



# Linear approximation of the Threshold AutoRegressive model: an application to order estimation

Francesco Giordano<sup>1</sup> · Marcella Niglio<sup>1</sup>  · Cosimo Damiano Vitale<sup>1</sup>

Accepted: 3 April 2022 / Published online: 10 May 2022  
© The Author(s) 2022, corrected publication 2022

## Abstract

This paper proposes a linear approximation of the nonlinear Threshold AutoRegressive model. It is shown that there is a relation between the autoregressive order of the threshold model and the order of its autoregressive moving average approximation. The main advantage of this approximation can be found in the extension of some theoretical results developed in the linear setting to the nonlinear domain. Among them is proposed a new order estimation procedure for threshold models whose performance is compared, through a Monte Carlo study, to other criteria largely employed in the nonlinear threshold context.

**Keywords** Linear approximation · Threshold model ·  $L^2$  norm · AIC

**Mathematics Subject Classification** 37M10 · 47A58

## 1 Introduction

The complexity of nonlinear time series models often makes it difficult to analyze the main features of their dynamics, the derivation of their statistical properties and the identification and estimation of the models. In this case the issues developed in the linear domain cannot be further adapted and new tools need to be introduced.

We herein propose a bridge between the linear AutoRegressive Moving Average (ARMA) model and the nonlinear Self-Exciting Threshold AutoRegressive (SETAR) model (Tong and Lim 1980; Tong 1990). We introduce a linear approximation of the threshold model that offers the opportunity to investigate the dynamics of the generating process.

---

✉ Marcella Niglio  
mniglio@unisa.it

<sup>1</sup> Department of Economics and Statistics, Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

In more detail, we show that the proposed linear approximation of the nonlinear SETAR process is given by an ARMA model. Even though one might expect that this approximation is given by an autoregressive (AR) process, we show that there is also a moving average (MA) component.

The linear approximation is obtained using the best (in  $L^2$  norm) one-step-ahead predictor of the SETAR process, and the corresponding coefficients of the ARMA approximation are theoretically derived.

Then, leveraging these results which allow us to derive the ARMA model that best approximates the threshold structure, we introduce a new procedure by which to estimate the order of the SETAR process, in what can be considered one of the potential applications of the proposed linear approximation.

Before going into the details of our proposal and to better introduce our contribution, we need to clarify the distinction between *linear representation* and *linear approximation* of a nonlinear process. The *linear representation* of nonlinear models was extensively presented in Francq and Zakoïan (1998), and in the references therein. It is based on a new parametrization of the nonlinear model that leads to a *weak* ARMA representation, where the adjective *weak* relates to the fact that the assumptions, usually given on the noises of the ARMA model (such as independence), are not satisfied by the linear ARMA representation. (For an example of linear representation of the subdiagonal bilinear model, see Pham 1985).

In this domain, more recently, Kokoszka and Politis (2011) used the definition of weak linear time series, and even showed that Autoregressive Conditional Heteroscedasticity-type and Stochastic Volatility-type processes do not belong to this class.

The definition of *linear approximation* of  $X_t$  is instead based on different issues. The aim is to introduce of an approximation that allows one to distinguish two components in the process  $X_t$ :

$$X_t = X_t^L + Y_t$$

with  $X_t^L$  the linear approximation and  $Y_t$  the nonlinear component, where  $X_t^L$  is given by the linear causal process:

$$X_t^L = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (1)$$

with  $\psi_0 = 1$ ,  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  and  $\{\varepsilon_t\}$  a sequence of independent and identically distributed (i.i.d.) random variables, with  $E[\varepsilon_t] = 0$  and  $E[\varepsilon_t^2] = \sigma^2$ .

The derivation of the linear approximation  $X_t^L$  is associated in the literature with the use of the Volterra series expansion of  $X_t$ , where  $X_t^L$  is obtained considering the first-order term of the expansion of  $X_t$ . (For an example of the bilinear model, see Priestley 1988, p. 58.)

The spirit of the Volterra expansion has inspired various contributions to linear approximation, with applications in many domains (among others see Huang et al. 2009 and Schoukens et al. 2016).

In this study we obtain a *linear approximation* of  $X_t$  by borrowing the definition of the best mean square one-step-ahead predictor (see among others Brockwell and Davis 1991, Sect. 2.7).

The first results are presented in Proposition 1, whereas the elements included in  $X_t^L$  are further discussed in Remark 1 that leads to the proposed linear approximation of  $X_t$ . Moreover, we show that the proposed linear approximation is an ARMA process and we use this result in the order estimation of the threshold process, thus confining the computational effort needed to estimate the order of a nonlinear process to the linear domain.

In particular, in Sect. 2 we establish the conditions under which a linear approximation of  $X_t$  can be obtained. In Sect. 3 the linear approximation is used to implement a new order estimation procedure whose consistency property is shown; in Sect. 4 some extensions of the linear approximation are given after removing some conditions fixed in the previous pages, and the role played by the SETAR parameters in the approximation is discussed; in Sect. 5 the order estimation procedure is evaluated and compared to other information criteria, through a Monte Carlo study. Some final comments are given at the end, and all proofs are given in the Appendix.

## 2 Linear approximation of the SETAR process: main results

Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a SETAR nonlinear process, given by:

$$X_t = \sum_{j=1}^k \left( \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} X_{t-i} \right) \mathbb{I}(X_{t-d} \in \mathbb{R}_j) + \varepsilon_t, \tag{2}$$

where  $\{\varepsilon_t\}$  is a sequence of continuous i.i.d.  $\mathbb{R}$ -valued random variables with mean zero and  $E[\varepsilon_t^2] = \sigma_\varepsilon^2 < \infty$ ,  $\mathbb{I}(\cdot)$  is the indicator function,  $k$  is the number of regimes,  $p_j$  is the order of the autoregressive regimes,  $X_{t-d}$  is the threshold variable,  $d$  is an integer representing the threshold delay,  $p_j$  is a nonnegative integer,  $k$  and  $d$  are positive integers,  $\mathbb{R}_j = (r_{j-1}, r_j]$  for  $j = 1, \dots, k-1$  with  $-\infty = r_0 < r_1 < \dots < r_{k-1} < r_k = \infty$ , and  $\mathbb{R}_k = (r_{k-1}, +\infty)$  are subsets of the real line such that  $\mathbb{R} = \bigcup_{j=1}^k \mathbb{R}_j$ .

In the following, to find a stochastic linear approximation of the process  $X_t$ , we use the  $L^2$  norm given by  $\|X_t - X_t^L | \mathcal{I}_{t-1}\|_{L^2} = E \left[ (X_t - X_t^L)^2 | \mathcal{I}_{t-1} \right]^{1/2}$ , with  $\mathcal{I}_{t-1}$  the set of information on  $X_t$  available up to time  $t - 1$ .

To show the theoretical results, we take advantage of an alternative representation of the threshold process that, to avoid heavy notation (which would not help in understanding the issues), is assumed to have  $k = 2$  regimes (the case with  $k > 2$  is discussed in Sect. 4.3) and null intercepts ( $\phi_0^{(j)} = 0$ , for  $j = 1, 2, \dots, k$ ). This guarantees the following form to process (2):

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} \mathbb{1}(X_{t-d} \leq r_1) + \Phi_2 \mathbf{X}_{t-1} [1 - \mathbb{1}(X_{t-d} \leq r_1)] + \boldsymbol{\varepsilon}_t, \quad (3)$$

where

$$\mathbf{X}_t = \begin{bmatrix} X_t \\ \dots \\ X_{t-p+1} \end{bmatrix}, \quad \Phi_j = \begin{bmatrix} \phi_1^{(j)} & \dots & \phi_{p-1}^{(j)} & \phi_p^{(j)} \\ & & \mathbf{I}_{p-1} & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ \mathbf{0} \end{bmatrix} \quad (4)$$

with  $\mathbf{I}$  the identity matrix,  $\mathbf{0}$  the null vector, and the autoregressive order  $p$  (the assumption that the two regimes have the same autoregressive order simplifies the presentation of results and can be easily met including null coefficients in the model (see Sect. 4.2)). Without loss of generality, we can suppose that  $d = 1$  and  $r_1 = 0$  (these assumptions will be further discussed in Sect. 4.1); then, the process (3) can be written as

$$\mathbf{X}_t = \mathbf{A}_{I_{t-1}} \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (5)$$

where

$$\mathbf{A}_{I_{t-1}} = \begin{cases} \Phi_1 & \text{if } X_{t-1} \leq 0 \\ \Phi_2 & \text{if } X_{t-1} > 0. \end{cases} \quad (6)$$

Note that all results given in the following are based on the assumptions of strict stationarity (referred to as stationarity in the next pages) and ergodicity of the process. In the threshold domain the milestone in this framework is given by Petrucci and Woolford (1984), who state a set of sufficient and necessary conditions for the stationarity and ergodicity of a SETAR(2;1) model; sufficient conditions for the more general SETAR(2; $p$ ) process, are given in Chan and Tong (1985) and Bec et al. (2004).

There are at least two difficulties inherent in obtaining the linear approximation of the SETAR model; these make this model different from other nonlinear models, such as Markov Switching structures (Hamilton 1989). First, the SETAR process (5) has stochastic coefficients that depend on the process  $X_t$  itself: it implies, among others things, that it is not easy to obtain its moments (whose sufficient conditions for their existence are given in Lemma 2 in the Appendix). Second, the SETAR process does not fulfil many regularity conditions that can assist in obtaining the linear approximation. (For example, the first derivative of  $X_t$ , carried out to derive the coefficients of the first-order Volterra expansion (see Priestley 1988, p. 26), may not exist when the skeleton of the process assumes a value equal to the threshold value.)

Starting from these two points, to derive a linear approximation of a SETAR(2;  $p$ ) process, we need to introduce an additional notation that simplifies the presentation.

Let  $X_t = \mathbf{e}_1^T \mathbf{X}_t$  be the SETAR(2; $p$ ) process (5), with  $\mathbf{e}_1$  a  $(p \times 1)$  vector with 1 as its first element and all remaining  $(p - 1)$  elements are zero, whereas  $\mathbf{A}^T$  is the transpose of  $\mathbf{A}$ . Further, let

$$\boldsymbol{\pi} = \begin{pmatrix} \pi \\ 1 - \pi \end{pmatrix} \otimes \mathbf{I}_p \quad \text{and} \quad \mathbf{P} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \tag{7}$$

be two matrices where  $\boldsymbol{\pi}$  is a  $(2p \times p)$  matrix, with  $\pi = Pr(X_t \leq 0)$ ,  $\mathbf{I}_p$  is an identity matrix of order  $p$  and  $\otimes$  is the Kronecker product, whereas  $\mathbf{P}$  is a  $(2 \times 2)$  matrix, with  $\pi_{ij} = Pr(X_t \in \mathbb{R}_j | X_{t-1} \in \mathbb{R}_i)$ , for  $i, j = 1, 2$  and  $\mathbb{R}_1 = (-\infty, 0]$ ,  $\mathbb{R}_2 = (0, \infty)$ . Moreover, consider the following matrix:

$$\mathbf{K} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_2 \end{pmatrix} (\mathbf{P} \otimes \mathbf{I}_p), \tag{8}$$

with  $\boldsymbol{\Phi}_i$ , for  $i = 1, 2$ , defined in (4).

