

## FAST CALIBRATION OF WEAK FARIMA MODELS

SAMIR BEN HARIZ<sup>1</sup>, ALEXANDRE BROUSTE<sup>1</sup>, YOUSSEF ESSTAFI<sup>1,\*</sup>  
AND MARIUS SOLTANE<sup>2</sup>

**Abstract.** In this paper, we investigate the asymptotic properties of Le Cam’s one-step estimator for weak Fractionally AutoRegressive Integrated Moving-Average (FARIMA) models. For these models, noises are uncorrelated but neither necessarily independent nor martingale differences errors. We show under some regularity assumptions that the one-step estimator is strongly consistent and asymptotically normal with the same asymptotic variance as the least squares estimator. We show through simulations that the proposed estimator reduces computational time compared with the least squares estimator. An application for providing remotely computed indicators for time series is proposed.

**Mathematics Subject Classification.** 62M10, 62M15, 91B84.

Received June 27, 2022. Accepted December 15, 2022.

### 1. INTRODUCTION

Time series often exhibit linear and/or nonlinear dependence. When large correlations for small lags are detected, short memory processes can suffice to model the dependence structure of the series. The ARMA processes (see *e.g.* [6]) or VARMA for the multivariate framework (see [28]) are examples of short memory processes.

However, in many scientific disciplines and many applied fields, including hydrology, climatology, economics, finance and computer science, the autocorrelations of some time series decrease very slowly. This phenomenon may be due to several factors, in particular nonstationarity and/or long-range dependence.

A large audience of mathematicians has been attracted by long memory processes and their various applications (see for instance [29–33]), Granger in economics [18], Dobrushin in physics [12] and, even earlier, Hurst in hydrology [22]. Long-range dependent processes constitute currently one of the popular areas of statistical research and occupy a central place in the time series literature (see [10, 13, 18, 21], among others).

The fractional autoregressive integrated moving-average (FARIMA, for short) model is widely used to model the long memory phenomenon. This model was first introduced by [18] and then generalized to take into account short-term fluctuations in time series by [21]. FARIMA models therefore have the advantage of jointly modeling the long memory behavior of time series and their short-term dynamics through a fractional integration

---

*Keywords and phrases:* Weak FARIMA models, Le Cam’s one-step estimator, least squares estimator, consistency, asymptotic normality.

<sup>1</sup> Le Mans Université, Institut du Risque et de l’Assurance, Laboratoire Manceau de Mathématiques, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France.

<sup>2</sup> LMAC EA 2222 Université de Technologie de Compiègne CS 60319 - 57 Avenue de Landshut 60203 Compiègne Cedex, France.

\* Corresponding author: [youssef.esstafa@univ-lemans.fr](mailto:youssef.esstafa@univ-lemans.fr)

parameter  $d$  and autoregressive and moving-average parameters respectively. Their fame is partly due to their structure similar to the one of standard ARIMA models in which the differentiation exponent  $d$  is an integer. FARIMA models are generally used with strong assumptions on the noise that limit their generality. We call strong FARIMA the standard models in which the error term is assumed to be an independent and identically distributed sequence (iid for short), and we speak about weak FARIMA models when the errors are uncorrelated but neither necessarily independent nor martingale differences. It is common in the time series literature to talk about the subclass of semi-strong FARIMA models when the associated innovation process is a semi-strong white noise, that is a stationary martingale difference. An example of semi-strong white noise is the generalized autoregressive conditional heteroscedastic (GARCH) model (see [15]). The distinction between strong, semi-strong or weak FARIMA models is therefore only a matter of noise assumptions with the following inclusions:

$$\{\text{strong FARIMA}\} \subset \{\text{semi-strong FARIMA}\} \subset \{\text{weak FARIMA}\}.$$

The independence of the noise in strong FARIMA models is often considered to be very restrictive for many time series with general nonlinear dependencies. Weak FARIMA models correct this problem by allowing the noise to contain very general nonlinear dependencies of often unidentified structures. They therefore have the great advantage of providing linear modeling to nonlinear processes.

The asymptotic theory of estimation is mainly limited to strong and semi-strong FARIMA models. Whittle's estimator (see [39]) is commonly used to estimate the parameters of FARIMA models (see *e.g.* [10, 13, 16] and [38]). The study of the asymptotic properties of this estimator is developed in the case where the errors are assumed to be independent and identically distributed and in the framework where the noise is considered to be a martingale difference (see [1–3, 20, 27, 34], among others). All this work is limited to the case where the nonlinear dependency is absent or with a well-identified structure. For example, in financial time series modeling, in order to capture conditional heteroscedasticity, it is common that innovations in FARIMA models are assumed to have a GARCH structure (see *e.g.* [1, 20]).

For weak FARIMA models, the asymptotic normality has been obtained for Whittle's estimator [36, 37] and the LSE [5]. In this paper, we propose a new calibration of weak FARIMA models based on Le Cam's one-step approach (see [25]). In Le Cam's one-step procedure, an initial guess estimator is corrected by a single step of Newton gradient descent method on loglikelihood function. We adapt in this work the Le Cam one-step procedure. Firstly, we propose the LSE on subsample as an initial estimator. Secondly, since we do not specify the distribution of the weak white noise, the single Newton step is done on the least squares functional. This estimator greatly reduces the computation time and preserves the same asymptotic properties as the LSE. One-step procedure has shown its efficiency in terms of computation time and precision for diffusion processes [17, 23], ergodic Markov chains [24], fractional Gaussian noise observed at high frequency [8] or stable noise [7] and inhomogeneous Poisson processes [9].

The paper is organized as follows. In Section 2, we introduce the model and the notations used in the sequel and we give the asymptotic properties of Le Cam's one-step estimator of the parameters of weak FARIMA models. In Section 3, we provide some numerical illustrations to show the performance of the proposed estimator on finite sample sizes. All the technical proofs are gathered in Section 4.

## 2. LE CAM'S ONE-STEP ESTIMATION OF WEAK FARIMA MODELS

In this section we present the parametrization that is used in the sequel and we study the asymptotic properties of the Le Cam one-step estimator of long memory FARIMA processes induced by uncorrelated but not independent error terms. We also recall the results on the asymptotic behavior of the least squares estimator of weak FARIMA models obtained by [5]. This estimator will be used as the initial estimator in Le Cam's one-step procedure.

### 2.1. Statement of the problem and notations

Let  $(X_t)_{t \in \mathbb{Z}}$  be a long memory second-order stationary process satisfying a weak FARIMA( $p, d_0, q$ ) representation of the form

$$a(L)(1-L)^{d_0}X_t = b(L)\epsilon_t, \quad (2.1)$$

where  $d_0 \in (0, 1/2)$  is the long memory parameter and  $(\epsilon_t)_{t \in \mathbb{Z}}$  is a sequence of uncorrelated random variables defined on some probability space  $(\Omega, \mathcal{T}, \mathbb{P})$  with zero mean and common variance  $\sigma_\epsilon^2$ . The back-shift operator is denoted  $L$ . We define  $a(L) = 1 - \sum_{i=1}^p a_i L^i$  and  $b(L) = 1 - \sum_{i=1}^q b_i L^i$  the autoregressive and the moving-average operators. These operators represent the short memory part of the model and are supposed to have all their roots outside the unit disk with no zero in common to ensure the invertibility of the model and the unique identifiability of the parameters.

The fractional difference operator  $(1-L)^{d_0}$  is defined by

$$(1-L)^{d_0} = \sum_{j \geq 0} \alpha_j(d_0) L^j,$$

where for all  $j \geq 0$ ,  $\alpha_j(d_0) = \Gamma(j-d_0) / \{\Gamma(j+1)\Gamma(-d_0)\}$  and  $\Gamma(\cdot)$  is the Gamma function. It can be shown that for large  $j$ ,  $\alpha_j(d_0) \sim j^{-d_0-1} / \Gamma(-d_0)$  (see [3]). It is therefore clear that the fractional difference operator impacts the speed of convergence to 0 of the coefficients in the AR( $\infty$ ) and MA( $\infty$ ) representations of Model (2.1) compared with standard short-memory ARMA models where this operator is absent. This loss of speed, compared to the exponential one of ARMA models, implies that the series of autocovariances of the process  $(X_t)_{t \in \mathbb{Z}}$  defined in (2.1) is not absolutely summable.

