Efficient On-Line Estimation of Autoregressive Parameters

Teo Sharia

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX
e-mail: t.sharia@rhul.ac.uk

Abstract

New procedures for estimating autoregressive parameters in $AR(m)$ models are proposed. The proposed method allows for incorporation of auxiliary information into the estimation process and, under certain regularity conditions are consistent and asymptotically efficient. Also, these procedures are naturally on-line and do not require storing all the data. Theoretical results are presented in the case when $m = 1$. Two important particular cases are considered in details: linear procedures and likelihood procedures with the LS truncations. A specific example is also presented to briefly discuss some practical aspects of applications of the procedures of this type.

Keywords: Asymptotic efficiency, Least squares, Parameter estimation, Recursive likelihood procedures, Stochastic approximation

1 Introduction

In 1951, Robbins and Monro [16] created a method to estimate a root of an unknown function when the function can only be observed with random errors. The method was given a name of stochastic approximation (SA). Their work triggered a series of important advances in a broad spectrum of statistical science including systems control, optimization, and signal processing. A comprehensive survey of the related results and of some more recent developments is given in [10]. The technique of SA has been exploited by a number of authors to develop and analyze on-line estimation procedures for the classical parametric setting. The application of SA technique to the classical parametric estimation for the iid models was greatly influenced by [7] and [9] (see also [14], [15], and references therein). In [8], the authors developed a new SA algorithm for maximum likelihood estimation in the incomplete-data setting.

Considerable work has been done in non-i.i.d. models as well, see e.g., [1], [3], [4], [6], [13], [18]–[22]). Nevertheless, we feel that the full potential of the SA technique, especially in the time series context has not yet been fully exploited. This paper is another step in this direction.

In this article, we propose a new class of on-line estimating procedures for the $AR(m)$ processes. The advantage of our approach is that, it provides a simple tool to achieve an efficient use of information available in the estimation process. This information can be the exact distribution of the innovation process or/and auxiliary information about the parameters (e.g., a consistent but not necessarily efficient auxiliary estimator, or a set, possibly time dependent, that is known to contain the value of the unknown parameters). Rigorous results are presented for the proposed procedures in AR(1) models including convergence, rate of convergence, and asymptotic linearity leading to asymptotic efficiency. Although preliminary results are encouraging, working with AR(m) processes is technically much more challenging with additional issues concerning matrix-valued normalising processes.

The paper is organized as follows. In Section 2, we give background information for a reader not necessarily familiar with SA technique, and a heuristic justification of the proposed procedures. In sections 3 and 4, we present sufficient conditions for convergence, rate of convergence and asymptotic efficiency of the corresponding estimators. The proofs of all results are postponed to Appendix. In section 5, we consider two important special cases: linear procedures and likelihood procedures with the least squares truncations. In the same section we present a specific example and discuss some practical aspects of applications of the procedures of this type.

2 Stochastic approximation type estimation algorithms

Consider an AR(1) process

$$X_t = \theta X_{t-1} + \xi_t, \tag{2.1}$$

where ξ_t is a sequence of random variables (r.v.'s) with mean zero. The ordinary least squares (LS) estimator of θ can be written recursively as

$$\begin{aligned} \hat{\theta}_t^{LS} &= \hat{\theta}_{t-1}^{LS} + \hat{I}_t^{-1} X_{t-1} \left(X_t - \hat{\theta}_{t-1}^{LS} X_{t-1} \right), \\ \hat{I}_t &= \hat{I}_{t-1} + X_{t-1}^2, \qquad t = 1, 2, \ldots \end{aligned} \tag{2.2}$$

where $\hat{\theta}_0 = 0$ and $\hat{I}_0 = 0$. This can easily be verified by subtracting two successive terms of $\hat{\theta}_t^{LS} = \sum_{s=1}^t X_s X_{s-1} / \sum_{s=1}^t X_{s-1}^2$ and simple algebra (note also that $\hat{I}_t = \sum_{s=1}^t X_{s-1}^2$). It is well-known that in the case when ξ_t is a sequence of Gaussian i.i.d. r.v.'s, the LS estimators are consistent and asymptotically efficient. However, in the case of non-Gaussian ξ_t's the LS estimators fail to be efficient.

Suppose now that ξ_t is a sequence of i.i.d. r.v.'s and the probability density function of ξ_t w.r.t. Lebesgue's measure is $g(x)$. Consider an estimator defined recursively as

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \hat{I}_t^{-1} i_g^{-1} X_{t-1} \frac{g'(X_t - \hat{\theta}_{t-1} X_{t-1})}{g(X_t - \hat{\theta}_{t-1} X_{t-1})}, \qquad t = 1, 2, \dots \tag{2.3}$$

where $i_g = \int \left(g'(z)/g(z)\right)^2 g(z)\, dz, \quad t \geq 1$ and $\hat{\theta}_0 \in \mathbb{R}$ is an arbitrary starting point. The procedure (2.3) will be referred to as the recursive likelihood procedure. This name can be justified by the fact that under certain conditions, the estimators defined by (2.3) are asymptotically equivalent to MLEs in the sense that they have the same asymptotic properties as the MLE's, in particular consistency and asymptotic efficiency. A heuristic justification of the estimation procedures of this type in a more general setting will be given later in the paper.

Let us now consider a class of estimation procedures defined by

$$\hat{\theta}_t = \Big[\hat{\theta}_{t-1} + \Gamma_t^{-1} \gamma(X_{t-1}) \phi_t(X_t - \hat{\theta}_{t-1} X_{t-1})\Big]_{\alpha_t}^{\beta_t}, \qquad t = 1, 2, \dots \tag{2.4}$$

with suitably chosen functions ϕ_t and γ, and a normalising process Γ_t. Here α_t and β_t are random variables with $-\infty \leq \alpha_t \leq \beta_t \leq \infty$ and $[v]_{\alpha_t}^{\beta_t}$ is the truncation operator, that is,

$$[v]_{\alpha_t}^{\beta_t} = \begin{cases} \alpha_t & \text{if } v < \alpha_t \\ v & \text{if } \alpha_t \leq v \leq \beta_t \\ \beta_t & \text{if } v > \beta_t. \end{cases}$$

The truncation interval $U_t = [\alpha_t, \beta_t]$ represents our auxiliary knowledge about the unknown parameter which is incorporated in the procedure through the truncation operator. For example, if $\theta \in \Theta = [\alpha, \beta]$, then one can take $\alpha_t = \alpha$ and $\beta_t = \beta$. In the case of the open interval $\Theta = (\alpha, \beta)$ we may choose to consider truncations with moving bounds to avoid possible singularities at the endpoints of the interval. That is, we can take $U_t = [\alpha_t, \beta_t]$ with some sequences $\alpha_t \downarrow \alpha$ and $\beta_t \uparrow \beta$.

The most interesting case arises when a consistent, but not necessarily efficient auxiliary estimator $\tilde{\theta}_t$ is available having a rate d_t. Then one can use $\tilde{\theta}_t$ to truncate the recursive procedure in a neighbourhood of θ by taking $U_t = [\tilde{\theta}_t - \varepsilon_t, \tilde{\theta}_t + \varepsilon_t]$ with $\varepsilon_t \to 0$. Such a procedure is obviously consistent since $\hat{\theta}_t \in [\tilde{\theta}_t - \varepsilon_t, \tilde{\theta}_t + \varepsilon_t]$ and $\tilde{\theta}_t \pm \varepsilon_t \to \theta$. However, since the main goal is to construct an efficient estimator, care should be taken to ensure that the truncation intervals do not shrink to θ too rapidly, for otherwise $\hat{\theta}_t$ will have the same asymptotic properties as $\tilde{\theta}_t$ (see Remark 4.4 for details).

An example of possible applications of (2.4) is a likelihood procedure with LS truncations, that is,

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} - \hat{I}_t^{-1} i_g^{-1} X_{t-1} \frac{g'(X_t - \hat{\theta}_{t-1} X_{t-1})}{g(X_t - \hat{\theta}_{t-1} X_{t-1})}\right]_{\hat{\theta}_t^{LS} - c\hat{I}_t^{-\varepsilon}}^{\hat{\theta}_t^{LS} + c\hat{I}_t^{-\varepsilon}}, \quad t = 1, 2, \dots \tag{2.5}$$

where $\hat\theta_t^{LS}$ and $\hat I_t$ are defined by (2.2) and c and ε are positive constants.

Let us now consider a general time series model given by a sequence $X_1,\dots,X_t$ of r.v.'s with the joint distribution depending on an unknown parameter $\theta\in\mathbb{R}^m$. Recall that an M-estimator of θ is defined as a solution of the estimating equation

$$\sum_{s=1}^{t}\psi_s(v)=0, \tag{2.6}$$

where $\psi_s(v)=\psi_s(X_1^s;v)$, $s=1,2,\dots,t$, are suitably chosen functions which may, in general, depend on the vector $X_1^s=(X_1,\dots,X_s)$ of all past and present observations. If $f_s(x,\theta)=f_s(x,\theta|X_1,\dots,X_{s-1})$ is the conditional probability density function (pdf) or probability function (pf) of the observation X_s given $X_1,\dots,X_{s-1}$, then one can obtain a MLE by choosing $\psi_s(v)=f_s'(X_s,v)/f_s(X_s,v)$. Besides MLEs, the class of M-estimators includes estimators with special properties such as robustness. Under certain regularity and ergodicity conditions, there exists a consistent sequence of solutions of (2.6) which has the property of local asymptotic linearity.

If ψ-functions are nonlinear, it is rather difficult to work with the corresponding estimating equations. Note that for a linear estimator, e.g., for the sample mean $\hat\theta_t=\bar X_t$, we have $\bar X_t=(t-1)\bar X_{t-1}/t+X_t/t$, that is $\hat\theta_t=\hat\theta_{t-1}(t-1)/t+X_t/t$, which means that the estimator $\hat\theta_t$ at each step t can be obtained recursively using the estimator at the previous step $\hat\theta_{t-1}$ and the new information X_t. Such an exact recursive relation may not hold for nonlinear estimators.

In general, to find a possible form of an approximate recursive relation consider $\hat\theta_t$ defined as a root of the estimating equation (2.6). Denoting the left hand side of (2.6) by $M_t(v)$ and assuming that the difference $\hat\theta_t-\hat\theta_{t-1}$ is "small" we can write $M_t(\hat\theta_t)\approx M_t(\hat\theta_{t-1})+M_t'(\hat\theta_{t-1})(\hat\theta_t-\hat\theta_{t-1})$ and

$$0=M_t(\hat\theta_t)-M_{t-1}(\hat\theta_{t-1})\approx M_t'(\hat\theta_{t-1})(\hat\theta_t-\hat\theta_{t-1})+\psi_t(\hat\theta_{t-1}).$$

Therefore,

$$\hat\theta_t\approx\hat\theta_{t-1}-\frac{\psi_t(\hat\theta_{t-1})}{M_t'(\hat\theta_{t-1})},$$

where $M_t'(\theta)=\sum_{s=1}^t\psi_s'(\theta)$. Now, depending on the nature of the underlying model, $M_t'(\theta)$ can be replaced by a simpler expression. For instance, in the i.i.d. models with $\psi(x,v)=f'(x,v)/f(x,v)$ (the MLE case), by the strong law of large numbers,

$$\frac{M_t'(\theta)}{t}=\frac{1}{t}\sum_{s=1}^{t}\left(f'(X_s,\theta)/f(X_s,\theta)\right)'\approx E_\theta\left[\left(f'(X_1,\theta)/f(X_1,\theta)\right)'\right]=-i(\theta)$$

for large t's, where $i(\theta)$ is the one-step Fisher information. So, in this case, one can consider

$$\hat\theta_t=\hat\theta_{t-1}+\frac{1}{t\,i(\hat\theta_{t-1})}\frac{f'(X_t,\hat\theta_{t-1})}{f(X_t,\hat\theta_{t-1})},\qquad t\ge1, \tag{2.7}$$

to construct an estimator which is "asymptotically equivalent" to the MLE (see also [9] and [12]).

Motivated by the above argument, one can consider a class of estimators

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \Gamma_t^{-1}(\hat{\theta}_{t-1})\psi_t(\hat{\theta}_{t-1}), \qquad t \geq 1, \tag{2.8}$$

where ψ_t is a suitably chosen vector process, Γ_t is a normalizing matrix process, and $\hat{\theta}_0 \in \mathbb{R}^m$ is some initial value. In particular, if $\psi_s(\theta) = f_s'(X_s,\theta)/f_s(X_s,\theta)$, where $f_s(x,\theta) = f_s(x,\theta|X_1,\ldots,X_{s-1})$ is the conditional pdf/pf of the observation X_s given $X_1,\ldots,X_{s-1}$, we obtain

$$\hat{\theta}_t = \hat{\theta}_{t-1} + I_t^{-1}(\hat{\theta}_{t-1})\frac{f_t'^T(X_t,\hat{\theta}_{t-1})}{f_t(X_t,\hat{\theta}_{t-1})}, \qquad t \geq 1, \tag{2.9}$$

where, $I_t(\theta)$ is the conditional Fisher information matrix, f_t' is the row-vector of partial derivatives of f_t w.r.t. the components of θ (here T means transposition).

