ESTIMATION IN AUTOREGRESSIVE MODEL WITH MEASUREMENT ERROR

JÉRÔME DEDECKER$^{(1)}$, ADELINE SAMSON$^{(1)}$, MARIE-LUCE TAUPIN$^{(2)}$

Abstract. Consider an autoregressive model with measurement error: we observe $Z_i = X_i + \varepsilon_i$, where X_i is a stationary solution of the equation $X_i = f_{\theta^0}(X_{i-1}) + \xi_i$. The regression function f_{θ^0} is known up to a finite dimensional parameter θ^0. The distributions of X_0 and ξ_1 are unknown whereas the distribution of ε_1 is completely known. We want to estimate the parameter θ^0 by using the observations $Z_0, \ldots, Z_n$. We propose an estimation procedure based on a modified least square criterion involving a weight function w, to be suitably chosen. We give upper bounds for the risk of the estimator, which depend on the smoothness of the errors density f_ε and on the smoothness properties of wf_θ.

(1) Laboratoire MAP5 UMR CNRS 8145, Université Paris Descartes

(2) Laboratoire Statistique et Génome UMR CNRS 8071- USC INRA, Université d'Évry Val d'Essonne

May 6, 2011

Keywords: autoregressive model, markov chain, weak dependent, deconvolution, semi-parametric nonlinear autoregressive model.

AMS 2000 MSC: Primary 62J02, 62F12, Secondary 62G05, 62G20.

1. Introduction

We consider the autoregressive model with measurement error where one observes $Z_0, \cdots, Z_n$ satisfying

$$\left\{ \begin{array}{ll} Z_i & = X_i + \varepsilon_i, \\ X_i & = f_{\theta^0}(X_{i-1}) + \xi_i \end{array} \right. \tag{1.1}$$

involving unobserved random variables $\xi_i, X_i, \varepsilon_i$, and a regression function f_{θ^0}. This regression function is known up to a finite dimensional parameter θ^0, belonging to the interior of a compact set $\Theta \subset \mathbb{R}^d$. The centered innovations $(\xi_i)_{i\geq 1}$ and the errors $(\varepsilon_i)_{i\geq 0}$ are independent and identically distributed (i.i.d.) random variables with finite variances $\mathrm{Var}(\xi_1) = \sigma_\xi^2$ and $\mathrm{Var}(\varepsilon_0) = \sigma_\varepsilon^2$. We assume that ε_0 admits a known density with respect to the Lebesgue measure, denoted by f_ε. Furthermore we assume that the random variables X_0, $(\xi_i)_{i\geq 0}$ and $(\varepsilon_i)_{i\geq 0}$ are independent. The distribution of ξ_1 is unknown and does not necessarily admit a density with respect to the Lebesgue measure. We assume that $(X_i)_{i\geq 1}$ is strictly stationary, which means that the initial distribution of X_0 is an invariant distribution for the transition kernel of the homogeneous Markov chain $(X_i)_{i\geq 0}$.

In this context we want to estimate the finite dimensional parameter θ^0 in the presence of two functional nuisance parameters: the distribution of X_0 and the distribution of ξ_1.

Previously known results. When the function f_θ is linear (in both θ and x) let us mention the works of Andersen and Deistler (1984), Nowak (1985), Chanda (1995, 1996), Staudenmayer and Buonaccorsi (2005), and Costa *et al.* (2010). Note that, in this specific case, Model (1.1) can also be written as an ARMA model (see Section 5.3 for further details). Consequently, all previously known estimation procedures for ARMA models can be applied here (see for instance Chapter 8 in Brockwell and Davis (1991)).

For a general regression function, Model (1.1) is a particular Hidden Markov Model with possibly a non compact, continuous state space, and with unknown innovation distribution. When the innovation ξ distribution is known up to a finite dimensional parameter, the model (1.1) is fully parametric. In that case, various results are already stated. The parameters can be estimated by maximum likelihood, and consistency, asymptotic normality and efficiency have been proved. For further references on estimation in fully parametric hidden Markov models, we refer for instance to Leroux (1992), Bickel *et al.* (1998), Jensen and Petersen (1999), Douc and Matias (2001), Douc *et al.* (2004), Fuh (2006), Genon-Catalot and Laredo (2006), Na *et. al.* (2006), and Douc *et al.* (2011).

In this paper, we consider the case where the innovation distribution is unknown, and thus the model is not fully parametric.

Our results. Our aim is to estimate θ^0 for a large class of functions f_θ, whatever the known error distribution, and without the knowledge of the ξ_i's distribution. The distribution of the innovations being unknown, this model belongs to the family of semi-parametric models.

In this general context there are few results. To our knowledge, the only paper which gives a consistent estimator is the paper by Comte and Taupin (2001). The authors propose an estimation procedure based on a modified least squares minimization. The resulting estimator is a consistent estimator of θ^0. They also give an upper bound for its rate of convergence in probability that depends on the smoothness of the regression function and on the smoothness of the error distribution ε. Those results are obtained by assuming that X_0 admits a density f_X with respect to the Lebesgue measure and that the stationary Markov chain $(X_i)_{i\geq 0}$ is absolutely regular (β-mixing).

More precisely, Comte and Taupin (2001) propose an estimation criterion based on the following least squares contrast

$$\mathbb{E}\left[(Z_1 - \mathbb{E}(f_\theta(X_0)|Z_0))^2 W(Z_0)\right], \tag{1.2}$$

where W is a compactly supported weight function. The conditional expectation $\mathbb{E}(f_\theta(X_{i-1})|Z_{i-1})$ is estimated using tools related to density deconvolution. The parameter θ^0 is estimated by

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\left[n^{-1}\sum_{i=1}^{n} W(Z_{i-1})[Z_i - \widehat{\Gamma}_{f_\theta}(Z_{i-1})/\widehat{f_Z}(Z_{i-1})]^2\right],$$

where f_z is the density of Z_0 and where

$$\widehat{\Gamma}_f(Z_{i-1}) = \int f(x) f_\varepsilon(Z_{i-1} - x) \widehat{f_X}(x) dx, \quad \text{and} \quad \widehat{f_X}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{n,C_n}(x - Z_i).$$

The density deconvolution kernel K_{n,C_n} is defined via its Fourier transform K^*_{n,C_n} by

$$K^*_{n,C_n}(t) = \frac{K^*(t/C_n)}{f^*_\varepsilon(t)} := \frac{K^*_{C_n}(t)}{f^*_\varepsilon(t)}, \tag{1.3}$$

where K^* has a compact support and C_n tends to infinity with n. Their procedure has the great advantage to provide a consistent estimator in a general context. The main drawbacks are the followings. First the estimation criterion and consequently, the resulting upper bounds for the rate of convergence are not explicit. Secondly, this procedure allows to achieve the parametric rate for very few couples of regression function/error distribution and once again those couples are not often explicit. Lastly, the dependency conditions are stated under β-mixing conditions which are quite restrictive.

We propose here a new estimation procedure based on the new contrast function

$$S_{\theta^0,P_X}(\theta) = \mathbb{E}[(Z_1 - f_\theta(X_0))^2 \, w(X_0)],$$

where w is a weight function to be chosen. We estimate θ^0 by $\widehat{\theta} = \arg\min_{\theta \in \Theta} S_n(\theta)$, where

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \big((Z_i - f_\theta)^2 w\big) \star K_{n,C_n}(Z_{i-1}),$$

with $C_n \to \infty$, K_{n,C_n} is defined by (1.3), and $\star$ denotes the convolution product.

This estimation provides a consistent estimator in a very general context. We consider two different theoretical frameworks.

- We first give conditions under which the parametric rate of convergence as well as the asymptotic normality can be stated. Those results hold under weak dependency conditions as introduced in Dedecker and Prieur (2005).

- Secondly, we state a consistency result under weaker conditions on f_θ, and we give an upper bound for the quadratic risk, that relates the smoothness properties of the regression function and the ones of f_ε. These last results are stated under α-mixing conditions.

The asymptotic properties of our estimator are illustrated through a simulation study. It confirms that our estimator performs well in various contexts, even in cases where the Markov chain $(X_i)_{i \geq 0}$ is not β-mixing (and not even irreducible), when the ratio signal to noise is small or large, for various sample sizes, and for different types of errors ε distribution. Our estimator always better performs than the so-called naive estimator (built by replacing the non-observed X by Z in the usual least squares criterion). Our estimation procedure depends on the choice of a weight function. The influence of this weight function is also studied in the simulations.

This procedure is clearly simpler than that of Comte and Taupin (2001). The resulting rate is more explicit and links directly the smoothness of the regression function to that of f_ε. Our new estimator is asymptotically Gaussian for a large class of regression functions, which was not the case in Comte and Taupin (2001)

The paper is organized as follows. In Section 2, we present the dependency context, the related technical tools, the mixing properties of autoregressive models, and some examples of non irreducible Markov chains to which our results apply. In Section 3 we present our estimation procedure. The theoretical properties are stated in Section 4, followed by the simulation context and results (Sections 5 and 6). The proofs are gathered at the end of the paper.

2. Dependence properties of autoregressive models

2.1. **Measures of dependence.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let Y be a random variable with values in a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. Denote by $\Lambda_\kappa(\mathbb{B})$ the set of κ-Lipschitz functions, *i.e.* the functions f from $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ to $\mathbb{R}$ such that $|f(x)-f(y)| \leq \kappa \parallel x-y \parallel_{\mathbb{B}}$. Let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let $\mathbb{P}_{Y|\mathcal{M}}$ be a conditional distribution of Y given $\mathcal{M}$, $\mathbb{P}_Y$ the distribution of Y, and $\mathcal{B}(\mathbb{B})$ the Borel σ-algebra on $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$.

Define now the two coefficients

$$\begin{aligned} \alpha(\mathcal{M}, \sigma(Y)) &= \frac{1}{2} \sup_{A \in \mathcal{B}(\mathbb{B})} \mathbb{E}(|\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)|)\,, \\ \text{and if } \mathbb{E}(\|Y\|_{\mathbb{B}}) < \infty, \quad \tau(\mathcal{M}, Y) &= \mathbb{E}\Big(\sup_{f \in \Lambda_1(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(f) - \mathbb{P}_Y(f)| \Big)\,. \end{aligned}$$

The coefficient $\alpha(\mathcal{M}, \sigma(Y))$ is the usual strong mixing coefficient introduced by Rosenblatt (1956). The coefficient $\tau(\mathcal{M}, Y)$ has been introduced by Dedecker and Prieur (2005).

Let $\mathbf{X} = (X_i)_{i \geq 0}$ be a strictly stationary Markov chain of real-valued random variables. On $\mathbb{R}^2$, we put the norm $\|x\|_{\mathbb{R}^2} = (|x_1| + |x_2|)/2$. For any integer $k \geq 0$, the coefficients $\alpha_{\mathbf{X}}(k)$ and $\tau_{\mathbf{X},2}(k)$ of the chain are defined by

$$\begin{aligned} \alpha_{\mathbf{X}}(k) &= \alpha(\sigma(X_0), \sigma(X_k)) \\ \text{and if } \mathbb{E}(|X_0|) < \infty, \quad \tau_{\mathbf{X},2}(k) &= \sup\{\tau(\sigma(X_0), (X_{i_1}, X_{i_2})), k \leq i_1 \leq i_2\}\,. \end{aligned}$$

Conditions on ξ_0 and f_{θ^0} such that the Markov chain $(X_i)_{i \geq 0}$ is α-mixing or τ-dependent are recalled in Section 2.3.

2.2. **Covariance inequalities and coupling.** We now present the main tools related to the weak dependence, which are the key arguments in the proofs.

Let us first recall a covariance inequality due to Rio (1993). For any positive random variable Z, let Q_Z be the inverse cadlag of the tail function $t \to \mathbb{P}(Z > t)$. Let X and Y be two real valued random variables such that $\text{Cov}(X, Y)$ is well defined. For α-mixing we use the following inequality

$$|\text{Cov}(Y, X)| \leq 4 \int_0^{\alpha(\sigma(Y), \sigma(X))} Q_{|X|}(u) Q_{|Y|}(u) du\,. \tag{2.1}$$

Under weak dependence, we shall use the coupling properties of τ: enlarging Ω if necessary, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$\tau(\mathcal{M}, X) = \mathbb{E}(\|X - X^*\|_{\mathbb{B}})\,. \tag{2.2}$$

2.3. **Dependence properties of autoregressive models.** We recall here the mixing properties of the autoregressive models

$$X_i = f_{\theta^0}(X_{i-1}) + \xi_i,$$

that have been described in particular in the papers by Mokkadem (1985) and Ango-Nzé (1998). For instance, assume that

- the law of ξ_0 has a density f_ξ such that $f_\xi > c > 0$ on a neighborhood of zero, and there exists $S \geq 1$ such that $\mathbb{E}(|\xi_0|^S) < \infty$.
- f_{θ^0} is continuous and there exist $R \geq 1$ and $\rho \in]0,1[$ such that: for any $|x| \geq R$, $|f_{\theta^0}(x)| \leq \rho|x|$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\alpha_{\mathbf{X}}(n) = O(\kappa^n)$ for any $\kappa \in]\rho, 1[$. Now if the second point is weakened to

- f_{θ^0} is continuous and there exist $R \geq 1$ and $\delta \in]0,1[$ such that: for any $|x| \geq R$, $|f_{\theta^0}(x)| \leq |x|(1 - |x|^{-\delta})$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\alpha_{\mathbf{X}}(n) = O(n^{1-S/\delta})$.

Now, if we do not assume that ξ_0 has a density, then the chain may be non irreducible. However, under appropriate assumptions on f_{θ^0}, it is still possible to obtain upper bounds for the coefficient τ. For instance assume that

- there exists $S \geq 1$ such that $\mathbb{E}(|\xi_0|^S) < \infty$.
- $|f_{\theta^0}(x) - f_{\theta^0}(y)| \leq \rho|x - y|$ for some $\rho \in]0,1[$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\tau_{\mathbf{X},2}(n) = O(\rho^n)$. Now if the second point is weakened to

- there exist δ in $[0,1[$ and C in $]0,1]$ such that $|f'(t)| \leq 1 - C(1 + |t|)^{-\delta}$ almost everywhere.

Then there exists a unique invariant probability measure, and for $S > 1 + \delta$ the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\tau_{\mathbf{X},2}(n) = O(n^{(\delta+1-S)/\delta})$.

2.4. **Examples.** To conclude this section, let us give two simple examples of non-mixing autoregressive process for which $\tau_{\mathbf{X},2}(n)$ may be easily computed:

- Assume that X_0 is uniformly distributed over $[0,1]$, and let $(\xi_i)_{i\geq 1}$ be a sequence of i.i.d. random variables, independent of X_0 and such that $\mathbb{P}(\xi_1 = -1/4) = \mathbb{P}(\xi_1 = 1/4) = 1/2$. Then the Markov chain defined for $n > 0$ by

$$X_n = \frac{1}{4} + \frac{1}{2}X_{n-1} + \xi_n \tag{2.3}$$

  is strictly stationary. Moreover $\alpha_{\mathbf{X}}(n) = 1/4$ (see for instance Bradley (1986), p. 180) and $\tau_{\mathbf{X},2}(n) = O(2^{-n})$.
- Assume that X_0 is uniformly distributed over the Cantor set, and let $(\xi_i)_{i\geq 1}$ be a sequence of i.i.d. random variables, independent of X_0 and such that $\mathbb{P}(\xi_1 = -1/3) = \mathbb{P}(\xi_1 = 1/3) = 1/2$. Then the Markov chain defined for $n > 0$ by

$$X_n = \frac{1}{3} + \frac{1}{3}X_{n-1} + \xi_n \tag{2.4}$$

is strictly stationary. Moreover $\alpha_{\mathbf{X}}(n) = 1/4$ and $\tau_{\mathbf{X},2}(n) = O(3^{-n})$.

