Annals of the Institute of Statistical Mathematics (2020) 72:235–264
https://doi.org/10.1007/s10463-018-0689-2

Sequential fixed accuracy estimation for nonstationary autoregressive processes

Victor Konev[1] · **Bogdan Nazarenko**[1]

Received: 30 January 2018 / Revised: 27 July 2018 / Published online: 8 September 2018

Abstract
For an autoregressive process of order p, the paper proposes new sequential estimates for the unknown parameters based on the least squares (LS) method. The sequential estimates use p stopping rules for collecting the data and presumes a special modification the sample Fisher information matrix in the LS estimates. In case of Gaussian disturbances, the proposed estimates have non-asymptotic normal joint distribution for any values of unknown autoregressive parameters. It is shown that in the i.i.d. case with unspecified error distributions, the new estimates have the property of uniform asymptotic normality for unstable autoregressive processes under some general condition on the parameters. Examples of unstable autoregressive models satisfying this condition are considered.

Keywords Unstable autoregressive process · Non-asymptotic distribution of estimates · Sequential least squares method · Uniform asymptotic normality of estimates

1 Introduction

Consider the autoregressive $AR(p)$ model

$$x_k = \theta_1 x_{k-1} + \ldots + \theta_p x_{k-p} + \varepsilon_k, \tag{1}$$

where $\{\varepsilon_k\}$ are independent and identically distributed (i.i.d.) unobservable random errors (noise) with $\mathsf{E}\varepsilon_1 = 0$ and $0 < \mathsf{E}\varepsilon_1^2 = \sigma^2 < \infty$; x_k is the observation with

This study was supported by the ministry of education and science of the Russian Federation, goszadanie no 2.3208.2017/4.6 and in part by RFBR under Grant 16-01-00121.

✉ Victor Konev
vvkonev@mail.tsu.ru

[1] Department of High Mathematics and Mathematical Modelling, Institute of Applied Mathematics and Computer Science, Tomsk State University, 36, Lenina Avenue, Tomsk, Russia 634050

the initial values $x_0, \ldots, x_{-p+1}$ independent of $\{\varepsilon_k\}$ and $\theta_1, \ldots, \theta_p$ are unknown parameters of the model.

Throughout the sequel, we shall assume that the distribution of ε_k is either unspecified or is Gaussian, that is $\varepsilon_k \sim \mathcal{N}(0, \sigma^2)$. Henceforth, for simplicity we set $\sigma^2 = 1$. Model (1), which is a general autoregressive time series, is said to be stable if all roots $\lambda_1, \ldots, \lambda_p$ of the characteristic polynomial

$$\phi(z) = z^p - \theta_1 z^{p-1} - \ldots - \theta_p \tag{2}$$

lie inside the unit circle, that is $|\lambda_i| < 1,\ i = \overline{1, p}$; it is said to be unstable if $\max_{1 \le i \le p} |\lambda_i| = 1$, and purely explosive if $|\lambda_i| > 1,\ i = \overline{1, p}$.

We shall let

$$\Lambda_p = \left\{(\theta_1, \ldots, \theta_p) : \ |\lambda_1| < 1, \ldots, |\lambda_p| < 1\right\},\ \partial\Lambda_p \text{ and } [\Lambda_p] \tag{3}$$

denote, respectively, the parametric stability region of the $AR(p)$ process, its boundary and its closure; $[\Lambda_p] = \Lambda_p + \partial\Lambda_p$.

Defining the p-dimensional vectors

$$\theta = (\theta_1, \ldots, \theta_p)',\ X_k = (x_k, \ldots, x_{k-p+1})',$$

Equation (1) can be written as

$$x_k = X'_{k-1}\theta + \varepsilon_k,\ k = 1, 2, \ldots; \tag{4}$$

(the transpose of any vector or matrix a is indicated by a').

A commonly used estimate for $\theta = (\theta_1, \ldots, \theta_p)'$ by observations $x_1, \ldots, x_n$ is the least squares estimate (LSE) or the maximum likelihood estimate (MLE) in the Gaussian case, given by

$$\widehat{\theta}(n) = G_n^{-1} \sum_{k=1}^{n} X_{k-1} x_k,\ \ G_n = \sum_{k=1}^{n} X_{k-1} X'_{k-1}. \tag{5}$$

Autoregressive models are widely used in engineering applications and time series analysis because they can generate a great variety of processes (see, e.g., Anderson 1971; Box et al. 2008).

Estimate (5) can be represented (Lai and Wei 1985) in the recursive form of the Kalman filter type providing a simple algorithm for its computation.

The statistical properties of $\hat{\theta}(n)$ have been well studied in the literature. The problem of strong consistency of estimate (5) for general autoregressive models, without any assumption on the roots of the characteristic polynomial (2) was solved by Lai and Wei (1983). The theory of asymptotic distributions, under different assumptions on the order p and the range of unknown parameters, has been developed by many authors (see Mann and Wald 1943; Anderson 1959, 1971; Ahtola and Tiao 1987; Brockwell and Davis 1991; Greenwood and Shiryaev 1992; Monsour and Mikulski

1998; Shiryaev and Spokoiny 2000; Liptser and Shiryaev 2001 and references therein). For general stable AR(p) processes, it was established by Anderson (1959) that the limiting distribution of normalized estimate (5) is multivariate standard normal in the case of i.i.d. errors $\{\varepsilon_k\}$.

For unstable autoregressive models, there is no one universal limiting distribution. It is well known (White 1958; Rao 1978; Shiryaev and Spokoiny 2000) that the least squares estimate for the AR(1) model has three different limiting distributions, each demanding its own normalizing factor.

For unstable AR(p) processes, Chan and Wei (1988) proved, by applying the functional central limit theorem approach, that the limiting distribution of (5) can be represented by ratios of certain Brownian functionals. The limiting distributions for the maximum likelihood estimates in the purely explosive Gaussian AR(p) process have been investigated in Anderson (1971), Mikulski and Monsour (1991).

In recent decades, remarkable theoretical advancements in statistical inference for stochastic regression models, including the autoregressive processes, were made by applying the sequential analysis approach. A distinguishing feature of sequential inference methods is that the sample size, they use in the procedures, is not fixed in advance and determined by special stopping rules.

Sequential methods are especially important in problems of constructing confidence intervals (regions of fixed size) for the unknown parameter with a prescribed coverage probability (see Novikov 1972; Lai and Siegmund 1983; Lee 1994; Shiryaev and Spokoiny 2000; Sriram and Iaci 2014) and in problems of constructing sequential point estimates of unknown parameters with prescribed mean square precision (see Borisov and Konev 1977; Konev and Pergamenshchikov 1981; Pergamenshchikov 1992; Konev and Lai 1995; Konev and Pergamenshchikov 1997; Konev and Le Breton 2000; Galtchouk and Konev 2001, 2006).

The present paper has two linked objectives. First, we address the problem of constructing sequential least squares estimates for parameters in AR(p) model with non-asymptotic normal joint distribution under the condition that the errors (ε_k) in (1) form a Gaussian white noise. The second problem is to study asymptotic properties of these estimates for the unstable AR(p) process in the case when $\{\varepsilon_k\}$ is an i.i.d. sequence of random variables with unspecified distribution and to find a general condition on the autoregressive parameters providing the uniform asymptotic normality of estimates.

The key idea of our approach is to use a special transformation of the sample Fisher information matrix in (5) to obtain the sequential estimates with the desired properties.

The paper is organized as follows. In Sect. 2, we construct sequential least square estimates. The exact distribution of these estimates in the case of Gaussian white noise has been found (Theorem 1). The heart of the proof of Theorem 1 is a probabilistic result for square integrable martingales with conditionally Gaussian increments established in Appendix (Theorem 3). In Sect. 3, we prove uniform asymptotic normality of the estimates in case of unstable AR(p) process under a general condition on the set of parameters. Three important examples are considered: AR(1), AR(2) and AR(3) unstable processes (Propositions 1–5). In Sect. 3.4, we consider the case of unstable AR(p) process under the assumption that the roots of the characteristic polynomial (2) are all different and prove uniform asymptotic normality (Proposition 4). In Proposition 5, we establish the uniform asymptotic normality of the proposed estimates in case of stable AR(p) process.

2 Sequential least squares estimates. Distribution in the case of Gaussian noise

In this section, a new sequential sampling scheme for estimating unknown parameters $\theta_1, \ldots, \theta_p$ in an autoregressive model of order p, specified by Equation (1), is developed. The goal is to construct sequential least square estimates with known joint distribution, in the case of Gaussian white noise. To define a sequential counterpart of least square estimate (5), we introduce a system of special stopping rules and do some modifications of the sample Fisher information matrix G_n. We set

$$G_{n,l} = \sum_{k=n+1}^{l} X_{k-1} X'_{k-1}, \tag{6}$$

where $0 \leq n < l < \infty$. For every $h > 0$, we define p stopping times $\tau_i(h)$, $1 \leq i \leq p$,

$$\tau_i(h) = \inf \left\{ n > \tau_{i-1}(h) : \sum_{k=\tau_{i-1}(h)+1}^{n} x_{k-i}^2 \geq h \right\}, \; i = \overline{1, p}, \tag{7}$$

where $\tau_0(h) \equiv 0$; $\inf\{\emptyset\} = +\infty$.

Further we introduce a system of the modified sample Fisher information matrices

$$\widehat{G}_{i,\tau_i(h)} = \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} X_{k-1} X'_{k-1}, \; i = \overline{1, p}, \tag{8}$$

where

$$\beta_{i,k}(h) = \begin{cases} 1 & if\ k < \tau_i(h), \\ \alpha_i(h) & if\ k = \tau_i(h); \end{cases}$$

$\alpha_1(h), \ldots, \alpha_p(h)$ are the correction multipliers ($0 < \alpha_i(h) \leq 1$) defined by the equations

$$\sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)-1} x_{k-i}^2 + \alpha_i(h) x_{\tau_i(h)-i}^2 = h, \; i = \overline{1, p}.$$

The matrix $\widehat{G}_{i,\tau_i(h)}$ can possibly be regarded as the sample Fisher information matrix computed by the observations at the (random) time period $(\tau_{i-1}, \tau_i]$.

