Journal of Forecasting
J. Forecast. **30**, 31–50 (2011)
Published online 6 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/for.1188

Identification of TAR Models Using Recursive Estimation

MIGUEL ÁNGEL BERMEJO,* DANIEL PEÑA AND ISMAEL SÁNCHEZ
Departamento de Estadística, Universidad Carlos III de Madrid, Spain

ABSTRACT
This paper proposes an automatic procedure to identify threshold autoregressive models and specify the values of thresholds. The proposed procedure is based on the time-varying estimation of the parameters using an arranged autoregression. The proposed method not only allows for the automatic identification of the thresholds, but also has a superior identification performance than the competitors. The performance of the proposed procedure is illustrated using Monte Carlo experiments and real data. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS nonlinear time series; recursive estimation; arranged autoregression; TAR models; nonlinearity tests

INTRODUCTION

Thanks to the methodology developed by Box and Jenkins (1970), autoregressive moving average (ARMA) models have been the most successful models for analysing and forecasting linear time series. Part of the impact of the Box–Jenkins methodology can be explained by the use of simple graphical tools, based on the sample autocorrelations, as an aid in the identification and diagnosis steps. In the modelling of nonlinear processes, there is a lack of these kinds of graphical tools. This paper fills this gap in the literature and proposes a graphical method to identify and model the self-exciting threshold autoregressive (TAR) models, proposed by Tong (1978, 1983) and Tong and Lim (1980). A time series y_t is a TAR(k; p, d) if it follows

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} y_{t-i} + e_t^{(j)}, \qquad r_{j-1} \le y_{t-d} < r_j \tag{1}$$

where $j = 1, \ldots, k$. The integer k is the number of regimes, y_{t-d} is the threshold variable and the values of the thresholds are $-\infty = r_0 < r_1 < \ldots < r_k = \infty$; d is called the delay parameter. In each regime, $e_t^{(j)}$ is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and finite and constant standard deviation $\sigma^{(j)}$.

*Correspondence to: Miguel Ángel Bermejo, Departamento de Estadística, Universidad Carlos III de Madrid, 126 28903 Getafe, Madrid, Spain. E-mail: miguelangel.bermejo@uc3m.es

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

The second main approach to detecting 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 has known properties. Petruccelli and Davis (1986) 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 LR test approach, we do not need to know the thresholds to compute the test. However, the tests do not provide any information about the value of the thresholds, which are eventually needed to estimate the TAR model. Tsay (1989) proposed some approximate graphical methods using a scatterplot to detect the thresholds manually. However, more accurate procedures to estimate the values of the thresholds are needed.

We use the idea of arranged autoregression to develop a graphical procedure based on recursive and time-varying estimation of the parameters. The proposed procedure allows us to detect TAR models and also to estimate the thresholds. The proposed procedure has a superior identification performance to previous proposals.

The article is organized as follows. The next section introduces the arranged autoregression and discusses when it is possible in time series. The third section introduces notation and discusses the recursive estimation method. The fourth section gives the proposed graphical procedure, which is called arranged recursive least squares (ARLS), and we illustrate the advantage of the ARLS method. The fifth section details an automatic procedure, which is called Aut-ARLS. Finally, the sixth section applies Aut-ARLS to real data.

ARRANGED AUTOREGRESSION

An AR(p) model can be written as $y_t = X_t'\phi + a_t$, where $X_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and $t = p + 1, \ldots, n$. Following the notation in Tsay (1989), we refer to (y_t, X_t') as a case. We then denote an arranged autoregression as an autoregression with the cases rearranged based on a particular criterion. 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 weak exchangeability, in the sense that the vector of error terms a_t of any rearrangement still maintains its covariance matrix unaltered (Wedlin, 1997).

Let us define S as the set of all possible orders of the time index $t = 1, \ldots, n$, and $s_{i,t}$ as the tth position, $t = 1, \ldots, n$, of the ith element of S, with $i = 1, \ldots, n!$. The subscript i will sometimes be omitted when we refer to a generic element of S. Let us denote π_t as the tth position for the particular case corresponding to arranging the cases in ascending order of the threshold variable y_{t-d}. That is, π_t is the time index of the tth smallest element of $(y_h, \ldots, y_{n-d})$, where $h = \max(1, p + 1 - d)$. To illustrate the arranged autoregressions we show a simple example. Let y_t be a time-varying

DOI: 10.1002/for

AR(1) process $y_t = \phi_{1t}y_{t-1} + a_t$. If we sort (y_t, y_{t-1}) using y_{t-d} as the threshold variable, we obtain the arranged autoregression

$$\begin{pmatrix} y_{\pi_h} \\ y_{\pi_{h+1}} \\ \vdots \\ y_{\pi_n} \end{pmatrix} = \begin{pmatrix} \phi_{1\pi_h} y_{\pi_h -1} \\ \phi_{1\pi_{h+1}} y_{\pi_{h+1}-1} \\ \vdots \\ \phi_{1\pi_n} y_{\pi_n -1} \end{pmatrix} + \begin{pmatrix} a_{\pi_h} \\ a_{\pi_{h+1}} \\ \vdots \\ a_{\pi_n} \end{pmatrix} \tag{2}$$

A TAR(2; 1, d) 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, d) model the sequence of parameters in (2) has a change point at the value r. We will use this property to estimate the parameters in (2) using a 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 (2) 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 procedure.

RECURSIVE METHODS FOR THE ESTIMATION OF TIME-VARYING PARAMETERS

Weighted least squares

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

$$y_{s_t} = X'_{s_t}\phi_{s_t} + a_{s_t}; t = h, \ldots, n \tag{3}$$

where, for any order s_i, 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 $E(a^2_{s_t}) = \sigma^2_t < \infty$. The vector $X_{s_t} = (1_t, y_{s_t-1}, \ldots, y_{s_t-p})'$ 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

$$C_{s_t}(\phi) = \sum_{j=1}^{t} \kappa(t, j)(y_{s_t} - X'_{s_t}\phi)^2 \tag{4}$$

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

$$\kappa(t, j) = \prod_{i=j+1}^{t} \lambda_i, \; j < t \tag{5}$$

where $\kappa(t, t) = 1$ and $0 \le \lambda_t \le 1$ is called the forgetting factor. The forgetting factor can either be constant, $\lambda_t = \lambda$, or time varying. The forgetting factor causes a progressive reduction in the importance of old data in the estimation. For this reason, the estimation is time adaptive. The WLS estimator of (3) is

$$\hat{\phi}_{s_t} = (X'_{s_t}\Lambda_t X_{s_t})^{-1} X'_{s_t}\Lambda_t y_{s_t} \tag{6}$$

where X_{s_t} is the matrix $(X_{s_h}, X_{s_{h+1}}, \ldots, X_{s_t})'$, Λ_t is a diagonal matrix with the elements λ_h, $\lambda_h\lambda_{h+1}, \ldots,$ $\lambda_h\lambda_{h+1} \ldots \lambda_t$ in the main diagonal, and $y_{s_t} = (y_{s_h}, \ldots, y_{s_t})'$. This estimator can be calculated recursively by means of (see, for instance, Ljung and Söderström, 1983)

$$\hat{\phi}_{s_t} = \hat{\phi}_{s_{t-1}} + M_{s_t}^{-1} X_{s_t} \hat{a}_{s_t} \tag{7}$$

where $\hat{a}_{s_t} = y_{s_t} - X_{s_t}'\hat{\phi}_{s_{t-1}}$ is the one-step-ahead prediction error and $M_{s_t} = (X_{s_t}'\Lambda_t X_{s_t})$. Expression (7) is the recursive least squares (RLS) estimator. The gain matrix $M_{s_t}^{-1}$ can also be calculated recursively as

$$M_{s_t}^{-1} = \frac{1}{\lambda_t}\left(M_{s_{t-1}}^{-1} - \frac{M_{s_{t-1}}^{-1} X_{s_t} X_{s_t}' M_{s_{t-1}}^{-1}}{\lambda_t + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}} \right) \tag{8}$$