Now, it is easy to see that (2.3) is of the form of (2.9), since in this case, $f_t(x,\theta) = f_t(x,\theta|X_{t-1}) = g(x-\theta X_{t-1})$ and $I_t(\theta) = i_g\hat{I}_t = i_g\sum_{i=1}^t X_{s-1}^2$.

It should be noted that at first glance, recursions (2.7) and (2.9) resemble the Newton-Raphson or the one-step Newton-Raphson iterative procedures. In the i.i.d. case, the Newton-Raphson iteration for the likelihood equation is

$$\vartheta_k = \vartheta_{k-1} + J^{-1}(\vartheta_{k-1})\sum_{s=1}^t \frac{f'(X_s,\vartheta_{k-1})}{f(X_s,\vartheta_{k-1})}, \qquad k \geq 1, \tag{2.10}$$

where $-J(v)$ is the second derivative of the log-likelihood function, that is, $\sum_{s=1}^t \frac{\partial}{\partial v}\left(f'(X_s,v)/f(X_s,v)\right)$ or its expectation, that is, $-ti(v)$. In the latter case, the iterative scheme is often called the method of scoring. The main feature of the scheme (2.10) is that t is fixed, and ϑ_k, at each step $k = 1,2,\ldots$, is the k'th approximation to a root, say $\tilde{\theta}_t$, of the likelihood equation $\sum_{s=1}^t \left(f'(X_s,v)/f(X_s,v)\right) = 0$. Also, if a new $(t+1)$st observation is available, the whole procedure has to be repeated again. Note also, that the one-step Newton-Raphson is a simplified version of (2.10) when an auxiliary $\sqrt{t}$-consistent estimator, say $\tilde{\theta}_t$ is available. Then, the one-step Newton-Raphson improves $\tilde{\theta}_t$ in one step (that is, $k = 1$) by

$$\hat{\theta}_t = \tilde{\theta}_t + J^{-1}(\tilde{\theta}_t)\sum_{s=1}^t \frac{f'(X_s,\tilde{\theta}_t)}{f(X_s,\tilde{\theta}_t)}. \tag{2.11}$$

As one can see the procedure (2.7) is quite different. It does not require an auxiliary estimator and it adjusts the value of the estimator at each instant of time with the arrival of the new observation. A theoretical implication of this is that by studying the procedures (2.7), or in general (2.8), we study the asymptotic behaviour of the estimator. As far as applications are concerned, there are advantages in using (2.7), (2.8), or (2.9), since these procedures are easy to use and, unlike other methods, do not require storing all the data. Also, these procedures naturally allow for on-line implementation, which is particulary convenient for sequential data processing.

Note that the recursive procedure (2.8) is not a numerical solution of (2.6). Nevertheless, recursive estimator (2.8) and the corresponding M-estimator are expected to have the same asymptotic properties under quite mild conditions.

To understand how the procedure works, consider the likelihood recursive procedure (2.9) in the one-dimensional case. Denote $\Delta_t = \hat{\theta}_t - \theta$, rewrite the above recursion as

$$\Delta_t = \Delta_{t-1} + I_t^{-1}(\theta + \Delta_{t-1}) \frac{f_t'(X_t, \theta + \Delta_{t-1})}{f_t(X_t, \theta + \Delta_{t-1})}$$

and let

$$b_t(\theta, u) = E_\theta \left\{ \frac{f_t'(X_t, \theta + u)}{f_t(X_t, \theta + u)} \mid \mathcal{F}_{t-1} \right\},$$

where $\mathcal{F}_t$ is the σ-field generated by the random variables $X_1, \dots, X_t$. Then,

$$E_\theta \left\{ \hat{\theta}_t - \hat{\theta}_{t-1} \mid \mathcal{F}_{t-1} \right\} = E_\theta \left\{ \Delta_t - \Delta_{t-1} \mid \mathcal{F}_{t-1} \right\} = I_t^{-1}(\theta + \Delta_{t-1}) b_t(\theta, \Delta_{t-1}).$$

Under usual regularity conditions (see [20] Remark 3.2 for details), $b_t(\theta, 0) = 0$ and $\frac{\partial}{\partial u} b_t(\theta, u) \mid_{u=0} = -i_t(\theta) < 0$, implying that

$$u b_t(\theta, u) < 0 \tag{2.12}$$

for small values of $u \neq 0$. Now, assuming that (2.12) holds for all $u \neq 0$, suppose that at time $t-1$, $\hat{\theta}_{t-1} < \theta$, that is $\Delta_{t-1} < 0$. Then, by (2.12), $E_\theta \left\{ \hat{\theta}_t - \hat{\theta}_{t-1} \mid \mathcal{F}_{t-1} \right\} > 0$. So, the next step $\hat{\theta}_t$ will be in the direction of θ. If at time $t-1$, $\hat{\theta}_{t-1} > \theta$, then by the same reason, $E_\theta \left\{ \hat{\theta}_t - \hat{\theta}_{t-1} \mid \mathcal{F}_{t-1} \right\} < 0$. So, on average, at each step the procedure moves towards θ. However, the magnitude of the jumps $\hat{\theta}_t - \hat{\theta}_{t-1}$ should decrease, for otherwise, $\hat{\theta}_t$ may oscillate around θ without approaching it. On the other hand, care should be taken to ensure that the jumps do not decrease too rapidly to avoid failure of $\hat{\theta}_t$ to reach θ.

Note also that in the iid case, (2.7) can be regarded as a stochastic iterative scheme, i.e., a classical stochastic approximation procedure, to detect the root of an unknown function when the latter can only be observed with random errors. (see, e.g., [7] and [9] and references therein). Note that the idea of using auxiliary estimators in these schemes also goes back to [7] and [9] (see also [5]).

Let us now consider an AR(m) process

$$X_t = \theta^{(1)} X_{t-1} + \dots + \theta^{(m)} X_{t-m} + \xi_t = \theta^T X_{t-m}^{t-1} + \xi_t,$$

where $X_{t-m}^{t-1} = (X_{t-1}, \dots, X_{t-m})^T$, $\theta = (\theta^{(1)}, \dots, \theta^{(m)})^T$ and ξ_t is a sequence of r.v.'s with mean zero.

For a set $U \subseteq \mathbb{R}^m$, let us define a truncation operator as a function $\Phi_U : \mathbb{R}^m \longrightarrow \mathbb{R}^m$, such that

$$\Phi_U(v) = \begin{cases} v & \text{if } v \in U \\ v^* & \text{if } v \notin U. \end{cases}$$

where $v*$ is a point in the closure of U, that minimizes the distance to v ($v*$ is unique if U is convex). Assume that $\hat{\theta}_0 \in \mathbb{R}^m$ is some starting value and consider the estimator

$$\hat{\theta}_t = \Phi_{U_t}\left(\hat{\theta}_{t-1} + \Gamma_t^{-1}\gamma(X_{t-m}^{t-1})\phi_t(X_t - \hat{\theta}_{t-1}^T X_{t-m}^{t-1})\right), \tag{2.13}$$

where $\phi_t(z)$ and Γ_t^{-1} ($z \in \mathbb{R}^m$) are respectively vector and matrix processes of suitable dimensions, $\gamma(z)$ is a vector function, $U_t \subseteq \mathbb{R}^m$ is a sequence of sets such that U_t for each t may depend on $X_1, \ldots, X_t$ and $\theta \in U_t$ for large t's (a.s.).

If the pdf of ξ_t w.r.t. Lebesgue's measure is $g(x)$, the conditional pdf is $f_t(\theta, x_t \mid x_1^{t-1}) = g(x_t - \theta^T x_{t-m}^{t-1})$. In this case one may consider the likelihood procedure, that is, (2.13) with

$$\gamma(X_{t-m}^{t-1})\phi_t(x) = \frac{f_t'^T(\theta, x \mid X_1^{t-1})}{f_t(\theta, x \mid X_1^{t-1})} = -\frac{g'(x)}{g(x)} X_{t-m}^{t-1}, \tag{2.14}$$

and $\Gamma_t = I_t$, where I_t is the conditional Fisher information matrix

$$I_t = i^g \sum_{s=1}^{t} X_{s-m}^{s-1}(X_{s-m}^{s-1})^T. \tag{2.15}$$

The set U_t represents auxiliary knowledge about the unknown parameter which is incorporated in the procedure through the truncation operator Φ. For example, for an $AR(2)$ process, if the roots of the corresponding polynomial lie outside of the unit circle, one can take $U_t = U$ where U is a triangle defined by

$$U = \{\ (\theta^{(1)}, \theta^{(2)})\ :\ |\theta^{(2)}| < 1,\ \theta^{(1)} + \theta^{(2)} < 1,\ \theta^{(2)} - \theta^{(1)} < 1\ \}.$$

In the case when a consistent but not necessarily efficient auxiliary estimator $\tilde{\theta}_t$ is available, one can consider $U_t = S(\tilde{\theta}_t, \varepsilon_t)$, where S is the ball in $\mathbb{R}^m$ with the center at $\tilde{\theta}_t$ and the radius $\varepsilon_t \to 0$. For example, if $\tilde{\theta}_t$ is the LS estimator, in the case of the likelihood procedure a possible choice is $\varepsilon_t = c\|I_t\|^{-\varepsilon}$ where $\varepsilon < 1/2$ and $c > 0$.

Another example of application of (2.13) is that of robust estimation, e.g., estimation in the presence of outliers with the function ϕ_t equal to the Huber or the Hampell function. Although the details of this example are not given here, but we had it in mind when deriving Corollary 3.4 and Corollary 3.6 below.

There are three main problems arising concerning the behaviour of the estimating procedures of type (2.13): the global convergence, that is convergence for any starting point $\hat{\theta}_0$; the rate of the convergence; and the asymptotic distribution.

Note that in the case of an auxiliary consistent estimator, the procedure (2.13) is automatically globally convergent. In general, given that usual regularity conditions are satisfied (e.g., conditions similar to (2.12) with an appropriate rate of the normalising sequence), the construction of the procedure guarantees local convergence. In other words, the estimator will converge to θ, provided that the values of the procedure "stay" in a sufficiently small neighbourhood of θ. To ensure global convergence, one needs to impose conditions of the global type on the corresponding functions, e.g. conditions that guarantee a property of type (2.12) for any u, and also conditions on the growth of the corresponding functions at infinity (see [20] for details). Once the convergence is secured, the rate of convergence and the asymptotic distribution depend on the local behaviour of the corresponding functions (like differentiability of higher order) and the ergodicity of the model (see [20]–[22]). The statistical model described in [20]–[22] are quite general with no specific requirements on the dependence structure and the distribution of the underline process. The conditions in these works are given in terms of the conditional distributions. The downside of this generality is that these conditions are often difficult to check. In this paper, we give an explicit set of conditions suitable for the models under consideration. More importantly, the paper introduces a new class of on-line procedures to achieve an efficient use of information available in the estimating process.

3 Convergence

Everywhere in the present work we assume that X_t is a process defined by (2.1) where $\theta \in \mathbb{R}$ and $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$ denotes the σ-field generated by the random variables $X_0, \dots, X_t$. Conditional expectations are meant to be calculated as integrals w.r.t. the conditional probability densities. For example, if ϕ is a measurable function,

$$E_\theta \left\{ \phi(X_t) \mid \mathcal{F}_{t-1} \right\} = \int \phi(z) f_t(\theta, z \mid x_0^{t-1}) \mu(dz),$$

where $f_t(\theta, z \mid x_0^{t-1})$ is the conditional pdf of X_t given $X_0^{t-1} = x_0^{t-1}$. In particular, if ξ_t are i.i.d. with the pdf $g(z)$, we have $f_t(\theta, z \mid x_0^{t-1}) = g(z - \theta x_{t-1})$.

Without loss of generality we assume that all random variables are defined on a probability space $(\Omega, \mathcal{F})$ and denote by $\{P^\theta, \ \theta \in \Theta\}$ the family of the corresponding distributions on $(\Omega, \mathcal{F})$.

Recall that a process ζ_t is called a martingale difference, if $E_\theta(\zeta_t \mid \mathcal{F}_{t-1}) = 0$. A process ζ_t is called predictable if the random variable ζ_t is $\mathcal{F}_{t-1}$ measurable for each $t \geq 1$.