Let  $\Theta^*$  be the parameter space

$$\Theta^* = \left\{ \theta = (\theta_1, \theta_2, \dots, \theta_{p+q}) \in \mathbb{R}^{p+q}; a_\theta(z) = 1 - \sum_{i=1}^p \theta_i z^i \text{ and } b_\theta(z) = 1 - \sum_{j=1}^q \theta_{p+j} z^j \right. \\ \left. \text{have all their zeros outside the unit disk} \right\}.$$

Denote by  $\Theta$  the Cartesian product  $\Theta^* \times (0, 1/2)$ . Note that the unknown parameter of interest  $\theta_0 = (a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q, d_0)'$  belongs to the parameter space  $\Theta$ .

For all  $\theta = (\theta_1, \theta_2, \dots, \theta_{p+q}, d)' \in \Theta$ , we define  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$  as the second-order stationary process which is the solution of

$$\epsilon_t(\theta) = \sum_{j \geq 0} \alpha_j(d) X_{t-j} - \sum_{i=1}^p \theta_i \sum_{j \geq 0} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^q \theta_{p+j} \epsilon_{t-j}(\theta). \quad (2.2)$$

Note that, for all  $t \in \mathbb{Z}$ ,  $\epsilon_t(\theta_0) = \epsilon_t$  a.s. Given a realization  $X_1, \dots, X_n$  of length  $n$ ,  $\epsilon_t(\theta)$  can be approximated, for  $0 < t \leq n$ , by  $\tilde{\epsilon}_t(\theta)$  defined recursively by

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} - \sum_{i=1}^p \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^q \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta), \quad (2.3)$$

with  $\tilde{\epsilon}_t(\theta) = X_t = 0$  if  $t \leq 0$ .

As shown in Proposition 4.1 (see Sect. 4), these initial values are asymptotically negligible and in particular it holds that  $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \rightarrow 0$  almost-surely as  $t \rightarrow \infty$  uniformly in  $\theta$ . Thus the choice of the initial values has no influence on the asymptotic properties of the model parameters estimator.

Let  $\Theta_\kappa^*$  denote the compact set

$$\Theta_\kappa^* = \left\{ \theta \in \mathbb{R}^{p+q}; \text{ the roots of the polynomials } a_\theta(z) \text{ and } b_\theta(z) \text{ have modulus } \geq 1 + \kappa \right\}.$$

We define the set  $\Theta_\kappa$  as the Cartesian product  $\Theta_\kappa^* \times [d_1, d_2]$ , where  $\kappa$  is a positive constant chosen such that  $\theta_0$  belongs to  $\Theta_\kappa$  and where  $[d_1, d_2] \subset (0, 1/2)$ .

For  $n \geq 1$  and  $\theta \in \Theta$ , consider the function

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\epsilon}_t^2(\theta), \quad (2.4)$$

where  $(\tilde{\epsilon}_t(\theta))_{t \in \mathbb{Z}}$  is given in (2.3). The Le Cam one-step estimator is defined, almost-surely, by

$$\bar{\theta}_n = \theta_n^* - \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) \right\}^{-1} \frac{\partial}{\partial \theta} Q_n(\theta_n^*), \quad (2.5)$$

where  $\theta_n^*$  is the least squares estimator of parameter  $\theta_0$  calculated over the first  $m = [n^\delta]$ , with  $0 < \delta \leq 1$ , observations  $X_1, \dots, X_m$ , *i.e.*

$$\theta_n^* = \underset{\theta \in \Theta_\kappa}{\operatorname{argmin}} Q_m(\theta), \text{ where } Q_m(\theta) = \frac{1}{[n^\delta]} \sum_{t=1}^{[n^\delta]} \tilde{\epsilon}_t^2(\theta). \quad (2.6)$$

The consideration of the initial LSE on a subsample of size  $m = [n^\delta]$  greatly reduces the computation time for the estimation of the parameters in the model. Furthermore, for  $\delta > 1/2$ , a sole Fisher scoring correcting step is sufficient to reach similar asymptotic properties as the LSE on the whole sample. If  $\delta \leq 1/2$ , the one-step estimator remains consistent but similar asymptotic normality as the LSE requires multiple Fisher scoring steps.

We will also propose alternative estimators where the matrix  $\partial^2 Q_n(\theta_n^*) / \partial \theta \partial \theta'$  takes other forms (see Rem. 2.3 and Sect. 2.3).

## 2.2. Asymptotic properties

The asymptotic properties of the least squares estimator of the parameters of weak FARIMA models have been established by [5]. The authors have showed, under some regularity assumptions on the noise, the consistency and the asymptotic normality of the least squares estimator. In this subsection, we study the asymptotic behavior of the Le Cam one-step estimator  $\bar{\theta}_n$ . We show, under the same assumptions, that the estimator  $\bar{\theta}_n$  converges not only in probability but almost-surely to the true parameter  $\theta_0$  and also satisfies a central limit theorem with a similar limit variance.

To ensure the strong consistency of the Le Cam one-step estimator  $\bar{\theta}_n$ , we assume that the innovation process in (2.1) satisfies the following condition:

**(A1):** The process  $(\epsilon_t)_{t \in \mathbb{Z}}$  is strictly stationary and ergodic.

Our first main result is stated in the following theorem.

**Theorem 2.1.** *(Strong consistency). Assume that  $(X_t)_{t \in \mathbb{Z}}$  satisfies (2.1). Let  $(\bar{\theta}_n)_{n \geq 1}$  be the sequence of Le Cam's one-step estimators defined by (2.5). Under Assumption (A1), we have*

$$\bar{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0.$$

The proof of this theorem is given in Section 4.

For the asymptotic normality of the Le Cam one-step estimator, additional assumptions are required. It is necessary to assume that  $\theta_0$  is not on the boundary of the parameter space  $\Theta_\kappa$ .

**(A2):** We have  $\theta_0 \in \overset{\circ}{\Theta}_\kappa$ , where  $\overset{\circ}{\Theta}_\kappa$  denotes the interior of  $\Theta_\kappa$ .

The stationary process  $(\epsilon_t)_{t \in \mathbb{Z}}$  is not supposed to be an independent sequence. So one needs to control its dependency by means of its strong mixing coefficients  $\{\alpha_\epsilon(h)\}_{h \geq 0}$  defined by

$$\alpha_\epsilon(h) = \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+h}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where  $\mathcal{F}_{-\infty}^t = \sigma(\epsilon_u, u \leq t)$  and  $\mathcal{F}_{t+h}^\infty = \sigma(\epsilon_u, u \geq t+h)$ .

We shall need an integrability assumption on the moments of the noise  $(\epsilon_t)_{t \in \mathbb{Z}}$  and a summability condition on the strong mixing coefficients  $\{\alpha_\epsilon(h)\}_{h \geq 0}$ .

**(A3):** There exists an integer  $\tau$  such that for some  $\nu \in (0, 1]$ , we have  $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$  and  $\sum_{h=0}^\infty (h+1)^{k-2} \{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$  for  $k = 1, \dots, \tau$ .

In order to state our asymptotic normality result, we define the function

$$O_n(\theta) = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2(\theta),$$

where the sequence  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$  is given by (2.2), and we consider the following information matrices

$$I(\theta) = \lim_{n \rightarrow \infty} \text{Var} \left\{ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta) \right\} \quad \text{and} \quad J(\theta) = \lim_{n \rightarrow \infty} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta) \right] \text{ a.s.}$$

The existence of these matrices and the invertibility of  $J(\theta_0)$  are proved in Lemmas 16 and 18 in [5] for weak FARIMA.

Our second main result is given in the next theorem.

**Theorem 2.2.** (*Asymptotic normality*). Assume that  $(X_t)_{t \in \mathbb{Z}}$  satisfies (2.1). Under **(A1)**, **(A2)** and **(A3)** with  $\tau = 4$ , the sequence  $\{\sqrt{n}(\hat{\theta}_n - \theta_0)\}_{n \geq 1}$  with  $\delta > 1/2$  has a limiting centered normal distribution with covariance matrix  $\Omega := J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ .

The detailed proof of this result is postponed to Section 4.

**Remark 2.3.** The quantity  $\partial^2 Q_n(\theta_n^*) / \partial \theta \partial \theta'$  in the definition of Le Cam's one-step estimator (2.5) can be replaced by  $J(\theta_n^*)$  since the matrix  $J(\theta_0)$  has an explicit expression in the framework of FARIMA models (see Sect. 2.3).