These two examples are studied by simulations in Sections 5 and 6.

3. Estimation procedure

In order to define more rigorously the criterion presented in the introduction, we first give some preliminary notations and assumptions.

3.1. **Notations.** We denote by

$$\parallel \varphi \parallel_1 = \int |\varphi(x)|dx, \ \parallel \varphi \parallel_2^2 = \int \varphi^2(x)dx, \ \text{ and } \ \parallel \varphi \parallel_\infty = \sup_{x\in\mathbb{R}} |\varphi(x)|.$$

The convolution product of two square integrable functions p and q is denoted by $p \star q(z) = \int p(z-x)q(x)dx$. For a function φ the Fourier tranform φ^* is defined by

$$\varphi^*(t) = \int e^{itx}\varphi(x)dx.$$

For $\theta \in \mathbb{R}^d$, $\parallel \theta \parallel_{\ell^2}^2 = \sum_{k=1}^d \theta_k^2$ and $\theta^\top$ is the transpose matrix of θ.

For a map $(\theta, u) \mapsto \varphi_\theta(u)$ from $\Theta \times \mathbb{R}$ to $\mathbb{R}$, the first and second derivatives with respect to θ are denoted by

$$\begin{aligned} \varphi_\theta^{(1)}(\cdot) &= \left(\varphi_{\theta,j}^{(1)}(\cdot)\right)_{1\le j\le d} \text{ with } \varphi_{\theta,j}^{(1)}(\cdot) = \frac{\partial \varphi_\theta(\cdot)}{\partial\theta_j} \text{ for } j \in \{1,\cdots,d\} \\ \text{and} \quad \varphi_\theta^{(2)}(\cdot) &= \left(\varphi_{\theta,j,k}^{(2)}(\cdot)\right)_{1\le j,k\le d} \text{ with } \varphi_{\theta,j,k}^{(2)}(\cdot) = \frac{\partial^2 \varphi_\theta(\cdot)}{\partial\theta_j\partial\theta_k}, \text{ for } j,k \in \{1,\cdots,d\}. \end{aligned}$$

¿From now, $\mathbb{P}$, $\mathbb{E}$ and Var denote respectively the probability $\mathbb{P}_{\theta^0,P_X}$, the expected value $\mathbb{E}_{\theta^0,P_X}$, and the variance $\text{Var}_{\theta^0,P_X}$, when the underlying and unknown true parameters are θ^0 and P_X.

3.2. **Assumptions.** We consider three types of assumptions.

Smoothness and moment assumptions

($\mathbf{A_1}$) On Θ°, the function $\theta \mapsto f_\theta$ admits continuous derivatives with respect to θ up to the order 3 .

($\mathbf{A_2}$) On Θ°, the quantity $w(X_0)(Z_1 - f_\theta(X_0))^2$, and the absolute values of its derivatives with respect to θ up to order 2 have a finite expectation.

Identifiability assumptions

(I1$_1$) The quantity $S_{\theta^0,P_X}(\theta)=\mathbb{E}[(f_{\theta^0}(X)-f_\theta(X))^2w(X)]$ admits one unique minimum at $\theta=\theta^0$.

(I1$_2$) For all $\theta\in\Theta^\circ$ the matrix $S^{(2)}_{\theta^0,P_X}(\theta)=\left(\dfrac{\partial^2 S_{\theta^0,P_X}(\theta)}{\partial\theta_j\partial\theta_j}\right)_{1\le i,j\le d}$ exists and the matrix $S^{(2)}_{\theta^0,P_X}(\theta^0)=\mathbb{E}\left[w(X)\left(f^{(1)}_{\theta^0}(X)\right)\left(f^{(1)}_{\theta^0}(X)\right)^\top\right]$ is positive definite.

Assumptions on f_ε

(N$_1$) The density f_ε belongs to $\mathbb{L}_2(\mathbb{R})$ and for all $x\in\mathbb{R}$, $f^*_\varepsilon(x)\neq 0$.

The assumption (**N$_1$**) is quite usual in density deconvolution. It ensures the existence of the estimator based on deconvolution tools.

3.3. **Estimation contrast.** As already mentionned in the introduction, the starting point of the estimation procedure is to construct an estimator of the least square contrast

$$S_{\theta^0,P_X}(\theta)=\mathbb{E}[(Z_1-f_\theta(X_0))^2\,w(X_0)],\tag{3.5}$$

based on the observations (Z_i) for $i=0,\dots,n$. The choice of the positive weight function w depends on some conditions precised later on. Its practical choice is discussed in the simulation study (Sections 5 and 6).

3.4. **General estimation criterion.** The key idea for this construction is the following: for any integrable function Φ, one has $\lim_{n\to\infty}n^{-1}\sum_{i=1}^n\Phi\star K_{n,C_n}(Z_i)=\mathbb{E}(\Phi(X))$, where K_{n,C_n} is a density kernel deconvolution defined via its Fourier transform by (1.3).

Hence we estimate $\mathbb{E}(\Phi(X))$ by $n^{-1}\sum_{i=1}^n\Phi\star K_{n,C_n}(Z_i)$ instead of $n^{-1}\sum_{i=1}^n\Phi(X_i)$ which is not available. We then propose to estimate $S_{\theta^0,P_X}(\theta)$ by

(3.6)
$$S_n(\theta)=\frac{1}{n}\sum_{i=1}^n\left((Z_i-f_\theta)^2w\right)\star K_{n,C_n}(Z_{i-1})=\frac{1}{n}\sum_{i=1}^n\int(Z_i-f_\theta(x))^2\,w(x)\,K_{n,C_n}(x-Z_{i-1})dx,$$

where $C_n\to\infty$. The kernel K belongs to $\mathbb{L}^2(\mathbb{R})$. Its Fourier transform K^* is compactly supported and satisfies $|1-K^*(t)|\le\mathbb{1}_{|t|\ge1}$. Using this empirical criterion we propose to estimate θ^0 by

$$\widehat{\theta}=\arg\min_{\theta\in\Theta}S_n(\theta).\tag{3.7}$$

3.5. **A simpler criterion.** The estimation criterion can be simplified when w is such that $(wf_\theta)^*/f^*_\varepsilon$ and $(wf^2_\theta)^*/f^*_\varepsilon$ are integrable. More precisely, we consider the following condition: assume that there exists a weight function w such that for all $\theta\in\Theta$,

(C$_1$) the functions (wf_θ) and (wf^2_θ) belong to $\mathbb{L}_1(\mathbb{R})$, and the functions w^*/f^*_ε, $(f_\theta w)^*/f^*_\varepsilon$, $(f^2_\theta w)^*/f^*_\varepsilon$ belong to $\mathbb{L}_1(\mathbb{R})$.

The first part of Condition (**C₁**) is not restrictive. The second part can be heuristically expressed as "one can find a weight function w such that wf_θ is smooth enough compared to f_ε". For a large number of regression functions, such a weight function can be easily exhibited. Nevertheless for some specific regression functions it seems not straightforward to find such a weight function. We refer to Butucea and Taupin (2008) for a more complete discussion on this point.

If (**C₁**) holds, the expectations $\mathbb{E}(w(X))$, $\mathbb{E}(w(X)f_\theta(X))$ and $\mathbb{E}(w(X)f_\theta^2(X))$ can be directly estimated without the deconvolution kernel K_n. Let φ be such that φ and $\varphi^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})$. For such a function, due to the independence between ε_0 and X_0 we have

$$\mathbb{E}\left(\frac{1}{2\pi}\int\frac{\varphi^*(t)e^{-itZ_0}}{f_\varepsilon^*(-t)}dt\right)=\mathbb{E}\left(\frac{1}{2\pi}\int\varphi^*(t)e^{-itX_0}dt\right)=\mathbb{E}[\varphi(X_0)].$$

We then propose to estimate $S_{\theta^0,P_X}(\theta)$ by

$$S_n(\theta)=\frac{1}{2\pi n}\sum_{k=1}^n\int\frac{\left((Z_k-f_\theta)^2w\right)^*(t)e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt,\tag{3.8}$$

which is the same as in (3.6) with $C_n=+\infty$. In this case $S_n(\theta)$ satisfies

$$\mathbb{E}(S_n(\theta))=\mathbb{E}[(Z_1-f_\theta(X_0))^2w(X_0)],$$

which is minimum when $\theta=\theta^0$ under the identifiablity assumption (**I1₁**). Using this empirical criterion we propose to estimate θ^0 by

$$\widehat{\theta}=\arg\min_{\theta\in\Theta}S_n(\theta).\tag{3.9}$$

4. Asymptotic properties

According to Sections 3.4 and 3.5, we consider two different theoretical frameworks.

First, in Section 4.1, we give conditions ensuring that the parametric rate of convergence as well as the asymptotic normality can be stated. Those results hold under weak dependency conditions.

Secondly, we state a consistency result under weaker conditions on f_θ, and we give an upper bound for the quadratic risk, that relates the smoothness properties of the regression function and the ones of f_ε.

4.1. **Asymptotic properties under conditions ensuring the $\sqrt{n}$-consistency.** We start with the case where a suitable choice of w ensures that Condition (**C₁**) holds. In this case, θ^0 is estimated by minimizing the simpler criterion (3.8).

The first result to mention is the consistency of our estimator. It holds under the following condition.

(**C₂**) the functions $\sup_{\theta\in\Theta}\left|\left(f_{\theta,i}^{(1)}w\right)^*/f_\varepsilon^*\right|$ and $\sup_{\theta\in\Theta}\left|\left(f_\theta f_{\theta,i}^{(1)}w\right)^*/f_\varepsilon^*\right|$ belong to $\mathbb{L}_1(\mathbb{R})$ for any $i\in\{1,\dots,d\}$.

Theorem 4.1. *Consider Model (1.1) under assumptions (**A₁**)-(**A₂**), (**I1₁**), (**I1₂**), (**N₁**), and conditions (**C₁**)-(**C₂**). Then $\widehat{\theta}$ defined by (3.9) converges in probability to θ^0.*

The following theorems show that under different dependence conditions, the estimator is $\sqrt{n}$-consistent and asymptotically Gaussian. These results require some additional conditions.

($\mathbf{C_3}$) the functions $\sup_{\theta\in\Theta}\left|\left(f^{(2)}_{\theta,i,j}w\right)^*/f^*_\varepsilon\right|$ and $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}(f^2_\theta w)\right)^*/f^*_\varepsilon\right|$ belong to $\mathbb{L}_1(\mathbb{R})$ for any $i,j\in\{1,\dots,d\}$;

($\mathbf{C_4}$) the functions $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^3(f_\theta w)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right)^*/f^*_\varepsilon\right|$ and $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}(f^2_\theta w)\right)^*/f^*_\varepsilon\right|$ belong to $\mathbb{L}_1(\mathbb{R})$, for $i,j,k\in\{1,\cdots,d\}$.

($\mathbf{C_5}$) The integrals $\int|t(f_{\theta^0}w)^*(t)|dt$ and $\int|t(f_{\theta^0}f^{(1)}_{\theta^0,k}w)^*(t)|dt$ are finite, for $k\in\{1,\cdots,d\}$.

We start by studying the $\sqrt{n}$-consistency and asymptotic normality under α-mixing conditions.

Theorem 4.2. *Consider Model (1.1) under assumptions ($\mathbf{A_1}$), ($\mathbf{A_2}$), ($\mathbf{I1_1}$), ($\mathbf{I1_2}$), ($\mathbf{N_1}$), and conditions ($\mathbf{C_1}$)-($\mathbf{C_4}$). Assume that*

$$\sum_{k\geq 1}\int_0^{\alpha_{\mathbf{X}}(k)}Q^2_{|X_1|}(u)du<\infty\,, \tag{4.10}$$

where the function $Q_{|X_1|}$ has been defined in Section 2.2. Then $\widehat{\theta}$ defined by (3.9) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies

$$\sqrt{n}(\widehat{\theta}-\theta^0)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Sigma_1),$$

where the covariance matrix Σ_1 is defined in equation (7.5).

We now extend the previous result to τ-dependency context.

Theorem 4.3. *Consider Model (1.1) under assumptions ($\mathbf{A_1}$), ($\mathbf{A_2}$), ($\mathbf{I1_1}$), ($\mathbf{I1_2}$), ($\mathbf{N_1}$), and conditions ($\mathbf{C_1}$)-($\mathbf{C_5}$). Let $G(t)=t^{-1}\mathbb{E}(X_1^2\mathbf{1}_{X_1^2>t})$, and let G^{-1} be the inverse cadlag of G. Assume that*

$$\sum_{k>0}G^{-1}(\tau_{\mathbf{X},2}(k))\tau_{\mathbf{X},2}(k)<\infty\,. \tag{4.11}$$

Then $\widehat{\theta}$ defined by (3.9) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies

$$\sqrt{n}(\widehat{\theta}-\theta^0)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\Sigma_1),$$

where the covariance matrix Σ_1 is defined in equation (7.5).

Remark. Assume that $\mathbb{E}(|X_0|^p)<\infty$ for some $p>2$. Then (4.10) is true provided that $\sum_{k>0}k^{2/(p-2)}\alpha_{\mathbf{X}}(k)<\infty$, and (4.11) is true provided that $\sum_{k>0}(\tau_{\mathbf{X},2}(k))^{(p-2)/p}<\infty$.

The choice of w is crucial here. Various weight functions can handle with Conditions $\mathbf{C_1}$-$\mathbf{C_5}$. The numerical properties of the resulting estimators will differ from one choice to another. This point is discussed in simulations in Sections 5 and 6.

4.2. **Asymptotic properties under general assumptions.** This section presents the asymptotic properties of $\widehat{\theta}$ defined by (3.7) under milder conditions than Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$, when one cannot exhibit a weight function w ensuring that these Conditions hold. In this context the estimator is still consistent, but with a rate which is not necessarily the parametric rate. For the sake of simplicity we shall only consider the case of α-mixing Markov chain.

We assume that

$(\mathbf{A_3})$ On Θ^0, the quantity $w^2(X_0)(Z_1 - f_\theta(X_0))^4$ and the absolute values of its derivatives with respect to θ up to order 2 have a finite expectation.

$(\mathbf{A_4})$ The quantity $\sup_n \sup_{j\in\{1,\cdots,d\}} \mathbb{E}\Big(\sup_{\theta\in\Theta^\circ}\Big|\dfrac{\partial}{\partial\theta_j}S_n(\theta)\Big|\Big)$ is finite.

$(\mathbf{A_5})$ $\sup_{\theta\in\Theta}|wf_\theta|$, $|w|$ and $\sup_{\theta\in\Theta}|wf_\theta^2|$ belong to $\mathbb{L}_1(\mathbb{R})$.