We will assume below that $\phi_t : \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ are measurable functions and the normalising sequence Γ_t is predictable, that is, the r.v. Γ_t are $\mathcal{F}_{t-1}$ measurable for each $t \geq 1$ (may only depend on the observations up to time $t-1$). We also assume that the truncation bounds α_t and β_t are adaptive, that is, the r.v.'s α_t and β_t are $\mathcal{F}_t$ measurable for each $t \geq 1$.

*Everywhere in the present work $\theta \in \mathbb{R}$ is an arbitrary but fixed value of the parameter. Convergence and all relations between random variables are meant with probability one w.r.t. the measure P^θ unless specified otherwise. A sequence of random variables $(\zeta_t)_{t\geq 1}$ has some property **eventually** if for every ω in a set Ω^θ of P^θ probability 1, ζ_t has this property for all t greater than some $t_0(\omega) < \infty$.*

We say that a truncation sequence (α_t, β_t) is ***admissible*** if $-\infty \leq \alpha_t \leq \beta_t \leq \infty$ (P^θ-a.s.) and $\theta \in [\alpha_t, \beta_t]$ eventually.

Theorem 3.1 *Suppose that the estimator $\hat{\theta}_t$ is defined by (2.4) with an admissible truncation sequence (α_t, β_t).*

(i) *Suppose that $\beta_t - \alpha_t \to 0$ P^θ-a.s. Then $\hat{\theta}_t$ is strongly consistent, that is, $\hat{\theta}_t \to \theta$ (P^θ-a.s.), for any initial value $\hat{\theta}_0$.*

(ii) *Suppose that*

(*N1*) *for $\Delta_t = \hat{\theta}_t - \theta$,*

$$\sum_{t=1}^{\infty} \frac{[\mathcal{N}_t(\Delta_{t-1})]^+}{1+\Delta_{t-1}^2} < \infty, \qquad P^\theta\text{-a.s.}$$

where

$$\begin{aligned}\mathcal{N}_t(u) &= 2u\Gamma_t^{-1}\gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - uX_{t-1}) \mid \mathcal{F}_{t-1}\right\} \\ &\quad +\Gamma_t^{-2}\gamma^2(X_{t-1})E_\theta\left\{\phi_t^2(\xi_t - uX_{t-1}) \mid \mathcal{F}_{t-1}\right\}.\end{aligned}$$

Then $|\hat{\theta}_t - \theta|$ converges (P^θ-a.s.) to a finite limit for any initial value $\hat{\theta}_0$.

(iii) *Suppose that the condition **(*N1*)** holds and*

(*N2*) *there exists a set $A \in \mathcal{F}$ with $P^\theta(A) > 0$ such that for each $\varepsilon \in (0,1)$,*

$$\sum_{t=1}^{\infty} \inf_{\substack{\varepsilon \leq |u| \leq 1/\varepsilon \\ \theta+u \in [\alpha_{t-1}, \beta_{t-1}]}} [\mathcal{N}_t(u)]^- = \infty, \qquad \text{on } A,$$

with the convention that the $\inf_{u\in U} v(u)$ of a function $v(u)$ is 1 whenever $U = \emptyset$.

Then the estimator $\hat{\theta}_t$ is strongly consistent, that is, $\hat{\theta}_t \to \theta$ (P^θ-a.s.), for any initial value $\hat{\theta}_0$.

Remark 3.2 *Condition (N1) above as well as (b4$'$) (see (3.5)) below, may look confusing since they are written in terms of the estimator we are to study. Rewriting these in a different form would unnecessarily strengthen them. For example, (N1) holds if the positive parts of functions $\mathcal{N}_t(u)$ do not increase faster than the square function as $u \to \infty$, that is, $[\mathcal{N}_t(u)]^+ \leq B_t(1+u^2)$, with some summable (w.r.t. t) coefficients B_t. Condition (b4$'$) (see (3.5)) obviously reduces to (b4) (see (3.4)) when ϕ_t functions are uniformly bounded.*

Corollary 3.3 *(Linear procedures) Suppose that* ξ_t *in* (2.1) *is a martingale-difference and*

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} + \frac{1}{\hat{\Gamma}_t D_t} X_{t-1}\left(X_t - \hat{\theta}_{t-1} X_{t-1}\right)\right]_{\alpha_t}^{\beta_t}, \tag{3.1}$$

where the truncation sequence (α_t, β_t) *is admissible,*

$$D_t = E_\theta \left\{\xi_t^2 \mid \mathcal{F}_{t-1}\right\}, \tag{3.2}$$

and

$$\hat{\Gamma}_t = \hat{\Gamma}_0 + \sum_{s=1}^{t} \frac{X_{s-1}^2}{D_s} \ \rightarrow \ \infty. \tag{3.3}$$

Then the estimator $\hat{\theta}_t$ *is strongly consistent for any initial value* $\hat{\theta}_0$ *and* $\hat{\Gamma}_0$.

Corollary 3.4 *Suppose that the estimator* $\hat{\theta}_t$ *is defined by* (2.4) *and*

(b1) *the r.v.'s* ξ_t *are independent and for each* t, ξ_t *has a bell-shaped, symmetric about zero probability density function* $g_t(z)$ *(that is,* $g_t(-z) = g_t(z)$, *and* $g_t \downarrow 0$ *on* $\mathbb{R}_+$*);*

(b2) *for each* t, $\phi_t(x)$ *is an odd function on* $\mathbb{R}$ *such that* $\phi_t(z) > 0$ *for* $z > 0$ *and* $\int_{\mathbb{R}} |\phi_t(z-w)| g_t(z) dz < \infty$ *for all* $w \in \mathbb{R}$*;*

(b3) *the r.v.'s* Γ_t *are non-negative and* $\gamma(u)$ *is a function of the form*

$$\gamma(u) = u h(u)$$

for some non-negative function h *of* u*;*

(b4) *the functions* $\phi_t(x)$ *are uniformly bounded and*

$$\sum_{t=1}^{\infty} \frac{X_{t-1}^2 h^2(X_{t-1})}{\Gamma_t^2} < \infty \tag{3.4}$$

P^θ*-a.s.*

Then the sequence $|\hat{\theta}_t - \theta|$ *converges* P^θ*-a.s. to a finite limit for any admissible truncation sequence* (α_t, β_t) *and any initial value* $\hat{\theta}_0$.

Remark 3.5 *It follows from inequality* (A.14) *in the proof of Corollary 3.4, that that condition (b4) in corollary 3.4 can be replaced by the following one.*

(b4′)

$$\sum_{t=1}^{\infty} \Gamma_t^{-2} X_{t-1}^2 h^2(X_{t-1}) \frac{E_\theta \left\{\phi_t^2(\xi_t - \Delta_{t-1} X_{t-1}) \mid \mathcal{F}_{t-1}\right\}}{1 + \Delta_{t-1}^2} < \infty \tag{3.5}$$

P^θ*-a.s.*

Corollary 3.6 *Suppose that conditions (b1)– (b3) in Corollary 3.4 and one of the conditions (b4) or (b4′) hold. Suppose also that*

(b5) *for each* $\varepsilon \in (0,1)$,

$$\sum_{t=1}^{\infty} \frac{1}{\Gamma_t} h\left(X_{t-1}\right) \inf_{\substack{\varepsilon \leq |u| \leq 1/\varepsilon \\ \theta+u \in [\alpha_{t-1}, \beta_{t-1}]}} G_t(uX_{t-1}) = \infty \qquad (3.6)$$

P^θ*-a.s., with the convention that the* $\inf_{u \in U} v(u)$ *of a function* $v(u)$ *is* 1 *whenever* $U = \emptyset$ *and*

$$G_t(w) = -w \int_{-\infty}^{\infty} \phi_t\left(z - w\right) g_t(z) dz.$$

Then the recursive estimator defined by (2.4) *is strongly consistent for any admissible truncation sequence* (α_t, β_t) *and any initial value* $\hat{\theta}_0$.

Corollary 3.7 *Suppose that* ξ_i *are i.i.d. random variables that are independent from* X_0 *and with a bell-shaped, symmetric about zero probability density function* $g(z)$ *(that is,* $g(-z) = g(z)$*, and* $g \downarrow 0$ *on* $\mathbb{R}_+$*). Suppose also that the function* g'/g *is bounded and continuous at zero. If*

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} - \hat{I}_t^{-1} K X_{t-1} \frac{g'(X_t - \hat{\theta}_{t-1}X_{t-1})}{g(X_t - \hat{\theta}_{t-1}X_{t-1})}\right]_{\alpha_t}^{\beta_t}, \qquad (3.7)$$

$$\hat{I}_t = \hat{I}_{t-1} + X_{t-1}^2,$$

then the sequence $|\hat{\theta}_t - \theta|$ *converges* P^θ*-a.s. to a finite limit for any admissible truncation sequence* (α_t, β_t)*, any constant* $K > 0$ *and any initial values* $\hat{\theta}_0$ *and* $\hat{I}_0$.

Furthermore, if X_t *is strongly stationary then the recursive estimator defined by* (3.7) *is strongly consistent for any admissible truncation sequence* (α_t, β_t)*, any constant* $K > 0$ *and any initial value* $\hat{\theta}_0$ *and* $\hat{I}_0$.

4 Asymptotic behaviour

In this section, we study the rate of convergence and asymptotic distribution of the estimators. Theorems below do not formally require consistency of $\hat{\theta}_t$. However, when interpreting the conditions, we may assume (using the results in the previous section, or otherwise) that the estimator $\hat{\theta}_t$ is consistent, that is, $\Delta_t = \hat{\theta}_t - \theta \to 0$ P^θ-a.s. or in probability P^θ.

Theorem 4.1 *Suppose that the truncation sequence in* (2.4) *is admissible and* a_t *is a non-decreasing predictable process with* $a_t \to \infty$ *as* $t \to \infty$*. Denote* $\triangle a_t = a_t - a_{t-1}$*,* $\Delta_t = \hat{\theta}_t - \theta$*, and suppose that*

(R1)

$$\lim_{t\to\infty} \frac{\triangle a_t}{a_{t-1}} = 0, \qquad P^\theta\text{-}a.s.;$$

(R2) *there exists a predictable non-negative scalar process* $\mathcal{P}_t$ *such that*

$$2\Delta_{t-1}\Gamma_t^{-1}\gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} + \mathcal{P}_t \le -\lambda_t(\theta)\Delta_{t-1}^2, \quad (4.1)$$

eventually, where $\{\lambda_t(\theta)\}$ *is a predictable process, satisfying*

$$\sum_{t=1}^{\infty}\left[\frac{\triangle a_t}{a_t} - \lambda_t(\theta)\right]^+ < \infty, \qquad P^\theta\text{-}a.s.; \quad (4.2)$$

(R3) *for each* $0 < \varepsilon < 1$,

$$\sum_{t=1}^{\infty} a_t^\varepsilon\left[\frac{1}{\Gamma_t^2}\gamma^2(X_{t-1})E_\theta\left\{\phi_t^2(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} - \mathcal{P}_t\right]^+ < \infty, \qquad P^\theta\text{-}a.s.$$

Then $\quad a_t^\delta|\hat\theta_t - \theta| \to 0 \quad (P^\theta\text{-}a.s.)$ *for any* $\delta \in]0, 1/2[$.

Corollary 4.2 *(Linear procedures) Suppose that* ξ_t *in* (2.1) *is a martingale-difference and* $\hat\theta_t$ *is defined by* (3.1), (3.2) *and* (3.3). *Suppose also that*

$$\lim_{t\to\infty}\frac{\triangle\hat\Gamma_t}{\hat\Gamma_{t-1}} = 0, \qquad P^\theta\text{-}a.s. \quad (4.3)$$

Then $\hat\Gamma_t^\delta|\hat\theta_t - \theta| \to 0$ P^θ*-a.s. for any* $\delta \in]0, 1/2[$, *any admissible truncation sequence* (α_t, β_t), *and any initial values* $\hat\theta_0$ *and* $\hat\Gamma_0$.

Corollary 4.3 *Let* X_t *be strongly stationary and* ξ_t *be i.i.d. and independent from* X_0. *Suppose that* ξ_t, *at each* t, *has a finite fourth moment and a probability density function* g.

Consider the recursive estimator defined by

$$\hat\theta_t = \left[\hat\theta_{t-1} - \hat I_t^{-1}KX_{t-1}\frac{g'(X_t - \hat\theta_{t-1}X_{t-1})}{g(X_t - \hat\theta_{t-1}X_{t-1})}\right]_{\hat\theta_t^{LS} - c\hat I_t^{-\varepsilon}}^{\hat\theta_t^{LS} + c\hat I_t^{-\varepsilon}} \quad (4.4)$$

$$\hat I_t = \hat I_{t-1} + X_{t-1}^2, \quad (4.5)$$

where $\hat\theta_t^{LS}$ *is defined by* (2.2), $\frac{1}{4} \le \varepsilon < \frac{1}{2}$, *and* $K > 0$ *and* $c > 0$ *are constants. Suppose also that*

$$\frac{d}{dw}\int_{-\infty}^{\infty}\frac{g'}{g}(z - w)\,g(z)dz\bigg|_{w=0} < -\frac{1}{2K} \quad (4.6)$$

and

$$\sup_{-\tau<w<\tau} \int_{-\infty}^{\infty} \left(\frac{g'}{g} (z-w) \right)^2 g(z)dz < \infty, \tag{4.7}$$

for some $\tau > 0$.