It can also be replaced by

$$\hat{J}_n(\theta_n^*) = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta} \tilde{\epsilon}_t(\theta_n^*) \right\} \left\{ \frac{\partial}{\partial \theta'} \tilde{\epsilon}_t(\theta_n^*) \right\}.$$

This is due to the fact that  $\hat{J}_n(\cdot)$  satisfies a stochastic Lipschitz condition similar to the one in Proposition 2.5 and that  $\hat{J}_n(\theta_n^*)$  converges almost-surely to  $J(\theta_0)$  (see Lem. 4.3). The ergodic theorem and the uncorrelatedness of  $(\epsilon_t)_{t \in \mathbb{Z}}$  are behind the intuition of the construction of the estimator  $\hat{J}_n(\theta_n^*)$ . More precisely, observe that

under **(A1)**, the matrix  $J(\theta_0)$  can be rewritten as

$$\begin{aligned}
J(\theta_0) &= \lim_{n \rightarrow \infty} \left\{ \frac{2}{n} \sum_{t=1}^n \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_t(\theta_0) + \frac{2}{n} \sum_{t=1}^n \epsilon_t(\theta_0) \frac{\partial^2}{\partial \theta \partial \theta'} \epsilon_t(\theta_0) \right\} \\
&= 2\mathbb{E} \left[ \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_t(\theta_0) \right] + 2\mathbb{E} \left[ \epsilon_t(\theta_0) \frac{\partial^2}{\partial \theta \partial \theta'} \epsilon_t(\theta_0) \right] \\
&= 2\mathbb{E} \left[ \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_t(\theta_0) \right] \text{ a.s.}
\end{aligned} \tag{2.7}$$

**Remark 2.4.** Under **(A1)**, **(A2)** and **(A3)** with  $\tau = 4$ , it can be shown (see [5], Lem. 18) that the sequence  $(\mathbb{E} [H_1(\theta_0) H'_{1+k}(\theta_0)])_{k \in \mathbb{Z}}$  where, for all  $t \in \mathbb{Z}$ ,  $H_t(\theta) = 2\epsilon_t(\theta) \frac{\partial}{\partial \theta} \epsilon_t(\theta) = (2\epsilon_t(\theta) \frac{\partial}{\partial \theta_1} \epsilon_t(\theta), \dots, 2\epsilon_t(\theta) \frac{\partial}{\partial \theta_{p+q+1}} \epsilon_t(\theta))'$ , is absolutely summable. Therefore, from the stationarity of the centered process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$ , we have

$$\begin{aligned}
I(\theta_0) &= \lim_{n \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) \right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \text{Cov} (H_t(\theta_0), H_s(\theta_0)) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1-n}^{n-1} (n - |k|) \text{Cov} (H_1(\theta_0), H_{1+k}(\theta_0)) \\
&= \sum_{k=-\infty}^{\infty} \mathbb{E} [H_1(\theta_0) H'_{1+k}(\theta_0)].
\end{aligned}$$

When the noise  $(\epsilon_t)_{t \in \mathbb{Z}}$  is assumed to be an iid sequence, one can use the orthogonality of  $\epsilon_t$  with any linear combination of  $(\epsilon_s)_{s \leq t-1}$  in particular  $\partial \epsilon_t(\theta_0) / \partial \theta$  (see Sect. 4.1) to obtain

$$\begin{aligned}
I(\theta_0) &= \mathbb{E} [H_1(\theta_0) H'_1(\theta_0)] + 2 \sum_{k=1}^{\infty} \mathbb{E} [H_1(\theta_0) H'_{1+k}(\theta_0)] \\
&= 4\mathbb{E} \left[ \epsilon_1^2 \frac{\partial}{\partial \theta} \epsilon_1(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_1(\theta_0) \right] + 4 \sum_{k=1}^{\infty} \mathbb{E} \left[ \epsilon_1 \frac{\partial}{\partial \theta} \epsilon_1(\theta_0) \epsilon_{1+k} \frac{\partial}{\partial \theta'} \epsilon_{1+k}(\theta_0) \right] \\
&= 2\sigma_\epsilon^2 J(\theta_0).
\end{aligned}$$

Thus, the asymptotic covariance matrix in the strong FARIMA case is reduced to  $\Omega_S := 2\sigma_\epsilon^2 J^{-1}(\theta_0)$ . Generally, when the noise is not an independent sequence, this simplification can not be made and we have  $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$ . The true asymptotic covariance matrix  $\Omega = J^{-1}(\theta_0) I(\theta_0) J^{-1}(\theta_0)$  obtained in the weak FARIMA framework can be very different from  $\Omega_S$ .

A key point allowing to establish the limit distribution of the one-step estimator of the parameters of the weak FARIMA model (2.1) is the fact that  $\theta \in \Theta_\kappa \rightarrow \partial^2 Q_n(\theta) / \partial \theta \partial \theta' \in \mathbb{R}^{(p+q+1) \times (p+q+1)}$  is a stochastic Lipschitz function.

**Proposition 2.5.** *Assume that  $(X_t)_{t \in \mathbb{Z}}$  satisfies (2.1). For any  $i, j \in \{1, \dots, p+q+1\}$  and all  $\theta^{(1)}, \theta^{(2)} \in \Theta_\kappa$ , one has*

$$\left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta^{(1)}) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta^{(2)}) \right| \leq \Delta_n \|\theta^{(1)} - \theta^{(2)}\|,$$

where  $\Delta_n$  is bounded in probability.

The proof of this proposition is detailed in Section 4.

### 2.3. Explicit computations of $J(\theta)$

The particular structure of FARIMA models allows an explicit calculation of the matrix  $J(\theta_0)$ . Thus, the use of the closed form of the limit matrix  $J(\theta_0)$  in (2.5) instead of the second derivative of the function  $Q_n(\cdot)$  further improves the computational performance of Le Cam's one-step estimator while maintaining the same asymptotic properties. The matrix  $J(\theta)$  clearly depends on the derivative of the process  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$ . This derivative can be expressed as an infinite linear combination of the past and present values of the true noise  $(\epsilon_t)_{t \in \mathbb{Z}}$ , which subsequently gives rise to a simple calculation of  $J(\theta_0)$  by exclusively exploiting the uncorrelatedness of the innovations process  $(\epsilon_t)_{t \in \mathbb{Z}}$ . Let us be more precise. The assumptions made on the autoregressive polynomials  $a_\theta$  and the moving average polynomials  $b_\theta$  imply that the power series  $a_\theta^{-1}$  and  $b_\theta^{-1}$  are well defined. Hence, the function  $\epsilon_t(\theta)$  satisfies

$$\begin{aligned} \epsilon_t(\theta) &= b_\theta^{-1}(L)a_\theta(L)(1-L)^d X_t \\ &= b_\theta^{-1}(L)a_\theta(L)(1-L)^{d-d_0} a_{\theta_0}^{-1}(L)b_{\theta_0}(L)\epsilon_t \\ &= \sum_{i=0}^{\infty} \lambda_i(\theta) \epsilon_{t-i}, \end{aligned}$$

where  $\lambda(\theta) = (\lambda_i(\theta))_{i \geq 0}$  is given by the sequence of the coefficients of the power series  $b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0} a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$ . Similarly, one can obtain

$$\frac{\partial \epsilon_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}(\theta) \epsilon_{t-i} \quad (2.8)$$

for all  $k \in \{1, \dots, p+q+1\}$ , where  $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i \geq 1}$  is the sequence of the coefficients in the power series

$$\frac{\partial}{\partial \theta_k} (b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0} a_{\theta_0}^{-1}(z)b_{\theta_0}(z)). \quad (2.9)$$

It then follows from equations (2.7) and (2.8) that

$$J(\theta_0) = 2\sigma_\epsilon^2 \sum_{i \geq 1} \begin{pmatrix} \dot{\lambda}_{i,1}^2(\theta_0) & \dot{\lambda}_{i,1}(\theta_0) \dot{\lambda}_{i,2}(\theta_0) & \cdots & \dot{\lambda}_{i,1}(\theta_0) \dot{\lambda}_{i,p+q+1}(\theta_0) \\ \dot{\lambda}_{i,2}(\theta_0) \dot{\lambda}_{i,1}(\theta_0) & \dot{\lambda}_{i,2}^2(\theta_0) & \cdots & \dot{\lambda}_{i,2}(\theta_0) \dot{\lambda}_{i,p+q+1}(\theta_0) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{\lambda}_{i,p+q+1}(\theta_0) \dot{\lambda}_{i,1}(\theta_0) & \dot{\lambda}_{i,p+q+1}(\theta_0) \dot{\lambda}_{i,2}(\theta_0) & \cdots & \dot{\lambda}_{i,p+q+1}^2(\theta_0) \end{pmatrix}.$$

Furthermore,  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient in (2.9) taken  $\theta = \theta_0$ . There are three cases.

