

## RECURSIVE ESTIMATION OF A DRIFTED AUTOREGRESSIVE PARAMETER

BY EDUARD BELITSER

*Leiden University*

Suppose the  $X_0, \dots, X_n$  are observations of a one-dimensional stochastic dynamic process described by autoregression equations when the autoregressive parameter is drifted with time, i.e. it is some function of time:  $\theta_0, \dots, \theta_n$ , with  $\theta_k = \theta(k/n)$ . The function  $\theta(t)$  is assumed to belong a priori to a predetermined nonparametric class of functions satisfying the Lipschitz smoothness condition. At each time point  $t$  those observations are accessible which have been obtained during the preceding time interval. A recursive algorithm is proposed to estimate  $\theta(t)$ . Under some conditions on the model, we derive the rate of convergence of the proposed estimator when the frequency of observations  $n$  tends to infinity.

**1. Introduction.** Consider an autoregression model with a drifted parameter

$$(1) \quad X_{k+1} = \theta_k X_k + \xi_{k+1}, \quad X_0 = 0, \quad k = 0, 1, \dots, n,$$

where  $\theta_k = \theta_{k,n} = \theta(k/n)$  for some function  $\theta(t)$  and  $\{\xi_k\}_{k=1}^n$  is a sequence of random variables satisfying the following conditions:

- (A1) the sequence  $\{\xi_k\}_{k=1}^n$  is uniformly bounded:  $\sup_{1 \leq k \leq n} |\xi_k| \leq M$  almost surely;
- (A2) there exists some positive  $\sigma_\xi$  such that  $\inf_{1 \leq k \leq n} E \xi_k^2 \geq \sigma_\xi^2 > 0$ ;
- (A3)  $\{\xi_k\}_{k=1}^n$  is a martingale difference sequence with respect to an increasing sequence of  $\sigma$ -fields  $\{\mathcal{F}_k\}_{k=0}^n$  ( $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n$ ), that is,  $\xi_k$  is measurable with respect to  $\mathcal{F}_k$ ,  $k = 1, \dots, n$  and

$$E\{\xi_k | \mathcal{F}_{k-1}\} = 0, \quad k = 1, \dots, n + 1.$$

For example, the case when  $\{\xi_k\}_{k=1}^n$  are independent identically distributed random variables such that  $E \xi_1 = 0$  and  $|\xi_1| \leq M$  almost surely fits in this framework. Here  $\mathcal{F}_k = \sigma(\xi_1, \dots, \xi_k)$ ,  $k = 1, \dots, n$ , is the  $\sigma$ -field generated by the random variables  $\xi_1, \dots, \xi_k$  and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$  is the trivial  $\sigma$ -field.

The unknown function  $\theta(t)$  is assumed to belong to the class  $\Sigma = \Sigma(q, L)$  for some  $q, 0 < q < 1$ ,

$$\Sigma(q, L) = \{\theta(\cdot) : \sup_{t \in [0,1]} |\theta(t)| \leq q; |\theta(t_1) - \theta(t_2)| \leq L|t_1 - t_2|, t_1, t_2 \in [0, 1]\}.$$

---

Received January 1998; revised January 2000.

AMS 1991 subject classifications. Primary 62M10, 62G20; secondary 60F99.

Key words and phrases. Autoregressive model, convergence rate, recursive algorithm.

Note that the process (1) is stable due to the condition  $q < 1$  [see Brockwell and Davis (1991)].

We will interpret the argument  $t$  as time and call  $\theta(t)$  a nonparametric signal. We focus on the problem of estimating a signal value  $\theta(t)$  at a fixed time point  $t$ ,  $t \in [0, 1]$ . Without loss of generality we take  $[0, 1]$  to be the time interval under consideration. The estimation problem on any finite time interval can be reduced to this case with some minor adjustments (infinite intervals can be treated as well).

An estimator  $\hat{\theta}(t) = \hat{\theta}_n(t) = \hat{\theta}_n(t, X_1, \dots, X_n)$  of the signal value  $\theta(t)$  is a measurable function of the observations. It is assumed that observations (1) appear successively so that at a fixed moment  $t \in [0, 1]$  only those observations  $X_k$  are accessible which have been obtained during the preceding time interval  $k/n \leq t$ . So, at a fixed time point  $t$  we consider only those estimators which are based on the observations  $X_0, X_1, \dots, X_k$  accessible at time point  $t$ , i.e.  $k/n \leq t$ . Without loss of generality one can assume that  $\mathcal{F}_k = \sigma(\xi_1, \dots, \xi_k)$ ,  $k = 1, \dots, n$ , is the  $\sigma$ -field generated by the random variables  $\xi_1, \dots, \xi_k$  and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ .

