Ann Inst Stat Math (2013) 65:959–992
DOI 10.1007/s10463-013-0402-4

Estimation in threshold autoregressive models with correlated innovations

P. Chigansky · Yu. A. Kutoyants

Received: 15 June 2012 / Revised: 17 November 2012 / Published online: 7 April 2013

Abstract Large sample statistical analysis of threshold autoregressive models is usually based on the assumption that the underlying driving noise is uncorrelated. In this paper, we consider a model, driven by Gaussian noise with geometric correlation tail and derive a complete characterization of the asymptotic distribution for the Bayes estimator of the threshold parameter.

Keywords Asymptotic statistics · Bayes estimator · Threshold autoregression · Hidden Markov models

1 Introduction: the setting and the main result

Let $(X_j)_{j\in\mathbb{Z}_+}$ be the sequence generated by the recursion

$$X_j = (\rho^+ \mathbf{1}_{\{X_{j-1}\geq\theta\}} + \rho^- \mathbf{1}_{\{X_{j-1}<\theta\}})X_{j-1} + \epsilon_j, \quad j \geq 1 \tag{1}$$

where $(\epsilon_j)_{j\in\mathbb{Z}_+}$ is a random process with known distribution, ρ^+ and ρ^- are known constants, and θ is the unknown threshold parameter to be estimated from the sample $X^n := (X_1, \ldots, X_n)$. The Eq. (1) is a basic instance of the threshold autoregression

P. Chigansky is supported by ISF grant 314/09.

P. Chigansky (✉)
Department of Statistics, The Hebrew University, Mount Scopus, 91905 Jerusalem, Israel
e-mail: pchiga@mscc.huji.ac.il

Y. A. Kutoyants
Laboratoire Manceau de Mathématiques, Faculté des Sciences et Techniques, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France
e-mail: kutoyants@univ-lemans.fr

(TAR) models, which play a considerable role in the theory and practice of time series. This type of models have been studied by statisticians for already more than three decades, producing interesting theory and finding many important applications, some of which can be traced in the early and more recent surveys (Tong 1983, 2011; Tsay 1989; Hansen 2011; Chan and Kutoyants 2010; Kutoyants 2012) [see also, e.g., Ling and Tong 2005; Ling et al. 2007 for the analysis of the related moving average threshold (TMA) models].

When it comes to the large sample asymptotic analysis of the estimators, the standard conditions imposed on the models such as (1) are

(i) strong ergodicity of the observed process (X_j)
(ii) independence of the driving random variables ϵ_j's.

Departure from these assumptions poses challenging problems. For the model (1), the condition (i) fails if the absolute value of either ρ^+ or ρ^- is greater or equal to 1. If the process (X_j) is null recurrent (e.g., $\rho_+ = 1$ and $|\rho_-| < 1$), characterization of the exact large sample asymptotic distribution of the likelihood-based estimators of the threshold parameter θ remains an open problem (see Remark 1 below). If $\theta \neq 0$ is assumed to be known, the asymptotic distribution of the coefficients' estimators in the non-ergodic case has been studied in Pham et al (1991), Caner and Hansen (2001), and Liu et al (2011).

TAR models beyond the independence assumption (ii) of the driving noise sequence has not yet been addressed. As we shall shortly see, in the dependent case, the problem falls into the framework of statistical inference of hidden Markov models (HMM), where the driving noise plays the role of the hidden state (see Ch. 10–12 in Cappé et al. 2005 and the references therein). However, most of the HMM literatures deal with locally asymptotically normal (LAN) experiments and, to the best of our knowledge, non-LAN models with partial observations have not yet been studied systematically.

In this paper, we consider the model (1) in which (ϵ_j) is a sequence with geometrically decaying correlation. More precisely, let $X = (X_j)_{j\in\mathbb{Z}_+}$ be generated by the recursion

$$X_j = (\rho^+\mathbf{1}_{\{X_{j-1}\geq\theta\}} + \rho^-\mathbf{1}_{\{X_{j-1}<\theta\}})X_{j-1} + \xi_{j-1} + \varepsilon_j, \tag{2}$$

subject to $X_0 \sim N(0, 1)$, where $\rho := |\rho^+| \vee |\rho^-| < 1$ and the unknown parameter θ takes values in an open bounded subset $\Theta \subset \mathbb{R}$. We shall consider the problem with discontinuous drift function $f(x,\theta) := (\rho^+\mathbf{1}_{\{x\geq\theta\}} + \rho^-\mathbf{1}_{\{x<\theta\}})x$, and thus assume $\rho^+ \neq \rho^-$ and $0 \notin \Theta$. The driving noises $(\varepsilon_j)_{j\in\mathbb{Z}_+}$ and $(\xi_j)_{j\in\mathbb{Z}_+}$ are assumed independent: the *white noise* component (ε_j) is a sequence of i.i.d. $N(0, 1)$ random variables and the *colored noise* (ξ_j) is the Gaussian AR(1) process, generated by the linear recursion

$$\xi_j = a\xi_{j-1} + \zeta_j, \quad j \geq 1, \tag{3}$$

where $(\zeta_j)_{j\in\mathbb{Z}_+}$ are i.i.d. $N(0, 1)$ random variables and a is a known constant $|a| < 1$, controlling the bandwidth of the noise.

All the aforementioned random variables are defined on a measurable space $(\Omega, \mathcal{F})$, with the family of probabilities $(\mathbb{P}_\theta)_{\theta\in\Theta}$, indexed by the unknown parameter. For integers $k \ge m$, we define $\mathcal{F}_{k,m} := \sigma\{\zeta_i, \varepsilon_i \ \ k \le i \le m\}$ and set $\mathcal{F}_m := \mathcal{F}_{0,m}$ and $\mathcal{F}_{k,\infty} := \bigvee_{i\ge k} \mathcal{F}_{k,i}$. All the processes in our problem are adapted to the filtration $(\mathcal{F}_j)_{j\in\mathbb{Z}_+}$ and we shall assume that $\mathcal{F} = \mathcal{F}_{0,\infty}$. Finally we define the observed filtration $\mathcal{F}^X_j := \sigma\{X_i,\ i \le j\} \subset \mathcal{F}_j$.

The recursions (2) and (3) form a conditionally Gaussian system, which means that the conditional law of ξ_n given X^n is Gaussian, and by Theorem 13.5 in Liptser and Shiryaev (2001)

$$X_j = f(X_{j-1},\theta) + \widehat{\xi}_{j-1}(\theta) + \sqrt{1+\gamma_{j-1}}\,\widehat{\varepsilon}_j, \tag{4}$$

where

$$\widehat{\varepsilon}_j := \frac{1}{\sqrt{1+\gamma_{j-1}}}(\xi_{j-1} - \widehat{\xi}_{j-1}(\theta) + \varepsilon_j), \quad j \ge 1$$

is the innovation sequence of i.i.d. $N(0,1)$ random variables. The process $\widehat{\xi}_j(\theta) := \mathbb{E}_\theta(\xi_j|\mathcal{F}^X_j)$ and the deterministic sequence $\gamma_j := \mathbb{E}_\theta(\xi_j - \widehat{\xi}_j(\theta))^2$ satisfy the generalized Kalman filter equations

$$\widehat{\xi}_j(\theta) = a\widehat{\xi}_{j-1}(\theta) + \frac{a\gamma_{j-1}}{1+\gamma_{j-1}}(X_j - f(X_{j-1},\theta) - \widehat{\xi}_{j-1}(\theta)) \tag{5}$$

$$\gamma_j = a^2\gamma_{j-1} + 1 - \frac{a^2\gamma^2_{j-1}}{1+\gamma_{j-1}}, \tag{6}$$

subject to $\widehat{\xi}_0 = 0$ and $\gamma_0 = \mathrm{var}(\xi_0)$.

To avoid inessential technicalities, we shall assume that ξ_0 and X_0 are independent and $\xi_0 \sim N(0,\gamma)$, where γ is the unique positive root of the equation

$$\gamma = a^2\gamma + 1 - \frac{a^2\gamma^2}{1+\gamma}.$$

In this case, the conditional mean $\widehat{\xi}_j(\theta)$ satisfies (5) with constant coefficients:

$$\widehat{\xi}_j(\theta) = a\widehat{\xi}_{j-1}(\theta) + \frac{a\gamma}{1+\gamma}(X_j - f(X_{j-1},\theta) - \widehat{\xi}_{j-1}(\theta)). \tag{7}$$

It can be seen that (γ_j) converges to γ exponentially fast and all the results claimed below hold for ξ_0 with an arbitrary Gaussian distribution.

Let $\bar{X}_0 := X_0$ and $\bar{X}_j := X_j - f(X_{j-1},\theta)$, $j \ge 1$ and note that $\mathcal{F}^{\bar{X}}_j = \mathcal{F}^X_j$ for all $j \ge 0$. By definition of the conditional expectation, $\xi_{j-1} - \widehat{\xi}_{j-1}$ is orthogonal to $\mathcal{F}^X_{j-1}$, and thus to $\mathcal{F}^{\bar{X}}_{j-1}$. Moreover, since the process $(\bar{X}_j, \xi_j, \widehat{\xi}_j)$ is Gaussian, $\xi_{j-1} - \widehat{\xi}_{j-1}$ is independent of $\mathcal{F}^{\bar{X}}_{j-1}$ and thus of $\mathcal{F}^X_{j-1}$ as well. Further, since $\sqrt{1+\gamma}\,\widehat{\varepsilon}_j = \xi_{j-1} -$

$\widehat{\xi}_{j-1} + \varepsilon_j$ and ε_j is independent of $\mathcal{F}_{j-1}$, independence of $\widehat{\varepsilon}_j$ and $\mathcal{F}_{j-1}^X$ follows and the representation (4) implies that the likelihood of the sample X^n is given by

$$L_n(X^n;\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}X_0^2\right)\left(\frac{1}{\sqrt{2\pi(1+\gamma)}}\right)^n \times\exp\left(-\frac{1}{2}\frac{1}{1+\gamma}\sum_{j=1}^n\left(X_j - f(X_{j-1},\theta) - \widehat{\xi}_{j-1}(\theta)\right)^2\right). \tag{8}$$

The likelihood function is discontinuous in θ and hence we are faced with an irregular statistical experiment. In such problems, the maximum likelihood estimator (MLE) is often asymptotically inferior to the Bayes estimator $\widetilde{\theta}_n$, while the latter is typically asymptotically efficient for arbitrary continuous positive prior densities in the following minimax sense (see Theorem 9.1, Ibragimov and Has'minskiĭ 1981):

$$\lim_{\delta\to 0}\varliminf_{n\to\infty}\inf_{T_n}\sup_{\theta:|\theta-\theta_0|\le\delta} n^2\mathbb{E}_\theta(T_n-\theta)^2 \ge \lim_{\delta\to 0}\varliminf_{n\to\infty}\sup_{\theta:|\theta-\theta_0|\le\delta} n^2\mathbb{E}_\theta(\widetilde{\theta}_n-\theta)^2,$$

where T_n's are $\mathcal{F}_n^X$-measurable statistics.

The Bayes estimator for the problem at hand has relatively low computational complexity, since the likelihood function is piecewise constant in θ and has at most n jumps at $\{X_0, \dots, X_{n-1}\}$. More precisely, for a prior density π, the Bayes estimator with respect to the quadratic risk is given by

$$\widetilde{\theta}_n := \frac{\int_\Theta \theta L_n(X^n;\theta)\pi(\theta)\mathrm{d}\theta}{\int_\Theta L_n(X^n;\theta)\pi(\theta)\mathrm{d}\theta} = \frac{\sum_{j:\{X_{(j-1)},X_{(j)}\}\cap\Theta\neq\emptyset} L_n(X^n;X_{(j)})\int_{X_{(j-1)}}^{X_{(j)}}\theta\pi(\theta)\mathrm{d}\theta}{\sum_{j:\{X_{(j-1)},X_{(j)}\}\cap\Theta\neq\emptyset} L_n(X^n;X_{(j)})\int_{X_{(j-1)}}^{X_{(j)}}\pi(\theta)\mathrm{d}\theta}, \tag{9}$$

where $X_{(j)}$ is the jth order statistic of X^n. If the prior π is chosen so that numerical integration in the right-hand side is avoided, the computation of $\widetilde{\theta}_n$ can be carried out in polynomial time of order $O(n^2)$.

The following property, whose proof is deferred to "Appendix" (see Lemma 7), plays a crucial role in the forthcoming analysis.

Proposition 1 *Assume*

$$|\rho^\pm| < 1 \quad and \quad |a| < 1, \tag{10}$$

then the Markov process (X_j, ξ_j) is geometrically ergodic under $\mathbb{P}_\theta$, $\theta \in \Theta$, with the unique invariant probability density $p(x, y; \theta)$.

Remark 1 Obviously, recurrence of (X_j) is necessary for consistent estimation of the threshold parameter. The condition (10) guarantees positive recurrence of the process (X_j), which is an essential ingredient in derivation of the large sample asymptotic of

Theorem 1 below. In the null recurrent case, i.e., when one of the coefficients have unit absolute value, consistent estimation of θ seems to be possible and some preliminary calculations show that the corresponding rate may depend on the distribution tail of the driving noise. The exact characterization of the large sample asymptotic in this setting remains an open problem, even for independent innovations.