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, however, we need the properties of RLS estimates under the assumption of a time-invariant AR process. We will use those properties to establish a benchmark with which to compare the estimates of an arranged autoregression like (2), which is clearly time varying, having 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, 1981; Kunitomo and Yamamoto, 1985)

$$\text{MSE}\left(\hat{\phi}_{\text{OLS}}\right) = E\left[\left(\hat{\phi}-\phi\right)\left(\hat{\phi}-\phi\right)'\right] = \frac{\sigma^2}{n}\Gamma^{-1} + O\left(n^{-3/2}\right) \tag{9}$$

where $\Gamma = E(X_t X_t')$, which can be estimated by $\hat{\Gamma} = n^{-1}(X_t'X_t)^{-1}$. The use of a forgetting factor can be interpreted as a shrinkage of the sample size. In OLS each data point has the same contribution in the estimation. However, in the estimator (7), the equivalent or effective sample size is lower than n. If we use, for simplicity, a constant forgetting factor, the equivalent sample size is $n_{\text{eq}} = 1 + \lambda + \cdots + \lambda^{n-1}$. If $n \to \infty$ the asymptotic equivalent sample size is usually termed the asymptotic memory length and is easily computed as

$$N_0 = \frac{1}{1-\lambda} \tag{10}$$

Consequently, the MSE of the RLS estimator is larger as λ is smaller, since it is as if we are a smaller sample size. The asymptotic MSE for the RLS estimator with forgetting factor can be written approximately as

$$\text{MSE}\left(\hat{\phi}_{\text{RLS}}\right) = \sigma^2 E\left[(X_t'\Lambda_t X_t)^{-1}\right] + O\left[(1-\lambda)^{-3/2}\right] \tag{11}$$

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

$$\text{MSE}\left(\hat{\phi}_{\text{RLS}}\right) = \sigma^2 E\left[(X_t'\Lambda_t X_t)^{-1}\right] \tag{12}$$

DOI: 10.1002/for

and is estimated by

$$\widehat{\mathrm{MSE}}\left(\hat{\phi}_{\mathrm{RLS}}\right) = \sigma_t^2 \left(X_t' \Lambda_t X_t\right)^{-1} \tag{13}$$

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

$$\hat{\sigma}_t^2 = \hat{\sigma}_{t-1}^2 + \frac{1}{t-p}\left(\hat{a}_t^2 - \hat{\sigma}_{t-1}^2\right) \tag{14}$$

Adaptive forgetting factors

The forgetting factor will control the influence of the old observations in the estimation. To illustrate its importance we can rewrite the expression (4) as

$$C_{s_t}(\phi) = (y_{s_t} - X_{s_t}'\phi)^2 + \lambda_t C_{s_{t-1}}(\phi) = (y_{s_t} - X_{s_t}'\phi)^2 + \lambda_t (y_{s_{t-1}} - X_{s_{t-1}}'\phi)^2 + \cdots + \lambda_t \lambda_{t-1} \cdots \lambda_2 (y_1 - X_1'\phi)^2 \tag{15}$$

It is easily seen that the influence of the past is downweighted exponentially. In this way a λ_t far away from 1 causes new observations to have a larger influence in the estimation. Consequently, changes in the estimation are quickly found. This higher speed of adaptation, however, increases variability. It can be seen in (8) that the gain matrix, which is a measure of the dispersion of the estimation, grows as λ_t decreases. For this reason, a correct choice of forgetting factor is a key issue for a good adaptive estimation. Several adaptive forgetting factors have been proposed, some of which are:

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

$$\lambda_t^{\mathrm{pre}} = 1 - \alpha \frac{\hat{a}_{s_t}^2}{1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}} \tag{16}$$

  where α is a user-defined 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

$$\lambda_t^{\mathrm{lev}} = 1 - \frac{X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}}{1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}} \tag{17}$$

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

$$\lambda_t^{\mathrm{Cook}} = \lambda_{\min} + (1 - \lambda_{\min}) P\left(\chi_m^2 > m D_t\right) \tag{18}$$

  where $\lambda_{\min}$ is a lower bound of the forgetting factor specified by the user, m is the number of parameters in (3) and D_t is a time-varying version of Cook's distance, calculated by

$$D_t = \frac{X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t} a_{s_t}^2}{m \hat{\sigma}_{s_{t-1}}^2 \left(1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}\right)} \tag{19}$$

DOI: 10.1002/for

where $\hat{\sigma}^2_{s_{t-1}}$ is an estimate of σ^2. Sánchez (2006) shows that this forgetting factor combines the advantages of (16) and (17).

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

ARRANGED RECURSIVE LEAST SQUARES APPLIED TO TAR MODELS

This section describes the proposed procedure for the identification of TAR structures. This procedure is called arranged recursive least squares (ARLS), and can easily be applied using graphical representations. For simplicity of notation, in this section we assume a TAR(2; 1, d) model and rewrite the model (1) as

$$y_t = (\phi + \delta I_{(y_{t-d} > r)})\, y_{t-1} + a_t \tag{20}$$

The main idea of the method is the estimation of the parameters of an arranged AR(p), as in (3), using a time-varying recursive estimation method, as described in (7). If the time-varying autoregression is arranged according to the threshold variable y_{t-d}, the sequence of estimates $\hat{\phi}_{\pi_t}$ will tend to show a structural change around the threshold value r. The procedure includes an analysis of the significance of such structural change.

Initial estimate

The arranged autoregression needs a value for the delay parameter d. Since it will be unknown, alternative values of d will be used. The final value of d can then be selected using, for instance, an information criterion. Once d is selected and the autoregression is arranged, we need an initial value $\hat{\phi}_0$ to initialize the recursive estimation (7). The selection of an appropriate initial value is important, since it can help in the identification of the threshold. A good option is to initialize the recursion using the OLS estimation based on the whole sample. By doing so, we ensure that we start the estimation sequence at an intermediate value between both regimes: the recursive estimation of $\hat{\phi}_{\pi_t}$ will then tend to trace out the shape of a knee around the threshold r that can be used for identification. To better see this point, let $y_t^{(1)}$ and $y_t^{(2)}$ be AR(1) processes with parameters ϕ and $\phi + \delta$, respectively. Define y_t as a time series composed of $y_t^{(1)}$ and $y_t^{(2)}$, that is, $y_t = (y_t'^{(1)}, y_t'^{(2)})'$. In the same manner we can define $X_t = (X_t'^{(1)}, X_t'^{(2)})'$. If we fit an AR(1) model to y_t using OLS, we will obtain, after some algebra:

$$\hat{\phi}_0 = (X_t'X_t)^{-1}X_t'y_t = \hat{\omega}\hat{\phi} + (1-\hat{\omega})\left(\widehat{\phi+\delta}\right) \tag{21}$$

where $\hat{\omega} = X_t'^{(2)}X_t^{(2)}(X_t'^{(1)}X_t^{(1)} + X_t'^{(2)}X_t^{(2)})^{-1}$ and $1 - \hat{\omega} = X_t'^{(1)}X_t^{(1)}(X_t'^{(1)}X_t^{(1)} + X_t'^{(2)}X_t^{(2)})^{-1}$. Therefore, $\hat{\phi}_0$ will tend to be between ϕ and $\phi + \delta$.

Choice of adaptive forgetting factor

After the initial estimate $\hat{\phi}_0$, we need the sequence of recursive estimates $\hat{\phi}_{\pi_t}$ to be as close as possible to the true values. This is attained by the use of a forgetting factor in (8). In our case, the use of an adaptive forgetting factor is apparent. Firstly, if $y_{t-d} \le r$ the model is $y_{\pi_t} = \phi X_{\pi_t} + a_t$. We then require that $\hat{\phi}_{\pi_t}$ moves from $\hat{\phi}_0$ to ϕ very quickly, which implies the use of a small forgetting factor. However, once $\hat{\phi}_{\pi_t}$ is close to ϕ we need the forgetting factor to increase so as to reduce sampling variability.

DOI: 10.1002/for

When $y_{t-d} > r$, the model is $y_{\pi_t} = (\phi + \delta)X_{\pi_t} + a_t$, and again we need a small forgetting factor to allow $\hat{\phi}_{\pi_t}$ to approach $\phi + \delta$. Finally, once $\hat{\phi}_{\pi_t}$ is close to $\phi + \delta$ we need the forgetting factor to increase to reduce variability. As a result, by the use of an appropriate adaptive forgetting factor we can obtain a sequence $\hat{\phi}_{\pi_t}$ with a knee around the threshold r.