Given the previous notation, we can now introduce the linear approximation of  $X_t$ . Let  $X_t(1)$  be the best (in  $L^2$  norm) one-step-ahead predictor for the SETAR(2; $p$ ) process,  $X_t(1) = E[X_t | \mathcal{I}_{t-1}]$ ; we can now state the following results.

**Proposition 1** *Let  $X_t$ ,  $t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2; $p$ ) process with  $E(\varepsilon_t^2) < \infty$  and  $\gamma_2 < 1$  (with  $\gamma_2$  given in Lemma 2 in the Appendix); then, given the linear process  $X_t^L$  as in (1), we have the following results in  $L^2$  norm:*

- (i)  $\arg \min_{\{g_j\}} \|X_t - X_t^L | \mathcal{I}_{t-1}\|_{L^2}^2 = g_j(\mathcal{I}_{t-1}) = \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1, \forall j \geq 1$ , -  
that is, the minimizer - say,  $X_{t|t-1}^L$ , is given by  $X_{t|t-1}^L = \sum_{j=1}^{\infty} g_j(\mathcal{I}_{t-1}) \varepsilon_{t-j}$ ;
- (ii)  $X_t(1) = \sum_{j=1}^{\infty} g_j(\mathcal{I}_{t-1}) \varepsilon_{t-j}$ ,

where  $\mathcal{I}_{t-1} = \{X_{t-1}, X_{t-2}, \dots\}$  and the subscript  $t - 1$  in  $X_{t|t-1}^L$  is due to the conditional set  $\mathcal{I}_{t-1}$ .

**Proof** See the Appendix.

The result of Proposition 1 is twofold. First it gives, for the SETAR(2;  $p$ ) model, a representation of the optimal one-step-ahead predictor,  $X_t(1)$ , which is a generalized linear process with coefficients  $g_j(\mathcal{I}_{t-1})$  that relate to the process itself. Second, this optimal predictor corresponds to  $X_{t|t-1}^L$  when the minimization is done with respect to the linear process  $X_t^L$ .

Now we need to emphasize the following key remark.

**Remark 1** Proposition 1 gives a representation of the best one-step-ahead predictor for the SETAR(2; $p$ ) process. However, the quantities  $g_j(\mathcal{I}_{t-1})$  are not easily managed because they are nonlinear functions of the observations. An easy way to derive a linear process, as defined in (1), is to consider the expectation of  $g_j(\mathcal{I}_{t-1})$  - that is

$$g_0 = 1, \quad g_j^{(2)} = E(g_j(\mathcal{I}_{t-1})) = E\left(\mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1\right), \quad j \geq 1. \tag{9}$$

If we suppose that  $\gamma_1 < 1$  (see Lemma 2) and use the same arguments as in the proof of Lemma 2, we have that  $g_j^{(2)} = O(\gamma_1^j)$ ,  $\forall j \geq 1$ . So, it follows that  $\sum_{j=0}^\infty |g_j^{(2)}| < \infty$ . Then, we have the following linear approximation of the SETAR(2; $p$ ) process

$$X_t^{(2)} = \varepsilon_t + \sum_{j=1}^\infty g_j^{(2)} \varepsilon_{t-j}. \tag{10}$$

□

Further note that Proposition 1 gives the best (in  $L^2$  norm) one-step-ahead predictor for the SETAR(2; $p$ ) process, whereas the linear process in (10) is a projection of the SETAR(2; $p$ ) predictor in the space of the linear predictors. Many features of this linear process will be analysed in the following; the extension to the SETAR(2; $p_1, p_2$ ) case, with  $p_1 \neq p_2$ , is discussed in Sect. 4.2.

First of all, we provide an example to illustrate the linear approximation in (10).

**Example 1** Let  $X_t$  be a stationary and ergodic SETAR(2;1) process given by

$$X_t = \phi_1 X_{t-1} \mathbb{1}(X_{t-1} \leq 0) + \phi_2 X_{t-1} (1 - \mathbb{1}(X_{t-1} \leq 0)) + \varepsilon_t.$$

To simplify the computations, we suppose that the transition matrix  $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , which implies that  $\pi = 1/2$  and that the two parameters  $\phi_1$  and  $\phi_2$  are negative. Moreover, the matrix  $\mathbf{K}$ , defined in (8), is

$$\mathbf{K} = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} \mathbf{P} = \begin{pmatrix} 0 & \phi_1 \\ \phi_2 & 0 \end{pmatrix}.$$

By (9) and (27), we have that

$$g_1^{(2)} = \frac{1}{2}(1; 1) \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad \text{and} \quad g_2^{(2)} = \frac{1}{2}(1; 1) \mathbf{K} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \phi_1 \phi_2.$$

Set  $j = 2u$  and  $j = 2u - 1$  for even and odd cases of  $j$ , respectively. Since

$$\mathbf{K}^{2u} = (\phi_1 \phi_2)^u \mathbf{I}_2 \quad \text{and} \quad \mathbf{K}^{2u-1} = (\phi_1 \phi_2)^{u-1} \mathbf{K},$$

it follows that

$$g_{2u}^{(2)} = (\phi_1 \phi_2)^u \quad \text{and} \quad g_{2u-1}^{(2)} = (\phi_1 \phi_2)^{u-1} g_1^{(2)} \quad \forall u \geq 1.$$

Finally, note that the coefficients  $g_j^{(2)}$  derived through Proposition 1 and Remark 1 exhibit the same ergodic conditions given in Petruccielli and Woolford (1984). ■

The next theorem shows a more precise characterization of the linear approximation (10) of the SETAR(2; $p$ ) process.

**Theorem 1** Let  $X_t, t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2;p) process with  $E(\varepsilon_t^2) < \infty$  and  $\gamma_2 < 1$ ; then the linear approximation (10) is the ARMA(2p, 2p) process.

**Proof** See the Appendix.

What is stated in Theorem 1 allows us to make some interesting remarks.

**Remark 2** The assumptions in Theorem 1,  $E(\varepsilon_t^2) < \infty$  and  $\gamma_2 < 1$ , are sufficient to have a finite second moment,  $E(X_t^2) < \infty$  (see Lemma 2 in the Appendix). Further,  $\gamma_r$  given by  $\gamma_r = \pi \|\Phi_1\|^r + (1 - \pi) \|\Phi_2\|^r$  is always less than one when the two regimes are also stationary, with  $\|\Phi_1\| < 1$  and  $\|\Phi_2\| < 1$ ; in the other cases, however, this condition needs to be verified case by case.  $\square$

**Remark 3** Theorem 1 provides the linear approximation of the SETAR(2;p) process, given by the ARMA(2p,2p) model. Looking at the proof of Theorem 1 in the Appendix (Eq. (30)) we have the exact expression of the linear process  $X_t^{(2)}$  in (10) that even allows to theoretically derive the coefficients of the ARMA(2p,2p) model. For simplicity, suppose that the matrix  $\mathbf{K}$ , in (8), has different eigenvalues - say  $\lambda_1, \dots, \lambda_{2p}$ . By (30) and after some easy but long algebra, we have two polynomials of degree 2p related to the autoregressive and moving average component of  $X_t^{(2)}$  whose coefficients are given respectively by:

$$\begin{aligned}
 a_i &= (-1)^{i-1} \sum_{|\mathbf{v}|=i} \prod_{u=1}^{2p} \lambda_u^{v_u}, \quad i = 1, \dots, 2p, \\
 b_j &= -a_j + (-1)^{j-1} \sum_{k=1}^{2p} c_k \sum_{\substack{|\mathbf{v}_{-k}|=j-1 \\ u \neq k}} \prod_{u=1}^{2p} \lambda_u^{v_u}, \quad j = 1, \dots, 2p,
 \end{aligned} \tag{11}$$

with  $|\mathbf{v}| = \sum_{u=1}^{2p} v_u, |\mathbf{v}_{-k}| = \sum_{\substack{u=1 \\ u \neq k}}^{2p} v_u$  for  $v_u \in \{0, 1\}$  and where the sums in (11)

are made with respect to all different  $i$  and  $j - 1$  groups of ones in the vectors  $\mathbf{v}$  and  $\mathbf{v}_{-k}$ , respectively.  $\square$

To emphasize the issues in Theorem 1 and illustrate how the coefficients (11) can be used, consider the following example.

**Example 2** Let  $X_t$  be a stationary and ergodic SETAR(2;1) process with the matrices of the coefficients  $\Phi_1 = \{\phi_1\}$  and  $\Phi_2 = \{\phi_2\}$ .

The matrix  $\mathbf{K}$  in (8) becomes

$$\mathbf{K} = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}.$$

For simplicity, suppose that the matrix  $\mathbf{K}$  has two different eigenvalues - say  $\lambda_1$  and

$\lambda_2$ . Write  $\mathbf{K} = \mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Gamma}^{-1}$  and set  $\mathbf{P}_1^T = \boldsymbol{\pi}^T\mathbf{\Gamma}$  and  $\mathbf{P}_2 = \mathbf{\Gamma}^{-1} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$  as in the proof of Theorem 1. Let  $c_k, k = 1, 2$ , be the component-wise product of the elements in the vectors  $\mathbf{P}_1$  and  $\mathbf{P}_2$ . By Theorem 1, we have that the linear approximation of  $X_t$  is given by an ARMA(2,2) model - that is,

$$X_t^{(2)} - a_1X_{t-1}^{(2)} - a_2X_{t-2}^{(2)} = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}, \tag{12}$$

where  $a_1 = \lambda_1 + \lambda_2$ ,  $a_2 = -\lambda_1\lambda_2$ ,  $b_1 = -a_1 + c_1 + c_2$ , and  $b_2 = -(a_2 + c_1\lambda_2 + c_2\lambda_1)$  by using (11) in Remark 3. ■

Furthermore note that from Remark 1 it follows that the coefficients of the approximation (10) are based on the expectation (9) that makes  $X_t^{(2)}$  a *suboptimal* linear approximation, in the sense that it is not the best linear one-step-ahead predictor of  $X_t$  (see Proposition 1). Moreover, it has the great advantage of establishing a clear correspondence between the ARMA and SETAR orders, as stated in Theorem 1. This correspondence is the key element that overshadows the need to evaluate how  $X_t^{(2)}$  approximates  $X_t$ , which becomes only a theoretical issue lacking any empirical application in identifying the SETAR model (as discussed in Sect. 3).

Another key remark needs to be introduced to discuss the order of the ARMA approximation.

**Remark 4** The order  $2p$  of the ARMA approximation is the maximum value of the AR and MA components, and this correspondence between the order of the ARMA approximation and the order of the autoregressive regimes of the SETAR process could be used to estimate the autoregressive order of the threshold process (usually left to information criteria). □

To emphasize this last remark and to give empirical evidence that  $2p$  is the maximum order of the ARMA approximation (where  $2p$  could be greater than the “true” ARMA order), consider the following example.