◇  $k = 1, \dots, p$ :

Since

$$\frac{\partial}{\partial \theta_k} (b_{\theta}^{-1}(z)a_{\theta}(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)) = -b_{\theta}^{-1}(z)z^k(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z),$$

we deduce that  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient of  $-z^k a_{\theta_0}^{-1}(z)$ .

◇  $k = p + 1, \dots, p + q$ :

We have

$$\frac{\partial}{\partial \theta_k} (b_{\theta}^{-1}(z)a_{\theta}(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)) = \left( \frac{\partial}{\partial \theta_k} b_{\theta}^{-1}(z) \right) a_{\theta}(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$$

and consequently  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ th coefficient of  $(\frac{\partial}{\partial \theta_k} b_{\theta}^{-1}(z))b_{\theta_0}(z)$ .

◇  $k = p + q + 1$ :

In this case,  $\theta_k = d$  and so we have

$$\frac{\partial}{\partial \theta_k} (b_{\theta}^{-1}(z)a_{\theta}(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)) = b_{\theta}^{-1}(z)a_{\theta}(z)\ln(1-z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$$

which implies that  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient of  $\ln(1-z)$  which is equal to  $-1/i$ .

**Example 2.6.** In this example, we illustrate the previous calculations in the case of weak FARIMA(1,  $d$ , 1) model (*i.e.* when  $p = q = 1$  in (2.1) and (2.2)). This model is widely used in practice. Since the modulus of the autoregressive parameter  $a_0$  and the moving-average parameter  $b_0$  are assumed to be strictly less than 1, one can easily obtain that  $a^{-1}(z) = (1 - a_0 z)^{-1} = \sum_{i \geq 0} a_0^i z^i$  and similarly  $b^{-1}(z) = \sum_{i \geq 0} b_0^i z^i$ . So, it can be shown that

$$\frac{\partial}{\partial \theta} \epsilon_t(\theta_0) = \begin{pmatrix} \frac{\partial}{\partial a} \epsilon_t(\theta_0) \\ \frac{\partial}{\partial b} \epsilon_t(\theta_0) \\ \frac{\partial}{\partial d} \epsilon_t(\theta_0) \end{pmatrix} = - \sum_{i \geq 1} \begin{pmatrix} a_0^{i-1} \\ -b_0^{i-1} \\ \frac{1}{i} \end{pmatrix} \epsilon_{t-i}$$

and consequently, we deduce that

$$J(\theta_0) = 2\sigma_{\epsilon}^2 \begin{pmatrix} 1/(1-a_0^2) & -1/(1-a_0 b_0) & -\ln(1-a_0)/a_0 \\ -1/(1-a_0 b_0) & 1/(1-b_0^2) & \ln(1-b_0)/b_0 \\ -\ln(1-a_0)/a_0 & \ln(1-b_0)/b_0 & \pi^2/6 \end{pmatrix}.$$

This explicit expression can be used in the one-step procedure to speed it up.

### 3. NUMERICAL ILLUSTRATIONS

We investigate in this section the behavior of the one-step estimator on finite sample sizes through Monte Carlo experiments. The numerical illustrations of this section are made with the open source statistical software R [35].

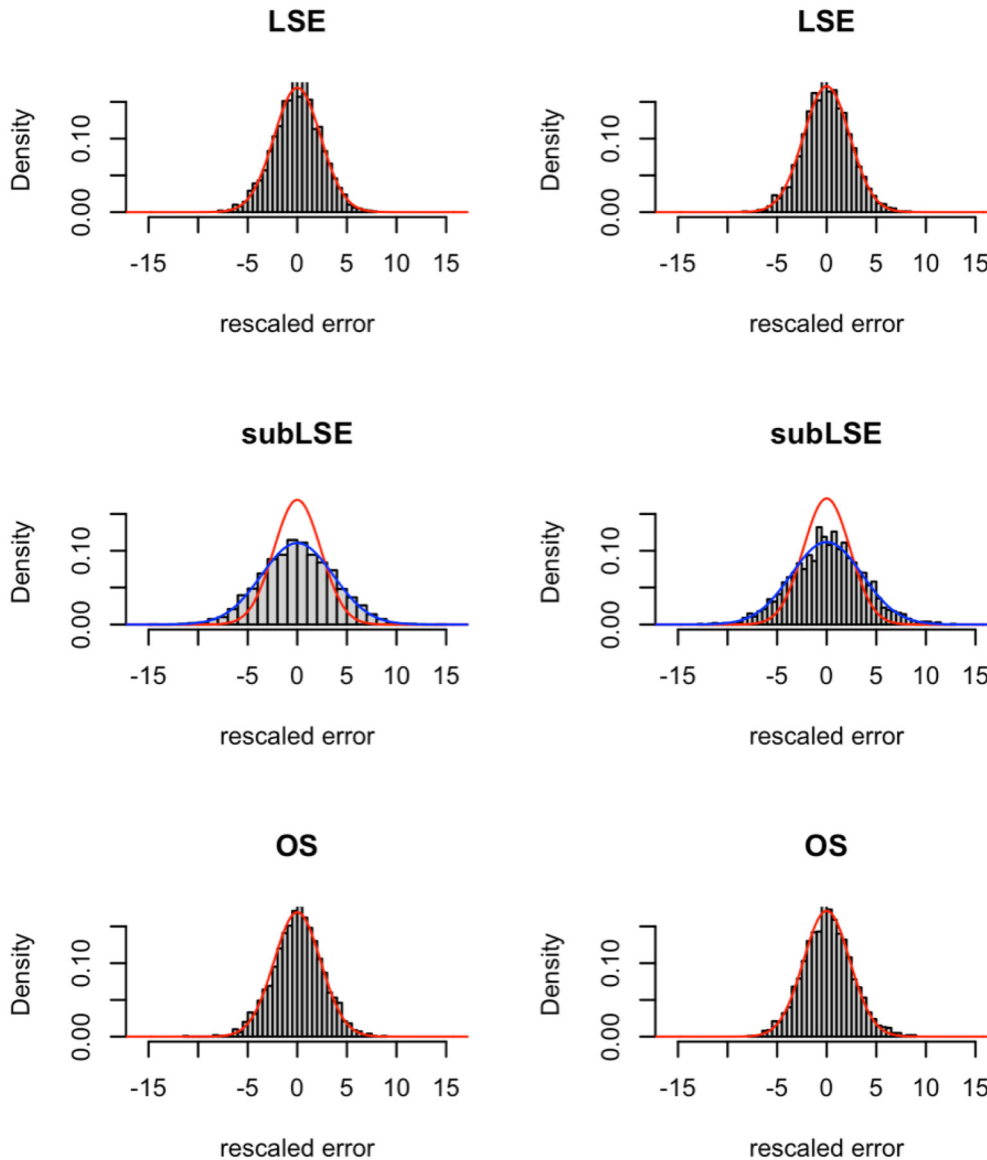


FIGURE 1. Histograms for the  $M = 2,000$  Monte Carlo simulations of the rescaled statistical error of the LSE (first line), the LSE calculated on the fraction  $\delta = 0.9$  of the data (second line) and one-step estimator (last line) of the memory parameter for Model (3.1) with  $(a, b, d)' = (0.2, 0.5, 0.3)'$  and  $n = 5,000$ . Superimposed red and blue lines are the theoretical centered Gaussian asymptotic distributions of the LSE and the subLSE, respectively.

