PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 23, Fasc. 1 (2003), pp. 61–76

ALMOST SURE PROPERTIES OF WEIGHTED VECTORIAL MARTINGALES TRANSFORMS WITH APPLICATIONS TO PREDICTION FOR LINEAR REGRESSION MODELS

BY

PEGGY CÉNAC (ORSAY)

Abstract. We establish new almost sure properties for powers of weighted martingale transforms. It allows us to deduce useful asymptotic results for cumulative prediction and estimation errors associated with linear regression models. We also provide two examples of applications on the linear and functional autoregressive models.

Notation. For any square matrix A, A^t denotes the transpose of A, $\operatorname{tr}(A)$ is the trace of A, and $\det(A)$ denotes the determinant of A. In addition, $\lambda_{\min} A$ and $\lambda_{\max} A$ are the minimum and the maximum eigenvalues of A, respectively.

AMS Subject Classification: 62J12, 62H12, 60G25, 60G42.

Key words and phrases: Linear regression models; martingales transforms; least squares estimators; cumulative prediction and estimation errors.

1. INTRODUCTION

Consider the linear regression model given, for all $n \geqslant 1$, by

$$X_{n+1} = \theta^t \Phi_n + \varepsilon_{n+1}, \tag{1.1}$$

where θ in $\boldsymbol{R}^d$ is the unknown parameter and X_n, Φ_n, ε_n are the scalar observation, the regression vector, and the scalar driven noise of the system, respectively. In the sequel, we shall assume that (ε_n) is a martingale difference sequence adapted to a filtration $\boldsymbol{F}$ with $\boldsymbol{F} = (\mathscr{F}_n)_{n \geqslant 0}$, where $\mathscr{F}_n$ is the σ-algebra of the events occurring up to time n. Our purpose is to establish asymptotic properties for cumulative prediction and estimation errors associated with the linear regression model (1.1). We shall also illustrate our results on the linear autoregressive model given, for all $n \geqslant 1$, by

$$X_{n+1} = \sum_{k=1}^{d} \theta_k X_{n-k+1} + \varepsilon_{n+1} \tag{1.2}$$

and on the parametric functional autoregressive model given, for $n \geqslant 1$, by

$$X_{n+1} = \sum_{k=1}^{d} \theta_k f_k(X_{n-k+1}) + \varepsilon_{n+1}. \tag{1.3}$$

For a reasonable sequence $(\hat{\theta}_n)$ of estimators of θ, we shall investigate the asymptotic performance of $\hat{\theta}_n^t \Phi_n$ as a predictor of X_{n+1}. More precisely, we shall focus our attention on the prediction error $X_{n+1} - \hat{\theta}_n^t \Phi_n$ and the estimation error $\hat{\theta}_n - \theta$. As well known (see e.g. [6]) it is more appropriate to consider the cumulative prediction error defined, for $p \geqslant 1$, by

$$C_n(p) = \sum_{k=0}^{n-1} (X_{k+1} - \hat{\theta}_k^t \Phi_k)^{2p} \tag{1.4}$$

and the estimation error defined, for $p \geqslant 1$, by

$$G_n(p) = \sum_{k=1}^{n} k^{p-1} \|\hat{\theta}_k - \theta\|^{2p}. \tag{1.5}$$

In the one-dimensional parameter case $d = 1$, under suitable moment conditions, asymptotic results on the cumulative prediction and estimation errors were established in [2], [4], [9] and [15]. In all these papers, the asymptotic properties were proved with the standard least squares (LS) estimator. Our goal is to extend these results in the multidimensional parameter case $d > 1$.

In the multidimensional framework, only in the particular case $p = 1$, the authors of [4], [15], [16], [12] and [17] have established asymptotic results with the standard LS estimator. The authors of [5], [10], [11] and [14] proved the strong consistency of the LS estimator for general linear autoregressive model and they studied the asymptotic behavior of the empirical estimator of the covariance associated with this model. In the case $p > 1$, in order to overcome the difficulties inherent in the multivariate framework, we have chosen to make use of the weighted least squares (WLS) estimator $\hat{\theta}_n$ of θ, introduced in [3].

The paper is organized as follows. Section 2 is devoted to new almost sure properties for powers of weighted martingales transforms. In Section 3, we propose some statistical applications to prediction and estimation for linear and functional autoregressive models. All technical proofs are collected in Appendices A and B.

2. ALMOST SURE PROPERTIES

We first propose new almost sure properties for powers of weighted vectorial martingales transforms. These properties are the keystone to understand the asymptotic behavior of cumulative prediction and estimation errors. For a sequence of random vectors (Φ_n) in R^d adapted to F, we define the weighted martingale transform

$$M_n = M_0 + \sum_{k=1}^{n} \alpha_{k-1} \Phi_{k-1} \varepsilon_k,$$

where M_0 can be arbitrarily chosen. We also set

$$S_n(\alpha) = \sum_{k=0}^{n} \alpha_k \Phi_k \Phi_k^t + S, \tag{2.1}$$

where S is a deterministic, symmetric and positive definite matrix. The weighting sequence (α_n) is adapted to F, non-increasing, with $0 \leqslant \alpha_n \leqslant 1$. Moreover, it is chosen so that

$$\sum_{n=0}^{\infty} \alpha_n f_n(\alpha) < \infty \text{ a.s.}, \tag{2.2}$$

where the explosion coefficient $f_n(\alpha)$ is given by

$$f_n(\alpha) = \alpha_n \Phi_n^t S_n^{-1}(\alpha) \Phi_n. \tag{2.3}$$

THEOREM 1. *Assume that*

$$\sup_{n \geqslant 0} E[\varepsilon_{n+1}^2 \mid \mathscr{F}_n] < \infty \text{ a.s.} \tag{2.4}$$

