes progressively a reduction in the importance of old data in the estimation. For this reason, the estimation is time-adaptive. The WLS estimator of (4) is

$$\hat{\phi}_{s_t} = \left(\mathbf{X}'_{s_t}\Lambda_t\mathbf{X}_{s_t}\right)^{-1}\mathbf{X}'_{s_t}\Lambda_t Y_{s_t}, \tag{7}$$

where $\mathbf{X}_{s_t}$ is the matrix $(X_{s_h}, X_{s_{h+1}}, ..., X_{s_t})'$, Λ_t is a diagonal matrix with the forgetting factors $\lambda_h, \lambda_{h+1}, ..., \lambda_t$ in the diagonal, and $Y_{s_t} = (y_{s_h}, ..., y_{s_t})'$. This estimator can be calculated recursively by means of (see, for instance, Ljung and Söderstrom (1983))

$$\hat{\phi}_{s_t} = \hat{\phi}_{s_{t-1}} + M_{s_t}^{-1}X_{s_t}\hat{a}_{s_t}, \tag{8}$$

where $\hat{a}_{s_t} = y_{s_t} - X'_{s_t}\hat{\phi}_{s_{t-1}}$ is the one-step-ahead prediction error and $M_{s_t} = \left(\mathbf{X}'_{s_t}\Lambda_t\mathbf{X}_{s_t}\right)$. The gain matrix $M_{s_t}^{-1}$ can also be calculated recursively as

$$M_{s_t}^{-1} = \frac{1}{\lambda_t}\left(M_{s_{t-1}}^{-1} - \frac{M_{s_{t-1}}^{-1}X_{s_t}X'_{s_t}M_{s_{t-1}}^{-1}}{\lambda_t + X'_{s_t}M_{s_{t-1}}^{-1}X_{s_t}}\right). \tag{9}$$

3.2 Properties of the RLS-estimator with forgetting

The properties of the RLS estimates with a variable forgetting factor are complex. The distribution of the parameter estimators for a general time-varying regression model is unknown. In this article we would need the properties of RLS estimates under the assumption of a time-invariant AR process. We will use those properties just to establish a benchmark to compare the estimates of an arranged autoregression like (3), that clearly is time-varying, with a shift in the parameters caused by the ordering of the variables using the threshold variable y_{t-d}.

In the case of a time-invariant AR process, with no forgetting, i.e., with $\lambda_t = 1$, the MSE of the OLS estimator is (Fuller and Hasza, 1985; Kunitomo and Yamamoto, 1985)

$$MSE[\hat{\phi}_{\text{OLS}}] = E\left[\left(\hat{\phi}-\phi\right)\left(\hat{\phi}-\phi\right)'\right] = \frac{\sigma^2}{n}\Gamma^{-1} + O(n^{-3/2}), \tag{10}$$

where $\Gamma = E(X_tX'_t)$, and that can be estimated by $\hat{\Gamma} = n^{-1}\left(\mathbf{X}'_t\mathbf{X}_t\right)^{-1}$. The use of a forgetting factor can be interpreted as a shrinkage of the sample size. In OLS each data has the same contribution in the estimation. However, in the estimator (8), the equivalent or efective sample size is lower than n. If we use, for simplicity, a constant forgetting factor, the equivalent sample size is $n_{eq} = 1 + \lambda + \cdots + \lambda^{n-1}$. If $n \to \infty$ the asymptotic equivalent sample size is usually denoted as asymptotic memory length and is easily computed as

$$N_0 = \frac{1}{1-\lambda}. \tag{11}$$

Consequently, the MSE of the RLS estimator is larger as λ is smaller, since the forgetting factor avoids that the estimator converges. The asymptotic MSE for the RLS estimator with forgetting factor can be written approximately as

$$MSE[\hat{\phi}_{\text{RLS}}] = \sigma^2 E[\left(\mathbf{X}'_t\Lambda_t\mathbf{X}_t\right)^{-1}] + O\left[(1-\lambda)^{-3/2}\right], \tag{12}$$

and if λ is close to 1, it can be approximated as

$$MSE[\hat{\phi}_{\text{RLS}}] = \sigma^2 E[\left(X'_t\Lambda_tX_t\right)^{-1}], \tag{13}$$

and is estimated with

$$\widehat{MSE}[\hat{\phi}_{\mathrm{RLS}}] = \hat{\sigma}_t^2 \left(\mathbf{X}_t' \Lambda_t \mathbf{X}_t\right)^{-1}, \tag{14}$$

with $\hat{\sigma}_t^2$ an estimate of σ^2 like, for instance, the recursive estimator

$$\hat{\sigma}_t^2 = \hat{\sigma}_{t-1}^2 + \frac{1}{t-p}\left(\hat{a}_t^2 - \hat{\sigma}_{t-1}^2\right). \tag{15}$$

3.3 Adaptive forgetting factors

The forgetting factor will control the influence of the old observations in the estimation. To ilustrate its importance we can rewrite the expression (5) as

$$\begin{aligned} C_{s_t}(\phi) &= \left(y_{s_t} - X_{s_t}'\phi\right)^2 + \lambda_t C_{s_{t-1}}(\phi) \\ &= \left(y_{s_t} - X_{s_t}'\phi\right)^2 + \lambda_t \left(y_{s_{t-1}} - X_{s_{t-1}}'\phi\right)^2 \\ &\quad + \lambda_t\lambda_{t-1}\left(y_{s_{t-2}} - X_{s_{t-2}}'\phi\right)^2 \\ &\quad + \cdots + \lambda_t\lambda_{t-1}\cdots\lambda_2\left(y_1 - X_1'\phi\right)^2. \end{aligned} \tag{16}$$

It is easily seen that the influence of the past is weighted down exponentially. In this way a λ_t far away from 1 causes a larger influence of new observations in the estimation. Consequently, changes in the estimation are quickly found. This higher speed of adaptation, however, increases variability. It can be seen in (9) that the gain matrix , that is a measure of the dispersion of the estimation, grows up as λ_t decreases. For this reason, a right election of the forgetting factor is a key issue for a good adaptive estimation. Several adaptive forgetting factors have been proposed in previous literature. Some of them are:

- Fortescue *et al.*(1981). This proposal is related to the prediction error. It is defined by

$$\lambda_t^{\mathrm{pre}} = 1 - \alpha \frac{\hat{a}_{s_t}^2}{1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}}, \tag{17}$$

  where α is a user-define parameter. This parameter is a problem for the implementation of this forgetting factor, since there is no fixed rule for selecting it.

- Landau *et al.*(1998). This proposal is related to the leverage of the new observations. It is defined by

$$\lambda_t^{\mathrm{lev}} = 1 - \frac{X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}}{1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}}. \tag{18}$$

- Sánchez (2006). This proposal is based on Cook's distance. It is defined by

$$\lambda_t = \lambda_{min} + (1 - \lambda_{min}) P\left(\chi^2_m > mD_t\right), \tag{19}$$

where λ_{min} is a lower bound of the forgetting factor specified by the user, m is the number of parameters in (4) and D_t is a time-varying version of the Cook's distance calculated by

$$D_t = \frac{X'_{s_t} M^{-1}_{s_{t-1}} X_{s_t} \hat{a}^2_{s_t}}{m\hat{\sigma}^2_{s_{t-1}} \left(1 + X'_{s_t} M^{-1}_{s_{t-1}} X_{s_t}\right)}, \tag{20}$$

where $\hat{\sigma}^2_{s_{t-1}}$ is an estimate of σ^2. Sánchez (2006) shows that this forgetting factor combines the advantages of (17) and (18).

