Annals of the Institute of Statistical Mathematics (2022) 74:685–711
https://doi.org/10.1007/s10463-021-00812-4

Fixed accuracy estimation of parameters in a threshold autoregressive model

Victor V. Konev[1] · Sergey E. Vorobeychikov[1]

Received: 26 February 2021 / Revised: 2 August 2021 / Accepted: 13 September 2021 /
Published online: 18 October 2021

Abstract

For parameters in a threshold autoregressive process, the paper proposes a sequential modification of the least squares estimates with a specific stopping rule for collecting the data for each parameter. In the case of normal residuals, these estimates are exactly normally distributed in a wide range of unknown parameters. On the base of these estimates, a fixed-size confidence ellipsoid covering true values of parameters with prescribed probability is constructed. In the i.i.d. case with unspecified error distributions, the sequential estimates are asymptotically normally distributed uniformly in parameters belonging to any compact set in the ergodicity parametric region. Small-sample behavior of the estimates is studied via simulation data.

Keywords TAR process · Sequential estimates · Fixed-size confidence ellipsoid

1 Introduction

The first step in the analysis of a time series is the selection of a suitable mathematical model for the data which allows one to obtain an understanding of the mechanism generating the series and make inference about the probabilistic mechanism of the underlying structure. Linear and nonlinear models are two main structures for statistical modeling. Linear models are usually more tractable for studies and interpretable, but their efficiency highly relies on the validity of the linearity assumption. The nonlinear models may be more flexible and more accurately capture the dependence between observations in some cases, but they often computationally challenging. In engineering applications related to automatic control, filtering, segmentation of signals, biomedical

✉ Sergey E. Vorobeychikov
sev@mail.tsu.ru

Victor V. Konev
vvkonev@mail.tsu.ru

[1] Institute of Applied Mathematics and Computer Science, Tomsk State University, Lenina str. 36, Tomsk 634050, Russia

signal processing, change-point detection, spectral analysis, finance, linear and nonlinear models are often completely specified by some stochastic difference and stochastic differential equations except for a finite number of unknown parameters. The classical methods of identifying unknown parameters, such as least-squares method, stochastic approximation, have been successfully developed for a great variety of linear models. As a rule, the estimation theory becomes more complicated and mostly asymptotic, in the case of processes with dependent values. For an autoregression process, the least squares estimate (LSE) of its parameters, based on fixed sample size, is nonlinear function of observations. Asymptotic properties of the LSE of the autoregressive parameters, such as strong consistency, asymptotic normality, have been proved, under general conditions, relatively by not so long ago (we refer the reader to Lai and Wei 1983 and references therein). In recent years, remarkable theoretical advancements in statistical inference for linear stochastic models are made by the development of sequential methods for processes with dependent observations. The idea of sequential analysis to sample until enough information is gathered about unknown parameters turned out to be fruitful as in the case of independent observations. As the sequential sampling methods developed for i.i.d. framework, sequential methods for dependent observations are applied either to compare asymptotic properties of estimates or to find a solution to the problem which is unsolvable within the estimation theory with fixed sample size. We cannot go to detail here and refer the reader to the literature (see, e.g., Lai and Siegmund 1983 for a first-order non-explosive autoregressive process proposed to use the special stopping rule based on the observed Fisher information in the least squares estimate to obtain the estimator which is asymptotically normal uniformly in parameter).

In this paper, we develop a sequential least square method for estimating parameters in a threshold autoregressive model of order one TAR(1), introduced by Tong (1978), obeying the equation

$$x_k = \theta_1 x_{k-1}^+ \;+\; \theta_2 x_{k-1}^- \;+\; \varepsilon_k, \quad k = 1, 2, \ldots, \tag{1}$$

where $\{\varepsilon_k\}$ are independent and identically distributed (i.i.d.) unobservable random errors (noise) with $E\varepsilon_k = 0$ and $0 < E\varepsilon_k^2 = \sigma^2 < \infty$;$\theta_1$ and θ_2 are unknown parameters; $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$;x_0 (not dependent on θ_1 and θ_2) is independent of $\{\varepsilon_k\}$. The parameters θ_1 and θ_2 are commonly estimated by the least squares estimates

$$\hat{\theta}_{1,n} = \frac{\sum_{k=1}^n x_k x_{k-1}^+}{\sum_{k=1}^n \left(x_{k-1}^+\right)^2}\,, \quad \hat{\theta}_{2,n} = \frac{\sum_{k=1}^n x_k x_{k-1}^-}{\sum_{k=1}^n \left(x_{k-1}^-\right)^2}. \tag{2}$$

Asymptotic properties of these estimates are well studied in the literature (see Chan 1993; Chigansky and Kutoyants 2013; Li and Ling 2012; Yau et al. 2015; Gao et al. 2013 for details and other references). Pham et al. (1991) have proved strong consistency of the least squares estimates under quite general conditions, for the first-order SETAR model with one threshold. Proposition 1 of their paper implies that estimates (2) for parameters in (1) are strongly consistent if and only if $\theta_1 \leq 1$ and $\theta_2 \leq 1$. In terms of the processes $\{x_k^+\}$ and $\{x_k^-\}$, $\hat{\theta}_{1,n}$ and $\hat{\theta}_{2,n}$ are strongly consistent if and only if simultaneously

$$P_\theta\left(\sum_{k\geq 1}\left(x_k^+\right)^2=+\infty\right)=1,\ \ P_\theta\left(\sum_{k\geq 1}\left(x_k^-\right)^2=+\infty\right)=1. \tag{3}$$

Petruccelli and Woolford (1984) have established that the process (1) is ergodic if and only if parameters $\theta=(\theta_1,\theta_2)$ satisfy

$$\theta_1<1,\ \ \theta_2<1,\ \ \theta_1\theta_2<1. \tag{4}$$

and they proved asymptotic normality of estimate (2), when the process (1) is ergodic.

In recent years, there has been growing an interest to developing sequential analysis approach to the problem of estimating parameters in a threshold TAR(1) model. Lee and Sriram (1999) have studied risk efficient sequential estimates of parameters in a threshold TAR(1) model. Keeping in mind a surprising uniform asymptotic normality result of Lai and Siegmund (1983) for AR(1) model, Sriram raised a theoretical conjecture about the limiting distribution of a sequential pivot quantity for a parameter vector in TAR(1) model and conducted an extensive numerical study that strongly suggests in favor of sequential procedure (see Sriram and Iaci 2014 for details).

The goal of this paper is to develop sequential point and sequential confidence region procedures for estimating parameters in a TAR(1) model with a given precision on the basis of the least squares method. The sequential point estimate for the vector $\theta=(\theta_1,\theta_2)$ in (1) is constructed by modifying the least squares estimates (2) and introducing specific stopping rule for each of them. We consider the estimation problem in non-asymptotic and asymptotic statements. First, we study case of normal residuals $\{\varepsilon_k\}$ in (1). We show that the proposed sequential estimates, under this assumption, have exactly normal joint distribution.

The present paper has two linked objectives. First, we address the problem of constructing sequential least squares estimates for TAR(1) model (1) with non-asymptotic normal joint distribution under the assumption that the errors $\{\varepsilon_n\}$ form a Gaussian white noise. The obtained results are applied to construction the fixed-size confidence ellipsoid for estimating parameters in a TAR(p) model. The second problem is to study asymptotic properties of these estimates for ergodic process (1) in the case when $\{\varepsilon_n\}$ is an i.i.d. sequence of random variables with unspecified distribution and to prove uniform asymptotic normality of estimates.

The paper is organized as follows. In Sect. 2, we construct sequential least squares estimates for TAR(1) model. The exact non-asymptotic normal distribution of these estimates has been derived under assumption that the errors in the model (1) are an i.i.d. sequence of normally distributed random variables (Theorem 1). In Sect. 3, we study the case of unspecified error distribution. The property of uniform asymptotic normality for sequential estimates is established. In Sect. 4, the proof of key theoretical result is presented. In Sect. 5, we construct sequential point and sequential confidence region procedures for estimating parameters in a TAR(p) model (40). In Sect. 6, the results of numerical simulations are given.

Proceeding from this result, the problem of constructing a fixed-size confidence ellipsoid with a given coverage probability has been solved in nonparametric statement (Propositions 1, 5).

2 Construction of sequential least squares estimates

In this section, it is assumed that the errors $\{\varepsilon_k\}$ are i.i.d. with the standard Gaussian distribution. We will construct sequential point estimates for θ_1 and θ_2 on the basis of estimates (2). For a fixed $h > 0$, we introduce two stopping rules

$$\begin{aligned}\tau_1(h) &= \inf\left\{ n \geq 1 \,:\, \sum_{k=1}^{n} \left(x_{k-1}^{+}\right)^2 \geq h \right\},\\ \tau_2(h) &= \inf\left\{ n \geq 1 \,:\, \sum_{k=1}^{n} \left(x_{k-1}^{-}\right)^2 \geq h \right\}.\end{aligned} \tag{5}$$

The parameter h defines the accuracy of the estimates (see Theorem 1).