The integer parameter  $n$  is the frequency of observations, i.e. the number of observations per time unit. The estimation problem is studied in an asymptotic setup as this parameter tends to infinity. In fact we study the sequence of models:

$$X_{k+1,n} = \theta_{k,n} X_{k,n} + \xi_{k+1,n}, \quad X_{0,n} = 0, \quad k = 0, 1, \dots, n, \quad n = 1, 2, \dots$$

To avoid typographical excess, it is convenient to omit  $n$  in subscripts: for example,  $\theta_k = \theta_{k,n}$ ,  $X_k = X_{k,n}$  etc. Sometimes we supply the probability and expectation signs with subscript  $\theta$ ,  $P_\theta$  and  $E_\theta$ , to emphasize that the corresponding measure depends on  $\theta$ .

The classical problem of estimating the constant parameter of an autoregression, that is,  $\theta_k = \theta$ , has been treated by a number of authors. The least squares estimator proved to be  $\sqrt{n}$ -consistent in the stable case. The properties of this estimator and its various modifications have been thoroughly studied [see, e. g., Brockwell and Davis (1991), Dmitrienko et al. (1997) and further references therein]. Poznyak (1979), Verulava (1981), Leonov (1988) considered the almost sure convergence of stochastic approximation algorithms in the problem of estimating an autoregressive parameter. The literature on recursive estimation is quite extensive; see Nevelson and Hasminskii (1973), Ljung (1977), Ljung and Söderström (1983), Ljung (1987), Kushner and Yin (1997) and further references therein. Dahlhaus (1997) elaborated on aspects of parametric inference for a rather general class of time series models that have an evolutionary spectral representation. This class contains the autoregressive model (1) as a special case, with function  $\theta$  known up to a finite dimensional parameter.

When the autoregressive parameter is regarded as expressing some internal relationship of the dynamical process under study, it is natural to consider the case when the parameter itself varies with time [cf. also with Dahlhaus (1997)], that is, it is a function of time. In this paper we pursue

the nonparametric formulation of the estimation problem: this function is assumed to satisfy a smoothness condition rather than to belong to some parametric family. Since the parameter of autoregression varies with time in the observation model considered, one should not expect the same accuracy of estimation as in case of a constant parameter. The smoothness condition on the function describing the drift of the autoregressive parameter enables us to propose a consistent estimator and specify its rate of convergence which turns out to be  $n^{1/3}/\log n$ . The proposed estimator is based on a stochastic approximation procedure and has a recursive form, which makes it easy to compute.

**2. A recursive estimator.** The following recursive algorithm gives an estimator for the values of the signal  $\theta_k = \theta(k/n)$ ,  $k = 1, \dots, n$ :

$$(2) \quad \hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_n X_k (X_{k+1} - \hat{\theta}_k X_k), \quad k = 0, 1, \dots, n-1,$$

with the initial condition  $\hat{\theta}_0 = 0$  and  $\gamma_n = n^{-2/3} \log n$ .

The estimator  $\hat{\theta}_k = \hat{\theta}_{k,n}$  can not be consistent for all signal values  $\theta(k/n)$ ,  $k \geq 0$ , since the true initial value  $\theta(0)$  is unknown. Nevertheless, this estimator is consistent beginning with a certain moment. Moreover, the rate of convergence of the risk of this estimator is described by the following theorem.

**THEOREM 1.** *Let  $J_n = \{i \in \mathbb{N} : 2\sigma_\xi^{-2} n^{2/3} \leq i \leq n\}$ , where  $\sigma_\xi$  appears in Condition (A2). Then, for any fixed  $\alpha > 0$ , any sequence  $\{k_n\}$  such that  $k_n \in J_n$  and any  $\theta(\cdot) \in \Sigma$ ,*

$$(3) \quad \lim_{n \rightarrow \infty} \frac{n^{1/3}}{(\log n)^{3/2+\alpha}} |\hat{\theta}_{k_n} - \theta_{k_n}| = 0 \quad \text{almost surely.}$$

Moreover, for some positive constant  $A$ , the relation

$$(4) \quad \lim_{n \rightarrow \infty} \max_{k \in J_n} \frac{n^{2/3}}{(\log n)^2} E_\theta (\hat{\theta}_k - \theta_k)^2 \leq A$$

holds uniformly in  $\theta(\cdot) \in \Sigma$ .

Notice that, for any  $0 < \varepsilon < 1$ ,  $\{k \in \mathbb{N} : \varepsilon \leq k/n \leq 1\} \subset J_n$  for sufficiently large  $n$ .