### 3.1. Simulation studies

The behavior of Le Cam's one-step estimator is numerically studied for FARIMA(1,  $d$ , 1) model of the form

$$(1 - L)^d(X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}, \quad (3.1)$$

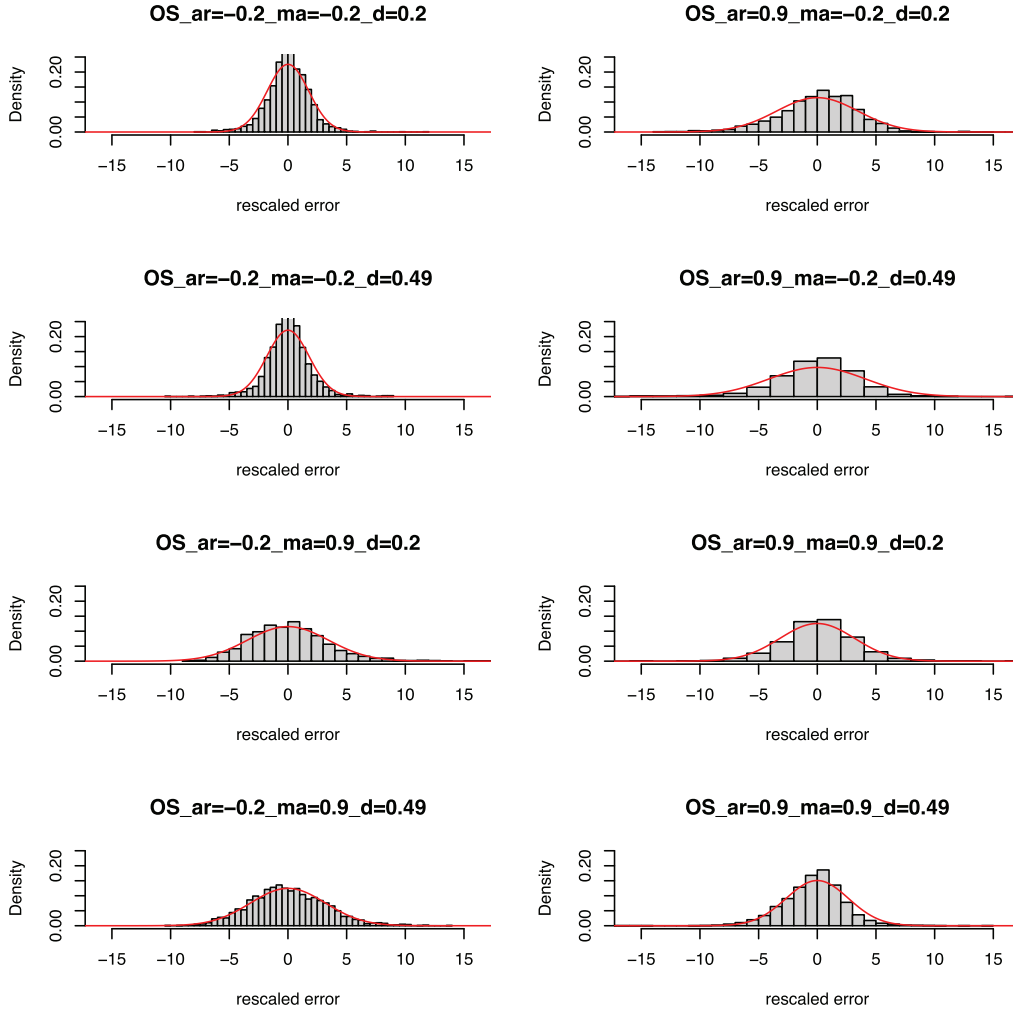


FIGURE 2. Empirical distribution over  $M = 2000$  simulations of the rescaled statistical error of the one-step estimator of the memory parameter for FARIMA(1,  $d$ , 1) model induced by the weak white noise (3.2) with  $n = 5000$  and different values of the parameters.

where the unknown parameter  $(a, b, d)'$  takes different values. We start by comparing the asymptotic properties of the one-step estimator and the LSE in both strong and weak frameworks. For this, firstly we consider that the innovation process  $(\epsilon_t)_{t \in \mathbb{Z}}$  in (3.1) is an iid centered Gaussian process with common variance 1 (which corresponds to the strong FARIMA case) and secondly that it is defined by

$$\epsilon_t = \eta_t^2 \eta_{t-1}, \quad (3.2)$$

where  $(\eta_t)_{t \in \mathbb{Z}}$  is a sequence of iid centered Gaussian random variables with variance 1. Note that the innovation process in (3.2) is a weak white noise which is not a martingale difference.

Figure 1 compares the empirical distributions of the LSE and of the Le Cam one-step estimator of the memory parameter  $d$  in the strong case (first column) and in the weak case (second column). We simulated  $M = 2,000$  independent trajectories of size  $n = 5,000$  of Model (3.1) with  $(a, b, d)' = (0.2, 0.5, 0.3)'$  endowed first by the

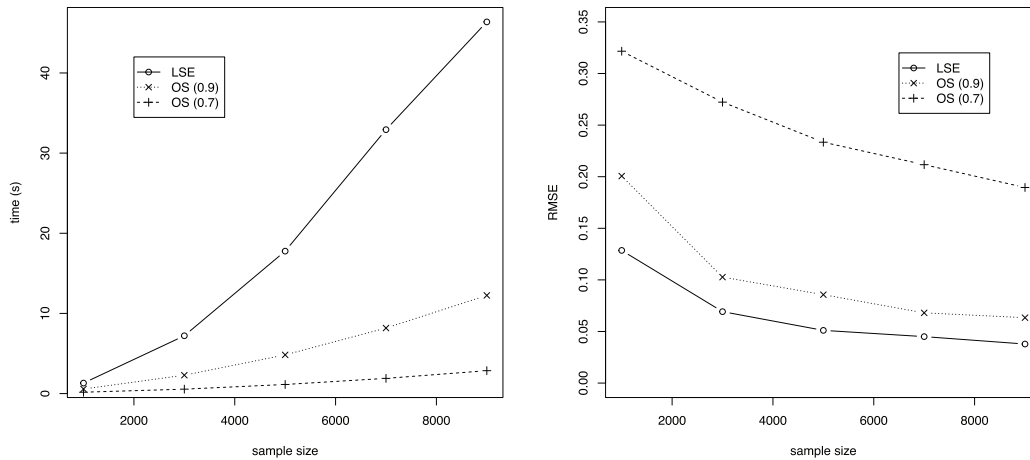


FIGURE 3. Comparison of the computation times (on the left) and RMSE (on the right) with respect to the sample size of the LSE and the one-step estimators of the parameters of Model (3.1) induced by Noise (3.2) with  $(a, b, d)' = (0.2, 0.5, 0.3)'$ . For each size  $n$ , 1000 replications are generated.

strong Gaussian noise and then by the weak noise (3.2). We considered that  $\delta = 0.9$ . Let us recall that the fraction  $\delta$  defines the size of the sample on which the initial estimator is calculated.

The LSE calculated on the fraction  $\delta$  of the data (see the two middle graphs) is computed faster than the LSE on the whole sample but is naturally less efficient. We can observe the similarity of the empirical distributions of the one-step estimator and the LSE on the whole sample. This perfectly illustrates the theoretical results presented in Section 2.2.

In Figure 2, we present the empirical distribution of the Le Cam one-step estimator of the memory parameter for different values of the parameters in (3.1) induced by the noise (3.2) with  $n = 5,000$  over the 2,000 replications. This graph highlights the adequacy of the empirical results (distributions over a finite sample size) and the theoretical results obtained in Theorem 2.2, even when the parameter is close to the boundary of the parameter space.

Finally, we compare in Figure 3 the computation time (in seconds) and the root mean square error (RMSE), with respect to the sample size, of the LSE and the one-step estimator of all the parameters, with two different fractions ( $\delta = 0.7$  and  $\delta = 0.9$ ) and  $(a, b, d)' = (0.2, 0.5, 0.3)'$ . For each size  $n$ , we simulated 1,000 replications to calculate the estimators. We observe that the one-step estimator outperforms the LSE in terms of computation time. It should also be noted that taking a small fraction  $\delta$  further reduces the calculation time with a slight degradation of the corresponding RMSE.

### 3.2. Illustrative example

We consider in this example the daily log returns of the Standard & Poor's 500 index called in the sequel by S&P 500 for short (see Fig. 4). The log returns (or simply returns) are defined by  $r_t = 100 \log(p_t/p_{t-1})$  where  $p_t$  is the price index of the S&P 500 at time  $t$ . The observations of the S&P 500 index cover the period from January 3, 1950 to February 14, 2019. The length of the series is  $n = 17,391$ . The data can be downloaded with the R package `quantmod`.