Let $\tau(h) = \tau_1(h) \vee \tau_2(h)$ be the observation time. We denote sequential least squares estimates for $\theta = (\theta_1, \theta_2)$ as the vector

$$\hat{\theta}(h) = (\hat{\theta}_1(h), \hat{\theta}_2(h)) \tag{6}$$

with coordinates

$$\begin{aligned}\hat{\theta}_1(h) &= \frac{1}{h^{+}} \sum_{k=1}^{\tau_1(h)} \beta_{1,k} x_{k-1}^{+} x_k,\\ \hat{\theta}_2(h) &= \frac{1}{h^{-}} \sum_{k=1}^{\tau_2(h)} \beta_{2,k} x_{k-1}^{-} x_k\end{aligned} \tag{7}$$

where $\beta_{1,k}$ and $\beta_{2,k}$ are weight coefficients of the form

$$\beta_{i,k} = \begin{cases} 1 & \text{if } \; k < \tau_i(h),\\ \sqrt{\alpha_{i,\tau_i}} & \text{if } \; k = \tau_i(h),\\ 0 & \text{if } \; k > \tau_i(h); \; i = 1, 2. \end{cases} \tag{8}$$

Here, α_{1,τ_1} and α_{2,τ_2} are correction factors compensating the overshots in (5), uniquely defined by the equations

$$\begin{aligned}\sum_{k=1}^{\tau_1(h)-1} \left(x_{k-1}^{+}\right)^2 + \alpha_{1,\tau_1(h)} \left(x_{\tau_1(h)-1}^{+}\right)^2 &= h,\\ \sum_{k=1}^{\tau_2(h)-1} \left(x_{k-1}^{-}\right)^2 + \alpha_{2,\tau_2(h)} \left(x_{\tau_2(h)-1}^{-}\right)^2 &= h.\end{aligned} \tag{9}$$

We set

$$h^+ = \sum_{k=1}^{\tau_1(h)} \beta_{1,k} \left(x_{k-1}^+\right)^2, \\ h^- = \sum_{k=1}^{\tau_2(h)} \beta_{2,k} \left(x_{k-1}^-\right)^2. \tag{10}$$

It will be noted that $0 < \alpha_{1,\tau_1(h)} \le 1$ and $0 < \alpha_{2,\tau_2(h)} \le 1$ for any $h > 0$. Therefore, $h^+/h \ge 1$ and $h^-/h \ge 1$. It is clear that these rations are close to one provided that contribution of the last addends to the corresponding sums in (10) is small.

The following theorem gives joint distribution of the standardized deviations of estimates (7).

Theorem 1 *Define $\hat{\theta}(h)$, $h > 0$ by (6), (7) and $\tau_1(h), \tau_2(h)$ by (5). If $\{\varepsilon_k\}$ are i.i.d. normal random variables with mean 0 and variance 1 and are independent of x_0, then for any $\theta = (\theta_1, \theta_2)$, $\theta_1 \le 1$, $\theta_2 \le 1$ the standardized deviations*

$$\xi_1 = \frac{h^+}{\sqrt{h}}\left(\hat{\theta}_1(h) - \theta_1\right), \quad \xi_2 = \frac{h^-}{\sqrt{h}}\left(\hat{\theta}_2(h) - \theta_2\right) \tag{11}$$

have the standard two-dimensional normal distribution $N_2(0, I)$, I is the unit matrix of the size 2×2, that is

$$P_\theta\left(\xi_1 < x,\ \xi_2 < y\right) = \Phi(x)\Phi(y), \ \ -\infty < x < \infty, \ \ -\infty < y < \infty.$$

The proof of Theorem 1 is given in Appendix.

Corollary 1 *Under conditions of Theorem 1 for every $x \in R$*

$$P_\theta\left(\frac{(h^+)^2}{h}\left(\hat{\theta}_1(h) - \theta_1\right)^2 + \frac{(h^-)^2}{h}\left(\hat{\theta}_2(h) - \theta_2\right)^2 \le x \right) = P\left(\chi_2^2 \le x\right), \tag{12}$$

where χ_2^2 is Chi-square random variable with two degrees of freedom.

This result is a direct consequence of Theorem 1.

Equation (12) enables one to construct a fixed-size confidence region for unknown parameters (θ_1, θ_2) with a prescribed coverage probability. Consider a family of confidence regions of elliptic form

$$G(r) = \left\{ (t_1, t_2) : \left(\frac{h^+}{h}\right)^2 \left(\hat{\theta}_1(h) - t_1\right)^2 + \left(\frac{h^{-2}}{h}\right)\left(\hat{\theta}_2(h) - t_2\right)^2 \le r^2 \right\}. \tag{13}$$

Proposition 1 *Let $\{\varepsilon_k\}$ in (1) be an i.i.d. sequence of standard normal variables, $\varepsilon_k \sim N(0, 1)$, and $\hat{\theta}_1(h)$, $\hat{\theta}_2(h)$ be defined by (7). Then, for any given $r > 0$, $0 < \alpha < 1$, $\theta_1 \le 1$, $\theta_2 \le 1$*

$$P_\theta\{\theta \in G(r)\} = 1 - \alpha \tag{14}$$

for stopping rules (5) *with* $h = 2\log\frac{1}{\alpha}/r^2$.

Proof Let $r > 0$ and $0 < \alpha < 1$ be fixed. From (12), (13), it follows that confidence region (13) covers true values of parameters θ_1 and θ_2 with probability $1 - \alpha$ if h satisfies the equation

$$P_\theta(\theta \notin G(r)) = P_\theta\left(\frac{\chi_2^2}{h} > r^2\right) = \alpha.$$

By making use of Chi-square density of distribution with two degrees of freedom

$$P_{\chi_2^2}(z) = \frac{1}{2}e^{-z/2}$$

one gets

$$P_\theta\left(\frac{\chi_2^2}{h} > r^2\right) = e^{-\frac{hr^2}{2}} = \alpha. \tag{15}$$

Solving this equation with respect to h, one comes to the desired result. This completes the proof of Proposition 1. □

3 The case of unspecified error distribution

Henceforth, we assume that $\{\varepsilon_k\}$ in (1) is a sequence of i.i.d. random variables with zero mean, variance $0 < \sigma^2 < \infty$ and $E\varepsilon_1^4 < \infty$, each having a strictly positive density $f(\cdot)$ on $R = (-\infty, \infty)$. We set for simplicity $\sigma^2 = 1$. In this section, we will study asymptotic behavior of the joint distribution of sequential estimates (7) as parameter h in stopping rules (5) tends to infinity and establish the property of uniform asymptotic normality in some parametric region. We need some extension of the probabilistic result for martingales of Lai and Siegmund (1983) (Proposition 2.1).

Proposition 2 *Let* $x_n,\ \varepsilon_n,\ n = 0, 1 \ldots$ *be random variables adapted to the increasing sequence of* σ*–algebras* $\{\mathcal{F}_n\}_{n\ge 0}$. *Let* $\left\{P_\theta, \theta \in \Theta\right\}$ *be a family of probabilistic measures such that under every* P_θ

(1) $\varepsilon_1,\ \varepsilon_2, \ldots$ *are i.i.d. with* $E_\theta\varepsilon_1 = 0,\ E_\theta\varepsilon_1^2 = 1$;
(2) $\sup_\theta E_\theta(\varepsilon_1^2; |\varepsilon_1| > a) \to\ 0$ *as* $a \to \infty$;
(3) ε_n *is independent of* $\mathcal{F}_{n-1}$ *for each* $n \ge 1$;
(4) $\sup_\theta P_\theta(x_n^2 > a) \to\ 0$ *as* $a \to \infty$ *for each* $n \ge 1$;

(5) $P_\theta\left(\sum_{k\geq 1}\left(x_k^+\right)^2=+\infty\right)=1,\ \ P_\theta\left(\sum_{k\geq 1}\left(x_k^-\right)^2=+\infty\right)=1,$ *where* $x^+=\max(x,0)$, $x^-=\min(x,0)$;

(6) *for each* $\delta>0$

$$\lim_{n\to\infty}\sup_\theta P_\theta\left((x_n^+)^2>\delta\sum_{k=1}^n(x_{k-1}^+)^2 \quad \text{for some} \quad n\geq m\right)=0;$$

$$\lim_{n\to\infty}\sup_\theta P_\theta\left((x_n^-)^2>\delta\sum_{k=1}^n(x_{k-1}^-)^2 \quad \text{for some} \quad n\geq m\right)=0.$$

For $h>0$ let

$$\begin{aligned}\tau_1(h)&=\inf\left\{n\geq 1\ :\ \sum_{k=1}^n\left(x_{k-1}^+\right)^2\geq h\right\},\\ \tau_2(h)&=\inf\left\{n\geq 1\ :\ \sum_{k=1}^n\left(x_{k-1}^-\right)^2\geq h\right\},\end{aligned}\tag{16}$$

$\tau(h)=\tau_1(h)\vee\tau_2(h)$.