The main result of this paper is the following characterization of the asymptotic distribution of the sequence of Bayes estimators:

Theorem 1 *Let $(\widetilde{\theta}_n)$ be the sequence of the Bayes estimators with respect to the quadratic loss function and a prior with continuous positive density π. Then for any continuous function ϕ with at most polynomial growth*

$$\lim_n \mathbb{E}_{\theta_0}\phi(n(\widetilde{\theta}_n-\theta_0)) = \mathbb{E}_{\theta_0}\phi(\tilde{u}),$$

uniformly on compacts from Θ, where

$$\tilde{u} = \frac{\int_{\mathbb{R}} uZ(u)\mathrm{d}u}{\int_{\mathbb{R}} Z(u)\mathrm{d}u}$$

and $\ln Z(u)$, $u\in\mathbb{R}$ *is the following two-sided compound Poisson process:*

$$\ln Z(u) = \begin{cases} \sum_{j=1}^{\Pi^+(u)}(\beta\varepsilon_j^+ - \frac{1}{2}\beta^2) & u\ge 0 \\ \sum_{j=1}^{\Pi^-(|u|)}(\beta\varepsilon_j^- - \frac{1}{2}\beta^2) & u<0. \end{cases} \tag{11}$$

Here Π^+, Π^- are i.i.d Poisson processes with the intensity

$$\varpi = \int_{\mathbb{R}} p(\theta_0, y;\theta_0)\mathrm{d}y,$$

$p(x,y;\theta_0)$ is the unique invariant probability density of the Markov process (X_j,ξ_j) under $\mathbb{P}_{\theta_0}$, $(\varepsilon_j^\pm)$ are i.i.d. $N(0,1)$ random variables, independent of Π^+ and Π^- and

$$\beta^2 = \left(\frac{\theta_0(\rho^+-\rho^-)}{\sqrt{1+\gamma}}\right)^2\left(1+\left(\frac{a\gamma}{1+\gamma}\right)^2\sum_{j=0}^{\infty}\left(\frac{a}{1+\gamma}\right)^{2j}\right)$$
$$= \theta_0^2(\rho^+-\rho^-)^2\frac{1+\gamma^3}{(1+\gamma)(1+\gamma^2)}.$$

1.1 Generalizations

(i) If ξ_{j-1} is replaced in (2) with ξ_j, i.e., if the colored noise component enters without the one-step delay, the model

$$X_j = f(X_{j-1},\theta) + a\xi_{j-1} + \zeta_j + \varepsilon_j.$$

is obtained. In this case, the Kalman filter equations take a slightly different form and the asymptotic analysis can be carried out exactly as in our setting.
On the other hand, if the white noise component ε_j is omitted, the observed process satisfies the equation

$$\begin{aligned} X_j &= f(X_{j-1},\theta) + \xi_j = f(X_{j-1},\theta) + a\xi_{j-1} + \zeta_j \\ &= f(X_{j-1},\theta) + a(X_{j-1} - f(X_{j-2},\theta)) + \zeta_j. \end{aligned}$$

Being a completely observed system, this model fits the setting of Chan (1993) or Chan and Kutoyants (2012) after a straightforward modification.

(ii) Our method is directly applicable to the models, where the colored noise is generated by a linear multivariate recursion:

$$\boldsymbol{\xi}_j = A\boldsymbol{\xi}_{j-1} + B\boldsymbol{\zeta}_j, \quad j \ge 1, \tag{12}$$

where $(\boldsymbol{\zeta}_j)$ are i.i.d. standard Gaussian vectors in $\mathbb{R}^M$ and A and B are $N \times N$ and $N \times M$ matrices, respectively. In this case, the observed process satisfies the scalar recursion

$$X_j = (\rho^+ \mathbf{1}_{\{X_{j-1}\ge\theta\}} + \rho^- \mathbf{1}_{\{X_{j-1}<\theta\}})X_{j-1} + C^\top \boldsymbol{\xi}_{j-1} + \varepsilon_j,$$

where C is a column vector of size N. In this setting, the Kalman filter equations read [cf. (5) and (6)]

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_j(\theta) &= A\widehat{\boldsymbol{\xi}}_{j-1}(\theta) + \frac{A\boldsymbol{\gamma}_{j-1}C}{1 + C^\top \boldsymbol{\gamma}_{j-1}C}(X_j - f(X_{j-1},\theta) - C^\top \widehat{\boldsymbol{\xi}}_{j-1}(\theta)) \\ \boldsymbol{\gamma}_j &= A\boldsymbol{\gamma}_{j-1}A^\top + BB^\top - \frac{A\boldsymbol{\gamma}_{j-1}CC^\top \boldsymbol{\gamma}_{j-1}A^\top}{1 + C^\top \boldsymbol{\gamma}_{j-1}C}, \end{aligned}$$

subject to $\widehat{\boldsymbol{\xi}}_0 = 0$ and $\boldsymbol{\gamma}_0 = \mathrm{cov}(\xi_0, \xi_0)$. If A is a stability matrix, i.e., the absolute values of its eigenvalues are strictly <1, and the pair (A, B) is controllable:

$$\mathrm{rank}(B \ AB \ \ldots \ A^{N-1}B) = N,$$

then the solution of the Riccati equation converges to the matrix $\boldsymbol{\gamma}$, which is the unique strictly positive definite root of the corresponding algebraic Riccati equation (see, e.g., Liptser and Shiryaev 2001)

$$\boldsymbol{\gamma} = A\boldsymbol{\gamma}A^\top + BB^\top - \frac{A\boldsymbol{\gamma}CC^\top\boldsymbol{\gamma}A^\top}{1 + C^\top\boldsymbol{\gamma}C}.$$

The statement of Theorem 1 holds with the rate

$$\varpi = \int_{\mathbb{R}^N} p(\theta_0, y; \theta_0)\mathrm{d}y,$$

where $p(x, y; \theta_0)$ is the invariant density of the process $(\boldsymbol{\xi}_j, X_j)$, and

$$\beta^2 = \left(\frac{\theta_0(\rho^+ - \rho^-)}{\sqrt{1 + C^\top \boldsymbol{\gamma} C}}\right)^2 \left(1 + \sum_{j=0}^{\infty} \left(C^\top \left(A - \frac{A \boldsymbol{\gamma} C C^\top}{1 + C^\top \boldsymbol{\gamma} C}\right)^j \frac{A^\top \boldsymbol{\gamma} C}{1 + C^\top \boldsymbol{\gamma} C}\right)^2\right).$$

The latter formula emerges in the proof of the Lemma 3 below, with the obvious adjustments to the multivariate setting.

The model (12) incorporates the case of the stationary ARMA(p, q) noise:

$$\xi_j = -\sum_{k=1}^{p} a_k \xi_{j-k} + \sum_{\ell=0}^{q} b_\ell \zeta_{j-\ell},$$

where $a_1, \ldots, a_p$ and $b_0, \ldots, b_q$ are constants, such that the roots of the polynomial $a_p z^p + \cdots + a_1 z + 1$ lie in the open unit disk of the complex plain. The canonical state space representation (12) is obtained through the usual state augmentation

$$\boldsymbol{\xi}_j^\top := (\xi_j, \ldots \xi_{j-p+1}, \varepsilon_j, \ldots \varepsilon_{j-q+1})^\top \in \mathbb{R}^{p+q}.$$

(iii) Assuming noises with Gaussian distribution is essential, since in this case, the filtering equations for the conditional density of ξ_j given $\mathcal{F}_j^X$ are finite dimensional and, moreover, the conditional mean $\widehat{\xi}_j$ satisfies the linear recursion, whose explicit solution is used on several occasions through the proof and appears in the expression for β^2. The result can be extended to more general conditionally Gaussian models, such as, e.g., higher order TAR with possibly heteroscedastic driving noise. We expect that for non-Gaussian noise, the limit likelihood will still be a two-sided compound Poisson process, but no neat closed form expression for β^2 will be available.

(iv) In principle, our technique is applicable to Gaussian sequences with non-Markov structure, such as fractional noises, etc. The analysis in this setting is more complicated, depending on the ergodic properties of the processes and the complexity of the filtering equations (whose linearity will be intact).

(v) Joint asymptotic analysis of the likelihood-based estimators of all the parameters in the model can be in principle carried out using the same weak convergence approach, used in the proof of Theorem 1 below (see a brief outline in Sect. 2.2). In this case, the likelihood (8) is considered as a function of the four unknown parameters $\rho^+, \rho^-, a \in (-1, 1)$ and $\theta \in \Theta$ and the corresponding normalized likelihood ratios read [cf. (14) below]

$$Z_n(y, w, v, u) := \frac{L_n\left(X^n; \rho_0^+ + \frac{y}{\sqrt{n}}, \rho_0^- + \frac{w}{\sqrt{n}}, a_0 + \frac{v}{\sqrt{n}}, \theta_0 + \frac{u}{n}\right)}{L_n(X^n; \rho_0^+, \rho_0^-, a_0, \theta_0)},$$

for a fixed value of the parameter vector $(\rho_0^+, \rho_0^-, a_0, \theta_0)$ and variables y, w, v, u taking values in appropriate sets. Note that the localizing scaling of θ differs from that of the other parameters, in which the likelihood function is smooth. It is possible to check the weak convergence of processes

$$Z_n(y, w, v, u) \xrightarrow{w} Z(y, w, v)Z(u), \quad \text{as } n \to \infty, \tag{13}$$

where $Z(u)$ is the same as in (11) and

$$Z(y, w, v) = \exp\left(\eta^\top \nu - \frac{1}{2}\eta^\top I\eta\right), \quad \eta := (y, w, v) \in \mathbb{R}^3,$$

with $\nu \sim N(0, I)$ and the Fisher information matrix I, whose explicit expression is cumbersome. The convergence (13) implies the weak convergence of errors for the corresponding Bayes estimators

$$(\sqrt{n}(\widetilde{\rho}_n^+ - \rho_0^+), \sqrt{n}(\widetilde{\rho}_n^- - \rho_0^-), \sqrt{n}(\widetilde{a}_n - a_0), n(\widetilde{\theta}_n - \theta_0)) \xrightarrow{w} (\tilde{y}, \tilde{w}, \tilde{v}, \tilde{u}),$$

with $\tilde{u}$ as in Theorem 1 and zero mean normal vector $(\tilde{y}, \tilde{w}, \tilde{v})$ with covariance I^{-1}, independent of $\tilde{u}$. Since the LAN property holds with respect to (ρ^+, ρ^-, a), the corresponding MLEs are also asymptotically efficient.

The proof of Theorem 1 is given in the next section and supplementary results, concerning the ergodic properties of the relevant processes, appear in "Appendix". Some simulations, demonstrating the contributions of this paper, are gathered in Sect. 3.

2 The proof of Theorem 1

2.1 The notations

The actual unknown value of the parameter will be denoted by θ_0 and will be assumed to belong to a generic compact $K \subset \Theta$. We shall use C_i, $i \in \mathbb{N}$ to denote absolute constants, whose values depend only on the known parameters of the model and the compact K and may change at each appearance. For random sequences (x_n), (y_n) and a positive real decreasing sequence (r_n), $x_n = y_n + O(r_n)$ means that $\sup_n |x_n - y_n|/r_n$ is a random variable with moments, bounded uniformly over K. Throughout, we reserve

$$b := \frac{a}{1+\gamma} \quad \text{and} \quad c := \frac{a\gamma}{1+\gamma}.$$

For an integer n, the quantities such as $n^{1/2}$ and $n^{1/4}$, are understood to be rounded to the nearest integer if needed. For $\ell < k$, we set $\sum_{j=k}^{\ell}(\ldots) = 0$ and $\prod_{j=k}^{\ell}(\ldots) = 1$. For a vector $z \in \mathbb{R}^d$, $\|z\|$ stands for the ℓ_1 norm. Finally, $\widetilde{\mathbb{P}}_{\theta_0}$ and $\widetilde{\mathbb{E}}_{\theta_0}$ denote the probability on $(\Omega, \mathcal{F})$ and the corresponding expectation, under which all the processes are stationary (the unique existence of such probability is argued in "Appendix").

2.2 Preliminaries

Consider the scaled sequence of likelihoods

$$Z_n(u) := \frac{L_n(X^n; \theta_0 + u/n)}{L_n(X^n; \theta_0)}, \quad u \in \mathbb{U}_n := n(\Theta - \theta_0), \quad n \geq 1. \tag{14}$$

The Bayes estimator of θ is given by

$$\begin{aligned}\widetilde{\theta}_n &= \frac{\int_\Theta \theta L_n(X^n; \theta)\pi(\theta)\mathrm{d}\theta}{\int_\Theta L_n(X^n; \theta)\pi(\theta)\mathrm{d}\theta} = \frac{\int_{\mathbb{U}_n} \left(\theta_0 + \frac{u}{n}\right) L_n\left(X^n; \theta_0 + \frac{u}{n}\right) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u}{\int_{\mathbb{U}_n} L_n\left(X^n; \theta_0 + \frac{u}{n}\right) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u} \\ &= \theta_0 + \frac{1}{n} \frac{\int_{\mathbb{U}_n} u L_n\left(X^n; \theta_0 + \frac{u}{n}\right) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u}{\int_{\mathbb{U}_n} L_n\left(X^n; \theta_0 + \frac{u}{n}\right) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u}\end{aligned}$$

and thus

$$n\left(\widetilde{\theta}_n - \theta_0\right) = \frac{\int_{\mathbb{U}_n} u Z_n(u) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u}{\int_{\mathbb{U}_n} Z_n(u) \pi\left(\theta_0 + \frac{u}{n}\right) \mathrm{d}u}.$$

The right-hand side is a functional of $Z_n(u), u \in \mathbb{U}_n$, which under appropriate tightness conditions, converges weakly to the random variable

$$\tilde{u} = \frac{\int_{\mathbb{R}} u Z(u)\mathrm{d}u}{\int_{\mathbb{R}} Z(u)\mathrm{d}u},$$

if the finite dimensional distributions of $Z_n(u)$ converge to those of $Z(u)$. More precisely, the result claimed in Theorem 1 follows from Theorem I.10.2 in Ibragimov and Has'minskiĭ (1981), whose assumptions we check in Sects. 2.3 and 2.4 below.