Then $t^\delta |\hat{\theta}_t - \theta| \to 0$ *(P^θ-a.s.) for any* $\delta \in]0, 1/2[$*, and any starring values* $\hat{\theta}_0$ *and* $\hat{I}_0$*.*

Remark 4.4 **(a)** In the remaining part of this section asymptotic linearity of recursive estimators is studied which leads to asymptotic normality and efficiency results. The key condition there will be (4.9). Note that this condition holds if the truncations in (2.4) do not occur for large t's. More precisely, (4.9) holds if the truncations in (2.4) do not occur if $t > T$, for some, possibly random T. If $\hat{\theta}_t \to \theta$, then (4.9) will obviously hold if, e.g., the admissible truncation sequence contains an open neighbourhood of θ, that is, $\theta \in (\alpha, \beta) \subset (\alpha_t, \beta_t)$ eventually, for some α and β.
(b) Let us now consider the case when $\beta_t - \alpha_t = \varepsilon_t \to 0$. Suppose e.g., that a consistent, but not necessarily efficient, auxiliary estimator $\tilde{\theta}_t$ is available. Then one can use $\tilde{\theta}_t$ to truncate the recursive procedure in a neighbourhood of θ by taking $(\alpha_t, \beta_t) = (\tilde{\theta}_t - \varepsilon_t, \tilde{\theta}_t + \varepsilon_t)$ with $\varepsilon_t \to 0$. Such a procedure is obviously consistent since $\hat{\theta}_t \in [\tilde{\theta}_t - \varepsilon_t, \tilde{\theta}_t + \varepsilon_t]$ and $\tilde{\theta}_t \pm \varepsilon_t \to \theta$. However, if ε_t vanishes too rapidly, condition (4.9) may fail to hold. Intuitively, it is quite obvious that if ε_t vanishes too rapidly, it may result in $\hat{\theta}_t$ having the same asymptotic properties as $\tilde{\theta}_t$. Note that (4.9) does not directly require admissiblilty of the truncation intervals. However, intuitively it is quite obvious that if we want the recursive procedure to be guided but not drugged by the auxiliary $\tilde{\theta}_t$, we must at least have that $\theta \in [\tilde{\theta}_t - \varepsilon_t, \tilde{\theta}_t + \varepsilon_t]$ for large t's (eventually). This happens if ε_t is a rate of $\tilde{\theta}_t$ in the sense that $\varepsilon_t^{-1} |\tilde{\theta}_t - \theta)| < 1$ eventually. In these circumstances, (4.9) will hold if the procedure generates the sequence $\hat{\theta}_t$ which converges to θ with a rate faster than ε_t.
(c) The considerations described in (b) lead to the following construction. Suppose that an auxiliary estimator $\tilde{\theta}_t$ has a rate d_t, that is d_t is a sequence of positive r.v.'s such that $d_t(\tilde{\theta}_t - \theta) \to 0 \quad P^\theta$-a.s. Let us consider (2.4) with

$$(\alpha_t, \beta_t) = \left(\tilde{\theta}_t - c(d_t^{-1} + |\Gamma_t|^{-\delta_0}),\ \tilde{\theta}_t + c(d_t^{-1} + |\Gamma_t|^{-\delta_0}) \right),$$

where c is any positive constant. Then the truncation sequence is obviously admissible since $|\tilde{\theta}_t - \theta| < c d_t^{-1}$ eventually. Now, if we can claim by Theorem 4.1 or otherwise that $|\Gamma_t|^{\delta_0} |\hat{\theta}_t - \theta| \to 0$, then (4.9) holds. Indeed, suppose that the truncations in (2.4) occur infinitely many times on a set A of positive probability. This would imply that $\hat{\theta}_t$ coincides with one of the endpoints of the truncation interval infinitely many times on A. Since $\theta \in (\tilde{\theta}_t - c d_t^{-1},\ \tilde{\theta}_t + c d_t^{-1})$ we obtain that $|\hat{\theta}_t - \theta| \geq c |\Gamma_t|^{-\delta_0}$ infinitely many times on A which contradicts our assumptions.

Another possible choice of the truncation sequence is

$$(\alpha_t, \beta_t) = \left(\tilde{\theta}_t - c\left(d_t^{-1} \bigwedge |\Gamma_t|^{-\delta_0}\right),\ \tilde{\theta}_t + c\left(d_t^{-1} \bigwedge |\Gamma_t|^{-\delta_0}\right)\right).$$

Now, if we can claim by Theorem 4.1 or otherwise that $|\Gamma_t|^{\delta_0}|\hat{\theta}_t - \theta| \to 0$, then (4.9) holds. Indeed, suppose that on a set A of positive probability the truncations in (2.4) occur infinitely many times. This would imply that $c(d_t^{-1} \bigwedge |\Gamma_t|^{-\delta_0}) = |\tilde{\theta}_t - \hat{\theta}_t|$ and

$$1 = c^{-1}(d_t \bigvee |\Gamma_t|^{\delta_0})|\tilde{\theta}_t - \hat{\theta}_t| \le c^{-1}(d_t \bigvee |\Gamma_t|^{\delta_0})|\tilde{\theta}_t - \theta| + c^{-1}(d_t \bigvee |\Gamma_t|^{\delta_0})|\hat{\theta}_t - \theta|$$

infinitely many times on A which contradicts our assumptions.

Note that if $|\Gamma_t|^{\delta_0}|\tilde{\theta}_t - \theta| \to 0$, then we can take

$$(\alpha_t, \beta_t) = (\tilde{\theta}_t - c|\Gamma_t|^{-\delta_0},\ \tilde{\theta}_t + c|\Gamma_t|^{-\delta_0}). \tag{4.8}$$

where c is any positive constant. Also, in this case, $|\Gamma_t|^{\delta_0}|\hat{\theta}_t - \theta| \le c$ eventually.

Theorem 4.5 *Suppose that the truncation sequence in* (2.4) *is admissible and*

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\Gamma_t}\gamma(X_{t-1})\phi_t(X_t - \hat{\theta}_{t-1}X_{t-1}), \quad P^\theta - \textit{eventually}. \tag{4.9}$$

Suppose also that $\Gamma_t \to \infty$ *in probability* P^θ *and for* $\Delta_t = \hat{\theta}_t - \theta$*,*

(1)

$$\lim_{t\to\infty} \frac{1}{\Gamma_t^{1/2}} \sum_{s=1}^{t} \left(\triangle\Gamma_s\Delta_{s-1} + \gamma(X_{s-1})E_\theta\left\{\phi_s(\xi_s - \Delta_{s-1}X_{s-1}) \mid \mathcal{F}_{s-1}\right\}\right) = 0$$

in probability P^θ*;*

(2)

$$\lim_{t\to\infty} \frac{1}{\Gamma_t^{1/2}} \sum_{s=1}^{t} \mathcal{E}_s(\theta) = 0$$

in probability P^θ*, where*

$$\mathcal{E}_s(\theta) = \gamma(X_{s-1})\left[\phi_s(\xi_s - \Delta_{s-1}X_{s-1}) - \phi_s(\xi_s) - E_\theta\left\{\phi_s(\xi_s - \Delta_{s-1}X_{s-1}) \mid \mathcal{F}_{s-1}\right\}\right].$$

Then

$$\hat{\theta}_t - \theta = \frac{1}{\Gamma_t}\sum_{s=1}^{t} \gamma(X_{s-1})\phi_s(\xi_s) + \delta_t^\theta,$$

where $\Gamma_t^{1/2}\delta_t^\theta \to 0$ *in probability* P^θ*.*

Proposition 4.6 **(a)** *Suppose that $\Gamma_t(\theta)$ in Theorem 4.5 is non-decreasing (w.r.t. t) and*

(L1)

$$\lim_{t\to\infty} \frac{1}{\Gamma_t} \sum_{s=1}^{t} \Gamma_s^{1/2} \left(\triangle\Gamma_s \Delta_{s-1} + \gamma(X_{s-1}) E_\theta \left\{\phi_s(\xi_s - \Delta_{s-1}X_{s-1}) \mid \mathcal{F}_{s-1}\right\}\right) = 0$$

in probability P^θ.

Then (1) in Theorem 4.5 holds.

(b) *Suppose that $\phi_t(\xi_t)$ in Theorem 4.5 is a martingale difference and*

(E) *there exists a non-random sequence G_t such that*

$$\frac{\Gamma_t}{G_t} \to \eta(\theta)$$

weakly w.r.t. P^θ, where $\eta(\theta)$ is a random variable with $\eta(\theta) \neq 0$ P^θ-a.s.;

(L2)

$$\lim_{t\to\infty} \frac{1}{\Gamma_t} \sum_{s=1}^{t} E_\theta \left\{\mathcal{E}_s^2(\theta) \mid \mathcal{F}_{s-1}\right\} = 0$$

in probability P^θ.

Then (2) in Theorem 4.5 holds.

(c) *Suppose that $\phi_t(\xi_t)$ in Theorem 4.5 is a martingale difference and $\Gamma_t(\theta)$ is non-decreasing. Suppose also that*

(LL2)

$$\sum_{s=1}^{\infty} \frac{E_\theta \left\{\mathcal{E}_s^2(\theta) \mid \mathcal{F}_{s-1}\right\}}{\Gamma_s(\theta)} < \infty \qquad P^\theta\text{-a.s.}.$$

Then (2) in Theorem 4.5 holds.

5 Examples

5.1 Linear procedures

Suppose that X_t is an AR(1) process defined by (2.1) where ξ_t is a martingale-difference, that is, $E_\theta\{\xi_t \mid \mathcal{F}_{t-1}\} = 0$. Suppose also that $D_t = E_\theta\{\xi_t^2 \mid \mathcal{F}_{t-1}\} > 0$ and consider the recursive estimator

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\hat{\Gamma}_t D_t} X_{t-1}\left(X_t - \hat{\theta}_{t-1}X_{t-1}\right) \tag{5.1}$$

where

$$\hat{\Gamma}_t = \hat{\Gamma}_{t-1} + \frac{X_{t-1}^2}{D_t}. \tag{5.2}$$

Corollary 5.1 *Suppose that*

$$\hat{\Gamma}_t = \hat{\Gamma}_0 + \sum_{s=1}^{t} \frac{X_{s-1}^2}{D_s} \;\to\; \infty, \qquad P^\theta\text{-}a.s.$$

Then the estimator $\hat{\theta}_t$ defined by (5.1) *is strongly consistent, that is, $\hat{\theta}_t \to \theta$ P^θ-a.s. for any initial values $\hat{\theta}_0$ and $\hat{\Gamma}_0$.*

Furthermore, if

$$\lim_{t\to\infty} \frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_{t-1}} = 0, \qquad P^\theta\text{-}a.s. \tag{5.3}$$

then $\hat{\Gamma}_t^\delta|\hat{\theta}_t - \theta| \to 0$ P^θ- a.s. for any $\delta \in]0, 1/2[$ and for any initial values of $\hat{\theta}_0$ and $\hat{\Gamma}_0$.

Remark 5.2 In the case of the i.i.d. innovations ξ_t we have $D_t = \sigma^2 = \mathrm{Var}(\xi_t)$ and so, $\hat{\Gamma}_t = \hat{\Gamma}_0 + \sum_{s=1}^t X_{s-1}^2/\sigma^2 \to \infty$ for any θ. This implies that $\hat{\theta}_t$, which reduces to (2.2) if $\hat{\Gamma}_0 = 0$ and $\hat{\theta}_0 = 0$, is strongly consistent for any value of the parameter θ. Also, it is easy to see that (5.3) holds if e.g., the limit $\hat{\Gamma}_t/t$ exists (a.s.) and is finite. For example, in the case of the i.i.d. innovations ξ_t, this will happen if $|\theta| \le 1$ implying that $t^\delta|\hat{\theta}_t - \theta| \to 0$ a.s. for any $\delta \in]0, 1/2[$ (see also Remark 5.5 below).