(7) *Let* α_{1,τ_1} *and* α_{2,τ_2} *be correction multipliers uniquely defined by the equations*

$$\sum_{k=1}^{\tau_1(h)-1}\left(x_{k-1}^+\right)^2+\alpha_{1,\tau_1(h)}\left(x_{\tau_1(h)-1}^+\right)^2=h,$$

$$\sum_{k=1}^{\tau_2(h)-1}\left(x_{k-1}^-\right)^2+\alpha_{2,\tau_2(h)}\left(x_{\tau_2(h)-1}^-\right)^2=h;$$

(8) *For each vector* $u=(u_1,u_2)$ *with real components and* $||u||=\sqrt{u_1^2+u_2^2}=1$, *let define the sequence of random variables*

$$y_{k-1}=\beta_{1,k}u_1x_{k-1}^+\ +\ \beta_{2,k}u_2x_{k-1}^-,\quad 1\leq k\leq\tau(h),$$

where

$$\beta_{i,k}=\begin{cases}1 & \text{if}\ \ k<\tau_i(h),\\ \sqrt{\alpha_{i,\tau_i}} & \text{if}\ \ k=\tau_i(h),\\ 0 & \text{if}\ \ k>\tau_i(h);\ \ i=1,2.\end{cases}$$

Then,

$$\lim_{h\to\infty}\sup_{\theta\in\Theta}\sup_{t\in R}\left|P_\theta\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{\tau(h)}y_{k-1}\varepsilon_k<t\right)-\Phi(t)\right|=0$$

where Φ is the standard normal distribution function.

The proof of Proposition 2 actually proceeds along the lines of the proof of Proposition 2.1 in Lai and Siegmund (1983). Some technical changes are caused by the presence of two stopping times instead of one. For sake of clarity and ease of reading, the proof of Proposition 2 is given in Appendix.

Now, we consider sequential least squares estimates for $\theta = (\theta_1, \theta_2)$ defined by (5)–(7) in the case of unspecified distribution of residuals ε_k in (1). Let the filtration $\{\mathcal{F}_{n\geq 0}\}$ be defined as

$$\mathcal{F}_0 = \sigma\{x_0\}, \ \ \mathcal{F}_n = \sigma\{x_0, \varepsilon_1, \dots, \varepsilon_n\}, \ n \geq 1. \tag{17}$$

Then, τ_1 and τ_2 introduced in (5) are stopping times with respect to the filtration (17). We assume that the process (1) is ergodic, that is the parametric region Θ given by (4). In order to apply Proposition 2 to the process (1), one has to verify conditions (6) because all other conditions are obvious.

Proposition 3 *Let $\{x_k\}$ obey equation (1) and $\{\varepsilon_k\}_{k\geq 1}$ be an i.i.d. sequence with $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = 1$. Then, for any $0 < \lambda < 1$ both conditions (6) of Proposition 2 hold for all $\theta \in \Theta_\lambda$ where*

$$\Theta_\lambda = \left\{\theta = (\theta_1, \theta_2) : \ |\theta_i| < \lambda, \ i = 1, 2\right\}. \tag{18}$$

The proof of this result is rather laborious and is postponed to the next section.

The following theorem establishes the property of uniform asymptotic normality for the sequential estimates (5)–(7), when the distribution of residuals (ε_k) in (1) is not specified.

Theorem 2 *Define $\hat\theta(h) = (\hat\theta_1(h), \hat\theta_2(h)), \ h > 0$, by (6, 7) and $\tau_1(h), \tau_2(h)$ by (5). If $\{\varepsilon_k\}_{k\geq 1}$ are i.i.d. with mean 0 and variance 1, and are independent of x_0, then for any $0 < \lambda < 1$*

$$\lim_{h\to\infty} \left| P_\theta\left(\frac{h^+}{\sqrt{h}}\left((\hat\theta_1(h) - \theta_1\right) \leq t_1, \ \frac{h^-}{\sqrt{h}}\left((\hat\theta_2(h) - \theta_2\right) \leq t_1 \right) - \Phi(t_1)\Phi(t_2)\right| = 0,$$

uniformly for $\theta \in \Theta_\lambda$ and $t \in R^2$; h^+ and h^- are given by (10).

Proof . Substituting (1) in (7) yields

$$\frac{h^+}{\sqrt{h}}\left((\hat\theta_1(h) - \theta_1\right) = \frac{1}{\sqrt{h}} \sum_{k=1}^{\tau_1(h)} \beta_{1,k} x_{k-1}^+ \varepsilon_k,$$
$$\frac{h^-}{\sqrt{h}}\left((\hat\theta_2(h) - \theta_2\right) = \frac{1}{\sqrt{h}} \sum_{k=1}^{\tau_2(h)} \beta_{2,k} x_{k-1}^- \varepsilon_k.$$

From here, it follows that

$$u_1 \cdot \frac{h^+}{\sqrt{h}}\left(\hat{\theta}_1(h) - \theta_1\right) + u_2 \cdot \frac{h^-}{\sqrt{h}}\left(\hat{\theta}_2(h) - \theta_2\right) = \frac{1}{\sqrt{h}} \sum_{k=1}^{\tau_(h)} y_{k-1}\varepsilon_k$$

where $\tau(h) = \tau_1(h) \vee \tau_2(h)$,

$$y_{k-1} = \beta_{1,k} u_1 x_{k-1}^+ + \beta_{2,k} u_2 x_{k-1}^-,$$

$u = (u_1, u_2) \in R^2$ with $||u|| = 1$. For $\theta \in \Theta$, thanks to Proposition 3, the process (1) satisfies all conditions of Proposition 2 which implies the desired result. Theorem 2 is proved. $\square$

A direct consequence of Theorem 2 is the following result.

Proposition 4 *Under conditions of Theorem 2, for* $0 < \lambda < 1$

$$\lim_{h\to\infty} P_\theta\left(\frac{(h^+)^2}{h}\left((\hat{\theta}_1(h) - \theta_1\right)^2 + \frac{(h^-)^2}{h}\left((\hat{\theta}_2(h) - \theta_2\right)^2 \le t \right) = P\left(\chi_2^2 \le t\right)$$

uniformly for $\theta \in \Theta_\lambda$ *and* $t \in R^2$.

4 Proof of Proposition 3

Let $\theta \in \Theta_\lambda = \left\{\theta = (\theta_1, \theta_2) : \; |\theta_i| < \lambda, \; i = 1, 2\right\}$. We have to verify the following limiting relations, for $TAR(1)$ process (x_k): for each $\delta > 0$

$$\lim_{n\to\infty} \sup_\theta P_\theta\left((x_n^+)^2 > \delta \sum_{k=1}^{n} (x_{k-1}^+)^2 \quad \text{for some} \quad n \ge m \right) = 0; \tag{19}$$

$$\lim_{n\to\infty} \sup_\theta P_\theta\left((x_n^-)^2 > \delta \sum_{k=1}^{n} (x_{k-1}^-)^2 \quad \text{for some} \quad n \ge m \right) = 0. \tag{20}$$

The proofs of relations (19) and (20) are similar. We will consider that of (19). Equation (1) can be written as

$$x_k^+ = \theta_1 x_{k-1}^+ + \varepsilon_k', \tag{21}$$

where $\varepsilon_k' = -x_k^- + \theta_2 x_{k-1}^- + \varepsilon_k$. Further, we apply Lemma 1 from the paper of Pergamenshchikov (1992) which claims that if $\left\{x_k\right\}_{k\ge0}$ is autoregressive process

$$x_k = \theta x_{k-1} + \varepsilon_k, \quad k \ge 1,$$

with $|\theta| \le 1$, then

$$\sum_{k=1}^{n}(x_k)^2 \geq c \sum_{k=1}^{n} \varepsilon_k^2, \tag{22}$$

where $c = 3 - 2\sqrt{2}$.

For Eq. (21), this inequality takes the form

$$\sum_{k=1}^{n}(x_k^+)^2 \geq c \sum_{k=1}^{n}(\varepsilon_k')^2. \tag{23}$$

It will be observed that

$$(\varepsilon_k')^2 \geq (\varepsilon_k')^2 \chi_{(x_{k-1}>0)} \chi_{(x_k>0)} = \varepsilon_k^2 \chi_{(x_{k-1}>0)} \chi_{(\theta_1 x_{k-1}+\varepsilon_k>0)} =: \zeta_k'.$$

From here and (23), one has

$$\sum_{k=0}^{n}(x_k^+)^2 \geq c \cdot \eta_n', \quad \eta_n' = \sum_{k=1}^{n} \zeta_k'. \tag{24}$$

Now, we represent the process $\{\eta_n'\}$ as

$$\begin{aligned} \eta_n' &= M_n' \ + \ T_n', \\ M_n' &= \sum_{k=1}^{n} \zeta_k' - T_n', \quad T_n' = \sum_{k=1}^{n} g_{\theta_1}'(x_{k-1}); \end{aligned} \tag{25}$$

$$g_{\theta_1}'(z) = E_\theta\left(\zeta_k' | x_{k-1} = z\right) = \chi_{(z>0)} \int_{-\theta_1 z}^{\infty} y^2 f(y) \mathrm{d}y. \tag{26}$$

The first step in estimating from above the probability of event

$$A_m(\delta) = \left\{ \frac{(x_n^+)^2}{\sum_{k=1}^{n}(x_{k-1}^+)^2} \geq \delta \ \text{ for some } \ n \geq m \right\}. \tag{27}$$

is the following result.