By making use of matrices (8), we construct a sequential counterpart of the sample Fisher information matrix (5) as

$$G_p(h) = \|\langle G_p\rangle_{i,j}(h)\|, \quad \langle G_p\rangle_{i,j}(h) = \langle \widehat{G}_{i,\tau_i(h)}\rangle_{ij} \tag{9}$$

where the symbol $\|A_{i,j}\|$ means a matrix having elements $A_{i,j}$ and $\langle A\rangle_{i,j}$ denotes the (i, j)-th element of a matrix A. Let $v(h) = (v_1(h), \ldots, v_p(h))'$ be a vector with the coordinates

$$v_i(h) = \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)}\langle X_{k-1}x_k\rangle_i \tag{10}$$

where $\langle b\rangle_i$ denotes the i-th component of vector b.

Finally we define the sequential least squares estimate for $\theta = (\theta_1, \ldots, \theta_p)'$ as

$$\theta^*(h) = (\theta_1^*(h), \ldots, \theta_p^*(h))' = G_p^{-1}(h)v(h). \tag{11}$$

The matrix $G_p(h)$ is assumed to be invertible for sufficiently large h. This conditions is checked in Proposition 5 for a stable AR(p) model.

Now that the construction of sequential estimate (11) is complete, the natural question arises: whether there are some arguments in favor of this estimate in comparison with the commonly used LSE (5). A certain justification of the more complicated form of the sequential estimate (11) is the fact that, in contrast to (5), it permits one to derive its exact distribution by samples of small volumes, in the case of Gaussian noise.

The desired result may be given as follows.

Theorem 1 *Suppose that $(\varepsilon_i)_{i\geq 1}$ are i.i.d. with the standard Gaussian distribution, $\varepsilon_i \sim \mathcal{N}(0, 1)$ and the estimate for $\theta = (\theta_1, \ldots, \theta_p)'$ be defined by (11). Then for any $\theta \in R^p$ and $h > 0$*

$$\mathrm{Law}\left(\frac{G_p(h)}{\sqrt{h}}(\theta^*(h) - \theta)\right) = \mathcal{N}(O, I_p)$$

that is the standardized deviation of $\theta^*(h)$ with normalization factor (9) has the p-dimensional standard normal distribution; I_p is the unit matrix of order p.

The proof of this theorem, given in Appendix, makes fundamental use of the properties of stopped square integrable martingales with the conditionally Gaussian increments (see Theorem 3 in Appendix).

3 Uniform asymptotic normality of the sequential plan

In this section, we will extend the developed sequential method to the problem of estimating parameters in AR(p) model (1) with unknown distribution of the disturbances $(\varepsilon_k)_{k\geq 1}$. In what follows, the noise $(\varepsilon_k)_{k\geq 1}$ is assumed to be a sequence of independent identically distributed random variables. We consider the case of unstable AR(p) process when all roots $\lambda_1, \ldots, \lambda_p$ of the characteristic polynomials (2) lie on or inside the unit circle, that is $|\lambda_i| \leq 1,\ i = \overline{1, p}$. The goal of this section is to prove the property of uniform asymptotic normality of sequential estimate (11) as h in the stopping rules (7) tends to infinity. To this end, we will need to impose an additional condition on the parametric set.

Let K be a compact subset in the closure $[\Lambda_p]$ of the stability region (3) satisfying the following condition.

Condition 1

$$\sup_{\theta \in K} \sup_{n \geq 1} \|A^n(\theta)\| =: \kappa_p < \infty \tag{12}$$

where $\|\cdot\|$ denotes a matrix norm and

$$A(\theta) = \begin{pmatrix} \theta_1 & \dots\ \theta_p \\ I_{p-1} & 0 \end{pmatrix}. \tag{13}$$

We will establish the following uniform asymptotic normality result.

Theorem 2 *Let $(\varepsilon_k)_{k\geq 1}$ in AR(p) model (1) be a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_k = 0$, and $0 < \mathsf{E}\varepsilon_k^2 = 1$ and $\theta^*(h)$ for $\theta = (\theta_1, \dots, \theta_p)'$ be given by (11).*

Then for any set $K \subset [\Lambda_p]$, satisfying condition 1,

$$\lim_{h\to\infty} \sup_{\theta \in K} \sup_{t \in R^p} \left| P_\theta \left(\frac{G_p(h)}{\sqrt{h}} (\theta^*(h) - \theta) \leq t \right) - \Phi_p(t) \right| = 0 \tag{14}$$

where $t = (t_1, \dots, t_p)'$, $\Phi_p(t)$ is the standard p-dimensional normal distribution function and Λ_p is given by (3).

The proof of this result is shown in Appendix.

Consider some applications of Theorem 2.

3.1 AR(1) model

Let $\{x_k\}_{k\geq 0}$ obey the first-order autoregressive equation

$$x_k = \theta_1 x_{k-1} + \varepsilon_k, \tag{15}$$

where the ε_k's are i.i.d. with $\mathsf{E}\varepsilon_k = 0,\ \mathsf{E}\varepsilon_k^2 = 1$.

In this case, estimate (11) of parameter θ has the form

$$\theta^*(h) = \frac{1}{\widetilde{h}} \left(\sum_{k=1}^{\tau(h)-1} x_{k-1}x_k + \sqrt{\alpha(h)}\, x_{\tau(h)-1} x_{\tau(h)} \right), \tag{16}$$

where

$$\tau(h) = \inf\left\{ n \geq 0 : \sum_{k=1}^{n} x_{k-1}^2 \geq h \right\}, \quad \widetilde{h} = \sum_{k=1}^{\tau(h)-1} x_{k-1}^2 + \sqrt{\alpha(h)}\, x_{\tau(h)-1}^2; \tag{17}$$

with the correction factor $0 < \alpha(h) \leq 1$ determined by the equation

$$\sum_{k=1}^{\tau(h)-1} x_{k-1}^2 + \alpha(h) x_{\tau(h)-1}^2 = h.$$

Similar to the sequential least squares estimate proposed by Lai and Siegmund (1983), Borisov and Konev (1977), estimate (16) has the property of uniform asymptotic normality for the unstable AR(1) process.

Proposition 1 *Let $(\varepsilon_k)_{k\geq 1}$ in (15) be an i.i.d. sequence of random variables with* $\mathsf{E}\varepsilon_k = 0$ *and* $\mathsf{E}\varepsilon_k^2 = 1$ *and* $\theta^*(h)$ *be defined by (16). Then*

$$\lim_{h\to\infty} \sup_{|\theta|\leq 1} \sup_{-\infty<t<\infty} \left| P_\theta \left(\frac{\widetilde{h}}{\sqrt{h}} (\theta^*(h) - \theta) \leq t \right) - \Phi(t) \right| = 0. \tag{18}$$

The proof of this result proceeds along the lines of Theorem 2.1 in Lai and Siegmund (1983) and is omitted.

If $\{\varepsilon_k\}_{k\geq 1}$ in (15) is a sequence of i.i.d. random variables with standard normal distribution, i.e., $\varepsilon_k \sim \mathcal{N}(0, 1)$, then, by applying Theorem 1 in Sect. 2, we get an exact non-asymptotic distribution of estimate (16): for any $h > 0$ and $-\infty < \theta < \infty$

$$P_\theta \left(\frac{\widetilde{h}}{\sqrt{h}} \left(\theta^*(h) - \theta\right) \leq t \right) = \Phi(t), \quad -\infty < t < \infty.$$

Some extensions of this result can be found in Konev (2016).

3.2 The case of unstable $AR(2)$ process

Consider an $AR(2)$ process satisfying the equation

$$x_k = \theta_1 x_{k-1} + \theta_2 x_{k-2} + \varepsilon_k, \quad k = 1, 2, \ldots \tag{19}$$

with $x_0 = x_{-1} = 0$. In this case, the closure of stability region Λ_2, given by (3), is the triangle:

$$[\Lambda_2] = \{\theta = (\theta_1, \theta_2)' : \ -1 + \theta_2 \leq \theta_1 \leq 1 - \theta_2, \ -1 \leq \theta_2 \leq 1\}. \tag{20}$$

Lemma 1 *Condition 1 holds for AR(2) process (19) for any compact set*

$$K \subset [\Lambda_2] \setminus \{(-2, -1), (2, -1)\}. \tag{21}$$

Proof. One can check that the powers of the matrix

$$A(\theta) = \begin{bmatrix} \theta_1 & \theta_2 \\ 1 & 0 \end{bmatrix}$$

are given by the formulae

$$A^n = \frac{1}{z_1 - z_2}\begin{pmatrix} z_1^{n+1} - z_2^{n+1} & -(z_1^n - z_2^n)z_1z_2 \\ z_1^n - z_2^n & -(z_1^{n-1} - z_2^{n-1})z_1z_2 \end{pmatrix}$$

if roots of the characteristic polynomial $\varphi(z) = z^2 - \theta_1 z - \theta_2$ are real and by

$$A^n = \frac{a^{n-1}}{sin\varphi}\begin{pmatrix} \sin(n+1)\varphi & -a^2 \sin n\varphi \\ \sin n\varphi & -\sin(n-1)\varphi \end{pmatrix}$$

if the roots are complex, that is $z_1 = ae^{i\varphi}$ and $z_2 = ae^{-i\varphi}$. Using these equations, one can easily verify (12). This completes the proof of Lemma 1. □

By applying Theorem 2 and Lemma 1, we arrive at the following result.

Proposition 2 *Let $(\varepsilon_n)_{n\geq 1}$ in $AR(2)$ model (19) be a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$ and the estimate $\theta^*(h)$ for $\theta = (\theta_1, \theta_2)'$ be given by (11) with $p = 2$.*

Then for any compact set K, satisfying (21),

$$\lim_{h\to\infty} \sup_{\theta\in K} \sup_{t\in R^2} \left| P_\theta\left(\frac{G_2(h)}{\sqrt{h}}(\theta^*(h) - \theta) \leq t\right) - \Phi_2(t)\right| = 0 \tag{22}$$

where $t = (t_1, t_2)'$, $\Phi_2(t) = \Phi(t_1)\Phi(t_2)$, Φ is the standard normal distribution function.