Then, for any $p \geqslant 1$, we have

$$\sum_{n=1}^{\infty} \left((M_n^t S_{n-1}^{-1}(\alpha) M_n)^p - (M_n^t S_n^{-1}(\alpha) M_n)^p \right) < \infty \text{ a.s.} \tag{2.5}$$

and

$$\sum_{n=1}^{\infty} (M_n^t S_{n-1}^{-1}(\alpha) M_n - M_n^t S_n^{-1}(\alpha) M_n)^p < \infty \text{ a.s.} \tag{2.6}$$

Proof. The proof is straightforward when using the two elementary inequalities given, for $x \geqslant y \geqslant 0$ and $p \geqslant 1$, by

$$(x-y)^p \leqslant x^p - y^p \leqslant p x^{p-1}(x-y). \tag{2.7}$$

Let us put, for any integer $p \geqslant 1$,

$$a_n(p) = (M_n^t S_{n-1}^{-1}(\alpha) M_n)^p - (M_n^t S_n^{-1}(\alpha) M_n)^p.$$

By choosing $x = M_n^t S_{n-1}^{-1}(\alpha) M_n$ and $y = M_n^t S_n^{-1}(\alpha) M_n$, we deduce from (2.7) that

$$(a_n(1))^p \leqslant a_n(p) \leqslant p V_n^{p-1} a_n(1) \quad \text{with } V_n = M_n^t S_{n-1}^{-1}(\alpha) M_n. \tag{2.8}$$

In addition, it has already been established in [3] and [4] that V_n is a.s. bounded and

$$\sum_{n=1}^{\infty} a_n(1) < \infty \quad \text{a.s.} \tag{2.9}$$

Consequently, (2.5) and (2.6) follow immediately from (2.8) and (2.9). ∎

Remark 1. In comparison with Theorem 2 of [2] in the scalar case, one can realize that our assumption (2.4) is really not restrictive. Thanks to our weighting sequence (α_n), this assumption is restricted only to a finite conditional moment of order 2.

Theorem 1 leads to useful information about the asymptotic properties of the cumulative prediction error $C_n(p)$, defined in (1.4). Indeed, from the Riccati formula we obtain

$$S_{n-1}^{-1}(\alpha) = S_n^{-1}(\alpha) + \alpha_n\bigl(1 - f_n(\alpha)\bigr) S_{n-1}^{-1}(\alpha)\, \Phi_n \Phi_n^t S_{n-1}^{-1}(\alpha),$$

which implies

$$M_n^t S_{n-1}^{-1}(\alpha) M_n - M_n^t S_n^{-1}(\alpha) M_n = \alpha_n\bigl(1 - f_n(\alpha)\bigr)\bigl(M_n^t S_{n-1}^{-1}(\alpha)\,\Phi_n\bigr)^2. \tag{2.10}$$

Moreover, the recall WLS estimator $\hat{\theta}_n$ of θ is given, for all $n \geqslant 1$, by

$$\hat{\theta}_n = S_{n-1}^{-1}(\alpha) \sum_{k=1}^{n} \alpha_{k-1}\, \Phi_{k-1} X_k, \tag{2.11}$$

where $S_n(\alpha)$ is defined in (2.1). The asymptotic properties of the WLS estimator were established in [3], [1] and [7]. The choice of the weighting sequence is of course crucial, and two possible choices are given in Section 3.

We clearly deduce from (1.1) and (2.11) that

$$\hat{\theta}_n - \theta = S_{n-1}^{-1}(\alpha) M_n \quad \text{with } M_0 = -S\theta.$$

Hence, we infer from (2.6) and (2.10) that

$$\sum_{n=1}^{\infty} \alpha_n^p \bigl(1 - f_n(\alpha)\bigr)^p \bigl(X_{n+1} - \hat{\theta}_n^t \Phi_n - \varepsilon_{n+1}\bigr)^{2p} < \infty \quad \text{a.s.} \tag{2.12}$$

It is often difficult to get asymptotic information on the explosion coefficient $f_n(\alpha)$. In the models considered in this paper, $f_n(\alpha)$ tends to zero a.s. Nevertheless, at this point, we only need a lower bound strictly positive for the quantity $1 - f_n(\alpha)$ to find that

$$\sum_{n=1}^{\infty} \alpha_n^p \bigl(X_{n+1} - \hat{\theta}_n^t \Phi_n - \varepsilon_{n+1}\bigr)^{2p} < \infty \quad \text{a.s.} \tag{2.13}$$

The proofs of the following corollaries rely on (2.13). We still study more precisely the consequences of this result for $C_n(p)$ and $G_n(p)$ in the statistical applications.

3. STATISTICAL APPLICATIONS

A possible application of Theorem 1 concerns the linear regression model given by (1.1). Our purpose is to investigate the asymptotic behavior of $C_n(p)$ and $G_n(p)$, defined in (1.4) and (1.5). By the same arguments, we shall estimate the moments of order p of the driven noise (ε_n).