2.3 Convergence of finite dimensional distributions

We shall prove that the characteristic functions of the finite dimensional distributions of log likelihoods $Y_n(u) := \ln Z_n(u)$ converge to those of the compound Poisson process in (11). To this end, we will show that for any $d \geq 1$ and real numbers $u_{-d} < \cdots < u_d$

$$\begin{aligned}\psi_n(\lambda) &:= \mathbb{E}_{\theta_0} \exp\left(\sum_{k=-d}^{d} \mathbf{i}\lambda_k \left(Y_n(u_k) - Y_n(u_{k-1})\right)\right) \\ &\xrightarrow{n\to\infty} \prod_{k=-d}^{d} \exp\left(\varpi(u_k - u_{k-1})(\mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k\beta^2 - \frac{1}{2}\beta^2\lambda_k^2} - 1)\right) =: \psi(\lambda), \quad \forall \lambda \in \mathbb{R}^d,\end{aligned} \tag{15}$$

uniformly over compacts from Θ, where ϖ and β are constants, defined in Theorem 1. Without loss of generality, we shall assume that $u_0 = 0$ and consider only positive

u_k's. The symmetric case of negative u_k's is treated similarly and independence of the emerging compound Poisson processes Π^+ and Π^- will be evident from the proof.

Let $\widehat{\Xi}_j$ be the vector with the entries

$$\widehat{\Xi}_j^k := \widehat{\xi}_j(\theta_0 + u_k/n), \quad k = 0, \ldots, d. \tag{16}$$

Using the expression (8) for the likelihood, we get

$$\begin{aligned}
&Y_n(u_k) - Y_n(u_{k-1}) \\
&= -\frac{1}{2}\frac{1}{1+\gamma}\sum_{j=1}^n \Big(X_j - f(X_{j-1}, \theta_0 + u_k/n) - \widehat{\Xi}_{j-1}^k\Big)^2 \\
&\quad + \frac{1}{2}\frac{1}{1+\gamma}\sum_{j=1}^n \Big(X_j - f(X_{j-1}, \theta_0 + u_{k-1}/n) - \widehat{\Xi}_{j-1}^{k-1}\Big)^2 \\
&= \sum_{j=1}^n -\frac{1}{2}\frac{1}{1+\gamma}\Big(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^k\}} - \widehat{\Xi}_{j-1}^k + \widehat{\Xi}_{j-1}^{k-1}\Big) \\
&\quad \times \Big(2X_j - f(X_{j-1}, \theta_0 + u_k/n) - f(X_{j-1}, \theta_0 + u_{k-1}/n) - \widehat{\Xi}_{j-1}^k - \widehat{\Xi}_{j-1}^{k-1}\Big)
\end{aligned} \tag{17}$$

where we set $D_n^k := [\theta_0 + \frac{u_{k-1}}{n}, \theta_0 + \frac{u_k}{n}]$ and used the identity

$$f(x, \theta_0 + u_{k-1}/n) - f(x, \theta_0 + u_k/n) = x(\rho^+ - \rho^-)\mathbf{1}_{\{x\in D_n^k\}}.$$

If we define

$$\delta_j^k := \widehat{\Xi}_j^{k-1} - \widehat{\Xi}_j^k, \quad \text{and} \quad \sigma_j^{k-1} := \widehat{\Xi}_j^0 - \widehat{\Xi}_j^{k-1},$$

the expression (17) takes the following form under $\mathbb{P}_{\theta_0}$

$$\begin{aligned}
&Y_n(u_k) - Y_n(u_{k-1}) \\
&= \sum_{j=1}^n -\frac{1}{1+\gamma}\Big(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^k\}} + \delta_{j-1}^k\Big) \\
&\quad \times \Big(\sqrt{1+\gamma}\,\widehat{\varepsilon}_j + X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in B_n^{k-1}\}} + \sigma_{j-1}^{k-1} \\
&\quad + \frac{1}{2}\Big(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^k\}} + \delta_{j-1}^k\Big)\Big),
\end{aligned} \tag{18}$$

with $B_n^{k-1} := [\theta_0, \theta_0 + \frac{u_{k-1}}{n}]$. The sequences (δ_j^k) and (σ_j^{k-1}) satisfy the recursions

$$\begin{aligned}
\delta_j^k &= b\delta_{j-1}^k - c(\rho^+ - \rho^-)X_{j-1}\mathbf{1}_{\{X_{j-1}\in D_n^k\}}, \quad j \ge 1 \\
\delta_0^k &= 0,
\end{aligned} \tag{19}$$

and

$$\begin{aligned}\sigma_j^{k-1} &= b\sigma_{j-1}^{k-1} - c(\rho^+ - \rho^-)X_{j-1}\mathbf{1}_{\{X_{j-1}\in B_n^{k-1}\}}, \quad j \geq 1\\ \sigma_0^{k-1} &= 0\end{aligned} \tag{20}$$

where we set $b := \frac{a}{1+\gamma}$, $c := \frac{a\gamma}{1+\gamma}$. In what follows, both representations (17) and (18) will be useful.

To prove the convergence (15), we shall partition the terms in the sum (17) or (18) into $n^{1/2}$ consecutive blocks of size $n^{1/2}$ and discard $n^{1/4}$ first terms in each block. As shown in the Lemma 1 below, discarding the total of $n^{1/4} \cdot n^{1/2}$ terms does not alter the limit of the sum and, by Lemma 2, the remaining blocks become approximately independent due to the fast mixing of the process $(X_j, \xi_j, \widehat{\Xi}_j)$. Moreover, in each remaining block, the probability of having exactly one of the events $\{X_{j-1} \in D_n^k\}$ occurred is of order $n^{1/2}$. Hence, the sum of $n^{1/2}$ such nearly independent blocks yields the compound Poisson limit of Lemma 3. This approach to Poisson limits dates back to at least Meyer (1973).

Denote by $s_{j,k}$ the summands in the right-hand side of (17) or (18). Set

$$S_n := \sum_{k=1}^{d} \lambda_k \sum_{j=1}^{n} s_{j,k},$$

and, for $m = 1, \ldots, n^{1/2}$, define

$$S_{n,m} = \sum_{k=1}^{d} \lambda_k \sum_{j=(m-1)n^{1/2}+n^{1/4}}^{mn^{1/2}} s_{j,k}.$$

For $\psi_n(\lambda)$ and $\psi(\lambda)$ defined in (15), the triangle inequality yields the bound

$$\begin{aligned}|\psi_n(\lambda) - \psi(\lambda)| \;&\leq \mathbb{E}_{\theta_0}\left|\exp(\mathbf{i}S_n) - \exp\left(\mathbf{i}\sum_{m=1}^{n^{1/2}} S_{n,m}\right)\right|\\ &\quad + \left|\mathbb{E}_{\theta_0}\exp\left(\mathbf{i}\sum_{m=1}^{n^{1/2}} S_{n,m}\right) - (\widetilde{\mathbb{E}}_{\theta_0}\exp(\mathbf{i}S_{n,1}))^{n^{1/2}}\right|\\ &\quad + \left|(\widetilde{\mathbb{E}}_{\theta_0}\exp(\mathbf{i}S_{n,1}))^{n^{1/2}} - \psi(\lambda)\right|,\end{aligned} \tag{21}$$

where $\widetilde{\mathbb{E}}_{\theta_0}$ stands for the expectation with respect to the probability $\widetilde{\mathbb{P}}_{\theta_0}$ on $(\Omega, \mathcal{F})$, under which the process $(X_j, \xi_j, \widehat{\Xi}_j)$ is stationary (see Lemma 8). In the following lemmas, we show that all three terms in the right-hand side of (21) vanish as $n \to \infty$, uniformly over θ_0 on compacts from Θ.

Lemma 1 *For any* $\lambda \in \mathbb{R}^d$ *and a compact* $K \subset \Theta$,

$$\lim_n \sup_{\theta_0 \in K} \mathbb{E}_{\theta_0} \left| S_n - \sum_{m=1}^{n^{1/2}} S_{n,m} \right| = 0, \tag{22}$$

and consequently

$$\lim_n \mathbb{E}_{\theta_0} \left| \exp(\mathbf{i} S_n) - \exp\left(\mathbf{i} \sum_{m=1}^{n^{1/2}} S_{n,m} \right) \right| = 0, \tag{23}$$

uniformly over $\theta_0 \in K$.

Proof We shall assume that n is large enough so that $\max_k |\theta_0 + u_k/n| \le \sup |K| + 1$ and hence on the events $\{X_{j-1} \in D_n^k\}$ and $\{X_{j-1} \in B_n^k\}$, we have $|X_{j-1}| \le \sup |K| + 1$. Using the representation (18), we get

$$\begin{aligned} \mathbb{E}_{\theta_0} \left| S_n - \sum_{m=1}^{n^{1/2}} S_{n,m} \right| &\le n^{3/4} d \max_k |\lambda_k| \max_{j \le n} \mathbb{E}_{\theta_0} \left| X_{j-1}(\rho^+ - \rho^-) \mathbf{1}_{\{X_{j-1} \in D_n^k\}} + \delta_{j-1}^k \right| \\ &\quad \times \Bigg| \sqrt{1+\gamma}\,\widehat{\varepsilon}_j + X_{j-1}(\rho^+ - \rho^-) \mathbf{1}_{\{X_{j-1} \in B_n^{k-1}\}} + \sigma_{j-1}^{k-1} \\ &\quad + \frac{1}{2} \left(X_{j-1}(\rho^+ - \rho^-) \mathbf{1}_{\{X_{j-1} \in D_n^k\}} + \delta_{j-1}^k \right) \Bigg|. \end{aligned} \tag{24}$$

By Jensen's inequality, it follows from (19), that

$$\mathbb{E}_{\theta_0} (\delta_j^k)^2 \le |b| \mathbb{E}_{\theta_0} (\delta_{j-1}^k)^2 + \frac{1}{1-|b|} c^2 (\rho^+ - \rho^-)^2 (\sup |K| + 1)^2 \mathbb{P}_{\theta_0} (X_{j-1} \in D_n^k),$$

which, in view of (44) and $\delta_0^k = 0$, implies $\max_{j \le n} \mathbb{E}_{\theta_0} (\delta_j^k)^2 \le C_1/n$. Further, since $\widehat{\varepsilon}_j$ is independent of $\mathcal{F}_{j-1}^X$, $\mathbb{E}_{\theta_0} |\delta_{j-1}^k| |\widehat{\varepsilon}_j| = \mathbb{E}_{\theta_0} |\delta_{j-1}^k| \mathbb{E}_{\theta_0} |\widehat{\varepsilon}_j| \le C_2/n$. Similarly $\mathbb{E}_{\theta_0} (\sigma_j^{k-1})^2 \le C_3/n$. Plugging these bounds into (24), we obtain (22)

$$\mathbb{E}_{\theta_0} \left| S_n - \sum_{m=1}^{n^{1/2}} S_{n,m} \right| \le C_4 n^{-1/4},$$

with a constant C_4, depending only on K. The uniform convergence in (23) follows, since $|\exp(\mathbf{i}x) - \exp(\mathbf{i}y)| \le |x - y|$. □

Lemma 2 *For any* $\lambda \in \mathbb{R}^d$,

$$\lim_n \left| \mathbb{E}_{\theta_0} \exp\left(\mathbf{i} \sum_{m=1}^{n^{1/2}} S_{n,m} \right) - (\widetilde{\mathbb{E}}_{\theta_0} \exp(\mathbf{i} S_{n,1}))^{n^{1/2}} \right| = 0,$$

uniformly over $\theta_0 \in K$ *for any compact* $K \subset \Theta$.