The phenomenon of long memory has been widely studied for financial series. [11] have shown that the positive powers of the absolute value of returns have more persistence than the returns themselves. We choose here the case of squared returns. The mean and the standard deviation of  $(r_t^2)_{t \geq 1}$  are 0,9347 and 5,0036. As in [26], we consider the centered series  $(X_t)_{t \geq 1}$  of the squared returns, that is,  $X_t = r_t^2 - 0,9347$ . The sample

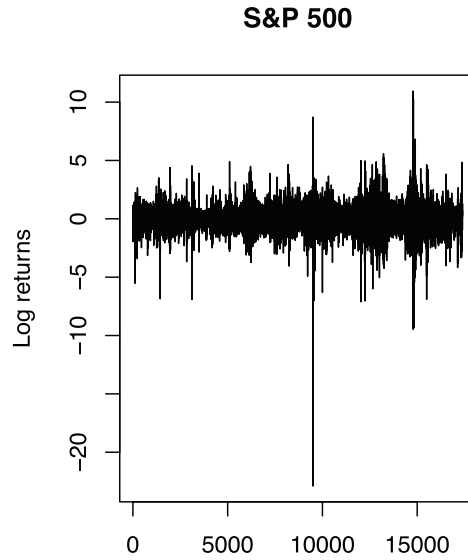


FIGURE 4. Returns of the S&P 500 index from January 3, 1950 to February 14, 2019.

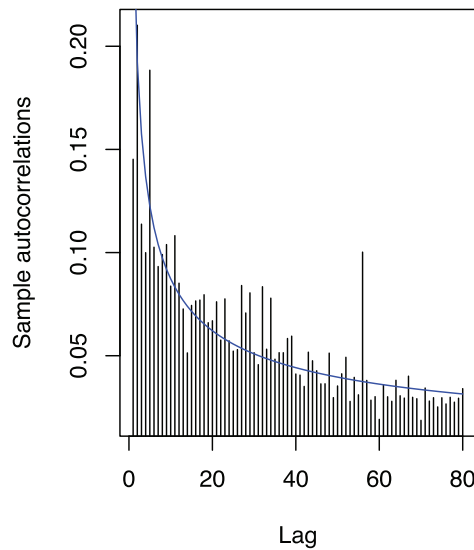


FIGURE 5. Sample autocorrelations of squared returns of the S&P 500 stock market index. The curve in blue is that of  $x \rightarrow 0.26/x^{0.49}$ .

autocorrelations of the series  $(X_t)_{t \geq 1}$  (see Fig. 5) decrease very slowly and are approximated by the function in blue (which is not integrable on  $\mathbb{R}$ ). This suggests that this series has a long memory.

It has been statistically validated in our previous work (see [4]) that the time series  $(X_t)_{t \geq 1}$  could be adjusted by a weak FARIMA(1,  $d$ , 1) model. It is worth emphasizing that the strong FARIMA fitting with the same orders is rejected for this series. The calibration of such models with the LSE is time consuming. Consequently, we propose a remote solution (API REST with the HTTP protocol) to provide fast and similar estimates with the same asymptotic properties as the LSE. In R, we can access to this API with the `httr` package using the `POST`

command and format our data into JSON (package `jsonlite`). We can find an example of a program at the address [https://www.ffi-stats.fr/files/API\\_wFARIMA.R](https://www.ffi-stats.fr/files/API_wFARIMA.R).

## 4. PROOFS

Before starting the proofs of our main results, we introduce in the next subsection some necessary results on estimations of the coefficients of formal power series that will arise in our study. In all our proofs,  $K$  is a positive constant that may vary from line to line.

### 4.1. Preliminary results

For the true value of the parameter, we have

$$a_{\theta_0}(L)(1-L)^{d_0}X_t = b_{\theta_0}(L)\epsilon_t. \quad (4.1)$$

The power series  $a_\theta^{-1}$  and  $b_\theta^{-1}$  are well defined. Thus, the functions  $\epsilon_t(\theta)$  given in (2.2) can be written as

$$\epsilon_t(\theta) = b_\theta^{-1}(L)a_\theta(L)(1-L)^d X_t \quad (4.2)$$

$$= b_\theta^{-1}(L)a_\theta(L)(1-L)^{d-d_0}a_{\theta_0}^{-1}(L)b_{\theta_0}(L)\epsilon_t \quad (4.3)$$

and if we denote  $\gamma(\theta) = (\gamma_i(\theta))_{i \geq 0}$  the sequence of coefficients of the power series  $b_\theta^{-1}(z)a_\theta(z)(1-z)^d$ , we may write

$$\epsilon_t(\theta) = \sum_{i \geq 0} \gamma_i(\theta) X_{t-i}. \quad (4.4)$$

Similarly, in view of (4.2), one has

$$X_t = (1-L)^{-d}a_\theta^{-1}(L)b_\theta(L)\epsilon_t(\theta)$$

and if we denote  $\eta(\theta) = (\eta_i(\theta))_{i \geq 0}$  the coefficients of the power series  $(1-z)^{-d}a_\theta^{-1}(z)b_\theta(z)$ , we obtain

$$X_t = \sum_{i \geq 0} \eta_i(\theta)\epsilon_{t-i}(\theta). \quad (4.5)$$

For large  $j$ , [19] have shown that the sequences  $\gamma(\theta)$  and  $\eta(\theta)$  satisfy

$$\frac{\partial^k \gamma_j(\theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} = O\left(j^{-1-d} \{\log(j)\}^k\right), \text{ for } k = 0, 1, 2, 3, \quad (4.6)$$

and

$$\frac{\partial^k \eta_j(\theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} = O\left(j^{-1+d} \{\log(j)\}^k\right), \text{ for } k = 0, 1, 2, 3. \quad (4.7)$$

Similar calculations can be provided to obtain

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{t-1} \gamma_i(\theta) X_{t-i} = \sum_{i \geq 0} \gamma_i^t(\theta) X_{t-i}, \quad (4.8)$$

where the truncated sequence  $\gamma^t(\theta) = (\gamma_i^t(\theta))_{i \geq 0}$  is given by

$$\gamma_i^t(\theta) = \begin{cases} \gamma_i(\theta) & \text{if } 0 \leq i \leq t-1, \\ 0 & \text{otherwise.} \end{cases}$$

In the following proposition, we show that the difference between  $\epsilon_t(\theta)$  and  $\tilde{\epsilon}_t(\theta)$  converges almost-surely to 0 as  $t \rightarrow \infty$  and this uniformly in  $\theta$ . This proposition shows that the convergence of the least squares estimator  $\theta_n^*$  in (2.6) studied in [5] is not only in probability but also almost-sure when  $d_0 \in (0, 1/2)$ . This last affirmation can be easily demonstrated by following line by line the proof of Theorem 1 in [14].

**Proposition 4.1.** *Let  $(X_t)_{t \in \mathbb{Z}}$  be the second-order stationary process given by (2.1). Under the standard assumptions of invertibility and identifiability on the autoregressive polynomial  $a$  and the moving-average polynomial  $b$ , we have almost-surely*

$$\lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_\kappa} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| = 0. \quad (4.9)$$

*Proof.* In view of (4.4) and (4.8), it can be shown that for all  $\theta \in \Theta_\kappa$  and any  $t \in \mathbb{Z}$ ,

$$\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) = \sum_{j \geq 0} \gamma_j(\theta) X_{t-j} - \sum_{j=0}^{t-1} \gamma_j(\theta) X_{t-j} = \sum_{j \geq t} \gamma_j(\theta) X_{t-j} = \sum_{k \geq 0} \gamma_{t+k}(\theta) X_{-k}.$$