Lemma 1 *For each natural number p, for $|\theta_i| \leq 1$ and for sufficiently large n,*

$$\frac{(x_n^+)^2}{\sum\limits_{k=1}^{n}(x_{k-1}^+)^2} \leq \frac{2}{p} \ + \frac{12p\left(\sum\limits_{1=1}^{p+1}(x_{n-i})^2 + \sum\limits_{1=1}^{p} \varepsilon_{n-i}^2\right)}{c\eta_n}. \tag{28}$$

Proof For a fixed number p, we define the number

$$l_n^{(p)} \in \left\{ l : \ x^+_{n-l} = \min_{1\le j\le p} x^+_{n-j}, \ 1 \le l \le p \right\}.$$

By making use of (21) repeatedly, one gets

$$x_n^+ = \theta_1^l x^+_{n-l} + \sum_{i=0}^{l-1} \theta_1^i \, \varepsilon'_{n-i}.$$

From here and (24), it follows that

$$\begin{aligned} \frac{(x_n^+)^2}{\sum_{k=1}^{n}(x^+_{k-1})^2} &\le \frac{2(x^+_{n-l_n^{(p)}})^2}{\sum_{k=1}^{n}(x^+_{k-1})^2} + \frac{2\left(\sum_{i=0}^{l_{n^{(p)}}-1} \theta_1^i \, \varepsilon'_{n-i}\right)^2}{\sum_{k=1}^{n}(x^+_{k-1})^2} \\ &\le \frac{2}{p} + \frac{2p\sum_{i=0}^{p}(\varepsilon'_{n-i})^2}{c\sum_{k=1}^{n}(\varepsilon'_k)^2}. \end{aligned} \tag{29}$$

In view of (21),

$$\begin{aligned} \sum_{i=0}^{p}(\varepsilon'_{n-i})^2 &\le \left(3\sum_{i=0}^{p} x^2_{n-i} + \sum_{i=0}^{p} x^2_{n-i-1} + \sum_{i=0}^{p} \varepsilon^2_{n-i}\right) \\ &\le 6\left(\sum_{i=0}^{p+1} x^2_{n-i} + \sum_{i=0}^{p} \varepsilon^2_{n-i}\right). \end{aligned} \tag{30}$$

Combining (29), (30) yields (28).

Lemma 1 is proved. □

Further, we will study the asymptotic behavior of the ratios $x_n/\sqrt{n}$ and η'_n/n as $n \to \infty$.

Lemma 2 *Let* $\{x_k\}$ *be defined by* (1) *and* Θ_λ *by* (18). *Then, for any* $0 < \lambda < 1$ *with probability one*

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_\lambda} \frac{|x_n|}{\sqrt{n}} = 0, \tag{31}$$

The proof of Lemma 2 is given in Appendix.

The analysis of each term in the sum

$$\frac{|\eta'_n|}{\sqrt{n}} = \frac{T'_n}{n} + \frac{M'_n}{n} \tag{32}$$

requires its own technical means. It will be observed that the process (1) is a homogeneous Markov chain, and moreover, it has the property of uniform geometric ergodicity. All necessary notions are reminded in Appendix. By applying general Theorem 1, cited ibidem, from the paper by Galtchouk and Pergamenshchikov (2014) we obtain

Lemma 3 *For any* $0 < \lambda < 1$

$$\sup_{\theta\in\Theta_\lambda}\sup_{x\in R}\frac{1}{V(x)}\left|E_x^\theta g'(x_k) - b'(\theta)\right| \le R^* e^{-\kappa^* k}, \quad k \ge 1, \tag{33}$$

where $V(x)$ *is the Lyapunov function* (.),

$$b'(\theta) = \int_R g'_{\theta_1}(z)\pi^\theta(z)dz,$$

π^θ *is the ergodic measure of the chain,* κ^* *and* R^* *are some positive parameters.*

From here, one can derive the following result.

Lemma 4 *For any* $0 < \lambda < 1$

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_\lambda}\sup_{\lambda\in R}\frac{1}{V(x)}\left|\frac{1}{n}E_x^\theta T'_n - b'(\theta)\right| = 0. \tag{34}$$

Proof We have

$$\begin{aligned}\left|\frac{1}{n}E_x^\theta T'_n - b'(\theta)\right| &= \left|\frac{1}{n}\sum_{k=1}^n E_x^\theta g'(x_{k-1}) - b'(\theta)\right| \\ &\le \left|\frac{1}{n}\frac{1}{V(x)}\sum_{k=1}^N E_x^\theta g'(x_{k-1}) - b'(\theta)\right| + \frac{(n-N)}{n}R^*\sum_{k=N+1}^n e^{-\kappa^*(k-1)} \\ &\le \left|\frac{1}{n}\frac{1}{V(x)}\sum_{k=1}^N E_x^\theta g'(x_{k-1}) - b'(\theta)\right| + \frac{(n-N)}{n}R^*\frac{e^{-N\kappa^*}}{1-e^{-\kappa^*}}.\end{aligned}$$

Limiting $n \to \infty$, then $N \to \infty$ and taking into account (33), we come to (34). Lemma 4 is proved. □

Now, we can estimate η'_n/n from below. For any $0 < L < \infty$ and $|x| \le L$, one has

$$
\begin{aligned}
\frac{1}{n} E_x^\theta T_n' &\geq \frac{1}{nV(x)} E_x^\theta T_n' = \frac{1}{V(x)} \left| \frac{1}{n} E_x^\theta T_n' - b'(\theta) + b'(\theta) \right| \\
&\geq \frac{b'(\theta)}{V(x)} - \frac{1}{V(x)} \left| \frac{1}{n} E_x^\theta T_n' - b'(\theta) \right| \\
&\geq \frac{1}{V(x)} \inf_{\theta \in \Theta_\lambda} b'(\theta) - \sup_{|x| \leq L} \sup_{\theta \in \Theta_\lambda} \frac{\left| \frac{1}{n} E_x^\theta T_n' - b'(\theta) \right|}{V(x)}.
\end{aligned} \tag{35}
$$

Denote

$$
\begin{aligned}
B(L, \lambda) &= \frac{1}{V(x)} \inf_{\theta \in \Theta_\lambda} b'(\theta), \\
t'(n, \lambda) &= \sup_{|x| \leq L} \sup_{\theta \in \Theta_\lambda} \frac{\left| \frac{1}{n} E_x^\theta T_n' - b'(\theta) \right|}{V(x)}.
\end{aligned}
$$

Combining (32) and (34) yields

$$
\frac{1}{n} \eta_n' \geq B(L, \lambda) - t'(n, \lambda) - \frac{|M_n'|}{n}. \tag{36}
$$

By making use of this estimate in Lemma 1, we obtain for sufficiently large n:

$$
\frac{(x_n^+)^2}{\sum_{k=1} n (x_{k-1}^+)^2} \leq \frac{2}{p} + \Delta(\lambda; n), \tag{37}
$$

where

$$
\Delta(\lambda; n) = \frac{12p \left(\frac{1}{n} \sum_{i=0}^{p+1} x_{n-i}^2 + \frac{1}{n} \sum_{i=0}^{p} \varepsilon_{n-i}^2 \right)}{C \left(B(L, \lambda) - t'(n, \lambda) - \frac{|M_n'|}{n} \right)}. \tag{38}
$$

Further, we will establish the following result.

Lemma 5 *Let $\Delta(\lambda, n)$ be defined by (38). Then, for any $0, \lambda < 1$ and $\mu > 0$*

$$
\lim_{m \to \infty} \sup_{\theta \in \Theta_\lambda} P_\theta \left\{ \sup_{n \geq m} \Delta(\lambda, n) \geq \mu \right\} = 0. \tag{39}
$$

The proof of Lemma 5 is postponed to Appendix.

Finally, we will show that (37) and (39) imply the desired result (19). Let $A_m(\delta)$ be defined by (27) and $\delta > 0$. In view of (37),

$$
\begin{aligned}
A_m(\delta) &= A_m(\delta) \bigcap \left\{ \frac{2}{p} + \Delta(\lambda, n) < \delta \ \text{ for all } \ n \geq m \right\} \\
&\quad + A_m(\delta) \bigcap \left\{ \frac{2}{p} + \Delta(\lambda, n) \geq \delta \ \text{ for some } \ n \geq m \right\} \\
&\subset \left\{ \frac{2}{p} + \Delta(\lambda, n) \geq \delta \ \text{ for some } \ n \geq m \right\} \\
&= \left\{ \frac{2}{p} + \Delta(\lambda, n) \geq \delta \ \text{ for some } \ n \geq m;\ \Delta(\lambda, n) < \mu \ \text{ for all } \ n \geq m \right\} \\
&\quad + \left\{ \frac{2}{p} + \Delta(\lambda, n) \geq \delta \ \text{ for some } \ n \geq m;\ \Delta(\lambda, n) < \mu \ \text{ for some } \ n \geq m \right\} \\
&\subset \left\{ \frac{2}{p} + \mu \geq \delta \right\} \bigcup \{\Delta(\lambda, n) < \mu \ \text{ for some } \ n \geq m\}.
\end{aligned}
$$

If m is large and μ is small enough, then $\frac{2}{p} + \mu < \delta$. Therefore,

$$
\sup_{\theta \in \Theta_\lambda} P_\theta\big(A_m(\delta)\big) \ \leq \ \sup_{\theta \in \Theta_\lambda} P_\theta\left(\bigcup_{n \geq m} \{\Delta(\lambda, n) \geq \mu\} \right).
$$