**Example 3** Let  $X_t \sim \text{SETAR}(2;1)$  with coefficients  $\phi_1^{(1)} = \phi_1^{(2)} = \phi_1$ , with  $|\phi_1| < 1$ . It is easy to note that this last equality implies the degeneration of the SETAR(2;1) model to an AR(1) structure, and that this degeneration needs to be found even in (10). This can be verified because, in this case, the transpose of the vector  $\boldsymbol{\pi}$  in (7) becomes (1, 0) (or (0, 1)), whereas  $\mathbf{P} = \mathbf{I}_2$  and then  $\mathbf{K} = \phi_1\mathbf{I}_2$ . From (27) it follows that  $g_j^{(2)} = \phi_1^j$  for  $j = 1, 2, \dots$ , and so:

$$X_t^{(2)} = \varepsilon_t + \sum_{j=1}^{\infty} \phi_1^j \varepsilon_{t-j},$$

which is the MA( $\infty$ ) representation of the AR(1) process. ■



The result given in Example 3 leads to an important evaluation: when  $X_t$  is a linear autoregressive process then  $X_t^{(2)}$  is identically equal to  $X_t$  and is no longer an approximation.

Another important consequence of the result of Theorem 1 and Remark 4 is the fact that we can build a SETAR(2; $p$ ) process whose linear approximation is null. This is important for two reasons. First we can subtract from  $X_t$  the linear approximation, and then we can mainly focus on the structure of the residuals. Second, by a proper parametrization, we can find a SETAR(2; $p$ ) process whose linear approximation (10) is null.

This last evaluation is summarized in the following Corollary.

**Corollary 1** *Let  $X_t, t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2; $p$ ) process. Under the assumptions of Theorem 1, there exists a SETAR(2; $p$ ) process such that its linear approximation (10) is null - that is  $g_j^{(2)} = 0, \forall j \geq 1$ .*

**Proof** See the Appendix.

To illustrate the result of this Corollary consider the following example.

**Example 4** Let  $X_t$  be an ergodic SETAR(2;1) process. As shown in the proof of Corollary 1, a sufficient condition to obtain a SETAR(2;  $p$ ) process with null linear approximation, is given by  $\phi_1 = -\phi_2$ , with  $\pi_{11} = \pi_{22} = 0.5$ . To give empirical evidence of this result, we generated  $n = 200$  artificial data from a SETAR(2;1) model with autoregressive coefficients  $\phi_1 = -0.58$  and  $\phi_2 = 0.58$ . The correlogram on the left side of Fig. 1 clearly shows the absence of the linear component whereas the correlogram of  $X_t^2$  on the right side, gives evidence of nonlinearity in the generating process. ■

Corollary 1 also plays an important role in the order estimation presented in the next section, because if we have a SETAR(2; $p$ ) model with  $p > 0$ , it can happen that the linear approximation in Theorem 1 is a white noise, and then the  $2p$  order of the ARMA process will not match that of the SETAR(2; $p$ ) model (as discussed in Remark 4). This implies that in this case, the results of Theorem 1 cannot be used in the order estimation. To manage this problem and broadly use the results of

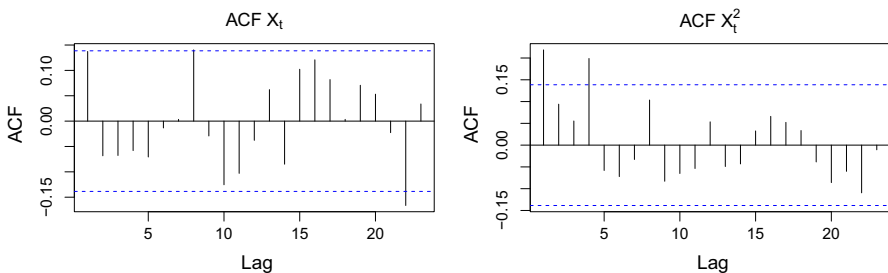


Fig. 1 Plot of the autocorrelation function of  $X_t$  (left) and  $X_t^2$  (right)

Theorem 1 to estimate the autoregressive order of the threshold model, some additional conditions need to be added (see Theorem 2).

In the following Remark we give a bound for the residual part, coming out from the proposed linear approximation,  $X_t^{(2)}$ , in  $L^2$  norm.

**Remark 5** First, for simplicity, suppose that  $X_t^{(2)} = \varepsilon_t$ , that is the linear approximation is null. Otherwise, we can always consider the process  $X_t - X_t^{(2)}$ . So, we have to evaluate the following quantity

$$\|X_t - \varepsilon_t\|_{L^2}^2. \quad (13)$$

By using the representation of the SETAR process as in Lemma 1, if  $\gamma_4 < 1$  the main upper bound (in the sense that we take into account the series of the squared terms) of (13) is

$$C \mathbf{e}_1^T \boldsymbol{\pi}^T (\mathbf{I}_{2p} - \mathbf{K}_2)^{-1} \begin{pmatrix} \boldsymbol{\Phi}_1^2 \\ \boldsymbol{\Phi}_2^2 \end{pmatrix} \mathbf{e}_1,$$

for some positive constant  $C$ ,  $\mathbf{K}_2 = \begin{pmatrix} \boldsymbol{\Phi}_1^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_2^2 \end{pmatrix} (\mathbf{P} \otimes \mathbf{I}_p)$  whereas the other quantities correspond to those defined in (8). This result follows by using the same arguments as in the proof of Theorem 1.  $\blacksquare$

Then, by focusing on the proposed linear approximation, the threshold autoregressive order can be consistently estimated, thus restricting attention to the linear time series domain. As will be discussed in Sect. 3, this has remarkable advantages not only from a theoretical point of view but even empirically, because the computational burden of the order estimation procedure of a nonlinear time series model is restricted to its linear (and less complex) approximation.

### 3 Order estimation

Identify the time series models is a crucial step in the ‘iterative stages in the selection of a model’ (Box and Jenkins 1976) that needs to preserve the parsimony of the model and makes a heavy impact on the computational effort needed to estimate the parameters.

In the linear domain, and in particular in the ARMA context, the identification has well-established results based on the relation between the ARMA parameters and the total/partial autocorrelation (at different lags) of the generating process.

As expected, these results - which relate strictly to the linear dependence - cannot be extended to the nonlinear domain. The complexity of the nonlinear time series structures has led to the model selection and identification largely discussed in the literature, which often focuses on information criteria and their performance (see Psaradakis et al. 2009; Emiliano et al. 2014; Rinke and Sibbertsen 2016). If we focus on order estimation in the SETAR domain, we find it looks different: Kapetanios (2001) proposes a consistent information criterion by which to estimate

the lag order of the autoregressive regimes; Wong and Li (1998) introduce a correction to the Akaike Information Criterion (AIC; Akaike 1974) to correct its bias in the presence of SETAR models; and De Gooijer (2001) proposes cross-validation criteria to select the autoregressive order of these nonlinear structures. Galeano and Peña (2007), modifying the model selection criteria introduced by Hurvich et al. (1990), propose the inclusion of a determinant term related to the estimated parameters of each regime; furthermore, more recently, Fenga and Politis (2015) evaluated a bootstrapped version of the AIC in the SETAR domain, and defined the procedure step by step.

The results given in Sect. 2 could be used in this context to introduce a new approach to estimating the autoregressive order of the threshold regimes. For simplicity, assume that both regimes have the same order. (As noted before, this is not a limitation because zeroes could be included in the vector of parameters.) The proposed procedure is based on the linear approximation of the SETAR model as given in Theorem 1. We call this procedure *Linear AIC* (L-AIC) and it is based on the two main steps that are discussed below and summarized with the pseudo-code in Algorithm 1.

Let  $X_t$  be a stationary SETAR(2;  $p$ ) process. The first step starts by fixing the maximum order for  $p$  ( $p_{max}$ ) and then for each  $p = 1, 2, \dots, p_{max}$ :

- (1.a) estimate the parameters of the SETAR(2;  $p$ ) model;
- (1.b) compute the eigenvalues of the matrix  $\mathbf{K}$  in (8)
- (1.c) compute the estimates of the autoregressive parameters - say  $\hat{a}_j$  - for  $j = 1, 2, \dots, 2p$ , of the linear approximation

$$X_t^{(2)} = a_1 X_{t-1}^{(2)} + \dots + a_{2p} X_{t-2p}^{(2)} + \varepsilon_t, \tag{14}$$

using the results in (11).

Given the  $p_{max}$ -estimated SETAR models, the second step is based on a parametric bootstrap approach with  $B$  bootstrap replications. In particular, for  $b = 1, \dots, B$ :

- (2.a) generate the bootstrap independent innovations  $\{\varepsilon_t^*\}$ ,  $t = 1, \dots, n + 2p_{max}$ , from a random variable with mean 0, variance  $\sigma^2$ , and with  $n$  the time series length;

for  $i = 1, \dots, p_{max}$  run the steps 2.b) and 2.c):

- (2.b) generate the artificial time series from the AR( $2i$ ) models by using the innovations  $\{\varepsilon_t^*\}$  in 2.a) and the coefficients estimated in 1.c). Then we have  $X_t^{*(2)}$  as

$$X_t^{*(2)} = \hat{a}_1 X_{t-1}^{*(2)} + \dots + \hat{a}_{2i} X_{t-2i}^{*(2)} + \varepsilon_t^*; \tag{15}$$

- (2.c) for each artificial time series  $i$ , select the autoregressive order  $\hat{p}_b^{(i)}$  such that  $\hat{p}_b^{(i)} = \arg \min_{j \in \{1, \dots, 2i\}} AIC(j)$ , with  $j$  the order of the autoregressive process fitted

to  $X_t^{*(2)}$ , whose coefficients are obtained through the Yule-Walker estimators;

- (2.d) using the  $\hat{p}_b^{(i)}$  obtained from steps 2.b) and 2.c), estimate the autoregressive order of the SETAR(2;  $\hat{p}_b$ ) model such that:

$$\hat{p}_b = \max\{i : |\hat{p}_b^{(i)} - \hat{p}_b^{(i+1)}| \neq 0\}, \quad \text{for } i = 1, 2, \dots, p_{max} - 1.$$

Given the  $B$  selected orders (one for each bootstrap replication), the SETAR model has autoregressive order  $\hat{p}$  such that

$$\hat{p} = \arg \max_{\hat{p}_b \in \{1, \dots, p_{max}\}} \{\#\hat{p}_b\}, \quad (16)$$

with  $\#\hat{p}_b$  being the empirical frequency of  $\hat{p}_b$ .

Note that: in step 1.b), given the ergodicity of the process  $X_t$ , the probabilities included in the matrix  $\mathbf{P}$  of (8) can be consistently estimated using the empirical frequencies  $n_{ij} = \sum_{t=2}^n \mathbb{1}(X_t \in \mathbb{R}_i | X_{t-1} \in \mathbb{R}_j)$ , for  $i, j = 1, 2$ , such that  $\hat{\pi}_{ij} = n_{ij} / \sum_{j=1}^2 n_{ij}$  (see among others Anderson and Goodman 1957); in step 1.c) and all the elements included in the second step of the procedure, only the linear AR model is involved; this is because, given the results of Theorem 1 and Remark 3, the AR and MA components have the same order  $2p$ . Moreover, the AR part of the ARMA approximation shares all the elements used in order estimation. Therefore, to simplify the procedure, we can only choose the AR component. (The extension to SETAR models with different autoregressive order is discussed in Sect. 4.2.)

Further, from the computational point of view, the use of theoretical issues related to linear models (instead of to nonlinear ones) in the second step, makes the algorithm quite fast.