Let  $\hat{\theta}(t) = \hat{\theta}_n(t)$  be a piecewise constant continuation of  $\hat{\theta}_k = \hat{\theta}(k/n)$ ,  $k = 0, 1, \dots, n$ , that is,  $\hat{\theta}(t) = \hat{\theta}_i$  for  $i/n \leq t < (i+1)/n$ ,  $i = 0, 1, \dots, n-1$  and  $\hat{\theta}(1) = \hat{\theta}_n$ . The following corollary follows immediately from Theorem 1 and the Lipschitz condition on functions from the class  $\Sigma$ .

**COROLLARY 1.** *Let  $T_n = \{t : 2\sigma_\xi^{-2} n^{-1/3} \leq t \leq 1\}$ . Then, for any fixed  $\alpha > 0$ ,  $t \in (0, 1]$  and  $\theta(\cdot) \in \Sigma$ ,*

$$\lim_{n \rightarrow \infty} \frac{n^{1/3}}{(\log n)^{3/2+\alpha}} |\hat{\theta}(t) - \theta(t)| = 0 \quad \text{almost surely.}$$

Moreover, for some positive constant  $A$ , the relation

$$\lim_{n \rightarrow \infty} \max_{t \in T_n} \frac{n^{2/3}}{(\log n)^2} E_\theta (\hat{\theta}(t) - \theta(t))^2 \leq A$$

holds uniformly in  $\theta(\cdot) \in \Sigma$ .

REMARK 1. The estimator (2) is in fact a particular one from the class of recursive estimators proposed by Poznyak (1979) [cf. also Verulava (1981) and Leonov (1988)] for the case of constant autoregressive parameter, with the difference that the sequence  $\gamma_n$  is specified. This estimator may be thought of as a recursive version of the least squares estimator. Indeed,  $\hat{\theta}_{k+1} = \hat{\theta}_k - \alpha_n \nabla J_k(\hat{\theta}_k)$ ,  $k = 0, 1, \dots, n-1$ , where  $\nabla J_k(u) = \partial(X_{k+1} - uX_k)^2 / \partial u$  and  $\alpha_n = \gamma_n/2$ .

REMARK 2. As is evident from the proof, we have in fact established a slightly stronger result instead of (3): for any fixed  $\alpha, \varepsilon > 0$  and any sequence  $\{k_n\}$  such that  $k_n \in J_n$ ,

$$\sum_{n=1}^{\infty} P_\theta \left\{ n^{1/3} (\log n)^{-(3/2+\alpha)} |\hat{\theta}_{k_n} - \theta_{k_n}| > \varepsilon \right\} < \infty$$

uniformly in  $\theta(\cdot) \in \Sigma$ . One can make the corresponding assertion from Corollary 1 slightly stronger as well: for any fixed  $\alpha, \varepsilon > 0$  and  $t \in (0, 1]$ ,

$$\sum_{n=1}^{\infty} P_\theta \left\{ n^{1/3} (\log n)^{-(3/2+\alpha)} |\hat{\theta}(t) - \theta(t)| > \varepsilon \right\} < \infty$$

uniformly in  $\theta(\cdot) \in \Sigma$ .

REMARK 3. Analyzing the proof of the theorem, one can see that the choice  $\gamma_n = n^{-2/3} \log n$  comes essentially from the balancing the terms in the upper bound for the risk of the estimator and the size of the set  $J_n$ . Note also that taking  $\gamma_n = Cn^{-2/3} \log n$  leads to  $J_n = \{i \in \mathbb{N} : 2\sigma_\xi^{-2} C^{-1} n^{2/3} \leq i \leq n\}$ . So, we can enlarge the set  $J_n$  by taking a bigger constant  $C$ , but the constant  $A$  becomes then bigger too.

REMARK 4. As an Associate Editor pointed out, in practice the choice of the step size  $\gamma$  should depend only on characteristics of the process. This raises the problem of adaptation of the recursive procedure to the local properties of the underlying signal  $\theta$ , for example by choosing the step size  $\gamma$  from the data alone. This is particularly important when the signal function is not homogeneously smooth. We give here some heuristic arguments. Assuming that the signal function  $\theta$  is differentiable (denote by  $\theta'_k$  the derivative of  $\theta$  at  $k/n$ ) and analyzing the proofs of Lemma 3 and the theorem (neglecting for the moment the log factor), one can see that the optimal  $\gamma$  at the  $k$ th step is proportional to  $(|\theta'_k|/n)^{2/3} \approx |\theta_k - \theta_{k-1}|^{2/3}$ , explaining why the optimal  $\gamma$  only depends on the characteristics of the process and not on the choice of  $n$ .

One may now insert some estimator of  $|\theta_k - \theta_{k-1}|^{2/3}$  to get a fully adaptive procedure.