3.3 The case of unstable AR(3) process

Consider an $AR(3)$ process satisfying the equation

$$x_k = \theta_1 x_{k-1} + \theta_2 x_{k-2} + \theta_3 x_{k-3} + \varepsilon_k, \; k \geq 1, \tag{23}$$

with $x_0 = x_{-1} = x_{-2} = 0$. In this case, the boundary of the stability region Λ_3, defined by (3) with $p = 3$, is the union of three surfaces

$$\delta\Lambda_3 = S_1 \bigcup S_2 \bigcup S_3$$

where S_1 is the triangle with apexes

$$B_1 = (1, 1, -1), \; B_2 = (-1, 1, 1), \; B_3 = (-3, -3, -1),$$

S_2 is the triangle with apexes B_1, B_2 and $B_4 = (3, -3, 1)$,

$$S_3 = \{(\theta_1, \theta_2, \theta_3) : \; \theta_1 = a + 2\cos\varphi, \; \theta_2 = -1 - 2a\cos\varphi, \; \theta_3 = a, \; |a| \leq 1, \; 0 \leq \varphi \leq \pi\}.$$

Lemma 2 *Condition 1 holds for $AR(3)$ process (23) for any compact set K satisfying the inclusion*

$$K \subset [\Lambda_3] \setminus ([B_1, B_4] \bigcup [B_2, B_3]) \tag{24}$$

where $[\Lambda_3]$ is the closure of the region Λ_3, $[B_1, B_4]$ and $[B_2, B_3]$ are the segments connecting the points B_1, B_4 and B_2, B_3, respectively.

For the sake of brevity, we omit the proof of Lemma 2.

Remark 1 It should be observed that the characteristic polynomial $\varphi_3(z) = z^3 - \theta_1 z^2 - \theta_2 z - \theta_3$ has multiple roots on the unit circle for all points $\theta = (\theta_1, \theta_2, \theta_3)$ in the segments $[B_1, B_4]$ and $[B_2, B_3]$.

Proposition 3 *Let $(\varepsilon_n)_{n\geq 1}$ in $AR(3)$ model (23) be a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$ and the estimate $\theta^*(h) = (\theta_1^*, \theta_2^*, \theta_3^*)'$ be given by (11) with $p = 3$.*

Then for any compact set K, satisfying (24)

$$\lim_{h\to\infty} \sup_{\theta\in K} \sup_{t\in R^3} \left| P_\theta \left(\frac{G_3(h)}{\sqrt{h}} (\theta^*(h) - \theta) \leq t \right) - \Phi_3(t) \right| = 0. \tag{25}$$

This result is a direct consequence of Lemma 2 and Theorem 2.

3.4 The case of unstable AR(p) process

Consider the autoregressive AR(p) model specified by Equation (1) with zero initial values $x_0 = x_{-1} = \ldots = x_{1-p} = 0$. Because the calculations in terms of the general Jordan canonical form are laborious, we shall check condition (12) under the assumption that the characteristic roots $\lambda_i = \lambda_i(\theta)$ of matrix (13) are all different. Moreover, we suppose that the vector of unknown parameters $\theta = (\theta_1, \ldots, \theta_p)'$ belongs to the set

$$K_\delta = \{\theta \in [\Lambda_p] : \ |\lambda_i(\theta) - \lambda_j(\theta)| \geq \delta \text{ for } i \neq j; \ i, j = \overline{1, n}\} \tag{26}$$

for some positive number δ. Then matrix (13) can be written as

$$A(\theta) = T(\theta) D(\theta) T^{-1}(\theta)$$

where $D(\theta)$ is a diagonal matrix, $D(\theta) = \mathrm{diag}\{\lambda_1(\theta), \ldots, \lambda_p(\theta)\}$ and $T(\theta)$ is the Vandermonde matrix

$$T = \begin{bmatrix} \lambda_1^{p-1}(\theta) & \ldots & \lambda_p^{p-1}(\theta) \\ \ldots & \ldots & \ldots \\ \lambda_1(\theta) & \ldots & \lambda_p(\theta) \\ 1 & \ldots & 1 \end{bmatrix}.$$

Noting that

$$\inf_{\theta\in K_\delta}|T(\theta)| = \inf_{\theta\in K_\delta}\left|\prod_{k=1}^{p-1}\prod_{i=k+1}^{p}(\lambda_k(\theta)-\lambda_i(\theta))\right| \geq \delta^{\frac{p(p-1)}{2}}$$

one gets

$$\sup_{\theta\in K_\delta}\sup_{n\geq 1}\|A^n(\theta)\| \leq \sup_{\theta\in K_\delta}\|T(\theta)\|\cdot\|T^{-1}(\theta)\| < \infty.$$

Therefore, condition (12) is satisfied and applying Theorem 2, we obtain the following result

Proposition 4 *Let $(\varepsilon)_{n\geq 1}$ in AR(p) model (1) be a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$ and $\theta = (\theta_1, \dots, \theta_p)' \in K_\delta$ where K_δ is defined in (26).*

Then estimate (11) is asymptotically uniformly normal on the set K_δ, that is

$$\lim_{h\to\infty}\sup_{\theta\in K_\delta}\sup_{t\in R^p}\left|P_\theta\left(\frac{G_p(h)}{\sqrt{h}}(\theta^*(h)-\theta)\leq t\right)-\Phi_p(t)\right| = 0. \tag{27}$$

It will be noted that in the case of stable autoregressive process (1) the property of the uniform asymptotic normality of sequential estimate (11) holds true without any restriction on multiplicity of the roots of the characteristic polynomials (2). The result is as follows.

Proposition 5 *Let $AR(p)$ model (1) be stable, $(\varepsilon_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$ and the estimate $\theta^*(h)$ for $\theta = (\theta_1, \dots, \theta_p)'$ be given by (11).*

Then for any compact set $K \subset \Lambda_p$

$$\lim_{h\to\infty}\sup_{\theta\in K}\sup_{t\in R^p}\left|P_\theta\left(\frac{G_p(h)}{\sqrt{h}}(\theta^*(h)-\theta)\leq t\right)-\Phi_p(t)\right| = 0 \tag{28}$$

where $t = (t_1, \dots, t_p)'$, $\Phi_p(t) = \Phi(t_1)\cdots\Phi(t_p)$, Φ is the standard normal distribution function.

Moreover,

$$\lim_{h\to\infty}\frac{G_p(h)}{h} = \frac{F}{\langle F\rangle_{11}}, \quad a.s., \tag{29}$$

$$\lim_{h\to\infty}\frac{\tau_p(h)}{h} = \frac{p}{\langle F\rangle_{11}}, \quad a.s. \tag{30}$$

where F is a positive definite matrix given in (63).

The proof of this proposition is given in Appendix.

4 Monte Carlo results

In this section, we report the results of Monte Carlo experiments, carried out for the autoregressive model AR(2), to compare the performance of the proposed sequential estimates (11) with the fixed sample size LSE counterparts given by (5). In all simulations of the $AR(2)$ process, obeying Equation (19), the distribution of the residuals $\{\varepsilon_k\}$ was taken to be $\mathcal{N}(0, 1)$ and $x_0 = 0,\ x_{-1} = 0$. According to (11), the sequential estimate $\theta^*(h) = (\theta_1^*(h), \theta_2^*(h))'$ for $\theta = (\theta_1, \theta_2)'$ has the form

$$\theta^*(h) = G_2^{-1}(h) v(h). \tag{31}$$

Here $G_2(h)$ is the modified sample Fisher information matrix of size 2×2 with the elements

$$\langle G_2(h)\rangle_{11} = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} x_{k-1}^2, \quad \langle G_2(h)\rangle_{12} = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} x_{k-1} x_{k-2},$$
$$\langle G_2(h)\rangle_{21} = \sum_{k=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,k}(h)} x_{k-2} x_{k-1}, \quad \langle G_2(h)\rangle_{22} = \sum_{k=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,k}(h)} x_{k-2}^2.$$

and $v(h) = (v_1(h), v_2(h))'$ is the vector with coordinates

$$v_1(h) = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} x_{k-1} x_k, \ v_2(h) = \sum_{k=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,k}(h)} x_{k-2} x_k,$$

where $\tau_1(h)$ and $\tau_2(h)$ are stopping rules given by

$$\tau_1(h) = \inf\left\{n \geq 1 : \sum_{k=1}^{n} x_{k-1}^2 \geq h\right\}, \ \tau_2(h)$$
$$= \inf\left\{n > \tau_1(h) : \sum_{k=\tau_1(h)+1}^{n} x_{k-2}^2 \geq h\right\}.$$

The values of parameters θ_1, θ_2, used in the simulations, are given in the first two columns of Tables 1, 2 and belong to set (20).

The experiment to access the performance of the sequential estimate (11) consisted of 10000 replications of the sequential procedure (31) with thresholds $h = 300$ and $h = 500$ for each pair of parameters (θ_1, θ_2). The columns of Table 1 with the headings $\theta_1^*,\ \theta_2^*$ and $\mathsf{E}\tau_2$ report, respectively: θ_1^* is the mean of the sequential estimate $\theta_1^*(h)$ for parameter θ_1, $\theta_2^*(h)$ is that for θ_2 and $\mathsf{E}\tau_2$ is the mean duration of sequential procedure (31), based on the results of 10 000 replications. To assess the accuracy of the least squares estimates (5) for $p = 2$ in the fixed sample case, the sample size n was chosen so that $\mathsf{E}_\theta \tau_2 \approx n$ for each $\theta = (\theta_1, \theta_2)'$. The last two columns with the headings $\widehat{\theta}_1$ and $\widehat{\theta}_2$ report the means of the LS estimates based on 10 000 replications.