We say that a function $\psi \in \mathbb{L}_1(\mathbb{R})$ satisfies (4.12) if for a sequence C_n we have

$$\min_{q=1,2} \parallel \psi^*(K^*_{C_n} - 1) \parallel_q^2 + n^{-1}\min_{q=1,2}\left\|\frac{\psi^* K^*_{C_n}}{f^*_\varepsilon}\right\|_q^2 = o(1). \tag{4.12}$$

Theorem 4.4. *Under assumptions* $(\mathbf{I1_1})$, $(\mathbf{I1_2})$, $(\mathbf{N_1})$, $(\mathbf{A_1})$ $(\mathbf{A_3})$ - $(\mathbf{A_5})$, *let* $\widehat{\theta}$ *be defined by (3.7) with* C_n *such that (4.12) holds for* w, wf_θ *and* wf_θ^2 *and their first derivatives with respect to* θ. *Assume that the sequence* (X_k) *is* α*-mixing that is*

$$\alpha_{X,2}(k) \underset{n\to\infty}{\longrightarrow} 0, \text{ as } k \underset{n\to\infty}{\longrightarrow} \infty.$$

Then $\mathbb{E}(\|\widehat{\theta} - \theta^0\|^2_{\ell^2}) = o(1)$, *as* $n \to \infty$ *and* $\widehat{\theta}$ *is a consistent estimator of* θ^0.

We shall now prove rates of convergence under two different types of assumptions:

$(\mathbf{A_6})$ X_0 admits a density f_X with respect to the Lebesgue measure and there exist two constants $C_1(f^2_{\theta^0})$ and $C_2(f_{\theta^0})$ such that $\parallel f_{\theta^0}f_X \parallel_2^2 \leq C_1(f_{\theta^0})$, and $\parallel f^2_{\theta^0}f_X \parallel_2^2 \leq C_2(f^2_{\theta^0})$.

$(\mathbf{A_7})$ $\sup_{z\in\mathbb{R}}\mathbb{E}[f^2_{\theta^0}(X_0)f_\varepsilon(z - X_0)]$ and $\sup_{z\in\mathbb{R}}\mathbb{E}[f_\varepsilon(z - X_0)]$ are finite.

Theorem 4.5. *Under assumptions of Theorem 4.4, assume that the sequence* $(X_k)_{k\geq 0}$ *is* α*-mixing with* $\sum_{k\geq 1}\sqrt{\alpha_{X,2}(k)} < \infty$. *Assume moreover that for all* $\theta \in \Theta$, w, $f_\theta w$ *and* $f_\theta^2 w$ *and their derivatives up to order 3 with respect to* θ *satisfy (4.12).*

1) *Assume that the sequence* X_0 *admits a density with respect to the Lebesgue measure and that Assumption* $(\mathbf{A_6})$ *holds. Then* $\widehat{\theta} - \theta^0 = O_p(\varphi_n^2)$ *with* $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi^2_{n,j} = B^2_{n,j} + V_{n,j}/n$, $j = 1\ldots, d$, *where*

$$B_{n,j} \;=\; \min\left\{B^{[1]}_{n,j}, B^{[2]}_{n,j}\right\} \text{ and } V_{n,j} = \min\left\{V^{[1]}_{n,j}, V^{[2]}_{n,j}\right\}$$

and for $q = 1, 2$

$$B^{[q]}_{n,j} \;=\; \left\|(wf^{(1)}_{\theta,j})^*(K^*_{C_n} - 1)\right\|_q^2 + \left\|(wf_{\theta^0}f^{(1)}_{\theta^0,j})^*(K^*_{C_n} - 1)\right\|_q^2,$$

and

$$V_{n,j}^{[q]} = \left\| (wf_{\theta^0,j}^{(1)})^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2 + \left\| (wf_{\theta^0} f_{\theta^0,j}^{(1)})^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2 .$$

2) Assume that $(\mathbf{A_7})$ *holds. Then* $\widehat{\theta} - \theta^0 = O_p(\varphi_n^2)$ *with* $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi_{n,j}^2 = B_{n,j}^2 + V_{n,j}/n$, $j = 1 \dots, d$, *where* $B_{n,j} = B_{n,j}^{[1]}$ *and* $V_{n,j} = \min\left\{V_{n,j}^{[1]}, V_{n,j}^{[2]}\right\}$.

This theorem states an upper bound for the quadratic risk under very general conditions. It holds under mild conditions on w, f_θ and f_ε. We refer to Table 1 in Butucea and Taupin (2008) for details on the resulting rates.

5. Simulation context

Simulations are based on two error distributions and three different regression functions that are presented in Sections 5.1 and 5.2. Then we present the new estimator and the classical ones (without noise) that are compared in each simulation case. We refer the reader to Section 6 for the simulation results.

5.1. Distribution of errors ε_i.

We consider two types of error distribution: the Laplace distribution and the Gaussian distribution. More precisely:

- **Case 1: Laplace distribution.** In that case ε_1 has the density

(5.13)
$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2}} \exp\Big(- \frac{\sqrt{2}}{\sigma_\varepsilon}|x|\Big), \text{ and } f_\varepsilon^*(x) = \frac{1}{1 + \sigma_\varepsilon^2 x^2/2}.$$

Hence, ε_1 is centered with variance σ_ε^2.

- **Case 2: Gaussian distribution.** In that case ε_1 has the density

(5.14)
$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\Big(- \frac{x^2}{2\sigma_\varepsilon^2}\Big), \text{ and } f_\varepsilon^*(x) = \exp(-\sigma_\varepsilon^2 x^2/2).$$

Hence, ε_1 is centered with variance σ_ε^2.

5.2. The regression functions.

We consider three regression functions:

- **Case A: Linear regression function.** We consider the model (2.3), with $f_\theta(x) = ax + b$, where $\theta = (a, b)^T$ and with binary innovations. The true parameter is $\theta^0 = (1/2, 1/4)^T$. The stationary distribution is the uniform distribution over $[0, 1]$ and consequently $\sigma_{X_0}^2 = 1/12$. For the simulation, we start with X_0 uniformly distributed over $[0, 1]$, so the simulated chain is stationary. Recall that this chain is non-irreducible and that $\alpha_{\mathbf{X}}(k) = 1/4$ and $\tau_{\mathbf{X},2}(k) = O(2^{-k})$.

- **Case B: Linear regression function.** We consider the model (2.4) with $f_\theta(x) = ax + b$, where $\theta = (a, b)^T$ and with binary innovations. The true parameter is $\theta^0 = (1/3, 1/3)^T$. The stationary distribution is the uniform distribution over the Cantor set and consequently $\sigma_X^2 = 1/8$. Recall that this chain is non-irreducible and that $\alpha_{\mathbf{X}}(k) = 1/4$ and $\tau_{\mathbf{X},2}(k) = O(3^{-k})$. For the simulation, we start with X_0 uniformly distributed over $[0, 1]$, and we consider that the chain is close to the stationary chain after 1000 iterations. We then set $X_k = X_{k+1000}$.

These two models are close, but in the second case, the stationary distribution is not absolutely continuous with respect to the Lebesgue measure.

- **Case C: Cauchy regression function.** We consider the model (1.1) with $f_\theta(x) = \theta/(1+x^2)$. The true parameter is $\theta^0 = 1.5$. For the law of ξ_0 we take ξ_0 such that $\mathbb{P}(\xi_0 = -1/2) = \mathbb{P}(\xi_0 = 1/2) = 1/2$. In this case, empirical studies show that σ_X^2 is about 0.1 and $\tau_{\mathbf{X},2}(k) = O(\kappa^k)$ for some $\kappa \in]0,1[$. For the simulation, we start with X_0 uniformly distributed over $[0,1]$, and we consider that the chain is close to the stationary chain after 1000 iterations. We then set $X_k = X_{k+1000}$.

In these three cases, we can find a weight function w that allow us to use the simpler criterion defined in (3.8). We recall that Theorem 4.3 holds and that the estimators are asymptoticaly Gaussian.

5.3. **The compared estimators in the two Linear autoregressive models.** The estimators are the same for cases A and B and are presented thereafter. We first give the details on the estimator for two choices of the weight function w. Then we recall the classic estimator when X is directly observed, the ARMA estimator and the so-called naive estimator.

We start with our estimation procedure. The choice of the weight function w depends on the regression function f_θ which is linear in that case. Thus, we consider the two following weight functions w defined by

$$w(x) = N(x) = \exp\{-x^2/(4\sigma_\varepsilon^2)\} \text{ and } w(x) = SC(x) = \Big(\frac{2 * \sin(x)}{x}\Big)^4. \tag{5.15}$$

These choices of weight ensure that Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$ hold and that our two estimators, denoted $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ respectively, converge to θ^0 with the parametric rate of convergence. There are two main differences between these two weight functions. First, N depends on the variance error σ_ε^2. Hence the estimator should be adaptive to the noise level. On the contrary, it may be sensitive to very small errors variance as it appears in the simulations (see Figure 1). Second, SC has strong smoothness properties since its Fourier transform is compactly supported.

The construction of these estimators is based on the criterion $S_n(\theta)$, which can be written as

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}[(Z_k^2 + b^2 - 2Z_k b)I_0(Z_{k-1}) + a^2 I_2(Z_{k-1}) - 2aI_1(Z_{k-1})],$$

with

$$I_j(Z) = \frac{1}{2\pi}\int (p_j w)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(u)}du, \tag{5.16}$$

where $p_j(x) = x^j$ for $j = 0, 1, 2$, w being either $w = N$ or $w = SC$. With the above notations, $\widehat{\theta}$ defined as

$$\widehat{\theta} \;=\; \begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} = \arg\min_{\theta\in\Theta} S_n(\theta)$$

satisfies

$$\widehat{a} = \frac{\sum_{k=1}^n Z_k I_1(Z_{k-1}) \sum_{k=1}^n I_0(Z_{k-1}) - \sum_{k=1}^n Z_k I_0(Z_{k-1}) \sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_2(Z_{k-1}) \sum_{k=1}^n I_0(Z_{k-1}) - \big(\sum_{k=1}^n I_1(Z_{k-1})\big)^2}, \tag{5.17}$$

$$\widehat{b} = \frac{\sum_{k=1}^n Z_k I_0(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})} - \widehat{a}\frac{\sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})}. \tag{5.18}$$

We now compute $I_j(Z)$ for $j = 0, 1, 2$ and the two weight functions. In the following we respectively denote $I_{j,N}(Z)$ and $I_{j,SC}(Z)$ the previous integrals when the weight function is either $w = N$ or $w = SC$. We start with $w = N$ and we give the details of the calculations for the two error distributions (Laplace and Gaussian). With this weight function, the calculations are explicit. Then, with the weight function $w = SC$, we present the calculations, which are not explicit whatever the error distribution f_ε.

- When $w = N$, Fourier calculations provide that

$$\begin{aligned} N^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2) \\ (Np_1)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 t/i\big), \\ (Np_2)^*(t) &= -\sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 + 4\sigma_\varepsilon^4 t^2\big). \end{aligned}$$

It follows that

$$\begin{aligned} I_{0,N}(Z) &= \frac{1}{2\pi}\int (N)^*(t)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt = \frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt, \\ I_{1,N}(Z) &= \frac{1}{2\pi}\int (Np_1)^*(t)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt = \frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 t/i\big)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt, \\ I_{2,N}(Z) &= \frac{1}{2\pi}\int (Np_2)^*(t)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt = \frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(2\sigma_\varepsilon^2 - 4\sigma_\varepsilon^4 t^2\big)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt. \end{aligned}$$

Case 1: If f_ε is the Laplace distribution (5.13), replacing f_ε^* by its expression we obtain

$$I_{0,N}(Z) = \frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}e^{-\sigma_\varepsilon^2 t^2}\frac{e^{-itZ}}{1/(1+\sigma_\varepsilon^2 t^2/2)}dt.$$

Then we use that

$$\frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}e^{-\sigma_\varepsilon^2 t^2}e^{-itZ}dt = N(Z),$$

to get that

$$I_{0,N}(Z) = e^{-Z^2/(4\sigma_\varepsilon^2)} - \frac{\sigma_\varepsilon^2}{2}\frac{\partial^2}{\partial Z^2}w(Z) = \big[5/4 - Z^2/(8\sigma_\varepsilon^2)\big]\, e^{-Z^2/(4\sigma_\varepsilon^2)}.$$

Using the same arguments we have

$$I_{1,N}(Z) = \big[7Z/4 - Z^3/(8\sigma_\varepsilon^2)\big]\, e^{-Z^2/(4\sigma_\varepsilon^2)}, \quad \text{and} \quad I_{2,N}(Z) = \big[-\sigma_\varepsilon^2 + 9Z^2/4 - Z^4/(8\sigma_\varepsilon^2)\big]\, e^{-Z^2/(4\sigma_\varepsilon^2)}.$$

Case 2: If f_ε is the Gaussian distribution (5.14), replacing f_ε^* by its expression we obtain

$$I_{0,N}(Z) = \frac{1}{2\pi}\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}e^{-\sigma_\varepsilon^2 t^2}\frac{e^{-itZ}}{e^{-\sigma_\varepsilon^2 t^2/2}}dt = \frac{1}{\pi}\int e^{-t^2/2}e^{-itZ/\sigma_\varepsilon}dt = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}.$$

In the same way

$$I_{1,N}(Z) = 2\sqrt{2}Ze^{-Z^2/(2\sigma_\varepsilon^2)} \quad \text{and} \quad I_{2,N}(Z) = \sqrt{2}(4Z^2 - 2\sigma_\varepsilon^2)e^{-Z^2/(2\sigma_\varepsilon^2)}.$$

Hence we deduce the expression of $\widehat{a}_N$ and $\widehat{b}_N$ by applying (5.17).

•When $w = SC$, Fourier calculations provide that

$$\begin{aligned} SC^*(t) &= \mathbb{1}_{[-4,-2]}(t)(t^3/6 + 2t^2 + 8t + 32/3) + \mathbb{1}_{[-2,0]}(t)(-t^3/2 - 2t^2 + 16/3) \\ &\quad + \mathbb{1}_{[2,4]}(t)(-t^3/6 + 2t^2 - 8t + 32/3) + \mathbb{1}_{[0,2]}(t)(t^3/2 - 2t^2 + 16/3) \\ (SCp_1)^*(t) &= \frac{\partial}{\partial t}SC^*(t)/i \text{ and } (SCp_2)^*(t) = \frac{\partial^2}{\partial t^2}SC^*(t)/(i^2). \end{aligned}$$

The integrals $I_{j,SC}(Z)$, defined for $j = 0, 1, 2$ by

$$I_{j,SC}(Z) = \frac{1}{2\pi}\int (SCp_j)^*(t)\frac{e^{-itZ}}{f_\varepsilon^*(t)}dt. \tag{5.19}$$

have not an explicit form, whatever the error distribution f_ε. It has to be numerically computed, using the IFFT Matlab function. More precisely, we consider a finite Fourier series approximation of $(SCp_j)^*(t)/f_\varepsilon^*(t)$ whose Fourier transfom is calculated using IFFT Matlab function. The result is taken as an approximation of $I_{j,SC}(Z)$. Finally we deduce the expression of $\widehat{a}_{SC}$ and $\widehat{b}_{SC}$ by applying (5.17).