Limiting $m \to \infty$ and taking into account (39), we derive (19). The proof of Proposition 3 has been completed. □

5 Sequential estimators of parameters of generalized TAR(p) process

Consider a TAR(p) process, obeying the equation

$$
x_k = \sum_{i=1}^{p} \theta_i g_i(x_{k-1}) I_i(x_{k-1}) + \varepsilon_k, \tag{40}
$$

where g_i are known functions (for example, $g_i(x) = x$),

$$
I_i(x) = \begin{cases} 1, & \textit{if} \quad r_{i-1} < x \leq r_i \\ 0, & \textit{otherwise} \end{cases},
$$

and $-\infty = r_0 < r_1 < \cdots < r_{p-1} < r_p = +\infty$ are some fixed points; it is assumed that the errors $\{\varepsilon_k\}$ are i.i.d. with the standard Gaussian distribution. For a fixed $h > 0$, we introduce stopping rules

$$
\tau_i(h) = \inf \left\{ n \geq 1 \, : \, \sum_{k=1}^{n} g_i^2(x_{k-1}) I_i(x_{k-1}) \geq h \right\}, \quad 1 \leq i \leq p. \tag{41}
$$

Denote sequential least squares estimates for $\theta = (\theta_1, \ldots, \theta_p)$ as the vector

$$
\hat{\theta}(h) = (\hat{\theta}_1(h), \ldots, \hat{\theta}_p(h)) \tag{42}
$$

with coordinates

$$\hat{\theta}_i(h) = \frac{1}{h_i} \sum_{k=1}^{\tau_i(h)} \beta_{i,k} g_i(x_{k-1}) I_i(x_{k-1}) x_k, \tag{43}$$

where $\beta_{i,k}$ are weight coefficients of the form

$$\beta_{i,k} = \begin{cases} 1 & \text{if } k < \tau_i(h), \\ \sqrt{\alpha_{i,\tau_i}} & \text{if } k = \tau_i(h), \\ 0 & \text{if } k > \tau_i(h); \; 1 \le i \le p. \end{cases} \tag{44}$$

Here, α_{i,τ_i} are correction factors compensating the overshots in (41), uniquely defined by the equations

$$\sum_{k=1}^{\tau_i(h)} g_i^2(x_{k-1}) I_i(x_{k-1}) + (\alpha_{i,\tau_i(h)} - 1) g_i^2(x_{k-1}) I_i(x_{k-1}) = h, \tag{45}$$

$$h_i = \sum_{k=1}^{\tau_i(h)} \beta_{i,k} g_i^2(x_{k-1}) I_i(x_{k-1}), \quad 1 \le i \le p. \tag{46}$$

The following theorem gives joint distribution of the standardized deviations of estimates (43).

Theorem 3 *For $h > 0$ define $\hat{\theta}_i(h)$ by (42), (43) and $\tau_i(h)$ by (41). Let the following conditions*

$$P_\theta\left(\sum_{k \ge 1} g_i^2(x_{k-1}) I_i(x_{k-1}) = +\infty \right) = 1, \quad 1 \le i \le p \tag{47}$$

are satisfied. If $\{\varepsilon_k\}$ are i.i.d. normal random variables with mean 0 and variance 1 and are independent of x_0, then

1) *the standardized deviations*

$$\xi_i = \frac{h_i}{\sqrt{h}} \left(\hat{\theta}_i(h) - \theta_i \right), \quad 1 \le i \le p \tag{48}$$

have the standard p—dimensional normal distribution $N_p(0, I)$, I is the unit matrix of the size $p \times p$, that is

$$P_\theta\left(\xi_1 < z_1, \dots, \xi_p < z_p \right) = \Phi(z_1) \cdots \Phi(z_p), \quad -\infty < z_i < \infty, \quad 1 \le i \le p.$$

$$\text{2)} \quad \text{for every} \quad x \in R \quad P_\theta\left(\sum_{i=1}^{p} \frac{h_i^2}{h} \left(\hat{\theta}_i(h) - \theta_i \right)^2 \le x \right) = P\left(\chi_p^2 \le x \right), \tag{49}$$

where χ_p^2 is Chi-square random variable with p degrees of freedom.

The proof of Theorem 3 actually proceeds along the lines of the proof of Theorem 1.

Equation (49) enables one to construct a fixed-size confidence region for unknown parameters $(\theta_1, \ldots, \theta_p)$ with a prescribed coverage probability. Consider a family of confidence regions of elliptic form

$$G(r) = \left\{ (t_1, \ldots, t_p) \in R^n \,:\, \sum_{i=1}^{p} \left(\frac{h_i}{h} \right)^2 \left(\hat{\theta}_i(h) - t_i \right)^2 \leq r^2 \right\}, r > 0. \tag{50}$$

Proposition 5 *Under conditions of Theorem 3 for any given $r > 0,\ 0 < \alpha < 1$*

$$P_\theta\{\theta \in G(r)\} = 1 - \alpha \tag{51}$$

for stopping rules (41) with $h = F^{-1}_{\chi_p^2}(1-\alpha)/r^2$, where $F_{\chi_p^2}$ denotes the distribution function of χ_p^2 random variable.

Proof Let $r > 0$ and $0 < \alpha < 1$ be fixed. From (49), (50), it follows that confidence region (50) covers true values of parameters $\theta_i,\ 1 \leq i \leq p$ with probability $1-\alpha$ if h satisfies the equation

$$P_\theta(\theta \in G(r)) = P\left(\frac{\chi_p^2}{h} \leq r^2 \right) = P\left(\chi_p^2 \leq h \cdot r^2 \right) = 1 - \alpha.$$

Solving this equation with respect to h, one comes to the desired result. □

6 Simulation results

In this section, we present some results of Monte Carlo simulation of constructing confidence regions for parameters TAR(1) described by Eq. (1). The value of x_0 in all cases was taken equal to zero. The noise $\{\varepsilon_k\}_{k\geq 1}$ is assumed to be an i.i.d. sequence of random variables with zero mean and unit variance. The estimates of unknown parameters $\theta_1,\ \theta_2$ are defined by formulas (5–10). The fixed-size confidence ellipsoid is constructed using Eqs. (13, 14)

$$G(r) = \left\{ (t_1, t_2) \,:\, \left(\frac{h^+}{h} \right)^2 \left(\hat{\theta}_1(h) - t_1 \right)^2 + \left(\frac{h^{-2}}{h} \right) \left(\hat{\theta}_2(h) - t_2 \right)^2 \leq r^2 \right\}, r > 0,$$

and the parameter $h = h(\alpha, r)$ is found from (15):

$$P\left(\frac{\chi_2^2}{h} \geq r^2\right) = \alpha.$$

According to Proposition 1,

$$P_\theta(\theta \in G(r)) = 1 - \alpha.$$

The results of Monte Carlo simulation are reported in Tables 1–3. The values r and α were equal to 0.2 and 0.1, respectively. The corresponding value of the parameter $h(\alpha, r)$ is equal to 115.13. The quantities $\tilde{E}_\theta \tau_1$ and $\tilde{E}_\theta \tau_2$ denote the observed averages of stopping times $\tau_1(h)$ and $\tau_2(h)$, respectively ; $\hat{\alpha}$ denotes the frequency count of the number of times when the confidence ellipsoid does not contain the true values (θ_1, θ_2).

All the results were obtained by 10, 000 replications. The parameters θ_1, θ_2 in Table 1 lie inside the region of ergodicity of the process TAR(1) and were chosen as in Sriram and Iaci (2014). The noises $\{\varepsilon\}_n$ in (1) were taken as $\varepsilon_1 \sim N(0, 1)$.

One can see that in all cases the averaged coverage probability is very close to that of theoretical one.

The similar results in Table 2 were obtained when both parameters θ_1, θ_2 are negative and lie outside the region of ergodicity. The obtained results show that the constructed fixed-size confidence ellipsoid guarantees the prescribed coverage probability.

To illustrate the asymptotic property established in Proposition 4 (case of unspecified error distribution), we present the simulation results for the case of double exponential noises ε_n. The values α, r stay the same. The results are presented in Table 3. The only difference with Table 1 consists in changing of averaged stopping times.