Table 1 Averages of sequential LSE and LSE with fixed sample size (10^4 replications)

$h = 300$						
θ_1	θ_2	θ_1^*	θ_2^*	$\mathsf{E}\tau_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
− 1.4	− 1	− 1.458	− 1.083	48	− 1.288	−0.893
− 0.6	− 1	− 0.576	− 0.906	62	− 0.562	−0.933
0	− 1	0.002	− 0.899	64	− 0.003	− 0.931
0.6	− 1	0.577	− 0.916	62	0.572	− 0.923
− 1.2	− 0.4	− 1.314	− 0.531	143	− 1.185	− 0.403
0.6	− 0.4	0.602	−0.398	415	0.595	−0.398
− 0.8	0	−0.860	−0.066	223	−0.790	−0.008
0.2	0	0.203	0.001	581	0.197	− 0.002
0	0.4	0.002	0.397	506	− 0.001	0.394
$h = 500$						
θ_1	θ_2	θ_1^*	θ_2^*	$\mathsf{E}\tau_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
− 1.4	− 1	− 1.439	− 1.056	62	− 1.331	− 0.927
− 0.6	− 1	− 0.579	− 0.924	79	− 0.571	− 0.94
0	− 1	0.000	− 0.924	80	− 0.002	− 0.943
0.6	− 1	0.583	− 0.936	79	0.581	− 0.950
0.2	− 0.6	0.198	− 0.595	385	0.199	− 0.594
− 1.2	− 0.4	−1.281	− 0.490	229	− 1.190	− 0.399
0.6	− 0.4	0.597	− 0.397	691	0.599	− 0.400
− 0.8	0	− 0.822	− 0.027	373	− 0.794	− 0.005
0.2	0	0.203	0.001	962	0.201	− 0.002
0	0.4	− 0.002	0.395	846	− 0.003	0.393

The values in Table 1 indicate that the performance of the fixed sample size procedure (5) is close to that of the sequential least squares estimates (31). It will be noted that both estimators for the fixed sample case and those for the sequential case perform also well when the process $AR(2)$ is unstable ($\theta_2 = -1$).

The simulation study included as well testing the hypothesis that the joint distribution of the sequential estimates $(\theta_1^*(h), \theta_2^*(h))'$ is normal. To this end, we used the tests of Mardia (1974), Henze and Zirkler (1990), Royston (1983) and the multivariate Shapiro–Wilk test which were realized , up to standard, in the program package R. The experiment consisted of 200 replications of sequential procedure (31) with $h = 300$ for each pair of parameters (θ_1, θ_2) tabulated in the first two columns of Table 2.

The computed p-values of all tests are reported in columns 3–6 of Table 2. It will be noted that each result for Mardia test consists of two p-values: p-value of the skewness statistic and p-value of the kurtosis statistic.

For the confidence level $\alpha = 0.05$, the results of tests can be interpreted as follows: the hypothesis is accepted if the corresponding p-value exceeds 0.05. For example, the first row in Table 2 indicates that, in case of $\theta_1 = 0.3$ and $\theta_2 = -0.1$, one accepts the hypothesis of bivariate normality of $(\theta_1^*(h), \theta_2^*(h))'$ by applying any of the tests of Mardia, Henze–Zirkler and Royston. The multivariate test of Shapiro-Wilk rejects

Table 2 Testing hypothesis of bivariate normality of sequential LSE, $h = 300$ (200 replications)

θ_1	θ_2	Mardia's test	Henze–Zirkler's test	Royston's test	Multivariate Shapiro–Wilk test
0.3	− 0.1	0.083/0.667	0.464	0.469	0.022
0	0.5	0.371/0.131	0.226	0.058	0.161
− 0.2	− 0.8	0.116/0.982	0.369	0.062	0.307
1.2	− 0.9	0.589/0.380	0.971	0.723	0.405
− 0.2	− 0.1	*0.037*/0.808	0.182	0.047	0.701
0.2	0.7	0.339/0.378	0.770	0.366	0.403
0	1	0.123/0.287	0.916	0.732	0.789
1	0	0.229/0.292	0.284	0.080	0.663
0.3	− 1	0.104/0.719	0.431	0.132	0.664

Table 3 Sample frequencies of $G_2(h)(\theta^*(h) - \theta)\sqrt{h}$ hitting the area $\{(x, y) : \ x \leq -0.3; \ y \leq 1.7\}$ for normal noises and different values of h

(θ_1, θ_2)	(0,1)	(− 1,0)	(1,0)	(0,− 1)	(1.9,− 1)	(− 1.9,− 1)
$h = 20$	0.3649	0.3549	0.3671	0.3641	0.3472	0.3027
$h = 50$	0.3679	0.3614	0.3619	0.3682	0.3642	0.3322
$h = 200$	0.3577	0.3675	0.3654	0.3675	0.3647	0.3626
$h = 500$	0.3611	0.3601	0.3633	0.3647	0.3626	0.3611

Table 4 Sample frequencies of $G_2(h)(\theta^*(h) - \theta)\sqrt{h}$ hitting the area $\{(x, y) : \ x \leq -0.3; \ y \leq 1.7\}$ for skew t-distribution of noises and different values of h

(θ_1, θ_2)	(0,1)	(− 1,0)	(1,0)	(0,− 1)	(1.9,− 1)	(− 1.9,− 1)
$h = 20$	0.3419	0.336	0.3271	0.3223	0.314	0.2962
$h = 50$	0.344	0.3318	0.3347	0.3251	0.3149	0.3015
$h = 200$	0.3467	0.3523	0.3412	0.3393	0.3392	0.3309
$h = 1000$	0.3509	0.3641	0.3455	0.3465	0.3415	0.3326

the hypothesis. Table 2 shows that the performance of the sequential estimates agrees with the statement of Theorem 1.

In Tables 3, 4, we compare the performance of sequential procedure (31) for the autoregressive model AR(2) in two cases when the distribution of the disturbances ε_k in (19) is normal and when it is a skewed t-distribution. We conduct a simulation study to assess the property of the uniform asymptotic normality of the standardized deviation of estimate (31)

$$G_2(h)(\theta^*(h) - \theta)/\sqrt{h}. \tag{32}$$

The experiment, for each type of noises, consisted of 10000 replications of the procedure (31) for different values of $\theta = (\theta_1, \theta_2)'$ given in the first rows of Tables 3, 4 and those of parameter h indicated in the first columns. It will be noted all values of

$\theta = (\theta_1, \theta_2)'$ lie on the boundary $\partial \Lambda_2$ of stability region (20). Each figure in Table 3 is the corresponding frequency count of the number of times vector (32) hits the region

$$S = \{(x, y) : \ x \leq -0.3; \ \ y \leq 1.7\} \tag{33}$$

in the case of normal noises. Note that for the bivariate standard normal vector $P(S) = 0.3648$.

Table 4 refers to the case of the skewed t-distribution. The figures in the rows of Tables 3, 4 indicate that sequential estimate (31) performs well, and the sample probabilities are close to the true probability value.

However, the convergence in the case of t-distribution becomes slower with respect to h and greater values of the threshold h in the stopping rules are needed to ensure the desired approximation. One can remark also that in both cases the quality of sequential estimates deteriorates for the point $\theta = (-1.9, -1)$ near the corner point $(-2, -1)$ of stability region (20) which corresponds to the root on the unit circle with multiplicity 2.

Remark 2 When modeling AR(2) model with skewed t-distribution of disturbances, we used the rST-function (with $df = 3$, $skew = 17$) in the betategarch library which is built in the R-package. The disturbance density was explicitly specified as

$$p_\gamma(x) = \frac{2}{\gamma + \frac{1}{\gamma}} \left\{ f_{df} \left(\frac{x}{\gamma} \right) I_{[0,\infty)}(x) + f_{df}\,(x\gamma)\, I_{(-\infty,0)}(x) \right\}$$

where f_{df} was a t-distribution with $df = 3$ degrees of freedom. The value of γ was chosen so that the skewness of p_γ would be equal to parameter $skew = 17$. Then the corresponding centering and normalization operations were applied.

5 Concluding remarks

In this paper, we have primarily focused on two issues concerning sequential estimation of parameters in autoregressive process. Our first issue is related to the $AR(p)$ model (1) with Gaussian disturbances (ε_k). We propose a new construction of sequential estimates on the basis of the LSE estimates. An important property of these estimates is that they have a non-asymptotic normal joint distribution for any values of unknown parameters $(\theta_1, \ldots, \theta_p)' \in R^p$ (Theorem 1).

The second issue is related to the unstable $AR(p)$ process when the noise distribution is unspecified. Under general condition (12) on the admissible parameter set, we prove uniform asymptotic normality of the proposed estimates (Theorem 2). This condition has been checked for $AR(1)$, $AR(2)$ and $AR(3)$ models and also for a general autoregressive process when all roots of the characteristic polynomial are simple.

Acknowledgements The authors are grateful to the Associate Editor and to an anonymous referee for their helpful comments which improved this paper.

A Additional probabilistic result for the square integrable martingales with conditionally Gaussian increments

In order to prove Theorem 1, we will establish first the following probabilistic result for the square integrable martingales with conditionally Gaussian increments.