Comparison with the case without noise. In the case $\varepsilon_0 = 0$, that is $(X_0, \dots, X_n)$ is observed without error, the parameters can be easily estimated by the usual least square estimators

$$\widehat{a}_X = \frac{n\sum_{i=1}^n X_iX_{i-1} - \sum_{i=1}^n X_i\sum_{i=1}^n X_{i-1}}{n\sum_{i=1}^n X_{i-1}^2 - (\sum_{i=1}^n X_{i-1})^2} \quad \text{and} \quad \widehat{b}_X = \frac{1}{n}\Big(\sum_{i=1}^n X_i\Big) - \widehat{a}_X\frac{1}{n}\Big(\sum_{i=1}^n X_{i-1}\Big).$$

Comparison with an ARMA estimator When the regression function is linear with $a \neq 0$, the model (1.1) can be seen as an ARMA(1,1) model of the form

$$Z_i - aZ_{i-1} = b + \eta_i - \beta\eta_{i-1}, \tag{5.20}$$

where η is the innovation and β is linked to the initial parameters. In this context, one can use the estimators obtained by maximizing the so-called Gaussian likelihood. These estimators are consistent, asymptotically Gaussian. Moreover they are efficient when both the innovations and the errors ε are Gaussian (see Hannan (1973), Brockwell and Davis (1991)). Their constructions do not require the knowledge of the error density f_ε, but they are no longer consistent if f is not linear. For its computation we use the function *arma* from R *tseries* package (see Trapletti and Hornik (2011)). The resulting estimators are denoted by $\widehat{a}_{arma}$ and $\widehat{b}_{arma}$.

Comparison with the so-called naive estimator The naive estimator is constructed by replacing the unobserved X_i by the observation Z_i in the expression of $\widehat{a}_X$ and $\widehat{b}_X$:

$$\widehat{a}_{naive} = \frac{n\sum_{i=1}^n Z_iZ_{i-1} - \sum_{i=1}^n Z_i\sum_{i=1}^n Z_{i-1}}{n\sum_{i=1}^n Z_{i-1}^2 - (\sum_{i=1}^n Z_{i-1})^2} \quad \text{and} \quad \widehat{b}_{naive} = \frac{1}{n}\Big(\sum_{i=1}^n Z_i\Big) - \widehat{a}_{naive}\frac{1}{n}\Big(\sum_{i=1}^n Z_{i-1}\Big).$$

Classical results show that $\widehat{\theta}_{naive}$ is a biased estimator of θ^0, which is confirmed by the simulation study.

5.4. **The compared estimators in the Cauchy regression model.** To our knowledge, the estimator $\widehat{\theta}$ is the first consistent estimator that has been proposed in the literature for this regression function. We first detail our estimator for two choices of the weight function w. Then we recall the classic estimator when X is directly observed and the so-called naive estimator.

The weight function depends on the regression function, which is $f_\theta(x) = \theta f(x)$ with $f(x) = 1/(1+x^2)$. Thus, we consider the two following weight functions:

$$N_c(x) = (1+x^2)^2 \exp\{-x^2/(4\sigma_\varepsilon^2)\} \text{ and } SC_c(x) = (1+x^2)^2\Big(\frac{2*\sin(x)}{x}\Big)^4, \tag{5.21}$$

with σ_ε^2 the variance of ε. This choice of w ensures that Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$ hold and our method allows to achieve the parametric rate of convergence. As in the linear case, these two weight functions differ by their dependence on σ_ε^2 and their smoothness properties. The construction of the estimator is based on the criterion $S_n(\theta)$, which can be written as

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}[Z_k^2 I_w(Z_{k-1}) + \theta^2 I_{wf^2}(Z_{k-1}) - 2\theta Z_i I_{wf}(Z_{k-1})],$$

where

$$\begin{aligned} I_w(Z) &= \frac{1}{2\pi}\int (w)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du, \qquad I_{wf}(Z)) = \frac{1}{2\pi}\int (wf)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du \\ \text{and} \quad I_{wf^2}(Z) &= \frac{1}{2\pi}\int (wf^2)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du. \end{aligned}$$

The estimator can be expressed as

$$\widehat{\theta} = \frac{\sum_{k=1}^n Z_k I_{wf}(Z_{k-1})}{\sum_{k=1}^n I_{wf^2}(Z_{k-1})}. \tag{5.22}$$

We now compute $I_{wf}(Z)$ and $I_{wf^2}(Z)$ for the two weight functions. In the following we respectively denote by $I_{wf,N_c}(Z)$, $I_{wf^2,N_c}(Z)$, $I_{wf,SC_c}(Z)$ and $I_{wf^2,SC_c}(Z)$ the previous integrals when the weight function is either $w = N_c$ or $w = SC_c$. In the same way we denote by $\widehat{\theta}_{N_c}$ and $\widehat{\theta}_{SC_c}$ the coresponding estimators of θ^0.

- When $w = N_c$, Fourier calculations provide that

$$\begin{aligned} (N_c f)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(1+2\sigma_\varepsilon^2(1-2\sigma_\varepsilon^2 t^2)\big) \\ \text{and } (N_c f^2)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2). \end{aligned}$$

Calculations of integrals $I_{wf,N_c}(Z)$ and $I_{wf^2,N_c}(Z)$ are explicit with this weight function, but depend on the error distributions. We give the details in both cases.

Case 1: If f_ε is the Laplace distribution (5.13), replacing f_ε^* by its expression we obtain

$$\begin{aligned} I_{wf,N_c}(Z) &= \exp(-Z^2/(4\sigma_\varepsilon^2))\left[Z^4 - 18Z^2\sigma_\varepsilon^2 + Z^2 + 8\sigma_\varepsilon^4 - 10\sigma_\varepsilon^2\right]/(8\sigma_\varepsilon^2), \\ \text{and } I_{wf^2,N_c}(Z) &= \exp(-Z^2/(4\sigma_\varepsilon^2))\big[1+\frac{1}{4}\big(1-\frac{Z^2}{2\sigma_\varepsilon^2}\big)\big]. \end{aligned}$$

Case 2: If f_ε is the Gaussian distribution (5.14), replacing f_ε^* by its expression we obtain

$$I_{wf,N_c}(Z) = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}(1-2\sigma_\varepsilon^2+4Z^2), \text{ and } \quad I_{wf^2,N_c}(Z) = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}.$$

- When $w = SC_c$, easy calculations show that

$$I_{wf,SC_c}(Z) = I_{0,SC}(Z) + I_{2,SC}(Z) \text{ and } I_{wf^2,SC_c}(Z) = I_{0,SC}(Z),$$

where $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ are defined by (5.19). As explained before, the integrals $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ have no explicit form, whatever the error distributions, and are numerically approximated via the IFFT function.

Comparison with the case without noise. In the case $\varepsilon_0 = 0$, that is $(X_0, \dots, X_n)$ is observed without errors, the parameters can be easily estimated by the usual least square estimator

$$\widehat{\theta}_X = \frac{\sum_{i=1}^n X_i f(X_{i-1})}{\sum_{i=1}^n f^2(X_{i-1})}.$$

Comparison with the so-called naive estimator. The idea for the construction of the naive estimators is to replace X_i by the observation Z_i in the expression of $\widehat{\theta}_X$ to get

$$\widehat{\theta}_{naive} = \frac{\sum_{i=1}^n Z_i f(Z_{i-1})}{\sum_{i=1}^n f^2(Z_{i-1})}.$$

Classical results show that $\widehat{\theta}_{naive}$ is a biased estimator of θ^0, which is confirmed by the simulation study.

6. SIMULATIONS RESULTS

In this section, we present the results of the simulation study introduced in the previous section. For each regression function (Case A, B or C) and each error distribution (Case 1 or 2), we simulate 100 samples with size n for various n ($n = 500$, $n = 5000$ and $n = 10000$) and for various values of σ_ε. More precisely, the error variance is chosen such that the ratio signal to noise $s2n = \sigma_\varepsilon^2/\mathrm{Var}(X)$ is $0.5, 1.5$ or 3.

The comparison of the five (resp. four) estimators in Cases A and B (resp. C) are based on bias, Mean Squared Error (MSE) and box plots illustrating the convergence of the estimators. If $\widehat{\theta}(k)$ denotes the value of the estimation for the k-th sample, the MSE is evaluated by the empirical mean over the 100 samples:

$$MSE(\widehat{\theta}) = \frac{1}{100}\sum_{k=1}^{100}(\widehat{\theta}(k) - \theta^0)^2.$$

For each estimator, Tables 1-6 provide the empirical mean over the 100 samples with the corresponding MSE in brackets.

	ratio				Estimator		
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_{N}(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_{X}(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.487 (0.008)	0.459 (0.020)	0.489 (0.002)	0.493 (0.001)	0.328 (0.030)
		b	0.257 (0.002)	0.262 (0.002)	0.255 (0.001)	0.253 (0.001)	0.336 (0.008)
	1.5	a	0.494 (0.015)	0.488 (0.013)	0.492 (0.006)	0.501 (0.001)	0.198 (0.092)
		b	0.251 (0.004)	0.253 (0.002)	0.253 (0.002)	0.249 (0.001)	0.399 (0.023)
	3	a	0.461 (0.044)	0.502 (0.029)	0.503 (0.026)	0.493 (0.001)	0.121 (0.145)
		b	0.270 (0.012)	0.249 (0.001)	0.249 (0.001)	0.253 (0.001)	0.440 (0.037)
5000	0.5	a	0.497 (0.001)	0.499 (0.004)	0.499 (0.001)	0.499 (0.001)	0.332 (0.028)
		b	0.252 (0.001)	0.251 (0.001)	0.251 (0.001)	0.251 (0.001)	0.334 (0.007)
	1.5	a	0.498 (0.003)	0.508 (0.003)	0.503 (0.002)	0.499 (0.001)	0.199 (0.091)
		b	0.250 (0.001)	0.247 (0.001)	0.248 (0.001)	0.250 (0.001)	0.399 (0.022)
	3	a	0.487 (0.008)	0.492 (0.004)	0.495 (0.004)	0.500 (0.001)	0.123 (0.143)
		b	0.256 (0.002)	0.253 (0.001)	0.252 (0.001)	0.250 (0.001)	0.437 (0.035)
10000	0.5	a	0.496 (0.001)	0.501 (0.002)	0.500 (0.001)	0.499 (0.001)	0.334 (0.028)
		b	0.252 (0.001)	0.250 (0.001)	0.250 (0.001)	0.250 (0.001)	0.333 (0.007)
	1.5	a	0.504 (0.002)	0.500 (0.001)	0.501 (0.001)	0.500 (0.001)	0.200 (0.090)
		b	0.248 (0.001)	0.250 (0.001)	0.250 (0.001)	0.250 (0.001)	0.401 (0.023)
	3	a	0.493 (0.003)	0.499 (0.001)	0.499 (0.002)	0.498 (0.001)	0.124 (0.142)
		b	0.254 (0.001)	0.250 (0.001)	0.250 (0.001)	0.251 (0.001)	0.438 (0.036)

TABLE 1. Estimation results for Linear Case A, Laplace error Case 1. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_{N}$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_{X}$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

6.1. **Linear regression function.** In the linear case we have five estimators to be compared which are $\widehat{\theta}_{arma}$ (ARMA estimator), $\widehat{\theta}_{N}$ (our procedure with $w = N$), $\widehat{\theta}_{SC}$ (our procedure with $w = SC$), $\widehat{\theta}_{X}$ (classical estimator without noise) and finally $\widehat{\theta}_{naive}$ (X replaced by Z). Results are presented in Figures 1-2 and Tables 1-4.

The first thing to notice is that, not surprisingly, $\widehat{\theta}_{naive}$ presents a bias, whatever the values of n, $s2n$ and the error distribution. The estimator $\widehat{\theta}_{X}$ has the good expected properties (unbiased and small MSE), but it is based on the observation of the X_i's. The previously known estimator $\widehat{\theta}_{arma}$ has good asymptotic properties. However its bias is often larger than biases of $\widehat{\theta}_{N}$ and $\widehat{\theta}_{SC}$, except when $s2n = 0.5$ and ε is Gaussian.

	ratio		Estimator				
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.483 (0.006)	0.539 (0.039)	0.496 (0.002)	0.495 (0.001)	0.331 (0.030)
		b	0.259 (0.002)	0.243 (0.003)	0.253 (0.001)	0.253 (0.001)	0.336 (0.008)
	1.5	a	0.497 (0.021)	0.516 (0.027)	0.507 (0.009)	0.499 (0.001)	0.200 (0.091)
		b	0.251 (0.005)	0.243 (0.005)	0.246 (0.002)	0.249 (0.001)	0.399 (0.023)
	3	a	0.456 (0.031)	0.521 (0.082)	0.481 (0.030)	0.501 (0.001)	0.120 (0.145)
		b	0.272 (0.008)	0.244 (0.016)	0.260 (0.007)	0.250 (0.001)	0.441 (0.037)
5000	0.5	a	0.497 (0.001)	0.492 (0.006)	0.499 (0.001)	0.498 (0.001)	0.333 (0.028)
		b	0.251 (0.001)	0.252 (0.001)	0.250 (0.001)	0.250 (0.001)	0.333 (0.007)
	1.5	a	0.490 (0.002)	0.510 (0.006)	0.502 (0.001)	0.499 (0.001)	0.120 (0.090)
		b	0.254 (0.001)	0.245 (0.001)	0.248 (0.001)	0.250 (0.001)	0.399 (0.022)
	3	a	0.471 (0.010)	0.512 (0.008)	0.503 (0.005)	0.498 (0.001)	0.124 (0.141)
		b	0.263 (0.002)	0.245 (0.002)	0.249 (0.001)	0.251 (0.001)	0.437 (0.035)
10000	0.5	a	0.504 (0.006)	0.500 (0.003)	0.498 (0.001)	0.499 (0.001)	0.331 (0.028)
		b	0.249 (0.001)	0.250 (0.001)	0.251 (0.001)	0.251 (0.001)	0.335 (0.007)
	1.5	a	0.495 (0.002)	0.501 (0.002)	0.499 (0.001)	0.501 (0.001)	0.200 (0.090)
		b	0.253 (0.001)	0.250 (0.001)	0.251 (0.001)	0.250 (0.001)	0.401 (0.023)
	3	a	0.492 (0.004)	0.498 (0.004)	0.500 (0.003)	0.500 (0.001)	0.126 (0.140)
		b	0.254 (0.001)	0.251 (0.001)	0.251 (0.001)	0.250 (0.001)	0.437 (0.009)

TABLE 2. Estimation results for Linear Case A, Gaussian error Case 2. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