REMARK 5. It is easy to see that  $E\{\hat{\theta}_{k+1} | X_0, X_1, \dots, X_k\} = \hat{\theta}_k + \gamma_n X_k^2 (\theta_k - \hat{\theta}_k)$ . In a way, this expresses the requirement for the algorithm (2) to update the estimator  $\hat{\theta}_k$  correctly in the sense of shifting the estimator in the right direction.

REMARK 6. Interestingly, although the requirement (A2) comes from the proof, an informal interpretation of this can be as follows:  $X_i$  should be “bounded away from zero” (recall that  $EX_i = 0$ ) so that the magnitude of the step in the algorithm (2) is not too small.

REMARK 7. Since we know that  $|\theta_k| < 1, k = 0, 1, \dots, n$ , one can use a truncated version of the estimator:  $\hat{\theta}_k^t = [\hat{\theta}_k]_{-1}^1 = \max\{-1, \min\{\hat{\theta}_k, 1\}\}, k = 0, 1, \dots, n$ . The estimator  $\hat{\theta}^t = (\hat{\theta}_0^t, \hat{\theta}_1^t, \dots, \hat{\theta}_n^t)$  is certainly not worse than the estimator  $\hat{\theta}$ .

REMARK 8. Since the unknown signal  $\theta(t)$  satisfies the Lipschitz condition, it is natural to take a continuous continuation of  $\theta_k = \hat{\theta}(k/n), k = 0, 1, \dots, n$ . Obviously, the corollary is also true for piecewise linear continuation

$$\hat{\theta}(t) = (i + 1)\hat{\theta}_i - i\hat{\theta}_{i+1} + n(\hat{\theta}_{i+1} - \hat{\theta}_i)t \quad \text{for } i/n \leq t < (i + 1)/n,$$

$$i = 0, 1, \dots, n - 1 \text{ and } \hat{\theta}(1) = \hat{\theta}_n.$$

**3. Proof of the theorem.** The idea of the proof is as follows. First, by backward recursion we bound from above the absolute value of the difference between the estimator  $\hat{\theta}_k$  and the function value  $\theta_k = \theta(k/n)$  by a sum of two terms which we call the “stable term” and the “martingale term.” The deeper the backward recursion level is, the smaller the first and the bigger the second term is. Similar to the bias-variance trade-off, we choose the level of recursion and the step size  $\gamma_n$  by balancing the terms against each other.

To start, we define the difference between the estimator and the signal value

$$Y_k = \hat{\theta}_k - \theta_k, \quad k = 0, 1, \dots, n.$$

Further, introduce the notation

$$Q_k = \gamma_n X_k \xi_{k+1} + \theta_k - \theta_{k+1}, \quad k = 0, 1, \dots, n,$$

and establish the following conventions throughout this paper:

$$(5) \quad \sum_{i=m+1}^m b_i = 0, \quad \prod_{i=m+1}^m b_i = 1 \text{ for any sequence } \{b_i\}.$$

The proof of the theorem is based on several lemmas below. The first lemma gives an upper bound for the absolute value of the difference  $Y_k$  by a sum of the above mentioned stable and martingale terms.

LEMMA 1. For any integers  $k$  and  $k_0$  such that  $0 \leq k_0 \leq k \leq n-1$ , the inequality

$$|Y_{k+1}| \leq |Y_{k_0}| \prod_{i=k_0}^k (1 - \gamma_n X_i^2) + 2 \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i Q_l \right|$$

is true for sufficiently large  $n$ , uniformly in  $\theta(\cdot) \in \Sigma$ .

PROOF. First notice that

$$X_{k+1} = \sum_{i=1}^{k+1} \xi_i \prod_{j=i}^k \theta_j, \quad k = 1, \dots, n.$$

From this and the conditions on the sequence  $\{\xi_k\}_{k=1}^n$ , it follows that

$$(6) \quad E_\theta X_{k+1}^2 \leq \frac{M^2 q^2}{1 - q^2} \quad \text{and} \quad |X_k| \leq \frac{Mq}{1 - q}, \quad k = 0, 1, \dots, n,$$

almost surely.