Proof We shall use the bound (50) of Lemma 8 and thus will need to establish the corresponding Lipschitz property. To this end, for fixed $x, y \in \mathbb{R}$ and $z \in \mathbb{R}^{d+1}$, let $(X_j(x), \xi_j(y), \widehat{\Xi}_j(z))$ be the solution of the recursions (2), (3) and [cf. (7) and (16)]

$$\widehat{\Xi}^k_j = b\widehat{\Xi}^k_{j-1} + c(X_j - f(X_{j-1}, \theta_0 + u_k/n)), \quad k = 0, \dots, d,$$

subject to the initial conditions x, y and z, respectively. The latter recursions give

$$\widehat{\Xi}^k_j(z) = b^j z_k + c\sum_{i=1}^{j}(X_i - f(X_{i-1}, \theta_0 + u_k/n))b^{j-i}. \tag{25}$$

Consider the random variable [cf. the right-hand side of (17)]

$$\begin{aligned}&\Phi_\ell(x, y, z)\\ &:= \sum_{k=1}^{d}\lambda_k\sum_{j=1}^{\ell}-\frac{1}{2}\frac{1}{1+\gamma}\Big(X_{j-1}(\rho^+-\rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^k\}}-\widehat{\Xi}^k_{j-1}(z)+\widehat{\Xi}^{k-1}_{j-1}(z)\Big)\\ &\quad\times\Big(2X_j-f(X_{j-1},\theta_0+u_k/n)-f(X_{j-1},\theta_0+u_{k-1}/n)-\widehat{\Xi}^k_{j-1}(z)-\widehat{\Xi}^{k-1}_{j-1}(z)\Big),\end{aligned}$$

where we dropped the dependence on x and y for brevity. Define the function $h(x, y, z) := \mathbb{E}_{\theta_0}\exp(\mathbf{i}\Phi_\ell(x, y, z))$. We aim to show that for $z, z' \in \mathbb{R}^{d+1}$,

$$\big|h(x, y, z) - h(x, y, z')\big| \le L(1 + |x| + |y| + \|z\| + \|z'\|)\|z - z'\| \tag{26}$$

for some constant L, independent of ℓ. Using the definition of $\Phi_\ell(x, y, z)$ and the explicit formula (25), a tedious but straightforward calculation gives

$$\begin{aligned}|\Phi_\ell(x, y, z)-\Phi_\ell(x, y, z')|\le&\sum_{k=1}^{d}|\lambda_k|(|z_k-z'_k|\\ &+|z_{k-1}-z'_{k-1}|)\left(\sum_{j=1}^{\ell}2|b|^j(|X_j|+|X_{j-1}|\right)\\ &+(\|z\|+\|z'\|)\frac{1}{1-|b|}+2\sum_{j=1}^{\ell}|b|^j\sum_{i=1}^{j}\Big((|X_i|+|X_{i-1}|+|\xi_{i-1}|)|b|^{j-i}\Big).\end{aligned}$$

Taking the expectation of both sides, we get

$$\begin{aligned}&|h(x, y, z)-h(x, y, z')| \le \mathbb{E}_{\theta_0}|\Phi_\ell(x, y, z)-\Phi_\ell(x, y, z')| \le \frac{2\|\lambda\|}{1-|b|}\|z - z'\|\\ &\times\left(\sum_{j=1}^{\ell}2|b|^j2\mathbb{E}_{\theta_0}|X_j|+(\|z\| + \|z'\|)+4\sum_{j=1}^{\ell}|b|^j\sum_{i=1}^{j}|b|^{j-i}\mathbb{E}_{\theta_0}(|X_{i-1}|+|\xi_{i-1}|)\right)\end{aligned}$$

and applying the bound (45), the inequality (26) follows. Note that by the Markov property, $\widetilde{\mathbb{E}}_{\theta_0} e^{\mathbf{i}S_{n,1}} = \widetilde{\mathbb{E}}_{\theta_0} h(X_0, \xi_0, \widehat{\Xi}_0)$, and for $m = 1, \ldots, n^{1/2}$

$$\mathbb{E}_{\theta_0}\left(e^{\mathbf{i}S_{n,m}} | \mathcal{F}_{(m-1)n^{1/2}}\right)$$
$$= \mathbb{E}_{\theta_0}\left(h\left(X_{(m-1)n^{1/2}+n^{1/4}-1}, \xi_{(m-1)n^{1/2}+n^{1/4}-1}, \widehat{\Xi}_{(m-1)n^{1/2}+n^{1/4}-1}\right) | \mathcal{F}_{(m-1)n^{1/2}}\right).$$

Hence, by the Lemma 8

$$\left|\mathbb{E}_{\theta_0}(e^{\mathbf{i}S_{n,m}} | \mathcal{F}_{(m-1)n^{1/2}}) - \widetilde{\mathbb{E}}_{\theta_0} e^{\mathbf{i}S_{n,1}}\right| \le C_1 q^{n^{1/4}},$$

with a positive constant $q < 1$ and C_1, independent of θ_0. Finally, considering the telescopic series, we get

$$\left|\mathbb{E}_{\theta_0} \exp\left(\mathbf{i}\sum_{m=1}^{n^{1/2}} S_{n,m}\right) - \left(\widetilde{\mathbb{E}}_{\theta_0} \exp(\mathbf{i}S_{n,1})\right)^{n^{1/2}}\right|$$
$$= \Bigg|\sum_{k=0}^{n^{1/2}-1}\Bigg(\mathbb{E}_{\theta_0}\prod_{m=1}^{n^{1/2}-k} e^{\mathbf{i}S_{n,m}} \prod_{m=n^{1/2}-k+1}^{n^{1/2}} \widetilde{\mathbb{E}}_{\theta_0} e^{\mathbf{i}S_{n,1}}$$
$$-\mathbb{E}_{\theta_0}\prod_{m=1}^{n^{1/2}-k-1} e^{\mathbf{i}S_{n,m}} \prod_{m=n^{1/2}-k}^{n^{1/2}} \widetilde{\mathbb{E}}_{\theta_0} e^{\mathbf{i}S_{n,1}}\Bigg)\Bigg|$$
$$\le \sum_{k=0}^{n^{1/2}-1} \mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(e^{\mathbf{i}S_{n,n^{1/2}-k}} | \mathcal{F}_{(n^{1/2}-k-1)n^{1/2}}\right) - \widetilde{\mathbb{E}}_{\theta_0} e^{\mathbf{i}S_{n,1}}\right| \le C_1 n^{1/2} q^{n^{1/4}} \to 0,$$

as claimed. □

Lemma 3 *For any $\lambda \in \mathbb{R}^d$,*

$$\lim_n \left|\left(\widetilde{\mathbb{E}}_{\theta_0} \exp(\mathbf{i}S_{n,1})\right)^{n^{1/2}} - \psi(\lambda)\right| = 0,$$

uniformly over $\theta_0 \in K$ for any compact $K \subset \Theta$.

Proof Let $D_n := \bigcup_{k=1}^d D_n^k$ and define the following events

(i) All samples avoid D_n:

$$A_0 = \bigcap_{j=1}^{n^{1/2}} \{X_{j-1} \notin D_n\}$$

(ii) Only $(j-1)$th sample falls in D_n^k:

$$A_{j,k} = \left\{X_{j-1} \in D_n^k\right\} \cap \bigcap_{i \neq j,\ i=1}^{n^{1/2}} \{X_{i-1} \notin D_n\}$$

(iii) Precisely one sample falls in D_n:

$$A_1 = \bigcup_{k=1}^{d} \bigcup_{j=1}^{n^{1/2}} A_{j,k}$$

(iv) Two or more samples fall in D_n:

$$A_{2+} = (A_0 \cup A_1)^c.$$

Note that

$$\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} = \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_0\}} + \sum_{k=1}^{d} \sum_{j=1}^{n^{1/2}} \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} + \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_{2+}\}}. \tag{27}$$

Below we shall show that

$$\sum_{j=1}^{n^{1/2}} \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} = \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1/2} + O(n^{-3/4}) \tag{28}$$

$$\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_0\}} = 1 - \sum_{k=1}^{d} (u_k - u_{k-1}) \varpi n^{-1/2} + O(n^{-3/4}) \tag{29}$$

$$\widetilde{\mathbb{P}}_{\theta_0}(A_{2+}) = O(n^{-3/4}), \tag{30}$$

where $\varpi := \int p(\theta_0, y; \theta_0) \mathrm{d}y$ and $p(x, y; \theta_0)$ is the unique invariant density of the chain (X_j, ξ_j) (see Lemma 7). Plugging these expressions into (27), we obtain

$$\begin{aligned}
\left(\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}}\right)^{n^{1/2}} &= \Bigg(1 - \sum_{k=1}^{d} (u_k - u_{k-1}) \varpi n^{-1/2} \\
&\quad + \sum_{k=1}^{d} \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1/2} + O(n^{-3/4})\Bigg)^{n^{1/2}} \\
&\xrightarrow{n\to\infty} \exp\left(\sum_{k=1}^{d} \left(\mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} - 1\right)(u_k - u_{k-1}) \varpi\right).
\end{aligned}$$

The claimed result follows, once we check that (28)–(30) hold uniformly in θ_0 on compacts from Θ.

To this end, note that on the event $A_{j,k}$, the equation (19) give

$$\delta_i^\ell = \delta_0^\ell b^i, \quad \ell \neq k$$

and

$$\delta_i^k = \delta_0^k b^i - c(\rho^+ - \rho^-) X_{j-1} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} b^{j-i} \mathbf{1}_{\{i \geq j\}},$$

where δ_0^ℓ, $\ell = 1, \ldots, k$ are bounded random variables under $\widetilde{\mathbb{P}}_{\theta_0}$. Similarly, since $D_n^k \subset B_n^{\ell-1}$ for $\ell > k$, and $\widetilde{\mathbb{P}}_{\theta_0}(X_{j-1} \in B_n^{\ell-1} \cap D_n^k) = 0$ for $\ell \leq k$,

$$\sigma_i^{\ell-1} = \sigma_0^{\ell-1} b^i, \quad \ell \in \{1, \ldots, k\}$$

and

$$\delta_i^{\ell-1} = \sigma_0^{\ell-1} b^i - c(\rho^+ - \rho^-) X_{j-1} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} b^{j-i} \mathbf{1}_{\{i \geq j\}}, \quad \ell \in \{k+1, \ldots, d\}.$$

Hence, for $n^{1/4} \leq j \leq n^{1/2}$ on the event $A_{j,k}$, by (18) we have

$$\begin{aligned}
S_{n,1} &= \sum_{\ell=1}^{d} \sum_{i=n^{1/4}}^{n^{1/2}} \lambda_\ell s_{i,\ell} = \lambda_k \sum_{i=n^{1/4}}^{n^{1/2}} s_{i,k} + O\left(b^{n^{1/4}}\right) \\
&= -\lambda_k \sum_{i=j}^{n^{1/2}} \frac{1}{1+\gamma} \left(X_{i-1}(\rho^+ - \rho^-) \mathbf{1}_{\{X_{i-1} \in D_n^k\}} + \delta_{i-1}^k \right) \\
&\quad \times \left(\sqrt{1+\gamma}\,\widehat{\varepsilon}_i + \frac{1}{2} \left(X_{i-1}(\rho^+ - \rho^-) \mathbf{1}_{\{X_{i-1} \in D_n^k\}} + \delta_{i-1}^k \right) \right) + O\left(b^{n^{1/4}}\right) \\
&= \lambda_k \sum_{i=j+1}^{n^{1/2}} \left(\frac{c(\rho^+ - \rho^-)}{\sqrt{1+\gamma}} X_{j-1} b^{j-i+1} \widehat{\varepsilon}_i - \frac{1}{2} \left(\frac{c(\rho^+ - \rho^-)}{\sqrt{1+\gamma}} X_{j-1} b^{j-i+1} \right)^2 \right) \\
&\quad - \lambda_k \left(\frac{\rho^+ - \rho^-}{\sqrt{1+\gamma}} X_{j-1} \widehat{\varepsilon}_j + \frac{1}{2} \left(\frac{\rho^+ - \rho^-}{\sqrt{1+\gamma}} \right)^2 X_{j-1}^2 \right) + O\left(b^{n^{1/4}}\right) \\
&= \lambda_k \underbrace{\sum_{i=j}^{n^{1/2}} \left(Q(j-i) X_{j-1} \widehat{\varepsilon}_i - \frac{1}{2} Q^2(i-j) X_{j-1}^2 \right)}_{:=\Psi_{j,k}} + O\left(b^{n^{1/4}}\right),
\end{aligned}$$

where we defined the kernel

$$Q(i) = \begin{cases} 0 & i < 0 \\ -\dfrac{\rho^+ - \rho^-}{\sqrt{1+\gamma}} & i = 0 \\ \dfrac{c(\rho^+ - \rho^-)}{\sqrt{1+\gamma}} b^{i-1} & i > 0. \end{cases}$$

By the triangle inequality

$$\begin{aligned} &\left| \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i} S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} - \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1} \right| \\ &\quad \le \left| \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{A_{j,k}\}} - \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \right| \\ &\quad\quad + \left| \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} - \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1} \right| + O\left(b^{n^{1/4}}\right). \end{aligned} \tag{31}$$

By (46), for $i < j$

$$\begin{aligned} &\widetilde{\mathbb{P}}_{\theta_0}(X_i \in D_n, X_j \in D_n^k) \\ &= \widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_i \in D_n\}} \int_{D_n^k} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x - f(X_{j-1}, \theta_0) - \xi_{j-1})^2} \mathrm{d}x \\ &\le \frac{u_k - u_{k-1}}{n} \widetilde{\mathbb{P}}_{\theta_0}(X_i \in D_n) \le C_1 n^{-2}, \end{aligned}$$

with a constant C_1, independent of θ_0. Similar bound holds for $j < i$ and hence, using the identity $\mathbf{1}_{\{A\}} - \mathbf{1}_{\{A \cap B\}} = \mathbf{1}_{\{A \setminus (A \cap B)\}} = \mathbf{1}_{\{A \cap B^c\}}$, we get

$$\begin{aligned} &\left| \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{A_{j,k}\}} - \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \right| \\ &\le \widetilde{\mathbb{P}}_{\theta_0} \left(\left\{ X_{j-1} \in D_n^k \right\} \cap \bigcup_{i \ne j,\, i = n^{1/4}}^{n^{1/2}} \{X_{i-1} \in D_n\} \right) \\ &\le \sum_{i \ne j,\, i = n^{1/4}}^{n^{1/2}} \widetilde{\mathbb{P}}_{\theta_0} \left(X_{i-1} \in D_n, X_{j-1} \in D_n^k \right) \le C_1 n^{-3/2}. \end{aligned} \tag{32}$$

Further, since $(\widehat{\varepsilon}_i)$ are i.i.d $N(0,1)$ and $\widehat{\varepsilon}_i$ is independent of $\mathcal{F}_{j-1}^X$ for $i \ge j$

$$\begin{aligned} &\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \\ &= \widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \exp\left(-\frac{\lambda_k^2}{2} X_{j-1}^2 \sum_{i=j}^{n^{1/2}} Q^2(i-j) - \frac{\mathbf{i}\lambda_k}{2} X_{j-1}^2 \sum_{i=j}^{n^{1/2}} Q^2(i-j) \right) \end{aligned}$$

$$
\begin{aligned}
&= \widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \exp\left(-\frac{\lambda_k^2 \theta_0^2}{2} \sum_{i=0}^{n^{1/2}-j} Q^2(i) - \frac{\mathbf{i}\lambda_k \theta_0^2}{2} \sum_{i=0}^{n^{1/2}-j} Q^2(i)\right) + O(n^{-2}) \\
&= \widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \exp\left(-\frac{\lambda_k^2 \beta^2}{2} - \frac{\mathbf{i}\lambda_k \beta^2}{2} + \frac{\theta_0^2}{2} \sum_{i=n^{1/2}-j+1}^{\infty} Q^2(i)(\lambda_k^2 + \mathbf{i}\lambda_k)\right) + O(n^{-2}).
\end{aligned}
$$