First of all, we have already mentioned that the choice of the weighting sequence (α_n) is crucial. We shall now propose two different choices for (α_n). Let

$$S_n = \sum_{k=0}^{n} \Phi_k \Phi_k^t + S$$

and write $s_n = \mathrm{tr}(S_n)$, $d_n = \det(S_n)$, and $d_n(\alpha) = \det(S_n(\alpha))$. Since $S_n(\alpha) \leqslant S_n$, we deduce from Corollary 7.7.4 of [8] that $d_n(\alpha) \leqslant d_n$. Consequently, as

$$f_n(\alpha) = \frac{d_n(\alpha) - d_{n-1}(\alpha)}{d_n(\alpha)},$$

it is not difficult to see using (2.3) that if $S > eI$ and

$$\alpha_n = \frac{1}{(\log s_n)^{1+\gamma}} \tag{3.1}$$

with $\gamma > 0$, then the convergence (2.2) holds, as well as the other hypothesis on the weighting sequence. Let us mention that $\alpha_n = s_n^{-\gamma}$ for some $\gamma > 0$ is another possible choice. One might observe that we can also replace s_n by d_n in (3.1). The different advantages of the choices of (α_n) are discussed in [1]. In all the sequel, we shall make use of weighting sequence defined in (3.1).

3.1. Moment estimation and prediction errors. For any $p \geqslant 1$, a natural estimator of the moment of order p of the driven noise (ε_n) is given by

$$\Gamma_n(p) = \frac{1}{n} \sum_{k=0}^{n-1} (X_{k+1} - \hat{\theta}_k^t \Phi_k)^p. \tag{3.2}$$

One can observe that $n\Gamma_n(2p) = C_n(p)$. The following corollary gives asymptotic properties of $\Gamma_n(p)$.

COROLLARY 1. *Assume that s_n increases a.s. to infinity and that*

$$\limsup_{n \to +\infty} f_n(\alpha) < 1 \quad a.s. \tag{3.3}$$

Moreover, assume that one can find $p \geqslant 2$ such that, for all $n \geqslant 1$,

$$E[\varepsilon_{n+1}^p \mid \mathscr{F}_n] = \sigma(p) \quad a.s. \tag{3.4}$$

Then, as soon as $\alpha_n^{-1} = \mathcal{O}(n)$ *a.s.,* $\Gamma_n(p)$ *is a strongly consistent estimator of* $\sigma(p)$ *and*

$$\left|\Gamma_n(p) - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^p\right|^2 = o\left(\frac{(\log s_n)^{1+\gamma}}{n}\right) \text{ a.s.} \tag{3.5}$$

In addition, if p is even and if for all $n \geqslant 1$, $E[\varepsilon_{n+1}^{p-1} \mid \mathscr{F}_n] = 0$ *a.s., then, as soon as* $\alpha_n^{-p} = \mathcal{O}(n^2)$ *a.s., we can improve* (3.5) *by*

$$\left|\Gamma_n(p) - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^p\right| = o\left(\frac{(\log s_n)^{p(1+\gamma)/2}}{n}\right) \text{ a.s.} \tag{3.6}$$

The proof is given in Appendix A.

Remark 2. Corollary 1 still holds if we replace (3.3) by

$$\|\Phi_n\|^2 = \mathcal{O}(\alpha_n^{-1}) \text{ a.s.} \tag{3.7}$$

Moreover, if we assume that $\lambda_{\min} S_n(\alpha)$ increases a.s. to infinity, then the hypothesis (3.7) implies that $f_n(\alpha)$ tends to zero a.s. One can notice that if $s_n = \mathcal{O}(n)$ a.s., then all the assumptions on α_n in Corollary 1 are automatically satisfied for the weighting sequence defined in (3.1).

We shall now deduce from Corollary 1 the asymptotic behavior of $C_n(p)$. If one can find $p \geqslant 1$ such that, for all $n \geqslant 1$, $E[\varepsilon_{n+1}^{2p} \mid \mathscr{F}_n] = \sigma(2p)$ a.s., then we infer from (3.5) that $C_n(p)/n$ converges a.s. to $\sigma(2p)$. Moreover, if (ε_n) has a finite conditional moment of order $a > 2p$, we infer from Chow's lemma (see e.g. [4], p. 22) that for c such that $2pa^{-1} < c < 1$

$$\left|\frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^{2p} - \sigma(2p)\right| = o(n^{c-1}) \text{ a.s.} \tag{3.8}$$

Consequently, if for all $n \geqslant 1$, $E[\varepsilon_{n+1}^{2p-1} \mid \mathscr{F}_n] = 0$ a.s. and $\alpha_n^{-p} = \mathcal{O}(n^{2c})$, we find from (3.6) together with (3.8) that

$$|n^{-1} C_n(p) - \sigma(2p)| = o(n^{c-1}) \text{ a.s.}$$

3.2. Estimation errors. We shall now focus our attention on the cumulative estimation error $G_n(p)$, defined in (1.5). We first need the following corollary.

COROLLARY 2. *Assume that one can find* $p \geqslant 2$ *such that* (3.4) *holds. Then*

$$\sum_{k=1}^{\infty} \alpha_k f_k(\alpha)\left((\hat{\theta}_k - \theta)^t S_k(\alpha)(\hat{\theta}_k - \theta)\right)^p < \infty \text{ a.s.} \tag{3.9}$$

Moreover, assume that one can find an invertible matrix L such that

$$\lim_{n\to+\infty} n^{-1} S_n = L \text{ a.s.} \tag{3.10}$$

Then we also have

$$\sum_{k=1}^{\infty} k^{p-1} \alpha_k^{p+1} \big((\hat{\theta}_k - \theta)^t L(\hat{\theta}_k - \theta)\big)^p < \infty \quad a.s. \tag{3.11}$$

The proof is given in Appendix B.