---

**Algorithm 1** L-AIC algorithm

---

- 1: Let  $X_t$  be a stationary SETAR(2;  $p$ ) process:
  - 2: **Input:**  $X_t \sim \text{SETAR}(2;p)$ , with  $p$  unknown; maximum order for  $p$ ,  $p_{max}$ ; bootstrap replicates  $B$ .
  - 3: **Output:** estimate of  $p$ :  $\hat{p}$ .
  - 4: **procedure** L-AIC( $X_t, p_{max}, B$ )
  - 5: **STEP 1**
  - 6:     **for**  $p = 1, \dots, p_{max}$  **do**
  - 7:         estimate  $\phi_i^{(1)}, \phi_i^{(2)}$ , for  $i = 1, \dots, p$  of model (3);
  - 8:         compute the eigenvalues of  $\mathbf{K}$ ;
  - 9:         compute the estimates  $\hat{a}_i$  by (11),  $i = 1, \dots, 2p$ ;
  - 10:     **end for**
  - 11: **STEP 2**
  - 12:     **for**  $b = 1, \dots, B$  **do**
  - 13:         generate the bootstrap innovations  $\{\varepsilon_t^*\}$ ;
  - 14:         **for**  $i = 1, \dots, p_{max}$  **do**
  - 15:             generate artificial time series  $X_t^{*(2)}$  from (15) using the coefficients obtained in step 9;
  - 16:             select the autoregressive order of  $X_t^{*(2)}, \hat{p}_b^i$ ;
  - 17:         **end for**
  - 18:         given all the  $\hat{p}_b^i$ 's of step 16, estimate the order  $p_b$  of the SETAR(2;  $p_b$ ) model:
 
$$\hat{p}_b = \max\{i : |\hat{p}_b^{(i)} - \hat{p}_b^{(i+1)}| \neq 0\}, \quad \text{for } i = 1, 2, \dots, p_{max} - 1;$$
  - 19:     **end for**
  - 20:     **return**  $\hat{p} = \arg \max_{\hat{p}_b \in \{1, \dots, p_{max}\}} \{\#\hat{p}_b\}$ .
  - 21: **end procedure**
- 

This procedure has two important aspects. First, the linear approximation process,  $X_t^{(2)}$ , is not observable, and so it is a sort of latent process. For this reason,

in step 2.a) we generate for each  $p$  a sequence of i.i.d. innovations (for example, from the standard Gaussian distribution) and we estimate the parameters of the linear approximation by using (11). In this way, we build the process in (15) that is always a well-defined AR(2i) for each  $i$ ; this is the key point for using the selection rule in step 2.d).

Moreover, our procedure is motivated by two main considerations. First, when we estimate a linear model in the SETAR(2;p) data generating process, the residuals are not a sequence of i.i.d. random variables, because they also capture the nonlinear structure of the SETAR process. Second, even if we could estimate the parameters of the linear approximation by using the results in Francq and Zakoïan (1998), it is not easy to derive the variance-covariance matrix of the estimators, in which case, the results (11) should be preferred.

Before stating the next results about the consistency of the L-AIC procedure, we need to report some regularity assumptions that refer to the Assumptions of Theorem 1 of Chan (1993).

*Assumptions (H1).* Let  $X_t$  be a nondegenerate SETAR(2;p) process where the coefficients of the two regimes are not equal and  $0 < \pi < 1$ . Moreover,

- (1) the process  $X_t$  is stationary and ergodic (see Chan and Tong 1985; Bec et al. 2004);
- (2) there exists the univariate density function of  $X_t$  with respect to its invariant distribution, which is positive everywhere.  $\square$

Finally, starting with the AIC behaviour in Shibata (1976), we define an asymptotically *type-AIC* consistent procedure if the asymptotic distribution of the chosen order is

$$\lim_{n \rightarrow \infty} Pr(\hat{p} = p) = 0 \quad \text{if } p < p_0;$$

$$\lim_{n \rightarrow \infty} Pr(\hat{p} = p) > 0 \quad \text{if } p \geq p_0;$$

with  $p_0$  the true order and  $\hat{p}$  the estimator of the autoregressive order that depends on the series length  $n$ .

**Theorem 2** Let  $X_t, t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2;p<sub>0</sub>), defined in (2). Under Assumptions (H1) and if  $\pi_{11} \neq \pi$ ,  $E(\varepsilon_t^2) < \infty$  and  $\gamma_2 < 1$ , then  $\hat{p}$  in (16) is type-AIC consistent.

**Proof** See the Appendix.

Note that the assumption  $\pi_{11} \neq \pi$  in Theorem 2 is relevant because it guarantees that the autoregressive process in Theorem 1 is exactly of order  $2p$  (see Remark 4) and then what stated in Corollary 1 is prevented.

The proposed procedure and its consistency were empirically evaluated in a Monte Carlo study (see Sect. 5) that considers various SETAR models characterized by various degrees of complexity and nonlinearity.

## 4 Extensions

In this section we examine the proposed linear approximation of the SETAR process and evaluate three different aspects. First, we show the role of the threshold and delay parameters in this linear approximation. Second, we study the case where regimes have different orders. Third, we generalize the result of Theorem 1 when the number of regimes is more than two.

### 4.1 Threshold and delay parameters

To write the linear approximation for any values of the threshold and delay parameters, we consider the results in Sect. 2 presented for the case of two regimes with the same order. For simplicity, we consider only the AR part of the linear approximation. So, the matrix  $\mathbf{K}$  in (8) becomes

$$\mathbf{K}_{r,d} = \begin{pmatrix} \Phi_1 & \mathbf{0} \\ \mathbf{0} & \Phi_2 \end{pmatrix} (\mathbf{P}_{r,d} \otimes \mathbf{I}_p),$$

where the threshold and delay parameters are  $r$  and  $d$ , respectively. The matrix  $\mathbf{P}_{r,d}$  is defined as

$$\mathbf{P}_{r,d} = \begin{pmatrix} \pi_{11,r}^{(d)} & \pi_{12,r}^{(d)} \\ \pi_{21,r}^{(d)} & \pi_{22,r}^{(d)} \end{pmatrix},$$

with  $\pi_{ij,r}^{(d)} = Pr(X_t \in \mathbb{R}_j | X_{t-d} \in \mathbb{R}_i)$ ,  $i, j = 1, 2$  and  $\mathbb{R}_1 = (-\infty, r]$ ,  $\mathbb{R}_2 = (r, +\infty)$ . All the other quantities are the same as in (8).

First, note that the delay parameter has the role of the order for the transition probabilities in  $\mathbf{P}_{r,d}$ . It follows that  $\mathbf{P}_{r,d} = (\mathbf{P}_{r,1})^d$ . Then, the matrix  $\mathbf{K}_{r,d}$  can be written as

$$\mathbf{K}_{r,d} = \begin{pmatrix} \Phi_1 & \mathbf{0} \\ \mathbf{0} & \Phi_2 \end{pmatrix} \left( (\mathbf{P}_{r,1})^d \otimes \mathbf{I}_p \right).$$

Now, looking at the form of the matrix  $\mathbf{K}_{r,d}$ , we can note that:

1. the parameters  $r$  and  $d$  involve only the probabilities in the matrix  $\mathbf{P}_{r,d}$ ;
2. in general, assuming that the SETAR model does not degenerate into an AR process, the rank of the matrix  $\mathbf{K}_{r,d}$  depends only on the matrices of the coefficients  $\Phi_1$  and  $\Phi_2$ ;
3. the procedure for the order estimation (proposed in Sect. 3) needs only evaluate the difference between two successive orders. Then, it mainly depends on the matrices of the coefficients and not on the matrix  $\mathbf{P}_{r,d}$ .

By using the previous considerations, we can argue that our procedure for the order estimation is independent of the threshold and delay parameters. This means that in

identifying the SETAR model, we can separate the estimation of the threshold and delay parameters from the order estimation of the regimes.

#### 4.2 Different orders in the regimes

Suppose that we have a SETAR model with two regimes with two different autoregressive orders - say  $p_1$  and  $p_2$ . Moreover, suppose that the threshold and delay parameters are zero and one, respectively. Set  $p_0 = \max\{p_1, p_2\}$ . Let  $p_1 < p_2$  and  $p_2 = p_1 + m$ . Then, the matrix  $\mathbf{K}$  is the same as in (8), except that matrix  $\Phi_1$  becomes:

$$\Phi_1 = \begin{pmatrix} \phi_1^{(1)} & \cdots & \phi_{p_1}^{(1)} & \mathbf{0}_{m-1} & 0 \\ & & \mathbf{I}_{p_0-1} & & \mathbf{0} \end{pmatrix}.$$

In this case,  $\mathbf{K}$  is a  $2p_0$  matrix with  $a_{2p_0} = a_{2p_0-1} = \dots = a_{2p_0-m+1} = 0$  by (11) and using the fact that the matrix  $\mathbf{K}$  has  $m$  eigenvalues equal to zero. This means that the AR part of the linear approximation is  $AR(p_1 + p_2)$ .

Note that the transition probabilities in the matrix  $\mathbf{P}$  are well defined in the sense that the irreducibility property of the Markov Chain with respect to the two regimes is still valid.

In this case, the procedure for the order estimation (see Sect. 3) is applied properly, changing Algorithm 1. The first step is modified by fixing a grid of candidate values for  $p_1$  and  $p_2$  such that  $p_1 = 1, \dots, p_{1,\max}$  and  $p_2 = 1, \dots, p_{2,\max}$ . In the second step, the  $p_{1,\max} \times p_{2,\max}$  estimated SETAR models are used to carry out the bootstrap replications, with  $p_{1,\max} + p_{2,\max}$  innovations in row 13. A double cycle given by  $i = 1, \dots, p_{1,\max}$  and  $s = 1, \dots, p_{2,\max}$  is considered in row 14, whereas  $i$  is replaced by  $i + s$  in row 16. All these changes imply that the maximum in row 17 becomes:

$$(\hat{p}_{b_1}, \hat{p}_{b_2}) = \max\{i, s : |\hat{p}_b^{(i,s)} - \hat{p}_b^{(i_*,s_*)}| \neq 0\}$$

for  $i = 1, 2, \dots, p_{1,\max} - 1$ ,  $s = 1, 2, \dots, p_{2,\max} - 1$  and with  $i_* \geq i$ ,  $s_* \geq s$ , such that at least  $i_*$  or  $s_*$  is strictly greater than the corresponding  $i$  or  $s$  value. Finally, the computation of the empirical frequencies of  $(\hat{p}_{b_1}, \hat{p}_{b_2})$  allows to conclude the algorithm.

#### 4.3 More regimes

Suppose that we have  $k$  regimes with  $k \geq 2$ . For simplicity, assume that all regimes have the same order - say  $p$ . Moreover, suppose that the delay is  $d = 1$  and the thresholds are  $r_i$ , for  $i = 1, \dots, k - 1$ , with  $-\infty < r_1 < \dots < r_{k-1} < \infty$ . Then, the quantities in (7), become



$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_k \end{pmatrix} \otimes \mathbf{I}_p \quad \text{and} \quad \mathbf{P} = \begin{pmatrix} \pi_{11} & \dots & \pi_{1k} \\ \vdots & \vdots & \vdots \\ \pi_{k1} & \dots & \pi_{kk} \end{pmatrix},$$

with  $\sum_{i=1}^k \pi_i = 1$ . The matrix  $\mathbf{K}$  in (8) becomes

$$\mathbf{K} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Phi}_k \end{pmatrix} (\mathbf{P} \otimes \mathbf{I}_p),$$

where  $\boldsymbol{\Phi}_i, i = 1, \dots, k$  are the matrices of coefficients of each regime.