Table 1 TAR(1) model with Gaussian noises. Region of ergodicity

$h = 115.13$		$r = 0.2$		$\alpha = 0.10$
θ_1	θ_2	$\tilde{E}_\theta \tau_1$	$\tilde{E}_\theta \tau_2$	$\hat{\alpha}$
0.9	0.5	51.8	320.5	0.0972
0.9	− 0.5	39.1	968.7	0.1017
0.9	− 10.0	11.5	3378.6	0.1021
0.1	− 0.5	162.2	311.1	0.1052
− 0.1	0.5	361.0	132.5	0.1016
− 0.1	− 5.0	14.5	207.3	0.0944
− 0.9	0.5	510.9	94.7	0.1018
− 0.9	− 0.5	155.3	106.9	0.1048
0.5	0.5	180.0	180.5	0.0935
− 0.5	− 0.5	176.6	177.1	0.1014
− 0.9	− 0.9	58.0	58.0	0.0969
− 0.2	− 4.9	11.7	50.5	0.1033
− 0.19	− 5	11.8	58.1	0.1056

Table 2 TAR(1) model. Negative parameters

$h = 115.13$		$r = 0.2$		$\alpha = 0.10$
θ_1	θ_2	$\tilde{E}_\theta \tau_1$	$\tilde{E}_\theta \tau_2$	$\hat{\alpha}$
− 5.0	− 0.2	46.8	11.4	0.1015
− 10.0	− 0.1	48.1	7.3	0.0983
− 10.0	− 0.2	11.8	6.5	0.0966
− 10.0	− 2.0	5.1	4.6	0.1002
− 10.0	− 0.1	48.1	7.4	0.0991
− 10.0	− 5.0	4.6	4.3	0.1022
− 1.0	− 1.0	32.7	32.7	0.0969
− 0.1	− 10.0	7.4	48.4	0.1020
− 0.2	− 5.0	11.4	47.0	0.1047

Table 3 TAR(1) model with double exponential noises

$h = 115.13$		$r = 0.2$		$\alpha = 0.10$
θ_1	θ_2	$\tilde{E}_\theta \tau_1$	$\tilde{E}_\theta \tau_2$	$\hat{\alpha}$
0.9	0.5	54.9	311.7	0.0979
0.9	− 0.5	41.7	783.1	0.1009
0.1	− 0.5	168.5	292.9	0.1023
− 0.1	0.5	331.1	141.5	0.0990
− 0.9	0.5	429.2	97.3	0.0970
− 0.9	− 0.5	158.4	108.1	0.0988
0.5	0.5	183.0	182.9	0.0955
− 0.5	− 0.5	178.9	178.8	0.1027
− 0.9	− 0.9	59.8	59.8	0.1005

Table 4 Simulation results of TAR(1) model based on 10000 replications

$h = 115.13$							
θ_1	θ_2	$E_\theta \tau_1$	$E_\theta \tau_2$	$\Delta_1^2(S)$	$\Delta_2^2(S)$	$\Delta_1^2(L)$	$\Delta_2^2(L)$
− 10.0	− 0.1	48.3	7.4	0.0087	0.0087	0.0120	0.0188
− 5.0	− 0.2	46.7	11.4	0.0088	0.0087	0.0126	0.0162
− 2.0	− 0.5	41.8	22.1	0.0087	0.0085	0.0116	0.0149
− 1.0	− 1.0	32.8	32.8	0.0086	0.0087	0.0151	0.0143
− 0.2	− 5.0	11.4	46.9	0.0088	0.0087	0.0192	0.0114
− 0.5	− 2.0	21.7	41.1	0.0084	0.0087	0.0167	0.0126
− 0.1	− 10.0	7.3	48.2	0.0088	0.0086	0.0305	0.0117

To compare the quality of the proposed sequential estimators and least squares estimators, an additional simulation was performed. The sample size n of the least squares estimators θ_i was chosen so that $\hat{E}_\theta \tau_i \approx n$ for each $i = 1, 2$. The results are given in Table 4. Here, $\Delta_1^2(S)$, $\Delta_2^2(S)$ and $\Delta_1^2(L)$, $\Delta_2^2(L)$ denote the averaged squared deviations of sequential and least squares estimators from the corresponding true values of the parameters θ_1, θ_2. The parameters lie on the hyperbolic boundary of the region of ergodicity.

Note that according to Theorem 1, the mean square deviation of the estimates $\hat{\theta}_j$, $j = 1, 2$ satisfies the inequality

$$E_\theta\left(\hat{\theta}_j - \theta_j\right)^2 \leq \frac{1}{h}.$$

One can see that the averaged squared deviations of sequential estimates are close to the value $1/h = 0,00868$. The quality of sequential estimates appeared to be essentially better as compared to least squares estimates.

Numerical results confirm the established properties of estimates.

7 Conclusion

In this paper, we consider the problem of estimating parameters in TAR process. We construct fixed-size confidence regions on the basis of the sequential point estimates. It is shown that this allows one to derive non-asymptotic confidence ellipsoid of prescribed size and coverage probability. The results may be useful for fitting nonlinear time series models.

Appendix

Proof of Theorem 1 Let the filtration $\{\mathcal{F}\}_{n\geq 0}$ be given by (17). We will show that the characteristic function of vector $\xi = (\xi_1, \xi_2)$ with the coordinates defined in (6) has the form

$$\varphi_\xi(u) = Ee^{i(u,\xi)} = Ee^{i(u_1\xi_1 + u_2\xi_2)} = e^{-\frac{u_1^2}{2}} e^{-\frac{u_2^2}{2}},$$

$u = (u_1, u_2)$, $-\infty < u_j < \infty$, $j = 1, 2$. Taking into account (11), we introduce two sequences

$$\xi_1(N) = \frac{1}{\sqrt{h}} \sum_{k=1}^{N} \chi_{(k\leq\tau_1(h))} \beta_{1,k} x_{k-1}^{+} \varepsilon_k,$$

$$\xi_2(N) = \frac{1}{\sqrt{h}} \sum_{j=1}^{N} \chi_{(j\leq\tau_2(h))} \beta_{2,j} x_{j-1}^{+} \varepsilon_j, \quad N \geq 1.$$

Consider the characteristic function of the vector $\xi(N) = (\xi_1(N), \xi_2(N))$:

$$\varphi_{\xi(N)}(u) = Ee^{i(u,\xi(N))} = E\exp\left(\sum_{k=1}^{N} \frac{i}{\sqrt{h}} y_{k-1}\varepsilon_k\right)$$

where

$$y_{k-1} = \chi_{(k\le\tau_1(h))}\beta_{1,k}u_1x^+_{k-1} \ + \ \chi_{(k\le\tau_2(h))}\beta_{2,k}u_2x^-_{k-1}.$$

Since

$$\lim_{N\to\infty}\xi_1(N) = \xi_1, \quad \lim_{N\to\infty}\xi_2(N) = \xi_2,$$

we have

$$\varphi_\xi(u) = \lim_{N\to\infty}\varphi_{\xi(N)}(u).$$

Now, we represent $\varphi_{\xi(N)}(u)$ as

$$\begin{aligned}\varphi_{\xi(N)}(u) &= E\exp\left(\left(\sum_{k=1}^N \frac{i}{\sqrt{h}}y_{k-1}\varepsilon_k + \frac{1}{2h}y^2_{k-1}\right) - \sum_{k=1}^N \frac{1}{2h}y^2_{k-1}\right)\\ &= \exp\left(-\frac{u_1^2}{2} - \frac{u_2^2}{2}\right)E\exp\left(\sum_{k=1}^N \frac{i}{\sqrt{h}}y_{k-1}\varepsilon_k + \frac{1}{2h}y^2_{k-1}\right) \ + \ R_N,\end{aligned} \tag{52}$$

where

$$\begin{aligned}R_N =& E\left[\exp(\eta_N)\cdot S_N\right],\\ \eta_N =& \sum_{k=1}^N\left(\frac{i}{\sqrt{h}}y_{k-1}\varepsilon_k + \frac{1}{2h}y^2_{k-1}\right),\\ S_N =& E\exp\left(\sum_{k=1}^N \frac{1}{2h}y^2_{k-1}\right) - \exp\left(-\frac{u_1^2}{2} - \frac{u_2^2}{2}\right).\end{aligned}$$

Taking repeatedly conditional expectation yields

$$\begin{aligned}Ee^{\eta(N)} &= E\left(E\left(e^{\eta(N)}|\mathcal{F}_{N-1}\right)\right)\\ &= E\exp\left(\sum_{k=1}^{N-1}\frac{i}{\sqrt{h}}y_{k-1}\varepsilon_k + \sum_{k=1}^N \frac{1}{2h}y^2_{k-1}\right)E\exp\left(\frac{i}{\sqrt{h}}y_{N-1}\varepsilon_N|\mathcal{F}_{N-1}\right)\\ &= E\left[\exp\left(\sum_{k=1}^{N-1}\left(\frac{i}{\sqrt{h}}y_{k-1}\varepsilon_k + \frac{1}{2h}y^2_{k-1}\right)\right)\right] = Ee^{\eta(N-1)} = \cdots = 1.\end{aligned} \tag{53}$$

Further, we note that

$$\begin{aligned}\sum_{k=1}^N \frac{1}{2h}y^2_{k-1} &\le \frac{1}{2}\left(u_1^2 + u_2^2\right),\\ \lim_{N\to\infty}\sum_{k=1}^N \frac{1}{2h}y^2_{k-1} &= \frac{1}{2}\left(u_1^2 + u_2^2\right).\end{aligned}$$

Using the estimate

$$Ee^{\eta(N)} \leq \exp\left(\frac{1}{2}\left(u_1^2 + u_2^2\right)\right)$$

and applying the theorem of dominated convergence, one gets

$$\lim_{N\to\infty} R_N = 0.$$

Substituting (53) in (52) and limiting $N \to \infty$, we arrive at the desired result. Thus, Theorem 2.1 is proved. □