The convergence (3.10) together with an Abel transform implies that

$$\lim_{n \to +\infty} \frac{1}{n\alpha_n} S_n(\alpha) = L \text{ a.s.} \tag{3.12}$$

Hence the explosion coefficient $f_n(\alpha)$ tends to zero a.s. In addition, since the application trace is linear, s_n increases a.s. to infinity. Finally, we infer from (3.11) together with Kronecker's lemma that

$$G_n(p) = o\big((\log s_n)^{(p+1)(1+\gamma)}\big) \text{ a.s.}$$

3.3. Applications. We now illustrate our results on the linear and functional autoregressive models given by (1.2) and (1.3). In both situations, we shall only consider the stable case.

3.3.1. *Linear autoregressive models.* The linear autoregressive model (1.2) is a particular case of (1.1) with $\theta^t = (\theta_1, \ldots, \theta_d)$ and $\Phi_n^t = (X_n, \ldots, X_{n-d+1})$. We shall focus on the stable case, that is, $\varrho(C) < 1$, where $\varrho(C)$ denotes the spectral radius of the companion matrix C, associated with (1.2),

$$C = \begin{pmatrix} \theta_1 & \theta_2 & \ldots & \theta_{d-1} & \theta_d \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}.$$

In the sequel, for all $n \geqslant 1$, $E[\varepsilon_{n+1}^2 \mid \mathcal{F}_n] = \sigma^2$ a.s. Let us set

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 \end{pmatrix}.$$

COROLLARY 3. *Assume that (ε_n) is either a sequence of independent and identically distributed variables or a martingale difference sequence with finite conditional moment of order greater than 2. In addition, assume that one can find $p \geqslant 2$ such that (3.4) holds. Then $\Gamma_n(p)$ is a strongly consistent* *estimator of* $\sigma(p)$ *and*

$$\left|\Gamma_n(p) - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^p\right|^2 = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \text{ a.s.} \tag{3.13}$$

Furthermore, we have

$$G_n(p) = o\left((\log n)^{(1+\gamma)(p+1)}\right) \text{ a.s.}$$

and (3.11) *also holds with*

$$L = \sum_{k=0}^{\infty} C^k \Gamma (C^t)^k. \tag{3.14}$$

Finally, if p *is even and if for all* $n \geqslant 1$, $E[\varepsilon_{n+1}^{p-1} | \mathscr{F}_n] = 0$ *a.s., then* (3.6) *is true after replacing* $\log s_n$ *by* $\log n$.

Proof. If (ε_n) are independent and identically distributed, then the convergence result (3.10) is true with L given by (3.14). More generally, if (ε_n) is a martingale difference sequence with finite conditional moment of order greater than 2, then (3.10) also holds with the same limit L (see e.g. [4] and [11]). Moreover, one can easily check that L is invertible. Consequently, $\log s_n$ is a.s. equivalent to $\log n$ and Corollaries 1 and 2 hold after replacing $\log s_n$ by $\log n$. ∎

Remark 3. Assume now that the linear autoregressive model (1.2) is unstable, that is, $\varrho(C) = 1$. If (3.4) holds with $p > 2$, we can deduce from Proposition 4.4.24 of [4] that $f_n(\alpha)$ converges a.s. to zero and $\log s_n = \mathcal{O}(\log n)$ a.s. Therefore, Corollary 1 still holds if we replace $\log s_n$ by $\log n$.

3.3.2. *Functional autoregressive model.* The functional autoregressive model (1.3) is also a particular case of the model (1.1) with $\Phi_n^t = (f_1(X_n), \ldots, f_d(X_{n-d+1}))$. In order to remain in a stable framework, it is necessary to impose several restrictive conditions. More precisely, we shall assume that for all $1 \leqslant k \leqslant d$ and for all real x

$$c_k|x| + d_k \leqslant |f_k(x)| \leqslant a_k|x| + b_k$$

with $a_k, b_k, c_k, d_k \geqslant 0$ and

$$0 < \sum_{k=1}^{d} a_k \theta_k < 1.$$

Moreover, we suppose that either one can find $1 \leqslant k \leqslant d$ such that $d_k > 0$ or one can find $1 \leqslant k \leqslant d$ such that $c_k > 0$.

In this situation, and with the above standard assumptions on (ε_n), we can show after some straightforward calculations that $n = \mathcal{O}(s_n)$ and $s_n = \mathcal{O}(n)$ a.s. Next, to ensure that (3.3) still holds, we are led to introduce more assumptions on (ε_n).

COROLLARY 4. *Assume that* (ε_n) *is either a Gaussian white noise or a locally generalized Gaussian martingale sequence with a constant conditional variance* σ^2, *which means that, for all* $n \geqslant 0$ *and for all* $t \geqslant 0$,

$$E[\exp(t\varepsilon_n) \mid \mathscr{F}_{n-1}] \leqslant \exp((\sigma^2 t^2)/2) \quad a.s. \tag{3.15}$$

Then, for any $p \geqslant 1$, $\Gamma_n(p)$ *is a strongly consistent estimator of* $\sigma(p)$ *and*

$$\left|\Gamma_n(p) - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^p\right|^2 = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad a.s. \tag{3.16}$$

In addition, if p *is even and if for all* $n \geqslant 1$, $E[\varepsilon_{n+1}^{p-1} \mid \mathscr{F}_n] = 0$ *a.s., then* (3.6) *is true after replacing* $\log s_n$ *by* $\log n$.

Remark 4. We refer the reader to [13], p. 257, for the standard properties of locally generalized Gaussian martingale sequences.