Repeating the proof of Theorem 1, we get the expression in (30) as the sum of  $k \cdot p$  quantities instead of  $2p$ . This implies that we have as a linear approximation the ARMA( $k \cdot p, k \cdot p$ ) process. Finally, in this case the procedure for the order estimation can be applied by replacing  $2p$  with  $k \cdot p$  in Algorithm 1.

### 5 Simulation study

To evaluate the L-AIC procedure for the order estimation of the SETAR(2;  $p$ ) process, we generated time series from four data generating processes:

- $M1 : X_t = -0.8X_{t-1}\mathbb{1}_{r-1} - 0.2X_{t-1}(1 - \mathbb{1}_{r-1}) + \varepsilon_t$
- $M2 : X_t = 0.5X_{t-1}\mathbb{1}_{r-1} - 0.5X_{t-1}(1 - \mathbb{1}_{r-1}) + \varepsilon_t$
- $M3 : X_t = (-0.4X_{t-1} - 0.5X_{t-2})\mathbb{1}_{r-1} + (0.2X_{t-1} - 0.6X_{t-2})(1 - \mathbb{1}_{r-1}) + \varepsilon_t$
- $M4 : X_t = (-0.7X_{t-1} - 0.2X_{t-2})\mathbb{1}_{r-1} + (0.3X_{t-1} - 0.6X_{t-2})(1 - \mathbb{1}_{r-1}) + \varepsilon_t$

with threshold value  $r_1 = 0$ , and  $\varepsilon_t \sim N(0, 1)$ . For each model we considered time series of length  $n = 50, 75, 150, 200, 500$  (with burn-in at 500 to discharge the effects of the initial conditions). The models  $M1$  and  $M2$  are those considered in Galeano and Peña (2007), where even for model  $M1$  the condition of Theorem 2,  $\pi \neq \pi_{11}$ , is satisfied; Models  $M3$  and  $M4$  were chosen to guarantee the assumptions of Theorem 1 and change the percentage of observations generated from each regime. (In  $M3$  this percentage is almost identical in the two regimes, whereas in  $M4$  the percentage of observations generated from the first regime is, on average, less than 40%.) Further, note that not all models include the intercepts: it makes it less easy to distinguishing between the two regimes, and the performance of the order estimation procedure could be affected from it.

After fixing  $p_{max} = 5$ , we estimated the SETAR(2; $p$ ) models, for  $p = 1, \dots, 5$  using conditional least squares estimators, with  $d = 1$  and  $r_1 = 0$ . For each estimated model we obtained the linear autoregressive approximation  $\hat{X}_t^{*(2)}$  and then we started, for each approximation, the  $B = 125$  bootstrap replicates with  $\varepsilon_t \sim N(0, 1)$ . The number of Monte Carlo runs is 1000.

In the Monte Carlo study the L-AIC procedure has been compared to two other approaches.

The SETAR-AIC (Tsay 1989):

$$AIC(p_1, p_2) = \sum_{j=1}^2 \left( T_j \ln(\hat{\sigma}_{\epsilon_j}^2) + 2p_j \right), \quad (17)$$

where  $T_j$ ,  $p_j$  and  $\hat{\sigma}_{\epsilon_j}^2$  are the number of observations, the candidate autoregressive order, and the residual variance of regime  $j$ , for  $j = 1, 2$ , respectively, whereas the second approach considered for the order estimation of the SETAR(2;  $p$ ) model is the  $\beta$ AIC of Fenga and Politis (2015), which is a bootstrapped version of the SETAR-AIC.

Comparisons of the L-AIC with both the SETAR-AIC and the  $\beta$ AIC are presented in the left side of Table 1 where the relative empirical frequency of the selection of the true autoregressive orders of the SETAR(2;  $p$ ) model are summarized. (The complement to one of the relative frequencies in the table is related to overparametrized models.)

It is widely known that the AIC tends to overfit models (see among others Koehler and Murphree 1988), whereas this overparametrization is penalized by the Bayesian Information Criterion (BIC), so we extended our procedure to the BIC domain giving rise to the *Linear BIC* (L-BIC) procedure.

For this purpose we considered steps 1.a) - 2.d) of the L-AIC procedure and replaced in step 2.c) the AIC with the BIC. Then we replicated the simulation study and compared the performance of the SETAR-BIC (Galeano and Peña 2007):

**Table 1** Relative empirical frequency of right selection over 1000 Monte Carlo runs.

$n$	M1: SETAR(2;1)					M2: SETAR(2;1)				
	L-AIC	AIC	$\beta$ AIC	L-BIC	BIC	L-AIC	AIC	$\beta$ AIC	L-BIC	BIC
50	0.70	0.45	0.68	0.81	0.73	0.75	0.45	0.73	0.84	0.70
75	0.71	0.49	0.70	0.86	0.79	0.76	0.49	0.73	0.86	0.79
150	0.74	0.53	0.71	0.89	0.89	0.82	0.56	0.75	0.93	0.90
200	0.81	0.56	0.73	0.92	0.93	0.81	0.60	0.78	0.96	0.92
500	0.84	0.62	0.77	0.95	0.98	0.86	0.67	0.81	0.98	0.98
$n$	M3 : SETAR(2;2)					M4 : SETAR(2;2)				
	L-AIC	AIC	$\beta$ AIC	L-BIC	BIC	L-AIC	AIC	$\beta$ AIC	L-BIC	BIC
50	0.71	0.54	0.71	0.76	0.69	0.74	0.48	0.71	0.78	0.68
75	0.75	0.59	0.72	0.82	0.83	0.83	0.56	0.81	0.86	0.79
150	0.80	0.56	0.78	0.88	0.87	0.92	0.65	0.83	0.97	0.94
200	0.83	0.63	0.80	0.94	0.91	0.92	0.69	0.85	0.98	0.96
500	0.90	0.72	0.85	0.97	0.98	0.96	0.75	0.88	0.99	0.99

L-AIC: Linear AIC; AIC: SETAR-AIC;  $\beta$ AIC: Fenga and Politis 2015; L-BIC: Linear BIC; BIC: SETAR-BIC.  $n = \{50, 75, 150, 200, 500\}$  is the time series length

$$BIC(p_1, p_2) = \sum_{j=1}^2 \left( T_j \ln(\hat{\sigma}_{\epsilon_j}^2) + \log(T_j) p_j \right) \tag{18}$$

to that of the L-BIC procedure. (Note that Fenga and Politis 2015 do not consider the BIC.) The results of each model are presented in the right side of Table 1.

Finally note that the L-BIC procedure is made consistent by using the same arguments as in the proof of Theorem 2, together with the consistency of the BIC measure in the linear domain.

The results in Table 1 clearly show the improvement in the L-AIC procedure with respect to the SETAR-AIC and the  $\beta$ AIC approaches in all considered cases and correspondingly the good performance of the L-BIC if compared to the BIC criterion, mainly for small values of  $n$ . In particular, one can be note that when the distinction between the two regimes is more marked, as in model  $M2$  where the autoregressive coefficients have opposite signs, there is an overall improvement in the competing order estimation approaches (mainly in the L-AIC case).

As the complexity of the data generating process grows with models  $M3$  and  $M4$ , all procedures confirm their pertinence to estimating the autoregressive order of the nonlinear threshold processes, but even in this case the frequency with which the L-AIC procedure correctly detects is generally higher than that of the competing approaches based on the Akaike Criterion and this result is confirmed in the L-BIC case, even if with less marked differences between the L-BIC and the SETAR-BIC.

From the computational point of view, if we compare the L-AIC procedure with the Fenga and Politis (2015) criterion, we can note that, when  $r_1, d$  are known and  $p_1 = p_2 = p$ , the effort (measured in terms of computing time) is heavier for the  $\beta$ AIC criterion; this is because in the bootstrap iterations, for each candidate  $p = 1, 2, \dots, p_{max}$ , we need to estimate a nonlinear threshold model instead of a linear  $AR(2p)$  model, as in the L-AIC case. In practice, a further advantage of the L-AIC procedure is that the computationally intensive steps of the bootstrap iterations are confined to the linear domain characterized by reduced complexity relative to the nonlinear domain.

Finally, to evaluate how the variability of the SETAR(2;  $p$ ) process is explained by the linear ARMA( $2p, 2p$ ) approximation, we have compared the variance of the SETAR generating process with the variance of its linear ARMA approximation. For all models,  $M1, M2, M3$  and  $M4$ , we have generated 1000 artificial time series and for each of them we have computed the ratio between the two variances. In Table 2 the means of these ratios and the corresponding standard deviations, are presented for  $n = \{200, 500\}$ . The variance explained by the ARMA approximation is higher for model  $M1$  whereas this ratio is lower for model  $M4$  that is characterized by high nonlinearity. The artificial time series have been further evaluated to investigate the dependence between  $X_t^{(2)}$  and the generating process  $X_t$ . For this aim, in Table 2 we have considered, for each model, the mean of the correlations computed between the ARMA approximation  $X_t^{(2)}$  and  $Y_t = X_t - X_t^{(2)}$ . The results show different correlations as the structure of the model changes.

**Table 2** Mean values, over 1000 Monte Carlo replicates, of the ratios between the variance of the ARMA(2*p*, 2*p*) approximation,  $X_t^{(2)}$ , and the variance of the corresponding generating process  $X_t \sim \text{SETAR}(2; p)$ ; mean values of the correlation between  $X_t^{(2)}$  and  $Y_t = X_t - X_t^{(2)}$ . In parenthesis the standard deviations

n	Index	M1	M2	M3	M4
200	$\frac{\text{var}(X_t^{(2)})}{\text{var}(X_t)}$	0.939 (0.044)	0.917 (0.045)	0.910 (0.056)	0.838 (0.078)
	$\text{cor}(X_t^{(2)}, Y_t)$	-0.354 (0.043)	-0.086 (0.054)	0.052 (0.079)	-0.010 (0.077)
500	$\frac{\text{var}(X_t^{(2)})}{\text{var}(X_t)}$	0.957 (0.027)	0.919 (0.031)	0.934 (0.041)	0.835 (0.057)
	$\text{cor}(X_t^{(2)}, Y_t)$	-0.356 (0.027)	-0.087 (0.035)	0.034 (0.063)	-0.004 (0.055)

## 6 Conclusion

It is widely known that when the behaviour of the autocorrelation function (ACF) is observed for a time series  $X_t$  generated by a SETAR(2; *p*) model, we can easily confuse the ACF of the threshold process with that of a linear ARMA structure. Different motivations can be cited for this empirical evaluation: among them is the fact that the linear autoregressive model is nested in the SETAR model because it can be seen as a degeneration of the SETAR structure when all data belong to the same regime.