Theorem 3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_k)_{k\geq 0}$. Let $(M_k^l, \mathcal{F}_k)_{k\geq 0}$, $l = \overline{1, p}$, be a family of p square integrable martingales with the quadratic characteristics $\{< M^{(l)} >_n\}_{n\geq 1}$ such that*

(a) $P(< M^{(l)} >_\infty = +\infty) = 1,\ l = \overline{1, p}$;
(b) $\mathrm{Law}(\Delta M_k^{(l)} | \mathcal{F}_{k-1}) = \mathcal{N}(0, \sigma_l^2(k-1)),\ k = 1, 2, \ldots,\ l = \overline{1, p}$, *i.e., the $\mathcal{F}_{k-1}$-conditional distribution of $\Delta M_k^l = M_k^{(l)} - M_{k-1}^{(l)}$ is Gaussian with parameters 0 and $\sigma_l^2(k-1) = \mathsf{E}((\Delta M_k^{(l)})^2 | \mathcal{F}_{k-1})$.*
For every $h > 0$, define the sequence of stopping times

$$\tau_l = \tau_l(h) = \inf\left\{ n > \tau_{l-1} : \sum_{k=\tau_{k-1}(h)+1}^{n} \sigma_l^2(j-1) \geq h \right\},\ j = \overline{1, p} \quad (34)$$

where $\tau_0 = \tau_0(h) = 0$, $\inf\{\emptyset\} = +\infty$, and the sequence of random variables

$$m_l(h) = \frac{1}{\sqrt{h}} \sum_{k=\tau_{l-1}+1}^{\tau_l} \sqrt{\beta_k(h,l)}\Delta M_k^{(l)},\ l = 1, \ldots, p,$$

where

$$\beta_k(h,l) = \begin{cases} 1 & if\ \tau_{l-1}(h) < k < \tau_l(h), \\ \alpha_l(h) & \qquad\qquad if\ k = \tau_l(h); \end{cases}$$

and $\alpha_l(h)$ are correcting factors, $0 < \alpha_j(h) \leq 1$ determined by the equations

$$\sum_{k=\tau_{l-1}+1}^{\tau_l-1} \sigma^2(k-1) + \alpha_j(h)\sigma_l^2(\tau_l(h)-1) = h.$$

Then, for any $h > 0$, the vector $m(h) = (m_1(h), \ldots, m_p(h))'$ has p-dimensional standard normal distribution, that is $m(h) \sim \mathcal{N}(0, I_p)$.

We will show that the characteristic function of the random vector $m(h) = (m_1(h), \ldots, m_p(h))'$ has the form

$$\varphi(u) = \varphi(u_1, \ldots, u_p) = \mathsf{E}\exp\left(i \sum_{l=1}^{p} m_l(h) u_l \right) = e^{-\frac{1}{2}\sum\limits_{l=1}^{p} u_l^2}$$

for any $u = (u_1, \ldots, u_p) \in R^p$.

Let $\mathcal{F}_{\tau_i}$ denote a $\sigma-$algebra of the events prior to stopping time τ_i, that is

$$\mathcal{F}_{\tau_i} = \{A \in \mathcal{F}_\infty : \quad A \cap (\tau_i \le k) \in \mathcal{F}_k \text{ for every } k \ge 0\}$$

where $\mathcal{F}_\infty = \sigma\left(\cup_{k\ge 0}\mathcal{F}_k\right)$. The family of $\sigma-$algebras $\{\mathcal{F}_{\tau_i}\}_{0\le i\le p}$ is non-decreasing, that is

$$\mathcal{F}_{\tau_0} \subset \mathcal{F}_{\tau_1} \subset \ldots \subset \mathcal{F}_{\tau_p}.$$

Since

$$\varphi(u) = \mathsf{E}\left\{\exp\left(i\sum_{l=1}^{p-1} m_l(h)u_l\right)\mathsf{E}\left(e^{im_p(h)u_p}\middle|\mathcal{F}_{\tau_p-1}\right)\right\},$$

one has to verify that

$$\mathsf{E}\left\{e^{im_1(h)u_1}\right\} = e^{-\frac{u_1^2}{2}}, \tag{35}$$

$$\mathsf{E}\left[e^{im_l(h)u_l}\middle|\mathcal{F}_{\tau_l-1}\right] = e^{-\frac{u_l^2}{2}},\ l = \overline{2,p}. \tag{36}$$

We introduce the sequence of truncated stopping times $\overline{\tau}_1(h,N) = \tau_1(h)\wedge N,\ N = 1, 2, \ldots$ and denote

$$\xi_N(h) = \frac{1}{\sqrt{h}}\sum_{k=1}^{\overline{\tau}_1(h,N)} \sqrt{\beta_k(h,1)}\Delta M_k^{(1)}.$$

Noting that

$$\lim_{N\to\infty} \xi_N(h) = m_1(h) \ \ a.s.,$$

one has

$$\mathsf{E}e^{im_1(h)u_1} = \lim_{N\to\infty} \mathsf{E}e^{iu_1\xi_N(h)}.$$

Further we use the equation

$$\mathsf{E}e^{iu_1\xi_N(h)} = e^{-\frac{u_1^2}{2}}\mathsf{E}e^{\xi_N^{(1)}(h,u_1)} + R_N, \tag{37}$$

where

$$R_N = \mathsf{E}e^{\xi_N^{(1)}(h,u_1)}\left(e^{-\xi_N^{(2)}(h,u_1)} - e^{-\frac{u_1^2}{2}}\right).$$

$$\xi_N^{(1)}(h, u_1) = \sum_{k=1}^{N} \left[\frac{i u_1 \sqrt{\beta_k(h,1)} \Delta M_k^{(1)}}{\sqrt{h}} \chi_{\{k \le \tau(h)\}} + \frac{u_1^2 \beta_k(h,1) \sigma_1^2(k-1)}{2h} \chi_{\{k \le \tau_1(h)\}} \right],$$

$$\xi_N^{(2)}(h, u_1) = \frac{u_1^2}{2h} \sum_{k=1}^{N} \beta_k(h,1) \sigma_1^2(k-1) \chi_{\{k \le \tau_1(h)\}}. \tag{38}$$

By the definition of stopping time $\tau_1(h)$ in (38), one obtains

$$\lim_{N \to \infty} \xi_N^{(2)}(h, u_1) = \frac{u_1^2}{2}.$$

Applying the theorem on dominated convergence yields

$$\lim_{N \to \infty} R_N = 0. \tag{39}$$

From here and (37), we come to (35).
Now we check (36). We have

$$\mathsf{E}\left[e^{i m_l(h) u_l} \middle| \mathcal{F}_{\tau_{l-1}} \right] = \lim_{n \to \infty} \mathsf{E}\left[e^{i m_l(h) u_l} \middle| \mathcal{F}_{\tau_{l-1} \wedge n} \right],$$

$$\begin{aligned} \mathsf{E}\left[e^{i m_l(h) u_l} \middle| \mathcal{F}_{\tau_{l-1} \wedge n} \right] &= \sum_{t=0}^{n} \mathsf{E}\left[e^{i m_l(h) u_l} \middle| \mathcal{F}_t \right] \chi_{\{\tau_{l-1} = t\}} \\ &= \sum_{t=0}^{n} \mathsf{E}\left[e^{\xi(h,l,t)} \middle| \mathcal{F}_t \right] \chi_{\{\tau_{l-1} = t\}}; \end{aligned} \tag{40}$$

where

$$\xi(h, l, t) = \frac{i}{\sqrt{h}} \sum_{k=t+1}^{\tau_l} \sqrt{\beta_k(h,l)} \Delta M_k^{(l)} u_l.$$

Introducing truncated stopping times

$$\tau_l \wedge N, \quad N = t+1, t+2, \ldots,$$

and the sequence of random variables

$$\xi_N(h, l, t) = \frac{i}{\sqrt{h}} \sum_{k=t+1}^{\tau_l \wedge N} \sqrt{\beta_k(h,l)} \Delta M_k^{(l)} u_l, \quad N = t+1, \ldots$$

and taking into account that

$$\lim_{N \to \infty} \xi_N(h, l, t) = \xi(h, l, t) \; a.s.,$$

we get

$$\mathsf{E}\left[e^{\xi(h,l,t)}\Big|\mathcal{F}_t\right]=\lim_{N\to\infty}\mathsf{E}\left[e^{\xi_N(h,l,t)}\Big|\mathcal{F}_t\right]=e^{-\frac{u_l^2}{2}}. \tag{41}$$

Further we represent $\xi_N(h,l,t)$ as

$$\xi_N(h,l,t)=\xi_N^{(1)}(h,l,t)-\xi_N^{(2)}(h,l,t),$$

where

$$\xi_N^{(1)}(h,l,t)=\sum_{k=t+1}^{N}\left[\frac{iu_l\sqrt{\beta_k(h,l)}}{\sqrt{h}}\Delta M_k^{(l)}\chi_{\{k\le\tau_l\}}+\frac{u_l^2\beta_k(h,l)}{2h}\sigma_l^2(k-1)\chi_{\{k\le\tau_l\}}\right],$$

$$\xi_N^{(2)}(h,l,t)=\frac{u_l^2}{2h}\sum_{k=t+1}^{N}\beta_k(h,l)\sigma_l^2(k-1)\chi_{\{k\le\tau_l(h)\}}.$$

Then

$$\mathsf{E}\left[e^{\xi_N(h,l,t)}\Big|\mathcal{F}_t\right]=e^{-\frac{u_l^2}{2}}\mathsf{E}\left[e^{\xi_N^{(1)}(h,l,t)}\Big|\mathcal{F}_t\right]+R_N, \tag{42}$$

where

$$R_N=\mathsf{E}\left\{e^{\xi_N^{(1)}(h,l,t)}\left[e^{-\xi_N^{(2)}(h,l,t)}-e^{-\frac{u_l^2}{2}}\right]\Big|\mathcal{F}_t\right\}.$$

Noting that

$$\mathsf{E}\left[e^{\xi_N^{(1)}(h,l,t)}\Big|\mathcal{F}_t\right]=\mathsf{E}\left[e^{\xi_{N-1}^{(1)}(h,l,t)}\Big|\mathcal{F}_t\right]=\ldots=1,$$

and tending $N\to\infty$ in (41) one gets

$$\mathsf{E}\left[e^{\xi(h,l,t)}\Big|\mathcal{F}_t\right]=e^{-\frac{u_l^2}{2}}.$$

In view of (39)

$$\mathsf{E}\left[e^{im_l(h)u_l}|\mathcal{F}_{\tau_{l-1}\wedge n}\right]=e^{-\frac{u_l^2}{2}}\chi_{\{\tau_{l-1}\le n\}}.$$

Limiting $n\to\infty$ we arrive at the desired results (35). Theorem 3 is proven. □

B Proofs of main theorems

B.1 Proof of Theorem 1

Substituting x_k from (1) in (10) yields

$$\begin{aligned}
v_i(h) &= \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} \langle X_{k-1}(X'_{k-1}\theta + \varepsilon_k)\rangle_i \\
&= \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} \langle X_{k-1}X'_{k-1}\theta\rangle_i + \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} \langle X_{k-1}\rangle_i \varepsilon_k \\
&= \left\langle \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} X_{k-1}X'_{k-1}\theta \right\rangle_i + \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} x_{k-i}\varepsilon_k \\
&= \langle \widehat{G}_{i,\tau_i(h)}\theta\rangle_i + \eta_i(h) = \langle G_p(h)\theta\rangle_i + \eta_i(h),
\end{aligned}$$

where $\eta_i(h) = \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{i,k}(h)} x_{k-i}\varepsilon_k$.