Two different problems can arise when using an adaptive estimator (see, for instance, Rao Sripada and Grant Fisher, 1987). Firstly, a large matrix M_t^{-1} in (7) will cause high variability in the sequence $\hat{\phi}_{\pi_t}$. As a result, a false structural change could be detected. This problem can appear when the input observations X_{π_t} are consecutively equal. This problem, however, can be diminished by using the previously variable forgetting factor (18) because in these cases the forgetting factor tends to increase.

The second problem is the opposite state. If M_t^{-1} is too small, the parameter updating in expression (7) could turn off. This might happen, for instance, if the forgetting factor is too high. Consequently, an existing structural change might not be detected. We then need the adaptive forgetting factor to be sensitive to changes in the parameters of the model when it is estimated from an arranged autoregression using y_{t-d}. The forgetting factor (17) proposed by Landau *et al*. (1998) is not a good choice for our purpose, because it is related to the leverage of the new observation. Since data are arranged according to increasing values of y_{t-d}, the new observations and the previous ones are similar. Consequently, the measure of leverage is not going to help detect any change in regime. Conversely, the forgetting factor proposed in Sánchez (2006) is especially convenient for our purpose. This forgetting factor is based on a recursive representation of Cook's distance, allowing the forgetting factor to be sensitive to changes in the parameters of the model.

Confidence intervals of ϕ_{s_t}

We need to evaluate whether a knee observed in the sequence of estimates $\hat{\phi}_{\pi_t}$ is significant, in the sense that it has been produced by a true change in regime. With this aim, we compare the observed trajectory $\hat{\phi}_{\pi_t}$ with the expected trajectories under the assumption of no change in regime. This comparison can be made by constructing a confidence region for the family of trajectories $\hat{\phi}_{s_t}$ when the model is linear, in a similar fashion to a control chart in statistical process control. The construction of such a confidence region can be complicated. One way of computing such a region is by computing the trajectories corresponding to all the elements s_i from the set S. This leads to a set of trajectories $\hat{\phi}_{s_{i,t}}$, $i = 1, \ldots, n!$. A confidence region of coverage $100 \times (1 - \alpha)\%$ can be obtained by keeping the percentiles $\alpha/2$ and $1 - \alpha/2$ of those trajectories. For simplicity, we will use constant confidence intervals along the time index. The two lines representing the percentiles of $\hat{\phi}_{s_t}$ can be interpreted as a confidence region in which the trajectory $\hat{\phi}_{\pi_t}$ will tend to lie under the linearity assumption. The trajectories $\hat{\phi}_{s_t}$ are computed using the same sequence of forgetting factors as in $\hat{\phi}_{\pi_t}$; that is, $\lambda_t \equiv \lambda_{\pi_t}$.

One problem with the computation of these confidence limits is that it can be infeasible to compute all the $n!$ trajectories. The alternative of computing only a random sample of elements of S is still computationally expensive. Therefore, we propose to use confidence intervals of ϕ_{s_t} based on some asymptotic approximation. Under the linearity assumption, the trajectories $\hat{\phi}_{s_t}$ and $\hat{\phi}_{\pi_t}$ can be taken to be random samples from the same population. Hence, for large samples, $\text{MSE}(\hat{\phi}_{s_{i,n}}) \approx \text{MSE}(\hat{\phi}_{\pi_n})$, $\forall s_i \in S$. If we have a sample size n sufficiently large, the asymptotic confidence interval of ϕ_{s_t} under the assumption of linearity can be approximated by

$$\hat{\phi}_0 \pm Z_{1-\alpha/2}\sqrt{\hat{\sigma}_n^2 M_n^{-1}} \tag{22}$$

Then, if the trajectory $\hat{\phi}_{\pi_t}$ has a knee outside the limits (22), we can conclude that there is a change in regime and that the knee identifies the value of the threshold r.

DOI: 10.1002/for

Finite sample performance of the asymptotic intervals

In this section, Monte Carlo experiments are used to evaluate the validity of the asymptotic intervals (22) as an approximation of the intervals obtained with the trajectories of estimates based on random elements of *S*. To that end we use simulated data from

$$y_t = (-0.6 + \delta I_{(y_{t-1}>1)})\, y_{t-1} + a_t \qquad (23)$$

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 a_t will be a $WN(0, 3)$ process.

Given model (23), if the sequence of recursive estimates $\hat{\phi}_{\pi_t}$ lies outside the confidence limits, we will conclude that the model is TAR. Moreover, the value of y_{t-1} corresponding to the value of $\hat{\phi}_{\pi_t}$ that is most distant from the limits will be the estimate of the threshold *r*. Figure 1 shows an example of the 95% confidence intervals based on a replication of model (23). The figure shows the trajectory of the estimates $\hat{\phi}_{\pi_t}$ along with some reference lines. The central solid line in Figure 1 is the OLS estimate $\hat{\phi}_0$ used as the initial estimate. The dotted central line is the average of 5000 trajectories $\hat{\phi}_{s_{i,t}}$ based on random elements of the set *S*. The solid limits are computed as in (22). The dotted limits show, for each *t*, the 2.5th and 97.5th empirical percentiles of those 5000 trajectories of $\hat{\phi}_{s_{i,t}}$.

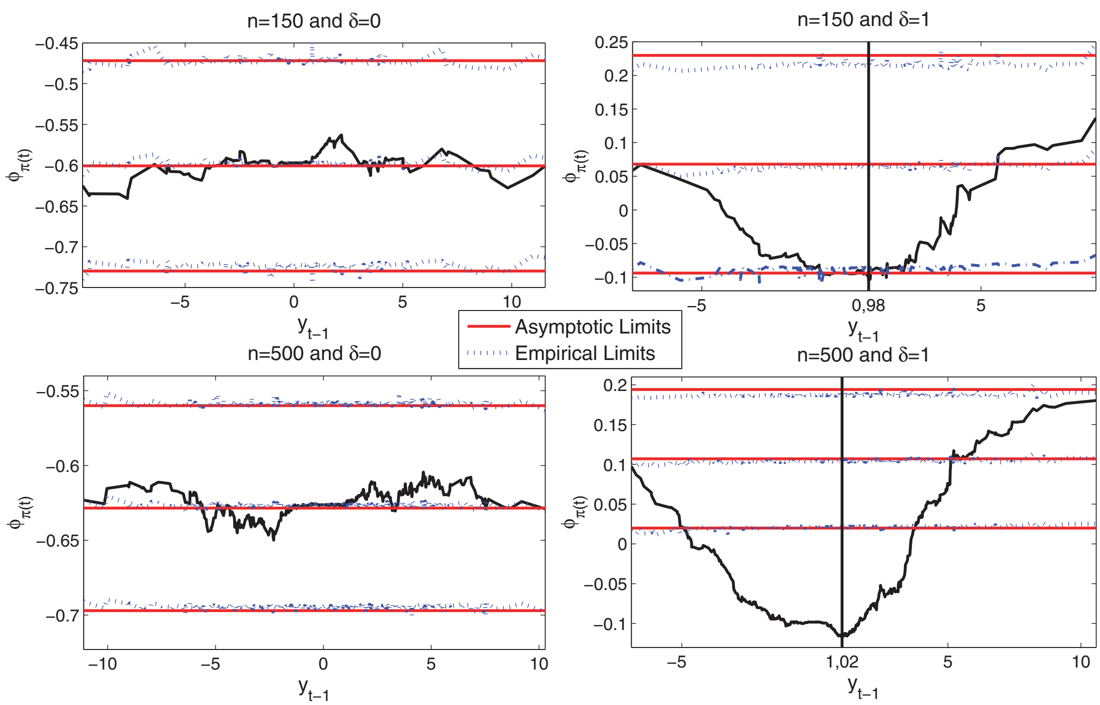

Figure 1. Recursive estimates $\hat{\phi}_{\pi_t}$ of the time-varying AR(1) model arranged according to y_{t-1} for simulated data from model (23) for alternative values of δ and n. The solid lines are the initial estimate $\hat{\phi}_0$ (centre line) and the 95% asymptotic confidence intervals, whereas the dotted lines are the empirical counterparts based on 5000 random trajectories $\phi_{s_{i,t}}$ with $i = 1, \ldots, 5000$. This figure is available in colour online at www.interscience.wiley.com/journal/for

DOI: 10.1002/for

Table I. Values of measures B^l, A, CV^l (where $l \equiv U$ (upper limit) or $l \equiv L$ (lower limit)) for simulated data from model (23) for alternative values of δ and n

δ	$n = 150$					$n = 500$				
	B^L	B^U	A	CV^L	CV^U	B^L	B^U	A	CV^L	CV^U
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

When $\delta = 0$, that is, when the process is linear, Figure 1 shows that the sequence of estimates $\hat{\phi}_{\pi_t}$ is always inside the limits. Conversely, when $\delta = 0.8$ the sequence $\hat{\phi}_{\pi_t}$ lies outside the limits in some periods, showing a change in the parameter. In this case, the most distant value from the limits is located around the value $y_{t-1} \approx 1$, which is the true value of the threshold.