Proof of Corollary 4. It is not difficult to see that under the assumption (3.15)

$$\limsup_{n \to +\infty} \frac{|\varepsilon_n|}{\sqrt{2\log n}} \leqslant \sigma \quad \text{a.s.},$$

which leads to

$$\sup_{k \leqslant n} |\varepsilon_k|^2 = \mathcal{O}(\log n) \quad \text{a.s.} \tag{3.17}$$

Furthermore, we deduce from the stabilization criteria 6.2.10 of [4] together with (3.17) that

$$\sup_{k \leqslant n} |X_k|^2 = \mathcal{O}(\log n) \quad \text{a.s.}$$

It implies that $\|\Phi_n\|^2 = \mathcal{O}(\log n)$ a.s. so that $\|\Phi_n\|^2 = o(\alpha_n^{-1})$ a.s. Consequently, since

$$f_n(\alpha) \leqslant \frac{\alpha_n \|\Phi_n\|^2}{\lambda_{\min} S_n(\alpha)} \leqslant \alpha_n \|\Phi_n\|^2,$$

we find that $f_n(\alpha)$ tends to zero a.s. and Corollary 1 holds after replacing $\log s_n$ by $\log n$. ∎

APPENDIX A

This appendix is concerned with the proof of Corollary 1 and Remark 2. The proof relies on Theorem 1.

Proof of Corollary 1. For all $n \geqslant 1$, let us set

$$\pi_n = (\theta - \hat{\theta}_n)^t \Phi_n = X_{n+1} - \hat{\theta}_n^t \Phi_n - \varepsilon_{n+1}.$$

It follows immediately from (2.13) together with (3.3) that, for any $q \geqslant 1$,

$$\sum_{n=1}^{\infty} \alpha_n^q \pi_n^{2q} < \infty \text{ a.s.} \tag{A.1}$$

In addition, since s_n increases a.s. to infinity, α_n tends a.s. to zero. Afterwards, we obtain by Kronecker's lemma (see e.g. [4], p. 19) that, for any $q \geqslant 1$,

$$\sum_{k=1}^{n} \pi_k^{2q} = o(\alpha_n^{-q}) \text{ a.s.} \tag{A.2}$$

By developing the expression of $\Gamma_n(p)$, we find that

$$\left|\Gamma_n(p) - \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k^p\right| \leqslant \frac{1}{n}\sum_{l=1}^{p} C_p^l \left|\sum_{k=0}^{n-1} \pi_k^l \varepsilon_{k+1}^{p-l}\right|. \tag{A.3}$$

On the one hand, in the particular case $l = p$, by the Cauchy–Schwarz inequality, we have

$$\Big(\sum_{k=0}^{n-1} \pi_k^p\Big)^2 = \mathcal{O}\Big(\Big(\sum_{k=0}^{n-1} \pi_k^2\Big)\Big(\sum_{k=0}^{n-1} \pi_k^{2(p-1)}\Big)\Big) \text{ a.s.,}$$

so that

$$\Big(\sum_{k=0}^{n-1} \pi_k^p\Big)^2 = o(\alpha_n^{-p}) \text{ a.s.}$$

On the other hand, in the case $l < p$, we decompose the right-hand side of (A.3) into a martingale term and a rest:

$$\sum_{k=0}^{n-1} \pi_k^l \varepsilon_{k+1}^{p-l} = \sum_{k=0}^{n-1} \pi_k^l e_{k+1}(p-l) + \sum_{k=0}^{n-1} \pi_k^l \sigma_k(p-l), \tag{A.4}$$

where, for any $q \geqslant 1$, $e_{n+1}(q) = \varepsilon_{n+1}^q - \sigma_n(q)$ and $\sigma_n(q) = E[\varepsilon_{n+1}^q \,|\, \mathscr{F}_n]$. By the standard strong law of large numbers for martingales, for all $\delta > 0$ we have

$$\Big(\sum_{k=1}^{n} \pi_{k-1}^l e_{k+1}(p-l)\Big)^2 = \mathcal{O}\Big(\sum_{k=0}^{n-1} \pi_k^{2l} \Big(\log\Big(\sum_{k=0}^{n-1} \pi_k^{2l}\Big)^{1+\delta}\Big)\Big) \text{ a.s.}$$

Consequently,

$$\Big(\sum_{k=1}^{n} \pi_{k-1}^l e_{k+1}(p-l)\Big)^2 = o(\alpha_n^{-p}) \text{ a.s.}$$

In addition, we clearly obtain from (3.4) that

$$\Big|\sum_{k=0}^{n-1} \pi_k^l \sigma_k(p-l)\Big| = \mathcal{O}\Big(\sum_{k=0}^{n-1} |\pi_k|^l\Big) \text{ a.s.}$$

Therefore, if l is even, we deduce from (A.2) that

$$\Big|\sum_{k=0}^{n-1} \pi_k^l \sigma_k(p-l)\Big| = o(\alpha_n^{-l/2}) \text{ a.s.}$$

Furthermore, if l is odd and $l > 1$, it follows from the Cauchy–Schwarz inequality that

$$\Big(\sum_{k=0}^{n-1} \pi_k^l\Big)^2 = \mathcal{O}\Big(\Big(\sum_{k=0}^{n-1} \pi_k^2\Big)\Big(\sum_{k=0}^{n-1} \pi_k^{2(l-1)}\Big)\Big) = o(\alpha_n^{-l}) \text{ a.s.}$$