Therefore

$$v(h) = G_p(h)\theta + \eta(h), \tag{43}$$

where $\eta(h) = (\eta_1(h), \ldots, \eta_p(h))'$.

Combining (11) and (43), one obtains

$$G_p(h)(\theta^*(h) - \theta) = \eta(h).$$

It remains to show that the random vector $\eta(h)/\sqrt{h}$ has p-dimensional standard normal distribution. To this end, we apply Theorem 3. First we introduce the natural filtration $(\mathcal{F}_k)_{k\geq 0}$ of process (1) defined as

$$\begin{aligned}
\mathcal{F}_0 &= \sigma(x_0, \ldots, x_{1-p}), \\
\mathcal{F}_k &= \sigma(x_0, \ldots, x_{1-p}; \varepsilon_1, \ldots, \varepsilon_k), k \geq 1.
\end{aligned} \tag{44}$$

It will be noted that the random variables $\{\tau_l(h)\}_{1\leq l\leq p}$, defined in (7), are stopping times with respect to this filtration for every $h > 0$.

Further we need the following p stochastic processes $\{M_t^{(l)}\}_{t\geq 0}$ defined as

$$\begin{aligned}
M_0^{(l)} &= 0, \\
M_t^{(l)} &= \sum_{k=1}^{t} x_{k-l}\varepsilon_k, \quad l = 1, \ldots, p.
\end{aligned}$$

These processes are martingales with respect to filtration (44) and, under the assumptions of Theorem 1, they satisfy both conditions (a), (b) of Theorem 3. Moreover, the stopping times (34) employed in Theorem 3 reduce, in the case of $AR(p)$ model, to those given by (7) because $\sigma_l^2(k-1) = x_{k-l}^2$.

Applying Theorem 3 to the vector $\eta(h)\Big/\sqrt{h} = \frac{1}{\sqrt{h}} G_p(h)(\theta^*(h) - \theta)$ with the coordinates

$$\frac{\eta_i(h)}{\sqrt{h}} = \frac{1}{\sqrt{h}} \sum_{k=\tau_{i-1}(h)+1}^{\tau_i(h)} \sqrt{\beta_{ik}(h)} x_{k-i} \varepsilon_k, \tag{45}$$

one comes to the desire result. This completes the proof of Theorem 1. □

B.2 Proof of Theorem 2

In view of the equation for the standardized deviation

$$\frac{G_p(h)}{\sqrt{h}}(\theta^*(h) - \theta) = \frac{\eta(h)}{\sqrt{h}},$$

one has to study the asymptotic distribution of the vector $\eta(h)/\sqrt{h}$ with coordinates (45).

We will show that, for every vector $\lambda = (\lambda_1, \ldots, \lambda_p)'$ with $\lambda_1^2 + \ldots + \lambda_p^2 = 1,\ \lambda_j \neq 0$, the linear combination $\lambda' \eta(h)/\sqrt{h}$ satisfies the limiting relation

$$\lim_{h\to\infty} \sup_{\theta\in K} \sup_{-\infty<t<\infty} \left| P_\theta\left(\frac{\lambda'\eta(h)}{\sqrt{h}} \le t\right) - \Phi(t)\right| = 0. \tag{46}$$

Let $\{y_j\}_{j\ge1}$ and $\{z_j\}_{j\ge1}$ be two sequences of random variables defined as

$$y_j = \begin{cases} \lambda_1\sqrt{\beta_{1,j+1}} x_j & if\ 0 \le j < \tau_1(h); \\ \ldots & \\ \lambda_{p-1}\sqrt{\beta_{p-1,j+1}(h)} x_{j-p+2} & if\ \tau_{p-2}(h) \le j < \tau_{p-1}(h); \\ \lambda_p x_{j-p+1} & if\ j \ge \tau_{p-1}(h); \end{cases} \tag{47}$$

$$z_j = \begin{cases} y_j & if\ 0 \le j < \tau_p(h) - 1; \\ \lambda_p\sqrt{\alpha_p(h)} x_{\tau_p - p} & if\ j = \tau_p(h) - 1. \end{cases} \tag{48}$$

Then $\lambda'\eta(h)\sqrt{h}$ can be written as

$$\frac{\lambda'\eta(h)}{\sqrt{h}} = \frac{1}{\sqrt{h}} \sum_{j=1}^{\tau_p(h)} z_{j-1}\varepsilon_j = Y(h) + \zeta(h) \tag{49}$$

where

$$Y(h) = \frac{1}{\sqrt{h}} \sum_{j=1}^{\tau_p(h)} y_{j-1}\varepsilon_j, \tag{50}$$

$$\zeta(h) = \frac{1}{\sqrt{h}} \left(\sqrt{\alpha_p(h)} - 1\right) y_{\tau_p(h)-1}\varepsilon_{\tau_p(h)}. \tag{51}$$

Further we establish the following results.

Proposition 6 *Under conditions of Theorem 2, for any set $K \subset [\Lambda_p]$ satisfying (12)*

$$\lim_{h\to\infty} \sup_{\theta\in K} \sup_{-\infty<t<\infty} |P_\theta(Y(h) \le t) - \Phi(t)| = 0 \tag{52}$$

where $Y(h)$ is given by (50).

Lemma 3 *Let $\zeta(h)$ be defined by (51). Then for any $\Delta > 0$ and any set $K \subset [\Lambda_p]$ satisfying (12)*

$$\lim_{h\to\infty} \sup_{\theta\in K} P_\theta\left(|\zeta(h)| > \Delta\right) = 0. \tag{53}$$

Combining these results with (49), one comes to (14). This completes the proof of Theorem 2. □

The proofs of Proposition 6 and Lemma 3 are rather tedious and given below in this section.

C Some technical results

C.1 Some properties of unstable AR(p) model

Here we establish some technical results for unstable AR(p) process and for the sequence of random variables (47) which are used below. Equation (1) can be written in the form

$$X_k = AX_{k-1} + \xi_k, \; k = 1, 2, \ldots \tag{54}$$

where $X_k = (x_k, x_{k-1}, \ldots, x_{k-p+1})'$, $\xi_k = (\varepsilon_k, 0, \ldots, 0)'$; $A = A(\theta)$ is given by (13).

Lemma 4 *Let the process $(X_k)_{k\ge 0}$ satisfy (54) with $\theta \in [\Lambda_p]$. Then for any $n \ge 1$*

$$\sum_{k=0}^{n} \|X_k\|^2 \ge c_p \sum_{k=1}^{n} \varepsilon_k^2 \tag{55}$$

where $[\Lambda_p]$ is the closure of region (3),

$$c_p = \inf_{\theta \in [\Lambda_p]} \Big((1 + \|A\|^2)^{1/2} - \|A\| \Big)^2,$$
$$\|A\|^2 = tr A A'.$$

Proof of Lemma 4. Using Equation (54), one gets

$$\sum_{k=1}^{n} \|X_k\|^2 = \sum_{k=1}^{n} X'_{k-1} A' A X_{k-1} + \sum_{k=1}^{n} \xi'_k A X_{k-1} + \sum_{k=1}^{n} X'_{k-1} A' \xi_k + \sum_{k=1}^{n} \|\xi_k\|^2.$$

From here it follows that

$$\begin{aligned} \sum_{k=1}^{n} \|\xi_k\|^2 &\le \sum_{k=0}^{n} \|X_k\|^2 + 2\|A\| \sum_{k=1}^{n} \|X_{k-1}\| \|\xi_k\| \le \sum_{k=0}^{n} \|X_k\|^2 \\ &+ 2\|A\| \left(\sum_{k=0}^{n} \|X_k\|^2 \right)^{1/2} \left(\sum_{k=1}^{n} \|\xi_k\|^2 \right)^{1/2} \\ &= s \left(\sqrt{\sum_{k=0}^{n} \|X_k\|^2} + \|A\| \sqrt{\sum_{k=1}^{n} \|\xi_k\|^2} \right)^2 - \|A\|^2 \sum_{k=1}^{n} \|\xi_k\|^2. \end{aligned}$$

This implies that

$$(1 + \|A\|^2) \sum_{k=1}^{n} \|\xi_k\|^2 \le \left(\sqrt{\sum_{k=0}^{n} \|X_k\|^2} + \|A\| \sqrt{\sum_{k=1}^{n} \|\xi_k\|^2} \right)^2 .$$

Therefore

$$\sqrt{\sum_{k=0}^{n} \|X_k\|^2} \ge \left[\left(1 + \|A\|^2\right)^{1/2} - \|A\| \right] \left(\sum_{k=1}^{n} \|\xi_k\|^2 \right)^{1/2}$$

and noting that $\sum_{k=1}^{n} \|\xi_k\|^2 = \sum_{k=1}^{n} \varepsilon_k^2$ one comes to (55). Hence Lemma 4. □

Lemma 5 *Let* $(\varepsilon_n)_{n \ge 1}$ *in AR(p) model (1) be a sequence of i.i.d. random variables with* $\mathsf{E}\varepsilon_n = 0$ *and* $\mathsf{E}\varepsilon_n^2 = 1$ *and* K *be a compact subset of* $[\Lambda_p]$ *satisfying (12). Then for any* $\delta > 0$ *and natural number* r

$$\lim_{m \to \infty} \sup_{\theta \in K} P_\theta \left(\|X_{n+r}\|^2 \ge \delta \sum_{k=1}^{n} \|X_{k-1}\|^2 \textit{ for some } n \ge m \right) = 0.$$

Proof of Lemma 5. Applying repeatedly, Equation (54) yields for any $l \geq 1$

$$X_{n+r} = A^{l+r} X_{n-l} + \sum_{i=0}^{l+r-1} A^i \xi_{n+r-i}. \tag{56}$$

Further, for each $s = 1, 2, \ldots$, we define a number $l_n^{(s)}$ such that

$$l_n^{(s)} \in \{l : \|X_{n-l}\| = \min_{1 \leq j \leq s} \|X_{n-j}\|, \ 1 \leq l \leq s\}.$$