5.2 Likelihood procedures

Corollary 5.3 *Let X_t be strongly stationary and ξ_t be i.i.d. and independent from X_0. Suppose that ξ_t have a finite fourth moment and a common probability density function g. Consider the recursive estimator defined by*

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} - \hat{I}_t^{-1} i_g^{-1} X_{t-1} \frac{g'(X_t - \hat{\theta}_{t-1}X_{t-1})}{g(X_t - \hat{\theta}_{t-1}X_{t-1})}\right]_{\hat{\theta}_t^{LS} - c\hat{I}_t^{-\varepsilon}}^{\hat{\theta}_t^{LS} + c\hat{I}_t^{-\varepsilon}} \tag{5.4}$$

$$\hat{\theta}_t^{LS} = \hat{\theta}_{t-1}^{LS} + \hat{I}_t^{-1} X_{t-1} \left(X_t - \hat{\theta}_{t-1}^{LS} X_{t-1} \right), \tag{5.5}$$
$$\hat{I}_t = \hat{I}_{t-1} + X_{t-1}^2,$$

where $1/4 \leq \varepsilon < 1/2$ *and*

$$0 < i_g = \frac{d}{dw} \int_{-\infty}^{\infty} \frac{g'}{g} (z - w)\, g(z) dz \Big|_{w=0} < \infty. \tag{5.6}$$

Suppose also that for some $\varepsilon_0 > 0$

$$\int_{-\infty}^{\infty} \frac{g'}{g} (z - w)\, g(z) dz = -i_g w + w^{1+\varepsilon_0} O(1) \tag{5.7}$$

as $w \to 0$,

$$\int_{-\infty}^{\infty} \left[\frac{g'}{g}(z) \right]^2 g(z) < \infty, \tag{5.8}$$

and

$$\int_{-\infty}^{\infty} \left[\frac{g'}{g}(z - w) - \frac{g'}{g}(z) \right]^2 g(z) dz \to 0 \tag{5.9}$$

as $w \to 0$.

Then $t^\delta |\hat{\theta}_t - \theta| \to 0$ P^θ*-a.s. for any* $\delta \in]0, 1/2[$ *and any starring values* $\hat{\theta}_0$ *and* $\hat{I}_0$*. Furthermore,* $\hat{\theta}_t$ *is asymptotically efficient in the sense that*

$$\mathcal{L}(I_t^{1/2}(\hat{\theta}_t - \theta)) \xrightarrow{d} \mathcal{N}(0, \ 1), \tag{5.10}$$

where $I_t = i_g \sum_{s=1}^t X_{t-1}^2$*, and also,*

$$\mathcal{L}(t^{1/2}(\hat{\theta}_t - \theta)) \xrightarrow{d} \mathcal{N}\left(0, \ \frac{(1-\theta^2)}{\sigma^2 i_g}\right) \tag{5.11}$$

where $\sigma^2 = var(\xi_t)$.

Remark 5.4 Note that under usual regularity assumptions,

$$i_g = \frac{d}{dw} \int_{-\infty}^{\infty} \frac{g'}{g} (z - w)\, g(z) dz \Big|_{w=0} = \int_{-\infty}^{\infty} \frac{g'}{g}(z) \, \frac{d}{dw} g(z + w) \Big|_{w=0} dz$$
$$= \int (g'(z)/g(z))^2 g(z) \, dz,$$

implying that $i_t = i_g X_{t-1}^2$ is the one step conditional Fisher information (see (2.15) with $m = 1$) and $I_t = i_g \sum_{s=1}^t X_{t-1}^2$ is the conditional Fisher information. So, (5.10) reflects the fact that $(\hat{\theta}_t - \theta)$ is asymptotically normal with asymptotic variance I_t^{-1}, where I_t is the conditional Fisher information.

Remark 5.5 Let us consider the case when ξ_t are i.i.d. Gaussian r.v.'s with zero mean σ^2 variance. Then, since $g'/g(z) = -z/\sigma^2$,

$$i^g = \int (g'(z)/g(z))^2 g(z)\, dz = \frac{1}{\sigma^2}.$$

It is easy to see that if $\hat{\theta}_0 = \hat{\theta}_0^{LS}$, then (5.4) reduces to (5.5) and all the conditions of the above corollary hold. Hence, it follows from (5.11) that $\mathcal{L}(t^{1/2}(\hat{\theta}_t - \theta)) \xrightarrow{d} \mathcal{N}\left(0,\ \ (1-\theta^2)\right)$.

5.3 An explicit example - AR(1) with Student innovations

Suppose that X_t is a strictly stationary and ξ_t are independent Student random variables with degrees of freedom α. So, the probability density functions of ξ_t is

$$g(z) = C_\alpha \left(1 + \frac{z^2}{\alpha}\right)^{-\frac{\alpha+1}{2}},$$

where $C_\alpha = \mathbf{\Gamma}((\alpha+1)/2)/(\sqrt{\pi\alpha}\ \mathbf{\Gamma}(\alpha/2))$. Since

$$\frac{g'(z)}{g(z)} = -(\alpha+1)\frac{z}{\alpha+z^2},$$

we have

$$\begin{aligned} i^g = \int \left(\frac{g'(z)}{g(z)}\right)^2 g(z)\, dz \qquad &= C_\alpha(\alpha+1)^2 \int \frac{z^2\, dz}{(\alpha+z^2)^2(1+\frac{z^2}{\alpha})^{\frac{\alpha+1}{2}}} \\ &= C_\alpha \frac{(\alpha+1)^2}{\sqrt{\alpha}} \int \frac{z^2\, dz}{(1+z^2)^{\frac{\alpha+5}{2}}} \\ &= C_\alpha \frac{(\alpha+1)^2}{\sqrt{\alpha}} \frac{\sqrt{\pi}\mathbf{\Gamma}((\alpha+5)/2 - 3/2)}{2\mathbf{\Gamma}((\alpha+5)/2)} \\ &= \frac{\alpha+1}{\alpha+3}. \end{aligned}$$

Consider a likelihood recursive procedure with $-\infty \le \alpha_t \le \beta_t \le \infty$:

$$\hat{\theta}_t = \left[\hat{\theta}_{t-1} + \hat{I}_t^{-1} i_g^{-1}(\alpha+1)X_{t-1}\frac{X_t - \hat{\theta}_{t-1}X_{t-1}}{\alpha + (X_t - \hat{\theta}_{t-1}X_{t-1})^2}\right]_{\alpha_t}^{\beta_t}, \qquad t \ge 1, \tag{5.12}$$

where

$$\hat{I}_t = \hat{I}_{t-1} + X_{t-1}^2$$

and $\hat{\theta}_0$ is any starting point. If $\alpha \ge 3$, $\hat{\theta}_t$ is strongly consistent provided that $\theta \in (\alpha_t, \beta_t)$ for large t's, in particular when $\alpha_t = -\infty$ and $\beta_t = \infty$.

Now consider (5.12) with the LS truncations, that is, when

$$\alpha_t = \hat{\theta}_t^{LS} - c\hat{I}_t^{-\varepsilon} \quad \text{and} \quad \beta_t = \hat{\theta}_t^{LS} + c\hat{I}_t^{-\varepsilon}. \tag{5.13}$$

It is not difficult to see that if $1/4 \le \varepsilon < 1/2$ and $\alpha \ge 5$, all the conditions of the proposition in the previous example are satisfied (all the improper integrals involved are uniformly convergent and all the corresponding functions are infinitely many times differentiable). Thus, if $\alpha \ge 5$, the recursive estimator defined by (5.12) and (5.13) with $1/4 \le \varepsilon < 1/2$, is strongly consistent with $t^\delta|\hat{\theta}_t - \theta| \to 0$.

Furthermore, $\hat{\theta}_t$ is asymptotically efficient, i.e.

$$\mathcal{L}\left(I_t^{1/2}(\hat{\theta}_t - \theta)\right) \xrightarrow{d} \mathcal{N}(0, \ \ 1),$$

where $I_t = (\alpha+1)(\alpha+3) \ \sum_{s=1}^t X_{t-1}^2$ is the conditional Fisher Information. Also, since $(\sigma^2 i_g)^{-1} = (\alpha+3)(\alpha-2)/((\alpha+1)\alpha) = 1 - 6/\alpha + 6/(1+\alpha)$,

$$\mathcal{L}\left(t^{1/2}(\hat{\theta}_t - \theta)\right) \xrightarrow{d} \mathcal{N}\left(0, \ \ (1-\theta^2)\left(1 - \frac{6}{\alpha} + \frac{6}{1+\alpha}\right)\right).$$

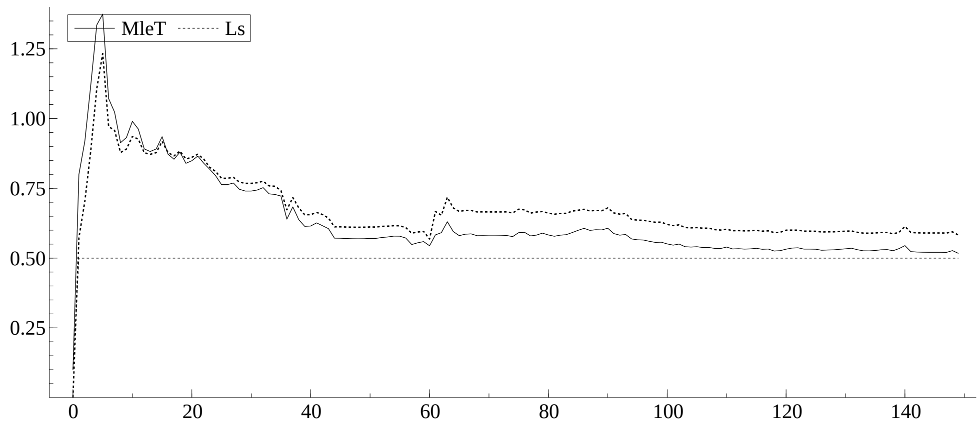

Figure 1: Realisations of $\hat{\theta}_t$ (MleT) and $\hat{\theta}_t^{LS}$ (Ls)

As far as the practical implementation of this procedure is concerned, it is important to note that the asymptotic behaviour of $\hat{\theta}_t$ will not change (including the rate of convergence), if we replace $\hat{I}_t$ in (5.12) (or, in general, in (3.7)) by $c_t\hat{I}_t$, where $c_t > 0$ are constants with $c_t = 1$ for large t's. In practice, c_t can be treated as tuning constants to control behaviour of the normalising sequence for the first several steps, especially when the number of observations is small or even moderately large. As it was mentioned above, at each step, the recursive procedure (5.12) (or, in general

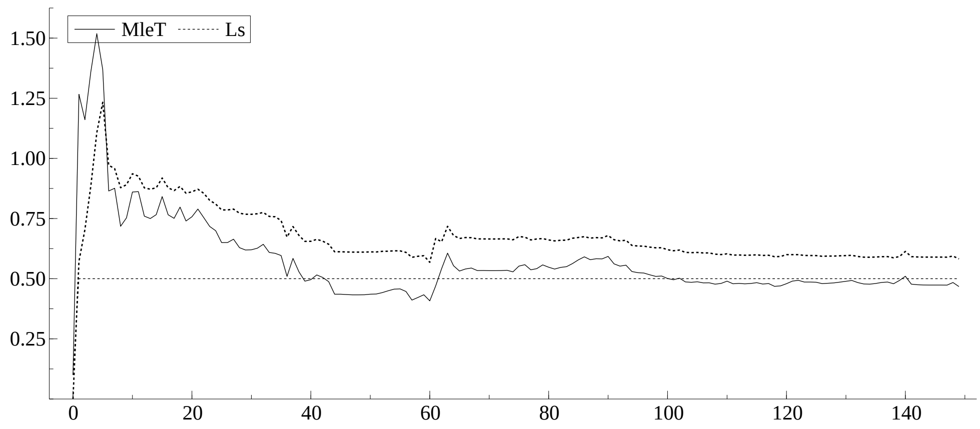

Figure 2: Realisations of $\hat{\theta}_t$ (MleT) and $\hat{\theta}_t^{LS}$ (Ls)

(3.7)) on average moves towards the parameter. Nevertheless, if the values of the normalizing sequence are too small for the first several steps, then the procedure will oscillate excessively around the true value of the parameter. On the other hand, too large values of the normalizing sequence will result in delay of the estimator reaching the value of the parameter. A good balance can be achieved by using the tuning constants.

Figures 1 and 2 show realisations of the estimators $\hat{\theta}_t$ and $\hat{\theta}_t^{LS}$ for $t = 0, \ldots, 150$, when the observations are from AR(1) process with the iid Student innovations with $\alpha = 5$, $\theta = 0.5$, $\hat{\theta}_0 = \hat{\theta}_0^{LS} = 0.1$ and $\hat{I}_0 = 0$. $\hat{\theta}_t$ is derived from (5.12) with (5.13) truncations where $\varepsilon = 1/4$ and $c = 1$. As we can see from Figure 1, $\hat{\theta}_t$ is moving downwards slowly. This may be due to the high values of the normalising sequence at the beginning of the procedure. Figure 2 shows the values of $\hat{\theta}_t$ for the same realisation but the normalising sequence $\hat{I}_t$ is replaced by $c_t\hat{I}_t$, where $c_t = 0.6$ for $t = 1, \ldots, 15$ and $c_t = 1$ otherwise. The path of this estimator has a better shape, that is a reasonable oscillation at the beginning of the procedure before settling down at a particular level. On other occasions, it may be desirable to increase the values of the normalising sequence for the first several steps. This happens when the procedure oscillates too excessively before settling down at a particular level. This can be dealt with by introducing a positive constant $\hat{I}_0 \neq 0$ and/or setting the values of c_t greater than one for the first several steps.