Figure 1 shows that the asymptotic limits can be a good approximation to the empirical ones. In order to validate this result, the experiment is repeated 1000 times for each n and δ. Several measures are used to compare the confidence intervals. Firstly, we define a measure of the relative difference between the asymptotic and empirical limits. It is called B^l and is calculated by

$$B^l = \frac{1}{R}\sum_{k=1}^{R} \frac{\frac{1}{n}\sum_{i=1}^{n} l_a^{(k)} - l_{e_i}^{(k)}}{\left|U_a^{(k)} - L_a^{(k)}\right|} \tag{24}$$

where l will have value U (upper limit) or L (lower limit), and R is the number of replications. The subscripts a and e are referred to as the asymptotic and empirical limits, respectively. This measure represents the distance between each empirical limit and its asymptotic approximation and is standardized by the amplitude of the interval. This standardization is necessary to give a better idea of the size of the bias with respect to the size of the interval we want to estimate. The second measure is defined by

$$A = \frac{1}{R}\sum_{k=1}^{R} \frac{\frac{1}{n}\sum_{i=1}^{n} \left|U_{e_i}^{(k)} - L_{e_i}^{(k)}\right|}{\left|U_a^{(k)} - L_a^{(k)}\right|} \tag{25}$$

This measure compares the amplitude of the empirical and asymptotic intervals. The target is then $A \approx 1$. The empirical limits have been estimated using 5000 random orders from the $n!$ possible orders. The coefficient of variation of the empirical limits has also been computed. Table I summarizes the results for different values of n and δ in model (23). The coefficients of variation are approximately 0, which means that 5000 random orders are enough to compute the empirical limits. The values of B^l are very small, of the order of 10^{-2} for every value of n and δ. Finally, the average value of A is close to 1. For $n = 150$ we have $A \approx 0.95$, which means that the amplitude of the asymptotic limits are about 5% larger than the empirical ones, so that the asymptotic intervals are slightly conservative. In conclusion, the results confirm that the asymptotic confidence intervals are a good approximation of the empirical ones.

PERFORMANCE OF THE ARLS METHOD IN FINITE SAMPLES

This section evaluates the efficiency of the ARLS method in detecting threshold nonlinearity using Monte Carlo experiments. The proposed ARLS method is compared with existing methods in the

DOI: 10.1002/for

Table II. Detection rate of different tests for simulated data from 100 replications of model (23) for different values of δ, $n = 289$ and $\alpha = 0.05$

Method	δ				
	0	0.25	0.5	0.75	1
McLeod and Li (1983)	0.06	0.04	0.07	0.13	0.22
Tsay (1986)	0.02	0.14	0.29	0.59	0.89
Tsay (1989)	0.03	0.33	0.60	0.94	1.00
BDS2	0.14	0.13	0.15	0.32	0.56
BDS3	0.12	0.15	0.16	0.24	0.44
BDS4	0.15	0.10	0.11	0.22	0.35
Hansen bootstrap homoskedastic	0.04	0.06	0.15	0.36	0.76
Hansen bootstrap heteroskedastic	0.03	0.04	0.12	0.27	0.68
Hansen asymptotic homoskedastic	0.19	0.29	0.40	0.62	0.90
Hansen asymptotic heteroskedastic	0.01	0.02	0.06	0.21	0.56
ARLS	0.06	0.41	0.96	1.00	1.00

literature. The competing methods are both general methods to detect nonlinearities as well as specific methods for TAR models. Among the general tests, we include in the evaluation the Tsay (1986) proposal, which is based on Tukey's one-degree-of-freedom test for non-additivity; the McLeod and Li (1983) proposal, which is a portmanteau test based on examining squared residuals; and the BDS test proposed by Brock *et al.* (1996), which is based on the correlation dimension. The tests for TAR models included in this evaluation are the Tsay (1989) and Hansen (1997, 1999a,b) proposals. Tsay (1989) shows that his test is more powerful than the Petruccelli and Davies (1986) test and, for this reason, we do not include this test in the comparison. Hansen (1997) tests are based on structural change tests proposed in Davies (1987) and Andrews and Ploberger (1994). Hansen (1999a,b) proposes four tests. Two are based on the asymptotic distribution and the other two are based on a bootstrap distribution. Each test is then adapted to homoskedastic or heteroskedastic errors. The significance level used in all tests is $\alpha = 0.05$. The proposed ARLS is based on the 95% asymptotic confidence interval.

The first experiment is based on model (23) for $\delta = (0, 0.25, 0.5, 0.75, 1)$. In this experiment, the Hansen (1999) tests are implemented using the code from the author's web page. For this reason, the sample size is set to $n = 289$. Owing to the computational cost of these tests, we restrict the number of replications of this experiment to 100. Table II shows the detection rate of each test. It can be seen that the best results are obtained by ARLS.

The experiment is repeated using 5000 replications from model (23) for alternative values of $\delta \in [0, 1.2]$ and $n = (150, 500)$, but without the Hansen tests, which were too time consuming to compute. Figure 2 displays the detection rates. The results confirm that the proposed ARLS method is the most efficient. It can also be seen that general nonlinear tests can have very low power to detect threshold nonlinearities.

A second experiment is based on Clements *et al.* (2003). They give results of the TAR tests proposed by Hansen (1997). The Monte Carlo experiment is based on simulated data from 10 models with different characteristics. For $n = 100$ the authors set the number of replications to 1000, and for $n = 200$ the number of replications is 500. Table III summarizes the best result obtained for the Hansen (1997) tests in Clements *et al.* (2003), and our results for the ARLS method and the Tsay (1989) test. The best detection rate is always obtained with ARLS.