Substituting $l_n^{(s)}$ for l in (56), one has

$$X_{n+r} = A^{l_n^{(s)}+r} X_{n-l_n^{(s)}} + \sum_{i=0}^{l_n^{(s)}+r-1} A^i \xi_{n+r-i}.$$

Using the elementary inequalities and taking into account (12), we obtain

$$\begin{aligned}
\|X_{n+r}\|^2 &\leq 2\|A^{l_n^{(s)}+r}\|^2 \|X_{n-l_n^{(s)}}\|^2 + 2\left(\sum_{i=0}^{l_n^{(s)}+r-1} \|A^i\| \|\xi_{n+r-i}\|\right)^2 \\
&\leq 2\|A^{l_n^{(s)}+r}\|^2 \|X_{n-l_n^{(s)}}\|^2 + 2\sum_{i=0}^{l_n^{(s)}+r-1} \|A^i\|^2 \sum_{i=0}^{l_n^{(s)}+r-1} \|\xi_{n+r-i}\|^2 \\
&\leq 2\kappa_p^2 \|X_{n-l_n^{(s)}}\|^2 + 2(r+s)\kappa_p^2 \sum_{i=0}^{s+r-1} \varepsilon_{n-i}^2.
\end{aligned}$$

From here and Lemma 4, it follows that

$$\begin{aligned}
\frac{\|X_{n+r}\|^2}{\sum_{k=0}^{n} \|X_k\|^2} &\leq \frac{2\kappa_p^2 \|X_n\|^2}{\sum_{k=0}^{n} \|X_k\|^2} + \frac{2(r+s)\kappa_p^2 \sum_{i=0}^{s+r-1} \varepsilon_{n-i}^2}{c_p \sum_{k=1}^{n} \varepsilon_k^2} \\
&\leq \frac{2\kappa_p^2}{s} + \frac{2(r+s)\kappa_p^2 \sum_{i=0}^{s+r-1} \varepsilon_{n-i}^2}{c_p \sum_{k=1}^{n} \varepsilon_k^2}, \quad s > r+1.
\end{aligned}$$

Using this estimate and the strong law of large numbers completes the proof of Lemma 5. □

Lemma 6 *Let $(y_j)_{j\geq 0}$ be defined by (47) and $\lambda = (\lambda_1, \ldots, \lambda_p)'$, $\lambda_i \neq 0$, $\lambda_1^2 + \ldots + \lambda_p^2 = 1$. Then for any set $K \subset [\Lambda_p]$ satisfying (12)*

$$\lim_{m\to\infty} \sup_{\theta\in K} P_\theta \left(y_n^2 \geq \delta \sum_{i=0}^{n-1} y_i^2 \textit{ for some } n \geq m \right) = 0.$$

Proof. Using (47), we represent y_{i-1}^2 as

$$y_{i-1}^2 = \sum_{k=1}^{p-1} \lambda_k^2 x_{i-k}^2 \chi_{(\tau_{k-1}+1\leq i<\tau_k)} + \sum_{k=1}^{p-1} \lambda_k^2 \alpha_k x_{i-k}^2 \chi_{i=\tau_k)} + \lambda_p^2 x_{i-p}^2 \chi_{(i\geq\tau_{p-1}+1)}.$$

From here, one has

$$\sum_{i=1}^{n} y_{i-1}^2 \geq \lambda_*^2 \left(\sum_{k=1}^{p-1} \sum_{i=1}^{n} x_{i-k}^2 \chi_{(\tau_{k-1}+1\leq i<\tau_k)} + \sum_{i=1}^{n} x_{i-p}^2 \chi_{(i\geq\tau_{p-1}+1)} \right)$$

where $\lambda_*^2 = \min(\lambda_1^2, \ldots, \lambda_p^2)$. Further we note that

$$\sum_{i=1}^{n} x_{i-k}^2 \chi_{(\tau_{k-1}+1\leq i<\tau_k)} = \sum_{j=0}^{n-k} x_j^2 \chi_{(\tau_{k-1}-k+1\leq j<\tau_k-k)},$$

$$\sum_{i=1}^{n} x_{i-p}^2 \chi_{(i\geq\tau_{p-1}+1)} = \sum_{j=0}^{n-p} x_j^2 \chi_{(j\geq\tau_{p-1}-p+1)}.$$

Therefore

$$\sum_{i=1}^{n} y_{i-1}^2 \geq \lambda_*^2 \left(\sum_{k=1}^{p-1} \sum_{j=0}^{n-k} x_j^2 \chi_{(\tau_{k-1}-k+1\leq j<\tau_k-k)} + \sum_{j=0}^{n-p} x_j^2 \chi_{(j\geq\tau_{p-1}-p+1)} \right)$$
$$\geq \lambda_*^2 \sum_{j=0}^{n-p} x_j^2 \left(\chi_{(0\leq j<\tau_{p-1}-p+1)} + \chi_{(j\geq\tau_{p-1}-p+1)} \right) = \lambda_*^2 \sum_{j=0}^{n-p} x_j^2. \quad (57)$$

In view of the identity

$$\sum_{k=0}^{n} \|X_k\|^2 = \sum_{i=0}^{p-1} \sum_{k=i}^{n} x_{k-i}^2 = p \sum_{l=0}^{n} x_l^2 - \sum_{i=0}^{p-1} \sum_{l=n-i+1}^{n} x_l^2,$$

we get

$$\sum_{l=0}^{n} x_l^2 \geq \frac{1}{p} \sum_{k=0}^{n} \|X_k\|^2. \quad (58)$$

Combining (57) and (58), one obtains

$$
\begin{aligned}
&P_\theta\left(y_n^2 \ge \delta \sum_{i=1}^{n-1} y_i^2 \text{ for some } n \ge m\right)\\
&\quad \le P_\theta\left(\max_{1\le k\le p} x_{n-k+1}^2 \ge \delta\lambda_*^2 \sum_{j=0}^{n-p} x_j^2 \text{ for some } n \ge m\right)\\
&\quad \le P_\theta\left(\|X_n\|^2 \ge \delta\cdot\lambda_*^2\cdot\frac{1}{p}\sum_{k=0}^{n-p}\|X_k\|^2 \text{ for some } n \ge m\right).
\end{aligned}
$$

It remains to apply Lemma 5 to arrive at the desired result. Hence Lemma 6. □

C.2 Proof of Proposition 6

First we note that the sequence (y_j) defined by (47) is adapted to the filtration $(\mathcal{F}_j)$ in (44). In order to show (52), we will use the following probabilistic result for martingales from the paper by Lai and Siegmund (1983).

Lemma 7 *(*Lai and Siegmund (1983), Proposition 2.1*) Let $\{x_n\}_{n\ge0}$ and $\{\varepsilon_n\}_{n\ge1}$ be sequences of random variables adapted to the increasing sequence of σ-algebras $(\mathcal{F}_n)_{n\ge0}$. Let $\{P_\theta, \theta\in\Theta\}$ be a family of probability measures such that under every P_θ*

A1: $\{\varepsilon_n\}$ are i.i.d. with $\mathsf{E}\varepsilon_1 = 0$, $\mathsf{E}\varepsilon_1^2 = 1$;
A2: $\sup_\theta \mathsf{E}_\theta\{\varepsilon_1^2|\varepsilon_1| > a|\} \to 0$ as $a\to\infty$;
A3: ε_n is independent of $\mathcal{F}_{n-1}$ for each $n\ge1$;
A4: $P_\theta\left(\sum_{i=0}^{\infty} x_i^2 = \infty\right) = 1$;
A5: $\sup_\theta P_\theta(x_n^2 > a) \to 0$ as $a\to\infty$ for each $n\ge0$;
A6: $\lim_{m\to\infty}\sup_\theta P_\theta\left(x_n^2 \ge \delta\sum_{i=0}^{n-1} x_i^2 \text{ for some } n\ge m\right) = 0$ for each $\delta>0$.

For $h>0$ let $T(h) = \inf\left\{n:\ \sum_{i=1}^{n} x_{i-1}^2 \ge h\right\}$, $\inf\{\emptyset\} = +\infty$. Then uniformly in $\theta\in\Theta$ and $-\infty<t<\infty$

$$
P_\theta\left\{\frac{1}{\sqrt{h}}\sum_{i=1}^{T(h)} x_{i-1}\varepsilon_i \le t\right\} \to \Phi(t) \text{ as } h\to\infty, \tag{59}
$$

where Φ is the standard normal distribution function.

We apply Lemma 7 to the sequence $(y_j)_{j\ge1}$ which, in view of Lemma 6, satisfies all the conditions A1-A6 for any parametric set K subjected to the restriction (12).

It is easy to check that, for the sequence (47), the stopping time

$$
T(h) = \inf\{n\ge1:\ \sum_{i=1}^{n} y_{i-1}^2 \ge h\} \tag{60}
$$

coincides with $\tau_p(h)$ defined in (7).

Therefore, $Y(h)$ in (50) can be written as

$$Y(h) = \frac{1}{\sqrt{h}} \sum_{j=1}^{T(h)} y_{j-1}\varepsilon_j$$

and, in virtue of Lemma 7, one comes to (52). Hence Proposition 6. $\square$

Remark 3 In spite of the fact that the sequence (y_j) in (47) depends on the parameter h, the proof of Lemma 7 proceeds along the lines of Proposition 2.1 in Lai and Siegmund (1983) and is omitted.