We now consider the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$. Recall that their construction requires the choice of w. Note first that, whatever the weight function w, the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ present good convergence properties. Their biases and MSEs decrease when n increases. When compared one to another, we can see that their numerical behaviors are not the same. Namely for not too large $s2n$, $\widehat{\theta}_{SC}$ has a MSE smaller than $\widehat{\theta}_N$ (see Figure 1 and Tables 1-4, when $s2n \leq 3$). With large $s2n$, the estimator $\widehat{\theta}_N$ seems to have better properties (see Figure 2 when $s2n = 6$). This is expected since N depends on σ_ε^2 and is thus more sensitive to small values of σ_ε^2. The errors ε distribution seems to have a slight infuence on the MSEs of the two estimators. The MSEs are often smaller when f_ε is the Laplace density. This may be related

	ratio				Estimator		
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.288 (0.021)	0.341 (0.013)	0.330 (0.002)	0.326 (0.001)	0.217 (0.015)
		b	0.354 (0.005)	0.331 (0.001)	0.333 (0.001)	0.335 (0.001)	0.389 (0.004)
	1.5	a	0.298 (0.050)	0.332 (0.009)	0.335 (0.007)	0.330 (0.001)	0.136 (0.040)
		b	0.349 (0.012)	0.331 (0.002)	0.329 (0.002)	0.335 (0.001)	0.429 (0.010)
	3	a	0.240 (0.127)	0.343 (0.017)	0.343 (0.018)	0.330 (0.001)	0.084 (0.063)
		b	0.385 (0.033)	0.333 (0.003)	0.333 (0.003)	0.338 (0.001)	0.465 (0.018)
5000	0.5	a	0.333 (0.004)	0.335 (0.003)	0.335 (0.001)	0.333 (0.001)	0.223 (0.012)
		b	0.333 (0.001)	0.332 (0.001)	0.332 (0.001)	0.334 (0.001)	0.388 (0.003)
	1.5	a	0.331 (0.011)	0.328 (0.002)	0.334 (0.001)	0.334 (0.001)	0.433 (0.041)
		b	0.334 (0.003)	0.334 (0.001)	0.329 (0.001)	0.332 (0.001)	0.132 (0.010)
	3	a	0.290 (0.030)	0.329 (0.003)	0.329 (0.004)	0.333 (0.001)	0.083 (0.063)
		b	0.355 (0.008)	0.335 (0.008)	0.335 (0.008)	0.334 (0.001)	0.459 (0.016)
10000	0.5	a	0.337 (0.002)	0.335 (0.002)	0.334 (0.001)	0.334 (0.001)	0.222 (0.012)
		b	0.331 (0.001)	0.332 (0.001)	0.332 (0.001)	0.332 (0.001)	0.388 (0.003)
	1.5	a	0.322 (0.006)	0.336 (0.001)	0.336 (0.001)	0.334 (0.001)	0.134 (0.040)
		b	0.339 (0.002)	0.332 (0.001)	0.332 (0.001)	0.333 (0.001)	0.433 (0.010)
	3	a	0.329 (0.010)	0.336 (0.002)	0.336 (0.002)	0.334 (0.001)	0.083 (0.063)
		b	0.335 (0.002)	0.332 (0.001)	0.332 (0.001)	0.332 (0.001)	0.457 (0.015)

TABLE 3. Estimation results for Linear Case B, Laplace error Case 1. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

with the theoretical properties in density deconvolution. In that context it is well known that the rate of convergence is slower when f_ε is the Gaussian density. The two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ have comparable numerical behaviors in the two linear autoregressive models. Let us recall that in both cases, the simulated chain X are non-mixing but are τ-dependent. In Case A, the stationary distribution of X is continuous whereas it is not the case in Case B. This explains the relative bad properties of $\widehat{\theta}_{arma}$ in Case B. Indeed, due to its construction, this estimator is expected to have good properties when the stationary distribution of the Markov Chain is close to the Gaussian distribution. On the contrary our estimators have similar behavior in both cases.

	ratio				Estimator		
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_{N}(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_{X}(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.327 (0.016)	0.349 (0.035)	0.330 (0.003)	0.326 (0.001)	0.218 (0.014)
		b	0.338 (0.004)	0.332 (0.002)	0.336 (0.001)	0.337 (0.001)	0.392 (0.004)
	1.5	a	0.290 (0.061)	0.355 (0.021)	0.345 (0.008)	0.332 (0.001)	0.133 (0.041)
		b	0.353 (0.015)	0.324 (0.004)	0.328 (0.002)	0.333 (0.001)	0.432 (0.010)
	3	a	0.234 (0.153)	0.329 (0.049)	0.329 (0.051)	0.326 (0.001)	0.077 (0.067)
		b	0.383 (0.040)	0.337 (0.010)	0.337 (0.010)	0.337 (0.001)	0.461 (0.017)
5000	0.5	a	0.329 (0.004)	0.341 (0.005)	0.333 (0.001)	0.332 (0.001)	0.220 (0.013)
		b	0.335 (0.001)	0.332 (0.001)	0.334 (0.001)	0.334 (0.001)	0.399 (0.003)
	1.5	a	0.329 (0.009)	0.331 (0.003)	0.332 (0.002)	0.333 (0.001)	0.132 (0.041)
		b	0.335 (0.002)	0.334 (0.001)	0.333 (0.001)	0.333 (0.001)	0.433 (0.010)
	3	a	0.315 (0.022)	0.348 (0.008)	0.348 (0.008)	0.334 (0.001)	0.084 (0.062)
		b	0.343 (0.006)	0.327 (0.002)	0.328 (0.002)	0.332 (0.001)	0.459 (0.016)
10000	0.5	a	0.330 (0.002)	0.333 (0.003)	0.333 (0.001)	0.332 (0.001)	0.221 (0.013)
		b	0.335 (0.001)	0.333 (0.001)	0.333 (0.001)	0.334 (0.001)	0.389 (0.003)
	1.5	a	0.328 (0.006)	0.336 (0.002)	0.334 (0.001)	0.333 (0.001)	0.132 (0.041)
		b	0.336 (0.002)	0.333 (0.001)	0.334 (0.001)	0.334 (0.001)	0.435 (0.010)
	3	a	0.312 (0.014)	0.334 (0.004)	0.334 (0.004)	0.333 (0.001)	0.083 (0.063)
		b	0.344 (0.003)	0.333 (0.001)	0.333 (0.001)	0.333 (0.001)	0.458 (0.016)

TABLE 4. Estimation results for Linear Case B, Gaussian error Case 2. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_{N}$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_{X}$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

6.2. **Cauchy regression function.** We propose to compare the two estimators $\widehat{\theta}_{N_c}$ (our procedure with $w = N_c$), $\widehat{\theta}_{SC_c}$ (our procedure with $w = SC_c$), with the classical estimator $\widehat{\theta}_X$ (without noise) and the naive $\widehat{\theta}_{naive}$ (X replaced by Z). Results are presented in Figure 3 and Tables 5-6.

The first thing to notice is that, not surprisingly, $\widehat{\theta}_{naive}$ presents a bias, whatever the values of n, $s2n$ and the errors distribution. Moreover it converges to (false) values which are different according to $s2n$ (see Tables (5)-(6)).

The estimator $\widehat{\theta}_X$ has the good expected properties (unbiased and small MSE), but it is based on the observation of the X_i's.

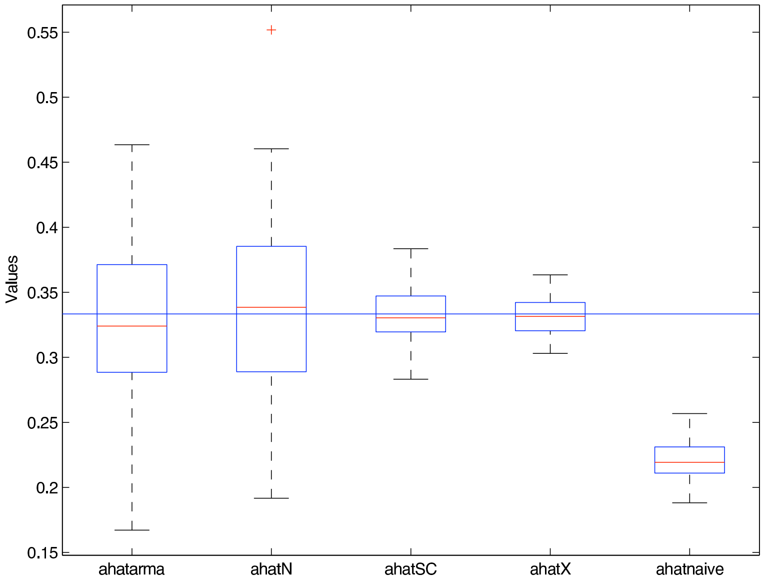

Figure 1. Results for linear Case B and Gaussian error Case 2, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 0.5$. Box plots of the five estimators $\widehat{a}_{arma}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{naive}$, from left to right, based on 100 replications. True value is 1/3 (horizontal line).

We now compare our two estimators illustrating the influence of w, $s2n$ and f_ε. Globally, whatever the weight function w, the two estimators $\widehat{\theta}$ present good convergence properties. Their biases and MSEs decrease when n increases. The MSEs of $\widehat{\theta}_{SC_c}$ increase when $s2n$ increases. This is not the case for the MSE of $\widehat{\theta}_{N_c}$. This is probably due to the fact that the weight function chosen for the construction of $\widehat{\theta}_{N_c}$ depends on σ_ε^2. This estimator is thus more adaptive to changes in $s2n$.

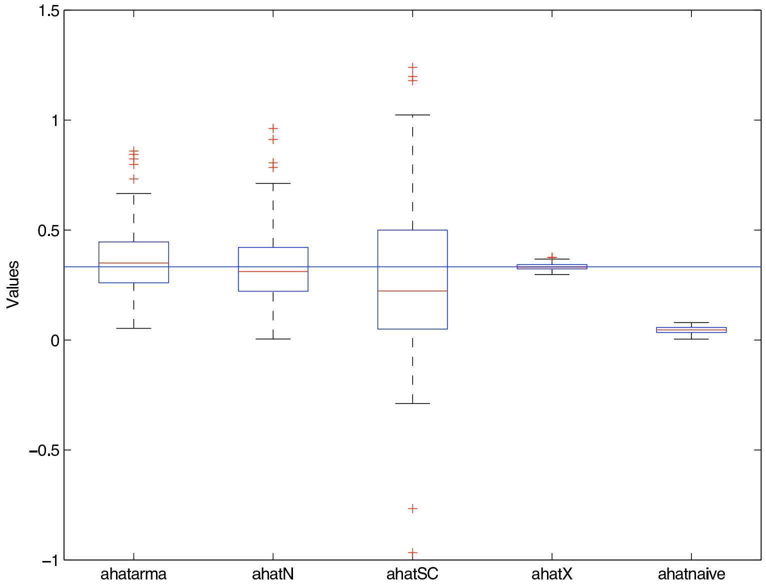

Figure 2. Results for linear Case B and Gaussian error Case 2, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 6$. Box plots of the five estimators $\widehat{a}_{arma}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{naive}$, from left to right, based on 100 replications. True value is 1/3 (horizontal line).

	ratio	Estimator			
n	s2n	$\widehat{\theta}_{N_c}(MSE)$	$\widehat{\theta}_{SC_c}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	1.5095 (0.0042)	1.5024 (0.0006)	1.5004 (0.0000)	1.4333 (0.0050)
	1.5	1.5006 (0.0021)	1.5005 (0.0013)	1.5002 (0.0000)	1.3657 (0.0190)
	3	1.5017 (0.0024)	1.5005 (0.0024)	1.5002 (0.0000)	1.3267 (0.0314)
5000	0.5	1.5045 (0.0008)	1.5005 (0.0001)	1.5003 (0.0000)	1.4320 (0.0047)
	1.5	1.5003 (0.0004)	1.4994 (0.0003)	1.4997 (0.0000)	1.3647 (0.0185)
	3	1.4989 (0.0005)	1.4992 (0.0005)	1.5000 (0.0000)	1.3223 (0.0318)
10000	0.5	1.5033 (0.0004)	1.5002 (0.0001)	1.5000 (0.0000)	1.4315 (0.0047)
	1.5	1.5000 (0.0002)	1.5000 (0.0001)	1.4998 (0.0000)	1.3650 (0.0183)
	3	1.4972 (0.0002)	1.4970 (0.0002)	1.4998 (0.0000)	1.3222 (0.0317)

Table 5. Estimation results for Cauchy Case C, Laplace error Case 1. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True value is $\theta^0 = 1.5$. MSE are given in brackets.

	ratio	Estimator			
n	s2n	$\widehat{\theta}_{N_c}(MSE)$	$\widehat{\theta}_{SC_c}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	1.4979 (0.0027)	1.4998 (0.0006)	1.5000 (0.0000)	1.4230 (0.0064)
	1.5	1.4995 (0.0029)	1.5001 (0.0015)	1.5005 (0.0000)	1.3336 (0.0287)
	3	1.5080 (0.0049)	1.5058 (0.0042)	1.4997 (0.0000)	1.2832 (0.0487)
5000	0.5	1.5033 (0.0006)	1.5011 (0.0001)	1.4999 (0.0000)	1.4250 (0.0057)
	1.5	1.5011 (0.0004)	1.5001 (0.0003)	1.4999 (0.0000)	1.3351 (0.0274)
	3	1.4998 (0.0009)	1.4996 (0.0008)	1.5002 (0.0000)	1.2767 (0.0501)
10000	0.5	1.5017 (0.0003)	1.4997 (0.0000)	1.4996 (0.0000)	1.4236 (0.0059)
	1.5	1.5025 (0.0003)	1.5027 (0.0002)	1.5001 (0.0000)	1.3375 (0.0265)
	3	1.5016 (0.0004)	1.5021 (0.0004)	1.5002 (0.0000)	1.2778 (0.0495)

TABLE 6. Estimation results for Cauchy Case C, Gaussian error Case 2. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True value is $\theta^0 = 1.5$. MSE are given in brackets.

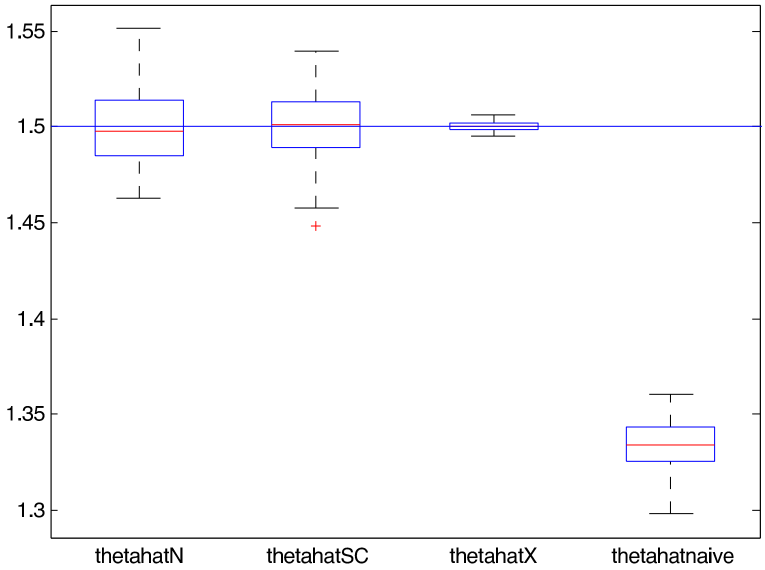

FIGURE 3. Results for Cauchy Case C and Gaussian error Case 2, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 1.5$. Box plots of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$, from left to right, based on 100 replications. True value is 1.5 (horizontal line).