In this study we investigated the relation between the SETAR and ARMA models. We showed that when  $X_t \sim \text{SETAR}(2; p)$ , under proper conditions, the linear approximation (10) of  $X_t$  is  $X_t^{(2)} \sim \text{ARMA}(2p, 2p)$ . We theoretically showed this result and even clarified when it could not be applied and how it can be generalized when the number of regimes is greater than two or the regimes have a different autoregressive order. Further, using linear approximation, we proposed an order estimation procedure called Linear-AIC to estimate the autoregressive order of the two regimes of  $X_t$ ; the consistency of this procedure was proved and the extension of these results to the BIC domain is discussed.

The L-AIC procedure is based on two main steps: in the first step we focus on the nonlinear generating process  $X_t$  and on its linear approximation  $X_t^{(2)}$ , whereas in the second step we estimate the autoregressive order of the SETAR model using a parametric bootstrap approach.

Our Monte Carlo study gives evidence of the performance of the L-AIC procedure and of its BIC extension (called L-BIC). The results highlight that the L-AIC generally does a better job than both the competing SETAR-AIC in (17) and Fenga and Politis (2015) criterion (and, correspondingly, the L-BIC performs better than the SETAR-BIC criterion), even in the presence of the parametrization of the SETAR model that makes it difficult to distinguish among regimes.

The results presented herein will serve as the topic of future research that further investigates the extensions discussed in Sect. 4; future research can also study how

the L-AIC and the L-BIC can be included in an overall identification procedure for SETAR models.

### Appendix

We report the proof of all Theorems, Proposition and Corollary of the paper and two technical Lemmas.

Before proofing Theorem 1, we need to introduce the following Lemma on the iterative representation of the SETAR model:

**Lemma 1** Let  $\mathbf{X}_t, t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2;p) process (4), with  $E[|\varepsilon_t|] < \infty, \gamma = \pi\|\Phi_1\| + (1 - \pi)\|\Phi_2\|$  and  $E[\mathbb{1}\{X_{t-1} \leq 0\}] = \pi$  (with  $0 \leq \pi \leq 1$ ), if  $\gamma < 1$  then  $\mathbf{X}_t$  can be given by:

$$\mathbf{X}_t = \sum_{i=1}^n \prod_{s=0}^{i-1} \mathbf{A}_{I_{t-1-s}} \varepsilon_{t-i} + \varepsilon_t, \quad a.s., \quad \text{for } n \rightarrow \infty, \tag{19}$$

where  $\mathbf{A}_{I_{t-1-s}}$  is defined in (5) and  $\|\cdot\|$  is any matrix norm.

**Proof** Using Theorem 1.1 in Bougerol and Picard (1992) and let

$$c_k = \frac{1}{k} E \left( \log \left\| \prod_{s=0}^{k-1} \mathbf{A}_{I_{t-1-s}} \right\| \right),$$

it is sufficient to show that  $c_k < 0$ , for any  $k \in \mathbb{N}$  and  $t \in \mathbb{Z}$ .

Let  $\|\cdot\|$  any matrix norm, now we have that

$$c_k \leq \log \left( E \left( \left\| \prod_{s=0}^{k-1} \mathbf{A}_{I_{t-1-s}} \right\|^{1/k} \right) \right) \leq \log \left( E \left( \frac{1}{k} \sum_{s=0}^{k-1} \|\mathbf{A}_{I_{t-1-s}}\| \right) \right) = \log(\gamma) < 0.$$

Given the ergodicity and stationarity of  $X_t$  and considered the representation (5), then  $E(\|\mathbf{A}_{I_{t-1-s}}\|) = \gamma$ , for any  $t$  and  $s$ . Noting that the assumptions of the present Lemma fit all conditions of Theorem 1.1 of Bougerol and Picard (1992), we have that  $\gamma < 1$ , for any  $k$ , and that this result holds for any matrix norm.  $\square$

The next lemma proves the existence of moments of the SETAR process.

**Lemma 2** Let  $X_t, t \in \mathbb{Z}$ , be a stationary and ergodic SETAR(2;p) process with  $E[|\varepsilon_t|^r] < \infty$ , and  $\gamma_r < 1$ , where  $\gamma_r = \pi\|\Phi_1\|^r + (1 - \pi)\|\Phi_2\|^r$ , for  $r \in [1, \infty)$  and  $E[\mathbb{1}\{X_{t-1} \leq 0\}] = \pi$  (with  $0 \leq \pi \leq 1$ ), then  $E[|X_t|^r] < \infty$  (where  $\|\cdot\|$  is any matrix norm).

**Proof** Let  $(\|\mathbf{X}_t\|)_{L^r} = E(\|\mathbf{X}_t\|)^{1/r}$ , be the  $L^r$  norm, as given in Stelzer (2009, Definition 4.1). We state with  $\|\cdot\|$  a generic matrix or vector norm on  $\mathbb{R}^p$ , since the result of this lemma holds for any matrix or vector norm. By Lemma 1 and using the same arguments given in the first part of the proof of Theorem 4.1 in Stelzer (2009), that can be applied even in this SETAR domain, we need to prove the sufficient condition that  $\mathbf{X}_t$  belongs to  $L^r(\mathbb{R}^p)$ , that is given by:

$$(E\|\boldsymbol{\varepsilon}_t\|^r)^{1/r} + \sum_{i=1}^{\infty} \left[ E \left\| \left( \prod_{s=0}^{i-1} \mathbf{A}_{t-1-s} \right) \boldsymbol{\varepsilon}_{t-i} \right\|^r \right]^{1/r} < \infty. \tag{20}$$

Define

$$Q_i = E \left\| \left( \prod_{s=0}^{i-1} \mathbf{A}_{t-1-s} \right) \boldsymbol{\varepsilon}_{t-i} \right\|^r, \quad c_{i+1} = \frac{1}{i+1} E \left( \log \left\| \left( \prod_{s=0}^{i-1} \mathbf{A}_{t-1-s} \right) \boldsymbol{\varepsilon}_{t-i} \right\|^r \right).$$

By using the same arguments as in the proof of Lemma 1, we have that

$$\begin{aligned} c_{i+1} &\leq \log \left( E \left\| \left( \prod_{s=0}^{i-1} \mathbf{A}_{t-1-s} \right) \boldsymbol{\varepsilon}_{t-i} \right\|^{r/(i+1)} \right) \leq \\ &\leq \log \left( E \left( \frac{1}{i+1} \sum_{s=0}^{i-1} \|\mathbf{A}_{t-1-s}\|^r + \frac{1}{i+1} \|\boldsymbol{\varepsilon}_{t-i}\|^r \right) \right) = \log(\gamma_r) \quad \text{as } i \rightarrow \infty, \end{aligned} \tag{21}$$

since  $E\|\mathbf{A}_{t-1-s}\|^r = \gamma_r$  by stationarity and ergodicity of  $X_t$  and  $E|\boldsymbol{\varepsilon}_t|^r < \infty$ . Then, by (21)

$$\exp((i+1)c_{i+1}) \leq Q_i \quad \text{and} \quad \exp((i+1)c_{i+1}) \leq (\gamma_r + \eta_{i+1})^{i+1}.$$

where  $\eta_{i+1} = O\left(\frac{1}{i+1}\right)$ .

Since  $\gamma_r < 1$  and  $Q_i = O(\gamma_r^i)$ , the result in (20) follows. Thus, we have that  $E[\|\mathbf{X}_t\|^r] < \infty$ . The proof is complete. □

**Proof of Proposition 1** Using the results of Lemma 1, we need to minimize the following quantity

$$\begin{aligned} Q(g_1, g_2, \dots; \mathcal{I}_{t-1}) &= \|\mathbf{X}_t - X_t^L | \mathcal{I}_{t-1}\|_{L^2}^2 = \\ &= E \left[ \left( \sum_{j=1}^{\infty} \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 - g_j \right) \boldsymbol{\varepsilon}_{t-j} \right)^2 \middle| \mathcal{I}_{t-1} \right]. \end{aligned} \tag{22}$$

Since by Lemma 2, the second moment exists, that is  $E(X_t^2) < \infty$ , then  $Q(g_1, g_2, \dots; \mathcal{I}_{t-1})$  is well defined.

Making the first partial derivatives of  $Q(g_1, g_2, \dots; \mathcal{I}_{t-1})$  with respect to  $g_k$ , with  $k \geq 1$ , and putting each equation equal to zero, we have

$$\frac{\partial Q}{\partial g_k} = E \left[ \left( \sum_{j=1}^{\infty} \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 - g_j \right) \varepsilon_{t-j} \varepsilon_{t-k} \right) \middle| \mathcal{I}_{t-1} \right] = 0, \quad \forall k \geq 1. \tag{23}$$

Since the process  $X_t$  is also measurable with respect to  $\tilde{\mathcal{I}}_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ , then it is equivalent to consider either  $\tilde{\mathcal{I}}_{t-1}$  in place of  $\mathcal{I}_{t-1}$  in the conditional means. Now, it is sufficient to derive  $g_j, \forall j \geq 1$ , in the following equations.

$$E \left[ \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 - g_j \right) \varepsilon_{t-j} \varepsilon_{t-k} \middle| \tilde{\mathcal{I}}_{t-1} \right] = 0 \quad \forall j, k \geq 1. \tag{24}$$

Without loss of generality, we can consider both  $\varepsilon_{t-j}$  and  $\varepsilon_{t-k}$  different from zero. So, we can write (24) as

$$\varepsilon_{t-j} \varepsilon_{t-k} E \left[ \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 - g_j \right) \middle| \tilde{\mathcal{I}}_{t-1} \right] = 0, \tag{25}$$

It follows that the solution in (25) is

$$g_j(\tilde{\mathcal{I}}_{t-1}) = g_j(\mathcal{I}_{t-1}) = E \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 \middle| \mathcal{I}_{t-1} \right) = \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1. \tag{26}$$

Since (22) is a square function, if we replace the solution (26) in (22), then the latter is zero. This means that we have a minimum. Then, the proof of the first part of Proposition is complete.

Now, for the second part, we know that the optimal one step ahead  $L^2$  predictor of  $X_t$  is  $X_t(1) = E(X_t | \mathcal{I}_{t-1})$ . By using again Lemma 1, we have that

$$\begin{aligned} X_t(1) &= E \left[ \varepsilon_t + \sum_{j=1}^{\infty} \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 \varepsilon_{t-j} \right) \middle| \mathcal{I}_{t-1} \right] = \\ &= \sum_{j=1}^{\infty} \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 \varepsilon_{t-j} = \sum_{j=1}^{\infty} g_j(\mathcal{I}_{t-1}) \varepsilon_{t-j}. \end{aligned}$$

The proof is so complete. □

Now we can prove Theorem 1.

**Proof of Theorem 1** We start from the linear process (10) with coefficients,  $g_j^{(2)} = E \left( \mathbf{e}_1^T \prod_{s=0}^{j-1} \mathbf{A}_{t-1-s} \mathbf{e}_1 \right), j \geq 1$ , defined in (9).