C.3 Proof of Lemma 3

Taking into account (60), one gets the following estimate

$$\begin{aligned} P_\theta\left(|\zeta(h)| > \Delta\right) &\le P_\theta\left(\left|\frac{y_{T(h)-1}\varepsilon_{T(h)}}{\sqrt{h}}\right| > \Delta\right) \le P_\theta\left(|\varepsilon_{T(h)}| > L\right) \\ &\quad + P_\theta\left(|y_{T(h)-1}| > \Delta\sqrt{h}L\right) \\ &\le \frac{1}{L^2}\mathsf{E}_\theta\varepsilon^2_{T(h)} + P_\theta\left(\sum_{k=1}^m y^2_{k-1} \ge h\right) + \alpha_\theta(m), \end{aligned}$$

where L is a positive constant,

$$\alpha_\theta(m) = P_\theta\left(y^2_{n-1} \ge \Delta^2 L^2 \sum_{k=1}^{n-1} y^2_{k-1} \text{ for some } n \ge m\right).$$

Since $(T(h) = m) \in \mathcal{F}_{m-1},\ m = 1, 2, \ldots,$ we compute

$$\begin{aligned} \mathsf{E}_\theta\varepsilon^2_{T(h)} &= \sum_{m\ge} \mathsf{E}_\theta\varepsilon^2_m\chi_{(T(h)=m)} = \mathsf{E}_\theta\sum_{m\ge}\mathsf{E}_\theta\left(\varepsilon^2_m\chi_{(T(h)=m)}|\mathcal{F}_{m-1}\right) \\ &= \mathsf{E}_\theta\sum_{m\ge}\chi_{(T(h)=m)}\mathsf{E}(\varepsilon^2_m|\mathcal{F}_{m-1}) = 1. \end{aligned}$$

Therefore

$$\sup_{\theta\in K} P_\theta\left(|\zeta(h)| > \Delta\right) \le \frac{1}{L^2} + \sup_{\theta\in K} P_\theta\left(\sum_{k=1}^m y^2_{k-1} \ge h\right) + \sup_{\theta\in K}\alpha_\theta(m).$$

Further, we note that

$$P_\theta\left(\sum_{k=1}^m y^2_{k-1} \ge h\right) \le P_\theta\left(\sum_{k=1}^m \|X_k\|^2 \ge h\right)$$

and that for each $\theta \in [\Lambda_p]$

$$\|X_k\|^2 \leq \left(\sum_{j=1}^{k} \|A^{k-j}\| \cdot |\varepsilon_j|\right)^2 \leq \sum_{j=1}^{K} \|A^{k-j}\|^2 \sum_{j=1}^{k} \varepsilon_j^2 \leq \kappa^2 k \sum_{j=1}^{k} \varepsilon_j^2.$$

Thus

$$P_\theta\left(\sum_{k=1}^{m} y_{k-1}^2 \geq h\right) \leq P\left(\kappa_p^2 \sum_{k=1}^{m} k \sum_{j=1}^{k} \varepsilon_j^2 \geq h\right).$$

From here it follows that

$$\lim_{h\to\infty} \sup_{\theta\in K} P_\theta\left(\sum_{k=1}^{m} y_{k-1}^2 \geq h\right) = 0. \tag{61}$$

Now limiting $h \to \infty$, $m \to \infty$ and $L \to \infty$ in (47) and taking into account (61) and Lemma (6), we come to (53). This completes the proof of Lemma 3. □

C.4 Proof of Proposition 5

In order to apply Theorem 2, one has to check condition (12).

Since for each $\theta \in \Lambda_p$

$$\sup_{n\geq 1} \|A^n(\theta)\|^2 \leq \sum_{j\geq 0} \|A^j(\theta)\|^2,$$

it suffices to show that for any compact set $K \subset \Lambda_p$

$$\sup_{\theta\in K} \sum_{j\geq 0} \|A^j\|^2 < \infty. \tag{62}$$

We set

$$\gamma(\theta) = \sum_{j\geq 0} A^j(\theta)(A'(\theta))^j.$$

This matrix series satisfies the equation

$$\gamma(\theta) - A(\theta)\gamma(\theta)A'(\theta) = I_p$$

where I_p is the $p \times p$ identity matrix. This equation can be rewritten as

$$(I_{p^2} - A(\theta) \otimes A(\theta))\mathrm{vec}\,\gamma(\theta) = \mathrm{vec}(I_p)$$

where the vector operation vec$(\cdot)$ is defined by

$$\text{vec}(v) = (v_{1,1}, \ldots, v_{p,1}, \ldots, v_{1,p}, \ldots, v_{p,p})'$$

and $U \otimes V = (U_{ij} \cdot V_{kl})_{1 \le i,j,k,l \le p}$ is the Kronecker product of $p \times p$ matrices $U = \|U_{ij}\|$ and $V = \|V_{ij}\|$. The matrix $I_{p^2} - A(\theta) \otimes A(\theta)$ is invertible because all eigenvalues of matrix $A(\theta) \otimes A(\theta)$ are less than one in modulus for any $\theta \in \Lambda_p$. Therefore

$$\text{vec}\, \gamma(\theta) = (I_{p^2} - A(\theta) \otimes A(\theta))^{-1} \text{vec}(I_p).$$

Now we prove (29) and (30). It is well known (Anderson 1971) that the sample Fisher information matrix (5) has the property

$$\lim_{h \to \infty} \frac{G_n}{n} = F \quad \text{a.s.,}$$

for each $\theta \in \Lambda_p$, where F is a positive definite matrix F satisfying the equation

$$F - AFA' = \|\delta_{1,i}\delta_{1,j}\|. \tag{63}$$

By making use of the definitions of stopping times (7), we obtain

$$\lim_{h \to \infty} \frac{\tau_i(h) - \tau_{i-1}(h)}{h} = \frac{1}{\langle F \rangle_{11}}.$$

Further we have

$$\lim_{h \to \infty} \frac{\hat{G}_{i,\tau_i(h)}}{\tau_i(h) - \tau_{i-1}(h)} = F.$$

Therefore

$$\lim_{h \to \infty} \frac{G_p(h)}{h} = \lim_{h \to \infty} \frac{\hat{G}_{i,\tau_i(h)}}{h} = \frac{F}{\langle F \rangle_{11}} \quad \text{a.s.}$$

$$\lim_{h \to \infty} \frac{\tau_p(h)}{h} = \frac{p}{\langle F \rangle_{11}}.$$

This completes the proof of Proposition 5. □

References

Ahtola, J., Tiao, G.C. (1987). Distributions of least squares estimators of autoregressive parameters for a process with complex roots on the unit circle. *Journal of Time Series Analysis*, *8*(1), 1–14.

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, *30*, 676–6877.

Anderson, T. W. (1971). *The statistical analysis of time series*. New York-London-Sydney: Wiley.

Borisov, V. Z., Konev, V. V. (1977). On sequential parameters estimation in discrete time processes. *Automation and Remote Control*, *6*, 58–64.

Box, G., Jenkins, G., Reinsel, G. (2008). *Time series analysis*: *Forecasting and control*. Hoboken, N.J.: John Wiley.

Brockwell, P. J., Davis, R. A. (1991). *Time series*: *Theory and methods* (2nd ed.). New York: Springer.

Chan, N. H., Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, *16*(1), 367–401.

Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(no. 366, part 1), 427–431.

Galtchouk, L., Konev, V. (2001). On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *Annals of Statistics*, *29*(5), 1508–1536.

Galtchouk, L., Konev, V. (2006). Sequential estimation of the parameters in unstable AR(2). *Sequential Analysis*, *25*(1), 25–43.

Greenwood, P. E., Shiryaev, A. N. (1992). Asymptotic minimaxity of a sequential estimator for a first order autoregressive model. *Stochastics*: *An International Journal of Probability and Stochastic Processes*, *38*(1), 49–65.

Henze, N., Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, *19*(10), 3595–3617.

Konev, V. V. (2016). On one property of martingales with conditionally Gaussian increments and its application in the theory of nonasymptotic inference. *Doklady Mathematics*, *94*, 676–680.

Konev, V. V., Lai, T. L. (1995). Estimators with prescribed precision in stochastic regression models. *Sequential Analysis*, *14*(3), 179–192.

Konev, V., Le Breton, A. (2000). Guaranteed parameter estimation in a first-order autoregressive process with infinite variance. *Sequential Analysis*, *19*(1–2), 25–43.

Konev, V. V., Pergamenshchikov, S. M. (1981). The sequential plans of parameters identification in dynamic systems. *Automation and Remote Control*, *7*, 84–92.

Konev, V., Pergamenshchikov, S. (1997). On guaranteed estimation of the mean of an autoregressive process. *Annals of Statistics*, *25*(5), 2127–2163.

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

Lai, T. L., Wei, C. Z. (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.

Lai, T. L., Wei, C. Z. (1985). *Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. Multivariate analysis VI* (Pittsburgh, Pa., 1983), (pp. 375–393), North-Holland, Amsterdam.

Lee, S. (1994). Sequential estimation for the parameters of a stationary autoregressive model. *Sequential Analysis*, *13*(4), 301–317.

Liptser, R. S., Shiryaev, A. N. (2001). *Statistics of Random Processes. II. Applications, Stochastic Modelling and Applied Probability* (2nd ed.). Springer: Berlin.

Mann, H. B., Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, *11*, 173–220.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B*: *The Indian Journal of Statistics*, *36*(2), 115–128.

Mikulski, P. W., Monsour, M. J. (1991). Optimality of the maximum likelihood estimator in first-order autoregressive processes. *Journal of Time Series Analysis*, *12*(3), 237–253.

Monsour, M. J., Mikulski, P. W. (1998). On limiting distributions in explosive autoregressive processes. *Statistics & Probability Letters*, *37*(2), 141–147.

Novikov, A. A. (1972). Sequential estimation of the parameters of processes of diffusion type. (Russian). *Matematicheskie Zametki*, *12*, 627–638.

Pergamenshchikov, S. M. (1992). Asymptotic properties of sequential design for estimating the parameter of a first-order autoregression. *Theory of Probability & Its Applications*, *36*, 36–49.

Rao, M. M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *Annals of Statistics*, *6*(1), 185–190.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro–Wilk. *Applied Statistics*, *32*, 121–133.

Shiryaev, A. N., Spokoiny, V. G. (2000). *Statistical experiments and decisions. Asymptotic theory. Advanced Series on Statistical Science & Applied Probability, 8.* World Scientific Publishing Co., Inc., River Edge, NJ.

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

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics*, *29*, 1188–1197.