In this article, we have used the forgetting factor proposed by Sánchez(2006). This election is justified in next sections.

4 Arranged Recursive Least Squares applied to TAR models

For simplicity of notation, in this section we assume a TAR(2;1,1) model. Then, we can rewrite the expression (1) as

$$y_t = (\phi + \delta I_{(y_{t-d} > r)}) y_{t-1} + a_t, \tag{21}$$

where $\delta \geq 0$. We have developed a graphical tool called Arranged Recursive Least Squares (ARLS). That tool will allow us to detect changes in the structure of the parameters caused by other variable z. In TAR models, the variable that cause the changes is the threshold variable y_{t-d}.

The main idea of the tool is to fit the time-varying arranged AR(p) showed in (4) using the recursive estimation method described in (8). Moreover, we have developed the necessary techniques to detect significant changes in the evolution of the estimation.

4.1 Threshold detection

The delay parameter d is unknown, but it is not a problem in our tool, because we can try different orders from different lags until that we find the true delay, if it exists. Then, assuming that delay parameter d is known, the difficulty will be to identify the possible existence of different regimes and

its respective threshold values r_k. We show how can detect those values. First we need to show a result.

Let $y_t^{(1)}$ and $y_t^{(2)}$ be AR(p) processes with parameter vectors $\Phi^{(1)}$ and $\Phi^{(2)}$, respectively. Define y_t as a time series composed of $y_t^{(1)}$ and $y_t^{(2)}$, that is, $y_t = [y_t^{\prime(1)}, y_t^{\prime(2)}]'$. In the same manner we can define $X_t = [X_t^{\prime(1)}, X_t^{\prime(2)}]'$. Then, if we fit an AR(p) model to y_t using OLS, we will obtain,

$$
\begin{aligned}
\hat{\Phi} &= (X_t' X_t)^{-1} X_t' y_t \\
&= \left(\begin{bmatrix} X_t^{\prime(1)} & X_t^{\prime(2)} \end{bmatrix} \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix} \right)^{-1} \begin{bmatrix} X_t^{\prime(1)} & X_t^{\prime(2)} \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} \\
&= \left(X_t^{\prime(1)} X_t^{(1)} + X_t^{\prime(2)} X_t^{(2)} \right)^{-1} \left(X_t^{\prime(1)} y_t^{(1)} + X_t^{\prime(2)} y_t^{(2)} \right) \\
&= M_t^{(2)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} \left(X_t^{\prime(1)} y_t^{(1)} + X_t^{\prime(2)} y_t^{(2)} \right) \\
&= M_t^{(2)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} X_t^{\prime(1)} y_t^{(1)} + M_t^{(2)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} X_t^{\prime(2)} y_t^{(2)} \\
&= M_t^{(2)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} X_t^{\prime(1)} y_t^{(1)} + M_t^{(1)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(2)} X_t^{\prime(2)} y_t^{(2)} \\
&= M_t^{(2)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} \hat{\Phi}^{(1)} + M_t^{(1)} \left(M_t^{(1)} + M_t^{(2)} \right)^{-1} \hat{\Phi}^{(2)} \\
&= \hat{\alpha} \hat{\Phi}^{(1)} + (1 - \hat{\alpha}) \hat{\Phi}^{(2)}.
\end{aligned}
$$

We check that, for large samples $\hat{\Phi}$ is a weighted average between $\hat{\Phi}^{(1)}$ and $\hat{\Phi}^{(2)}$, which makes sure that will always have an intermediate value.

Let y_t be a TAR process defined by (21). Let y_{π_t} be the arranged time series according to y_{t-d}. Now, we fit a time-varying arranged AR(1) using RLS method. To use an adaptive forgetting factor implies that RLS method does not have a general result of convergence. But really we do not need a fast convergence in the estimation, we only need to detect change in the trend of the estimation. That is, we only need that the estimation is not blow up. We can guarantee that using a lower bound in the forgetting factor.

If we initialize the recursive estimation (8) using OLS method on the full sample, we can make sure, as we showed previously in expression (22), that the starting estimation $\hat{\phi}_{\pi_0}$ will have a value between the real value of parameters ϕ and $\phi + \delta$. So, if

$$
\boldsymbol{y_{t-d} \leq r \Rightarrow y_{\pi_t} = \phi X_{\pi_t} + e_{\pi_t}}, \tag{22}
$$

in this way,

$$E[||\hat{\boldsymbol{\phi}}_{\boldsymbol{\pi_t}} - \phi||^2] < \epsilon,\ if\ t \to \infty. \tag{23}$$

For this reason, the recursive estimation will tend to go away from $\hat{\phi}_0$. Now, when

$$\boldsymbol{y_{t-d}} > r \Rightarrow \boldsymbol{y_{\pi_t}} = (\boldsymbol{\phi} + \boldsymbol{\delta})\boldsymbol{X_{\pi_t}} + \boldsymbol{e_{\pi_t}}, \tag{24}$$

in this way,

$$E[||\hat{\boldsymbol{\phi}}_{\boldsymbol{\pi_t}} - \hat{\phi}_0||^2] < \epsilon,\ if\ t \to \infty, \tag{25}$$

that is, the trend of the estimation will change in r to approach to $\hat{\phi}_0$. So, if we fit an arranged AR(1) using the RLS algorithm and initializing the estimation by means of OLS with all data, the observation in which the estimation changes its trend will be the threshold value r.

4.2 Confidence intervals of $\phi_{s(t)}$

The question is, how can be sure that the change in the trend is significant? The goal is to check that the realized ordered is different from any other one. To that end we use the confidence interval of any possible ordered that can be realized. In this way, we can calculate confidence intervals inside which must be any possible orderer, if that do not cause a structural change. Then, the confidence interval of ϕ_{s_t}, assuming normality, is

$$\hat{\phi}_{s_t} \pm Z_{1-\alpha}\sqrt{\widehat{MSE}[\hat{\phi}_{RLS}]}, \tag{26}$$

The problem is that to calculate (26) exactly is impossible due to the number of possible orderers is $n!$. Therefore, we will only be able to calculate an approximation of (26) using a finite number of random orderers. Even so calculate the approximation will be a slow calculation. Therefore, we propose to use the asymptotic confidence interval of ϕ_{s_t}. It is easily seen that,

$$\lim_{t\to\infty} M_{s_{it}} - M_t = 0,\ \forall i = 1, \ldots, n!. \tag{27}$$

Consequently, as we can see in expression (14) the estimation of the $\widehat{MSE}[\hat{\phi}_{RLS}]$ for each orderer is asymptotically equal to the $\widehat{MSE}[\hat{\phi}_{RLS}]$ for the unarranged variable. Then, the asymptotic confidence interval of $\hat{\phi}_{s_t}$ can be calculated by means of

$$\hat{\phi}_t \pm Z_{1-\alpha}\sqrt{\hat{\sigma}_t^2 M_t^{-1}}. \tag{28}$$

4.3 Choice of forgetting factor

In ARLS procedure is very important to have a good adaptive estimation. We have seen previously that the choice of the forgetting factor is a key factor when we are using adaptive estimation. For this reason, it is need to choose an adequate forgetting factor.