7. Proofs of Theorems

7.1. **Proof of Theorem 4.1.** The main point of the proof consists in showing the two following points

i) for any θ in Θ, $S_n(\theta) \xrightarrow[n\to\infty]{\mathbb{L}^1} S_{\theta^0,P_X}(\theta)$, with $S_{\theta^0,P_X}(\theta)$ admitting a unique minimum in $\theta=\theta^0$.

ii) For $\omega_2(n,\rho)$ defined as $\omega_2(n,\rho) = \sup\{|S_n(\theta)-S_n(\theta')| : \|\theta-\theta'\|_{\ell^2} \leq \rho\}$, there exists a sequence ρ_k tending to 0, such that

$$\mathbb{E}(\omega_2(n,\rho_k)) = O(\rho_k). \tag{7.1}$$

Let us start with the proof of i) by writing that

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}\Psi(Z_k,Z_{k-1}), \text{ with } \Psi(Z_1,Z_0) = \frac{1}{2\pi}\int \frac{\left(\left(Z_1-f_\theta\right)^2 w\right)^*(t)e^{-itZ_0}}{f_\varepsilon^*(-t)}dt,$$

that is seen as a fonction of a strictly stationary and ergodic sequence of random variables. By the ergodic theorem and Assumption $(\mathbf{A_2})$ we conclude that for any $\theta\in\Theta$,

$$S_n(\theta) \xrightarrow[n\to\infty]{\mathbb{L}^1} \mathbb{E}(\psi(Z_1,Z_0)) = S_{\theta^0,P_X}(\theta).$$

It remains now to check that there exist two sequences ρ_k and ϵ_k tending to 0, such that (7.1) holds. The result follows by the assumption $(\mathbf{C_2})$ and by writing that

$$\sup_{\|\theta-\theta'\|_{\ell^2}\leq\rho} |S_n(\theta)-S_n(\theta')| \leq \sup_{\|\theta-\theta'\|_{\ell^2}\leq\rho} \| \theta-\theta' \|_{\ell^2} \sup_{\theta\in\Theta^0} \| S_n^{(1)}(\theta) \|_{\ell^2} . \tag{7.2}$$

□

7.2. **Proof of Theorems 4.2.** By using a Taylor expansion based on the smoothness properties of $\theta\mapsto wf_\theta$ and the consistency of $\widehat{\theta}$, we obtain

$$0 = S_n^{(1)}(\widehat{\theta}) = S_n^{(1)}(\theta^0) + S_n^{(2)}(\theta^0)(\widehat{\theta}-\theta^0) + R_n(\widehat{\theta}-\theta^0),$$

with R_n defined by

$$R_n = \int_0^1 [S_n^{(2)}(\theta^0+s(\widehat{\theta}-\theta^0)) - S_n^{(2)}(\theta^0)]ds. \tag{7.3}$$

This implies that

$$\widehat{\theta}-\theta^0 = -[S_n^{(2)}(\theta^0)+R_n]^{-1}S_n^{(1)}(\theta^0). \tag{7.4}$$

Consequently, we have to check the three following points.

i) $\sqrt{n}S_n^{(1)}(\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_{0,1})$;

ii) $S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S_{\theta^0,P_X}^{(2)}(\theta^0)$;

iii) R_n defined in (7.3) satisfies $R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Note that the covariance matrix $\Sigma_{0,1}$ in i) satisfies $\Sigma_{0,1} = \Sigma/4\pi^2$, with Σ defined by the equation (7.6) below. Consequently, according to ii) and iii), the covariance matrix Σ_1 satisfies

$$\Sigma_1 = \frac{1}{4\pi^2}(S^{(2)}_{\theta^0,P_X}(\theta^0))^{-1}\Sigma(S^{(2)}_{\theta^0,P_X}(\theta^0))^{-1}, \quad \text{with } \Sigma \text{ defined by (7.6).} \tag{7.5}$$

Proof of i)

Under assumption $(\mathbf{C_2})$,

$$\Big(\sqrt{n}S^{(1)}_n(\theta^0)\Big)_i = \frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n \int \left(\frac{\partial}{\partial\theta_i}((Z_k - f_\theta)^2)w\Big|_{\theta=\theta^0}\right)^*(t)\frac{e^{-itZk-1}}{f^*_\varepsilon(t)}dt.$$

We have thus to prove that

$$\frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n \int \big(-2(Z_k - f_{\theta^0})f^{(1)}_{\theta^0})w\big)^*(t)\frac{e^{-itZk-1}}{f^*_\varepsilon(t)}dt \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_{0,1}).$$

We first use that $\mathbb{E}(S_n(\theta)) = S_{\theta^0,P_X}(\theta)$ and thus $\mathbb{E}(S^{(1)}_n(\theta^0)) = S^{(1)}_{\theta^0,P_X}(\theta^0) = 0$. Next we write

$$\sqrt{n}S^{(1)}_n(\theta^0) = \sqrt{n}S^{(1)}_n(\theta^0) - \mathbb{E}[\sqrt{n}S^{(1)}_n(\theta^0)] = \frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n T_k$$

with $T_k = -2W_{k,1} + 2W_{k,2}$, and

$$\begin{aligned} W_{k,1} &= Z_k\int (f^{(1)}_{\theta^0}w)^*(t)\frac{e^{-itZ_{k-1}}}{f^*_\varepsilon(t)}dt - \mathbb{E}\left[Z_k\int (f^{(1)}_{\theta^0}w)^*(t)\frac{e^{-itZ_{k-1}}}{f^*_\varepsilon(t)}dt\right] \\ W_{k,2} &= \int (f_{\theta^0}f^{(1)}_{\theta^0}w)^*(t)\frac{e^{-itZ_{k-1}}}{f^*_\varepsilon(t)}dt - \mathbb{E}\left[\int (f_{\theta^0}f^{(1)}_{\theta^0}w)^*(t)\frac{e^{-itZ_{k-1}}}{f^*_\varepsilon(t)}dt\right]. \end{aligned}$$

Let $\mathcal{M}_1 = \sigma(X_0, X_1, \varepsilon_0, \varepsilon_1)$. According to Dedecker and Rio (2000), $n^{-1/2}\sum_{k=1}^n T_k$ converges to a centered Gaussian vector with covariance matrix

$$\Sigma = \mathrm{Cov}(T_1, T_1) + 2\sum_{k>1}\mathrm{Cov}(T_1, T_k)\,, \tag{7.6}$$

as soon as for any (p, q) in $\{1, \cdots, d\} \times \{1, \cdots, d\}$

$$\sum_{k=3}^\infty \mathbb{E}|(T_1)_p\mathbb{E}((T_k)_q|\mathcal{M}_1)| < \infty. \tag{7.7}$$

For any (p, q) in $\{1, \cdots, d\} \times \{1, \cdots, d\}$ and any $i, j \in \{1, 2\}$, we shall give an upper bound for

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)|\,.$$

We first notice that the sequence $(\varepsilon_k, \varepsilon_{k-1})$ is independent of $\mathcal{M}_1 \vee \sigma(X_k, X_{k-1})$. It follows that for $i, j \in \{1, 2\}$,

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)| = \mathbb{E}\left|(W_{1,i})_p\mathbb{E}((\tilde{W}_{k,j})_q|\mathcal{M}_1)\right|,$$

with

$$\begin{aligned}(\tilde W_{k,1})_q &= X_k\int \big(f^{(1)}_{\theta^0,q}w\big)^*(t)e^{-itX_{k-1}}dt - \mathbb{E}\left[X_k\int \big(f^{(1)}_{\theta^0,q}w\big)^*(t)e^{-itX_{k-1}}dt\right]\\ (\tilde W_{k,2})_q &= \int \big(f_{\theta^0}f^{(1)}_{\theta^0,q}w\big)^*(t)e^{-itX_{k-1}}dt - \mathbb{E}\left[\int \big(f_{\theta^0}f^{(1)}_{\theta^0,q}w\big)^*(t)e^{-itX_{k-1}}dt\right].\end{aligned}$$

Next, since $\mathbb{P}_{(X_{k-1},X_k)|\sigma(\varepsilon_0,\varepsilon_1,X_0,X_1)} = \mathbb{P}_{(X_{k-1},X_k)|\sigma(X_1)}$, we infer that

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)| = \mathbb{E}\left|(W_{1,i})_p\mathbb{E}((\tilde W_{k,j})_q|X_1)\right|.$$

Next we use that under Condition ($\mathbf{C_2}$),

$$\begin{aligned}|(W_{1,1})_p| &\le |Z_1|\int\left|\big(f^{(1)}_{\theta^0,p}w\big)^*(t)\frac{e^{-itZ_0}}{f^*_\varepsilon(t)}\right|dt + \mathbb{E}\left\{|Z_1|\int\left|\big(f^{(1)}_{\theta^0,p}w\big)^*(t)\frac{e^{-itZ_0}}{f^*_\varepsilon(t)}\right|dt\right\}\\ &\le |Z_1|\int\left|\big(f^{(1)}_{\theta^0,p}w\big)^*(t)\frac{1}{f^*_\varepsilon(t)}\right|dt + \mathbb{E}\left\{|Z_1|\int\left|\big(f^{(1)}_{\theta^0,p}w\big)^*(t)\frac{1}{f^*_\varepsilon(t)}\right|dt\right\}\\ &\le C_1(|Z_1|+\mathbb{E}(|Z_1|)).\end{aligned}$$

In the same way we get that $|(W_{1,2})_p| \le C_2$.

Now, since ε_1 is independent of X_1, for $j\in\{1,2\}$

$$\begin{aligned}\mathbb{E}\left|(W_{1,1})_p\mathbb{E}((\tilde W_{k,j})_q|X_1)\right| &\le C_1\mathbb{E}\left[(|Z_1|+\mathbb{E}(|Z_1|))\left|\mathbb{E}((\tilde W_{k,j})_q|X_1)\right|\right]\\ &\le C\mathbb{E}\left[(|X_1|+\mathbb{E}(|X_1|))\left|\mathbb{E}((\tilde W_{k,j})_q|X_1)\right|\right].\end{aligned}\tag{7.8}$$

In the same way

$$\mathbb{E}\left|(W_{1,2})_p\mathbb{E}((\tilde W_{k,j})_q|X_1)\right| \le C\mathbb{E}\left|\mathbb{E}((\tilde W_{k,j})_q|X_1)\right|.\tag{7.9}$$

Note that

$$\mathbb{E}\left[(|X_1|+\mathbb{E}(|X_1|))\left|\mathbb{E}((\tilde W_{k,1})_q|X_1)\right|\right] = \mathrm{Cov}((|X_1|+\mathbb{E}(|X_1|))\mathrm{sign}(\mathbb{E}((\tilde W_{k,1})_q|X_1)),(\tilde W_{k,1})_q).$$

Now, we use the covariance inequality (2.1). Note first that

$$(|X_1|+\mathbb{E}(|X_1|))\mathrm{sign}(\mathbb{E}((\tilde W_{k,1})_q|X_1)) \le |X_1|+\mathbb{E}(|X_1|)$$

and

$$|(\tilde W_{1,1})_q| \le D(|X_1|+\mathbb{E}(|X_1|)).$$

Since $(X_i)_{i\ge0}$ is a strictly stationary Markov chain, it is well known that

$$\alpha(\sigma(X_1),\sigma(X_{k-1},X_k)) = \alpha(\sigma(X_1),\sigma(X_{k-1})) = \alpha_{\mathbf{X}}(k-2).\tag{7.10}$$

Hence, applying (2.1),

$$\mathbb{E}\left|(W_{1,1})_p\mathbb{E}((\tilde W_{k,1})_q|X_1)\right| \le C\int_0^{\alpha_{\mathbf{X}}(k-2)}Q^2_{|X_1|}(u)du.$$

We conclude that

$$\sum_{k\ge3}\mathbb{E}\,|(W_{1,1})_p\mathbb{E}((W_{k,1})_q|\mathcal{M}_1)| \le C\sum_{k\ge3}\int_0^{\alpha_{\mathbf{X}}(k-2)}Q^2_{|X_1|}(u)du.$$

Finally, using similar arguments for the three quantities $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,2})_p\mathbb{E}((W_{k,1})_q|\mathcal{M}_1)|$, $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,1})_p\mathbb{E}((W_{k,2})_q|\mathcal{M}_1)|$ and $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,2})_p\mathbb{E}((W_{k,2})_q|\mathcal{M}_1)|$ we conclude that

$$\sqrt{n}S_n^{(1)}(\theta^0) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0,\Sigma/(4\pi^2))$$

as soon as

$$\sum_{k\geq 1}\int_0^{\alpha_{\mathbf{X}}(k)} Q^2_{|X_1|}(u)du < \infty.$$

$\square$

Proof of ii)

Under condition $(\mathbf{C_3})$, for $j,k=1,\cdots,d$,

$$\text{(7.11)}\quad \left(S_n^{(2)}(\theta)\right)_{j,k} = -\frac{2}{n}\sum_{\ell=1}^{n}\int\left(-2Z_\ell\frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta w)+\frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta^2 w)\right)^*(t)\frac{e^{-itZ_{\ell-1}}}{f_\varepsilon^*(-t)}dt$$

and by applying the ergodic theorem we get that

$$S_n^{(2)}(\theta^0) \underset{n\to\infty}{\xrightarrow{\mathbb{P}}} S^{(2)}_{\theta^0,P_X}(\theta^0).$$

$\square$

Proof of iii)

Starting from (7.3) and (7.11), the point iii) follows from the assumption $(\mathbf{C_4})$ on the properties of the derivatives at order 3 of wf_θ and wf_θ^2. $\square$

7.3. **Proof of Theorem 4.3.** We follow the proofs of Theorems 4.2 and keep the same notations. We have to check that the condition (7.7) holds. We start from the inequalities (7.8) and (7.9). For clarity, let us write

$$(\tilde{W}_{k,1})_q = (\tilde{W}_{k,1})_q(X_k, X_{k-1})\,.$$

Let ψ_M be the truncating function defined by $\psi_M(x) = (x\wedge M)\vee(-M)$. Applying (2.2), let (X_k^*, X_{k-1}^*) be the random variable distributed as (X_k, X_{k-1}) and independent of X_1 such that

$$\frac{1}{2}(\|X_k - X_k^*\|_1 + \|X_{k-1}-X_{k-1}^*\|_1) = \tau(\sigma(X_1),(X_{k-1},X_k)) \leq \tau_{X,2}(k-2)\,.$$

Define the constants K_1 and K_2 by

$$K_1 = \int\left|(f^{(1)}_{\theta^0,q}w)^*(t)\right|dt<\infty\,,\quad K_2=\int|t|\left|(f^{(1)}_{\theta^0,q}w)^*(t)\right|dt<\infty\,.$$

Clearly

$$|X_1\mathbb{E}((\tilde{W}_{k,1})_q(X_k,X_{k-1})|X_1)| \leq M|\mathbb{E}((\tilde{W}_{k,1})_q(X_k,X_{k-1})|X_1)|+K_2|X_1|\mathbf{1}_{|X_1|>M}(|X_k|+\mathbb{E}(|X_k|))\,.$$

Now, since (X_k^*, X_{k-1}^*) is independent of X_1, one has that

$$|\mathbb{E}((\tilde{W}_{k,1})_q(X_k,X_{k-1})|X_1)| = |\mathbb{E}((\tilde{W}_{k,1})_q(X_k,X_{k-1})-(\tilde{W}_{k,1})_q(X_k^*,X_{k-1}^*)|X_1)|\,.$$

By definition of $(\tilde W_{k,1})_q(X_k, X_{k-1})$, there exists a constant C such that

$$|(\tilde W_{k,1})_q(X_k, X_{k-1}) - (\tilde W_{k,1})_q(X_k^*, X_{k-1}^*) - ((\tilde W_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde W_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*))| \leq C(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M}).$$