By using the stationarity and ergodicity assumption on  $X_t$ , it follows that also the indicator process,  $\mathbb{I}_t$ , is stationary and ergodic. Moreover, we can apply, on the expression for  $g_j^{(2)}$ , the Markov property and arguments similar to those given in

Timmermann (2000) on the recursive representation and generation of moments of the regimes switching models. So, it follows that

$$g_j^{(2)} = \mathbf{e}_1^T \boldsymbol{\pi}^T \mathbf{K}^{j-1} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1, \quad j \geq 1, \quad (27)$$

where  $\boldsymbol{\pi}$  and  $\mathbf{K}$  are defined in (7) and (8), respectively and  $\mathbf{K}^{j-1} = \mathbf{K} \cdot \mathbf{K} \cdot \dots \cdot \mathbf{K}$ , with the matrix  $\mathbf{K}$  multiplied  $(j-1)$  times. Then, we can write the linear process in (10) as

$$\begin{aligned} X_t^{(2)} &= \varepsilon_t + \mathbf{e}_1^T \boldsymbol{\pi}^T \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1 \varepsilon_{t-1} + \dots + \mathbf{e}_1^T \boldsymbol{\pi}^T \mathbf{K}^{j-1} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1 \varepsilon_{t-j} + \dots = \\ &= \mathbf{e}_1^T \left[ \mathbf{I}_p + \boldsymbol{\pi}^T \left( \sum_{j=1}^{\infty} B^j \mathbf{K}^{j-1} \right) \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \right] \mathbf{e}_1 \varepsilon_t, \end{aligned} \quad (28)$$

where  $B$  is the back-shift operator such that  $B^j \varepsilon_t = \varepsilon_{t-j}$ . For now, suppose that all the eigenvalues of the matrix  $\mathbf{K}$  are different. So, we can write  $\mathbf{K} = \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^{-1}$ , where  $\boldsymbol{\Lambda}$  is a diagonal matrix with the eigenvalues on the main diagonal and  $\boldsymbol{\Gamma}$  is the matrix with the corresponding eigenvectors. Therefore, we can write (28) as

$$X_t^{(2)} = \mathbf{e}_1^T \left[ \mathbf{I}_p + \boldsymbol{\pi}^T \boldsymbol{\Gamma} \left( \sum_{j=1}^{\infty} B^j \boldsymbol{\Lambda}^{j-1} \right) \boldsymbol{\Gamma}^{-1} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \right] \mathbf{e}_1 \varepsilon_t. \quad (29)$$

By Remark 1, it follows that the series  $\sum_{j=1}^{\infty} |g_j^{(2)}|$  is convergent. Note that in (29) we have a power series and so it follows that  $\rho(\mathbf{K}) < 1$ , with  $\rho(\mathbf{K})$  the maximum absolute eigenvalue of  $\mathbf{K}$ .

Set  $\mathbf{P}_1^T = \mathbf{e}_1^T \boldsymbol{\pi}^T \boldsymbol{\Gamma}$  and  $\mathbf{P}_2 = \boldsymbol{\Gamma}^{-1} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1$  and let  $c_k$  be the component-wise product of the corresponding elements in the vectors  $\mathbf{P}_1$  and  $\mathbf{P}_2$ ,  $k = 1, \dots, 2p$ . Now, we can write (29) as

$$X_t^{(2)} = \left( 1 + B \sum_{k=1}^{2p} \frac{c_k}{1 - \lambda_k B} \right) \varepsilon_t, \quad (30)$$

where  $\lambda_k$ ,  $k = 1, \dots, 2p$ , are the eigenvalues in the matrix  $\boldsymbol{\Lambda}$ . So, we can conclude, that:

$$X_t^{(2)} \sim ARMA(2p, 2p).$$

To complete the proof, we need to consider the case when two or more eigenvalues in the matrix  $\mathbf{K}$  are equal. In this case, we can apply the Jordan matrix decomposition of  $\mathbf{K}$  (among the others, see Lütkepohl (1996)) and the previous results still hold. The proof is so complete.  $\square$



Corollary 1 states that the linear approximation of the SETAR model could be null. We here give the proof.

**Proof of Corollary 1** Starting from (27) in the proof of Theorem 2, we have that

$$g_j^{(2)} = \mathbf{e}_1^T \boldsymbol{\pi}^T \mathbf{K}^{j-1} \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1, \quad j \geq 1,$$

and so it is sufficient to find one case such that  $g_j^{(2)} = 0, \forall j \geq 1$ . For this aim, suppose that the transition probability matrix (7) is

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Then  $\pi = 1/2$ .

Further, suppose that the parameters in the first regime of the SETAR(2;p) model are the same, in absolute value, to the corresponding parameters of the second regime but with the opposite sign. It is easy to verify that  $g_1^{(2)} =$

$$2^{-1} \mathbf{e}_1^T (\mathbf{I}_p, \mathbf{I}_p) \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1 = 0.$$

Now, by using the above assumptions, we can write  $g_j^{(2)}$  as

$$\begin{aligned} g_j^{(2)} &= 2^{-j} \mathbf{e}_1^T \left[ (\boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2)^{j-1}, (\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2)^{j-1} \right] \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} \mathbf{e}_1 \\ &= 2^{-j} \mathbf{e}_1^T [(\boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2)^j] \mathbf{e}_1, \end{aligned} \tag{31}$$

$\forall j \geq 1$ . Since the elements in the first row of the matrix  $\boldsymbol{\Phi}_1$  have opposite sign with respect to the elements in the first row of the matrix  $\boldsymbol{\Phi}_2$ , we can write  $\boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 = 2\mathbf{C}$ , where the matrix  $\mathbf{C}$  has all zeros in the first row and the same elements of  $\boldsymbol{\Phi}_1$  or  $\boldsymbol{\Phi}_2$  in the other rows. It is clear that the matrix  $\mathbf{C}$  is idempotent and so  $\mathbf{C}^j = \mathbf{C}, \forall j \geq 1$ . Then, we can write (31) as

$$g_j^{(2)} = 2^{-j} \mathbf{e}_1^T (2\mathbf{C})^j \mathbf{e}_1 = \mathbf{e}_1^T \mathbf{C}^j \mathbf{e}_1 = \mathbf{e}_1^T \mathbf{C} \mathbf{e}_1 = g_1^{(2)} = 0.$$

So, the result follows. □

We prove here the consistency of the identification procedure proposed.

**Proof of Theorem 2** Before to enter into the details of the proof, we need to discuss the assumptions of Theorem 2. For simplicity, suppose that the eigenvalues of the matrix  $\mathbf{K}$ , defined in (8), are all different. The assumption  $\pi_{11} \neq \pi$  guarantees that the linear approximation,  $X_t^{(2)}$ , shows, at least, one coefficient different from zero (in fact, by Corollary 1, we could also have that  $X_t^{(2)} = \varepsilon_t$ ).

In particular, from (11), it follows that  $a_{2p_0} = \prod_{k=1}^{2p_0} \lambda_k$ , with  $\lambda_k$ ,  $k = 1, \dots, 2p_0$ , the eigenvalues of  $\mathbf{K}$ . Then  $a_{2p_0}$  is the determinant of  $\mathbf{K}$  and it is different from zero, if and only if all  $\lambda_k$  are different from zero. Further,  $|\mathbf{K}| = \phi_{p_0}^{(1)} \phi_{p_0}^{(2)} (\pi_{11}\pi_{22} - \pi_{12}\pi_{21})$ , and then, given the generating process SETAR(2,  $p_0$ ) where  $\phi_{p_0}^{(1)} \neq 0$  and  $\phi_{p_0}^{(2)} \neq 0$ ,  $|\mathbf{K}|$  is different from zero if and only if  $\pi_{11}\pi_{22} - \pi_{12}\pi_{21} \neq 0$ , or in other words, if  $\pi_{11}\pi_{22} - (1 - \pi_{11})(1 - \pi_{22}) \neq 0$  (where the term on the left is the determinant of  $\mathbf{P}$  in (8)). Using the ergodicity of the indicator process  $\mathbb{1}_{t-1}$  such that  $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{P}$ , it follows that the determinant of  $\mathbf{P}$  (and of  $\mathbf{K}$ ) is different from zero if and only if  $\pi_{11} \neq \pi$ . Then, we have that  $a_{2p_0} \neq 0$  iff  $\pi_{11} \neq \pi$ .

The remaining assumptions of this Theorem imply that the Conditional Least Squares estimators of the parameters for the SETAR(2; $p_0$ ) process, are consistent by Theorem 1 of Chan (1993). In order to apply this Theorem we need that  $E(X_t^2) < \infty$  which follows by the assumptions  $\gamma_2 < 1$  and  $E(\varepsilon_t^2) < \infty$  (see Lemma 2).

Since the eigenvalues of  $\mathbf{K}$  are all different and by using the Dini's Theorem, it follows that the parameters  $a_i$ ,  $i = 1, \dots, 2p_0$  (given in (11)) are differentiable functions of the SETAR(2; $p_0$ ) parameters. So, they are also continuous functions of the SETAR(2; $p_0$ ) parameters. Then, we have that

$$\hat{a}_i - a_i = o_p(1), \quad (32)$$

where  $i = 1, \dots, 2p_0$ .

Now we focus the attention on the consistency of  $\hat{p}$ . Consider a  $\tilde{p}$  such that  $1 \leq \tilde{p} \leq p_{max}$ , where clearly  $\tilde{p}$  can be different from  $p_0$ . Moreover, (32) holds when  $\tilde{p}$  is not equal to  $p_0$ .

The second step of our procedure is based on a parametric bootstrap. In fact, we use the latter to estimate some quantities on the latent linear process  $X_t^{(2)}$  defined in (14). We build its bootstrap approximation as  $X_t^{*(2)}$  which is defined in (15). There are two main differences between  $X_t^{(2)}$  and  $X_t^{*(2)}$ . Noting that we can fix any distribution function such that the random variables (drawn from it) have zero mean and variance  $\sigma^2$ , the first difference is that the random variables  $\varepsilon_t^*$  are not the same as  $\varepsilon_t$  even if they are independently drawn from the same distribution. The second difference is that  $X_t^{*(2)}$  includes the estimates  $\hat{a}_i$  instead of the true values  $a_i$ , for  $i = 1, \dots, 2\tilde{p}$ , as in  $X_t^{(2)}$ . Now, we evaluate these two differences by using the Mallow's metric of order 2 (see Bickel and Freedman 1981), say  $d_2(F_1, F_2)$ , where  $F_1$  and  $F_2$  are two distributions. By Lemma 8.4 of Bickel and Freedman (1981), we have that

$$d_2(F_{\varepsilon,n}, F_\varepsilon) \rightarrow 0 \quad a.s. \quad (33)$$

where  $F_{\varepsilon,n}$  and  $F_\varepsilon$  are the empirical distribution and the true distribution function of  $\varepsilon_t$ , respectively. Since we use a parametric bootstrap, then  $\{\varepsilon_t^*\}$  and  $\{\varepsilon_t\}$  are two different samples from the same distribution function. So, by (33) and using the triangular property of the Mallow's metric, it follows that

$$d_2(F_{\varepsilon^*,n}, F_{\varepsilon,n}) \rightarrow 0 \quad a.s. \tag{34}$$

with  $F_{\varepsilon^*,n}$  the empirical distribution function of  $\varepsilon_t^*$ .

By Remark 1, the linear process  $X_t^{(2)}$  is stationary and ergodic. Further by ( ), also the linear process  $X_t^{*(2)}$  is stationary and ergodic with probability tending to one when  $n \rightarrow \infty$ . Then the process  $X_t^{*(2)} - X_t^{(2)}$  is still stationary and ergodic with probability tending to one when  $n \rightarrow \infty$ , since it is the difference of two (possibly different) AR( $2\tilde{p}$ ) process which are both stationary and ergodic, at least, in probability for large  $n$ .