According to (1) and (2), we have the following representation:

$$\begin{aligned} Y_{k+1} &= \hat{\theta}_k + \gamma_n X_k (X_{k+1} - \hat{\theta}_k X_k) - \theta_{k+1} \\ &= Y_k (1 - \gamma_n X_k^2) + \gamma_n X_k \xi_{k+1} + \theta_k - \theta_{k+1} = Y_k (1 - \gamma_n X_k^2) + Q_k. \end{aligned}$$

Iterating in the last relation gives

$$Y_{k+1} = Y_{k_0} \prod_{i=k_0}^k (1 - \gamma_n X_i^2) + \sum_{i=k_0}^k Q_i \prod_{j=i+1}^k (1 - \gamma_n X_j^2).$$

Using Abel's transformation for series, we have

$$\begin{aligned} & \sum_{i=k_0}^k Q_i \prod_{j=i+1}^k (1 - \gamma_n X_j^2) \\ (7) \quad &= \sum_{i=k_0}^k Q_i - \sum_{i=k_0}^{k-1} \sum_{j=k_0}^i Q_j \left( \prod_{l=i+2}^k (1 - \gamma_n X_l^2) - \prod_{l=i+1}^k (1 - \gamma_n X_l^2) \right) \\ &= \sum_{i=k_0}^k Q_i - \sum_{i=k_0}^{k-1} \sum_{j=k_0}^i Q_j \gamma_n X_{i+1}^2 \prod_{l=i+2}^k (1 - \gamma_n X_l^2). \end{aligned}$$

Now recall that  $|X_k|$ ,  $k = 0, 1, \dots, n$ , are all bounded by the same constant  $Mq(1-q)^{-1}$  almost surely. Let  $n$  be sufficiently large, so that  $\gamma_n X_l^2 \leq 1$  for all  $l = 0, 1, \dots, n$ . One can take for example  $n \geq N_0$ , with  $N_0 = \max\{M^6 q^6 / (1-q)^6, m_0\}$ ,  $m_0 = \min\{m : j^{1/3} \geq \log j \text{ for all } j \geq m\}$ . Thus, from the last two relations we obtain that for  $n \geq N_0$

$$|Y_{k+1}| \leq |Y_{k_0}| \prod_{i=k_0}^k (1 - \gamma_n X_i^2)$$

$$+ \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i Q_l \right| \left( 1 + \sum_{i=k_0}^{k-1} \gamma_n X_{i+1}^2 \prod_{l=i+2}^k (1 - \gamma_n X_l^2) \right).$$

Notice that the relation (7) holds for any sequence  $\{Q_i\}$ . So, taking in particular  $Q_{k_0} = 1$  and  $Q_i = 0$  for  $i > k_0$ , we derive the following relation:

$$\sum_{i=k_0}^{k-1} \gamma_n X_{i+1}^2 \prod_{l=i+2}^k (1 - \gamma_n X_l^2) = 1 - \prod_{j=k_0+1}^k (1 - \gamma_n X_j^2) \leq 1$$

for sufficiently large  $n$  (for example  $n \geq N_0$ ). Therefore, by combining the last two relations we obtain the assertion of the lemma.  $\square$

In the second lemma the expectation of the stable term is essentially evaluated.

LEMMA 2. *Let  $k$  and  $k_0$  be any integers such that  $0 \leq k_0 \leq k \leq n$ . Then, for sufficiently large  $n$ ,*

$$E_\theta \prod_{i=k_0}^k (1 - \gamma_n X_i^2)^2 \leq (1 - \gamma_n \sigma_\xi^2)^{k-k_0} \quad \text{and} \quad E_\theta \prod_{i=k_0}^k (1 - \gamma_n X_i^2) \leq (1 - \gamma_n \sigma_\xi^2)^{k-k_0}$$

uniformly in  $\theta(\cdot) \in \Sigma$ .

PROOF. Recall that, according to (A3),  $\{\xi_k\}_{k=1}^n$  is a martingale difference with respect to  $\{\mathcal{F}_k\}_{k=0}^n$  and, according to (6),  $|X_k| \leq Mq(1 - q)^{-1}$ ,  $k = 0, 1, \dots, n$ , almost surely. Let  $n$  be sufficiently large, so that

$$\gamma_n \sigma_\xi^2 \leq 1, \quad \gamma_n X_k^2 \leq \sigma_\xi^2, \quad \gamma_n X_k^4 \leq \sigma_\xi^2, \quad k = 0, 1, \dots, n,$$

almost surely. Since, by (A2), for any  $m = 1, \dots, n$ ,

$$\begin{aligned} E\{(1 - \gamma_n X_m^2)^2 \mid \mathcal{F}_{m-1}\} &= E\{1 - 2\gamma_n X_m^2 + \gamma_n^2 X_m^4 \mid \mathcal{F}_{m-1}\} \\ &\leq 1 - 2\gamma_n(\sigma_\xi^2 + \theta_{m-1}^2 X_{m-1}^2) + \gamma_n \sigma_\xi^2 \leq 1 - \gamma_n \sigma_\xi^2 \end{aligned}$$

and similarly

$$E\{(1 - \gamma_n X_m^2) \mid \mathcal{F}_{m-1}\} \leq 1 - \gamma_n \sigma_\xi^2$$