The first step of ARLS procedure is to arrange y_t according to y_{t-d}, assuming that the delay d is the threshold variable. The second step is to initialize the adaptive estimation $\hat{\phi}_{\pi_0}$ using OLS on full sample. Then, we start the recursive estimation. It is easily seen that, while $y_{t-d} \leq r$ the starting value is not adequate for the new observations. For this reason, we need that the adaptive estimation change toward the right value of the parameters. The adaptive estimation will change quicker if the forgetting factor is adequate.

The forgetting factor proposed by Landau(1998) is related to the leverage measurement. That is a measurement of the influence of the new observations in the previous ones. But, in our problem, the new observations and the previous ones are similar, because the arranged data come from the same regime, while $y_{t-d} \leq r$. For this reason, the leverage measurement does not notice changes until that $y_{t-d} > r$.

We need that the adaptive estimation changes quickly when the new observations have a great influence with respect to the last estimation. A traditional way to calculate that influence when we are using Least Squares estimation is the Cook's distance. For this reason, we think that the forgetting factor proposed by Sánchez(2006) is adequate for this task. In this way, if the new observation has a high value of Cook's distance, the forgetting factor will be a value far away from 1, and the adaptive estimation changes quicker.

4.4 Finite sample performance of the asymptotic intervals

We show an example of the graphical tool. To that end we use simulated data from,

$$y_t = (-0.6 + \delta I_{(y_{t-1}>1)})y_{t-1} + e_t, \tag{29}$$

where δ will have value 0 (AR(1) model) or 0.8 (TAR(2;1,1) model). The sample size will be $n = 150$ and $n = 500$ for each δ. The sequence e_t will be a $WN(0, 3)$ process.

Given the model (29), if the adaptive estimation is out of the confidence limits, we will detect a TAR model. Moreover, the observation most far away from the limits will be the threshold value 1. On the other hand, if the adaptive estimation is inside of the limits, we do not detect a TAR model. In figure 1 we display abovementioned simulations. When $\delta = 0$, that is, when the procces is linear, the estimation is always inside the limits. However, when $\delta = 0.8$ the estimation get out of the limits. Moreover, the observation most far away from the limits is approximately 1, the true value of the threshold.

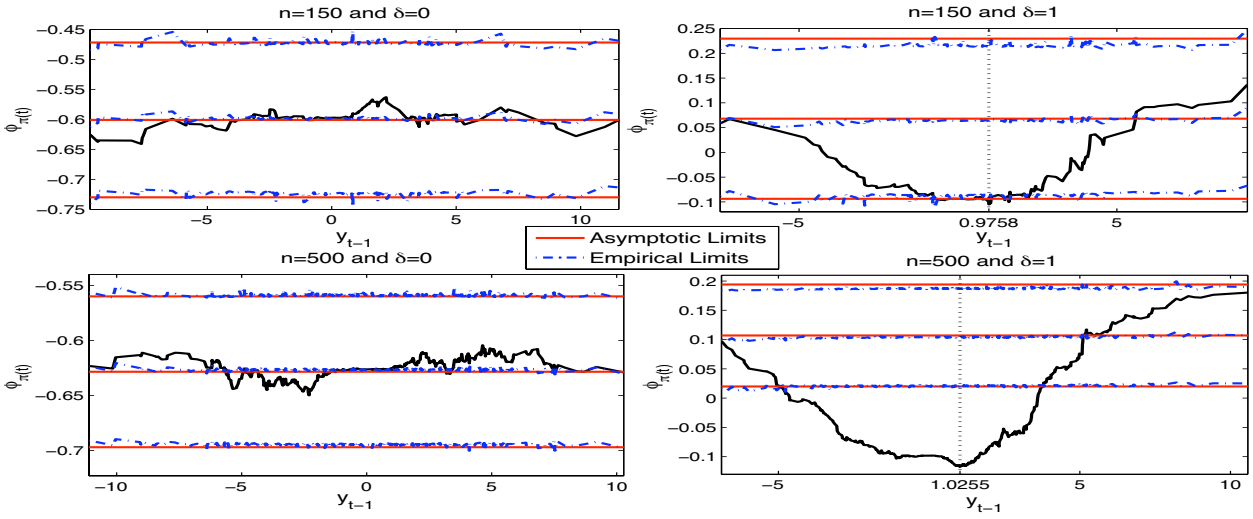


Figure 1: Examples of ARLS working

In figure 1 we have used the asymptotic confidence intervals (28) and the approximation of the empirical confidence intervals (26), using 5000 random orderers. Although apparently the asymptotic intervals are a good aproximation of the empirical ones, we must to check it. To that end, we make the following experiment,

1. Simulate a replica from the model 29.
2. Fit an arranged AR(1) using RLS method described in (8), obtaining $\hat{\phi}_{\pi_t}$ y Λ_t.
3. Calculate the asymptotic confidence interval using expression (28) and the matrix Λ_t with the forgetting factors.
4. Make 5000 random orders. We estimate $\hat{\phi}_{s_{it}}$ and $\widehat{MSE}[\hat{\phi}_{s_{it}}]$ for each one of them, using the

matrix Λ_t obtaining in step 3.

5. Calculate the approximated empirical confidence interval using expression (26).

6. Repeat steps 1 to 5 500 times, for each n and δ.

We have defined several measurements that we have used to compare the confidence intervals.

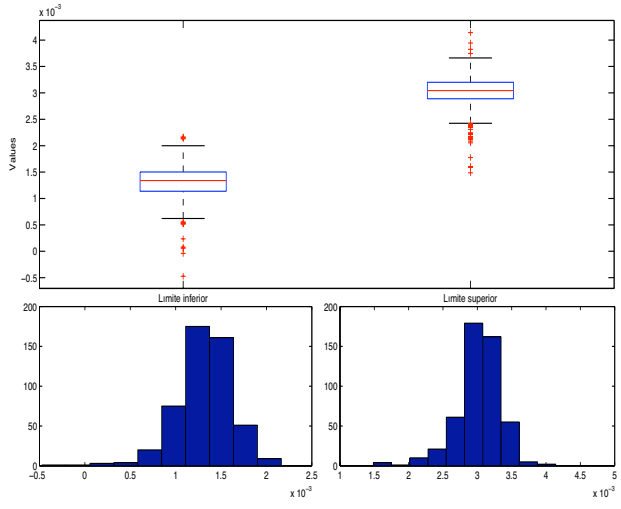
$$D^l = \frac{1}{REP} \sum_{k=1}^{REP} \frac{\frac{1}{n}\sum_{i=1}^{n} l_{ASY}^{(k)} - l_{EMP_i}^{(k)}}{|UL_{ASY}^{(k)} - LL_{ASY}^{(k)}|}, \tag{30}$$

where l will be able to have value UL (upper limit) or LL (lower limit) and REP is the number of replicas. D^l is the mean of the relative differences between asymptotical and empirical limits, therefore we hope a value close to 0. Other measurement is,