Hence

$$\begin{aligned} |\mathbb{E}((\tilde W_{k,1})_q(X_k, X_{k-1})|X_1)| &\leq |\mathbb{E}((\tilde W_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde W_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*)|X_1)| \\ &+ C(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M})\,. \end{aligned}$$

Since ψ_M is 1-Lipschitz and bounded by M, and since $x \to \exp(itx)$ is $|t|$-Lipschitz and bounded by 1, under Condition $(\mathbf{C_5})$, one has

$$|(\tilde W_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde W_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*)| \leq MK_2|X_{k-1} - X_{k-1}^*| + K_1|X_k - X_k^*|\,.$$

It follows that

$$\begin{aligned} |X_1\mathbb{E}((\tilde W_{k,1})_q(X_k, X_{k-1})|X_1)| &\leq K_2|X_1|\mathbf{1}_{|X_1|>M}(|X_k| + \mathbb{E}(|X_k|)) \\ &+ M^2K_2|X_{k-1} - X_{k-1}^*| + MK_1|X_k - X_k^*|) \\ &+ CM(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M})\,. \end{aligned}$$

Using that

$$|X_1|\mathbf{1}_{|X_1|>M}|X_k| \leq \frac{3}{2}X_1^2\mathbf{1}_{|X_1|>M} + \frac{1}{2}X_k^2\mathbf{1}_{|X_k|>M}\,,$$

we infer from (7.8) with $j = 1$ that there exists a positive constant K such that

$$\mathbb{E}\left[\left|(W_{1,1})_p\mathbb{E}((\tilde W_{k,1})_q|X_1)\right|\right] \leq K(L(M^2) + M(M+1)\tau_{2,X}(k-2))\,,$$

where $L(t) = \mathbb{E}(X_0^2\mathbf{1}_{X_0^2>t})$. Let then $G(t) = t^{-1}L(t)$, and let G^{-1} be the inverse cadlag of G. Choose then $M^2 = G^{-1}(\tau_{2,X}(k-2))$. We obtain that

$$\mathbb{E}\left[\left|(W_{1,1})_p\mathbb{E}((\tilde W_{k,1})_q|X_1)\right|\right] \leq 2K(2G^{-1}(\tau_{2,X}(k-2))\tau_{2,X}(k-2) + \sqrt{G^{-1}(\tau_{2,X}(k-2))}\tau_{2,X}(k-2))\,.$$

It follows that

$$\sum_{k\geq 3}\mathbb{E}\left[\left|(W_{1,1})_p\mathbb{E}((\tilde W_{k,1})_q|X_1)\right|\right] < \infty \quad \text{as soon as} \quad \sum_{k>0} G^{-1}(\tau_{2,X}(k))\tau_{2,X}(k) < \infty\,.$$

Easier control holds for the other terms in (7.8) and (7.9). Consequently (7.7) holds as soon as (4.11) holds, and the proof is complete.

7.4. **Consistency under α-mixing condition.**

Proof of Theorem 4.4. The proof of the consistency under the assumptions of Theorem 4.4 is quite different from the proof of the consistency under Conditions $(\mathbf{C_1})$-$(\mathbf{C_2})$ in Theorem 4.1. This comes from the fact that $S_n(\theta)$ is now a triangular array of the form

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^n \Psi_n(Z_k, Z_{k-1}) \text{ with } \Psi_n(Z_1, Z_0) = \frac{1}{2\pi}\int \frac{\left((Z_1 - f_\theta)^2 w\right)^*(t)e^{-itZ_0}K_{C_n}^*(t)}{f_\varepsilon^*(-t)}dt.$$

In this context we show that

i) For all θ in Θ, $\mathbb{E}[(S_n(\theta) - S_{\theta^0,P_X}(\theta))^2] = o(1)$ as $n \to \infty$.
ii) The control (7.1) holds.
Note first that ii) follows from the upper bound (7.2) and Assumption ($\mathbf{A_4}$).
For the proof of i) we check that for all $\theta \in \Theta$,

$$\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta) = o(1) \quad \text{and } \operatorname{Var}(S_n(\theta)) = o(1), \quad \text{as } n \to \infty. \tag{7.12}$$

Proof of the first part of (7.12). Since $Z_0 = X_0 + \varepsilon_0$, with ε_0 independent of (Z_1, X_0), it follows that

$$\mathbb{E}[S_n(\theta)] = \mathbb{E}\left[\left((Z_1 - f_\theta)^2\, w\right) \star K_{n,C_n}(Z_0)\right] = \mathbb{E}\left[\left((Z_1 - f_\theta)^2\, w\right) \star K_{C_n}(X_0)\right]$$

and hence

$$\begin{aligned}\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta) &= \frac{1}{2\pi}\iint (f^2_{\theta^0}(x) + \sigma^2_\xi + \sigma^2_\varepsilon) e^{-iux} w^*(u)(K^*_{C_n} - 1)(u)du P_X(dx)\\ &\quad -\frac{1}{\pi}\iint f_{\theta^0}(x) e^{-iux}(f_\theta w)^*(u)(K^*_{C_n} - 1)(u)du P_X(dx)\\ &\quad +\frac{1}{2\pi}\iint e^{-iux}(f^2_\theta w)^*(u)(K^*_{C_n} - 1)(u)P_X(dx)du.\end{aligned}$$

Now, arguing as in Butucea and Taupin (2008) we get that $|\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta)|^2 = o(1)$.

Proof of the second part of (7.12). Using that the Z_i's are strictly stationary we get that

$$\begin{aligned}\operatorname{Var}[S_n(\theta)] &= \operatorname{Var}\left[n^{-1}\sum_{k=1}^{n}((Z_k - f_\theta)^2\, w) \star K_{n,C_n}(Z_{k-1})\right]\\ &\leq \frac{1}{n}\operatorname{Var}\,(A_{1,0}) + \frac{2}{n}\sum_{i=2}^{n} |\operatorname{Cov}\big(A_{1,0}, A_{i,i-1}\big)|\\ &\leq \frac{3}{n}\operatorname{Var}\,(A_{1,0}) + \frac{2}{n}\sum_{k=3}^{n} |\operatorname{Cov}\big(A_{1,0}, A_{k,k-1}\big)|\end{aligned}$$

with

$$A_{k,k-1} = \Big[\big((Z_k - f_\theta)^2\, w\big) \star K_{n,C_n}(Z_{k-1})\Big].$$

Arguing as in Butucea and Taupin (2008) we obtain that $\lim_{n\to\infty} n^{-1}\operatorname{Var}\,(A_{1,0}) = 0$. It remains to study

$$\frac{1}{n}\sum_{k=3}^{n} |\operatorname{Cov}\big(A_{1,0}, A_{k,k-1}\big)|.$$

Lemma 7.1. *Let Ψ such that $\mathbb{E}(|\Psi(Z)|) < \infty$ and let Φ be an integrable function. Let*

$$B_{k,k-1} = \Psi(Z_k)\Phi \star K_{n,C_n}(Z_{k-1}).$$

Then for $k \geq 3$

$$\begin{aligned}\operatorname{Cov}(B_{k,k-1}, B_{1,0}) &= \operatorname{Cov}[\Psi(Z_k)\Phi \star K_{C_n}(X_{k-1}), \Psi(Z_1)\Phi \star K_{C_n}(X_0)]\\ &= \frac{1}{(2\pi)^2}\iint \Phi^*(t)\Phi^*(-s)\operatorname{Cov}\big(\Psi(Z_k)e^{itX_{k-1}}, \Psi(Z_1)e^{isX_0}\big)K^*_{C_n}(t)K^*_{C_n}(-s)dtds.\end{aligned}$$

Proof of Lemma 7.1: By stationarity we write

$$\mathrm{Cov}\big(B_{k,k-1},B_{1,0}\big)=\mathbb{E}(B_{k,k-1}B_{1,0})-\mathbb{E}(B_{k,k-1})\mathbb{E}(B_{1,0})=\mathbb{E}(B_{k,k-1}B_{1,0})-\big(\mathbb{E}(B_{1,0})\big)^2.$$

Now, we use that the sequences $(X_k)_{k\in\mathbb{Z}}$ and $(\varepsilon_k)_{k\in\mathbb{Z}}$ are independent. This implies that (Z_1,X_0) is independent of ε_0 and thus

$$\mathbb{E}(B_{1,0})=\frac{1}{2\pi}\int\Phi^*(t)\mathbb{E}[\Psi(Z_1)e^{itZ_0}]\frac{K^*_{C_n}(-t)}{f^*_\varepsilon(-t)}dt=\frac{1}{2\pi}\int\Phi^*(t)\mathbb{E}[\Psi(Z_1)e^{itX_0}]K^*_{C_n}(-t)dt.$$

In the same way, for $k\geq 3$,

$$\begin{aligned}\mathbb{E}(B_{k,k-1}B_{1,0}) &= \frac{1}{(2\pi)^2}\iint\Phi^*(-s)\Phi^*(t)\mathbb{E}[\Psi(Z_k)\Psi(Z_1)e^{-itZ_{k-1}}e^{isZ_0}]\frac{K^*_{C_n}(-t)}{f^*_\varepsilon(-t)}\frac{K^*_{C_n}(s)}{f^*_\varepsilon(s)}dtds\\ &= \frac{1}{(2\pi)^2}\iint\Phi^*(-s)\Phi^*(t)\mathbb{E}[\Psi(Z_k)\Psi(Z_1)e^{-itX_{k-1}}e^{isX_0}]K^*_{C_n}(-t)K^*_{C_n}(s)dtds\end{aligned}$$

and the lemma is proved. □

It follows from Lemma 7.1 that for $k\geq 3$,

$$\mathrm{Cov}\big(A_{k,k-1},A_{1,0}\big)=\mathrm{Cov}\Big[\big((Z_k-f_\theta)^2\,w\big)\star K_{C_n}(X_{k-1}),\big((Z_1-f_\theta)^2\,w\big)\star K_{C_n}(X_0)\Big]=C_1+C_2+C_3,$$

with

$$\begin{aligned}C_1 &= \frac{1}{(2\pi)^2}\iint\mathrm{Cov}(e^{-itX_{k-1}},e^{isX_0})(wf_\theta^2)^*(-t)(wf_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(-t)dtds,\\ C_2 &= \frac{-1}{\pi^2}\iint\mathrm{Cov}(X_ke^{-itX_{k-1}},X_1e^{isX_0})(wf_\theta)^*(-t)(wf_\theta)^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds,\\ C_3 &= \frac{1}{(2\pi)^2}\iint\mathrm{Cov}[(X_k^2+\varepsilon_k^2)e^{-itX_{k-1}},(X_1^2+\varepsilon_1^2)e^{isX_0}]w^*(-t)w^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds.\end{aligned}$$

Easy computations give

$$\begin{aligned}&\mathrm{Cov}[(X_k^2+\varepsilon_k^2)e^{-itX_{k-1}},(X_1^2+\varepsilon_1^2)e^{isX_0}]=\\ &\qquad\mathrm{Cov}(X_k^2e^{-itX_{k-1}},X_1^2e^{isX_0})+\sigma_\varepsilon^2\mathrm{Cov}(X_k^2e^{-itX_{k-1}},e^{isX_0})\\ &\qquad\qquad+\sigma_\varepsilon^2\mathrm{Cov}(e^{-itX_{k-1}},X_1^2e^{isX_0})+\sigma_\varepsilon^4\mathrm{Cov}(e^{-itX_{k-1}},e^{isX_0}),\end{aligned}$$

which induces the decomposition $\mathrm{Cov}\big(A_{k,k-1}, A_{1,0}\big) = E_1 + E_2 + E_3 + E_4$ with

$$\begin{aligned}
E_1 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(e^{-itX_{k-1}}, e^{isX_0})(wf_\theta^2 + \sigma_\varepsilon^4 w)^*(-t)(wf_\theta^2 + \sigma_\varepsilon^4 w)^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds,\\
E_2 &= C2 = \frac{-1}{\pi^2}\iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})(wf_\theta)^*(-t)(wf_\theta)^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds\\
E_3 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1^2 e^{isX_0})w^*(-t)w^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds\\
E_4 &= \frac{\sigma_\varepsilon^2}{(2\pi)^2}\Big[\iint \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{isX_0})w^*(-t)w^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds\\
&\qquad + \iint \mathrm{Cov}(e^{-itX_{k-1}}, X_1^2 e^{isX_0})w^*(-t)w^*(s)K^*_{C_n}(-t)K^*_{C_n}(s)dtds\Big].
\end{aligned}$$

Using (2.1) and (7.10), we have the upper bounds

$$\begin{aligned}
|\mathrm{Cov}(e^{-itX_{k-1}}, e^{isX_0})| &\le C\alpha_{\mathbf{X}}(k-1)\\
|\mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X|}(u)du\\
|\mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1^2 e^{isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^4_{|X|}(t)dt\\
|\mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^3_{|X|}(t)dt\\
|\mathrm{Cov}(e^{-itX_{k-1}}, X_1^2 e^{isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^3_{|X|}(t)dt\,.
\end{aligned}$$

Since $\mathbb{E}(X_1^4) < \infty$ and $\lim_{k\to\infty}\alpha_{\mathbf{X},2}(k) = 0$, we infer that $\lim_{k\to\infty}|\,\mathrm{Cov}\big(A_{k,k-1}, A_{1,0}\big)| = 0$. Now, by Cesaro's mean convergence theorem

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=3}^{n}|\,\mathrm{Cov}\big(A_{1,0}, A_{k,k-1}\big)| = 0\,.$$

This completes the proof of the consistency.

7.5. **Rate of convergence under α-mixing condition.**

Proof of 1) in Theorem 4.5. Starting from the decomposition (7.4) we shall check the three following points.

i) $\mathbb{E}\left[(S_n^{(1)}(\theta^0) - S^{(1)}_{\theta^0,P_X}(\theta^0))(S_n^{(1)}(\theta^0) - S^{(1)}_{\theta^0,P_X}(\theta^0))^\top\right] = O[\varphi_n\varphi_n^\top]$

ii) $S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S^{(2)}_{\theta^0,P_X}(\theta^0)$;

iii) R_n defined in (7.3) satisfies $R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

The rate of convergence of $\widehat{\theta}$ is thus given by the order of

$$\mathbb{E}\left[(S_n^{(1)}(\theta^0) - S^{(1)}_{\theta^0,P_X}(\theta^0))(S_n^{(1)}(\theta^0) - S^{(1)}_{\theta^0,P_X}(\theta^0))^\top\right].$$

Proof of i)

We first write

$$\begin{aligned}\left(S_n^{(1)}(\theta)\right)_i &= \frac{1}{n}\sum_{k=1}^n \frac{\partial}{\partial\theta_i}\left[((Z_k-f_\theta)^2 w)\star K_{n,C_n}(Z_{k-1}) - \mathbb{E}[(Z_k-f_\theta(X_{k-1}))^2 w(X_{k-1})]\right]\\ &= \frac{1}{n}\sum_{k=1}^n \left(\frac{\partial}{\partial\theta_i}(Z_k-f_\theta)^2 w\star K_{n,C_n}(Z_{k-1}) - \mathbb{E}\left[\frac{\partial}{\partial\theta_i}(Z_k-f_\theta(X_{k-1}))^2 w(X_{k-1})\right]\right).\end{aligned}$$

Study of the bias. As in Butucea and Taupin (2008), we get that

$$\left|\mathbb{E}\left[\left(S_n^{(1)}(\theta^0)\right)_j\right]\right| \le C_1(f_{\theta^0}, w, f_\varepsilon)\min\left[B^{[1]}_{n,j}B^{[2]}_{n,j}\right],$$

for $B^{[q]}_{n,j}$, $q=1,2$, defined in Theorem 4.5.