uniformly in  $\theta(\cdot) \in \Sigma$ , we have that

$$\begin{aligned} E \prod_{i=k_0}^k (1 - \gamma_n X_i^2)^2 &= E E \left\{ \prod_{i=k_0}^k (1 - \gamma_n X_i^2)^2 \mid \mathcal{F}_{k-1} \right\} \\ &= E \left\{ E [(1 - \gamma_n X_k^2)^2 \mid \mathcal{F}_{k-1}] \prod_{i=k_0}^{k-1} (1 - \gamma_n X_i^2)^2 \right\} \\ &\leq (1 - \gamma_n \sigma_\xi^2) E \prod_{i=k_0}^{k-1} (1 - \gamma_n X_i^2)^2 \end{aligned}$$

and analogously

$$E \prod_{i=k_0}^k (1 - \gamma_n X_i^2) \leq (1 - \gamma_n \sigma_\xi^2) E \prod_{i=k_0}^{k-1} (1 - \gamma_n X_i^2)$$

uniformly in  $\theta(\cdot) \in \Sigma$ . By iterating these relations, we obtain that

$$E \prod_{i=k_0}^k (1 - \gamma_n X_i^2)^2 \leq (1 - \gamma_n \sigma_\xi^2)^{k-k_0} E (1 - \gamma_n X_{k_0}^2)^2 \leq (1 - \gamma_n \sigma_\xi^2)^{k-k_0},$$

$$E \prod_{i=k_0}^k (1 - \gamma_n X_i^2) \leq (1 - \gamma_n \sigma_\xi^2)^{k-k_0}$$

for sufficiently large  $n$  and uniformly in  $\theta(\cdot) \in \Sigma$ . The lemma is proved.  $\square$

The next lemma provides an upper bound for the expectation of the martingale term.

LEMMA 3. *Let  $k$  and  $k_0$  be any integers such that  $0 \leq k_0 \leq k \leq n$ . Then, for some positive constants  $B_1$  and  $B_2$ ,*

$$E_\theta \max_{k_0 \leq i \leq k} \left\{ \sum_{l=k_0}^i Q_l \right\}^2 \leq B_1 \gamma_n^2 (k - k_0) + B_2 (k - k_0)^2 n^{-2}$$

uniformly in  $\theta(\cdot) \in \Sigma$ .

PROOF. Obviously

$$(8) \quad E \max_{k_0 \leq i \leq k} \left\{ \sum_{l=k_0}^i Q_l \right\}^2 \leq 2\gamma_n^2 E \max_{k_0 \leq i \leq k} \left\{ \sum_{l=k_0}^i X_l \xi_{l+1} \right\}^2 + 2 \max_{k_0 \leq i \leq k} (\theta_i - \theta_{k_0})^2.$$

Recall that the function  $\theta(t)$  satisfies the Lipschitz smoothness condition. Therefore,

$$(9) \quad \max_{k_0 \leq i \leq k} (\theta_i - \theta_{k_0})^2 \leq \max_{k_0 \leq i \leq k} L^2 (i - k_0)^2 n^{-2} = L^2 (k - k_0)^2 n^{-2}$$

uniformly in  $\theta(\cdot) \in \Sigma$ .

It remains to evaluate the first term in the right-hand side of (8). Notice that the sequence  $\{X_l \xi_{l+1}\}_{l=1}^n$  is a martingale difference relative to the nested sequence  $\{\mathcal{F}_l\}_{l=1}^n$ . Indeed, by (6),  $E|X_l \xi_{l+1}| \leq \sigma_\xi^2 q (1 - q^2)^{-1/2}$  and

$$E(X_l \xi_{l+1} | \mathcal{F}_l) = X_l E(\xi_{l+1} | \mathcal{F}_l) = 0, \quad l = 1, \dots, n.$$

Therefore  $\{\sum_{l=k_0}^i X_l \xi_{l+1}\}_{i=k_0}^k$  is a martingale. By applying Doob's maximal inequality for martingales, we get that

$$E \max_{k_0 \leq i \leq k} \left\{ \sum_{l=k_0}^i X_l \xi_{l+1} \right\}^2 \leq 4E \left\{ \sum_{l=k_0}^k X_l \xi_{l+1} \right\}^2 \leq B_3 (k - k_0)$$



uniformly in  $\theta(\cdot) \in \Sigma$ . The last inequality is due to the martingale difference property:  $E\{X_l \xi_{l+1} X_m \xi_{m+1}\} = 0$  if  $l \neq m$ . Combining the last relation with (8) and (9) completes the proof of the lemma.  $\square$

Now we proceed to prove the theorem. First notice that, according to (2),

$$\hat{\theta}_k = \sum_{i=1}^{k-1} \gamma_n X_i X_{i+1} \prod_{j=i+1}^{k-1} (1 - \gamma_n X_j^2).$$

Since  $\{X_i\}_{i=0}^n$  are uniformly bounded,  $|\hat{\theta}_k| \leq C_1 n \gamma_n = C_1 n^{1/3}$  for some positive constant  $C_1$  and sufficiently large  $n$ . Consequently, we have the following preliminary uniform rough estimate:

$$(10) \quad |Y_m| \leq C_1 n^{1/3} + q \leq C_2 n^{1/3}, \quad m = 0, 1, \dots, n,$$

for sufficiently large  $n$ .