By (46),

$$
\begin{aligned}
&\widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \exp\left(-\frac{1}{2}\lambda_k^2 \beta^2 - \mathbf{i}\frac{1}{2}\lambda_k \beta^2\right) \\
&\quad = \exp\left(-\frac{1}{2}\lambda_k^2 \beta^2 - \mathbf{i}\frac{1}{2}\lambda_k \beta^2\right) \varpi(u_k - u_{k-1}) n^{-1} + O(n^{-2}),
\end{aligned}
$$

and thus we have

$$
\begin{aligned}
&\left|\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} - \mathrm{e}^{-\frac{1}{2}\lambda_k^2 \beta^2 - \mathbf{i}\frac{1}{2}\lambda_k \beta^2} \varpi(u_k - u_{k-1}) n^{-1}\right| \\
&\quad \le \left|\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}\lambda_k \Psi_{j,k}} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} - \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{-\frac{1}{2}\lambda_k^2 \beta^2 - \mathbf{i}\frac{1}{2}\lambda_k \beta^2} \mathbf{1}_{\{X_{j-1} \in D_n^k\}}\right| + O(n^{-2}) \\
&\quad \le \widetilde{\mathbb{E}}_{\theta_0} \mathbf{1}_{\{X_{j-1} \in D_n^k\}} \left|\exp\left(\frac{1}{2}\theta_0^2 \sum_{i=n^{1/2}-j+1}^{\infty} Q^2(i)(\lambda_k^2 + \mathbf{i}\lambda_k)\right) - 1\right| + O(n^{-2}) \\
&\quad \le \widetilde{\mathbb{P}}_{\theta_0}\left(X_{j-1} \in D_n^k\right) C_2 b^{2(n^{1/2}-j)} + O(n^{-2}) \le C_3 n^{-1} b^{n^{1/2}-j} + O(n^{-2}).
\end{aligned}
$$

Plugging this bound and (32) into (31), we obtain

$$
\left|\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} - \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1}\right| \le C_3 n^{-1} b^{n^{1/2}-j} + C_4 n^{-3/2},
$$

and in turn (28):

$$
\begin{aligned}
&\left|\sum_{j=1}^{n^{1/2}} \widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} - \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1/2}\right| \\
&\quad \le \sum_{j=n^{1/4}}^{n^{1/2}} \left|\widetilde{\mathbb{E}}_{\theta_0} \mathrm{e}^{\mathbf{i}S_{n,1}} \mathbf{1}_{\{A_{j,k}\}} - \mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k \beta^2 - \frac{1}{2}\beta^2 \lambda_k^2} (u_k - u_{k-1}) \varpi n^{-1}\right|
\end{aligned}
$$

$$+\left|\sum_{j<n^{1/4}}\widetilde{\mathbb{E}}_{\theta_0}\mathrm{e}^{\mathbf{i}S_{n,1}}\mathbf{1}_{\{A_{j,k}\}}\right|+\left|\sum_{j<n^{1/4}}\mathrm{e}^{-\frac{\mathbf{i}}{2}\lambda_k\beta^2-\frac{1}{2}\beta^2\lambda_k^2}(u_k-u_{k-1})\varpi n^{-1}\right|$$
$$\leq C_3n^{-1}\sum_{j=n^{1/4}}^{n^{1/2}}b^{n^{1/2}-j}+C_4n^{1/2}n^{-3/2}+n^{1/4}\widetilde{\mathbb{P}}_{\theta_0}(A_{1,k})+C_5n^{-3/4}\leq C_6n^{-3/4}.$$

By setting all λ_k's to zero, we also get

$$\widetilde{\mathbb{P}}_{\theta_0}(A_1)=\sum_{k=1}^{d}\sum_{j=1}^{n^{1/2}}\widetilde{\mathbb{P}}_{\theta_0}(A_{j,k})=\sum_{k=1}^{d}(u_k-u_{k-1})\varpi n^{-1/2}+O(n^{-3/4}). \tag{33}$$

Further,

$$\begin{aligned}\widetilde{\mathbb{P}}_{\theta_0}(A_0)&=1-\widetilde{\mathbb{P}}_{\theta_0}\left(\bigcup_{j=1}^{n^{1/2}}\{X_{j-1}\in D_n\}\right)\\&\geq 1-\sum_{k=1}^{d}\sum_{j=1}^{n^{1/2}}\widetilde{\mathbb{P}}_{\theta_0}(X_{j-1}\in D_n^k)\\&\overset{\dagger}{=}1-\sum_{k=1}^{d}\sum_{j=1}^{n^{1/2}}(u_k-u_{k-1})\varpi n^{-1}+O(n^{-3/2})\\&=1-\sum_{k=1}^{d}(u_k-u_{k-1})\varpi n^{-1/2}+O(n^{-3/2})\end{aligned} \tag{34}$$

where in the equality † we used (46). On the other hand, $\widetilde{\mathbb{P}}_{\theta_0}(A_0)\leq 1-\widetilde{\mathbb{P}}_{\theta_0}(A_1)$ and the estimate (29) follows from (33) and (34) and the asymptotic

$$\widetilde{\mathbb{E}}_{\theta_0}\mathrm{e}^{\mathbf{i}S_{n,1}}\mathbf{1}_{\{A_0\}}=\widetilde{\mathbb{P}}_{\theta_0}(A_0)+O(b^{n^{1/4}}).$$

Finally, (30) follows since $\widetilde{\mathbb{P}}_{\theta_0}(A_{2+})=1-\widetilde{\mathbb{P}}_{\theta_0}(A_0)-\widetilde{\mathbb{P}}_{\theta_0}(A_1)=O(n^{-3/4})$. □

2.4 Tightness

In Lemmas 4 and 5, we check the tightness conditions (1.1) and (1.2) of Theorem I.10.2, Ibragimov and Has'minskiĭ (1981), respectively.

Lemma 4 *For any compact $K\subset\Theta$, there is a constant $C>0$, such that*

$$\sup_{|u_1|\leq R,|u_2|\leq R}|u_2-u_1|^{-1}\mathbb{E}_{\theta_0}\left(\sqrt{Z_n(u_2)}-\sqrt{Z_n(u_1)}\right)^2\leq C(1+R^2),$$
$$forall\quad \theta_0\in K\quad and\quad R\geq 0. \tag{35}$$

Proof Suppose $u_2 \geq u_1$, then using the elementary inequality $\ln \frac{1}{x} \geq 2(1-\sqrt{x})$, $x > 0$ we get

$$\mathbb{E}_{\theta_0}\left(\sqrt{Z_n(u_2)}-\sqrt{Z_n(u_1)}\right)^2 = \mathbb{E}_{\theta_0+u_1/n}\left(\sqrt{\frac{Z_n(u_2)}{Z_n(u_1)}}-1\right)^2$$
$$= 2\mathbb{E}_{\theta_0+u_1/n}\left(1-\sqrt{\frac{Z_n(u_2)}{Z_n(u_1)}}\right) \leq \mathbb{E}_{\theta_0+u_1/n} \ln \frac{Z_n(u_1)}{Z_n(u_2)}.$$

Similarly to (17), we find that under $\mathbb{P}_{\theta_0+u_1/n}$,

$$\ln \frac{Z_n(u_1)}{Z_n(u_2)} = \frac{1}{1+\gamma}\sum_{j=1}^n \left(X_{j-1}(\rho^+-\rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^2\}}+\delta_{j-1}^2\right)$$
$$\times\left(\sqrt{1+\gamma}\widehat{\varepsilon}_j + \frac{1}{2}\left(X_{j-1}(\rho^+-\rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^2\}}+\delta_{j-1}^2\right)\right),$$

where (δ_j^2) is defined in (19). Note that $\widehat{\varepsilon}_j$ is independent of $\mathcal{F}_{j-1}^X$ under $\mathbb{P}_{\theta_0+u_1/n}$ and, as in the proof of Lemma 1, $\mathbb{E}_{\theta_0+u_1/n}\left(\delta_j^2\right)^2 \leq C_1(u_2-u_1)n^{-1}$. Hence,

$$\mathbb{E}_{\theta_0}\left(\sqrt{Z_n(u_2)}-\sqrt{Z_n(u_1)}\right)^2$$
$$\leq \sum_{j=1}^n \mathbb{E}_{\theta_0+u_1/n}\left(X_{j-1}(\rho^+-\rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^2\}}+\delta_{j-1}^2\right)^2 \leq C_2|u_2-u_1|,$$

where C_2 depends only on K. By symmetry, the same inequality holds when $u_2 \leq u_1$ and (35) follows. □

Lemma 5 *For any $p \geq 1$, there is a constant $C(p)$, such that*

$$\mathbb{E}_{\theta_0} Z_n^{1/2}(u) \leq \frac{C(p)}{|u|^p}, \quad u \in \mathbb{U}_n := n(\Theta-\theta_0).$$

Proof The proof is an adaptation of the analogous Lemma 2.2 in Chan and Kutoyants (2012). We shall assume $\theta_0 > 0$ and $u > 0$, omitting the similar complementary cases (recall that $0 \notin \Theta$). Note that for a constant $c > 0$,

$$\mathbb{E}_{\theta_0} Z_n^{1/2}(u) = \mathbb{E}_{\theta_0} Z_n^{1/2}(u)\mathbf{1}_{\{Z_n^{1/2}(u)\geq \mathrm{e}^{-cu}\}} + \mathbb{E}_{\theta_0} Z_n^{1/2}(u)\mathbf{1}_{\{Z_n^{1/2}(u)<\mathrm{e}^{-cu}\}}$$
$$\leq \left(\mathbb{E}_{\theta_0} Z_n(u)\right)^{1/2}\left(\mathbb{P}_{\theta_0}\left(Z_n^{1/2}(u)\geq \mathrm{e}^{-cu}\right)\right)^{1/2} + \mathrm{e}^{-\frac{c}{2}u}$$
$$= \left(\mathbb{P}_{\theta_0}\left(\ln Z_n^{1/2}(u) \geq -cu\right)\right)^{1/2} + \mathrm{e}^{-\frac{c}{2}u},$$

and hence it is enough to check the large deviation bound

$$\mathbb{P}_{\theta_0}\left(\frac{1}{2}\ln Z_n(u) \geq -cu\right) \leq \frac{C(p)}{u^p}, \tag{36}$$

for some positive constants c and $C(p)$ and all $p \geq 1$.

For $u_1 := u > 0$, the formula (18) gives

$$\begin{aligned}\ln Z_n(u) = &-\frac{1}{1+\gamma}\sum_{j=1}^{n}\left(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^1\}} + \delta_{j-1}\right)\\ &\times\left(\sqrt{1+\gamma}\,\widehat{\varepsilon}_j + \frac{1}{2}\left(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^1\}} + \delta_{j-1}\right)\right)\\ &= \sum_{j=1}^{n}\left(\widehat{\varepsilon}_j V_{j-1} - \frac{1}{2}V_{j-1}^2\right)\end{aligned}$$

where $D_n^1 = [\theta_0, \theta_0 + u/n]$, the sequence (δ_j) is generated by (19) with $k = 1$, and

$$V_{j-1} := -\frac{1}{\sqrt{1+\gamma}}\left(X_{j-1}(\rho^+ - \rho^-)\mathbf{1}_{\{X_{j-1}\in D_n^1\}} + \delta_{j-1}\right).$$

Further,

$$\begin{aligned}&\mathbb{P}_{\theta_0}\left(\frac{1}{2}\ln Z_n(u) \geq -cu\right)\\ &= \mathbb{P}_{\theta_0}\left(\sum_{j=1}^{n}\left(\widehat{\varepsilon}_j\left(\frac{1}{2}V_{j-1}\right) - \frac{1}{2}\left(\frac{1}{2}V_{j-1}\right)^2\right) - \frac{1}{8}\sum_{j=1}^{n}V_{j-1}^2 \geq -cu\right)\\ &\leq \mathbb{P}_{\theta_0}\left(\sum_{j=1}^{n}\left(\widehat{\varepsilon}_j\left(\frac{1}{2}V_{j-1}\right) - \frac{1}{2}\left(\frac{1}{2}V_{j-1}\right)^2\right) \geq cu\right) + \mathbb{P}_{\theta_0}\left(-\frac{1}{8}\sum_{j=1}^{n}V_{j-1}^2 \geq -2cu\right),\end{aligned}$$

and since

$$\begin{aligned}&\mathbb{P}_{\theta_0}\left(\sum_{j=1}^{n}\left(\widehat{\varepsilon}_j\left(\frac{1}{2}V_{j-1}\right) - \frac{1}{2}\left(\frac{1}{2}V_{j-1}\right)^2\right) \geq cu\right)\\ &\qquad\leq \mathrm{e}^{-cu}\mathbb{E}_{\theta_0}\exp\left(\sum_{j=1}^{n}\left(\widehat{\varepsilon}_j\left(\frac{1}{2}V_{j-1}\right) - \frac{1}{2}\left(\frac{1}{2}V_{j-1}\right)^2\right)\right) = \mathrm{e}^{-cu},\end{aligned}$$

the bound (36) holds, if we show that for some positive constant c and all $p \geq 1$,

$$\mathbb{P}_{\theta_0}\left(\sum_{j=0}^{n-1} V_j^2 \le cu\right) \le \frac{C(p)}{u^p}. \tag{37}$$

To this end, we shall split the consideration to the cases $u \le n^s$ and $u > n^s$, where $s \in (0, 1)$ is a constant to be chosen later on, depending on p in (37).