Finally, it follows once again from the Cauchy–Schwarz inequality that

$$\Big(\sum_{k=0}^{n-1} \pi_k\Big)^2 \leqslant n \sum_{k=0}^{n-1} \pi_k^2 \text{ a.s.},$$

which leads to

$$\Big(\sum_{k=0}^{n-1} \pi_k\Big)^2 = o(n\alpha_n^{-1}) \text{ a.s.}$$

Piecing together all these contributions, we infer from (A.3) and (A.4) that (3.5) holds. Hereafter, if p is even and if for all $n \geqslant 1$, $E[\varepsilon_{n+1}^{p-1} \mid \mathscr{F}_n] = 0$ a.s., we have immediately

$$\sum_{k=0}^{n-1} \pi_k \sigma_k(p-1) = 0 \text{ a.s.},$$

which clearly implies (3.6), and completes the proof of Corollary 1. ∎

Proof of Remark 2. We have already seen that, for all $q \geqslant 1$,

$$\sum_{n=1}^{+\infty} \alpha_n^q \big(1-f_n(\alpha)\big)^q \pi_n^{2q} < +\infty \text{ a.s.} \tag{A.5}$$

Hence, in order to obtain the convergence (A.1), it is necessary to find a lower bound for $1-f_n(\alpha)$. We know from Proposition 4.2.12 of [4] that

$$1-f_n(\alpha) = \big(1+g_n(\alpha)\big)^{-1}, \quad \text{where } g_n(\alpha) = \alpha_n \Phi_n^t S_{n-1}^{-1}(\alpha)\Phi_n. \tag{A.6}$$

Then, since $\lambda_{\min} S_{n-1}(\alpha) \geqslant 1$, we obtain

$$\alpha_n\big(1-f_n(\alpha)\big) \geqslant (\alpha_n^{-1} + \|\phi_n\|^2)^{-1}.$$

Finally, (3.7) and (A.5) imply (A.1), which is sufficient to prove Corollary 1.

APPENDIX B

This appendix is concerned with the proof of Corollary 2. Recalling that $\hat{\theta}_n - \theta = S_{n-1}^{-1}(\alpha) M_n$, $V_n = M_n^t S_{n-1}^{-1}(\alpha) M_n$ and that $\pi_n = (\theta - \hat{\theta}_n)^t \Phi_n$, we get

$$\text{(B.1)}\quad \sum_{k=1}^{n} \alpha_k f_k(\alpha)\big((\hat{\theta}_k - \theta)^t S_k(\alpha)(\hat{\theta}_k - \theta)\big)^p = \sum_{k=1}^{n} \alpha_k f_k(\alpha)(V_k + \alpha_k \pi_k^2)^p$$
$$= \sum_{q=0}^{p} C_p^q \sum_{k=1}^{n} f_k(\alpha)\alpha_k^{q+1}\pi_k^{2q} V_k^{p-q}.$$

According to the proof of Theorem 1, V_n is a.s. bounded. Consequently,

$$\sum_{k=1}^{n} f_k(\alpha)\alpha_k^{q+1}\pi_k^{2q} V_k^{p-q} = \mathcal{O}\Big(\sum_{k=1}^{n} f_k(\alpha)\alpha_k^{q+1}\pi_k^{2q}\Big) \text{ a.s.}$$

For all $q \geqslant 1$, as $f_n(\alpha) \leqslant 1$ and $\alpha_n \leqslant 1$, we infer from (A.1) that

$$\text{(B.2)}\quad \sum_{k=1}^{\infty} f_k(\alpha)\alpha_k^{q+1}\pi_k^{2q} < \infty \text{ a.s.}$$

By the convergence (2.2), (B.2) also holds in the particular case $q = 0$. Thus, the average convergence (3.9) follows from (B.1) and (B.2). For the second part of Corollary 2, let us set

$$\text{(B.3)}\quad q_n(\alpha) = \frac{\alpha_n}{d_n(\alpha)}\big(M_n^t S_{n-1}^{-1}(\alpha) S_n(\alpha) S_{n-1}^{-1}(\alpha) M_n\big)^p.$$

We deduce from (2.3) and (3.9) that

$$\text{(B.4)}\quad \sum_{n=1}^{\infty} \big(d_n(\alpha) - d_{n-1}(\alpha)\big) q_n(\alpha) < \infty \text{ a.s.}$$

In addition, we infer from (3.12) that

$$\text{(B.5)}\quad \lim_{n \to +\infty} \frac{d_n(\alpha)}{n^d \alpha_n^d} = \delta \text{ a.s.},$$

where $\delta = \det(L) > 0$. Using an Abel transform together with the decomposition

$$d_n(\alpha) = \delta v_n + \delta_n(\alpha), \qquad \text{where } v_n = \frac{n^d}{\max\big((\log n)^{d(1+\gamma)}, 1\big)},$$

we find that

$$\text{(B.6)}\quad \Delta_n = \sum_{k=1}^{n} \big(d_k(\alpha) - d_{k-1}(\alpha)\big) q_k(\alpha)$$
$$= d_n(\alpha) q_n(\alpha) - d_0(\alpha) q_1(\alpha) + r_n + \delta \sum_{k=1}^{n-1} v_k\big(q_k(\alpha) - q_{k+1}(\alpha)\big),$$

where

$$r_n = \sum_{k=1}^{n-1} \delta_k(\alpha)\big(q_k(\alpha)-q_{k+1}(\alpha)\big).$$