DOI: 10.1002/for

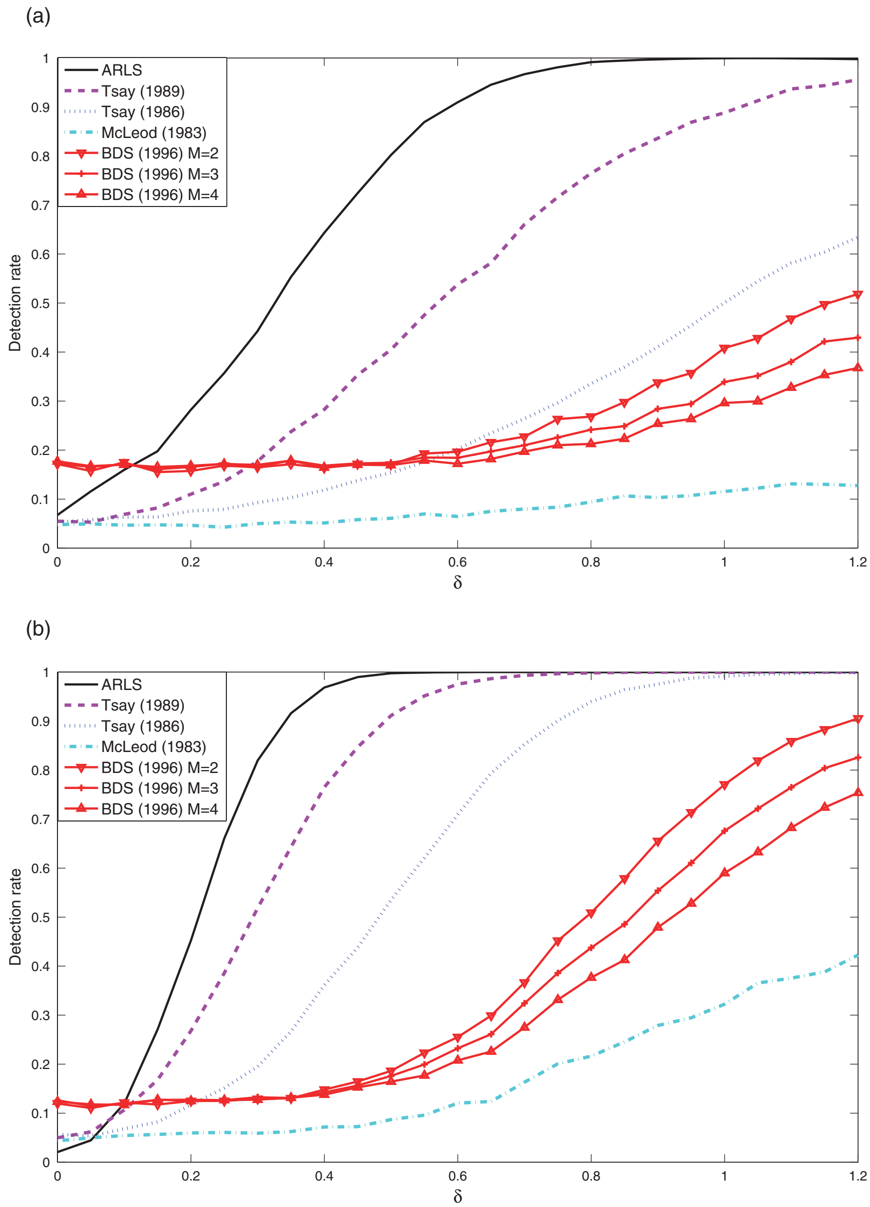

Figure 2. Detection rate of ARLS method and McLeod (1983), BDS and Tsay (1986, 1989) tests for simulated data from 5000 replications of model (23) for alternative values of δ and n [(a) $n = 150$; (b) $n = 500$]. This figure is available in colour online at www.interscience.wiley.com/journal/for

Table III. Detection rates of ARLS method and Hansen (1997) and Tsay (1989) tests on simulated data from model (1) for $d = 1$ and $p = 1$. Hansen's results have been extracted from Clements *et al*. (2003)

Regime 1			Regime 2			r	n = 100 and 1000 replications			n = 200 and 500 replications		
ϕ_0^1	ϕ_1^1	σ^1	ϕ_0^2	ϕ_1^2	σ^2		Hansen (1997)	Tsay (1989)	ARLS	Hansen (1997)	Tsay (1989)	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	0.193	0.040	0.374	0.312	0.075	0.338
−1.25	0.3	1	0	0.3	1	−0.97	0.741	0.212	0.744	0.982	0.376	0.990
−1.25	0.3	1	0	0.3	2	−1.25	0.391	0.126	0.480	0.716	0.239	0.793
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	0.826	0.734	0.902	0.984	0.969	1
0	−0.7	1	0	0.3	2	0.49	0.865	0.694	0.944	0.998	0.964	1
−1.25	−0.7	1	0	0.3	1	−0.2	0.913	0.858	0.980	0.996	0.982	0.998
−1.25	−0.7	1	0	0.3	2	−0.1	0.876	0.694	0.996	1	0.974	1
−1.25	−0.7	2	0	0.3	1	0.15	0.958	0.946	0.980	0.998	0.999	1

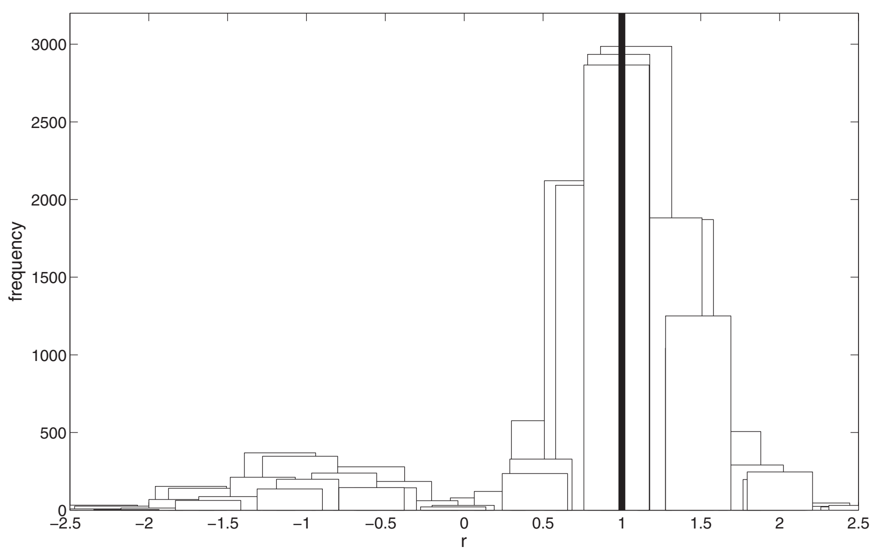

Figure 3. Histogram of the values of thresholds estimated with the ARLS method on simulated data from 5000 replications of model (23) for alternative values of δ and $n = 500$

Finally, the efficiency of the ARLS method in estimating the values of the thresholds is evaluated. 5000 replications from model (23) are simulated for values of $\delta \in [0, 1]$ and $n = 500$. When the ARLS method detects a TAR model, it estimates a threshold $\hat{r}$. The histogram of these estimated values $\hat{r}$ is then computed for each δ. Figure 3 displays the results, showing that the histograms have a sharper peak around 1, which is the true value of the threshold.

DOI: 10.1002/for

AUTOMATIC PROCEDURE TO IDENTIFY TAR MODELS

This section gives an automatic procedure for detecting and modelling TAR models, based on the results shown in previous sections. The proposed procedure is called Aut-ARLS and is as follows.

Step 1. Select the AR order p by considering the partial autocorrelation function and the Akaike information criteria. Select a maximum value $d_{\max}$ for the delay parameter.

Step 2. Fit a time-varying AR(p) arranged in ascending order according to y_{t-d} for each value of $d = 1, \ldots, d_{\max}$, and using the forgetting factor (18). Compute the 95% asymptotic confidence interval using expression (22) for each parameter of the vector $\hat{\Phi}$. If there are values of the sequence of estimates $\hat{\phi}_{j\pi_t}$, $j = 1, \ldots, p$, lying outside the limits, the coresponding values y_{t-d} are kept as a set of possible thresholds $r^{\text{asc}} = \left\{ y_{t-d} \| \hat{\Phi}_{\pi_t} - \hat{\Phi}_0 \right| > Z_{1-\alpha/2}\sqrt{\hat{\sigma}_n^2 M_n^{-1}} \right\}$. Moreover, the corresponding distances from the asymptotic limits are also kept at $h^{\text{asc}} = \{|\hat{\Phi}_{\pi_t} - \hat{\Phi}_0 \| y_{t-d} \in r^{\text{asc}}\}$.

Step 3. Repeat step 2 with y_t arranged in descending order of y_{t-d}. Let us denote by η_t the tth position of the element of S corresponding to arranging the cases in descending order of the variable y_{t-d}. We will denote the sequence of adaptive estimates as $\hat{\phi}_{j\eta_t}$, $j = 1, \ldots, p$. The use of both sequences of estimates, based on ascending or descending orders of y_{t-d}, can be useful when the threshold is near to an extreme of the threshold variable, avoiding the possibility that the threshold is masked by the sampling variability of the initial estimates. As before, we obtain the set of possible thresholds $r^{\text{desc}} = \left\{ y_{t-d} \| \hat{\Phi}_{\eta_t} - \hat{\Phi}_0 \right| > Z_{1-\alpha/2}\sqrt{\hat{\sigma}_n^2 M_n^{-1}} \right\}$ and their respective distances $h^{\text{desc}} = \{|\hat{\Phi}_{\eta_t} - \hat{\Phi}_0 \| y_{t-d} \in r^{\text{desc}}\}$.

Step 4. Choose the threshold r for each value of d among the candidate values r^{asc} and r^{desc} as follows. The sets r^{asc} and r^{desc}, and their respective distances h^{asc} and h^{desc}, are merged. If there are coincident candidates in r^{asc} and r^{desc}, their corresponding distances in h^{asc} and h^{desc} are added. The selected threshold r is then the candidate with the largest distance.