Case $u \le n^s$. In this case, $D_n^1 \subseteq \left[\theta_0, \theta_0 + \frac{1}{n^{1-s}}\right]$ and since

$$\delta_j = b\delta_{j-1} - c(\rho^+ - \rho^-)X_{j-1}\mathbf{1}_{\{X_{j-1}\in D_n^1\}}, \quad j \ge 1,$$

subject to $\delta_0 = 0$, it follows that

$$\begin{aligned} |\delta_j| &\le |c||\rho^+ - \rho^-| \left(\theta_0 + n^{s-1}\right) \frac{1}{1-|b|} \\ &= \frac{|a|\gamma}{1-|a|+\gamma}|\rho^+ - \rho^-| \left(\theta_0 + n^{s-1}\right) \le |a||\rho^+ - \rho^-| \left(\theta_0 + n^{s-1}\right), \end{aligned}$$

where we used the definitions $c := \frac{a\gamma}{1+\gamma}$ and $b := \frac{a}{1+\gamma}$ and the assumption $|a| < 1$. Further,

$$\begin{aligned} V_j^2 &= \frac{1}{1+\gamma}\left(X_j(\rho^+ - \rho^-)\mathbf{1}_{\{X_j\in D_n^1\}} + \delta_j\right)^2 \\ &\ge \frac{1}{1+\gamma}\left(X_j(\rho^+ - \rho^-) + \delta_j\right)^2 \mathbf{1}_{\{X_j\in D_n^1\}} \\ &\ge \frac{(\rho^+ - \rho^-)^2}{1+\gamma}\left(\theta_0 - |a|\left(\theta_0 + n^{s-1}\right)\right)^2 \mathbf{1}_{\{X_j\in D_n^1\}} \\ &\ge \frac{(\rho^+ - \rho^-)^2}{1+\gamma}\frac{1}{4}\theta_0^2\left(1-|a|\right)^2 \mathbf{1}_{\{X_j\in D_n^1\}} := C_1\mathbf{1}_{\{X_j\in D_n^1\}} \end{aligned}$$

where the latter inequality holds for all n large enough. Consequently,

$$\mathbb{P}_{\theta_0}\left(\sum_{j=0}^{n-1} V_j^2 \le cu\right) \le \mathbb{P}_{\theta_0}\left(\sum_{j=0}^{n-1} \mathbf{1}_{\{X_j\in D_n^1\}} \le \frac{c}{C_1}u\right) \le \mathbb{P}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1} \mathbf{1}_{\{X_j\in D_n^1\}} \le \frac{c}{C_1}u\right).$$

By Lemma 7, the process (X_j) is geometric mixing and we have

$$\begin{aligned} \mathbb{P}_{\theta_0}(X_j \in D_n^1) &= \mathbb{E}_{\theta_0}\int_{\theta_0}^{\theta_0+u/n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(t-\xi_{j-1}-f(X_{j-1},\theta_0))^2}\mathrm{d}t \\ &\ge \frac{u}{n}\frac{1}{\sqrt{2\pi}}\mathbb{E}_{\theta_0}\inf_{t\in[\theta_0,\theta_0+1]} \mathrm{e}^{-\frac{1}{2}\left(t-\xi_{j-1}-f(X_{j-1},\theta_0)\right)^2} \ge C_2\frac{u}{n}, \end{aligned}$$

where the constant $C_2 > 0$ can be chosen for all $\theta_0 \in K$ to be independent of j and n by the ergodic properties of (X_j, ξ_j) from Lemma 7. Hence, with $c := \frac{1}{2}C_1C_2$, for an integer $p \geq 1$,

$$\begin{aligned}
\mathbb{P}_{\theta_0}\left(\sum_{j=0}^{n-1} V_j^2 \leq cu\right) &\leq \mathbb{P}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1} \mathbf{1}_{\{X_j \in D_n^1\}} \leq \frac{c}{C_1}u\right) \\
&\leq \mathbb{P}_{\theta_0}\left(\left|\sum_{j=n^{1/2}}^{n-1}\left(\mathbf{1}_{\{X_j \in D_n^1\}} - \mathbb{E}_{\theta_0}\mathbf{1}_{\{X_j \in D_n^1\}}\right)\right| \geq \frac{1}{3}C_2 u\right) \\
&\leq (C_2/3)^{-2p}\frac{1}{u^{2p}}\mathbb{E}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p},
\end{aligned}$$

where we defined $\eta_j := \mathbf{1}_{\{X_j \in D_n^1\}} - \mathbb{E}_{\theta_0}\mathbf{1}_{\{X_j \in D_n^1\}}$. Since $|\eta_j| \leq 2$, by Lemma 7,

$$\begin{aligned}
&\left|\mathbb{E}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p} - \widetilde{\mathbb{E}}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p}\right| \\
&\quad\leq n^{2p}\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(\left(\frac{1}{n}\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p}\middle|\mathcal{F}_0\right) - \widetilde{\mathbb{E}}_{\theta_0}\left(\frac{1}{n}\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p}\right| \leq Cn^{2p}r^{n^{1/2}} \leq 1,
\end{aligned}$$

for all n large enough. Hence, it is enough to check

$$\widetilde{\mathbb{E}}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p} \leq C(p)u^p. \tag{38}$$

To estimate the latter expectation, we shall apply the covariance inequality (8.1) from Dedecker and Doukhan (2003):

$$\begin{aligned}
\widetilde{\mathbb{E}}_{\theta_0}\left(\sum_{j=n^{1/2}}^{n-1}\eta_j\right)^{2p} &= \widetilde{\mathbb{E}}_{\theta_0}\left(\sum_{j=0}^{n-n^{1/2}-1}\eta_j\right)^{2p} \\
&\leq \left(4pn\sum_{j=0}^{n}\left(\widetilde{\mathbb{E}}_{\theta_0}\left|\eta_0\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_0)\right|^p\right)^{1/p}\right)^p.
\end{aligned} \tag{39}$$

Since $|\eta_0| \leq 2$,

$$\widetilde{\mathbb{E}}_{\theta_0}\left|\eta_0\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_0)\right|^p \leq 4^p\,\widetilde{\mathbb{E}}_{\theta_0}\left|\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_0)\right|^p \leq 4^p\left(\frac{u}{n}C_4Cr^j\right)^p, \tag{40}$$

where the latter inequality holds by Lemma 7, since

$$\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_0) = \widetilde{\mathbb{E}}_{\theta_0}\left(\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_{j-1})\big|\mathcal{F}_0\right)$$

and $\left|\widetilde{\mathbb{E}}_{\theta_0}(\eta_j|\mathcal{F}_{j-1})\right| \le C_4\frac{u}{n}$. Plugging the bound (40) into (39), we obtain (38) and consequently (37) for $u \le n^s$.

Case $u > n^s$. For a fixed integer k and all $j > k$, define

$$\Gamma_{j-1,k} := \bigcap_{i=j-k}^{j-1} \{X_i \notin \Theta\}$$

and note that on $\Gamma_{j-1,k}$ we have

$$|\delta_j|/|\rho^+ - \rho^-| = \left|c\sum_{i=1}^{j-k} b^{j-i} X_{i-1}\mathbf{1}_{\{X_{i-1}\in D_n^1\}}\right| \le |c| \sup|\Theta| \frac{|b|^k}{1-|b|} =: C_1 b^k.$$

Now let k' be such that, $C_1 b^{k'} \le \frac{1}{2}\theta_0$, then (recall that both θ_0 and u are positive)

$$\begin{aligned} V_j^2 &\ge \frac{1}{1+\gamma}\left(X_j(\rho^+ - \rho^-) + \delta_j\right)^2 \mathbf{1}_{\{X_j\in D_n^1\}}\mathbf{1}_{\{\Gamma_{j-1,k'}\}} \\ &\ge \frac{(\rho^+ - \rho^-)^2}{1+\gamma}\frac{1}{4}\theta_0^2 \mathbf{1}_{\{X_j\in D_n^1\}}\mathbf{1}_{\{\Gamma_{j-1,k'}\}} =: C_2\mathbf{1}_{\{X_j\in D_n^1\}}\mathbf{1}_{\{\Gamma_{j-1,k'}\}}. \end{aligned}$$

Define $W_j := \mathbf{1}_{\{X_j\in D_n^1\}}\mathbf{1}_{\{\Gamma_{j-1,k'}\}}$ and note that

$$\begin{aligned} \mathbb{E}_{\theta_0} W_j &= \mathbb{E}_{\theta_0}\mathbf{1}_{\{\Gamma_{j-1,k}\}}\mathbb{P}_{\theta_0}\left(X_j \in D_n^1 | X_{j-1}, \xi_{j-1}\right) \\ &\ge \frac{u}{n}\mathbb{E}_{\theta_0}\mathbf{1}_{\{\Gamma_{j-1,k'}\}}\frac{1}{\sqrt{2\pi}}\inf_{t\in\Theta} \mathrm{e}^{-\frac{1}{2}(t-\xi_{j-1}-f(X_{j-1};\theta_0))^2} \ge C_3\frac{u}{n}, \end{aligned}$$

where the positive constant C_3 can be chosen independent of j and n due to the ergodic properties of (X_j, ξ_j) from Lemma 7. Hence, with $c := \frac{1}{2}C_3C_2$ and any integer $p > 1$,

$$\begin{aligned} \mathbb{P}_{\theta_0}\left(\sum_{j=0}^{n-1} V_j^2 \le cu\right) &\le \mathbb{P}_{\theta_0}\left(\sum_{j=k'}^{n-1} W_j \le \frac{1}{2}C_3 u\right) \\ &\le \mathbb{P}_{\theta_0}\left(\left|\sum_{j=k'}^{n-1}\left(W_j - \mathbb{E}_{\theta_0}W_j\right)\right| \ge \frac{1}{3}uC_3\right) \\ &\le (3/C_3)^{2p+2}\frac{1}{u^{2p+2}}\mathbb{E}_{\theta_0}\left(\sum_{j=k'}^{n-1}\left(W_j - \mathbb{E}_{\theta_0}W_j\right)\right)^{2p+2} \end{aligned}$$

for all sufficiently large n. Now (37) follows if we show that for a positive constant $C(p)$,

$$\mathbb{E}_{\theta_0}\left(\sum_{j=k'}^{n-1}\left(W_j - \mathbb{E}_{\theta_0} W_j\right)\right)^{2p+2} \leq C(p)u^{p+2}, \quad \forall\, u > n^s. \tag{41}$$

To this end, we shall use the Marcinkiewicz–Zygmund inequality from Doukhan and Louhichi (1999). For a sequence of random variables $(\eta_j)_{j\in\mathbb{N}}$, the coefficient of weak dependence is defined

$$C_{t,q} := \sup \left|\operatorname{cov}\left(\eta_{t_1}\ldots\eta_{t_m}, \eta_{t_{m+1}}\ldots\eta_{t_q}\right)\right|,$$

where the supremum is taken over all $\{t_1, \ldots, t_q\}$, such that $1 \leq t_1 \leq \ldots \leq t_q$ and m, t satisfy $t_{m+1} - t_m = t$.

Theorem 2 (Theorem 1, Doukhan and Louhichi 1999) *Let $(\eta_j)_{j\in\mathbb{N}}$ be a sequence of central random variables such that for a fixed integer $q \geq 2$,*

$$C_{t,q} = O(t^{-q/2}) \quad as \quad t \to \infty. \tag{42}$$

Then there exists a positive constant B, independent of n, for which

$$\left|\mathbb{E}\left(\sum_{j=0}^{n-1}\eta_j\right)^q\right| \leq Bn^{q/2}. \tag{43}$$

We shall apply this theorem to the bounded sequence $\eta_j := W_j - \mathbb{E}_{\theta_0} W_j$. Since η_j is a function of $(X_{j-k'}, \ldots, X_j)$, it inherits the mixing property (49). More precisely, with

$$\begin{aligned}
h(x, y) &:= \mathbb{E}_{\theta_0}\left(\eta_{t_1}\ldots\eta_{t_m} | X_{t_m} = x, \xi_{t_m} = y\right),\\
g(x, y) &:= \mathbb{E}_{\theta_0}\left(\eta_{t_{m+1}}\ldots\eta_{t_q} | X_{t_{m+1}-k'} = x, \xi_{t_{m+1}-k'} = y\right)
\end{aligned}$$

by the Markov property of (X_j, ξ_j), for $t \geq k'$

$$\begin{aligned}
C_{t,q} &= \left|\mathbb{E}_{\theta_0}\eta_{t_1}\ldots\eta_{t_q} - \mathbb{E}_{\theta_0}\eta_{t_1}\ldots\eta_{t_m}\mathbb{E}_{\theta_0}\eta_{t_{m+1}}\ldots\eta_{t_q}\right|\\
&= \left|\mathbb{E}_{\theta_0} h(X_{t_m}, \xi_{t_m}) g(X_{t_{m+1}-k'}, \xi_{t_{m+1}-k'})\right.\\
&\quad \left. - \mathbb{E}_{\theta_0} h(X_{t_m}, \xi_{t_m})\mathbb{E}_{\theta_0} g(X_{t_{m+1}-k'}, \xi_{t_{m+1}-k'})\right|\\
&\leq 2^q C r^{t-k'}(1 + r^{t_m}) \leq 2^{q+1} C r^{t-k'},
\end{aligned}$$

which clearly satisfies (42). Since $n < u^{1/s}$, (43) now gives

$$\mathbb{E}_{\theta_0}\left(\sum_{j=1}^{n}\left(W_j-\mathbb{E}_{\theta_0}W_j\right)\right)^{2p+2}\le Bn^{p+1}\le Bu^{(p+1)/s},$$

and the bound (41) follows, if we choose $s := (p+1)/(p+2)$. □

3 Simulated experiments

The objective of this section is to illustrate the results of Theorem 1 by means of a simulation. To this end, we fixed the following values of the parameters

$$\theta_0 = 1.5,\ \rho^+ = 0.9,\ \rho_- = -0.5,\ a = 0.9$$

and estimated the root mean square errors of the Bayes estimator $\widetilde{\theta}_n$ and the MLE $\widehat{\theta}_n$ by averaging over a large number of Monte Carlo trials. This has been done in two ways: by computing the estimators, based on simulated data, and computing the corresponding limit quantities, based on simulated process from the limit experiment. The practical advantage, offered by Theorem 1, is that the latter simulation requires much less CPU time than the former.