Once again by an Abel transform, we have

$$\sum_{k=1}^{n-1} v_k\big(q_k(\alpha)-q_{k+1}(\alpha)\big) = v_1\, q_1(\alpha) - v_{n-1}\, q_n(\alpha) + t_n,$$

where

$$t_n = \sum_{k=2}^{n-1} q_k(\alpha)(v_k - v_{k-1}).$$

Piecing together these two identities, we obtain

$$\text{(B.7)}\qquad \Delta_n = \delta t_n + r_n + q_n(\alpha)\big(d_n(\alpha)-\delta v_{n-1}\big) - q_1(\alpha)\big(d_0(\alpha)-\delta v_1\big).$$

On the one hand, we claim that the last terms in (B.7) are bounded. Indeed, according to (B.5) we get

$$q_n(\alpha)\big(d_n(\alpha)-\delta v_{n-1}\big) = o\big(q_n(\alpha)\, d_n(\alpha)\big) \text{ a.s.}$$

Moreover, going back to the expression of $q_n(\alpha)$, we infer from (A.6) that

$$q_n(\alpha)\, d_n(\alpha) \leqslant \alpha_n V_n^p \big(\lambda_{\max}\big(S_{n-1}^{-1/2}(\alpha)\, S_n(\alpha)\, S_{n-1}^{-1/2}(\alpha)\big)\big)^p,$$

which implies

$$q_n(\alpha)\, d_n(\alpha) \leqslant \alpha_n V_n^p \big(1+g_n(\alpha)\big)^p$$

and, consequently,

$$q_n(\alpha)\, d_n(\alpha) \leqslant \alpha_n V_n^p \big(1-f_n(\alpha)\big)^{-p}.$$

Besides, we have already seen from Remark 2 that $f_n(\alpha)$ tends to zero a.s. Consequently, $q_n(\alpha)\, d_n(\alpha) = o(1)$ a.s., which leads to

$$q_n(\alpha)\big(d_n(\alpha)-\delta v_{n-1}\big) = o(1) \text{ a.s.}$$

On the other hand, we can also prove that r_n is bounded. As a matter of fact, we have

$$r_n = \sum_{k=1}^{n-1} d_k(\alpha)\, \varrho_k\big(q_k(\alpha)-q_{k+1}(\alpha)\big) \quad \text{with } \varrho_n = 1-\delta\left(\frac{v_n}{d_n(\alpha)}\right).$$

The convergence (B.5) implies immediately that ϱ_n tends to zero a.s. Thus, we deduce from (B.3) that

$$|r_n| \leqslant \left|\sum_{k=1}^{n-1} \varrho_k\left(\alpha_k V_k^p - \frac{d_k(\alpha)}{d_{k+1}(\alpha)}\alpha_{k+1} V_{k+1}^p\right)\right|$$
$$+\sum_{j=1}^{p} C_p^j \sum_{k=1}^{n-1} |\varrho_k| \left(\alpha_k^{j+1}\, \pi_k^{2j}\, V_k^{p-j} + \frac{d_k(\alpha)}{d_{k+1}(\alpha)}\alpha_{k+1}^{j+1}\, \pi_{k+1}^{2j}\, V_{k+1}^{p-j}\right).$$

Moreover, as $d_{n+1}(\alpha)$ is a.s. equivalent to $d_n(\alpha)$, (A.1) ensures that

$$\sum_{j=1}^{p} C_p^j \sum_{k=1}^{n-1} |\varrho_k| \left(\alpha_k^{j+1} \pi_k^{2j} V_k^{p-j} + \frac{d_k(\alpha)}{d_{k+1}(\alpha)} \alpha_{k+1}^{j+1} \pi_{k+1}^{2j} V_{k+1}^{p-j} \right) = \mathcal{O}(1) \text{ a.s.}$$

Therefore, we need only to show that

$$\left| \sum_{k=1}^{n-1} \varrho_k \left(\alpha_k V_k^p - \frac{d_k(\alpha)}{d_{k+1}(\alpha)} \alpha_{k+1} V_{k+1}^p \right) \right| = \mathcal{O}(1) \text{ a.s.} \tag{B.8}$$

We have the decomposition

$$\begin{aligned} &\sum_{k=1}^{n-1} \varrho_k \left(\alpha_k V_k^p - \frac{d_k(\alpha)}{d_{k+1}(\alpha)} \alpha_{k+1} V_{k+1}^p \right) \\ &\quad = \sum_{k=2}^{n-1} \alpha_k V_k^p \left(\varrho_k - \varrho_{k-1} \frac{d_{k-1}(\alpha)}{d_k(\alpha)} \right) + \alpha_1 V_1^p \varrho_1 - \varrho_{n-1} \frac{d_{n-1}(\alpha)}{d_n(\alpha)} \alpha_n V_n^p. \end{aligned} \tag{B.9}$$

First, the last two terms on the right-hand side of (B.9) are bounded. Next, one can easily show that

$$\varrho_n - \varrho_{n-1} \frac{d_{n-1}(\alpha)}{d_n(\alpha)} = f_n(\alpha) + \frac{d_{n-1}(\alpha)}{d_n(\alpha)} (1 - \varrho_{n-1})(1 - \tau_n) \quad \text{with } \tau_n = \frac{v_n}{v_{n-1}}.$$

Using the elementary fact that the function $x/(\log x)^{1+\gamma}$ is increasing for $x \geqslant e^{1+\gamma}$, we have, for n large enough, $\tau_n > 1$. Therefore, we find that, for n large enough,

$$\left| \varrho_n - \varrho_{n-1} \frac{d_{n-1}(\alpha)}{d_n(\alpha)} \right| \leqslant f_n(\alpha) + \frac{d_{n-1}(\alpha)}{d_n(\alpha)} (1 - \varrho_{n-1})(\tau_n - 1). \tag{B.10}$$