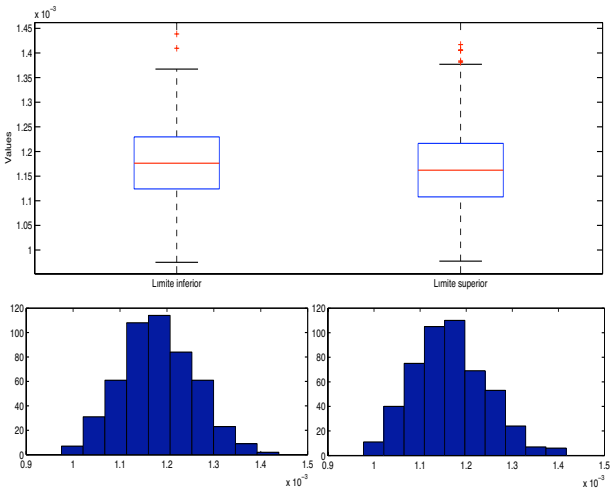
$$A = \frac{1}{REP} \sum_{k=1}^{REP} \frac{\frac{1}{n}\sum_{i=1}^{n} |UL_{EMP_i}^{(k)} - LL_{EMP_i}^{(k)}|}{|UL_{ASY}^{(k)} - LL_{ASY}^{(k)}|}. \tag{31}$$

It is a ratio of the range of the limits, that is, $A \approx 1$ is a good result. Finally, we calculate the coefficient of variation of the empirical limits to check its variability.

In figure 2 we display the results for $n = 150$ and $\delta = 0.8$ in model 29. We can see the boxplot of the difference between asymptotical and empirical limits, and their respective histograms.



(a) Boxplot and histogram of the differences between asymptotical and empirical limits



(b) Boxplot and histogram of the deviation of the empirical limits

Figure 2: Example of performed experiment for $n = 150$ and $\delta = 0.8$

In figure 2a we see that the differences are approximately of the order 10^{-3}, when the average range of the limits is 0.12. Moreover, we must check the variability of the empirical limits, because we have calculated it using only 5000 random sorts. In figure 2b we display the deviation, and they are approximately 0. That means that 5000 random sorts are enough to obtain a significant estimation of (26). In table 1 we show the different results for different values of n and δ.

δ \ n	150					500				
	D^{LI}	D^{LS}	A	CV^{LI}	CV^{LS}	D^{LI}	D^{LS}	A	CV^{LI}	CV^{LS}
0	0.007	0.039	0.955	0.026	0.012	0.003	0.013	0.989	0.003	0.002
0.8	0.008	0.050	0.954	0.014	0.008	0.011	0.025	0.985	0.002	0.002

Table 1: Summarizing results

We can see that the differences are little for any sample size or any δ. Moreover, the average ratio of the ranges is close to 1. Finally, the coefficients of variation are close to 0. This results confirm that the empirical (26) and the asymptotical (28) confidence interval converge.

4.5 Performance of the graphical tool in finite samples

We study the efficiency of the ARLS tool in detect TAR models via simulation. To that end, we make simulations using the following TAR(2;1,1) model,

$$y_t = (-0.6 + \delta I_{(y_{t-1}>1)})y_{t-1} + e_t, \tag{32}$$

where $e_t \sim WN(0, 3)$.

We have used the following test to compare with our proposal.

- **Tests to detect TAR models.** In Section 1, we describe the two main ways to construct tests to detect TAR models. We have choosed one proposal of each way. The first one is Tsay(1989) proposal, which we denote as *Tar-F*. The second one is Hansen(1999) proposal.

- **Test to detect nonlinearity.** The alternative hypothesis of these tests it is not a particular

model. We have used the Tsay(1986) proposal, which we denote as *Ori-F*, the McLeod and Li (1983) proposal and the BDS test proposed by Brock, Dechert and Scheinkman(1996).

In the first experiment we simulate 100 replicas of each δ value in model 32. The sample size was $n = 400$. In figure 3 is displayed the decision rate of each test. The results suggest that the present tool is more efficient.

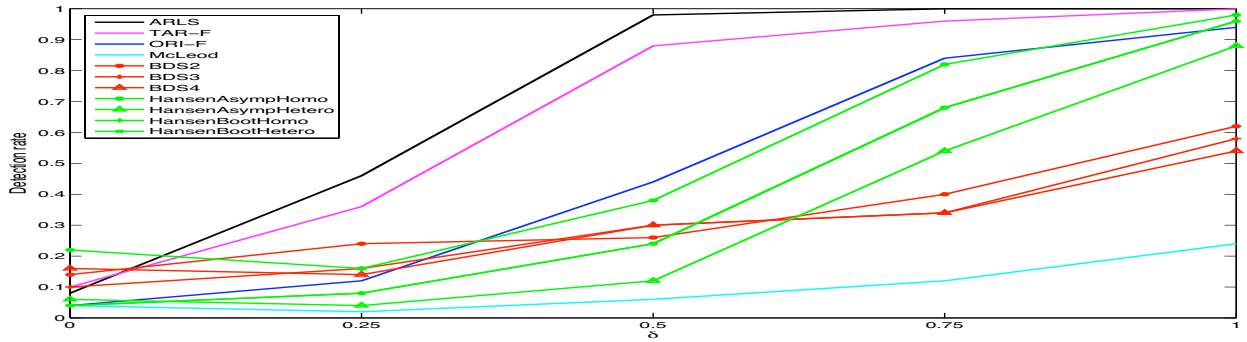


Figure 3: Detection rate with $n = 400$ and 100 replicas

To implement Hansen's tests we have used the Matlab code available in the web page of Professor Hansen[14]. We found that ones are slow and computationally expensive. We tried to repeat that experiment with $n = 150$, but we did not find logical results for Hansen's tests. The code of Professor Hansen is designed to be applied on sunspot data. That data has 289 observations. For this reason, we repeat the first experiment using $n = 289$. In figure 4 is displayed the results.

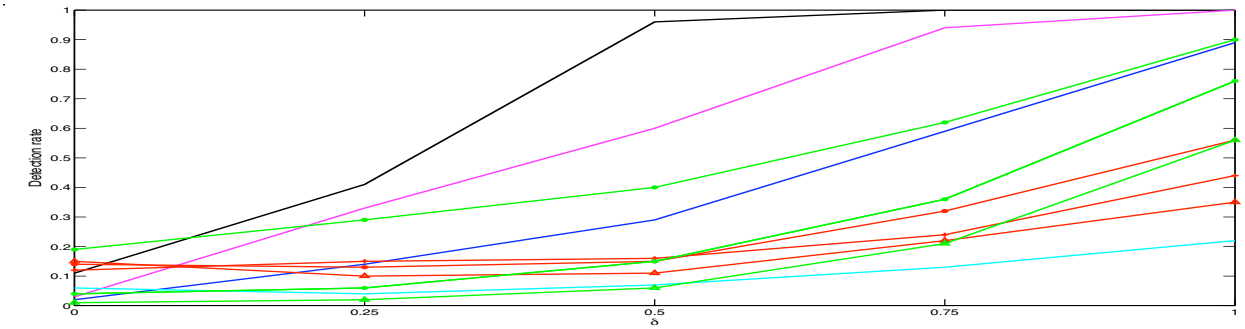


Figure 4: Detection rate with $n = 289$ and 100 replicas

Clements et al. (2003) give results of a previous TAR tests proposed by Hansen(1996). They used models with different characteristic: models with regime heteroskedasticity, with changes in the

intercept, with changes in the slope, among others. We have applied those models on ARLS tool and Tsay's test. We compare our results with the Clements's results for Hansen's test. In Table 5 are summarized the results. We can see that the best results in every model is of ARLS tool.