Step 5. For each value of d, fit a TAR model using the corresponding estimates of r, should it exist. The AR order in each regime is selected using the model selection criteria proposed by Galeano and Peña (2007), where a modification of AIC that improves the selection in TAR models is proposed.

Step 6. Repeat steps 2–5 on the residual series obtained in step 5, when $y_{t-d} \le r$ and $y_{t-d} > r$. Since the variance can be different in each regime, this step can be performed by splitting the residuals of each regime and analysing each regime separately. The procedure is repeated until no additional thresholds are found.

Step 7. Select the final TAR model. If several TAR models are detected for different d values, the final TAR model is selected using the Galeano and Peña (2007) criteria.

The efficiency of Aut-ARLS is evaluated via simulation. The first experiment consists of the simulation of 1000 replications from the TAR(3; [2, 3, 1], 2) model:

$$y_t = \begin{cases} -0.7y_{t-1} + 0.1y_{t-2} + a_t^{(1)}, & \text{if } y_{t-2} \le -0.5 \\ 0.2y_{t-1} + 0.6y_{t-2} - 0.3y_{t-3} + a_t^{(2)}, & \text{if } -0.5 < y_{t-2} \le 2 \\ 0.8y_{t-1}a_t^{(3)}, & \text{if } y_{t-2} > 2 \end{cases} \tag{26}$$

where the sequence a_t^j will be a $WN(0, \sigma^{(j)})$ process, with $\sigma^{(j)} = (3, 2, 5)$. The sample size is $n = 500$. Aut-ARLS is applied to the simulated data, obtaining for each replication the delay parameter

DOI: 10.1002/for

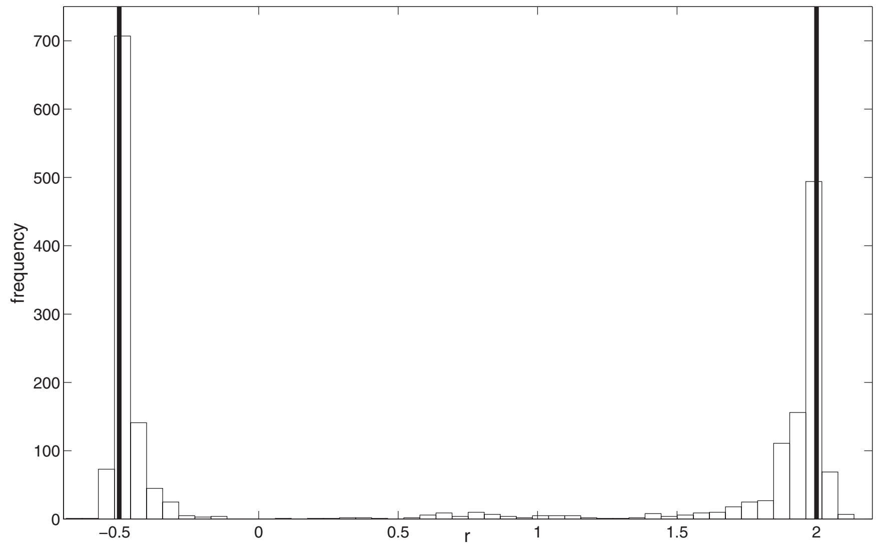

Figure 4. Histogram of the threshold values detected by Aut-ARLS for simulated data from model (26)

d and the threshold *r*. The detection rate of a TAR model using Aut-ARLS is 100%. Moreover, the percentage of replications where Aut-ARLS detects the right threshold variable is also 100%. Finally, we calculate the mean of the number of thresholds detected by Aut-ARLS, obtaining 2.02. Figure 4 displays the histogram of the thresholds estimated for each replication. The histogram has two peaks around −0.5 and 2, which are the true values of the thresholds.

In the second experiment data are simulated from

$$y_t = \left(-0.6 + \delta_1 I_{(-0.5<y_{t-2}\le 0.5)} + \delta_2 I_{(0.5<y_{t-2})}\right) y_{t-1} + a_t \tag{27}$$

where $(\delta_1, \delta_2) \in [0, 1]$ and $n = 500$. The sequence a_t will be a $WN(0, 3)$ process. The experiment consists of the simulation of 500 replications for alternative values of δ_1 and δ_2. Figure 5(a) displays the detection rate of threshold nonlinearities. Figure 5(b) gives the percentage of replications where Aut-ARLS detects the true threshold variable. From the plots, it is clear that Aut-ARLS works correctly.

APPLICATIONS

In this section the proposed Aut-ARLS method is applied to some real examples. We have used the Canadian lynx data and the Sunspot data, which have been extensively studied in the literature: see Tong (1990, Ch. 7) for a summary.

Canadian lynx data

The Canadian lynx data consist of the lynx trapped in the Mackenzie River district of Canada. There are 114 observations. The data are in Tong (1990, p. 470). We follow Moran (1953) and make a log transformation. The logged data are displayed in Figure 6.

DOI: 10.1002/for

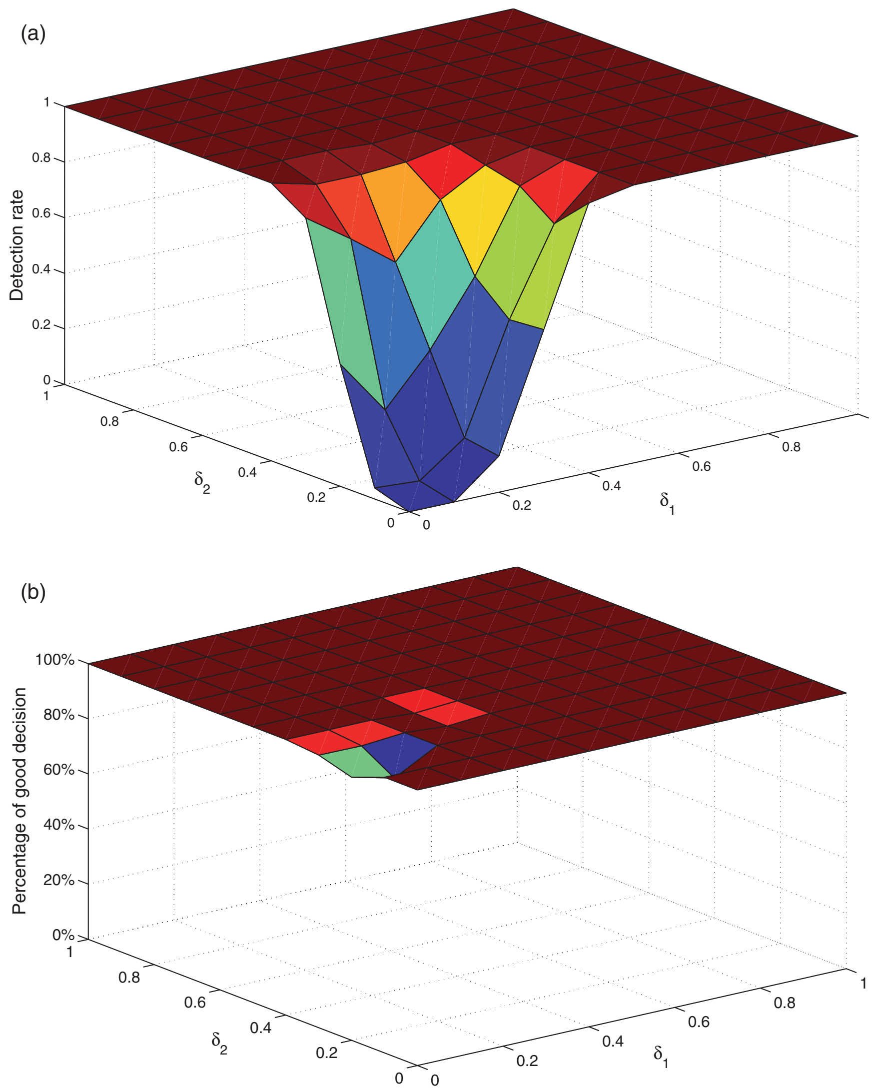

Figure 5. Summary of results for the second experiment based on model (27) for values $0 \leq (\delta_1, \delta_2) \leq 1$ and $n = 500$. (a) Detection rate of TAR model; (b) percentage of detection of the true threshold variable. This figure is available in colour online at www.interscience.wiley.com/journal/for

We apply Aut-ARLS to these logged data, $p = 2$ being selected as the AR order and y_{t-2} as the threshold variable. Figure 7 shows that the threshold detected is 3.2639.