Fix  $k_0, 0 \leq k_0 \leq n - 1$ . Then the last relation and Lemma 1 imply that, for  $0 \leq k_0 < k \leq n$  and sufficiently large  $n$ ,

$$(11) \quad |Y_{k+1}| \leq C_2 n^{1/3} \prod_{i=k_0}^k (1 - \gamma_n X_i^2) + 2 \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i Q_l \right|$$

uniformly in  $\theta(\cdot) \in \Sigma$ .

Choose the level of backward recursion  $k_0$  as follows:  $k_0 = k_0(k) = \lceil k - 2\sigma_\xi^{-2} \gamma_n^{-1} \log n \rceil$ , where  $\lceil b \rceil$  denotes the smallest whole number which is equal to or greater than  $b$ . So,

$$2\sigma_\xi^{-2} \gamma_n^{-1} \log n - 1 \leq k - k_0 \leq 2\sigma_\xi^{-2} \gamma_n^{-1} \log n$$

and therefore  $k_0(k)$ 's are properly defined for those  $k$ 's which satisfy

$$k + 1 \in \{i \in \mathbb{N} : 2\sigma_\xi^{-2} \gamma_n^{-1} \log n \leq i \leq n\} = J_n$$

(and also for  $k = n$ ). Now it is easy to derive the following bound:

$$(12) \quad (1 - \gamma_n \sigma_\xi^2)^{k-k_0} \leq C_3 (1 - \gamma_n \sigma_\xi^2)^{2\sigma_\xi^{-2} \gamma_n^{-1} \log n} \leq C_4 e^{-2 \log n} = C_4 n^{-2}$$

for sufficiently large  $n$  and uniformly over  $k$  such that  $k + 1 \in J_n$ .

Denote for brevity  $\phi_n = n^{1/3} / (\log n)^{3/2+\alpha}$ . Since  $k - k_0 \leq 2\sigma_\xi^{-2} \gamma_n^{-1} \log n$  and

$$\max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i Q_l \right| \leq \gamma_n \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i X_l \xi_{l+1} \right| + \max_{k_0 \leq i \leq k} |\theta_i - \theta_{k_0}|,$$

from (11) it follows that, for sufficiently large  $n$  and any  $k$  such that  $k + 1 \in J_n$ ,

$$\begin{aligned} P\{ \phi_n |Y_{k+1}| > \varepsilon \} &\leq P\left\{ C_2 \phi_n n^{1/3} \prod_{i=k_0}^k (1 - \gamma_n X_i^2) > \varepsilon/3 \right\} \\ &\quad + P\left\{ 2\phi_n \gamma_n \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i X_l \xi_{l+1} \right| > \varepsilon/3 \right\} \end{aligned}$$

$$+P \left\{ 2\phi_n \max_{k_0 \leq i \leq k} |\theta_i - \theta_{k_0}| > \varepsilon/3 \right\}.$$

Now use Markov’s inequality, Lemma 2 and (12) to bound from above the first term in the right hand side of the last inequality by

$$3\varepsilon^{-1}C_2\phi_n n^{1/3} E \prod_{i=k_0}^k (1 - \gamma_n X_i^2) \leq \frac{3C_2 n^{2/3} (1 - \gamma_n \sigma_\xi^2)^{k-k_0}}{\varepsilon(\log n)^{3/2+\alpha}} \leq \frac{C_5}{n^{4/3}(\log n)^{3/2+\alpha}}$$

for sufficiently large  $n$ , uniformly in  $\theta(\cdot) \in \Sigma$  and over  $k$  such that  $k + 1 \in J_n$ . The third term is zero for  $n$  large enough, uniformly in  $\theta(\cdot) \in \Sigma$  and over  $k$  such that  $k + 1 \in J_n$ , because

$$\phi_n \max_{k_0 \leq i \leq k} |\theta_i - \theta_{k_0}| \leq \frac{L(k - k_0)n^{1/3}}{n(\log n)^{3/2+\alpha}} \leq \frac{C_6}{(\log n)^{3/2+\alpha}}.$$