The random variable  $\sup_{\theta \in \Theta_\kappa} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)|$  converges almost-surely to 0 as soon as

$$\sup_{k \geq t} \sup_{\theta \in \Theta_\kappa} |\epsilon_k(\theta) - \tilde{\epsilon}_k(\theta)|$$

converges in probability to 0. For any  $\beta > 0$  and for large  $t$ , one can use (4.6) to get

$$\begin{aligned} \mathbb{P} \left( \sup_{k \geq t} \sup_{\theta \in \Theta_\kappa} |\epsilon_k(\theta) - \tilde{\epsilon}_k(\theta)| > \beta \right) &= \mathbb{P} \left( \sup_{k \geq t} \sup_{\theta \in \Theta_\kappa} \left| \sum_{j \geq 0} \gamma_{k+j}(\theta) X_{-j} \right| > \beta \right) \\ &\leq \mathbb{P} \left( \sum_{j \geq 0} \sup_{k \geq t} \sup_{\theta \in \Theta_\kappa} |\gamma_{k+j}(\theta)| |X_{-j}| > \beta \right) \\ &\leq \frac{K}{\beta} \left( \sup_{t \in \mathbb{Z}} \mathbb{E} |X_t| \right) \sum_{j \geq 0} \left( \frac{1}{t+j} \right)^{1+d_1} \\ &\leq \frac{K \text{Var}(X_1)}{\beta d_1} (t-1)^{-d_1} \xrightarrow[t \rightarrow \infty]{} 0, \end{aligned}$$

which completes the proof of the convergence in (4.9).  $\square$

**Remark 4.2.** Since, for large  $j$ ,  $\partial \gamma_j(\theta) / \partial \theta_{k_1} = O(j^{-1-d} \log(j))$  and  $\partial^2 \gamma_j(\theta) / \partial \theta_{k_1} \partial \theta_{k_2} = O(j^{-1-d} \{\log(j)\}^2)$ , this last proposition remains valid for the first and second derivatives of  $\epsilon_t(\theta)$ . Following the argument developed in the proof of Proposition 4.1, we have, almost-surely and for any  $i, j \in \{1, \dots, p+q+1\}$ ,

$$\lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial \epsilon_t(\theta)}{\partial \theta_i} - \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta_i} \right| = 0$$

and

$$\lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial^2 \epsilon_t(\theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial \theta_i \partial \theta_j} \right| = 0.$$

#### 4.2. Proof of Theorem 2.1

We use (2.5) and a Taylor expansion of the function  $\partial Q_n(\cdot)/\partial \theta$  around  $\theta_0$  to obtain

$$\bar{\theta}_n - \theta_0 = (\theta_n^* - \theta_0) - \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) \right\}^{-1} \left\{ \frac{\partial}{\partial \theta} Q_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\tilde{\theta}_{n,i,j}) \right] (\theta_n^* - \theta_0) \right\}, \quad (4.10)$$

where the  $\tilde{\theta}_{n,i,j}$ 's are between  $\theta_n^*$  and  $\theta_0$ . In the two following lemmas, we show respectively the almost-sure convergence of  $\partial^2 Q_n(\theta_n^*)/\partial \theta \partial \theta'$  to  $J(\theta_0)$  and that of  $\partial Q_n(\theta_0)/\partial \theta$  to 0.

**Lemma 4.3.** *Under the assumptions of Theorem 2.1, we have almost-surely*

$$\lim_{n \rightarrow \infty} \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) = J(\theta_0).$$

The proof of this lemma is similar to the proof of Lemma 17 in [5], so it is omitted.

**Lemma 4.4.** *Under the assumptions of Theorem 2.1, we have almost-surely*

$$\lim_{n \rightarrow \infty} \frac{\partial}{\partial \theta} Q_n(\theta_0) = 0.$$

*Proof.* Observe first that, for any  $i \in \{1, \dots, p+q+1\}$ ,

$$\frac{\partial}{\partial \theta_i} Q_n(\theta_0) = \frac{2}{n} \sum_{t=1}^n \left( \tilde{\epsilon}_t(\theta_0) \frac{\partial \tilde{\epsilon}_t(\theta_0)}{\partial \theta_i} - \epsilon_t \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_i} \right) + \frac{2}{n} \sum_{t=1}^n \epsilon_t \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_i}. \quad (4.11)$$

Since  $\epsilon_t$  and  $\partial \epsilon_t(\theta_0)/\partial \theta_i$  are uncorrelated, the ergodic theorem in (A1) implies that the second term in the right hand side of (4.11) converges almost-surely to 0 as  $n \rightarrow \infty$ .

Note now that

$$\tilde{\epsilon}_t(\theta_0) \frac{\partial \tilde{\epsilon}_t(\theta_0)}{\partial \theta_i} - \epsilon_t \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_i} = (\tilde{\epsilon}_t(\theta_0) - \epsilon_t) \frac{\partial \tilde{\epsilon}_t(\theta_0)}{\partial \theta_i} + \epsilon_t \left( \frac{\partial \tilde{\epsilon}_t(\theta_0)}{\partial \theta_i} - \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_i} \right).$$

We use Proposition 4.1 and Remark 4.2 to complete the proof of the lemma. □

The proof of the theorem follows from Proposition 4.1, Lemma 16 of [5] and Lemmas 4.3 and 4.4.

#### 4.3. Proof of Theorem 2.2

In view of (2.5) and by Taylor expansion of the function  $\partial Q_n(\cdot)/\partial \theta$  around  $\theta_0$ , we have

$$\sqrt{n}(\bar{\theta}_n - \theta_0) = \sqrt{n}(\theta_n^* - \theta_0) - \sqrt{n} \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) \right\}^{-1} \left\{ \frac{\partial}{\partial \theta} Q_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\tilde{\theta}_{n,i,j}) \right] (\theta_n^* - \theta_0) \right\},$$

where the  $\tilde{\theta}_{n,i,j}$ 's are between  $\theta_n^*$  and  $\theta_0$ . Hence, it follows that

$$\begin{aligned} \sqrt{n}(\bar{\theta}_n - \theta_0) &= \left\{ \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_n^*) \right\}^{-1} n^{\delta/2} \left\{ \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_n^*) - \left[ \frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n(\tilde{\theta}_{n,i,j}) \right] \right\} n^{\delta/2} (\theta_n^* - \theta_0) n^{1/2-\delta} \\ &\quad - \left\{ \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_n^*) \right\}^{-1} \sqrt{n} \frac{\partial}{\partial\theta} Q_n(\theta_0). \end{aligned} \quad (4.12)$$

Thanks to Lemmas 16–19 of [5], the second term on the right hand side of equation (4.12) converges in law to a centered Gaussian distribution with variance matrix  $J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ . We show that the first term converges in probability to 0. Indeed, the quantity  $n^{\delta/2}(\theta_n^* - \theta_0)$  is bounded in probability due to the  $n^{\delta/2}$ -consistency of the initial estimator. The random matrix  $n^{\delta/2}\{\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_n^*) - [\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n(\tilde{\theta}_{n,i,j})]\}$  is also bounded in probability due to Proposition 2.5 and the aforementioned property of the initial estimator. The normalization rate  $n^{1/2-\delta}$  makes it converge to 0 for  $\delta > 1/2$ . Slutsky's theorem allows to complete the proof.

#### 4.4. Proof of Proposition 2.5

For any  $\theta^{(1)}, \theta^{(2)} \in \Theta$ , the mean value theorem gives

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n(\theta^{(1)}) - \frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n(\theta^{(2)}) = \frac{\partial^3}{\partial\theta\partial\theta_i\partial\theta_j} Q_n((1-c)\theta^{(1)} + c\theta^{(2)}) \cdot \{\theta^{(1)} - \theta^{(2)}\},$$

with some  $c$  between 0 and 1.