To check the efficiency more deeply, we repeat the experiment without Hansen's proposals. Now, we simulate 5000 replicas of different δ values in model 32. In figure 6 are displayed the detection rates for the different tests. The results confirm that ARLS tool is the most powerful, to detect TAR models.

Finally, we check the capacity of the detection of threshold values. We have stored each threshold value detected in the last experiment, and we calculate the histogram of the values for each δ. Figure 7 displays the result. The histograms are very peakedness around 1.

5 Automatic procedure for complex TAR models

In last section we have showed that ARLS works correctly with easy models. For this reason, we develop an automatic procedure to detect and to model complex TAR models. The proposed procedure is as follows.

1. **Select the best order p, using model selection criterias.**

2. **Select a set of possible delay parameters $d = 1, 2, \ldots, d_{max}$. Repeat the next steps for each d.**

   2.1 **Order ascending y_t according to dth lag.**

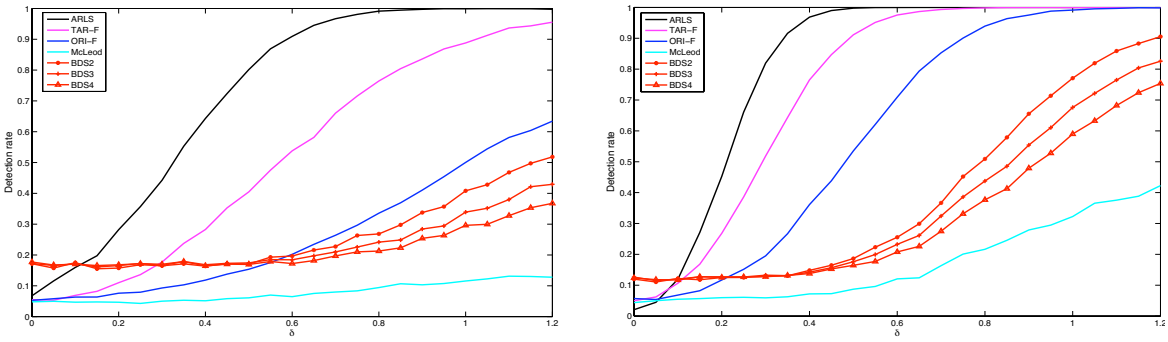
   2.2 **Fit a time-varying arranged AR(p).** If the estimation is out of the limits we obtain a set of possible threshold values C_a with their respective distances from the limits.

   2.3 **Order descending y_t according to dth lag.** The type of orderer should not be significant. But if we use both sorts, we can refine the detection of the threshold value, especially if we have a large sample size.

   2.4 **Fit a time-varying arranged AR(p).** We obtain another set of possible threshold values C_d.

Regime 1			Regime 2			r	n=100 and 500 replicas			n=200 and 1000 replicas		
ϕ_0	ϕ_1	σ	ϕ_0	ϕ_1	σ		Hansen	Tar-F	ARLS	Hansen	Tar-F	ARLS
0	0.3	1	0	0.3	1	-	0.045	0.040	**0.066**	0.054	0.049	**0.063**
-0.75	0.3	1	0	0.3	2	-0.76	0.106	0.040	**0.374**	0.186	0.075	**0.338**
-1.25	0.3	1	0	0.3	1	-0.97	0.503	0.212	**0.524**	0.886	0.376	**0.890**
-1.25	0.3	1	0	0.3	2	-1.25	0.289	0.126	**0.480**	0.578	0.239	**0.593**
0	-0.3	1	0	0.3	2	0.25	0.347	0.218	**0.564**	0.718	0.493	**0.856**
0	-0.7	1	0	0.3	1	0.34	0.806	0.734	**0.902**	0.984	0.969	**1**
0	-0.7	1	0	0.3	2	0.49	0.861	0.694	**0.944**	0.996	0.964	**1**
-1.25	-0.7	1	0	0.3	1	-0.2	0.838	0.858	**0.980**	0.994	0.982	**0.998**
-1.25	-0.7	1	0	0.3	2	-0.1	0.848	0.694	**0.996**	0.998	0.974	**1**
-1.25	-0.7	2	0	0.3	1	0.15	0.912	0.946	**0.980**	0.998	0.999	**1**

Figure 5: Detection rates of ARLS tool, and Hansen(1996) and Tsay(1989) proposals



(a) n=150 **(b)** n=500

Figure 6: Detection rate in 5000 replicas

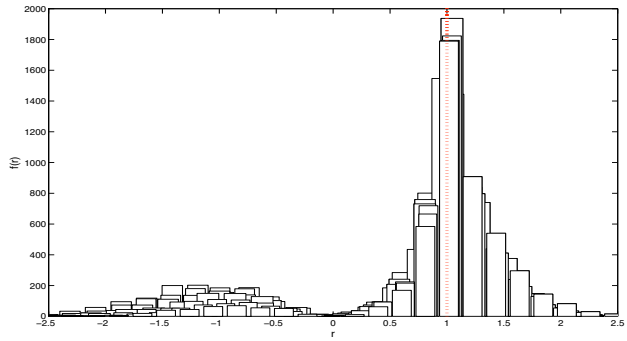


Figure 7: Histogram of the threshold values detected in each δ.

2.5 **Choose the threshold value** r**.** We join the sets C_a and C_d. If there are repeated candidates, we sum the distance. We choose as threshold value r the candidate with higher accumulated distance.

2.6 **TAR model detected?** If we do not have detected a TAR model, we go to step 2. If we have detected we need to continue.

2.7 **Fit a TAR model.** Now, we have the threshold value and the threshold variable. Then, we can fit the TAR model. To select the AR order in each regime, we have used the model selection criteria proposed by Galeano and Peña(2007). They proposed a modification in AIC criteria that improve the selection in TAR models.

2.8 **Are there more regimes?** We repeat the steps 2.1-2.5 on the residual series if $y_{t-d} \leq r$ and $y_{t-d} > r$. If there is not more regimes, the residuals of each regime should be noise, that is, we should not detect more regimes. If we detect more threshold values, we repeat

the procedure until that we do not detect any possible threshold values.

3. **Model selection.** If we have detected some TAR models for differents d values, we choose the TAR model using Galeano and Peña(2007) criteria.

We check the Automatic-ARLS, using a TAR(3;2,3,1) model,

$$y_t = \begin{cases} -0.7y_{t-1} + 0.1y_{t-2} + e_t^{(1)}, & if\ y_{t-2} \leq -0.5; \\ 0.2y_{t-1} + 0.6y_{t-2} - 0.3y_{t-3} + e_t^{(2)}, & if\ -0.5 < y_{t-2} \leq 2; \\ 0.8y_{t-1}e_t^{(3)}, & if\ y_{t-2} > 2. \end{cases} \tag{33}$$

We make 1000 replicas from the model 33 and we repeat the described procedure for each one of them, obtaining:

Percentage of replicas which we detect the right threshold variable: 100%.

Mean of threshold values detected: 2.02.