A Proofs

Everywhere below (see the convention in Section 3), convergence and all relations between random variables are meant with probability one w.r.t. the measure P^θ

unless specified otherwise.

Lemma A.1 *Suppose that for an estimator* $\hat{\theta}_t$ *we have*

$$|\Delta_t| \leq |\Delta_{t-1} + \Gamma_t^{-1}\gamma(X_{t-1})\phi_t(\xi_t - \Delta_{t-1}X_{t-1})| \tag{A.1}$$

eventually, where $\Delta_t = \hat{\theta}_t - \theta$. *Suppose also that* C_t *is a predictable non-negative process and*

$$\sum_{t=1}^{\infty} \left(1 + C_{t-1}\Delta_{t-1}^2\right)^{-1} [\mathcal{K}_t]^+ < \infty, \qquad P^\theta\text{-a.s.}. \tag{A.2}$$

where

$$\begin{aligned} \mathcal{K}_t &= \triangle C_t \; \Delta_{t-1}^2 + 2C_t\Delta_{t-1}\Gamma_t^{-1}\gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} \\ & \quad + C_t\Gamma_t^{-2}\gamma^2(X_{t-1})E_\theta\left\{\phi_t^2(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\}, \end{aligned} \tag{A.3}$$

with $\triangle C_t = C_t - C_{t-1}$.

Then $C_t(\hat{\theta}_t - \theta)^2$ *converges (*P^θ*-a.s.) to a finite limit for any initial value* $\hat{\theta}_0$ *and*

$$\sum_{s=1}^{\infty} [\mathcal{K}_s]^- < \infty, \qquad (P^\theta\text{-a.s.}) \tag{A.4}$$

Proof Squaring the both sides of (A.1), using $X_t - \theta X_{t-1} = \xi_t$, and taking the conditional expectation yields

$$\begin{aligned} E_\theta\left\{\Delta_t^2 \mid \mathcal{F}_{t-1}\right\} &\leq \Delta_{t-1}^2 + 2\Delta_{t-1}\Gamma_t^{-1}\gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} \\ & \quad + \Gamma_t^{-2}\gamma^2(X_{t-1})E_\theta\left\{\phi_t^2(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\}, \end{aligned} \tag{A.5}$$

Multiplying both sides on C_t and using $C_t = C_{t-1} + \triangle C_t$ we obtain

$$E_\theta\left\{C_t\Delta_t^2 \mid \mathcal{F}_{t-1}\right\} \leq C_{t-1}\Delta_{t-1}^2 + \mathcal{K}_t$$

eventually. Now, using $\mathcal{K}_t = [\mathcal{K}_t]^+ - [\mathcal{K}_t]^-$, the previous inequality can be rewritten as

$$E_\theta\left\{C_t\Delta_t^2 \mid \mathcal{F}_{t-1}\right\} \leq C_{t-1}\Delta_{t-1}^2(1 + B_t) + B_t - [\mathcal{K}_t]^- \tag{A.6}$$

eventually, where

$$B_t = \left(1 + C_{t-1}\Delta_{t-1}^2\right)^{-1} [\mathcal{K}_t]^+.$$

By (A.2),

$$\sum_{t=1}^{\infty} B_t < \infty. \tag{A.7}$$

According to Lemma B.1 in Appendix B (with $X_t = C_t\Delta_t^2$, $\beta_{t-1} = \xi_{t-1} = B_t$ and $\zeta_{t-1} = [\mathcal{K}_t]^-$), inequalities (A.6) and (A.7) imply that (A.4) holds and $C_t\Delta_t^2$ converges to some finite limit. $\diamondsuit$

Proof of Theorem 3.1
(i) The proof is trivial. Indeed, since the sequence (α_t, β_t) is admissible for θ it follows that $\theta \in [\alpha_t, \beta_t]$ eventually. So $|\hat\theta_t - \theta| \le \beta_t - \alpha_t \to 0$.

(ii) Let us show that the conditions of Lemma A.1 are satisfied with $C_t = 1$. Since $\theta \in [\alpha_t, \beta_t]$ eventually, we have that

$$|\hat\theta_t - \theta| \le |\hat\theta_{t-1} - \theta + \Gamma_t^{-1}\gamma(X_{t-1})\phi_t(X_t - \hat\theta_{t-1}X_{t-1})| \tag{A.8}$$

eventually. So, (A.1) holds. Also, since $\triangle C_t = 0$, we have $K_t = \mathcal{N}_t(\Delta_{t-1})$ and so, ***(N1)*** implies (A.2). Hence Δ_t^2 converges to some finite limit and also

$$\sum_{s=1}^{\infty}[\mathcal{N}_s(\Delta_{s-1})]^- < \infty. \tag{A.9}$$

(iii) Since ***(N1)*** is satisfied it follows that that $\Delta_t^2 \to r \ge 0$ and (A.9) holds. Suppose that there exists a set A with $P^\theta(A) > 0$, such that $r > 0$ on A. Then there exists $\varepsilon > 0$ and (possibly random) t_0, such that if $t \ge t_0$, $\varepsilon \le |\Delta_{t-1}| \le 1/\varepsilon$ on A. Note also that $\theta + \Delta_{t-1} = \hat\theta_{t-1} \in [\alpha_{t-1}, \beta_{t-1}]$. By ***(N2)***, these would imply that

$$\sum_{s=t_0}^{\infty}[\mathcal{N}_s(\Delta_{s-1})]^- \ge \sum_{s=t_0}^{\infty} \inf_{\substack{\varepsilon\le|u|\le 1/\varepsilon \\ \theta+u\in[\alpha_{s-1},\beta_{s-1}]}} [\mathcal{N}_s(u)]^- = \infty$$

on A, which contradicts (A.9). Hence, $\Delta_t^2 \to r = 0$ (a.s.). $\diamondsuit$

Proof of Corollary 3.3 By Theorem 3.1 (iii), it is sufficient to show that ***(N1)*** and ***(N2)*** hold with $\phi_t(u) = u/D_t$, $\gamma(u) = u$ and $\Gamma_t = \hat\Gamma_t$. Since ξ_t is a martingale-difference,

$$E_\theta\left\{\xi_t - uX_{t-1} \mid \mathcal{F}_{t-1}\right\} = -uX_{t-1} \tag{A.10}$$

and

$$E_\theta\left\{(\xi_t - uX_{t-1})^2 \mid \mathcal{F}_{t-1}\right\} = D_t + u^2X_{t-1}^2. \tag{A.11}$$

So,

$$\mathcal{N}_t(u) = -\frac{1}{\hat\Gamma_t D_t}2u^2X_{t-1}^2 + \frac{1}{\hat\Gamma_t^2 D_t}X_{t-1}^2 + \frac{1}{\hat\Gamma_t^2 D_t^2}X_{t-1}^4u^2.$$

Also, since $\Delta\hat\Gamma_t = \hat\Gamma_t - \hat\Gamma_{t-1} = X_{t-1}^2/D_t$, we obtain

$$\begin{aligned}\mathcal{N}_t(u) \quad &= -\frac{\Delta\hat\Gamma_t}{\hat\Gamma_t}2u^2 + \frac{\Delta\hat\Gamma_t}{\hat\Gamma_t^2} + \frac{\Delta\hat\Gamma_t^2}{\hat\Gamma_t^2}u^2, \\ &= -\frac{\Delta\hat\Gamma_t}{\hat\Gamma_t}\delta u^2 - \frac{\Delta\hat\Gamma_t}{\hat\Gamma_t}u^2\left((2-\delta) - \frac{\Delta\hat\Gamma_t}{\hat\Gamma_t}\right) + \frac{\Delta\hat\Gamma_t}{\hat\Gamma_t^2} \end{aligned} \tag{A.12}$$

for some $0 < \delta < 1$. Note that $\Delta\hat{\Gamma}_t/\hat{\Gamma}_t \leq 1$, implying that the middle term in (A.12) is non-positive. Also, $\hat{\Gamma}_t$ is non-decreasing with $\hat{\Gamma}_t \to \infty$. So, using Proposition B.3 (b), we obtain

$$\sum_{t=1}^{\infty} \frac{[\mathcal{N}_t(\Delta_{t-1})]^+}{(1+\Delta_{t-1}^2)} \leq \sum_{t=1}^{\infty} [\mathcal{N}_t(\Delta_{t-1})]^+ \leq \sum_{t=1}^{\infty} \frac{\Delta\hat{\Gamma}_t}{\hat{\Gamma}_t^2} < \infty$$

which implies ***(N1)***. Using the obvious inequalities $[a]^- \geq -a$, the fact that the middle term in (A.12) is non-positive, and Proposition B.3, we obtain

$$\sum_{t=1}^{\infty} \inf_{\substack{\varepsilon \leq |u| \leq 1/\varepsilon \\ \theta+u \in [\alpha_{t-1},\beta_{t-1}]}} [\mathcal{N}_t(u)]^- \geq \sum_{t=1}^{\infty} \min\left(1, \inf_{\varepsilon \leq |u| \leq 1/\varepsilon} \left(\frac{\Delta\hat{\Gamma}_t}{\hat{\Gamma}_t}\delta u^2 - \frac{\Delta\hat{\Gamma}_t}{\hat{\Gamma}_t^2}\right)\right) = \infty,$$

which implies ***(N2)***. $\diamondsuit$

Proof of Corollary 3.4 By Theorem 3.1(ii) it sufficient to show that ***(N1)*** holds. It follows from (b1), (b2) and Lemma B.4 that if $w \neq 0$,

$$G_t(w) = -w \int_{-\infty}^{\infty} \phi_t\,(z-w)\,g_t(z)dz > 0.$$

Therefore,

$$\begin{aligned} u\gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - uX_{t-1}) \mid \mathcal{F}_{t-1}\right\} &= uX_{t-1}h(X_{t-1})E_\theta\left\{\phi_t(\xi_t - uX_{t-1}) \mid \mathcal{F}_{t-1}\right\} \\ &= -h\left(X_{t-m}^{t-1}\right)G_t(uX_{t-1}) \leq 0 \qquad \text{(A.13)} \end{aligned}$$

So,

$$\frac{[\mathcal{N}_t(\Delta_{t-1})]^+}{1+\Delta_{t-1}^2} \leq \Gamma_t^{-2}X_{t-1}^2h^2(X_{t-1})\frac{E_\theta\left\{\phi_t^2(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\}}{1+\Delta_{t-1}^2}. \qquad \text{(A.14)}$$

Also, since ϕ_t functions are uniformly bounded,

$$\frac{[\mathcal{N}_t(\Delta_{t-1})]^+}{1+\Delta_{t-1}^2} \leq C^\theta \frac{X_{t-1}^2h^2(X_{t-1})}{\Gamma_t^2}.$$

for some positive constant C^θ. Now, **(b4)** implies ***(N1)*** of Theorem 3.1. $\diamondsuit$

Proof of Corollary 3.6 By Theorem 3.1(iii), it is sufficient to show that ***(N2)*** holds. Denote the second term in $\mathcal{N}_t(u)$ by η_t. Then using the obvious inequality $[a]^- \geq -a$ we obtain

$$\begin{aligned} [\mathcal{N}_t(u)]^- &\geq -u\Gamma_t^{-1}\gamma(X_{t-1})E_\theta\left\{\phi_t\,(\xi_t - uX_{t-1}) \mid \mathcal{F}_{t-1}\right\} - \eta_t \\ &= -u\Gamma_t^{-1}X_{t-1}h(X_{t-1})\int_{-\infty}^{\infty}\phi_t\,(z - uX_{t-1})\,g(z)dz - \eta_t \\ &= \Gamma_t^{-1}h(X_{t-1})G_t(uX_{t-1}) - \eta_t. \end{aligned}$$

Since η_t coincides with the r.h.s. of (A.14), by (b4), we have $\sum_{t=1}^{\infty} \eta_t < \infty$. Condition ***(N2)*** now obviously follows from (b5). $\diamondsuit$

Proof of Corollary 3.7 It follows from the conditions of the corollary that $-g'/g$ is an odd function on $\mathbb{R}$ such that $\phi(z) > 0$ for $z > 0$. So, it easy to see that the conditions (b1)–(b3) of Corollary 3.4 are satisfied with

$$\phi_t = -\frac{g'}{g}, \quad h(u) = 1 \quad \text{and} \quad \Gamma_t = \hat{I}_t/K = (\hat{I}_0 + \sum_{s=1}^{t} X_{t-1}^2)/K.$$

Since $\hat{I}_t \to \infty$ for any $\theta \in \mathbb{R}$ (see, e.g, Shiryayev [23], Ch.VII, §5) it follows that Γ_t is also non-decreasing with $\Gamma_t \to \infty$. Also, since $\Delta\Gamma_t = X_{t-1}^2/K$, Proposition B.3 (b) immediately implies (b4) of Corollary 3.4 which proves the first part.