Furthermore, since

$$\tau_n = \frac{n^d}{(\log n)^{d(1+\gamma)}} \frac{(\log(n-1))^{d(1+\gamma)}}{(n-1)^d} \leqslant \left(1 + \frac{1}{n-1}\right)^d,$$

and, for all $x \leqslant 1$, $(1+x)^d \leqslant 1 + (2^d - 1)x$, we have

$$\tau_n - 1 \leqslant \frac{2^d - 1}{n-1}. \tag{B.11}$$

We have already seen that α_n^{-1} is a.s. equivalent to $(\log n)^{1+\gamma}$ a.s. Consequently, as

$$\sum_{n=2}^{+\infty} \frac{1}{n (\log n)^{1+\gamma}} < \infty,$$

we deduce from (2.2), (B.10) and (B.11) that

$$\left| \sum_{k=2}^{n-1} \alpha_k V_k^p \left(\varrho_k - \varrho_{k-1} \frac{d_{k-1}(\alpha)}{d_k(\alpha)} \right) \right| = \mathcal{O}(1) \text{ a.s.,}$$

which leads to (B.8). Then (B.4) and (B.7) imply that

$$\left|\sum_{n=1}^{\infty} q_n(\alpha)(v_n - v_{n-1})\right| < \infty \text{ a.s.} \tag{B.12}$$

It is easy to prove that for n large enough

$$\frac{v_{n-1}}{v_n} \leqslant \left(\frac{n-1}{n}\right)^{d-1}\left(1-\frac{2}{n}\right),$$

which ensures that

$$v_n - v_{n-1} \geqslant v_n \left(\frac{n-1}{n}\right)^{d-1} \frac{1}{2n}.$$

Therefore, we infer from (B.12) that

$$\sum_{n=1}^{\infty} q_n(\alpha)\left(\frac{n-1}{n}\right)^{d-1} \frac{1}{2n} v_n < \infty,$$

which means that

$$\sum_{n=1}^{\infty} q_n(\alpha)\frac{n^{d-1}}{(\log n)^{d(1+\gamma)}} < \infty \text{ a.s.}$$

Finally, we infer from (B.5) together with (B.3) that

$$\sum_{n=1}^{\infty} \alpha_n^{p+1} n^{p-1}\left((\hat{\theta}_n - \theta)^t \frac{S_n(\alpha)}{n\alpha_n}(\hat{\theta}_n - \theta)\right)^p < \infty \text{ a.s.},$$

which, by the convergence (3.12) with L invertible, completes the proof of Corollary 2. ∎

Acknowledgements. Thanks to Bernard Bercu for his invaluable help, Bruno Portier for his careful reading of the manuscript and his advice, and also Marie Duflo for her advice.

REFERENCES

[1] B. Bercu, *Weighted estimation and tracking for ARMAX models*, SIAM J. Control Optim. 33, No. 1 (1995), pp. 89–106.

[2] B. Bercu, *On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications*, submitted for publication (2002).

[3] B. Bercu and M. Duflo, *Moindres carrés pondérés et poursuite*, Ann. Inst. H. Poincaré 28, No. 3 (1992), pp. 403–430.

[4] M. Duflo, *Random Iterative Methods*, Springer, 1997.

[5] M. Duflo, R. Senoussi and A. Touati, *Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autoregressif vectoriel*, Ann. Inst. H. Poincaré 27, No. 1 (1991), pp. 1–25.

[6] G. C. Goodwin and K. S. Sin, *Adaptative Filtering Prediction and Control*, Prentice-Hall, Englewood Cliffs, N. J., 1984.

[7] L. Guo, *Self-convergence of weighted least-squares with applications to stochastic adaptive control*, IEEE Trans. Automat. Control 41, No. 1 (1996), pp. 79–89.

[8] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[9] C. K. Ing, *A note on mean-squared prediction errors of the least squares predictors in random walk models*, J. Time Ser. Anal. 22, No. 6 (2001), pp. 711–724.

[10] T. L. Lai and C. Z. Wei, *Least-squares estimates in stochastic regression models with applications to identification and control of dynamic systems*, Ann. Statist. 10, No. 1 (1982), pp. 154–166.

[11] T. L. Lai and C. Z. Wei, *Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters*, J. Multivariate Anal. 13 (1983), pp. 1–23.

[12] R. Lipster and V. Spokoiny, *Deviation probability bound for martingales with applications to statistical estimation*, Statist. Probab. Lett. 46, No. 4 (2000), pp. 347–357.

[13] W. F. Stout, *Almost Sure Convergence*, Academic Press, 1974.

[14] C. Z. Wei, *Asymptotic properties of least-squares estimates in stochastic regression models*, Ann. Statist. 13, No. 4 (1985), pp. 1498–1508.

[15] C. Z. Wei, *Adaptative prediction by least squares predictors in stochastic regression models with applications to time series*, Ann. Statist. 15, No. 4 (1987), pp. 1667–1682.

[16] C. Z. Wei, *On predictive least squares principles*, Ann. Statist. 20, No. 1 (1992), pp. 1–42.

[17] J. F. Yao, *On least squares estimation for stable nonlinear AR processes*, Ann. Inst. Statist. Math. 52, No. 2 (2000), pp. 316–331.

Laboratoire de Mathématiques
UMR 8628 CNRS
Bâtiment 425, Université Paris-Sud
91405 Orsay Cedex, France
E-mail: peggy.cenac@math.u-psud.fr

Received on 10.12.2002