In view of (2.4) and for all  $\theta \in \Theta$ , a simple calculation of derivative leads to

$$\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} Q_n(\theta) = T_{1,n}(\theta) + T_{2,n}(\theta) + T_{3,n}(\theta) + T_{4,n}(\theta),$$

where

$$\begin{aligned} T_{1,n}(\theta) &= \frac{2}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial\theta_i\partial\theta_j} \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial\theta_k}, \\ T_{2,n}(\theta) &= \frac{2}{n} \sum_{t=1}^n \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial\theta_j} \frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial\theta_i\partial\theta_k}, \\ T_{3,n}(\theta) &= \frac{2}{n} \sum_{t=1}^n \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial\theta_i} \frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial\theta_j\partial\theta_k}, \end{aligned}$$

and

$$T_{4,n}(\theta) = \frac{2}{n} \sum_{t=1}^n \tilde{\epsilon}_t(\theta) \frac{\partial^3 \tilde{\epsilon}_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}.$$

We use equations (4.8) and (4.6) to obtain

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\theta \in \Theta_\kappa} |T_{1,n}(\theta)| \right] &\leq \frac{2}{n} \sum_{t=1}^n \mathbb{E} \left[ \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial \theta_i \partial \theta_j} \right| \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta_k} \right| \right] \\
&= \frac{2}{n} \sum_{t=1}^n \mathbb{E} \left[ \sup_{\theta \in \Theta_\kappa} \left| \sum_{\ell=1}^{\infty} \frac{\partial^2 \gamma_\ell^t(\theta)}{\partial \theta_i \partial \theta_j} X_{t-\ell} \right| \sup_{\theta \in \Theta_\kappa} \left| \sum_{\ell=1}^{\infty} \frac{\partial \gamma_\ell^t(\theta)}{\partial \theta_k} X_{t-\ell} \right| \right] \\
&\leq \frac{2}{n} \sum_{t=1}^n \sum_{\ell_1, \ell_2=1}^{t-1} \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial^2 \gamma_{\ell_1}(\theta)}{\partial \theta_i \partial \theta_j} \right| \sup_{\theta \in \Theta_\kappa} \left| \frac{\partial \gamma_{\ell_2}(\theta)}{\partial \theta_k} \right| \mathbb{E} [|X_{t-\ell_1} X_{t-\ell_2}|] \\
&\leq 2 \sup_{t \in \mathbb{Z}} \mathbb{E} [X_t^2] \left( \sum_{\ell=1}^{\infty} \ell^{-1-d_1} \{\log(\ell)\}^2 \right) \left( \sum_{\ell=1}^{\infty} \ell^{-1-d_1} \log(\ell) \right) < \infty.
\end{aligned}$$

Thanks to Markov's inequality, we deduce that

$$\sup_{\theta \in \Theta_\kappa} |T_{1,n}(\theta)| = O_{\mathbb{P}}(1).$$

Similar calculation can be done to show that  $T_{2,n}(\theta)$ ,  $T_{3,n}(\theta)$  and  $T_{4,n}(\theta)$  are bounded in probability uniformly in  $\theta$ . It follows then that

$$\sup_{\theta \in \Theta_\kappa} \left\| \frac{\partial^3}{\partial \theta \partial \theta_i \partial \theta_j} Q_n(\theta) \right\| = O_{\mathbb{P}}(1).$$

This is enough to complete the proof.

*Acknowledgements.* The authors sincerely thank the Associate Editor and the two anonymous referees for their valuable comments. This research benefited from the support of the ANR project "Efficient inference for large and high-frequency data" (ANR-21-CE40-0021), the "Chair Risques Émergents ou Atypiques en Assurance", under the aegis of Fondation du Risque, a joint initiative by Le Mans University and MMA company, member of Covea group and the "Chair Impact de la Transition Climatique en Assurance", under the aegis of Fondation du Risque, a joint initiative by Le Mans University and Groupama Centre-Manche company, member of Groupama group.

## REFERENCES

- [1] R.T. Baillie, C.-F. Chung and M.A. Tieslau, Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *J. Appl. Econometr.* **11** (1996) 23–40.
- [2] J. Beran, Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B* **57** (1995) 659–672.
- [3] J. Beran, Y. Feng, S. Ghosh and R. Kulik, Long-memory processes. Springer, Heidelberg (2013).
- [4] Y. Boubacar Maïnassara, Y. Esstafa and B. Sausseureau, Diagnostic checking in FARIMA models with uncorrelated but non-independent error terms (2021). DOI: [10.48550/ARXIV.1912.00013](https://doi.org/10.48550/ARXIV.1912.00013).
- [5] Y. Boubacar Maïnassara, Y. Esstafa and B. Sausseureau, Estimating FARIMA models with uncorrelated but non-independent error terms. *Stat. Inference Stoch. Process.* **24** (2021) 549–608.
- [6] G.E.P. Box and G.M. Jenkins, Time series analysis. Forecasting and control. Holden-Day, San Francisco, Calif.-London-Amsterdam (1970).
- [7] A. Brouste and H. Masuda, Efficient estimation of stable Lévy process with symmetric jumps. *Stat. Inference Stoch. Process.* **21** (2018) 289–307.
- [8] A. Brouste, M. Soltane and E. Votsi, One-step estimation for the fractional Gaussian noise model at high-frequency. *ESAIM: PS* **24** (2020) 827–841.
- [9] A. Dabye, A. Gounoung and Y. Kutoyants, Method of moments estimators and multi-step MLE for Poisson processes. *J. Contemp. Math. Anal.* **53** (2018) 187–196.
- [10] R. Dahlhaus, Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** (1989) 1749–1766.

- [11] Z. Ding, C.W. Granger and R.F. Engle, A long memory property of stock market returns and a new model. *J. Empir. Finance* **1** (1993) 83–106.
- [12] R.L. Dobrushin and P. Major, Non-central limit theorems for non-linear functional of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50** (1979) 27–52.
- [13] R. Fox and M.S. Taqqu, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** (1986) 517–532.
- [14] C. Francq and J.-M. Zakoïan, Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* **68** (1998) 145–165.
- [15] C. Francq and J.-M. Zakoïan, *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley (2010).
- [16] L. Giraitis and D. Surgailis, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle’s estimate. *Probab. Theory Related Fields* **86** (1990) 87–104.
- [17] A. Gloter and N. Yoshida, Adaptive estimation for degenerate diffusion processes. *Electr. J. Stat.* **15** (2021) 1424–1472.
- [18] C.W.J. Granger and R. Joyeux, An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** (1980) 15–29.
- [19] M. Hallin, M. Taniguchi, A. Serroukh and K. Choy, Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27** (1999) 2054–2080.
- [20] M. Hauser and R. Kunst, Fractionally integrated models with ARCH errors: with an application to the Swiss 1-month Euromarket interest rate. *Rev. Quant. Finance Account.* **10** (1998) 95–113.
- [21] J.R.M. Hosking, Fractional differencing. *Biometrika* **68** (1981) 165–176.
- [22] H.E. Hurst, Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Eng.* **116** (1951) 770–799.
- [23] K. Kamatani and M. Uchida, Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.* **18** (2015) 177–204.
- [24] Y. Kutoyants and A. Motrunich, On multi-step MLE-process for Markov sequences. *Metrika* **79** (2016) 705–724.
- [25] L. Le Cam, On the asymptotic theory of estimation and testing hypothesis, Proc. 3rd Berkeley Sympos. *Math. Statist. Probability* **1** (1956) 129–156.
- [26] S. Ling, Adaptive estimators and tests of stationary and nonstationary short- and long-memory ARFIMA-GARCH models. *J. Am. Statist. Assoc.* **98** (2003) 955–967.
- [27] S. Ling and W.K. Li, On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Am. Statist. Assoc.* **92** (1997) 1184–1194.
- [28] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, Berlin Heidelberg (2007).
- [29] B.B. Mandelbrot, Une classe de processus stochastiques homothétiques à soi; application à la loi climatologique de H. E. Hurst. **260** (1965) 3274–3277.
- [30] B.B. Mandelbrot and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** (1968) 422–437.
- [31] B.B. Mandelbrot and J.R. Wallis, Computer experiments with fractional Gaussian noises: Part 1. Averages and variances. *Water Resour. Res.* **5** (1969) 228–241.
- [32] B.B. Mandelbrot and J.R. Wallis, Computer experiments with fractional Gaussian noises: Part 2. Rescaled ranges and spectra. *Water Resour. Res.* **5** (1969) 242–259.
- [33] B.B. Mandelbrot and J.R. Wallis, Computer experiments with fractional Gaussian noises. Part 3: Mathematical appendix. *Water Resour. Res.* **5** (1969) 260–267.
- [34] W. Palma, *Long-memory time series*, Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ (2007), theory and methods.
- [35] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2021).
- [36] X. Shao, Nonstationarity-extended Whittle estimation. *Econ. Theory* **26** (2010) 1060–1087.
- [37] X. Shao, Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.* **39** (2012) 772–783.
- [38] M.S. Taqqu and V. Teverovsky, Robustness of Whittle-type estimators for time series with long-range dependence. *Commun. Statist. Stoch. Models* **13** (1997) 723–757.
- [39] P. Whittle, Estimation and information in stationary time series. *Ark. Mat.* **2** (1953) 423–434.

Please help to maintain this journal in open access!



This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting [subscribers@edpsciences.org](mailto:subscribers@edpsciences.org).

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.