Suppose now that the process is strongly stationary. By Corollary 3.6, it is sufficient to prove that for each $\varepsilon \in (0,1)$,

$$\sum_{t=1}^{\infty} \frac{1}{\hat{I}_t} \inf_{\substack{\varepsilon \le |u| \le 1/\varepsilon \\ \theta+u \in [\alpha_{t-1}, \beta_{t-1}]}} G(uX_{t-1}) = \infty \tag{A.15}$$

P^θ-a.s., with the convention that the $\inf_{u\in U} v(u)$ of a function $v(u)$ is 1 whenever $U = \emptyset$ and

$$G(w) = -w \int_{-\infty}^{\infty} \frac{g'}{g}(z-w)\, g(z)dz.$$

By Lemma B.4, $\inf_{\varepsilon \le |u| \le 1/\varepsilon} G(ux) > 0$ for any $x \neq 0$. Now, it is easy to see that

$$\inf_{\substack{\varepsilon \le |u| \le 1/\varepsilon \\ \theta+u \in [\alpha_{t-1}, \beta_{t-1}]}} G(ux) \ge \min\left(\inf_{\varepsilon \le |u| \le 1/\varepsilon} G(ux),\ 1 \right) > 0$$

for any $x \neq 0$. Since the process is strongly stationary, it follows from the ergodic theorem that in probability P^θ,

$$\lim_{t\to\infty} \frac{1}{t}\hat{I}_t \quad > 0 \qquad \text{and} \qquad \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \min\left(\inf_{\varepsilon \le |u| \le 1/\varepsilon} G(uX_{s-1}),\ 1 \right) \quad > 0.$$

These imply that (see e.g., Proposition 4 in [20]) that

$$\sum_{t=1}^{\infty} \frac{\min\left(\inf_{\varepsilon \le |u| \le 1/\varepsilon} G(uX_{s-1}),\ 1\right)}{\hat{I}_t} = \infty$$

P^θ-a.s., which proves (A.15). $\diamondsuit$

Proof of Theorem 4.1 We have to show that the conditions of Lemma A.1 are satisfied with $C_t = a_t^{2\delta}, \quad \delta \in]0, 1/2[$. Since $\hat{\theta}_t$ satisfies (A.8) it follows that (A.1) holds. The proof that (R1)– (R3) imply (A.2) is exactly same as the corresponding proof in [21], Corollary 3.1 (with $C_\theta = 1$ and $V_t(u) = C_t u^2$). $\diamondsuit$

Proof of Corollary 4.2 Let us show that the conditions of Theorem 4.1 hold with $a_t = \Gamma_t = \hat{\Gamma}_t$, $\phi_t(u) = u/D_t$ and $\gamma(u) = u$. Condition (R1) immediately follows from (5.3). Let us consider (R2) with

$$\mathcal{P}_t = \hat{\Gamma}_t^{-2} D_t^{-2} \Delta_{t-1}^2 X_{t-1}^4.$$

Then using (A.10), the left hand side of (4.1) can be rewritten as

$$-2\Delta_{t-1}^2 \hat{\Gamma}_t^{-1} D_t^{-1} X_{t-1}^2 + \hat{\Gamma}_t^{-2} D_t^{-2} \Delta_{t-1}^2 X_{t-1}^4 = -\Delta_{t-1}^2 \lambda_t(\theta),$$

where

$$\lambda_t(\theta) = 2\frac{X_{t-1}^2}{\hat{\Gamma}_t D_t} - \frac{X_{t-1}^4}{\hat{\Gamma}_t^2 D_t^2}.$$

So, since $\triangle\hat{\Gamma}_t = X_{t-1}^2/D_t$, we obtain

$$\frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t} - \lambda_t(\theta) = \frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t} - 2\frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t} + \frac{\triangle\hat{\Gamma}_t^2}{\hat{\Gamma}_t^2} = -\frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t}\left(1 - \frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t}\right) < 0$$

which implies (R2). To check (R3) let us use (A.11) and write

$$\begin{aligned}\frac{1}{\hat{\Gamma}_t^2} X_{t-1}^2 E_\theta\left\{\left(\frac{\xi_t - \Delta_{t-1}X_{t-1}}{D_t}\right)^2 \mid \mathcal{F}_{t-1}\right\} - \mathcal{P}_t &= \frac{1}{\hat{\Gamma}_t^2} X_{t-1}^2 \frac{D_t + \Delta_{t-1}^2 X_{t-1}^2}{D_t^2} - \mathcal{P}_t \\ &= \frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t^2}.\end{aligned}$$

Now, since

$$\hat{\Gamma}_t^\varepsilon \left[\frac{1}{\hat{\Gamma}_t^2} X_{t-1}^2 E_\theta\left\{\left(\frac{\xi_t - \Delta_{t-1}X_{t-1}}{D_t}\right)^2 \mid \mathcal{F}_{t-1}\right\} - \mathcal{P}_t\right]^+ = \frac{\triangle\hat{\Gamma}_t}{\hat{\Gamma}_t^{2-\varepsilon}},$$

it remains only to apply Proposition B.3 (b). $\diamondsuit$

Proof of Corollary 4.3 Let us show that the conditions of Theorem 4.1 hold with $a_t = \hat{I}_t$, $\phi_t = g'/g$ and $\gamma(u) = u$ and $\mathcal{P}_t = 0$. Since $\Gamma_t = \hat{I}_t/K = \sum_{s=1}^t X_s^2/K$, is follows from the ergodic theorem that $\hat{I}_t/t \to a > 0$ and so Γ_t and $\hat{I}_t$ have the rate t, implying that the condition (R1) holds. From the ergodic theorem,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^t X_s^4 < \infty,$$

which implies that $X_t^4/t \to 0$. It follows from Corollary 4.2 that $\hat{I}_t^\varepsilon|\theta_t^{LS} - \theta| \to 0$. So, $\theta \in [\hat{\theta}_t^{LS} - c\hat{I}_t^{-\varepsilon}, \hat{\theta}_t^{LS} + c\hat{I}_t^{-\varepsilon}]$ eventually. Since $\hat{\theta}_t \in [\hat{\theta}_t^{LS} - c\hat{I}_t^{-\varepsilon}, \hat{\theta}_t^{LS} + c\hat{I}_t^{-\varepsilon}]$ we obtain that $\hat{I}_t^\varepsilon|\hat{\theta}_t - \theta| = \hat{I}_t^\varepsilon|\Delta_t| \le 2c$ eventually. Now, since $\varepsilon \ge 1/4$, we have

$$X_{t-1}\Delta_{t-1} = X_{t-1}\hat{I}_{t-1}^{-\varepsilon}\hat{I}_{t-1}^{\varepsilon}\Delta_{t-1} \to 0. \tag{A.16}$$

Let us now check condition (R2). Denote

$$Q(w) = \int_{-\infty}^{\infty} \phi\,(z-w)\,g(z)dz.$$

Since $Q(0) = 0$, we have $Q(w) = Q'(0)w + o(w)$ where $o(w)/w \to 0$ as $w \to 0$. From (4.6), there exists $d > 0$ such that $Q'(0) < -\frac{1}{2K} - d$. So,

$$wQ(w) \le -(\frac{1}{2K} + d)w^2 + wo(w).$$

This implies that $wQ(w) \le -\frac{1}{2K}w^2 - w^2(d + o(w)/w) \le -\frac{1}{2K}w^2$ for small w's. Now, since $\mathcal{P}_t = 0$, for the l.h.s. of (4.1) we obtain

$$2\Delta_{t-1}\Gamma_t^{-1}X_{t-1}Q(\Delta_{t-1}X_{t-1}) \le -\lambda_t(\theta)\Delta_{t-1}^2 \tag{A.17}$$

eventually, where $\lambda_t(\theta) = X_{t-1}^2/(K\Gamma_t)$. Now, (4.2) trivially holds since $\triangle a_t/a_t = X_{t-1}^2/(K\Gamma_t)$ as well.

Let us prove now that (R3) holds with $\mathcal{P}_t = 0$. We have,

$$\frac{1}{\Gamma_t^2}X_{t-1}^2 E_\theta\left\{\phi_t^2(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} = \frac{1}{\Gamma_t^2}X_{t-1}^2\hat{Q}(\Delta_{t-1}X_{t-1}),$$

where

$$\hat{Q}(w) = \int_{-\infty}^{\infty}\left(\frac{g'}{g}\,(z-w)\right)^2 g(z)dz.$$

Since $\Delta_{t-1}X_{t-1} \to 0$, $\Gamma_t \to \infty$ and $\triangle\Gamma_t = X_{t-1}^2/K$, (R3) follows from (4.7) and B.3 (b). $\diamondsuit$

Proof of Theorem 4.5 Let us denote

$$\Delta_t^* = \frac{1}{\Gamma_t}\sum_{s=1}^{t}\gamma(X_{s-1})\phi_s(\xi_s).$$

Then we have to show that

$$\Gamma_t^{1/2}\delta_t^\theta = \Gamma_t^{1/2}(\Delta_t - \Delta_t^*) \to 0$$

in probability P^θ. Using (4.9) we obtain

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\Gamma_t}\gamma(X_{t-1})\phi_t(X_t - \hat{\theta}_{t-1}X_{t-1}) + \mathcal{I}_t,$$

P^θ-a.s., where $\mathcal{I}_t = 0$ P^θ-eventually. So,

$$\Delta_t = \Delta_{t-1} + \frac{1}{\Gamma_t}\gamma(X_{t-1})\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) + \mathcal{I}_t,$$

which can be rewritten as

$$\Delta_t = \left(\mathbf{1} - \Gamma_t^{-1}\triangle\Gamma_t\right)\Delta_{t-1} + \Gamma_t^{-1}\left[\triangle\Gamma_t\Delta_{t-1} + R_t\right] + \Gamma_t^{-1}\varepsilon_t + \Gamma_t^{-1}\Gamma_t\mathcal{I}_t, \qquad \text{(A.18)}$$

where

$$R_t = \gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\}$$

and

$$\varepsilon_t = \gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\} - R_t$$

is a P^θ-martingale difference.

Denote

$$\mathcal{H}_t := \sum_{s=1}^{t}\left(\triangle\Gamma_s\Delta_{s-1} + R_s\right),$$

$$\bar{M}_t := \sum_{s=1}^{t}\varepsilon_s \quad \text{and} \quad \tilde{\mathcal{I}}_t := \sum_{s=1}^{t}\Gamma_s\mathcal{I}_s.$$

Then the expression

$$\Delta_t = \Gamma_t^{-1}\left\{\bar{M}_t + \mathcal{H}_t + \tilde{\mathcal{I}}_t + \Delta_0\right\}, \qquad t \geq 1$$

can easily be obtained by inspecting the difference between t'th and $(t-1)$'th term of this sequence to check that (A.18) holds. Therefore,

$$\delta_t^\theta = \Delta_t - \Delta_t^* = \Gamma_t^{-1}\left\{M_t + \mathcal{H}_t + \tilde{\mathcal{I}}_t + \Delta_0\right\}, \qquad t \geq 1,$$

where

$$M_t := \sum_{s=1}^{t}\left(\varepsilon_s - \gamma(X_{s-1})\phi_s(\xi_s)\right).$$

Now, (1) implies that $\Gamma_t^{-1/2}\mathcal{H}_t \to 0$ in probability P^θ. By (2),

$$\Gamma_t^{-1/2}M_t = \Gamma_t^{-1/2}\sum_{s=1}^{t}\mathcal{E}_s(\theta) \to 0$$

in probability P^θ. Also, since $\mathcal{I}_t = 0$ eventually and $\Gamma_t \to \infty$, we obtain that $\Gamma_t^{-1/2}\tilde{\mathcal{I}}_t \to 0$. So, we conclude that $\Gamma_t^{1/2}\delta_t^\theta \to 0$ in probability P^θ. $\diamondsuit$

Proof of Proposition 4.6 The proof is the same as that of Proposition 2 in [22], with $A_t = \Gamma_t^{1/2}$ and $R_t(\theta, \Delta_{t-1}) = \gamma(X_{t-1})E_\theta\left\{\phi_t(\xi_t - \Delta_{t-1}X_{t-1}) \mid \mathcal{F}_{t-1}\right\}$, with only exception, that in the Lenglart-Rebolledo inequality ${A_t}^{(jj)}$ should be replaced by $G_t^{1/2}$ and then (c) will follow from the condition (E). $\diamondsuit$