3.1 Simulated data

Using the recursions (2) and (3), we generated a large number ($M = 20.000$) of sample paths. For each path, we computed the Bayes estimator $\widetilde{\theta}_n$, using the formula (9) and the uniform prior on the interval $\Theta := (1, 2)$. Then we calculated the normalized empirical root mean square error for a number of sample sizes

$$n\sqrt{\widehat{\mathbb{E}}_{\theta_0}\left(\widetilde{\theta}_n-\theta_0\right)^2},$$

where $\widehat{\mathbb{E}}_{\theta_0}$ denotes averaging over the paths.

Similarly, we computed the (central) MLE and the *pseudo* MLE, which assumes independent innovations with the same variance $1 + 1/(1 - a^2)$. The results, depicted at Fig. 1, indicate that the errors converge as the sample size n increases and that the Bayes estimator performs better than the others for smaller sample sizes as well.

3.2 Simulated limit experiment

While the distribution of the random variable $\tilde{u}$, defined in Theorem 1, cannot be computed in a closed form, it is easy to sample and the expectations can be approximated by averaging over Monte Carlo trials. To this end, we estimated the value of

$$\varpi = \int_{\mathbb{R}} p(\theta_0, y; \theta_0)\mathrm{d}y \approx 0.0576$$

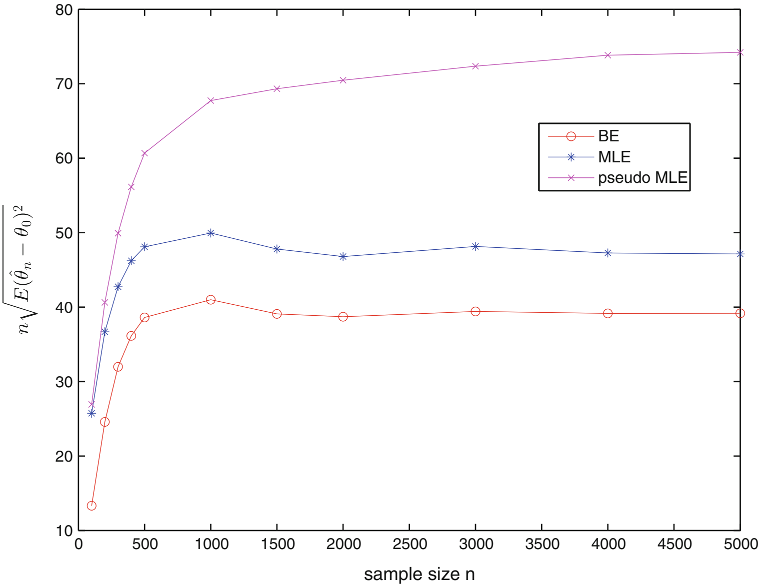

Fig. 1 The normalized empirical root mean squares of the BE, MLE and pseudo MLE versus the sample sizes

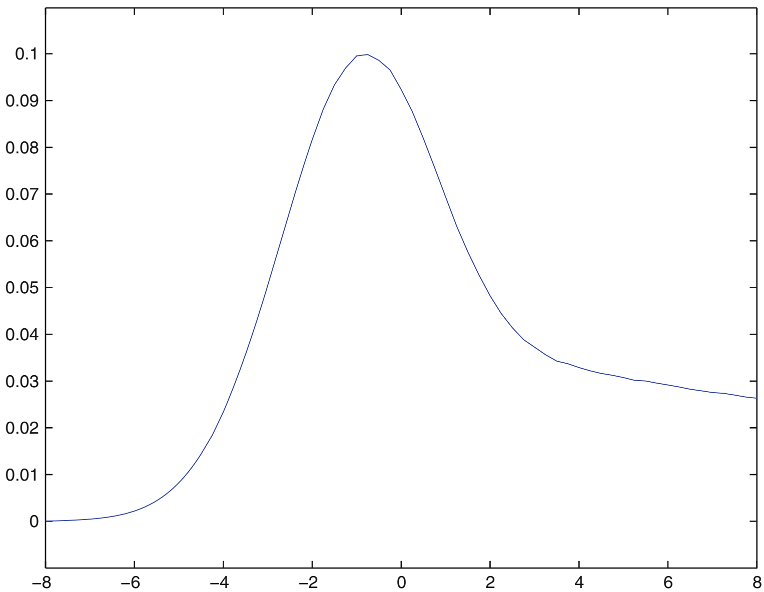

Fig. 2 The estimated stationary density of (X_j)

using standard kernel estimator, applied to a single long trajectory of (X_j). Figure 2 depicts the estimated marginal density

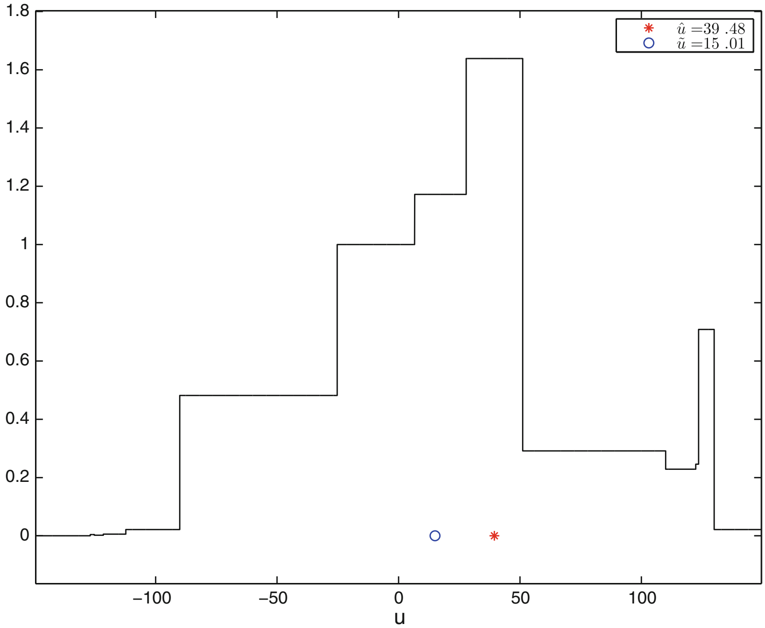

Fig. 3 A typical sample path of the limit likelihood ratio process $Z(u)$, $u \in \mathbb{R}$. The marks are the positions of $\hat{u}$ and $\tilde{u}$

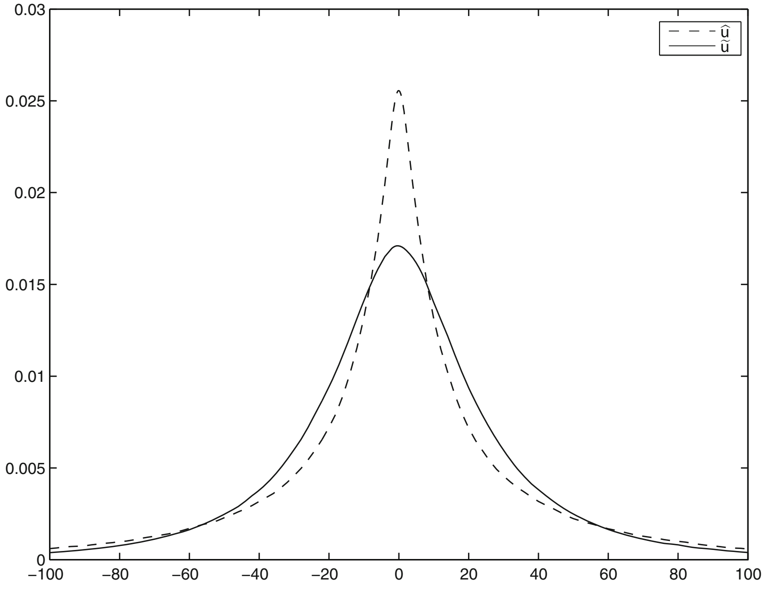

Fig. 4 The estimated probability densities of $\tilde{u}$ and $\hat{u}$

$$\int_{\mathbb{R}} p(x, y; \theta_0)\mathrm{d}y, \quad x \in [-8, 8].$$

Next, we generated a large number ($M = 10^6$) of samples from the compound Poisson process $Z(u)$, defined in (11) and computed the approximate root mean square errors

$$\sqrt{\widehat{\mathbb{E}}_{\theta_0}(\tilde{u})^2} \approx 38.64 \quad \text{and} \quad \sqrt{\widehat{\mathbb{E}}_{\theta_0}(\hat{u})^2} \approx 46.88,$$

where $\hat{u}$ is the central maximizer of $Z(u)$ and $\widehat{\mathbb{E}}_{\theta_0}$ denotes the empirical expectation. Note that the obtained estimates are at good correspondence with the plots at Fig. 1.

The typical realization of $Z(u)$ along with $\tilde{u}$ and $\hat{u}$ are plotted at Fig. 3. The densities of $\hat{u}$ and $\tilde{u}$, whose kernel estimates are depicted at Fig. 4, appear to be heavy tailed.

Appendix: Ergodic lemmas used in the proof

The proofs in Sect. 2 use the ergodic properties of the processes, summarized in the following lemmas. Our standing assumption is (10).

Lemma 6 *For all integers $j \geq 0$ and $p \geq 1$,*

$$\mathbb{P}_{\theta_0}\left(X_j \in [\theta_0, \theta_0 + v/n]\right) \leq \frac{|v|}{n}, \tag{44}$$

and

$$\mathbb{E}_{\theta_0}\left(|X_j|^p + |\xi_j|^p \middle| X_0 = x, \xi_0 = y\right) \leq r_1^j R_1\left(|x|^p + |y|^p\right) + R_2, \tag{45}$$

with a positive constant $r_1 < 1$ and constants R_1 and R_2, independent of θ_0.

Proof For $j \geq 1$,

$$\mathbb{P}_{\theta_0}\left(X_j \in [\theta_0, \theta_0 + v/n]\right) = \mathbb{E}_{\theta_0}\int_{\theta_0}^{\theta_0+v/n} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-f(X_{j-1},\theta_0)-\xi_{j-1})^2}\mathrm{d}x \leq \frac{v}{n}.$$

Further, by Jensen's inequality

$$|\xi_j|^p \leq \left(|a|\xi_{j-1} + (1-|a|)\frac{1}{1-|a|}|\zeta_j|\right)^p \leq |a||\xi_{j-1}|^p + (1-|a|)^{1-p}|\zeta_j|^p,$$

and hence

$$\mathbb{E}_{\theta_0}\left(|\xi_j|^p \middle| \xi_0 = y\right) \leq |a|^j|y|^p + C_1.$$

Similarly, with $\rho := |\rho^+| \vee |\rho^-|$

$$|X_j|^p \leq \left(\rho|X_{j-1}| + |\xi_{j-1}| + |\varepsilon_j|\right)^p = \rho|X_{j-1}|^p + \left(\frac{2}{1-\rho}\right)^{p-1}\left(|\xi_{j-1}|^p + |\varepsilon_j|^p\right),$$

and

$$\mathbb{E}_{\theta_0}\left(|X_j|^p\Big|X_0=x,\xi_0=y\right)\le \rho^j|x|^p+C_2|a|^j|y|^p+C_3,$$

which gives (45). □

Lemma 7 *The Markov chain (X_j,ξ_j) has the unique invariant measure under $\mathbb{P}_{\theta_0}$, with uniformly bounded probability density $p(x,y;\theta_0)$ satisfying*

$$\begin{aligned}&\widetilde{\mathbb{P}}_{\theta_0}\left(X_j\in[\theta_0,\theta_0+v/n]\right)\\&\quad=\int_{\theta_0}^{\theta_0+v/n}\int_{\mathbb{R}}p(x,y;\theta_0)\mathrm{d}x\mathrm{d}y=\frac{v}{n}\int_{\mathbb{R}}p(\theta_0,y;\theta_0)\mathrm{d}y+O(n^{-2}),\end{aligned}\tag{46}$$

where $\widetilde{\mathbb{P}}_{\theta_0}$ is the corresponding stationary probability on $(\Omega,\mathcal{F})$.

Moreover, the chain is geometrically ergodic, i.e., there exist positive constants C and $r<1$, such that for a measurable function $|h|\le 1$ and $m\ge k$

$$\left|\mathbb{E}_{\theta_0}\left(h(X_m,\xi_m)|X_k=x,\xi_k=y\right)-\widetilde{\mathbb{E}}_{\theta_0}h(X_k,\xi_k)\right|\le Cr^{m-k}(|x|+|y|),\quad x,y\in\mathbb{R},\tag{47}$$

and consequently, for an $\mathcal{F}_{m,\infty}$-measurable random variable $|H|\le 1$

$$\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(H|\mathcal{F}_k\right)-\widetilde{\mathbb{E}}_{\theta_0}H\right|\le Cr^{m-k}.\tag{48}$$

Finally, (X_j,ξ_j) is geometrically mixing, i.e., for measurable functions $|g|\le 1$, $|h|\le 1$

$$\left|\mathbb{E}_{\theta_0}g(X_k,\xi_k)h(X_{k+m},\xi_{k+m})-\mathbb{E}_{\theta_0}g(X_k,\xi_k)\mathbb{E}_{\theta_0}h(X_{k+m},\xi_{k+m})\right|\le Cr^m(1+r^k).\tag{49}$$

In particular, (48) and (49) hold with the stationary expectation $\widetilde{\mathbb{E}}_{\theta_0}$.