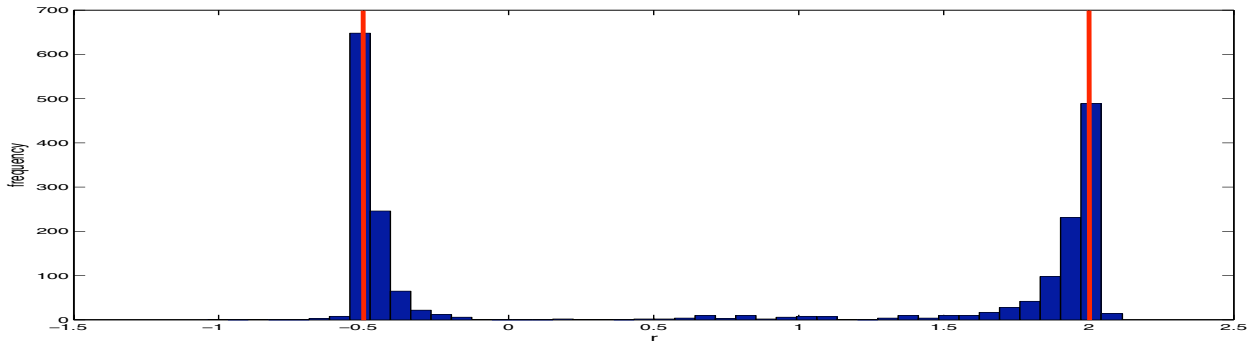


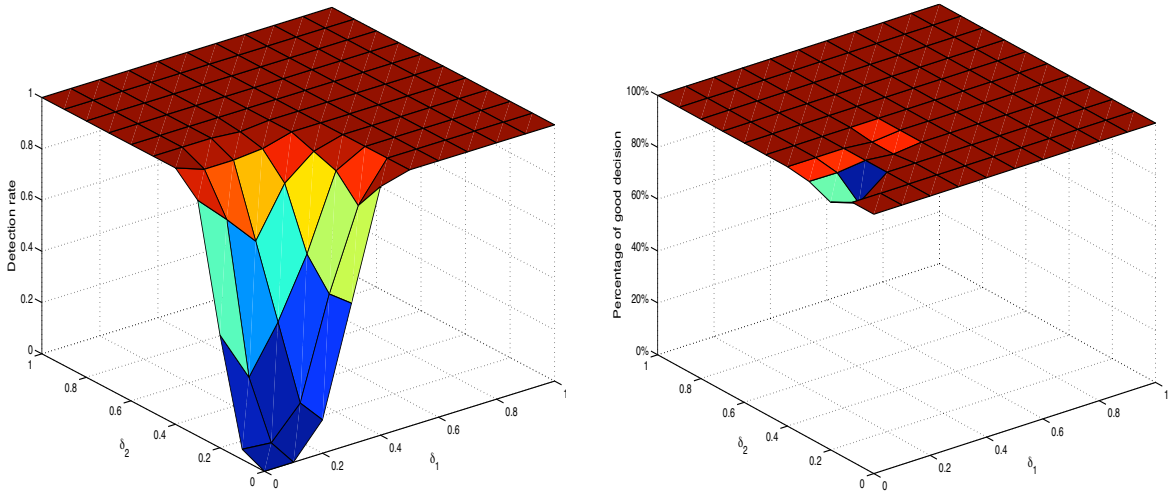
Figure 8: Histogram of the threshold values detected

The numerical results suggest that the Automatic-ARLS is working correctly, because the procedure have always choosed the true threshold variable and it has found an average of 2.02 threshold values. Moreover, in figure 8 is displayed the histogram of the threshold values detected in each replica. We can see two peaks around -0.5 and 2, the true threshold values.

Finally, we check the Automatic-ARLS making the follow experiment. We simulate data from

$$y_t = (-0.6 + \delta_1 I_{(-0.5<y_{t-2}\leq 0.5)} + \delta_2 I_{(0.5<y_{t-2})})y_{t-1} + e_t, \tag{34}$$

and we make 500 replicas for each δ_1 and each δ_2. Lastly, we apply the Automatic-ARLS in each replica. In figure 9a is displayed the detection rate of a TAR model. Figure 9b gives the percentage of replicas which Automatic-ARLS detect the right TAR model. From the plots, it is clear that the Automatic-ARLS is working correctly.



(a) Detection rate of TAR model

(b) Percentage of detection of the right model

Figure 9: Example of performed experiment for $0 \geq (\delta_1, \delta_2) \leq 1$ and $n = 500$

6 Applications

In this section we apply the proposed automatic procedure to some real examples. We have used the Canadian lynx data and the sunspot data. Those data set have been extensively studied. See Tong(1990, Ch.7) for a summary.

6.1 Canadian lynx data

The Canadian lynx data consists of the lynx trapped in the Mackenzie River district of Canada. There are 114 observations. We follow Moran(1953) and make a log transformation. The logged data is displayed in figure 10. The data are in Tong(1990, p. 470).

We apply the Automatic-ARLS to logged data. That selects $p = 2$ and it detects y_{t-2} as the threshold variable. In figure 11 we can see that the threshold detected is 3.2639.

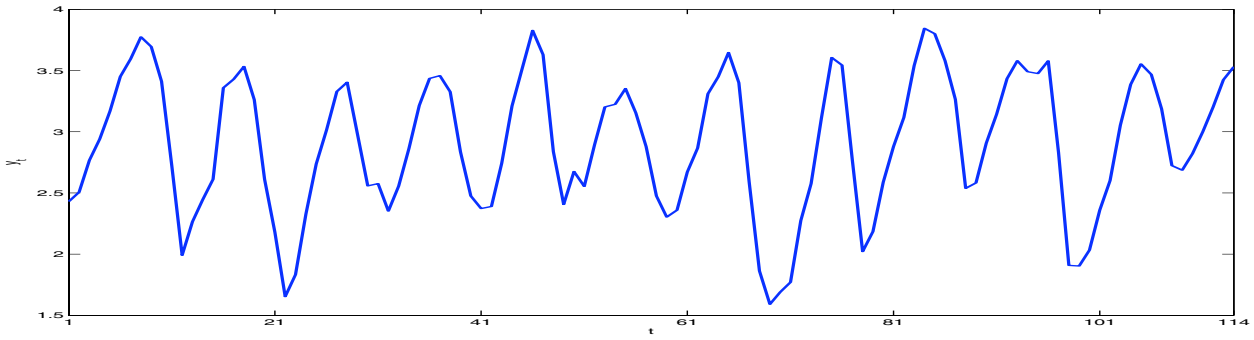


Figure 10: Logged Annual lynx trapped, 1821-1934

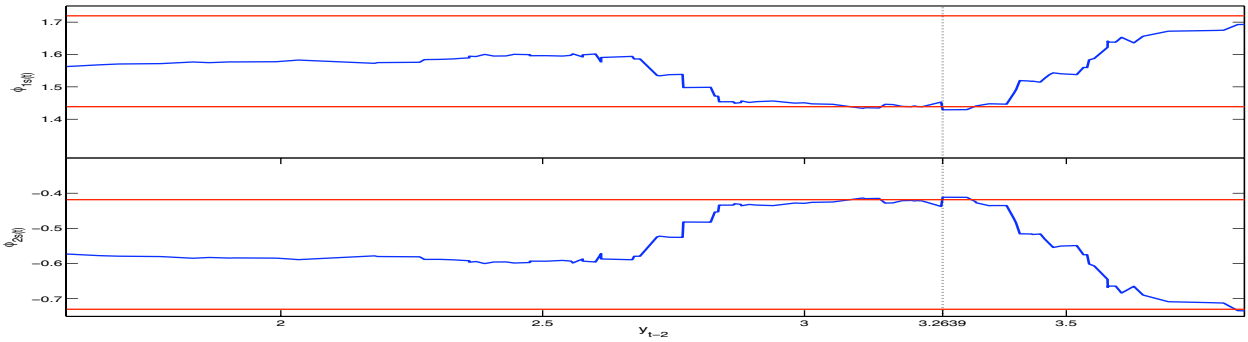


Figure 11: Adaptive estimation of $y_{s(t)}$ arranged according to y_{t-2}.