Proof of Corollary 5.1 The result is immediate from Corollary 3.3 and 4.2 with $\alpha_t = -\infty$ and $\beta_t = \infty$. $\diamondsuit$

Proof of Corollary 5.3 It is easy to see that the conditions of Corollary 4.3 are satisfied. This implies that $t^\delta \Delta_t \to 0$. Also, it follows from the proof of Corollary 4.3 that $\Delta_{t-1}X_{t-1} \to 0$ and $X_{t-1}/t^{1/4} \to 0$ and $\hat{I}_t/t \to a > 0$. Also, since $\Gamma_t/t \to i_g a > 0$, we have $\Gamma_t^\delta \Delta_{t-1} \to 0$ ($0 < \delta < 1/2$) and $X_{t-1}/\Gamma_t^{1/4} \to 0$. To proof the asymptotic normality, let us check that the conditions of Theorem 4.5 hold with $\Gamma_t = \hat{I}_t i_g$, $\phi_t = g'/g$ and $\gamma(u) = u$. It is easy to see that (4.9) holds (see Remark 4.4 (c)). Let us check (L1). Denoting

$$Q(w) = \int_{-\infty}^{\infty} \frac{g'}{g}\,(z-w)\,g(z)dz.$$

and using (5.7), we can rewrite

$$\Gamma_s^{1/2}\left(\triangle\Gamma_s\Delta_{s-1} + \gamma(X_{s-1})E_\theta\left\{\phi_s(\xi_s - \Delta_{s-1}X_{s-1}) \mid \mathcal{F}_{s-1}\right\}\right)$$

as

$$\begin{aligned}
\Gamma_s^{1/2}\left(i_g X_{s-1}^2\Delta_{s-1} + X_{s-1}Q(X_{s-1}\Delta_{s-1})\right) &= \Gamma_s^{1/2}X_{s-1}\left(X_{s-1}\Delta_{s-1}\right)^{1+\varepsilon_0}O(1)\\
&= \Gamma_s^{1/2}\Gamma_s^{-\delta(1+\varepsilon_0)}\left(\Gamma_s^\delta\Delta_{s-1}\right)^{1+\varepsilon_0}X_{s-1}^{2+\varepsilon_0}O(1)\\
&= \left(\frac{X_{s-1}}{\Gamma_s^{1/4}}\right)^{\varepsilon_0}\Gamma_s^{1/2-\delta(1+\varepsilon_0)+\varepsilon_0/4}X_{s-1}^2 O(1)\\
&= \Gamma_s^{1/2-\delta(1+\varepsilon_0)+\varepsilon_0/4}X_{s-1}^2 O(1)\\
&= \Gamma_s^{1/2-\delta(1+\varepsilon_0)+\varepsilon_0/4}\triangle\Gamma_s O(1).
\end{aligned}$$

Note that we can take $\delta > (2+\varepsilon_0)/(4(1+\varepsilon_0)) < \frac{1}{2}$. Then $1/2 - \delta(1+\varepsilon_0) + \varepsilon_0/4 < 0$, and (L1) now follows from the Toeplitz lemma.

Condition (E) of Proposition 4.6 trivially holds. Let us show that (L2) is satisfied. Denoting $w_s = X_s\Delta_s$, we have

$$\mathcal{E}_s(\theta) = X_{s-1}\left[\frac{g'}{g}(\xi_s - w_{s-1}) - \frac{g'}{g}(\xi_s) - Q(w_{s-1})\right]$$

Since

$$\mathcal{E}_s^2(\theta) \le 2X_{s-1}^2\left[\frac{g'}{g}(\xi_s - w_{s-1}) - \frac{g'}{g}(\xi_s)\right]^2 + 2X_{s-1}^2 Q^2(w_{s-1}),$$

we obtain

$$\begin{aligned}
E_\theta\left\{\mathcal{E}_s^2(\theta) \mid \mathcal{F}_{t-1}\right\} \;&\le\; 2X_{s-1}^2\int_{-\infty}^{\infty}\left[\frac{g'}{g}(z - w_{s-1}) - \frac{g'}{g}(z)\right]^2 g(z)dz\\
&+\; 2X_{s-1}^2 Q^2(w_{s-1}).
\end{aligned}$$

Since $\triangle\Gamma_t = i_g X_{s-1}^2$ and $w_s \to 0$, (L2) follows from (5.9), continuity of the function $Q(w)$ and the Toeplitz lemma. hence,

$$\hat{\theta}_t - \theta = \frac{1}{\hat{I}_t i_g} \sum_{s=1}^{t} X_{s-1} \frac{g'}{g}(\xi_s) + \delta_t^\theta,$$

where $\hat{I}_t^{1/2} \delta_t^\theta \to 0$ in probability P^θ. Also,

$$E_\theta \left\{ \left(X_{s-1} \frac{g'}{g}(\xi_s) \right)^2 \mid \mathcal{F}_{s-1} \right\} = X_{s-1}^2 E_\theta \left\{ \left(\frac{g'}{g} \right)^2 (\xi_s) \right\} = X_{s-1}^2 i_g$$

and

$$E_\theta \left\{ \left(X_{s-1} \frac{g'}{g}(\xi_s) \right)^2 \right\} = E_\theta \left\{ X_{s-1}^2 \right\} i_g = i_g \sigma^2/(1-\theta^2).$$

It follows from the central limit theorem for martingales by Billingsley and Ibragimov (see e.g., Theorem 1.4, Appendix 1 in [2]), that

$$\frac{1}{t^{1/2}} \sum_{s=1}^{t} X_{s-1} \frac{g'}{g}(\xi_s) \xrightarrow{d} \mathcal{N}\left(0, \;\; i_g \sigma^2/(1-\theta^2)\right).$$

Also, it follows from the ergodic theorem that

$$\frac{1}{t} \sum_{s=1}^{t} X_{s-1}^2 = \hat{I}_t/t \to \sigma^2/(1-\theta^2) > 0,$$

which implies

$$t^{1/2}(\hat{\theta}_t - \theta) = \frac{t}{\hat{I}_t i_g} \frac{1}{t^{1/2}} \sum_{s=1}^{t} X_{s-1} \frac{g'}{g}(\xi_s) \xrightarrow{d} \mathcal{N}\left(0, \;\; \frac{(1-\theta^2)}{\sigma^2 i_g}\right)$$

and (5.10). $\diamondsuit$

B Auxiliary results

Lemma B.1 *Let $\mathcal{F}_0, \mathcal{F}_1, \ldots$ be a non-decreasing sequence of σ-algebras and $X_n, \beta_n, \xi_n, \zeta_n \in \mathcal{F}_n$, $n \geq 0$, are nonnegative r.v.'s such that*

$$E(X_n|\mathcal{F}_{n-1}) \leq X_{n-1}(1+\beta_{n-1}) + \xi_{n-1} - \zeta_{n-1}, \quad n \geq 1$$

eventually. Then

$$\{\sum_{i=1}^{\infty} \xi_{i-1} < \infty\} \cap \{\sum_{i=1}^{\infty} \beta_{i-1} < \infty\} \subseteq \{X \to\} \cap \{\sum_{i=1}^{\infty} \zeta_{i-1} < \infty\} \quad (P\text{-}a.s.),$$

where $\{X \to\}$ denotes the set where $\lim_{n\to\infty} X_n$ exists and is finite.

Remark B.2 The proof can be found in [17]. Note also that this lemma is a special case of the theorem on the convergence sets nonnegative semimartingales (see, e.g., [11]).

Proposition B.3 *If d_n is a nondecreasing sequence of positive numbers such that $d_n \to +\infty$, then*

$$(a) \quad \sum_{n=1}^{\infty} \triangle d_n / d_n = +\infty$$

and

$$(b) \quad \sum_{n=1}^{\infty} \triangle d_n / d_n^{1+\varepsilon} < +\infty$$

for any $\varepsilon > 0$.

Proof These can easily be obtained by elementary arguments (see, e.g., [21], Appendix 2). $\diamondsuit$

Lemma B.4 *Suppose that $g \not\equiv 0$ is a nonnegative even function on $\mathbb{R}$ and $g \downarrow 0$ on $\mathbb{R}_+$. Suppose also that ϕ is a measurable odd function on $\mathbb{R}$ such that $\phi(z) > 0$ for $z > 0$ and $\int_{\mathbb{R}} |\phi(z-w)| g(z) dz < \infty$ for all $w \in \mathbb{R}$. Then*

$$w \int_{-\infty}^{\infty} \phi\left(z - w\right) g(z) dz < 0$$

for any $w \neq 0$. Furthermore, if $g(z)$ is continuous, then for any $\varepsilon \in (0, 1)$

$$\sup_{\varepsilon \le |w| \le 1/\varepsilon} w \int_{-\infty}^{\infty} \phi\left(z - w\right) g(z) dz < 0.$$

Remark B.5 The proof of this lemma is given [20] (see Lemma A2 in Appendix A).

References

[1] Aase, K.K. R(1983). Recursive estimators in non-linear time series models of autoregressive type. *J. R. Stat. Soc., ser. B* **45**, 228–237.

[2] Basawa, I.V., Prakasa Rao, B.L.S. (1980). *S*tatistical Inference for Stochastic Processes. Springer-Verlag, New York.

[3] Belister, E. (2000). Recursive estimation of the drifted autoregressive parameter. *Ann. Statist.* **28**, 860-870.

[4] CAMPBELL, K. (1982). Recursive computation of M-estimates for the parameters of a finite autoregressive process. *Ann. Statist.* **10**, 442-453.

[5] DIPPON, J. (1998). Globally convergent stochastic optimazation with optimal asymptotic distribution. *J. Appl. Prob.* **35**, 395-406.

[6] ENGLUND, J.-E., HOLST, U., AND RUPPERT, D. (1989). Recursive estimators for stationary, strong mixing processes – a representation theorem and asymptotic distributions. *Stochastic Processes Appl.* **31**, 203–222.

[7] FABIAN, V. (1978). On asymptotically efficient recursive estimation. *Ann. Statist.* **6**, 854-867.

[8] GU, M.G. and LI S. (1998). A stochastic approximation algorithm for maximum-likelihoood estimation with incomplete data. *The Canadian Journal of Statistics* **26**, 567-582.

[9] KHAS'MINSKII, R.Z., NEVELSON, M.B. (1972). *Stochastic Approximation and Recursive Estimation.* Nauka, Moscow.

[10] LAI, T.L. (2003). Stochastic approximation, *Ann. Statist.* **31**, 391-406.

[11] LAZRIEVA, N., SHARIA, T. AND TORONJADZE, T. (1997). The Robbins-Monro type stochastic differential equations. I. Convergence of solutions. *Stochastics and Stochastic Reports.* **61**, 67–87.

[12] LAZRIEVA, N. AND TORONJADZE, T. (1987). Ito-Ventzel's formula for semimartingales, asymptotic properties of MLE and recursive estimation, *Lect. Notes in Control and Inform. Sciences, 96, Stochast. diff. systems, H.J, Engelbert, W. Schmidt (Eds.), Springer,* 346–355.

[13] LEONOV, S.L. (1988). On recurrent estimation of autoregression parameters, *Avtomatika i Telemekhanika,* **5**, 105-116.

[14] LJUNG, L. PFLUG, G. and WALK, H. (1992). *Stochastic Approximation and Optimization of Random Systems,* Birkhäuser, Basel.

[15] LJUNG, L. and SODERSTROM, T. (1987). *Theory and Practice of Recursive Identification,* MIT Press.

[16] ROBBINS, H. and MONRO, S.: A stochastic approximation method, *Ann. Statist.* **22** (1951), 400–407.

[17] ROBBINS, H. AND SIEGMUND, D. (1971). A convergence theorem for non-negative almost supermartingales and some applications. *Optimizing Methods in Statistics.* ed. J.S. Rustagi Academic Press, New York, 233–257.

[18] Sharia, T. (1997). Truncated recursive estimation procedures, *Proc. A. Razmadze Math. Inst.* **115**, 149–159.

[19] Sharia, T. (1998). On the recursive parameter estimation for the general discrete time statistical model. *Stochastic Processes Appl.* **73**, **2**, 151–172.

[20] Sharia, T. (2008). Recursive parameter estimation: Convergence. *Statistical Inference for Stochastic Processes.* **11**, 2, pp. 157 – 175.

[21] Sharia, T. (2007). Rate of convergence in recursive parameter estimation procedures. *Georgian Mathematical Journal.* **14**, 4, pp. 721–736.

[22] Sharia, T. (2008). Recursive parameter estimation: Asymptotic expansion. *The Annals of The Institute of Statistical Mathematics* (DOI: 10.1007/s10463-008-0179-z).

[23] Shiryayev, A.N. (1984). *Probability,* Springer-Verlag, New York.