To evaluate the second term, we apply the Azuma-Hoeffding inequality [see, e.g., Williams (1991); the inequality can also be derived from a general result in de la Peña and Giné (1999)] for a martingale  $\{M_l\}_{l=1}^m$  whose increments  $d_l = M_l - M_{l-1}$  (with  $M_0 = 0$ ) are bounded in absolute value by  $c_l$ :  $P\{\sup_{1 \leq l \leq m} M_l \geq x\} \leq \exp\{-x^2/(2\sum_{l=1}^m c_l^2)\}$ . As we already know from the proof of Lemma 3,  $\{\sum_{l=k_0}^i X_l \xi_{l+1}\}_{i=k_0}^k$  is a martingale with bounded increments  $|X_l \xi_{l+1}| \leq M^2 q/(1 - q) = B$ . So, applying the Azuma-Hoeffding inequality gives

$$P \left\{ 2\phi_n \gamma_n \max_{k_0 \leq i \leq k} \left| \sum_{l=k_0}^i X_l \xi_{l+1} \right| > \varepsilon/3 \right\} \leq 2 \exp \left\{ -\frac{\varepsilon^2 n^{2/3} (\log n)^{1+2\alpha}}{72B^2(k - k_0)} \right\} \\ \leq 2 \exp \{-C_7(\log n)^{1+2\alpha}\} \leq C_8 n^{-2}$$

uniformly in  $\theta(\cdot) \in \Sigma$  and over  $k$  such that  $k+1 \in J_n$ . Since the above estimates for the all three terms are uniform in  $\theta(\cdot) \in \Sigma$  and over  $k$  such that  $k + 1 \in J_n$  (i.e., the constants  $C_5, C_6$  and  $C_8$  do not depend on  $k$ ), we conclude that, for any sequence  $\{k_n\}$  such that  $k_n \in J_n$ ,

$$\sum_{n=1}^{\infty} P\{\phi_n |Y_{k_n}| > \varepsilon\} < \infty$$

uniformly in  $\theta(\cdot) \in \Sigma$ . The relation (3) follows by the Borel-Cantelli lemma.

As  $\gamma_n = n^{-2/3} \log n$  and  $k - k_0 \leq 2\sigma_\xi^{-2} \gamma_n^{-1} \log n$ , using (11), Lemmas 2 and 3 yield that, for sufficiently large  $n$  and uniformly over  $k$  such that  $k + 1 \in J_n$ ,

$$E Y_{k+1}^2 \leq 2C_2^2 n^{2/3} (1 - \gamma_n \sigma_\xi^2)^{k-k_0} + C_9 [\gamma_n^2 (k - k_0) + (k - k_0)^2 n^{-2}] \\ \leq C_{10} n^{2/3} n^{-2} + C_{11} (\log n)^2 n^{-2/3}$$

uniformly in  $\theta(\cdot) \in \Sigma$ , which establishes (4). The proof of the theorem is complete.  $\square$

**Acknowledgment.** The author is grateful to an Associate Editor for helpful comments and suggestions.

## REFERENCES

- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.
- DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37.
- DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling. From Dependence to Independence. Randomly Stopped processes. U-Statistics and Processes. Martingales and Beyond*. Springer, New York.
- DMITRIENKO, A. A., KONEV V. V. and PERGAMENSHCHIKOV, S. M. (1997). Sequential generalized least squares estimator for an autoregressive parameter. *Sequential Anal.* **16** 25–46.
- KUSHNER, H. J. and YIN, G. G. (1997). *Stochastic Approximation Algorithms and Applications*. Springer, New York.
- LEONOV, S. L. (1988). Recurrent estimation of autoregression parameters. *Automat. Remote Control* **49** 633–642.
- LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Autom. Control* **22** 551–575.
- LJUNG, L. (1987). *System Identification—Theory for the User*. Prentice Hall, Englewood Cliffs, NJ.
- LJUNG, L. and SÖDERSTRÖM, T. (1983). *Theory and Practice of Recursive Identification*. MIT Press.
- NEVELSON, M. B. and HASMINSKII, R. Z. (1973). *Stochastic Approximation and Recursive Estimation*. Amer. Math. Soc, Providence, RI.
- POZNYAK, A. S. (1979). Convergence of stochastic approximation algorithms in parameter identification of dynamic systems. *Automat. Remote Control* **40** 1254–1258.
- VERULAVA, Y. S. (1981). Convergence of the stochastic approximation algorithm for estimating the autoregression parameter. *Automat. Remote Control* **42** 943–947.
- WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge Univ. Press.

MATHEMATICAL INSTITUTE  
LEIDEN UNIVERSITY  
NIELS BOHRWEG 1  
2333 CA LEIDEN  
THE NETHERLANDS  
E-MAIL: belitser@math.leidenuniv.nl