Proof The transition kernel of the process (X_j,ξ_j) has a positive density with respect to the Lebesgue measure:

$$(P\mathbf{1}_{\{A\}})(x,y):=\int_A\frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(u-f(x,\theta_0)-y\right)^2-\frac{1}{2}(v-ay)^2\right)\mathrm{d}u\mathrm{d}v$$

and hence in the terminology of Meyn and Tweedie (2009), it is ψ-irreducible and aperiodic. Further, a ball B_R of radius $R>0$ around the origin is a small set with respect to, e.g., the measure

$$\nu(\mathrm{d}x,\mathrm{d}y):=\mathrm{e}^{-\frac{1}{2}(u^2+v^2)-(1+R)(|u|+|v|)}\mathrm{d}u\mathrm{d}v$$

and $V(x, y) = |x| + |y|$ satisfies the drift condition

$$(PV)(x, y) - V(x, y) \le -\frac{1}{2}(1 - \rho \wedge |a|)V(x, y) + 2\mathbf{1}_{\{(x,y)\in B_R\}}, \quad (x, y) \in \mathbb{R}^2,$$

for sufficiently large R. By Theorem 15.0.1 in Meyn and Tweedie (2009), it follows that there exists a unique invariant probability measure π and for any measurable $h(x, y) \le V(x, y)$,

$$\left| P^n h - \int h \mathrm{d}\pi \right| \le C r^n V(x, y),$$

with positive constants C and $r < 1$, i.e., (47) holds. Since $\widetilde{\mathbb{E}}_{\theta_0} H = \widetilde{\mathbb{E}}_{\theta_0} h(X_m, \xi_m)$ and $\mathbb{E}_{\theta_0}(H|\mathcal{F}_k) = \mathbb{E}_{\theta_0}(h(X_m, \xi_m)|\mathcal{F}_k)$ with $h(x, y) := \mathbb{E}_{\theta_0}(H|X_m = x, \xi_m = y)$, the claim (48) follows from (45) and (47). Since the transition kernel P has a bounded continuously differentiable density with respect to the Lebesgue measure, so does the invariant measure π and (46) follows. The mixing inequality (49) follows from Theorem 16.1.5 in Meyn and Tweedie (2009). □

The theory, used in the proof of the previous lemma, does not directly apply to the Markov chain $(X_j, \xi_j, \widehat{\Xi}_j)$ [see (16) for the definition of $\widehat{\Xi}_j$], since it is generated by a $(3 + d)$-dimensional recursion, driven by two dimensional noise. This typically excludes ψ-irreducibility. Fortunately, for our purposes the following weaker properties are sufficient:

Lemma 8 *The Markov process $(X_j, \xi_j, \widehat{\Xi}_j)$ has the unique invariant measure. Let $\widetilde{\mathbb{P}}_{\theta_0}$ denote the corresponding stationary probability (by uniqueness, the stationary probabilities $\widetilde{\mathbb{P}}_{\theta_0}$, introduced in Lemmas 7 and 8, coincide). Then for a measurable function $h(x, y, z)$, satisfying $|h(x, y, z)| < 1$ and the Lipschitz condition*

$$\begin{aligned} &|h(x, y, z) - h(x, y, z')| \le L\left(1 + |x| + |y| + \|z\| + \|z'\|\right)\|z - z'\|, \\ &\quad x, y \in \mathbb{R}, \ z, z' \in \mathbb{R}^{d+1}, \end{aligned}$$

with a positive constant L,

$$\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m, \xi_m, \widehat{\Xi}_m)\middle|\mathcal{F}_\ell\right) - \widetilde{\mathbb{E}}_{\theta_0} h(X_0, \xi_0, \widehat{\Xi}_0)\right| \le C q^{m-\ell} \tag{50}$$

for some positive constants C and $q < 1$ and all integers $m \ge \ell \ge 0$.

Proof Under the stationary measure $\widetilde{\mathbb{P}}_{\theta_0}$ from Lemma 7, we can extend the definition of (X_j, ξ_j) to the negative integers and define

$$\widehat{\Xi}_0^k := c \sum_{i=-\infty}^{0} b^i \left(X_i - f(X_{i-1}, \theta_0 + u_k/n)\right), \quad k = 0, \ldots, d \tag{51}$$

where $b := \frac{a}{1+\gamma}$ and $c := \frac{a\gamma}{1+\gamma}$. The distribution of $(X_0, \xi_0, \widehat{\Xi}_0)$ is invariant. To establish uniqueness, let μ and μ' be two invariant measures and note that by Lemma 7 their (X, ξ) marginals coincide. Hence,

$$\mu(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = \nu(\mathrm{d}x, \mathrm{d}y)\mu(x, y; \mathrm{d}z), \quad \mu'(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = \nu(\mathrm{d}x, \mathrm{d}y)\mu'(x, y; \mathrm{d}z)$$

where ν is the invariant measure of the process (X_j, ξ_j) and $\mu(x, y; \mathrm{d}z)$ and $\mu'(x, y; \mathrm{d}z)$ are corresponding regular conditional probabilities. Let $(X_j, \xi_j, \widehat{\Xi}_j)$ and $(X_j, \xi_j, \widehat{\Xi}'_j)$ be the solutions of the recursions (2), (3) and (7) with $u := u_k$, $k = 0, \ldots, d$ subject to the initial conditions $(X_0, \xi_0, \widehat{\Xi}_0)$ and $(X_0, \xi_0, \widehat{\Xi}'_0)$, where (X_0, ξ_0) is sampled from ν and $\widehat{\Xi}_0$ and $\widehat{\Xi}'_0$ are sampled from $\mu(X_0, \xi_0; \mathrm{d}z)$ and $\mu'(X_0, \xi_0; \mathrm{d}z)$. Note that $\widehat{\Xi}'_j - \widehat{\Xi}_j = b^j(\widehat{\Xi}'_0 - \widehat{\Xi}_0)$ and hence for any uniformly continuous function g

$$\left|\int g\mathrm{d}\mu - \int g\mathrm{d}\mu'\right| \le \mathbb{E}_{\theta_0}\left|g(X_j, \xi_j, \widehat{\Xi}_j) - g(X_j, \xi_j, \widehat{\Xi}'_j)\right| \xrightarrow{j\to\infty} 0.$$

Since uniformly continuous functions form a measure defining class, the uniqueness follows.

To derive the bound (50), note that for $\ell \le m$

$$\begin{aligned}
\widehat{\Xi}^k_m &= \widehat{\Xi}^k_\ell b^{m-\ell} + c\sum_{j=\ell+1}^{m}\big(X_j - f(X_{j-1}, \theta_0 + u_k/n)\big)\, b^{m-j} \\
&= b^{\frac12(m-\ell)}\left(\widehat{\Xi}^k_\ell b^{\frac12(m-\ell)} + c\sum_{j=\ell+1}^{\frac12(m+\ell)}\big(X_j - f(X_{j-1}, \theta_0 + u_k/n)\big)\, b^{\frac12(m+\ell)-j}\right) \\
&\quad + c\sum_{j=\frac12(m+\ell)+1}^{m}\big(X_j - f(X_{j-1}, \theta_0 + u_k/n)\big)\, b^{m-j} =: b^{\frac12(m-\ell)}J^k_1 + J^k_2.
\end{aligned}$$

Using the bound (45), we get

$$\mathbb{E}_{\theta_0}\left|J^k_1\right|^2 \le 2\mathbb{E}_{\theta_0}|\widehat{\Xi}^k_\ell|^2 + C_1\sum_{j=\ell+1}^{\frac12(m+\ell)}\mathbb{E}_{\theta_0}\left(|X_j| + |X_{j-1}|\right)^2 |b|^{\frac12(m+\ell)-j} \le C_2. \tag{52}$$

By the triangle inequality

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m, \xi_m, \widehat{\Xi}_m)\middle|\mathcal{F}_\ell\right) - \widetilde{\mathbb{E}}_{\theta_0}h(X_0, \xi_0, \widehat{\Xi}_0)\right| \\
&\quad = \mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m, \xi_m, \widehat{\Xi}_m)\middle|\mathcal{F}_\ell\right) - \widetilde{\mathbb{E}}_{\theta_0}h(X_m, \xi_m, \widehat{\Xi}_m)\right| \\
&\quad \le \mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m, \xi_m, b^{\frac12(m-\ell)}J_1 + J_2)\middle|\mathcal{F}_\ell\right) - \mathbb{E}_{\theta_0}\left(h(X_m, \xi_m, J_2)\middle|\mathcal{F}_\ell\right)\right|
\end{aligned}$$

$$+\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m,\xi_m,J_2)\middle|\mathcal{F}_\ell\right)-\widetilde{\mathbb{E}}_{\theta_0}h(X_m,\xi_m,J_2)\right|$$
$$+\left|\widetilde{\mathbb{E}}_{\theta_0}h(X_m,\xi_m,J_2)-\widetilde{\mathbb{E}}_{\theta_0}h(X_m,\xi_m,b^{\frac{1}{2}(m-\ell)}J_1+J_2)\right|. \tag{53}$$

Note that $h(X_m,\xi_m,J_2)$ is measurable with respect to $\mathcal{F}_{\frac{1}{2}(m+\ell),\infty}$ and by (48)

$$\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m,\xi_m,J_2)\middle|\mathcal{F}_\ell\right)-\widetilde{\mathbb{E}}_{\theta_0}h(X_m,\xi_m,J_2)\right|\le Cr^{\frac{1}{2}(m-\ell)}.$$

By the Lipschitz property of h and (52), we have

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left|\mathbb{E}_{\theta_0}\left(h(X_m,\xi_m,b^{\frac{1}{2}(m-\ell)}J_1+J_2)\middle|\mathcal{F}_\ell\right)-\mathbb{E}_{\theta_0}\left(h(X_m,\xi_m,J_2)\middle|\mathcal{F}_\ell\right)\right|\\
&\le \mathbb{E}_{\theta_0}\left|h(X_m,\xi_m,b^{\frac{1}{2}(m-\ell)}J_1+J_2)-h(X_m,\xi_m,J_2)\right|\\
&\le b^{\frac{1}{2}(m-\ell)}\mathbb{E}_{\theta_0}L\left(1+|X_m|+|\xi_m|+\|\widehat{\Xi}_m\|+\|J_2\|\right)\|J_1\|\\
&\le b^{\frac{1}{2}(m-\ell)}L\left(\mathbb{E}_{\theta_0}\left(1+|X_m|+|\xi_m|+\|\widehat{\Xi}_m\|+\|J_2\|\right)^2\right)^{1/2}\left(\mathbb{E}_{\theta_0}\|J_1\|^2\right)^{1/2}\\
&\le C_3 b^{\frac{1}{2}(m-\ell)}.
\end{aligned}$$

Similar bound holds for the last term in (53) and the claim follows with $q:=\sqrt{|b|\vee r}$. □

References

Caner, M., Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, *69*(6), 1555–1596.

Cappé, O., Moulines, E., Rydén, T. (2005). *Inference in hidden Markov models*. New York: Springer Series in Statistics, Springer.

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.

Chan, N. H., Kutoyants, Y. A. (2010). Recent developments of threshold estimation for nonlinear time series. *Journal of the Japan Statistical Society*, *40*(2), 277–303.

Chan, N. H., Kutoyants, Y. A. (2012). On parameter estimation of threshold autoregressive models. *Statistical Inference for Stochastic Processes*, *15*(1), 81–104.

Dedecker, J., Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Processes and their Applications*, *106*(1), 63–80.

Doukhan, P., Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*(2), 313–342.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface*, *4*(2), 123–127.

Ibragimov, I. A., Has'minskiĭ, R. Z. (1981). Statistical estimation. Asymptotic theory. In *Applications of Mathematics* (Vol. 16). New York: Springer.

Kutoyants, Y. A. (2012). On identification of the threshold diffusion processes. *Annals of the Institute of Statistical Mathematics*, *64*(2), 383–413.

Ling, S., Tong, H. (2005). Testing for a linear MA model against threshold MA models. *The Annals of Statistics*, *33*(6), 2529–2552.

Ling, S., Tong, H., Li, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli*, *13*(1), 161–168.

Liptser, R. S., Shiryaev, A. N. (2001). Statistics of random processes. II Applications. In *Applications of Mathematics (New York)* (Vol. 6, expanded ed.). Berlin: Springer.

Liu, W., Ling, S., Shao, Q. M. (2011). On non-stationary threshold autoregressive models. *Bernoulli*, *17*(3), 969–986.

Meyer, R. M. (1973). A Poisson-type limit theorem for mixing sequences of dependent “rare” events. *Annals of Probability*, *1*, 480–483.

Meyn, S., Tweedie, R. L. (2009). *Markov chains and stochastic stability* (2nd ed.). Cambridge: Cambridge University Press.

Pham, D. T., Chan, K. S., Tong, H. (1991). Strong consistency of the least squares estimator for a nonergodic threshold autoregressive model. *Statistica Sinica*, *1*(2), 361–369.

Tong, H. (1983). Threshold models in nonlinear time series analysis. In *Lecture Notes in Statistics* (Vol. 21). New York: Springer.

Tong, H. (2011). Threshold models in time series analysis—30 years on. *Statistics and its Interface*, *4*(2), 107–118.

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