Study of the variance. For the variance term, note first that

$$\operatorname{Var}\left(\left(S_n^{(1)}(\theta^0)\right)_j\right) \le \frac{3}{n}\operatorname{Var}(D_{1,0}) + \frac{2}{n}\sum_{k=3}^n |\operatorname{Cov}(D_{1,0}, D_{k,k-1})|,$$

with

$$D_{k,k-1} = \Big(\big(-2Z_k f^{(1)}_{\theta^0,j} + 2f_\theta f^{(1)}_{\theta^0,j}\big)w\Big)\star K_{n,C_n}(Z_{k-1}).$$

The first part in $\operatorname{Var}\big[(S_n^{(1)}(\theta^0))_j\big]$ is controlled as in Butucea and Taupin (2008) by

$$\frac{1}{n}\operatorname{Var}(D_{1,0}) \le \frac{C(\sigma_\xi^2, f_{\theta^0}, f^{(1)}_{\theta^0,j}, w, f_\varepsilon)}{n}\min\{V^{[1]}_{n,j}(\theta^0), V^{[2]}_{n,j}(\theta^0)\} \tag{7.13}$$

with $V^{[q]}_{n,j}$, $q=1,2$ defined in Theorem 4.5. We now control the term

$$\frac{1}{n}\sum_{k=3}^n |\operatorname{Cov}(D_{1,0}, D_{k,k-1})|.$$

Applying again Lemma 7.1, we obtain that

$$\operatorname{Cov}(D_{1,0}, D_{k,k-1}) = F_1+F_2+F_3+F_4$$

with

$$\begin{aligned}F_1 &= \frac{4}{2\pi}\iint \operatorname{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(-t)\big(f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(-s)dtds\\ F_2 &= \frac{4}{2\pi}\iint \operatorname{Cov}(e^{-itX_{k-1}}, e^{isX_0})\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(-t)\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(-s)dtds\\ F_3 &= \frac{4}{2\pi}\iint \operatorname{Cov}(X_k e^{-itX_{k-1}}, e^{isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(-t)\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(-s)dtds\\ F_4 &= \frac{4}{2\pi}\iint \operatorname{Cov}(e^{-itX_{k-1}}, X_1 e^{isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(-t)\big(f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(-s)dtds.\end{aligned}$$

Using (2.1) and (7.10) we have the upper bounds

$$\begin{aligned} |\mathrm{Cov}(e^{-itX_{k-1}}, e^{isX_0})| &\leq C\alpha_{\mathbf{X}}(k-1) \\ |\mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X|}(u)du \\ |\mathrm{Cov}(X_k e^{itX_{k-1}}, e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-1)} Q_{|X|}(u)du \\ |\mathrm{Cov}(e^{itX_{k-1}}, X_1 e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q_{|X|}(u)du. \end{aligned}$$

Since $\mathbb{E}(X_1^4) < \infty$, we infer that $Q_{|X|}(u) \leq Cu^{-1/4}$, and consequently all the covariance terms are $O(\sqrt{\alpha_{\mathbf{X}}(k)})$. Finally, if $\sum_{k>0}\sqrt{\alpha_{\mathbf{X}}(k)} < \infty$, then

$$\frac{1}{n}\sum_{k=3}^{n}|\mathrm{Cov}(D_{1,0}, D_{k,k-1})| \leq \frac{C}{n}\,.$$

This, together with (7.13), implies that

$$\mathrm{Var}\left[\big(S_n^{(1)}(\theta^0)\big)_j\right] \leq \frac{C}{n}\min\{V_{n,j}^{[1]}(\theta^0), V_{n,j}^{[2]}(\theta^0)\}\,.$$

Proof of ii)

The proof of **ii)** starts form the expression of the second derivative of the estimation criterion

$$(7.14)\Big(S_n^{(2)}(\theta)\Big)_{j,k} = -\frac{2}{n}\sum_{\ell=1}^{n}\int\left(-2Z_\ell\frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta w) + \frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta^2 w)\right)^*(t)\frac{K^*_{C_n}(t)e^{-itZ_{\ell-1}}}{f^*_\varepsilon(-t)}dt.$$

Following the same lines as for the consistency we prove that

$$S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S^{(2)}_{\theta^0,P_X}(\theta^0).$$

□

Proof of iii)

The proof of **iii)** follows from (7.14), from the smoothness properties of wf_θ and from Assumption ($\mathbf{A_4}$).

□

Proof of 2) in Theorem 4.5. The proof of 2) in theorem 4.5 is quite similar to the proof of 1). The main differences appear in the control of the bias and variance of $S_n^{(1)}(\theta^0)$. More precisely, we start from

$$S_n^{(1)}(\theta) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\partial}{\partial\theta}(Z_k - f_\theta)^2 w\right)\star K_{n,C_n}(Z_{k-1}) - \mathbb{E}\left[\frac{\partial}{\partial\theta}(Z_k - f_\theta(X_{k-1}))^2 w(X_{k-1})\right].$$

Study of the bias Since $P_{Z,X}(z,z) = P_X(x)f_\varepsilon(z-x)$ we obtain that $\mathbb{E}[S_n^{(1)}(\theta^0)] - S^{(1)}_{\theta^0,P_X}(\theta^0)$ is equal to

$$-2\mathbb{E}\left[f_{\theta^0}(X_0)(f^{(1)}_{\theta^0}w)\star K_{C_n}(X_0) - f_{\theta^0}(X_0)f^{(1)}_{\theta^0}(X_0)w(X_0)\right] \\ + 2\mathbb{E}\left[(f^{(1)}_{\theta^0}f_{\theta^0}w)\star K_{C_n}(X_0) - (f^{(1)}_{\theta^0}f_{\theta^0}w)(X_0)\right],$$

that is $\mathbb{E}[S_n^{(1)}(\theta^0)] - S^{(1)}_{\theta^0,P_X}(\theta^0)$ is equal to

$$-2\iint f_{\theta^0}(x)e^{-iux}(f^{(1)}_{\theta^0}w)^*(u)(K^*_{C_n}(u)-1)P_X(dx)\,du \\ + 2\iint e^{-iux}(f_{\theta^0}f^{(1)}_{\theta^0}w)^*(u)(K^*_{C_n}(u)-1)P_X(dx)\,du.$$

It follows that for $j = 1,\cdots,d$,

$$\left|\mathbb{E}[(S_n^{(1)}(\theta^0))_j] - (S^{(1)}_{\theta^0,P_X}(\theta^0))_j\right| \\ \leq \mathbb{E}|f_{\theta^0}(X_0)|\int |(f^{(1)}_{\theta^0,j}w)^*(u)(K^*_{C_n}(u)-1)|du + \int |(f_{\theta^0}f^{(1)}_{\theta^0,j}w)^*(u)(K^*_{C_n}(u)-1)|du.$$

Study of the variance For the study of the variance we combine the proof in Butucea and Taupin (2008) and the proof of **1)** of Theorem 4.5. For these reasons we only give a sketch of the proof, with details only for specific parts. As for the proof of **1)** we start from

$$\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] = \frac{1}{n}\mathrm{Var}\left[\left(\frac{\partial[-2Z_kf_\theta w + f_\theta^2 w]}{\partial\theta_j}\,|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_{k-1})\right] \\ + \frac{2}{n^2}\sum_{1\leq j<k\leq n}\mathrm{Cov}(D_{k,k-1},D_{j,j-1}),$$

with $D_{k,k-1}$ defined in (7.5). The control of $(2/n^2)\sum_{1\leq j<k\leq n}\mathrm{Cov}(D_{k,k-1},D_{j,j-1})$ is done as in the proof of **1)**. We now control of the first part of $\mathrm{Var}\big((S_n^{(1)}(\theta^0))_j\big)$.

$$\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] \leq \frac{C}{n}\mathbb{E}\left[\left(\frac{\partial[-2Z_if_\theta w + f_\theta^2 w]}{\partial\theta_j}\,|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i)\right]^2.$$

In other words,

$$\begin{aligned}\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] &\leq \frac{C}{n}\mathbb{E}\left[\left(Z_if^{(1)}_{\theta^0}w + f_{\theta^0}f^{(1)}_{\theta^0}w\right)\star K_{n,C_n}(Z_i)\right]^2 \\ &= \frac{C}{n}\mathbb{E}\left[\left((f^2_{\theta^0}(X_0)+\sigma^2_\xi)f^{(1)}_{\theta^0}w + f_{\theta^0}f^{(1)}_{\theta^0}w\right)\star K_{n,C_n}(Z_0)\right]^2.\end{aligned}$$

Now, write that

$$\mathbb{E}\left[\left((f^2_{\theta^0}(X_0)+\sigma^2_\xi)f^{(1)}_{\theta^0}w + f_{\theta^0}f^{(1)}_{\theta^0}w\right)\star K_{n,C_n}(Z_0)\right]^2 = II_1 + II_2,$$

with

$$\begin{aligned} II_1 &= \iint f_\varepsilon(z-x)(f^2_{\theta^0}(x)+\sigma^2_\xi)\left(\int (f^{(1)}_{\theta^0}w)(u)K_{n,C_n}(z-u)du\right)^2 P_X(dx)dz \\ II_2 &= \iint f_\varepsilon(z-x)\left(\int (f_{\theta^0}f^{(1)}_{\theta^0}w)(u)K_{n,C_n}(z-u)du\right)^2 P_X(dx)dz. \end{aligned}$$

We apply Hölder Inequality and obtain that

$$II_1 \leq \sup_{z\in\mathbb{R}} \mathbb{E}[(f^2_{\theta^0}(X_0)+\sigma^2_\xi)f_\varepsilon(z-X_0)] \parallel (f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2,$$

and that II_1 is also less than

$$\mathbb{E}[(f^2_{\theta^0}(X_0)+\sigma^2_\xi)] \parallel (f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_\infty^2$$

In the same way we have

$$II_2 \leq \sup_{z\in\mathbb{R}} \mathbb{E}[f_\varepsilon(z-X_0)] \parallel (f_{\theta^0}f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2, \text{ and } II_2 \leq \parallel (f_{\theta^0}f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_\infty^2 .$$

Consequently we have

$$\mathrm{Var}\left[(S^{(1)}_n(\theta^0))_j\right] \leq \frac{C(\sigma^2_\xi, f_{\theta^0}, f_\varepsilon)}{n}\left[\parallel (f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2 + \parallel (f_{\theta^0}f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2\right] \tag{7.15}$$

and

$$\mathrm{Var}\left[(S^{(1)}_n(\theta^0))_j\right] \leq \frac{C_1(f_{\theta^0})}{n}\left[\parallel (f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_2^2 + \parallel (f_{\theta^0}f^{(1)}_{\theta^0}w)\star K_{n,C_n} \parallel_1^2\right]. \tag{7.16}$$

By combining (7.15) and (7.16), we get that

$$\mathrm{Var}\left[(S^{(1)}_n(\theta^0))_j\right] \leq \frac{C((f_{\theta^0}, \sigma^2_\xi, f_\varepsilon)}{n}\min\{V^{[1]}_{n,j}(\theta^0), V^{[2]}_{n,j}(\theta^0)\}$$

with $V^{[q]}_{n,j}$, $q=1,2$ defined in Theorem 4.5.

□

References

Anderson, B. D. O. and M. Deistler (1984). Identifiability in dynamic errors-in-variables models. *J. Time Ser. Anal. 5*(1), 1–13.

Ango Nze, P. (1998). Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne. *C. R. Acad. Sci. Paris Sér. I Math. 326*(3), 371–376.

Bickel, P. J., Y. Ritov, and T. Rydén (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist. 26*(4), 1614–1635.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in probability and statistics (Oberwolfach, 1985)*, Volume 11 of *Progr. Probab. Statist.*, pp. 165–192. Boston, MA: Birkhäuser Boston.

Brockwell, P. J. and R. A. Davis (1991). *Time series: theory and methods* (Second ed.). Springer Series in Statistics. New York: Springer-Verlag.

Butucea, C. and M.-L. Taupin (2008). New M-estimators in semiparametric regression with errors in variables. *Ann. Inst. Henri Poincar, Probab. Stat. 44*(3), 393–421.

Chanda, K. C. (1995). Large sample analysis of autoregressive moving-average models with errors in variables. *J. Time Ser. Anal. 16*(1), 1–15.

Chanda, K. C. (1996). Asymptotic properties of estimators for autoregressive models with errors in variables. *Ann. Statist. 24*(1), 423–430.

Comte, F. and M.-L. Taupin (2001). Semiparametric estimation in the (auto)-regressive β-mixing model with errors-in-variables. *Math. Methods Statist. 10*(2), 121–160.

Costa, M. and T. Alpuim (2010). Parameter estimation of state space models for univariate observations. *J. Statist. Plann. Inference 140*(7), 1889–1902.

Dedecker, J. and C. Prieur (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Relat. Fields 132*(2), 203–236.

Dedecker, J. and E. Rio (2000). On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist. 36*(1), 1–34.

Douc, R. and C. Matias (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli 7*(3), 381–420.

Douc, R., E. Moulines, J. Olsson, and R. van Handel (2011). Consistency of the maximum likelihood estimator for general hidden markov models. *Ann. Statist. 39*, 474–513.

Douc, R., É. Moulines, and T. Rydén (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist. 32*(5), 2254–2304.

Fuh, C.-D. (2006). Efficient likelihood estimation in state space models. *Ann. Statist. 34*(4), 2026–2068.

Genon-Catalot, V. and C. Laredo (2006). Leroux's method for general hidden Markov models. *Stochastic Process. Appl. 116*(2), 222–243.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probability 10*, 130–145, corrections, ibid. 10 (1973), 913.

Jensen, J. L. and N. V. Petersen (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist. 27*(2), 514–535.

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl. 40*(1), 127–143.

Mokkadem, A. (1985). Le modèle non linéaire AR(1) général. Ergodicité et ergodicité géométrique. *C. R. Acad. Sci. Paris Sér. I Math. 301*(19), 889–892.

Na, S., S. Lee, and H. Park (2006). Sequential empirical process in autoregressive models with measurement errors. *J. Statist. Plann. Inference 136*(12), 4204–4216.

Nowak, E. (1985). Global identification of the dynamic shock-error model. *J. Econometrics 27*(2), 211–219.

Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist. 29*(4), 587–597.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA 42*, 43–47.

Staudenmayer, J. and J. P. Buonaccorsi (2005). Measurement error in linear autoregressive models. *J. Amer. Statist. Assoc. 100*(471), 841–852.

Trapletti, A. and K. Hornik (2011). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-25.

(1) Laboratoire MAP5, UMR CNRS 8145Université Paris Descartes, 45, rue des Saints Pères, 75270 Paris Cedex 06, France

(2) Laboratoire Statistique et Génome, Université d'Évry Val d'Essonne, UMR CNRS 8071- USC INRA, Tour Evry II, 523 Place des Terrasses, 91000 Évry, France