Proof of Lemma 2 Noting that

$$|x_k| \;\leq\; \lambda|x_{k-1}| + |\varepsilon_k|, \quad k \geq 1,$$

and applying this inequality repeatedly, one gets

$$|x_k| \;\leq\; \lambda^n|x_0| + \sum_{j=1}^{n} \lambda^{n-j}|\varepsilon_j.$$

This implies the following estimate $N < n$

$$\begin{aligned}\frac{|x_k|}{\sqrt{n}} \;&\leq\; \frac{\lambda^n}{\sqrt{n}}|x_0| + \frac{1}{\sqrt{n}}\sum_{j=0}^{N} \lambda^{n-j}|\varepsilon_j| + \frac{1}{\sqrt{n}}\sum_{j=N+1}^{n} \lambda^{n-j}|\varepsilon_j| \\ &\leq \frac{\lambda^n}{\sqrt{n}}|x_0| + \frac{1}{\sqrt{n}}\sum_{j=0}^{N} \lambda^{n-j}|\varepsilon_j| + \frac{1}{1-\lambda}\sup\frac{|\varepsilon_j|}{\sqrt{j}}.\end{aligned}$$

Limiting $n \to \infty$, then $N \to \infty$ and thanks the strong law of large numbers one comes to (31). □

Proof of Lemma 5 For each $\theta \in \Theta_\lambda$, the process M_n' in decomposition (25) is a square integrable martingale subjected to the strong law of large numbers:

$$\lim_{n\to\infty} \frac{M_n'}{n} \to 0 \;\; (P_\theta - \;\; \text{a.s.}).$$

Moreover, this convergence is uniform in $\theta \in \Theta_\lambda$, i.e., for any $\mu > 0$

$$\sup_{\theta\in\Theta_\lambda} P_\theta\left\{\sup_{n\geq m}\frac{|M_n'|}{n} \geq \mu\right\} \to \;0 \;\;\text{a.s.}\;\; m \to \infty. \tag{54}$$

This can be checked by making use of the inequality (see, e.g., Shiryaev (1996))

$$\begin{aligned}\mu^2 P_\theta\left\{\sup_{n\geq m}\frac{|M'_n|}{n}\geq\mu\right\} &= \mu^2\lim_{l\to\infty}P_\theta\left\{\max_{m\leq n\leq l}\frac{(M'_n)^2}{n^2}\geq\mu^2\right\}\\ &\leq \frac{1}{m^2}E_\theta(M'_n)^2 \;+\; \sum_{n\geq m}E_\theta\big((M'_n)^2-(M'_{n-1})^2\big)\\ &= \sum_{n\geq m+1}\left(\frac{1}{(n-1)^2}-\frac{1}{n^2}\right)E_\theta(M_{n-1})^2.\end{aligned} \tag{55}$$

From the definition of M'_n, it follows that

$$E_\theta(M'_j)^2=\sum_{i=1}^{j}E_\theta(\Delta M'_j)^2\leq j\cdot E\varepsilon_1^4.$$

Using the estimate in (54), one gets

$$\sup_{\theta\in\Theta_\lambda}P_\theta\left\{\sup_{n\geq m}\frac{|M'_n|}{n}\geq\mu\right\}\;\leq\;\frac{2E\varepsilon_1^4}{\mu^2 m}. \tag{56}$$

This inequality provides the rate of convergence in (54). Now, we are ready to show (39). It remains to notice that the numerator of (38), thanks to Lemma 2 and the strong law of large numbers, tends to zero uniformly in $\theta\in\Theta_\lambda$ and the denominator of (38), in view of (54), is bounded away from zero below by some positive constant uniformly in $\theta\in\Theta_\lambda$. Thus, we arrive at (39). Lemma 5 is proved. □

Proof of Proposition 2

As in Lai and Siegmund (1983), we need the following martingale central limit theorem from Freedman (1983), pages 90–92.

Lemma 6 *Let* $0<\delta<1$ *and* $r>0$. *Assume that* $(u_n,\mathcal{F}_n)_{n\geq0}$ *is a martingale difference sequence satisfying*

$$|u_n|\leq\delta\quad\text{for all}\quad n$$

and

$$\sum E\big(u_n^2|\mathcal{F}_{n-1}\big)>r\quad\text{a.s.}$$

Let

$$\tau(h)=\inf\left\{n\geq1\;:\;\sum_{k=1}^{n}E\big(u_k^2|\mathcal{F}_{k-1}\big)\geq r\right\}.$$

There exists a function $\rho\,:\,(0,\infty)\to[0,2]$, *not depending on the distribution of martingale difference sequence*, *such that* $\lim\rho(x)=0$ *as* $x\to0$ *and*

$$\sup_{x\in R}\left|P\left(\sum_{k=1}^{\tau} u_k \le x\right) - \Phi\left(\frac{x}{\sqrt{\rho}}\right)\right| \le \rho\left(\frac{\delta}{\sqrt{\rho}}\right).$$

Proof of Proposition 2.

For each $0 < \delta < 1$, we define truncated versions for both processes $\{x_k^+\}_{k\ge 0}$ and $\{x_k^-\}_{k\ge 0}$:

$$\tilde{x}_k^+ = \begin{cases} x_k^+ & \text{if } (x_k^+)^2 \le \delta^2 h, \\ \delta\sqrt{h} & \text{if } (x_k^+)^2 > \delta^2 h; \end{cases}$$
$$\tilde{x}_k^- = \begin{cases} x_k^- & \text{if } (x_k^-)^2 \le \delta^2 h, \\ -\delta\sqrt{h} & \text{if } (x_k^-)^2 > \delta^2 h. \end{cases}$$

Then, we introduce the counterparts of stopping times (16) as

$$\begin{aligned} T_1(h) &= \inf\left\{ n \ge 1 \,:\, \sum_{k=1}^{n} \left(\tilde{x}_{k-1}^+\right)^2 \ge h \right\}, \\ T_2(h) &= \inf\left\{ n \ge 1 \,:\, \sum_{k=1}^{n} \left(\tilde{x}_{k-1}^-\right)^2 \ge h \right\}, \\ T(h) &= T_1(h) \vee T_2(h). \end{aligned} \tag{57}$$

Let $\tilde{\alpha}_{1,T_1}$ and $\tilde{\alpha}_{2,T_2}$ be correction factors compensating the overshots in (57) computed from the equations

$$\sum_{k=1}^{T_1(h)-1} \left(x_{k-1}^+\right)^2 + \tilde{\alpha}_{1,T_1(h)}\left(x_{T_1(h)-1}^+\right)^2 = h,$$
$$\sum_{k=1}^{T_2(h)-1} \left(x_{k-1}^-\right)^2 + \tilde{\alpha}_{2,T_2(h)}\left(x_{T_2(h)-1}^-\right)^2 = h.$$

Denote

$$\tilde{y}_{k-1} = \tilde{\beta}_{1,k} u_1 \tilde{x}_{k-1}^+ \; + \; \tilde{\beta}_{2,k} u_2 \tilde{x}_{k-1}^-, \quad 1 \le k \le \tau(h),$$

where

$$\tilde{\beta}_{i,k} = \begin{cases} 1 & \text{if } k < T_i(h), \\ \sqrt{\tilde{\alpha}_{i,T_i}} & \text{if } k = T_i(h), \\ 0 & \text{if } k > T_i(h); \; i = 1, 2, \end{cases}$$

and

$$\tilde{\varepsilon}_k = \varepsilon_k \chi_{(|\varepsilon_k| \le 1/\sqrt{\delta})}, \quad \tilde{\tilde{\varepsilon}}_k = \varepsilon_k - \tilde{\varepsilon}_k.$$

Then, under $P_\theta \left\{ \frac{1}{\sqrt{h}} \tilde{y}_{k-1}\left(\tilde{\varepsilon}_k - E\tilde{\varepsilon}_k\right), \; \mathcal{F}_k \right\}_{k\ge 0}$ is a martingale difference such that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{h}}\tilde{y}_{k-1}\left(\tilde{\varepsilon}_k - E\tilde{\varepsilon}_k\right)\right| \\
&\quad = \left|\frac{1}{\sqrt{h}}\left(\tilde{\beta}_{1,k}u_1\tilde{x}^+_{k-1} \;+\; \tilde{\beta}_{2,k}u_2\tilde{x}^-_{k-1}\right)\left(\tilde{\varepsilon}_k - E\tilde{\varepsilon}_k\right)\right| \\
&\quad \le \frac{1}{\sqrt{h}}2\delta\sqrt{h}\frac{2}{\sqrt{\delta}} = 4\sqrt{\delta}.
\end{aligned}$$

By Lemma 6

$$\left|P_\theta\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)}\tilde{y}_{k-1}\left(\tilde{\varepsilon}_k - E\tilde{\varepsilon}_k\right) \le t\right) - \Phi\left(\frac{t}{\sqrt{v_\theta(\delta)}}\right)\right| \le \rho\left(4\frac{\delta}{\sqrt{v_\theta(\delta)}}\right), \tag{58}$$

where $v_\theta(\delta) = \operatorname{Var}_\theta\tilde{\varepsilon}_1 \to 1$ uniformly in Θ as $\delta \to 0$. We need the following sets

$$\begin{aligned}
\Omega_{1,h} =& \left\{x^+_k = \tilde{x}^+_k \;\text{ for all }\; k < \tau_1(h)\right\}, \\
\Omega_{2,h} =& \left\{x^-_k = \tilde{x}^-_k \;\text{ for all }\; k < \tau_2(h)\right\}, \\
\Omega_h =& \Omega_{1,h} \bigcap \Omega_{2,h}.
\end{aligned}$$