Table IV summarizes the model proposed by Aut-ARLS, and the models proposed by Tong (1990) and Tsay (1989). The minimum AIC and BIC correspond to the proposed Aut-ARLS.

Sunspot data

This popular dataset consists of the annual sunspot index from 1700 to 2008. We have used observations from 1700 to 1920, so that we have the same information as used in Tong (1983). Figure 8

DOI: 10.1002/for

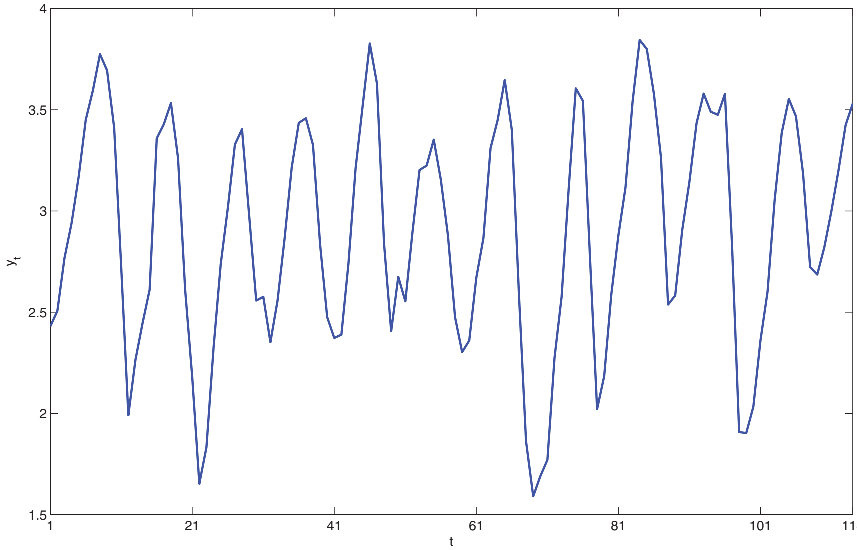

Figure 6. Logged annual lynx trapped, 1821–1934. This figure is available in colour online at www.interscience.wiley.com/journal/for

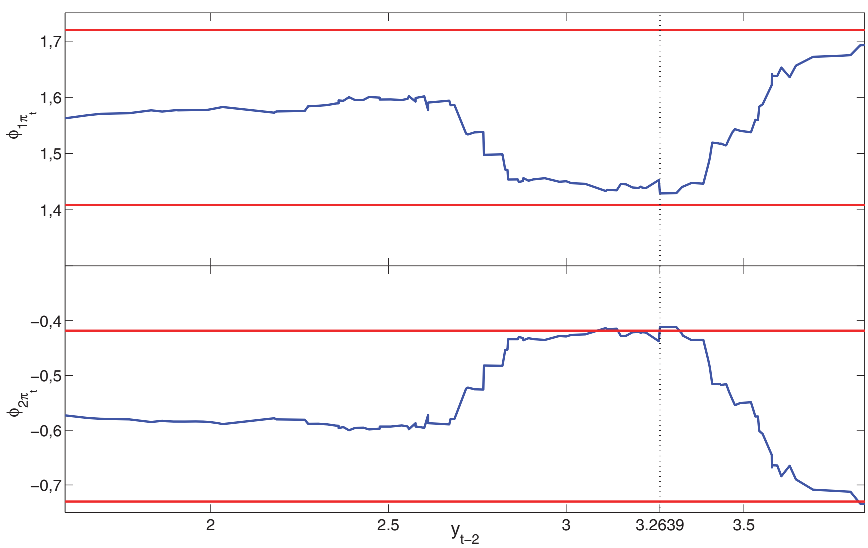

Figure 7. Recursive estimates $\hat{\phi}_{\pi_t}$ of the time-varying AR(2) model arranged according to y_{t-2} for logged lynx data. This figure is available in colour online at www.interscience.wiley.com/journal/for

displays the annual sunspot data. The data can be accessed from the National Geophysical Data Center web site.

Aut-ARLS selects an AR order $p = 3$ and detects y_{t-3} as threshold variable. Figure 9 displays the recursive estimates $\hat{\phi}_{\pi_t}$ of the time-varying AR(3) arranged according to y_{t-3}. Figure 9(a) and (b) display the estimates in ascending and descending order, respectively. In ascending order, 30.7 is

Table IV. TAR models proposed by Tsay (1989), Tong (1990) and Aut-ARLS for logged lynx data

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

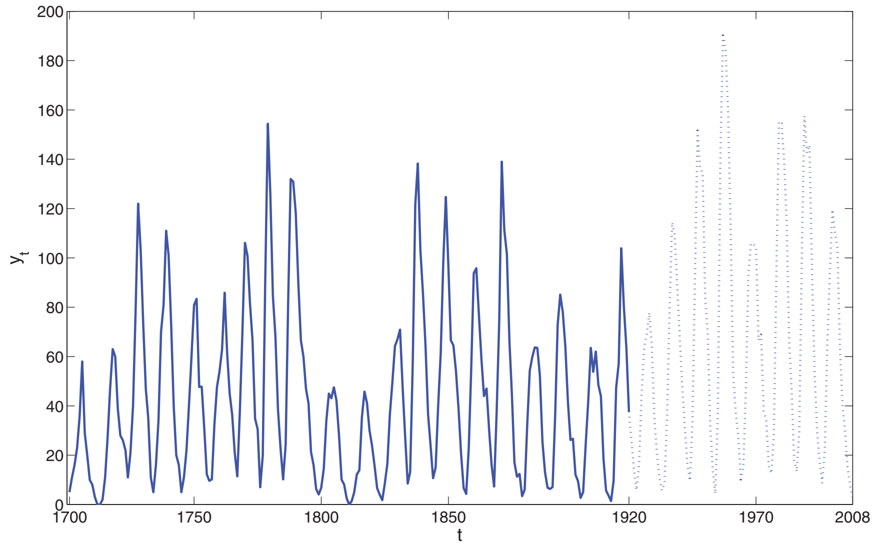

Figure 8. Annual sunspot data from 1700 to 2008. This figure is available in colour online at www.interscience.wiley.com/journal/for

the main candidate to be selected as threshold, whereas in descending order there are several candidates. Aut-ARLS selects as threshold value 30.7, using the criterion explained above.

Once the threshold has been selected, a TAR model can be estimated. Aut-ARLS looks for more possible thresholds in the residuals of each regime, but no additional thresholds were found.

We have used the TAR models proposed by Tsay (1989), Tong (1983) and Aut-ARLS to obtain out-of-sample forecasts of sunspot index during the period 1921–2008. Table V summarizes the proposed models, the AIC values reported by Tsay (1989) with data from 1700 to 1920 and the mean absolute error (MAE) and root mean square error (RMSE) for forecast horizon $h = 1, 2$. The proposed Aut-ARLS has lower AIC, MAE and RMSE values than Tong and Tsay proposals.