Now, we evaluate the differences between the two linear processes  $X_t^{*(2)}$  and  $X_t^{(2)}$  by using  $d_2(F_n^*, F_n)$  where  $F_n^*$  and  $F_n$  are two univariate empirical distributions with respect to  $X_t^{*(2)}$  and  $X_t^{(2)}$ , respectively. We show that  $d_2(F_n^*, F_n) \xrightarrow{p} 0$ . In order to get this result, it is sufficient that

$$\frac{1}{n - 2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_t^{*(2)} - X_t^{(2)} \right)^2 \xrightarrow{p} 0. \tag{35}$$

Since in our procedure both  $X_t^{*(2)}$  and  $X_t^{(2)}$  are two (possibly different) AR( $2\tilde{p}$ ) processes, we can write

$$X_t^{*(2)} - X_t^{(2)} = \sum_{i=1}^{2\tilde{p}} \hat{a}_i \left( X_{t-i}^{*(2)} - X_{t-i}^{(2)} \right) + \sum_{i=1}^{2\tilde{p}} (\hat{a}_i - a_i) X_{t-i}^{(2)} + \varepsilon_t^* - \varepsilon_t,$$

where  $\hat{a}_i$  and  $a_i$ ,  $i = 1, \dots, 2\tilde{p}$ , are in (15) and (14), respectively.

By (32), it follows that

$$\sum_{i=1}^{2\tilde{p}} (\hat{a}_i - a_i)^2 \frac{1}{n - 2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_t^{(2)} \right)^2 \xrightarrow{p} 0, \tag{36}$$

since  $\frac{1}{n-2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_t^{(2)} \right)^2 \xrightarrow{p} \gamma_X(0) < \infty$ , which is the variance of  $X_t^{(2)}$ . So, by (34) and (36), we can write

$$\begin{aligned} \frac{1}{n - 2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_t^{*(2)} - X_t^{(2)} \right)^2 &= \sum_{i=1}^{2\tilde{p}} \hat{a}_i \frac{1}{n - 2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_{t-i}^{*(2)} - X_{t-i}^{(2)} \right)^2 + \\ &+ \sum_{i \neq j}^{2\tilde{p}} \hat{a}_i \hat{a}_j \frac{1}{n - 2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_{t-i}^{*(2)} - X_{t-i}^{(2)} \right) \left( X_{t-j}^{*(2)} - X_{t-j}^{(2)} \right) + o_p(1). \end{aligned} \tag{37}$$

Since  $X_t^{*(2)} - X_t^{(2)}$  is a stationary and ergodic process with finite variance in probability, for large  $n$ , then  $\frac{1}{n-2\tilde{p}} \sum_{t=2\tilde{p}+1}^n \left( X_{t-i}^{*(2)} - X_{t-i}^{(2)} \right)^2 \xrightarrow{p} Q_1$ , for each integer  $i \geq 0$ , with  $0 \leq Q_1 < \infty$ . For now, suppose that  $Q_1 > 0$ , then, by (32), (37) asymptotically becomes  $Q_1 [1 - h(a_1, \dots, a_{2\tilde{p}})] = 0$  with  $h(a_1, \dots, a_{2\tilde{p}})$  a continuous function and it is less than one because  $1 - h(a_1, \dots, a_{2\tilde{p}})$  is the same expression we get to derive

$\gamma_X(0)$ . So, we can conclude that  $Q_1$  cannot be positive. The unique solution is  $Q_1 = 0$ . Thus, (35) is shown.

By Bickel and Freedman (1981), this result implies that

$$\frac{1}{n} \sum_{t=\max(2\tilde{p}, i)+1}^n X_t^{*(2)} X_{t-i}^{*(2)} - \frac{1}{n} \sum_{t=\max(2\tilde{p}, i)+1}^n X_t^{(2)} X_{t-i}^{(2)} \xrightarrow{P} 0, \quad (38)$$

for each integer  $i \geq 0$ .

By Theorem 8.1.1 of Brockwell and Davis 1991, the Yule-Walker estimators of the parameters in  $X_t^{(2)}$ , say  $\tilde{a}_i$ ,  $i = 1, \dots, 2\tilde{p}$ , (we omit the order  $2\tilde{p}$ ) are consistent as well as  $S_{\varepsilon}^2$ , the estimator of  $\sigma^2$ , the variance of the innovation process. By (38) and using the continuous mapping Theorem,  $\tilde{a}_i^* - \tilde{a}_i \xrightarrow{P} 0$ ,  $i = 1, \dots, 2\tilde{p}$ , where  $\tilde{a}_i^*$ ,  $i = 1, \dots, 2\tilde{p}$ , are the Yule-Waker estimators of parameters in  $X_t^{*(2)}$ . Moreover,  $S_{\varepsilon}^{2*} - S_{\varepsilon}^2 \xrightarrow{P} 0$ , with  $S_{\varepsilon}^{2*}$  the estimator of  $\sigma^2$  on  $X_t^{*(2)}$ . Therefore, the AIC measure on the process  $X_t^{*(2)}$  is equal in probability, for large  $n$ , to the AIC measure on  $X_t^{(2)}$ . Then, in the following, we can consider the  $X_t^{(2)}$  process.

First, suppose an AR( $2\tilde{p}$ ) process for  $X_t^{(2)}$  with  $\tilde{p} < p_0$  and  $2\tilde{p}$  the maximum order of the AR process. By Shibata (1976), it follows that  $\lim_{n \rightarrow \infty} P(2\hat{p} = 2\tilde{p}) = 1$  but with  $\tilde{p} < p_0$ .

Now, consider an AR( $2\tilde{p}$ ) process for  $X_t^{(2)}$ , with  $\tilde{p} = p_0$  and  $2\tilde{p}$  the maximum order for the AR process. In this case we have that  $\lim_{n \rightarrow \infty} Pr(2\hat{p} = 2p_0) = 1$  by Shibata (1976). So, we have the estimate of  $\hat{p} = p_0$  by step 3.c in our procedure with a probability tending to one.

Finally, suppose an AR( $2\tilde{p}$ ) process for  $X_t^{(2)}$  with  $\tilde{p} > p_0$  with  $2\tilde{p}$  the maximum order of the AR process. In this case, it follows that  $a_i = 0$  for  $i = 2\tilde{p}, \dots, 2p_{max}$  and by Shibata (1976), we have that  $\lim_{n \rightarrow \infty} Pr(2\hat{p} = 2p_0) > 0$ .

Putting all together, we have that our procedure gives  $\lim_{n \rightarrow \infty} Pr(\hat{p} = \tilde{p}) = 0$  if  $\tilde{p} < p_0$  and  $\lim_{n \rightarrow \infty} Pr(\hat{p} = \tilde{p}) > 0$  if  $\tilde{p} \geq p_0$ . So,  $\hat{p}$  is asymptotically *type-AIC* consistent. The proof is complete.  $\square$

**Acknowledgements** The authors gratefully acknowledge the Editor, the Associate Editor and the two anonymous Referees for the insightful comments and constructive suggestions.

**Funding** Open access funding provided by Università degli Studi di Salerno within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly

from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Control* AC 19:716–723
- Anderson TW, Goodman LA (1957) Statistical inference about Markov chains. *Ann Math Stat* 28:89–110
- Bec F, Salem BM, Carrasco M (2004) Tests for unit roots versus threshold specification with an application to the PPP. *J Bus Econ Statist* 22:382–395
- Bickel PJ, Freedman DA (1981) Some asymptotic theory for the bootstrap. *Ann Stat* 3:1196–1217
- Bougerol P, Picard N (1992) Strict stationarity of generalized autoregressive processes. *Ann Probab* 20:1714–1730
- Box GEP, Jenkins GM (1976) *Time series analysis: forecasting and control*. Holden-Day, New York
- Brockwell PJ, Davis RA (1991) *Time series: theory and methods*. Springer, New York
- Chan KS (1993) Consistency and limiting of the least squares estimator of a threshold autoregressive model. *Ann Stat* 21:520–533
- Chan KS, Tong H (1985) On the use of Lyapunov functions for the ergodicity of stochastic difference equations. *Adv Appl Probab* 17:666–678
- De Gooijer J (2001) Cross validation criteria for SETAR model selection. *J Time Ser Anal* 22:267–281
- Emiliano PC, Vivanco MJF, de Menezes FS (2014) Information criteria: How do they behave in different models? *Comput Stat Data Anal* 69:141–153
- Fenga L, Politis DN (2015) Bootstrap order selection for SETAR models. *J Stat Comput Simul* 85:235–250
- Franco C, Zakoian JM (1998) Estimating linear representation of nonlinear processes. *J Statist Plan Inference* 68:145–165
- Galeano P, Peña D (2007) Improved model selection criteria for SETAR time series models. *J Statist Plan Inference* 137:2802–2814
- Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384
- Huang R, Xu F, Chen R (2009) General expression for linear and nonlinear time series models. *Front Mech Eng China* 4:15–24
- Hurvich CM, Shumway R, Tsai C (1990) Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika* 77:709–719
- Kapetanios G (2001) Model selection in threshold models. *J Time Ser Anal* 22:733–754
- Koehler AB, Murphree ES (1988) A comparison of the Akaike and Schwarz criteria for selecting model order. *J R Stat Soc Ser C* 37:187–195
- Kokoszka PS, Politis DN (2011) Nonlinearity of arch and stochastic volatility models and Bartlett's formula. *Probab Math Stat* 31:47–59
- Lütkepohl H (1996) *Handbook of matrices*. Wiley, Chichester
- Petrucelli JD, Woolford SW (1984) A Threshold AR(1) model. *J Appl Probab* 21:270–286
- Pham DT (1985) Bilinear Markovian representation and bilinear models. *Stochast Process Appl* 20:295–306
- Priestley MB (1988) *Non-linear and non-stationary time series analysis*. Academic Press, London
- Psaradakis Z, Sola M, Spagnolo F, Spagnolo N (2009) Selecting nonlinear time series models using information criteria. *J Time Ser Anal* 30:369–394
- Rinke S, Sibbertsen P (2016) Information criteria for nonlinear time series models. *Stud Nonlinear Dyn Econ* 20:325–341
- Schoukens J, Vaes M, Pintelon R (2016) Linear system identification in a nonlinear setting: nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation. *IEEE Control Syst Mag* 36:38–69

- Shibata R (1976) Selection of the order of an autoregressive order by Akaike information criterion. *Biometrika* 63:117–126
- Stelzer R (2009) On Markov-switching ARMA processes - stationarity, existence of moments and geometric ergodicity. *Economet Theor* 25:43–62
- Timmermann A (2000) Moments of Markov switching models. *J Econ* 96:75–111
- Tong H (1990) *Non-linear time series: a dynamical system approach*. Oxford University Press, New York
- Tong H, Lim KS (1980) Threshold Autoregression, limit cycles and cyclical data. *J R Stat Soc (B)* 42:245–292
- Tsay R (1989) Testing and modeling threshold autoregressive processes. *J Am Stat Assoc* 84:231–240
- Wong CS, Li WK (1998) A note on the corrected Akaike information criterion for threshold autoregressive models. *J Time Ser Anal* 19:113–124

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.