In Table 2, we compare the proposed models of Tong(1990) and Tsay(1989). The best model is our proposal.

Proposal	Delay	Threshold	TAR orders	AIC	BIC
Tong(1990)	2	3.116	(7,2)	-337.6	-315.2
Tsay(1989)	2	(2.373, 3.154)	(1,7,2)	-347.7	-322.4
Automatic-ARLS	2	3.2639	(3,2)	-353.1	-339.0

Table 2: Different proposals to lynx data

6.2 Sunspot data

The annual sunspot is one of the most analyzed data in time series analysis. We have used data from 1700 to 1920. Thus, we have the same information of Tong's proposition. In figure 12 is displayed the annual sunspot data. We have used the blue observations to detect the TAR model. Later, we have

added the red observations to compare the models proposed by Tong, Tsay and ARLS. The data has been founded in the web page of the National Geophysical Data Center (*www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html*).

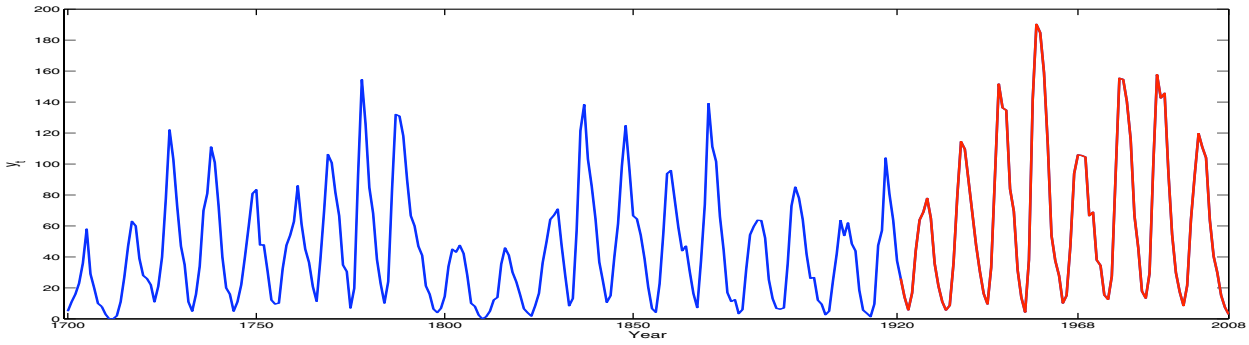


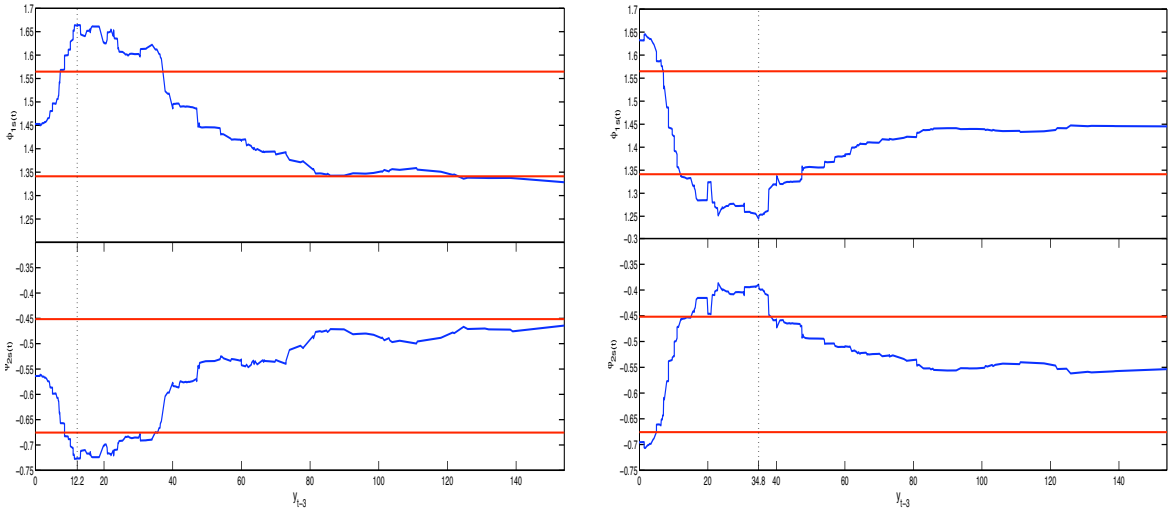
Figure 12: Annual sunspot data from 1700-1988

Automatic-ARLS tool selects $p = 2$ and it detects y_{t-3} as threshold variable. Figure 13 displays the first step. In figure 13a we can see that it chooses as threshold value 12.2 in the ascendant orderer. Whereas, in figure 13b is displayed the descendant orderer. In this case the tool detects as favourite candidate the observation 34.8. Using the criterion explained earlier, Automatic-ARLS detects as threshold value 12.2.

In figure 13 we can see that we detect two possible threshold values depending of the type of sort realized. The procedure selects the observation 12.2 as threshold value. In figure 14 is displayed the next step, which it is selected 34.8.

Now, Automatic-ARLS looks for more possible threshold values in residuals. In figure 14 we can see that we detect the same threshold values in both orderers. Then, the procedure selects the observation 34.8. Now, we can understand why Automatic-ARLS detected two different possible threshold in first step. The procedure detect the first threshold value that it finds.

In Table 3, we compare the proposed models of Tong(1983) and Tsay(1989). We showed the AIC values calculated by Tsay(1989) with data from 1700-1920. Moreover, we have recalculated the AIC values using data from 1700-2008.



(a) Ascending orderer **(b)** Descending orderer

Figure 13: Adaptive estimation of y_{π_t} arranged according to y_{t-3}.