CONCLUDING REMARKS

One of the limitations in using TAR models is the lack of a simple procedure, like the Box–Jenkins methodology for linear ARMA models, that assists in the identification. This article proposes a method to remove that limitation. One of the main problems in the use of TAR models is the

DOI: 10.1002/for

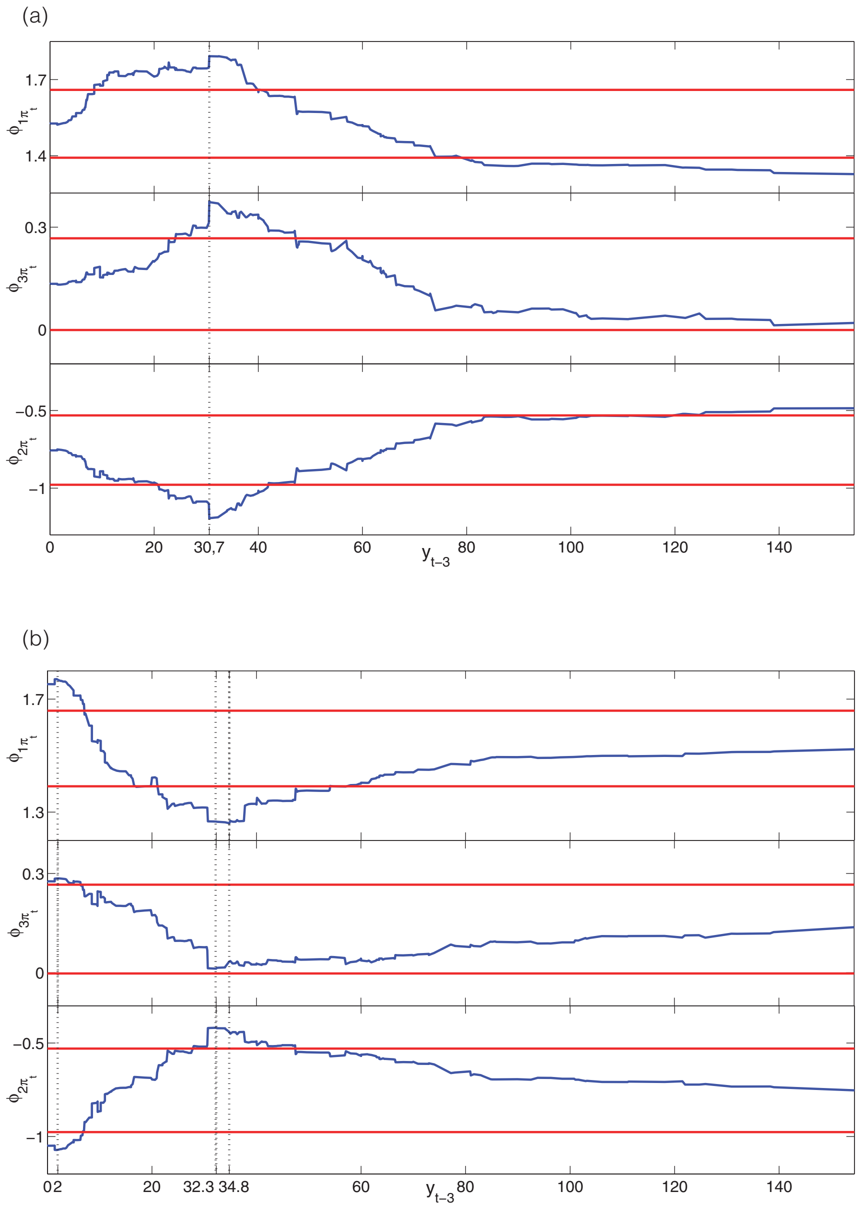

Figure 9. Recursive estimates $\hat{\phi}_{\pi_t}$ and $\hat{\phi}_{\eta_t}$ of the time-varying AR(3) arranged according to y_{t-3} for sunspot data. (a) Ascending order; (b) descending order. This figure is available in colour online at www.interscience.wiley.com/journal/for

difficulty of identifying the values of the thresholds. Similar to the use of correlograms in Box and Jenkins (1970) to identify the orders of ARMA models, we used a graphical tool to identify TAR models and to select the thresholds. The proposed procedure, besides its simplicity, is more efficient than the usual tests for TAR detection.

DOI: 10.1002/for

Table V. TAR models proposed by Tsay (1989), Tong (1983) and Aut-ARLS for sunspot data

Proposal	Delay	Threshold	TAR orders	$AIC_{1700-1920}$	$MAE_{1920-2008}$		$RMSE_{1920-2008}$	
					$h = 1$	$h = 2$	$h = 1$	$h = 2$
Tong (1983)	3	36.6	(3, 11)	1083.8	12.31	12.42	16.59	16.69
Tsay (1989)	2	(34.8, 70.9)	(11, 10, 10)	1064.1	11.74	11.87	15.52	15.61
Aut-ARLS	3	30.7	(7, 11)	1045.9	11.37	11.47	15.33	15.42

ACKNOWLEDGEMENTS

This research has been supported by MICINN Grant SEJ2007-64500. The authors gratefully acknowledge the detailed comments and suggestions of the editors of this Special Issue, Terry Mills and Ruey Tsay. They have been very helpful in improving the presentation of our work.

REFERENCES

Andrews DWK, Ploberger W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**: 1383–1414.

Box GEP, Jenkins GM. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day: San Francisco, CA.

Brock WA, Scheinkman JA, Dechert WD, LeBaron B. 1996. A test for independence based on the correlation dimension. *Econometric Reviews* **15**: 197–235.

Chan KS. 1990. Testing for threshold autoregression. *Annals of Statistics* **18**: 1886–1894.

Chan KS, Tong H. 1990. On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B* **52**: 469–476.

Clements MP, Franses PH, Smith J, Van Dijk D. 2003. On SETAR non-linearity and forecasting. *Journal of Forecasting* **22**: 359–375.

Davies RB. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**: 33–43.

Fortescue TR, Kershenbaum LS, Ydstie BE. 1981. Implementation of self-tuning regulators with variable forgetting factors. *Automatica* **17**: 831–835.

Fuller WA, Hasza DP. 1981. Properties of predictors for autoregressive time series. *Journal of the American Statistical Association* **76**: 155–161.

Galeano P, Peña D. 2007. Improved model selection criteria for SETAR time series models. *Journal of Statistical Planning and Inference* **137**: 2802–2814.

Hansen B. 1997. Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* **2**: 1–14.

Hansen B. 1999a. Testing for linearity. *Journal of Economic Surveys* **13**: 551–576.

Hansen B. 1999b. Testing for Linearity. Matlab Programs. Available: http://www.ssc.wisc.edu/bhansen/progs/joes_99.html [9 May 2010].

Kunitomo N, Yamamoto T. 1985. Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* **80**: 941–950.

Landau ID, Lozano R, M'Saad M. 1998. *Adaptive Control.* Springer: London.

Ljung L, Söderström T. 1983. *Theory and Practice of Recursive Identification*. MIT Press: Cambridge, MA.

McLeod AI, Li WK. 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4**: 269–273.

Moran PAP. 1953. The statistical analysis of the Canadian lynx cycle. I. Structure and prediction. *Australian Journal of Zoology* **1**: 163–173.

Petruccelli JD, Davies N. 1986. A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* **73**: 687–694.

Rao Sripada N, Grant Fisher D. 1987. Improved least squares identification. *International Journal of Control* **46**: 1889–1913.
Sánchez I. 2006. Recursive estimation of dynamic models using Cook's distance, with application to wind energy forecast. *Technometrics* **48**: 61–73.
Tong H. 1978. On a threshold model. In *Pattern Recognition and Signal Processing*, Chen CH (ed.). Sijhoff & Noordhoff: Amsterdam; 101–141.
Tong H. 1983. *Threshold Models in Nonlinear Time Series Analysis*. Springer: New York.
Tong H. 1990. *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.
Tong H, Lim KS. 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* **42**: 245–292.
Tsay RS. 1986. Nonlinearity tests for time series. *Biometrika* **73**: 461–466.
Tsay RS. 1989. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* **84**: 231–240.
Wedlin A. 1997. On the notion of second-order exchangeability. *Erkenntnis* **45**: 177–194.

Authors' biographies:
Miguel Ángel Bermejo is currently a PhD student in Statistics at the Universidad Carlos III de Madrid. His research interests are nonlinear time series, wind power forecasting and adaptive estimation.

Daniel Peña is Professor of Statistics at the Universidad Carlos III de Madrid. He is a Fellow of ASA and IMS. He has held visiting positions at the University of Wisconsin-Madison and the University of Chicago. His papers have appeared in JASA, Annals of Statistics, JRSS-B, Biometrika and Technometrics, among other journals. His research interests are time series, Bayesian statistics and robust methods.

Ismael Sánchez is associate professor in the Department of Statistics. His main research areas are non-linear time series, multivariate statistical process control as well as industrial applications. He is especially involved in the development of statistical methods for the prediction of renewable energies, like wind or solar power, where efficient non-linear methods are needed.

Authors' addresses:
Miguel Ángel Bermejo and **Ismael Sánchez**, Departamento de Estadística, Universidad Carlos III de Madrid, Av. de la Universidad 30 28911 Leganés, Madrid, Spain.

Daniel Peña, Departamento de Estadística, Universidad Carlos III de Madrid, C. Madrid 126 28903 Getafe, Madrid, Spain.

DOI: 10.1002/for