We will show that

$$\lim_{h\to\infty}\sup_{\theta\in\Theta} P_\theta\left(\Omega^c_h\right) = 0.$$

It suffices to check that

$$\lim_{h\to\infty}\sup_{\theta\in\Theta} P_\theta\left(\Omega^c_{i,h}\right) = 0, \quad i = 1, 2. \tag{59}$$

For all θ and $h > 0$, one has the inequality

$$\begin{aligned}
P_\theta\left(\Omega^c_{i,h}\right) &= P_\theta\left\{x^+_k \neq \tilde{x}^+_k \;\text{ for some }\; k < \tau_1(h)\right\} \\
&\le \sum_{k=1}^{m} P_\theta\left\{\left(x^+_{k-1}\right)^2 > \delta^2 h\right\} \;+\; P_\theta\left\{\tau_1(h),\, x^+_k \neq \tilde{x}^+_k \;\text{ for some }\; m \le k < \tau_1(h)\right\} \\
&\le \sum_{k=1}^{m} P_\theta\left\{\left(x^+_{k-1}\right)^2 > \delta^2 h\right\} \;+\; P_\theta\left\{\left(x^+_n\right)^2 \ge \delta^2\sum_{k=1}^{n}\left(x^+_{k-1}\right)^2 \;\text{ for some }\; n \ge m\right\}.
\end{aligned}$$

From here, it follows

$$\sup_{\theta\in\Theta} P_\theta\left(\Omega^c_{i,h}\right) \le \sum_{k=1}^{m} \sup_{\theta\in\Theta} P_\theta\left\{\left(x^+_{k-1}\right)^2 > \delta^2 h\right\}$$
$$+ \sup_{\theta\in\Theta} P_\theta\left\{\left(x^+_n\right)^2 \ge \delta^2 \sum_{k=1}^{n}\left(x^+_{k-1}\right)^2 \ \text{for some} \ n \ge m\right\}.$$

Limiting $h \to \infty$ and then $m \to \infty$ and taking into account conditions (4) and (6) and comes to (59) with $i = 1$. Similarly, one obtains (59) for $i = 2$.

It will be noted that, on the set Ω_h one has $y_{k-1} = \tilde{y}_{k-1}$, $\tau(h) = T(h)$,

$$\frac{1}{\sqrt{h}}\sum_{k=1}^{\tau(h)} y_{k-1}\varepsilon_k = \frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)} \tilde{y}_{k-1}\varepsilon_k.$$

This implies the equation

$$P_\theta\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{\tau(h)} y_{k-1}\varepsilon_k \le t\right) = P_\theta\left(\frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)} \tilde{y}_{k-1}\varepsilon_k \le t\right) + r_\theta(h)$$

where $r_\theta(h)$ is such that

$$\sup_{\theta\in\Theta} |r_\theta(h)| \le \sup_{\theta\in\Theta} P_\theta\left(\Omega^c_h\right) \to 0 \quad \text{as} \quad h \to \infty.$$

Using the presentation

$$\frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)} \tilde{y}_{k-1}\varepsilon_k = \xi_h + \eta_h$$

where

$$\xi_h = \frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)} \tilde{y}_{k-1}\left(\tilde{\varepsilon}_k - E\tilde{\varepsilon}_k\right),$$
$$\eta_h = \frac{1}{\sqrt{h}}\sum_{k=1}^{T(h)} \tilde{y}_{k-1}\left(\tilde{\tilde{\varepsilon}}_k - E\tilde{\varepsilon}_k\right)$$

one can show that

$$P_\theta\left(\xi_h + \eta_h \le t\right) \le P_\theta\left(\xi_h \le t + \Delta\right) + P_\theta\left(|\eta_h| \ge \Delta\right),$$
$$P_\theta\left(\xi_h + \eta_h \le t\right) \ge P_\theta\left(\xi_h \le t - \Delta\right) - P_\theta\left(|\eta_h| \ge \Delta\right),$$

where $\Delta > 0$. Taking into account (58), one gets

$$P_\theta\left(\xi_h + \eta_h \le t\right) - \Phi\left(\frac{t}{v_\theta(\delta)}\right) \le \Phi\left(\frac{t+\Delta}{v_\theta(\delta)}\right) - \Phi\left(\frac{t}{v_\theta(\delta)}\right)$$
$$+ P_\theta\left(\xi_h \le t+\Delta\right) - \Phi\left(\frac{t+\Delta}{v_\theta(\delta)}\right) + P_\theta\left(|\eta_h| \ge \Delta\right)$$
$$\le \omega\left(\Phi; \frac{\Delta}{v_\theta(\delta)}\right) + \rho\left[4\left(\frac{\delta}{v_\theta(\delta)}\right)\right] + P_\theta\left(|\eta_h| \ge \Delta\right)$$

where $\omega(\Phi;\delta)$ is the oscillation of function Φ of radius δ. Similarly, one derives

$$P_\theta\left(\xi_h + \eta_h \le t\right) - \Phi\left(\frac{t}{v_\theta(\delta)}\right) \ge -\omega\left(\Phi; \frac{\Delta}{v_\theta(\delta)}\right) - \rho\left[4\left(\frac{\delta}{v_\theta(\delta)}\right)\right]$$
$$- P_\theta\left(|\eta_h| \ge \Delta\right).$$

Combining these inequalities yields

$$\left|P_\theta\left(\xi_h + \eta_h \le t\right) - \Phi\left(\frac{t}{v_\theta(\delta)}\right)\right| \le \omega\left(\Phi; \frac{\Delta}{v_\theta(\delta)}\right) + \rho\left[4\left(\frac{\delta}{v_\theta(\delta)}\right)\right]$$
$$+ P_\theta\left(|\eta_h| \ge \Delta\right),$$

where

$$P_\theta\left(|\eta_h| \ge \Delta\right) \le \frac{1}{\Delta^2} E_\theta \eta_h^2 = \frac{1}{\Delta^2}\left(1 - v_\theta(\delta)\right).$$

Therefore,

$$\left|P_\theta\left(\frac{1}{\sqrt{h}} \sum_{k=1}^{\tau(h)} y_{k-1}\varepsilon_k \le t\right) - \Phi(t)\right| \le \omega\left(\Phi; \frac{\Delta}{v_\theta(\delta)}\right) + \rho\left[4\left(\frac{\delta}{v_\theta(\delta)}\right)\right]$$
$$+ |r_\theta(h)| + P_\theta\left(|\eta_h| \ge \Delta\right) + \sup_{t\in R}\left|\Phi\left(\frac{\delta}{\sqrt{v_\theta(\delta)}}\right) - \Phi(t)\right|.$$

Taking supremum with respect to θ in both sides of this inequality and limiting $h \to \infty$, $\delta \to 0$ and then $\Delta \to 0$, one arrives at the desired result. □

Acknowledgements Research was supported by RSF, Project No 20-61-47043.

References

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics, 21*(1), 520–533.

Chigansky, P., Kutoyants, Y. A. (2013). Estimation in threshold autoregressive models with correlated innovations. *Annals of the Institute of Statistical Mathematics, 65*(5), 959–992.

Freedman, D. (1983). *Markov chains*. New York: Springer.

Galtchouk, L., Pergamenshchikov, S. (2014). Geometric ergodicity for classes of homogenous Markov chains. *Stochastic Processes and Their Applications, 124*(10), 3362–3391.

Gao, J., Tjostheim, D., Yin, J. (2013). Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics, 172*(1), 1–13.

Lai, T., Siegmund, D. (1983). Fixed accuracy estimation of an autoregressive parameter. *Annals of Statistics, 11*(2), 478–485.

Lai, T., Wei, C. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis, 13*(1), 1–23.

Lee, S., Sriram, T. (1999). Sequential point estimation of parameters in a threshold AR(1) model. *Stochastic Processes and Their Applications, 84*(2), 343–355.

Li, D., Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics, 167*(1), 240–253.

Pergamenshchikov, S. (1992). Asymptotic properties for estimating the parameter in a first-order autoregression. *Theory of Probability and Its Applications, 36*(1), 36–46.

Petruccelli, J., Woolford, S. (1984). A threshold AR(1) model. *Journal of Applied Probability, 21*(2), 270–286.

Pham, D., Chan, K., Tong, H. (1991). Strong consistency of the least squares estimator for a non-ergodic threshold autoregressive model. *Statistica Sinica, 1*(2), 381–369.

Shiryaev, A. (1996). *Probability* (2nd ed.). New York: Springer.

Sriram, T., Laci, R. (2014). Editors's special invited paper: Sequential estimation for time series models. *Sequential Analysis*, *33*(2), 136–157.

Tong, H. (1978). On a threshold model. In Chen, C. (Ed.), *Pattern recognition and signal processing*, pp. 575–586. Alphen aan den Rijn, The Netherlands: Sijthoff & Noordhof.

Yau, C. Y., Tang, C. M., Lee, T. C. M. (2015). Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association, 110*(511), 1175–1186.