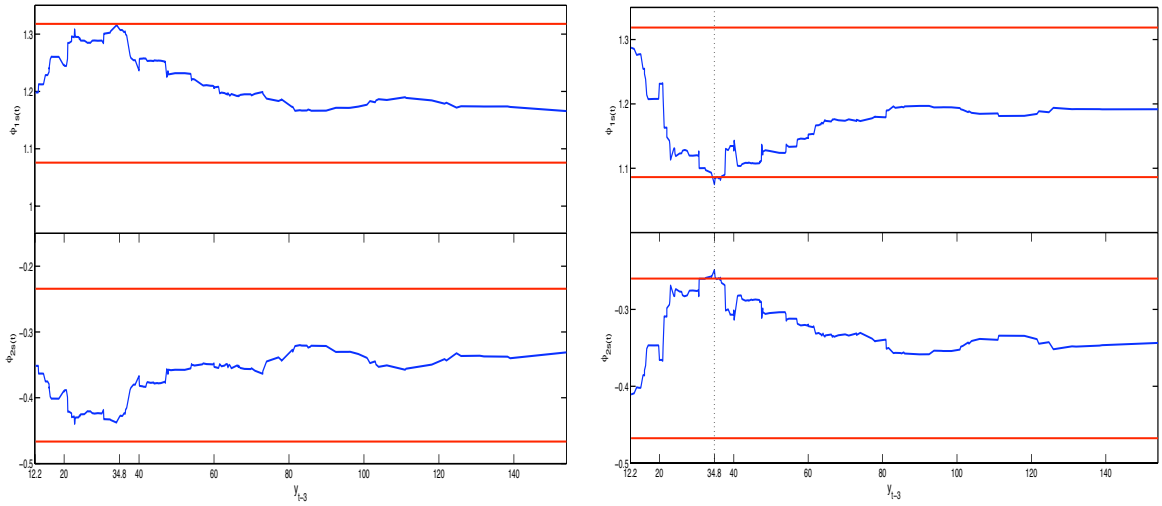
Proposal	Delay	Threshold	TAR orders	$AIC_{1700-1920}$	$AIC_{1700-2008}$
Tong(1983)	3	36.6	(3,11)	1083.8	1513.4
Tsay(1989)	2	(34.8, 70.9)	(11,10,10)	1064.1	1523.8
SALS	3	(30.6, 60, 101.6)	(7,11,1,6)	1009.5	1475.2

Table 3: Different proposals to sunspot data

7 Concluding remarks

One of the main problems to use TAR models is the difficulty to estimate the threshold values. In this paper, we have proposed an easy automatic procedure to solve that problem. Moreover, we have shown that the ARLS tool is more efficient that the previous proposals to detect TAR models.

Recursive estimation is the main technique used in the proposed procedure. For this reason, ARLS tool needs a good estimation to ensure a proper working. The estimation sensibility depend, to a great extent, on the lower bound of the forgetting factor. In practice, a value of $\lambda_{min} = 0.9$ is usually correct. Further research to improve the adaptive estimation could improve the efficiency of ARLS tool.



(a) Ascending orderer **(b)** Descending orderer

Figure 14: Adaptive estimation of e_{π_t} arranged according to $y_{t-3} > 12.2$.

References

[1] Brock, W.A., Scheinkman, J.A., Dechert, W.D. and LeBaron, B. (1996), "A test for independence based on the correlation dimension," *Econometric Reviews,* 15, 3, 197-235.

[2] Chan, K.S. (1990), "Testing for threshold autoregression," *The Annals of Statistics,* 18, 4, 1886-1894.

[3] Chan, K.S. and Tong, H. (1990), "On likelihood ratio tests for Threshold Autoregression," *Journal of the Royal Statistical Society, series B,* 52, 3, 469-476.

[4] Clements, M.P.; Franses, P.H.; Smith, J. and Van Dijk, D. (2003), "On SETAR non-linearity and forecasting," *Journal of Forecasting,* 22, 5, 359-375.

[5] Davidson, R. and MacKinnon, J.G. (1993), *Estimation and inference in econometrics,* New York, NY: Oxford University Press.

[6] De Finetti, B. (1990), *Theory of Probability, Volume 1 & 2,* Chichester, UK: Wiley.

[7] Doob, J. L. (1990), *Stochastic processes,* New York, NY: Wiley Classics Library.

[8] Fan, J. and Yao, Q. (2005), *Nonlinear time series. Nonparametric and parametric methods,* New York, NY: Springer-Verlag.

[9] Fortescue, T.R.; Kershenbaum, L.S. and Ydstie, B.E. (1981), "Implementation of Self-tuning regulators with variable forgetting factors," *Automatica,* 17, 6, 831-835.

[10] Fuller, W. A. and Hasza, D. P. (1981), "Properties of Predictors for Autoregressive Time Series," *Journal of the American Statistical Association,* 76, 373, 155-161.

[11] Galeano, P. and Peña, D. (2007), "Improved model selection criteria for SETAR time series models," Journal of Statistical Planning and Inference, 137, 9, 2802-2814.

[12] Goldstein, M. and Woof, D.A. (1998), "Adjusting exchangeable belifs," *Biometrika,* 85, 1, 39-54.

[13] Hansen, B. (1999), "Testing for Linearity," Journal of Economic Surveys, 13, 5, 551-576.

[14] Hansen, B. (1999), "Testing for Linearity. Matlab Programs," [on line], $<$ *http* : *//www.ssc.wisc.edu/ bhansen/progs/joes*$_9$9.*html* $>$

[15] Keenan, D.M.R. (1985), "A Tukey nonadditivity-type test for time series nonlinearity," *Biometrika,* 72, 1, 39-44.

[16] Kunitomo, N., and Yamamoto, T., (1985), "Properties of Predictors in Misspecified Autoregressive Time Series Models," *Journal of the American Statistical Association,* 80, 392, 941-950.

[17] Landau, I.D.; Lozano, R. and M'Saad, M. (1998), *Adaptive Control,* London, UK: Springer-Verlag.

[18] Lindoff, B. (1997), "On the Optimal Choice of the Forgetting Factor in the Recursive Least Squares Estimator," working paper, Lund Institute of Technology.

[19] Ljung, L. and Söderström, T. (1983), *Theory and practice of recursive identification,* Cambridge, MA: MIT Press.

[20] Maravall, A. (1983), "An application of nonlinear time series forecasting," *Journal of Business and Economic Statistics,* 1, 1, 66-74.

[21] McLeod, A.I. and Li, W.K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis,* 4, 4, 269-273.

[22] Niedzwiecki, M. (2000), *Identification of time-varying processes,* West Sussex, U.K.: Wiley.

[23] Peña, D. and Rodríguez, J. (2005), "Detecting nonlinearity in time series by model selection criteria," *International Journal of Forecasting,* 21, 4, 731-748.

[24] Petruccelli, J.D. and Davies, N. (1986), "A Portmanteau Test for Self-Exciting Threshold Autoregressive-Type Nonlinearity in Time Series," Biometrika, 73, 3, 687-694.

[25] Sánchez, I. (2006), "Recursive estimation of dynamic models using Cook's distance, with application to wind energy forecast," *Technometrics,* 48, 1, 61-73.

[26] Stout, W.F. (1974), *Almost sure convergence,* New York, NY: Academic Press.

[27] Tong, H. (1990), *Nonlinear time series. A dynamical system approach,* Oxford, UK: Oxford University Press.

[28] Tong, H. and Lim, K.S. (1980), "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society, series B,* 42, 3, 245-292.

[29] Tsay, R.S. (1986), "Nonlinearity tests for time series," *Biometrika,* 73, 2, 461-466.

[30] Tsay, R.S. (1989), "Testing and modeling threshold autoregressive processes," *Journal of the American Statistical Association,* 84, 405, 231-240.

[31] Wedlin, A. (1997), "On the notion of second-order exchangeability," *Erkenntnis,* 45, 2-3, 177-194.

[32] White, H. (2000), *Asymptotic theory for Econometricians,* San Diego, CA: Academic Press.

[33] Young, P. (1984), *Recursive estimation and time-series analysis,* New York, NY: Springer-Verlag.